

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

FELIPE CORDEIRO ALVES DIAS

**Caracterização de eventos de exceção e de seus respectivos impactos no
sistema de transporte público por ônibus da cidade de São Paulo**

São Paulo

2019

FELIPE CORDEIRO ALVES DIAS

**Caracterização de eventos de exceção e de seus respectivos impactos no
sistema de transporte público por ônibus da cidade de São Paulo**

Versão original

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação.

Área de concentração: Metodologia e Técnicas da Computação

Orientador: Prof. Dr. Daniel de Angelis Cordeiro

São Paulo

2019

Ficha catalográfica

Dissertação de autoria de Felipe Cordeiro Alves Dias, sob o título “**Caracterização de eventos de exceção e de seus respectivos impactos no sistema de transporte público por ônibus da cidade de São Paulo**”, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Metodologia e Técnicas da Computação.

A minha esposa, Laísa Dias, pelo amor, compreensão e companheirismo.

Agradecimentos

Ao professor Daniel Cordeiro por todo o apoio, confiança e dedicação. O presente trabalho faz parte do INCT, do projeto *Future Internet for Smart Cities* com o apoio do CNPq proc. 465446/2014-0, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, FAPESP proc. 14/50937-1 e FAPESP proc. 15/24485-9.

“Confie, mas verifique.”

(provérbio russo)

Resumo

DIAS, Felipe Cordeiro Alves. **Caracterização de eventos de exceção e de seus respectivos impactos no sistema de transporte público por ônibus da cidade de São Paulo.** 2019. 256 f. Dissertação (Mestrado em Ciências) - Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2019.

A cidade de São Paulo é o município mais populoso do Brasil, caracterizado por uma segregação urbana responsável por inúmeros problemas relacionados a mobilidade urbana. As ações atuais para resolver os problemas de mobilidade urbana têm pouco aprofundamento em questões tecnológicas e melhorias dos sistemas computacionais existentes — como as necessárias ao Sistema Integrado de Monitoramento e Transporte (SIM), utilizado para gestão e monitoramento do transporte público por ônibus de São Paulo. Uma das possíveis melhorias é integrar o SIM às Redes Sociais. Com essa perspectiva de integração, esse trabalho tem como objetivo utilizar tweets e dados do SIM na caracterização de eventos de exceção e de seus respectivos impactos no sistema de transporte público por ônibus da cidade de São Paulo. Para alcançar tal objetivo, esse trabalho propõe utilizar tweets publicados por instituições governamentais responsáveis por reportar eventos de exceção, dados dos módulos AVL (*Automatic Vehicle Location*) do SIM, responsáveis por rastrear e localizar os ônibus do município e GTFS da (*General Transit Feed Specification*) da SPTrans. Visando alcançar o objetivo proposto, classificamos manualmente 60.984 tweets e treinamos diferentes modelos por meio de algoritmos de aprendizado de máquina supervisionado para identificar eventos de exceção. Além disso, propomos uma nova metodologia para extrair e geolocalizar os endereços dos eventos de exceção, por meio de Processamento de Linguagem Natural e Expressão Regular. Com isso, demonstramos que é possível correlacionar os dados desses eventos com os dados históricos do SIM e da GTFS, para caracterizar como o transporte público por ônibus da cidade de São Paulo é impactado nesses cenários. Adicionalmente, propomos uma arquitetura distribuída para exploração e visualização de grandes volumes de dados relacionados a transporte público.

Palavras-chaves: Cidades Inteligentes. Transporte Público. Sistemas de Transporte Inteligentes. Eventos de exceção.

Abstract

DIAS, Felipe Cordeiro Alves. **Characterization of exception events and their respective impacts on the public transport system by bus of the city of São Paulo.** 2019. 256 p. Dissertation (Master of Science) - School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2019.

The city of São Paulo is the most populous municipality in Brazil, characterized by an urban segregation responsible for numerous problems related to urban mobility. The current actions to solve the problems of urban mobility have little deepening in technological issues and improvements of existing computer systems — such as those required for the Integrated Monitoring and Transport System (in the Portuguese acronym: SIM), used for the management and monitoring of public transport by buses of the city of São Paulo. One of the possible improvements is integrating the SIM with Social Networks. With this perspective of integration, this work aims to use tweets and data from SIM in the characterization of exception events and their respective impacts on the public transport system by buses of the city of São Paulo. In order to achieve this objective, this work proposes to use tweets published by governmental institutions responsible for reporting exception events, data from SIM' Automatic Vehicle Location (AVL) modules, responsible for the tracking and locating of urban buses and data from SPTrans' GTFS (General Transit Feed Specification). In order to reach the proposed goal, we manually classified 60,984 tweets and trained different models through supervised machine learning algorithms to identify exception events. In addition, we propose a new methodology to extract and geolocalize the addresses of the exception events, through Natural Language Processing and Regular Expression. Using that approaches, we show that it is possible to correlate the data of these events with the historical data of the SIM and GTFS, to characterize how the public transport by bus of the city of São Paulo is impacted in these scenarios. Additionally, we propose a distributed architecture for exploration and visualization of large volumes of data related to public transport.

Keywords: Smart Cities. Public Transportation. Intelligent Transport Systems. Exception events.

Lista de figuras

| | |
|---|-----|
| Figura 1 – Fluxograma do processo do aprendizado supervisionado | 45 |
| Figura 2 – Processo de Filtragem | 59 |
| Figura 3 – Quantidade de artigos publicados por ano | 60 |
| Figura 4 – Porcentagem dos artigos publicados por ano | 61 |
| Figura 5 – Nuvem de palavras das palavras chaves dos artigos selecionados . | 61 |
| Figura 6 – Arquitetura usada no estudo de caso para visualização e exploração dos dados AVL da SPTrans | 87 |
| Figura 7 – Quantidade de dados enviados por dia por ônibus (selecionados aleatoriamente) em janeiro de 2017 | 88 |
| Figura 8 – Distribuição da quantidade de dados enviados por ônibus (selecio- nados aleatoriamente) em janeiro de 2017 | 89 |
| Figura 9 – Localizações enviadas em Janeiro de 2017 de uma linha de ônibus selecionada aleatoriamente | 90 |
| Figura 10 – Localizações dos ônibus referente a movimentação de Janeiro de 2017 | 90 |
| Figura 11 – Metodologia baseada em <i>tweets</i> para encontrar linhas de ônibus impactadas por eventos de exceção na cidade de São Paulo | 92 |
| Figura 12 – Histograma da variação dos tamanhos das sentenças dos <i>tweets</i> existentes no <i>Corpus Twitter</i> | 98 |
| Figura 13 – Distribuição das classes dos eventos de exceção do <i>Corpus Twitter</i> | 99 |
| Figura 14 – Matriz de confusão relacionada a classificação dos <i>tweets</i> em eventos de exceção por meio do algoritmo <i>Multi-layer Perceptron</i> | 100 |
| Figura 15 – Endereços mais impactados por eventos de exceção | 102 |
| Figura 16 – Distribuição dos eventos de exceção na região central de São Paulo | 102 |
| Figura 17 – Distribuição do número de eventos de exceção geolocalizados . . | 104 |
| Figura 18 – Distribuição das classes de eventos de exceção geolocalizados ao longo dos meses do ano de 2017 | 105 |
| Figura 19 – Processo para correlação entre os dados AVL, GTFS e <i>tweets</i> para análise do impacto dos eventos de exceção | 105 |

| | |
|--|-----|
| Figura 20 – Velocidades inesperadas dos ônibus impactados por eventos de exceção a 100 m e 1.000 m dos pontos de parada, ao longo dos meses do ano de 2017 | 114 |
| Figura 21 – Velocidades inesperadas dos ônibus impactados por eventos de exceção a 100 m e 1.000 m dos pontos de rota, ao longo dos meses do ano de 2017 | 114 |
| Figura 22 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo Árvore de Decisão . . . | 205 |
| Figura 23 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo <i>Complement Naive Bayes</i> | 206 |
| Figura 24 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo Florestas Aleatórias . . | 207 |
| Figura 25 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo <i>Multinomial Naive Bayes</i> | 208 |
| Figura 26 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo Regressão Logística . . | 209 |
| Figura 27 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo Máquina de Vetores de Suporte | 210 |

Lista de tabelas

| | |
|---|-----|
| Tabela 1 – Descrição e nome dos perfis selecionados do Twitter | 31 |
| Tabela 2 – Quantidades de artigos coletados e fontes de busca | 59 |
| Tabela 3 – Arquivos e número de registros especificados na GTFS pela SPTrans | 76 |
| Tabela 4 – Detalhamento dos arquivos da GTFS | 77 |
| Tabela 5 – Metadados dos dados AVL da SPTrans | 79 |
| Tabela 6 – Descrição do conjunto de dados AVL | 80 |
| Tabela 7 – Intervalo de tempo e número de <i>tweets</i> coletados | 83 |
| Tabela 8 – Métricas das avaliações dos algoritmos utilizados para classificação dos <i>tweets</i> em eventos de exceção | 99 |
| Tabela 9 – Quantidade de eventos extraídos por classe | 101 |
| Tabela 10 – Linhas de ônibus mais impactadas por eventos de exceção | 103 |
| Tabela 11 – Porcentagem de ônibus dos grupos de linhas afetadas por eventos de exceção, a 1.000 m e 100 m de distância, respectivamente, que tiveram a velocidade mediana reduzida nos meses do ano de 2017 | 108 |
| Tabela 12 – Porcentagem de impacto na velocidade média dos grupos de linhas afetadas por eventos de exceção a 1.000 m e 100 m de distância, respectivamente, nos meses do ano de 2017 | 109 |
| Tabela 13 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans | 113 |
| Tabela 14 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados aos eventos de exceção (a distância de 100 m e 1.000 m, respec- tivamente, dos pontos de parada de ônibus) dos meses do ano de 2017 | 115 |
| Tabela 15 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados aos eventos de exceção (a distância de 100 m e 1.000 m, respec- tivamente, dos pontos de rota dos ônibus) dos meses do ano de 2017 | 116 |
| Tabela 16 – Tabela de logradouros com abreviaturas | 131 |

| | |
|---|-----|
| Tabela 17 – Detalhamento dos campos do arquivo <i>agency.txt</i> da GTFS | 136 |
| Tabela 18 – Detalhamento dos campos do arquivo <i>stops.txt</i> da GTFS | 137 |
| Tabela 19 – Detalhamento dos campos do arquivo <i>routes.txt</i> da GTFS | 142 |
| Tabela 20 – Detalhamento dos campos do arquivo <i>trips.txt</i> da GTFS | 144 |
| Tabela 21 – Detalhamento dos campos do arquivo <i>stop_times.txt</i> da GTFS . . | 147 |
| Tabela 22 – Detalhamento dos campos do arquivo <i>calendar.txt</i> da GTFS . . . | 154 |
| Tabela 23 – Detalhamento dos campos do arquivo <i>calendar_dates.txt</i> da GTFS | 157 |
| Tabela 24 – Detalhamento dos campos do arquivo <i>fare_attributes.txt</i> da GTFS | 158 |
| Tabela 25 – Detalhamento dos campos do arquivo <i>fare_rules.txt</i> da GTFS . . | 159 |
| Tabela 26 – Detalhamento dos campos do arquivo <i>shapes.txt</i> da GTFS | 160 |
| Tabela 27 – Detalhamento dos campos do arquivo <i>frequencies.txt</i> da GTFS . . | 161 |
| Tabela 28 – Detalhamento dos campos do arquivo <i>transfer.txt</i> da GTFS | 164 |
| Tabela 29 – Detalhamento dos campos do arquivo <i>feed_info.txt</i> da GTFS . . . | 166 |
| Tabela 30 – Linhas de ônibus impactadas por eventos de exceção | 169 |
| Tabela 31 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Janeiro | 223 |
| Tabela 32 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Fevereiro | 223 |
| Tabela 33 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Março | 224 |
| Tabela 34 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Abril | 224 |
| Tabela 35 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Maio | 225 |
| Tabela 36 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Junho | 225 |

| | |
|--|-----|
| Tabela 37 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Julho | 226 |
| Tabela 38 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Agosto | 226 |
| Tabela 39 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Setembro | 227 |
| Tabela 40 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Outubro | 227 |
| Tabela 41 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Novembro | 228 |
| Tabela 42 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Dezembro | 228 |
| Tabela 43 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de janeiro de 2017 | 229 |
| Tabela 44 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de fevereiro de 2017 | 230 |
| Tabela 45 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de março de 2017 | 230 |

| | |
|---|-----|
| Tabela 46 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de abril de 2017 | 231 |
| Tabela 47 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de maio de 2017 | 231 |
| Tabela 48 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de junho de 2017 | 232 |
| Tabela 49 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de julho de 2017 | 232 |
| Tabela 50 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de agosto de 2017 | 233 |
| Tabela 51 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de setembro de 2017 | 233 |
| Tabela 52 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de outubro de 2017 | 234 |
| Tabela 53 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de novembro de 2017 | 234 |

| | |
|---|-----|
| Tabela 54 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de dezembro de 2017 | 235 |
| Tabela 55 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de janeiro de 2017 | 236 |
| Tabela 56 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de fevereiro de 2017 | 237 |
| Tabela 57 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de março de 2017 | 237 |
| Tabela 58 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de abril de 2017 | 238 |
| Tabela 59 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de maio de 2017 | 238 |
| Tabela 60 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de junho de 2017 | 239 |
| Tabela 61 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de julho de 2017 | 239 |

| | |
|--|-----|
| Tabela 62 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de agosto de 2017 | 240 |
| Tabela 63 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de setembro de 2017 | 240 |
| Tabela 64 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de outubro de 2017 | 241 |
| Tabela 65 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de novembro de 2017 | 241 |
| Tabela 66 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de dezembro de 2017 | 242 |
| Tabela 67 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de janeiro de 2017 | 243 |
| Tabela 68 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de fevereiro de 2017 | 244 |
| Tabela 69 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de março de 2017 | 244 |

| | |
|---|-----|
| Tabela 70 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de abril de 2017 | 245 |
| Tabela 71 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de maio de 2017 | 245 |
| Tabela 72 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de junho de 2017 | 246 |
| Tabela 73 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de julho de 2017 | 246 |
| Tabela 74 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de agosto de 2017 | 247 |
| Tabela 75 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de setembro de 2017 | 247 |
| Tabela 76 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de outubro de 2017 | 248 |
| Tabela 77 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de novembro de 2017 | 248 |

| | |
|---|-----|
| Tabela 78 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de dezembro de 2017 | 249 |
| Tabela 79 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de rota) aos eventos de exceção do mês de janeiro de 2017 | 250 |
| Tabela 80 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de rota) aos eventos de exceção do mês de fevereiro de 2017 | 251 |
| Tabela 81 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de março de 2017 | 251 |
| Tabela 82 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de abril de 2017 | 252 |
| Tabela 83 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de maio de 2017 | 252 |
| Tabela 84 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de junho de 2017 | 253 |
| Tabela 85 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de julho de 2017 | 253 |

| | |
|---|-----|
| Tabela 86 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de agosto de 2017 | 254 |
| Tabela 87 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de setembro de 2017 | 254 |
| Tabela 88 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de outubro de 2017 | 255 |
| Tabela 89 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de novembro de 2017 | 255 |
| Tabela 90 – Análise <i>Apriori</i> aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de dezembro de 2017 | 256 |

Lista de abreviaturas e siglas

| | |
|-------|--|
| ACC | <i>Accuracy</i> |
| ACM | <i>Association for Computing Machinery</i> |
| ANN | <i>Artificial Neural Networks</i> |
| API | <i>Application Programming Interface</i> |
| APTS | <i>Advanced Public Transportations Systems</i> |
| ATIS | <i>Advanced Travelers Information Systems</i> |
| ATMS | <i>Advanced Traffic Management System</i> |
| AVCS | <i>Advanced Vehicles Control Systems</i> |
| AVL | <i>Automatic Vehicle Location</i> |
| BN | <i>Bayesian Network</i> |
| BP | <i>Back Propagation</i> |
| CCOI | Centro de Controle Integrado 24 Horas da Cidade de São Paulo |
| CE | Centro Expandido |
| CETSP | Companhia de Engenharia de Tráfego de SP |
| CIMU | Central Integrada de Mobilidade Urbana |
| CP | Cinturão Periférico |
| CPTM | Companhia Paulista de Trens Metropolitanos |
| CRF | <i>Conditional Random Field</i> |
| CSV | <i>Comma-separated values</i> |
| CVO | <i>Commercial Vehicles Operation</i> |
| DAG | <i>Directed Acyclic Graph</i> |
| ETL | <i>Extract, Tranform and Load</i> |

| | |
|------------|--|
| GPRS | <i>General Packet Radio Services,</i> |
| GPS | Global Positioning System |
| GTFS | <i>General Transit Feed Specification</i> |
| HDM | <i>Human Driven Method</i> |
| HP | Hipótese de Pesquisa |
| HTTP | <i>Hypertext Transfer Protocol</i> |
| IDF | Inverse Document Frequency |
| IEEE | <i>Institute of Electrical and Electronics Engineers</i> |
| ITS | <i>Intelligent Transport System</i> |
| K-NN | <i>K-Nearest Neighbour</i> |
| LDA | <i>Latent Dirichlet Allocation</i> |
| LISA | <i>Local Indicators of Spatial Association</i> |
| LR | <i>Logistic Regression</i> |
| MLP | <i>Multi-layer Perceptron</i> |
| NB | <i>Naive Bayesian</i> |
| NER | <i>Named Entity Recognition</i> |
| NLP | <i>Natural Language Processing</i> |
| NLTK | Natural Language Toolkit |
| PAC | Programa de Aceleração do Crescimento |
| PcD | Pessoas com Deficiência |
| PlanMob/SP | Plano de Mobilidade Urbana de São Paulo |
| PMESP | Polícia Militar do Estado de São Paulo |
| PPV | <i>Positive Predictive Value</i> |

| | |
|---------|--|
| PTCS | Sistema de Calibração de Trajetórias Privadas |
| QP | Questão de Pesquisa |
| RDBMS | <i>Relational Database Management Systems</i> |
| RL | Regressão Linear |
| RS | Revisão Sistemática |
| RTPI | <i>Real Time Passenger Information</i> |
| SARIMA | <i>Seasonal Autoregressive Integrated Moving Average</i> |
| SBD | <i>Sentence Boundary Disambiguation</i> |
| SC | <i>Smart Cities</i> |
| SIM | Sistema Integrado de Monitoramento e Transporte |
| SMT | Secretaria Municipal de Transportes |
| SPCEDEC | Defesa Civil do Estado de São Paulo |
| SPTrans | São Paulo Transportes |
| SVM | <i>Support Vector Machine</i> |
| TDM | <i>Technology Driven Method</i> |
| TF | <i>Term Frequency</i> |
| TF-IDF | <i>Term Frequency - Inverse Document Frequency</i> |
| TIC | Tecnologias da Informação e Comunicação |
| TPR | <i>True Positive Rate</i> |
| URL | <i>Uniform Resource Locator</i> |
| WSD | <i>Word Sense Disambiguation</i> |

Sumário

| | | |
|----------|--|----|
| 1 | Introdução | 27 |
| 1.1 | <i>Motivação</i> | 27 |
| 1.2 | <i>Definição do problema</i> | 29 |
| 1.3 | <i>Objetivos</i> | 30 |
| 1.4 | <i>Hipóteses</i> | 30 |
| 1.5 | <i>Organização do documento</i> | 32 |
| | | |
| 2 | Fundamentação Teórica | 33 |
| 2.1 | <i>Cidades Inteligentes</i> | 33 |
| 2.2 | <i>Sistemas de Transporte Inteligentes</i> | 35 |
| 2.3 | <i>Conceitos relacionados ao transporte público</i> | 37 |
| 2.3.1 | <i>Acessibilidade</i> | 38 |
| 2.3.2 | <i>Mobilidade</i> | 38 |
| 2.3.3 | <i>Viagem e modais de transporte</i> | 39 |
| 2.4 | <i>Processamento de Linguagem Natural</i> | 40 |
| 2.5 | <i>Feature Engineering</i> | 42 |
| 2.6 | <i>Algoritmos de aprendizado de máquina</i> | 43 |
| 2.6.1 | <i>Algoritmos de aprendizado supervisionado</i> | 44 |
| 2.6.2 | <i>Validação dos modelos de aprendizado supervisionado</i> | 51 |
| 2.7 | <i>Term frequency–Inverse document frequency</i> | 51 |
| 2.8 | <i>Algoritmo Apriori</i> | 52 |
| | | |
| 3 | Revisão Sistemática | 54 |
| 3.1 | <i>Planejamento da Revisão Sistemática</i> | 54 |
| 3.1.1 | <i>Justificativa da Revisão Sistemática</i> | 55 |
| 3.2 | <i>Questões de Pesquisa</i> | 55 |
| 3.3 | <i>Coleta de dados</i> | 58 |
| 3.4 | <i>Avaliação de Dados</i> | 59 |
| 3.5 | <i>Análise e Interpretação</i> | 61 |

| | | |
|----------|---|----|
| 3.5.1 | Tipos de problemas urbanos abordados utilizando o processamento <i>tweets</i> (QP1) | 62 |
| 3.5.2 | Casos de uso relacionados ao transporte público (QP2) | 65 |
| 3.5.3 | Técnicas estatísticas utilizadas no processamento de <i>tweets</i> (QP3) | 67 |
| 3.5.4 | Paradigmas de processamento (QP4) | 69 |
| 3.5.5 | Eventos de exceção relacionados ao transporte público (QP5) . | 69 |
| 3.5.6 | Técnicas de Aprendizado de Máquina utilizadas no processamento de <i>tweets</i> (QP6) | 70 |
| 3.6 | <i>Considerações finais sobre a revisão sistemática</i> | 73 |
| 4 | Dados abertos relacionados ao transporte público e eventos de exceção | 75 |
| 4.1 | <i>Corpus SPTrans</i> | 75 |
| 4.1.1 | Dados da <i>General Transit Feed Specification</i> da SPTrans | 75 |
| 4.1.2 | Dados AVL da SPTrans | 78 |
| 4.1.3 | Identificação de incosistências e indisponibilidade na base de dados AVL da SPTrans | 78 |
| 4.2 | <i>Corpus Twitter</i> | 81 |
| 4.2.1 | Processo de coleta dos <i>tweets</i> | 82 |
| 5 | Exploração e visualização de grandes volumes de dados . . . | 84 |
| 5.1 | <i>Trabalhos relacionados</i> | 84 |
| 5.2 | <i>Druid</i> | 85 |
| 5.2.1 | Real-time nodes | 85 |
| 5.2.2 | Historical nodes | 86 |
| 5.2.3 | Broker nodes | 86 |
| 5.2.4 | Coordinator nodes | 86 |
| 5.3 | <i>Arquitetura utilizada para visualização e exploração dos dados AVL da SPTrans</i> | 86 |
| 5.4 | <i>Estudo de caso com os dados AVL da SPTrans</i> | 87 |
| 5.5 | <i>Consideração sobre a arquitetura utilizada para exploração e visualização dos dados AVL da SPTrans</i> | 91 |

| | | |
|----------|---|-----|
| 6 | Uma metodologia baseada em tweets para encontrar linhas de ônibus impactadas por eventos de exceção na cidade de São Paulo | 92 |
| 6.1 | <i>Pré-processamento</i> | 92 |
| 6.2 | <i>Extração de endereço e geolocalização</i> | 94 |
| 6.3 | <i>Processamento de tweets</i> | 95 |
| 6.4 | <i>Classificação manual do Corpus Twitter</i> | 95 |
| 6.5 | <i>Modelo de classificação de tweets relacionados a eventos de exceção</i> | 96 |
| 6.6 | <i>Encontrando linhas de ônibus afetadas por eventos de exceção</i> | 96 |
| 6.7 | <i>Resultados</i> | 97 |
| 6.8 | <i>Considerações finais sobre a metodologia desenvolvida</i> | 103 |
| 7 | Correlação dos eventos de exceção com os dados AVL da SP-Trans | 104 |
| 8 | Identificação de padrões de velocidade média dos dados AVL | 110 |
| 8.1 | <i>Trabalhos relacionados</i> | 110 |
| 8.2 | <i>Resultados</i> | 111 |
| 9 | Conclusão | 117 |
| 9.1 | <i>Contribuições</i> | 117 |
| 9.2 | <i>Trabalhos publicados</i> | 117 |
| 9.3 | <i>Trabalhos submetidos</i> | 117 |
| 9.4 | <i>Trabalhos futuros</i> | 118 |
| | Referências | 119 |

| | |
|---|------------|
| APÊNDICES | 127 |
| Apêndice A – Exemplos de tweets | 128 |
| Apêndice B – Logradouros utilizados | 131 |
| Apêndice C – Detalhamento dos campos da GTFS | 136 |

| | |
|--|-----|
| Apêndice D – Linhas de ônibus impactadas por eventos de exceção | 169 |
| Apêndice E – Matrizes de confusão | 205 |
| Apêndice F – Parametrizações dos algoritmos | 211 |
| F.1 <i>Árvore de Decisão</i> | 211 |
| F.2 <i>Floresta Aleatória</i> | 213 |
| F.3 <i>K-ésimo Vizinho mais Próximo</i> | 215 |
| F.4 <i>Máquina de Vetores de Suporte</i> | 216 |
| F.5 <i>Naive Bayes</i> | 217 |
| F.6 <i>Redes Neurais</i> | 218 |
| F.7 <i>Régressão Logística</i> | 220 |
| Apêndice G – Análise Apriori | 223 |
| G.1 <i>Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans, referentes aos meses do ano de 2017</i> | 223 |
| G.2 <i>Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada), referentes aos meses do ano de 2017</i> | 229 |
| G.3 <i>Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada), referentes aos meses do ano de 2017</i> | 236 |
| G.4 <i>Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota), referentes aos meses do ano de 2017</i> | 243 |
| G.5 <i>Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de rota), referentes aos meses do ano de 2017</i> | 250 |

1 Introdução

1.1 Motivação

A cidade de São Paulo é o município mais populoso do Brasil, que passou por um rápido processo de urbanização e tem população atual estimada em 12.106.920 milhões de habitantes (com data de referência em 1º de julho de 2017)¹. Desse total de habitantes, 10% vivem na área do Centro Expandido (CE) e 90% no Cinturão Periférico (CP) (SÁ, T. H. et al., 2017), o que caracteriza uma segregação urbana responsável por inúmeros problemas relacionados a mobilidade urbana.

Um desses problemas é conhecido como o movimento pendular, no qual longas distâncias são percorridas diariamente pelos moradores do CP para acessar os locais de emprego, educação e serviços localizados em maioria no CE. Além disso, o movimento pendular torna o CP uma região dormitória, com parte de seus respectivos moradores dependentes do Sistema de Transporte Público para acessar o CE.

Devido aos problemas de mobilidade urbana existentes no Brasil, como os da cidade de São Paulo, a Lei Federal 12.587/2012², relacionada ao Programa de Aceleração do Crescimento³ (PAC), obrigou os municípios a enviarem seus respectivos planos de mobilidade urbana até o final do ano de 2015. O objetivo dessa obrigatoriedade é o de promover o desenvolvimento sustentável com a mitigação dos custos ambientais e socioeconômicos dos deslocamentos de pessoas. Em resposta a essa lei, o Plano de Mobilidade Urbana de São Paulo (*PlanMob/SP 2015*) foi instituído pelo Decreto 56.834⁴, como instrumento de planejamento e gestão do Sistema Municipal de Mobilidade Urbana para os próximos 15 anos.

No *PlanMob/SP 2015*, a Secretaria Municipal de Transportes (SMT) propõe criar uma central de monitoramento conhecida como Central Integrada de Mobilidade Urbana (CIMU), que tem como objetivo integrar as áreas de trânsito e transporte subordinadas à SMT. Nessa proposta, observam-se as outras questões

¹ <https://agenciadenoticias.ibge.gov.br/media/com_mediaibge/arquivos/9bc1a0065c49fd6f81dc785b2b8d8c35.xlsx>. Acessado em 29 de outubro de 2017.

² <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12587.htm>. Acessado em 29 de outubro de 2017.

³ <<http://www.pac.gov.br>>. Acessado em 29 de outubro de 2017.

⁴ <<http://www.prefeitura.sp.gov.br/cidade/secretarias/transportes/planmob>>. Acessado em 29 de outubro de 2017.

que se abordadas trariam benefícios ao CIMU: (I) a CIMU não processa conteúdo de Redes Sociais, (II) não aborda melhoria dos sistemas computacionais já existentes e (III) será integrada com o defasado (CONSULO et al., 2016) Sistema Integrado de Monitoramento e Transporte (SIM), da São Paulo Transportes (SPTrans), responsável pelo monitoramento da infraestrutura de ônibus.

O SIM utiliza a tecnologia *Automatic Vehicle Location* (AVL) para localizar e rastrear os ônibus, fornecer informações em tempo real aos passageiros (*Real Time Passenger Information* (RTPI)), monitorar 1.353 rotas de ônibus⁵, 10 corredores de ônibus⁶, 28 terminais de ônibus⁷ e 19.933 mil paradas de ônibus⁵ que serviram em 2016 a aproximadamente 8 milhões de passageiros por dia⁸. Apesar da importância do SIM, há inúmeras defasagens tecnológicas (que causam discrepância nas informações recebidas pelos usuários, dentre outros problemas) (CONSULO et al., 2016), que precisariam ser resolvidas antes de integrá-lo ao CIMU.

Sistemas como o SIM são classificados como Sistemas de Transporte Inteligente (ITS — *Intelligent Transport System*), e normalmente estão presentes nas Cidades Inteligentes (SC — *Smart Cities*). Por definição, ITS utilizam Tecnologias da Informação e Comunicação (TIC) para explorar dados capazes de contribuir com a melhoria da segurança, do gerenciamento, eficiência dos transportes e redução do impacto ambiental (ANTTIROIKO, 2013). Com isso, nota-se que ITS são essenciais para os objetivos mencionados na Lei Federal 12.587/2012 e no PlanMob/SP 2015.

No entanto, a lei de mobilidade urbana (12.587/2012) e o *PlanMob/SP 2015* não mencionam explicitamente ITS e TIC. O conteúdo de ambos os documentos tem um viés político-urbano, com pouco aprofundamento em questões tecnológicas e melhorias dos sistemas já existentes. Esse cenário é diferente em alguns países, nos quais existem planejamentos para o transporte e mobilidade urbana que estão explicitamente relacionados ao desenvolvimento e uso de novas tecnologias.

Por exemplo, os EUA têm o plano estratégico para 2015-2019 em ITS, abordando temas como veículos conectados, automação, uso de tecnologias emergentes (para apoiar decisões em tempo real), integração de dados corporativos, interoperabilidade (comunicação entre diferentes sistemas) e entrega acelerada de projetos

⁵ <<http://www.sptrans.com.br/desenvolvedores>>. Acessado em 29 de outubro de 2017.

⁶ <<http://www.sptrans.com.br/terminais/corredores.aspx>>. Acessado em 29 de outubro de 2017.

⁷ <<http://www.sptrans.com.br/terminais>>. Acessado em 29 de outubro de 2017.

⁸ <<http://www.sptrans.com.br/indicadores>>. Acessado em 29 de outubro de 2017.

(United States Department of Transportation, 2017). Já a União Européia e o Japão estão centrados em padronizações de tecnologias em ITS, com o objetivo de serem referências nesse setor (CONSULO et al., 2016).

O contraste entre os dois parágrafos anteriores talvez seja devido ao fato de a legislação brasileira e os planos para mobilidade urbana terem sido estabelecidos como consequência do crescimento urbano acelerado e sem planejamento. Ou seja, como solução paliativa para um problema urbano, o que difere dos planos em ITS mencionados, que têm como foco otimizar o transporte e criar padrões tecnológicos.

Apesar dessas diferenças políticas e sociais, o transporte público pode se beneficiar ao explorar ITS (NELSON; MULLEY, 2013) e ao integrar as Redes Sociais com o planejamento, gestão e as atividades operacionais dos transportes públicos, abordando seus respectivos fatores sócio-técnicos (KUFLIK et al., 2017). Por exemplo, um dos benefícios possíveis é o de se conseguir detectar o impacto dos eventos de exceção na operação do sistema de transporte público por ônibus na cidade de São Paulo, usando dados do SIM (AVL), da GTFS (*General Transit Feed Specification*) da SPTrans e de Redes Sociais.

1.2 Definição do problema

Eventos de exceção tais como acidentes, greves, falhas na operação do metrô, manifestações, enchentes, eventos sociais, dentre outros, podem comprometer muitos trechos do sistema de transporte público e, dependendo da proporção do impacto causado pela exceção, inúmeras pessoas podem ser afetadas. Tais eventos de exceção e seus respectivos impactos possuem características que podem ser identificadas visando melhor gestão dessas ocorrências.

Com a identificação dessas características é possível conhecer previamente quais seriam os impactos decorrentes de um determinado evento de exceção no funcionamento normal do transporte público. Tais características podem ser obtidas analisando o histórico do funcionamento do sistema de transportes, e utilizadas posteriormente em simulações de como o sistema responderia a determinados eventos de exceção.

Os dados históricos existentes para essa análise são os do SIM, obtidos utilizando AVL. No entanto, analisá-los envolve problemas como o (I) grande volume

de dados, em virtude da frequência com que são enviados (II) e os referentes ao comprometimento da qualidade dos dados enviados, como consequência dos problemas e limitações do *hardware* responsável pela transmissão; interferências e questões meteorológicas.

O uso de conteúdo de Redes Sociais pode ajudar a abordar os problemas anteriormente mencionados, o qual delimitaria o escopo da análise histórica para a identificação das características dos eventos de exceção e dos seus respectivos impactos. Usar o conteúdo de Redes Sociais envolve alguns desafios como o de (I) identificar eventos de exceção nas publicações, (II) geolocalizá-los, (III) determinar seus *timestamps* e, (IV) correlacioná-las com a base histórica.

1.3 Objetivos

O objetivo geral desse projeto de pesquisa é a caracterização de eventos de exceção e de seus respectivos impactos no sistema de transporte público por ônibus da cidade de São Paulo. Visando alcançar esse objetivo, serão coletados *tweets* das contas oficiais das instituições governamentais responsáveis por reportar eventos de exceção na cidade de São Paulo. Todas as contas selecionadas do *Twitter* estão listadas na Tabela 1. Também, serão utilizados os dados históricos dos módulos AVL do SIM e os dados do sistema de transporte por ônibus da cidade de São Paulo, especificados de acordo com a GTFS.

Além disso, temos como objetivos específicos:

- Identificar os eventos de exceção, quando existentes, dos *tweets* coletados.
- Extrair os endereços dos eventos de exceção identificados e geolocalizá-los.
- Criação de plataforma para exploração e visualização dos dados coletados e processados das fontes citadas na Tabela 1 e da SPTrans.

1.4 Hipóteses

Com base na Revisão Sistemática apresentada no Capítulo 3, os eventos de exceção presentes nos *tweets* podem ser caracterizados, não exaustivamente, em:

Tabela 1 – Descrição e nome dos perfis selecionados do Twitter

| Descrição do perfil no Twitter | Perfil no Twitter |
|---|--------------------------|
| Comando do Corpo de Bombeiros da PMESP ^a | @BombeirosPMESP |
| Companhia de Engenharia de Tráfego de SP | @CETSP_ |
| Companhia Paulista de Trens Metropolitanos | @CPTM_oficial |
| Defesa Civil do Estado de São Paulo | @SPCEDEC |
| Governo do Estado de São Paulo | @governosp |
| Metrô de São Paulo | @metrosp_oficial |
| Polícia Civil do Estado de São Paulo | @Policia_Civil |
| Polícia Militar do Estado de São Paulo | @PMESP |
| São Paulo Agora — CCOI ^b | @saopaulo_agora |
| São Paulo Transporte | @sptrans_ |
| São Paulo Turismo | @TurismoSaoPaulo |
| Secretaria Municipal de Transportes de São Paulo | @smtsp_ |

^a Polícia Militar do Estado de São Paulo (PMESP).

^b Centro de Controle Integrado 24 Horas da Cidade de São Paulo.

Fonte: Elaborado pelo autor

1. Acidentes (ITOH et al., 2016):

- a) acidentes nas estações transporte;
- b) incêndio.

2. Espaço-temporais (CHEN et al., 2016):

- a) dia da semana;
- b) hora do dia.

3. Eventos sociais (CHEN et al., 2016; LECUE et al., 2014; GAL-TZUR et al., 2014; ITOH et al., 2016):

- a) feiras de rua;
- b) festivais;
- c) jogos esportivos;
- d) passeatas e maratonas.

4. Eventos urbanos (CHEN et al., 2016; LECUE et al., 2014):

- a) relacionados ao tráfego.

5. Desastres naturais (ITOH et al., 2016):

- a) tempestades;
- b) terremoto;

c) tufões.

6. Metereológicos (CHEN et al., 2016):

- a) dia claro, nublado, chuvoso, nevando, com neblina;
- b) temperatura do ar.

Dito isso, espera-se que seja possível identificar tais características utilizando Processamento de Linguagem Natural (NLP — *Natural Language Processing*) em conjunto com dicionários auxiliares para o contexto dos eventos de exceção mencionados.

Após a identificação dos eventos de exceção, temos como hipótese que seja possível extrair, com confiabilidade, os endereços dos tweets utilizando a técnica de Expressão Regular. Uma análise preliminar mostra que o conteúdo das contas selecionadas, citadas na Tabela 1, utilizam padrões de formatação para os endereços publicados. Com isso, podemos afirmar que esses tweets apresentam a característica de serem semi-estruturados, diferentemente dos tweets não estruturados publicados pelos usuários comuns do Twitter; o que consequentemente simplifica o processamento necessário para geolocalizar os eventos de exceção.

1.5 Organização do documento

O documento inicia no Capítulo 1 com a introdução de aspectos gerais do trabalho; o Capítulo 2 trata sobre os conceitos fundamentais e necessários para melhor entendimento do conteúdo apresentado por essa dissertação; o Capítulo 3 apresenta a revisão sistemática realizada na literatura com o objetivo de encontrar trabalhos que utilizam tweets para tratar problemas relacionados a Cidades Inteligentes; o Capítulo 4 descreve o processo de coleta dos dados abertos relacionados ao transporte público e a eventos de exceção; os capítulos 5, 6, 7 e 8 discorrem sobre os experimentos realizados para atingir os objetivos já detalhados; o Capítulo 9 apresenta a conclusão dos trabalhos desenvolvidos e os apêndices A, B, C, D, E, F e G finalizam com os detalhes dos experimentos.

2 Fundamentação Teórica

Neste capítulo são apresentados fundamentos teóricos sobre os conceitos Cidades Inteligentes, Sistemas de Transporte Inteligentes, relacionados ao transporte público, Processamento de Linguagem Natural, *Feature Engineering*, Aprendizado de Máquina, a *Term frequency–Inverse document frequency* e ao algoritmo *Apriori*.

2.1 Cidades Inteligentes

Os problemas abordados por este trabalho, definidos na Seção 1.2, estão situados no contexto de Cidades Inteligentes. Embora não haja consenso, nesta dissertação, definimos o conceito de Cidades Inteligentes (SC — *Smart Cities*) como cidades sustentáveis e socialmente inclusivas (WANG; SINNOTT; NEPAL, 2016), que utilizam Tecnologias da Informação e Comunicação (TICs) para gerir eficientemente seus respectivos recursos naturais, de energia, transporte, lixo, dentre outros (AHVENNIEMI et al., 2017). As SC podem ter viés tecnológico (*TDM* — *Technology Driven Method*, top-down; de fornecimento) ou humano (*HDM* — *Human Driven Method*, bottom-up, de demanda) (KUMMITHA; CRUTZEN, 2017).

O aspecto humano das Cidades Inteligentes começou a ser explorado recentemente, após críticas referentes aos poucos indicadores humanos existentes para SC (AHVENNIEMI et al., 2017) (FINGER; RAZAGHI, 2017). A abordagem humana das SC foca questões sociais e qualidade de vida, tais como governança participativa, segurança, cultura, lazer, sustentabilidade, desenvolvimento de capital humano, dentre outras (AHVENNIEMI et al., 2017). Na perspectiva tecnológica de SC, argumenta-se que apenas o uso de TICs seja capaz viabilizar o desenvolvimento de capital humano e de soluções para os problemas da cidade (KUMMITHA; CRUTZEN, 2017).

Independentemente dos vieses humano e tecnológico, a cidade pode ser conceituada como um complexo e dinâmico sistema sócio-técnico. Ou seja, uma cidade (região metropolitana) é composta por sistemas urbanos, com espaços físicos para a vida cotidiana e com sistemas de infraestrutura (para transporte, energia, água e tratamento de água, moradia, telecomunicações e áreas verdes). Os sistemas

urbanos por natureza nunca estão em equilíbrio, possuem subsistemas imprevisíveis (FINGER; RAZAGHI, 2017).

Apesar disso, as TICs permeiam os sistemas urbanos e espaços físicos, o que tem sido acentuado com o crescente número de sensores e dispositivos conectados à Internet (*IoT – Internet of Things*), como os dispositivos móveis que permitem que pessoas enviem dados voluntários e publiquem conteúdo em Redes Sociais sobre os acontecimentos da cidade. Tais fontes heterogêneas geram grandes volumes de dados, utilizados para enriquecer sistemas já existentes e na composição de novos serviços de Cidades Inteligentes (FINGER; RAZAGHI, 2017) (ANG et al., 2017), como os elucidados na revisão sistemática apresentada no Capítulo 3.

O desenvolvimento de serviços de SC envolve desafios relacionados a conectividade (infraestrutura de rede, interoperabilidade e padrões, consumo de energia e escalabilidade) e aos dados (capacidade e local de armazenamento, extração, tratamento, processamento, análise, integração e agregação dos dados) (ANG et al., 2017), (XIAO; LIM; PONNAMBALAM, 2017). Além disso, a análise de dados pode tanger problemas referentes a correlação e inferência de dados de diferentes domínios, aprendizado de máquina, processamento em tempo real e propostas de novo uso para dados provenientes de infraestruturas já existentes (ANG et al., 2017).

Por fim, a seguir estão elencadas algumas frentes de estudo e de desenvolvimento de serviços de SC que ilustram iniciativas em Cidades Inteligentes:

- **Edifícios Inteligentes (*Smart Buildings*)** (TALARI et al., 2017), (MORENO et al., 2017), (ANG et al., 2017), (FINGER; RAZAGHI, 2017), (SANTOS et al., 2017), (KUMMITHA; CRUTZEN, 2017).
- **Comunidades Inteligentes (*Smart Citizen / Community / People*)** (TALARI et al., 2017), (SANTOS et al., 2017), (KUMMITHA; CRUTZEN, 2017), (BARTH et al., 2017), (AHVENNIEMI et al., 2017).
- **Econômias Inteligentes (*Smart Economy*)** (SANTOS et al., 2017), (KUMMITHA; CRUTZEN, 2017), (BARTH et al., 2017), (XIAO; LIM; PONNAMBALAM, 2017), (AHVENNIEMI et al., 2017).
- **Ambientes Inteligentes (*Smart Environment*)** (energia, lixo, água e espaços verdes) (SANTOS et al., 2017), (FINGER; RAZAGHI, 2017), (TALARI et

al., 2017), (ANG et al., 2017), (KUMMITHA; CRUTZEN, 2017), (BARTH et al., 2017), (AHVENNIEMI et al., 2017).

- **Governança Inteligente (*Smart Governance*)** (TALARI et al., 2017), (SANTOS et al., 2017), (KUMMITHA; CRUTZEN, 2017), (BARTH et al., 2017), (AHVENNIEMI et al., 2017).
- **Estilo de vida Inteligente (*Smart Living*)** (*educação, saúde, segurança, cultural*) (SANTOS et al., 2017), (TALARI et al., 2017), (KUMMITHA; CRUTZEN, 2017), (BARTH et al., 2017), (XIAO; LIM; PONNAMBALAM, 2017), (AHVENNIEMI et al., 2017).
- **Transportes Inteligentes (*Smart transportation / mobility*)** (TALARI et al., 2017), (MORENO et al., 2017), (ANG et al., 2017), (FINGER; RAZAGHI, 2017), (SANTOS et al., 2017), (KUMMITHA; CRUTZEN, 2017), (BARTH et al., 2017), (AHVENNIEMI et al., 2017).

2.2 Sistemas de Transporte Inteligentes

Os dados AVL que usamos neste trabalho têm como fonte de origem os inúmeros módulos instalados na frota de ônibus da SPTrans, esses equipamento fazem parte de um sistema de localização automática de veículos, tecnologia conhecida como *Automatic Vehicles Location*, definida mais adiante. A tecnologia AVL pertence a um conjunto de tecnologias desenvolvidas para transportes, conhecidas como Sistemas de Transporte Inteligentes (ITS — *Intelligent Transportation Systems*). Tais tecnologias são uma das mais antigas presentes em Cidades Inteligentes (MENOUAR et al., 2017), que têm como fim utilizar TICs para resolver problemas relacionados ao transporte, tais como congestionamento, segurança, eficiência e conservação ambiental (FIGUEIREDO et al., 2001).

É importante notar que o termo *intelligent*, contido em ITS, normalmente é utilizado em abordagens de cidades inteligentes orientadas à tecnologia, ou seja, que se preocupam principalmente no uso em si de determinada tecnologia, não necessariamente no uso como consequência de uma demanda dos cidadãos. Nos contextos de cidades inteligentes com um viés humano, utiliza-se o termo *smart*, devido a isso *Smart Transportation / Mobility* é conceitualmente diferente de *Intelligent Transportation Systems* (ALBINO; BERARDI; DANGELICO, 2015).

Na ausência de um problema real apontado pela população, é possível que o sistema seja implantado apenas para gerenciamento de frota, o que não tem impacto significativo no cotidiano das pessoas, como menor tempo de viagem. Dessa forma, os estudos realizados neste trabalho utilizam dados de um sistema ITS (no caso, o SIM) para caracterizar os impactos de eventos de exceção no transporte público por ônibus da cidade de São Paulo. Portanto, tais experimentos têm aplicabilidade em cidades inteligentes, mais especificamente em transportes inteligentes *Smart Transportation*. A seguir, detalhamos algumas das categorias de ITS:

1. Sistemas Avançados de Gerenciamento de Tráfego (ATMS — Advanced Traffic Management System) — são sistemas utilizados para melhorar a

qualidade do serviço de tráfego e redução de atrasos (FIGUEIREDO et al., 2001), por meio de:

- a) Equipe de coleta de dados: equipe de pessoas responsáveis por monitorar e coletar dados das condições de tráfego.
- b) Sistemas de suporte: conjunto de câmeras, semáforos, sensores, dentre outros dispositivos auxiliares para gerenciar e controlar o tráfego em tempo real.
- c) Sistemas de controle de tráfego em tempo real: sistemas utilizados para com base nos dados coletados controlar acesso a avenidas, semáforos, envio de mensagens para os dispositivos de monitoramento.

2. Sistemas Avançados de Informações ao Viajante (ATIS — Advanced Travelers Information Systems) — são sistemas utilizados para fornecer informação em tempo real aos viajantes (FIGUEIREDO et al., 2001).

3. Operação de Veículos Comerciais (CVO — Commercial Vehicles Operation) — são sistemas utilizados para a segurança de veículos comerciais e frotas, por meio de tecnologias relacionadas a gerenciamento de tráfego, controle e gerenciamento de veículos e informações aos viajantes (FIGUEIREDO et al., 2001), tais como:

- a) Identificação Automática de Veículos (*Automatic Vehicles Identification*);
- b) Classificação Automática de Veículos (*Automatic Vehicles Classification*);
- c) Automatic Vehicles Location (*Automatic Vehicles Location*);
- d) Detecção de Movimento Pedestre (*Pedestrian Movement Detection*);

- e) Comptadores de Bordo (*Board Computers*);
- f) Transmissões de Tráfego em Tempo Real (*Real Time Traffic Transmissions*).

4. Sistemas Avançados de Transportes Públicos (APTS — *Advanced Public Transportations Systems*) — são sistemas que utilizam ATMS e ATIS para melhorar a eficiência e operação do transporte público coletivo (FIGUEIREDO et al., 2001). É importante observar que APTS também podem utilizar CVO.

5. Sistemas Avançados de Controle de Veículos (AVCS — *Advanced Vehicles Control Systems*) — são sistemas compostos por sensores, computadores e sistemas de controle para auxiliar e alertar motoristas, com o objetivo de melhorar a segurança e reduzir congestionamentos (FIGUEIREDO et al., 2001).

As categorias mencionadas anteriormente representam parte da primeira geração de tecnologias em ITS. A próxima geração, ainda em desenvolvimento, tem como foco veículos autônomos e conectados, capazes de trocarem informações entre si em tempo real para melhorar a segurança dos condutores (MENOUAR et al., 2017). As categorias dessas tecnologias mais recentes não são listadas por estarem fora do escopo desse trabalho, as da primeira geração de ITS foram listadas para melhor contextualização dos dados AVL utilizados nos experimentos apresentados mais adiante.

2.3 Conceitos relacionados ao transporte público

Esta seção define os conceitos relacionados ao transporte público, de acordo com a perspectiva do Plano de Mobilidade Urbana do Município de São Paulo — PlanMob/SP 2015⁴. Tais conceitos são importantes para entendermos a capacidade de impacto dos eventos de exceção (evidenciada nos experimentos desse trabalho) na acessibilidade, mobilidade, nas viagens e nos diferentes modais de transporte, principalmente em cidades com segregação urbana, como a de São Paulo.

Por exemplo, os eventos de exceção podem dificultar ou mesmo restringir a acessibilidade (e acessibilidade universal) a determinadas regiões da cidade, como em caso de inundações, capazes de danificar infraestruturas de acesso ao espaço urbano. Além disso, tais eventos podem afetar a mobilidade quando impactam

no aumento da quantidade de viagens (por meio de diferentes modais ou não) necessárias para chegar a um destino, devido as possíveis mudanças de rotas ocasionadas por esses cenários atípicos.

2.3.1 Acessibilidade

A acessibilidade pode ser considerada como um atributo do espaço urbano, a qual é diretamente proporcional a abrangência e adequação das infraestruturas de acesso ao espaço urbano. As regiões da cidade têm diferentes padrões de infraestrutura de transporte e deslocamento, portanto, são diferenciadas no aspecto de acessibilidade. Além disso, a acessibilidade atua como instrumento de acesso as oportunidades socioeconômicas da cidade. Observa-se que a acessibilidade não é entendida como um atributo econômico relacionado ao valor das tarifas do transporte, ou, as condições de uso (como o congestionamento viário).

Uma qualidade específica do espaço urbano é a acessibilidade universal, a qual o caracteriza como acessível a pessoas com deficiência (PcD). A acessibilidade universal é garantida ao eliminar as barreiras físicas que impedem a participação plena e efetiva das PcD ao espaço urbano.

2.3.2 Mobilidade

A mobilidade pode ser entendida como um atributo do indivíduo, o qual está relacionado a sua capacidade de se deslocar pelo território da cidade e a sua respectiva renda (dimensão econômica); ou seja, pessoas ou famílias de maior renda tendem a ter maior número de viagens. Além disso, observa-se que a restrição da mobilidade devido a má qualidade das infraestruturas urbanas é considerada como falta de acessibilidade ao espaço e não como perda de mobilidade do indivíduo.

A condição de mobilidade pode ser calculada pelo indicador conhecido como taxa ou índice de mobilidade, determinado pelo quociente entre o total de viagens realizadas e o total da população residente em uma região. Tal indicador pode ser especializado de acordo o tipo de mobilidade, por exemplo, ao considerar apenas as viagens motorizadas, obtém-se o índice de mobilidade motorizada; e ser caracterizado como crescente ou decrescente de acordo com fatores socioeconômicos.

Além da mobilidade como atributo do indivíduo, existe a mobilidade como atributo da cidade, conhecida como mobilidade urbana. A mobilidade urbana considera um conjunto de fatores de uma aglomeração urbana que tornam a mobilidade mais qualificada e eficiente, tais como:

1. Transporte público coletivo;
2. transporte de alta capacidade;
3. acessibilidade universal nos passeios e edificações;
4. prioridade ao transporte coletivo no sistema viário;
5. terminais de transporte intermodais;
6. rede de transporte coletivo por ônibus (com acessibilidade universal);
7. rede cicloviária;
8. bicicletários e paraciclos;
9. legibilidade dos sistemas de orientação;
10. comunicação eficaz com os usuários;
11. modicidade tarifária;
12. logística eficiente no transporte de carga, dentre outros itens.

2.3.3 Viagem e modais de transporte

O conceito de viagem no setor de transportes é definido como o deslocamento de uma pessoa entre dois pontos de interesse (origem e destino), com um motivo definido e por meio de um modal de transporte. A saber, os modais de transporte considerados no *PlanMob/SP 2015* estão enumerados a seguir:

1. A pé.
 - a) Independentemente do deslocamento percorrido caso o motivo seja escola ou trabalho;
 - b) superior a 500 metros de deslocamento.
2. Coletivos.
 - a) Metrô;
 - b) ônibus;
 - c) ônibus fretado;

- d) ônibus escolar e lotação;
 - e) trem.
3. Individuais.
- a) Automóveis (bicicleta, carro particular, caminhão, moto e táxi).

2.4 Processamento de Linguagem Natural

O processamento automático de *tweets* para identificação de eventos de exceção envolve o Processamento de Linguagem Natural (NLP — *Natural Language Processing*), que explora como computadores podem ser utilizados para entender e manipular texto ou fala em linguagem natural (LIU; LI; THOMAS, 2017), o que envolve conhecimento interdisciplinar principalmente entre as áreas de ciência da computação, linguística e estatística. A seguir são detalhados alguns dos problemas relacionadas a NLP, divididos em baixo e alto nível (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011):

1. Baixo nível (problemas comuns a NLP) (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011).
 - a) **Desambiguação do limite de sentença (SBD — Sentence Boundary Disambiguation)**: processamento para identificação do início e fim de uma sentença (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011).
 - b) **Tokenização (Tokenization)**: processamento realizado para obtenção das palavras (*tokens*) que compõem uma sentença, inclui a remoção de números, pontuações e caracteres que não pertencem ao alfabeto (SETIAWAN; WIDYANTORO; SURENDRO, 2017).
 - c) **Marcação de parte da fala (part-of-speech tagging)**: processamento para identificação das classificações gramaticais (verbo, sujeito, adjetivo, etc.) das palavras em uma sentença, considerando seus respectivos significados e contexto no qual estão inseridas (ROY; MAJUMDER; NATH, 2017).
 - d) **Decomposição morfológica**: processamento para decomposição morfológica de uma determinada palavra para a sua forma inflexionada, usando *lematização* (lemmatization — identificação do lema da palavra) ou extra-

ção da raiz da palavra usando heurísticas para determinar a localização de sua respectiva flexão (processo de *stemming* — sem tradução direta para o português brasileiro) (SETIAWAN; WIDYANTORO; SURENDRO, 2017), (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011), (KORENIUS et al., 2004).

- e) **Análise superficial da fala (*Shallow parsing (chunking)*)** — conceitos sem tradução direta para o português brasileiro: processamento para identificação de segmentos de uma sentença, tais como frases verbais, nominais, etc., com base nos *tokens* que constituem a *part-of-speech* (COLLOBERT et al., 2011), (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011). O termo superficial se refere a análise superficial de como as classes gramaticais são combinada entre si.
2. Alto nível (aplicação de NLP a problemas específicos, com base nos problema de baixo nível) (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011).
- a) **Identificação e recuperação de erros ortográficos e gramaticais:** processamento iterativo para identificação e correção de erros gramaticais e de digitação. (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011).
 - b) **Reconhecimento de entidade nomeada (NER — *Named Entity Recognition*)**: processamento para identificação e categorização de palavras ou frases específicas (entidades) (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011).
 - c) **Desambiguação do sentido da palavra (WSD — *Word Sense Disambiguation*)**: processamento para identificação do sentido de uma palavra numa sentença (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011).
 - d) **Negação e identificação de incerteza**: processamento para inferir se uma entidade está presente ou não numa sentença, assim como quantificar a quantidade de incerteza da inferência realizada (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011).
 - e) **Extração de relacionamentos**: processamento para identificar relacionamentos entre entidades e eventos (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011).

- f) **Extração de relacionamento / inferência temporal:** processamento para inferência de expressões e relacionamentos temporais (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011).
- g) **Extração de informação:** processamento para extração e transformação para uma forma estruturada de informações específicas a um problema (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011).

Para esta pesquisa, utilizamos o processo de *tokenização* implementado pela classe *TweetTokenizer*¹ da NLTK (*Natural Language Toolkit* — biblioteca utilizada nos experimentos deste trabalho para NLP) para extrair os *tokens* dos *tweets* (*features* utilizadas para treinar os modelos de classificações) e a classe *RSLPStemmer*² para redução do espaço de *features* (por meio de *stemming*), além da remoção de palavras vazias (*stopwords*^{3,4}) — palavras comuns do português brasileiro.

2.5 Feature Engineering

A extração de *tokens* comentada no capítulo anterior envolve o processo de *Feature Engineering*, o qual é iterativo e utiliza o conhecimento do domínio dos dados e de suas métricas para criar (*feature construction*), extrair (*feature extraction*) e selecionar *features* (*feature selection*) para serem utilizadas em algoritmos de aprendizado de máquina. Um conjunto de dados pode ser representado por um número fixo de *features* (variáveis) binárias, categóricas ou contínuas. Antes do processo de *Feature Engineering*, os dados podem ser pré-processados usando técnicas de padronização, normalização, remoção de ruído, redução de dimensionalidade, discretização, expansão, entre outros; é importante notar que informações podem ser perdidas ao realizar essas transformações (GUYON; ELISSEEFF, 2006).

No experimento abordado no Capítulo 6 usamos uma fase de pré-processamento, explicada na Seção 6.1, e um processo para *feature extraction* (explicado adiante) realizado por meio de uma função que utiliza NLP para preparar os *tweets* coletados

¹ <<https://www.nltk.org/api/nltk.tokenize>>. Acessado em 15 de maio de 2018.

² <https://www.nltk.org/_modules/nltk/stem/rslp>. Acessado em 15 de maio de 2018.

³ <http://www.nltk.org/howto/portuguese_en>. Acessado em 15 de maio de 2018.

⁴ Palavras com alta ou baixa frequência no corpus — comuns ou raras — ou removidas por meio de *feature selection* — <http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html>. Acessado em 03 de junho de 2018.

para a tarefa de treinamento. As fases de *feature construction* e *feature selection* não são utilizadas pelos experimentos deste trabalho, porém, são mencionadas para um melhor entendimento.

Sendo assim, na fase de *feature construction*, é realizado um processo para descobrir informações ausentes sobre as relações entre as *features* e para aumentar o espaço de *features*, inferindo ou criando novas *features* com o objetivo de melhorar a precisão dos algoritmos de classificação, entender os dados e obter dados ocultos, etc. (MOTODA; LIU, 2002). Neste estágio, de um conjunto de n *features* A_1, A_2, \dots, A_n , é possível construir *features* adicionais $A_{n+1}, A_{n+2}, \dots, A_{n+m}$, por meio de heurísticas, operadores lógicos, algoritmos, etc. (MOTODA; LIU, 2002).

Por fim, no processo de extração de *features*, usa uma função de mapeamento para extrair um conjunto mínimo de novas *features* com base nas *features* originais e em métricas de desempenho, diferentemente da análise das relações entre *features* na fase de *feature construction* (MOTODA; LIU, 2002). Assim, com um conjunto inicial de n *features* A_1, A_2, \dots, A_n é possível extrair novas *features* $B_1, B_2, \dots, B_m (m < n)$, $B_i = F_i(A_1, A_2, \dots, A_n)$, onde F_i é a função de mapeamento (MOTODA; LIU, 2002). Analogamente, no processamento de *tweets* realizado no Capítulo 6, o espaço de *features* é composto inicialmente por cada palavra extraída do processo de *Tokenization*, o qual posteriormente é reduzido pelas funções responsáveis pelos processos de *stemming* e remoção de *stopwords*.

2.6 Algoritmos de aprendizado de máquina

O processo de classificação automatizada dos *tweets* em eventos de exceção envolve o treinamento de algoritmos de aprendizado de máquina, além dos processos já mencionados de NLP e *feature extraction*. Os algoritmos de Aprendizado de Máquina podem ser (I) supervisionados, nos quais relações com resultados conhecidos são criadas com base nas características de entrada; (II) não-supervisionado, nos quais são conhecidas as características de entrada, mas não os resultados; (III) semi-supervisionados, nos quais podem ser definidas algumas das relações entre dados de entrada e resultados; (IV) por reforço, nos quais são estabelecidas ações com o foco em maximizar determinado ganho.

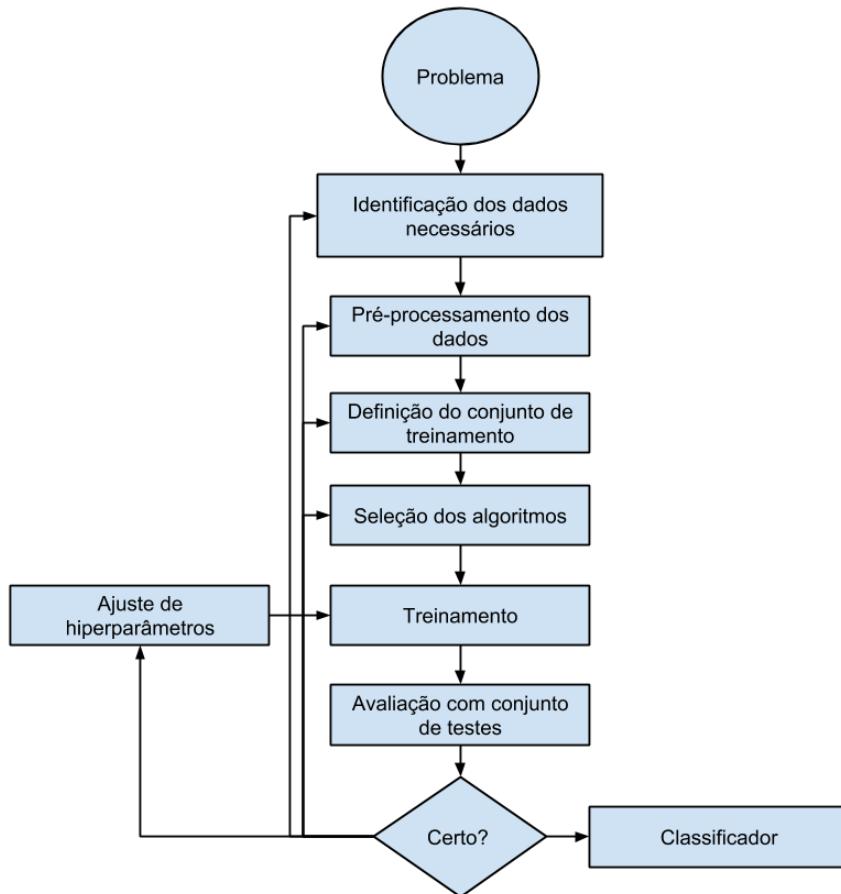
No contexto desse trabalho, os dados de entrada são conhecidos e foram classificados manualmente. Devido a isso usamos aprendizado de máquina supervisionado para o desenvolvimento do modelo de classificação, abordagem a qual também possui melhor desempenho para a tarefa de classificação textual (DWIVEDI; ARYA, 2016). Com base nisso, realizamos uma revisão não sistemática e, de acordo com a literatura, os seguintes algoritmos são os mais utilizados para aprendizado supervisionado (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007; DWIVEDI; ARYA, 2016; NARAYANAN et al., 2017):

- Árvore de Decisão (*Decision Tree*).
- Floresta Aleatória (*Random Forest*).
- K-ésimo Vizinho mais Próximo (K-NN — *K-Nearest Neighbour*).
- Máquina de Vetores de Suporte (SVM — *Support Vector Machine*).
- *Naive Bayes*.
- Redes Neurais (*Neural Networks*).
- Regressão Logística (*Logistic Regression*).

2.6.1 Algoritmos de aprendizado supervisionado

De acordo com a Figura 1, a aplicação de algoritmos de aprendizado supervisionado a um problema passa por algumas fases. As primeiras fases se referem aos processos de identificação dos dados necessários e pré-processamento, descritas respectivamente no Capítulo 4 e na Seção 6.1, as demais fases, explicadas na Seção 6.5, são relacionadas a definição do conjunto de treinamento; seleção dos algoritmos; treinamento; validação com o conjunto de teste e escolha do classificador. É importante observar que não faz parte do escopo deste trabalho afinar os parâmetros dos algoritmos (hiperparâmetros) mencionados na Seção 2.6 (fase *parameter tuning*), devido a isso as parametrizações padrões são utilizadas e descritas no apêndice F. Nas seções seguintes apresentamos os algoritmos de aprendizado supervisionado utilizados no trabalho.

Figura 1 – Fluxograma do processo do aprendizado supervisionado



Fonte: (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007)

Árvore de Decisão

As Árvores de Decisão podem ser utilizadas principalmente para problemas relacionados à classificação de instâncias — quando as variáveis alvo são categóricas — e à regressão — quando as variáveis alvo são contínuas, além disso o algoritmo em si não tem premissas sobre os dados de entrada, ou seja, é não paramétrico. Os nós internos da árvore representam as variáveis de entrada e os nós-folha as classes (variáveis alvo ou de saída) que podem ser utilizadas para classificação. As arestas, por sua vez, determinam as conjunções utilizadas para as ligações entre os diferentes nós, formando assim os caminhos possíveis entre a raiz e os nós-folha (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007).

O uso de uma árvore de decisão envolve o processo de construção de uma árvore de decisão binária ótima, que é conhecido como um problema NP-completo.

Devido a isso, existem inúmeras heurísticas eficientes para construir árvores de decisão quase ótimas, tais como a de ganho de informação, índice de gini, redução de variância, etc (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007).

Algumas das vantagens da árvore de decisão estão relacionadas a fácil interpretação do aprendizado (*white box*), devido ao fato de ser possível visualizar e interpretar a árvore de decisão e bom desempenho com grandes volumes de dados; dentre as desvantagens, estão o alto custo computacional para grandes quantidades de variáveis e possibilidade de sobreajuste quando a árvore atinge sua altura máxima (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007; DWIVEDI; ARYA, 2016).

Floresta Aleatória

Florestas aleatórias ou florestas de decisão aleatórias são um método de aprendizado conjunto (*ensemble*) utilizados principalmente para classificação e regressão. O conceito geral de classificação em conjunto é o de combinar classificadores fracos, por meio de árvores de decisão, para formar um classificador com melhores métricas de desempenho. Duas abordagens comuns para a classificação em conjunto são a de *boosting* e *bagging*, que podem ser implementadas como árvores impulsionadas (*Boosted Trees*) e florestas aleatórias (RF — *Random Forests*), respectivamente (MCDONALD et al., 2014).

O processo que utiliza *boosting* ajusta (*fit*) o algoritmo a todos os dados de entrada, em seguida encontra o conjunto de pontos classificados erroneamente e ajusta outro algoritmo (escolhido por meio de voto ponderado) aos pontos classificados incorretamente. Tal processo é repetido recursivamente com conjuntos de dados menores até que o erro fique abaixo de um determinado limiar. Por sua vez, o processo que utiliza *bagging* ajusta um algoritmo selecionando aleatoriamente do conjunto de dados original vários conjuntos de instâncias para treinamento com *substituição* (um elemento pode aparecer várias vezes na amostra), ajustando em seguida um algoritmo simples (escolhidos por meio de votação majoritária) a cada uma dessas amostras (MCDONALD et al., 2014; DOGRU; SUBASI, 2018).

Em resumo, o algoritmo de florestas aleatórias constroi um conjunto de árvores de decisão e as une para obter previsões mais precisas e estáveis. Dife-

rentemente das árvores de decisão, o algoritmo RF previne sobreajuste por meio da criação aleatória de conjuntos de dados menores, o que implica também em árvores de menor altura. Há também a possibilidade de redução de desempenho, dependendo da quantidade de árvores criadas durante o processo de aprendizado (DOGRU; SUBASI, 2018).

K-ésimo Vizinho mais Próximo

O algoritmo *K-ésimo Vizinho mais Próximo* (*k*-NN — *k*-Nearest Neighbors) é uma abordagem para classificação e regressão não-paramétrica, na qual o processo de aprendizado é caracterizado por encontrar um grupo de *k* amostras que estão mais próximas de amostras desconhecidas, por exemplo, com base em funções de distância. A partir dessas *k* amostras, as classes das amostras desconhecidas são determinadas com base nas classes mais próximas de um conjunto de pontos previamente rotulados (SINGH; THAKUR; SHARMA, 2016; NOI; KAPPAS, 2018).

Devido a característica de determinar os rótulos desconhecidos com base nos *k* mais próximos, o k-NN é considerado um método de aprendizado “preguiçoso” (*lazy learning*), com alto custo computacional durante a fase de classificação e baixo na fase de treinamento. Além disso, a eficiência do algoritmo depende da escolha de um bom valor para o *k*, é afetado por ruídos, variáveis irrelevantes e pelo tamanho do conjunto de dados que precisa ser revisatado (SINGH; THAKUR; SHARMA, 2016; KIBANOV et al., 2018).

Máquina de Vetores de Suporte

O algoritmo *Máquina de Vetores de Suporte* (SVM — *Support Vector Machines*) é uma abordagem de aprendizado para tarefas de classificação e regressão, que funciona em torno do conceito de “margem” — de cada lado de um hiperplano responsável por separar duas classes de dados. Dessa forma, o algoritmo tem como objetivo maximizar a margem entre o hiperplano de separação e as instâncias de ambos os lados (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006; SINGH; THAKUR; SHARMA, 2016).

Ao contrário do k-NN, a precisão e o desempenho do SVM são independentes do tamanho do conjunto de dados, mas dependentes do número de ciclos de treinamento. Sua complexidade não é afetada pelo tamanho do conjunto de dados de treinamento (o número de vetores de suporte selecionados pelo SVM é geralmente pequeno) (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006; SINGH; THAKUR; SHARMA, 2016).

O SVM é muito utilizado em problemas de classificação de texto, que normalmente possuem altos espaços dimensionais e tem boa capacidade de generalização. No entanto, a velocidade de treinamento é menor em relação ao k-NNN e seu desempenho depende dos hiperparâmetros escolhidos (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006; SINGH; THAKUR; SHARMA, 2016).

Dependendo do conjunto de dados, o SVM pode não conseguir localizar um hiperplano de separação devido a instâncias atribuídas incorretamente. O problema pode ser resolvido usando uma margem flexível que aceita algumas classificações erradas das instâncias de treinamento (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006).

Naive Bayes

Uma Rede Bayesiana (BN — *Bayesian Network*) é um modelo gráfico para relações de probabilidade entre um conjunto de variáveis. A estrutura de rede bayesiana S é um grafo acíclico direcionado (DAG — *Directed Acyclic Graph*) e os nós em S possuem uma correspondência um-para-um com o conjunto de variáveis X . As arestas, por sua vez, representam as influências casuais entre as variáveis, quando não existem arestas entre dois nós, não significa que eles sejam completamente independentes, pois podem ser conectados através de outros nós. Tais nós podem, no entanto, tornar-se dependentes ou independentes, dependendo da evidência que é definida em outros nós. Além disso, um nó é condicionalmente independente de seus não descendentes (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006).

Redes Naive Bayesianas (NB — *Naive Bayesian*) são redes bayesianas muito simples que são compostas de DAGs com apenas um pai (representando o nó não observado) e vários filhos (correspondentes a nós observados) com uma forte suposição de independência entre os nós descendentes no contexto de seu pai (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006). A suposição de independência entre

nós descendentes comumente está errada e, por essa razão, os classificadores bayesianos geralmente são menos precisos do que outros algoritmos de aprendizado mais sofisticados (como o de Redes Neurais). Apesar disso, há evidências de que em determinados cenários a abordagem NB possui acurácia melhor do que algoritmos do estado da arte (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006).

A principal vantagem do NB é seu curto tempo computacional para treinamento, além disso, ao contrário das Redes Neurais ou SVM, não há hiperparâmetros a serem definidos, o que o torna mais simples de ser aplicado a uma grande variedade de tarefas. Apesar disso, o NB não é aplicável quando há necessidade de se considerar interações entre as variáveis (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006; SINGH; THAKUR; SHARMA, 2016).

Redes Neurais

O aprendizado por meio de Redes de Neurais depende de três aspectos fundamentais: dados de entrada e função de ativação do neurônio; arquitetura da rede e o peso de cada conexão. Dado que os dois primeiros aspectos são fixos, o comportamento da rede é definido pelos valores dos pesos. A função de ativação mais simples é popularmente conhecida como perceptron (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006).

O conceito de perceptron mapeia um conjunto de entrada de x_1 a x_n para um valor de saída $f(x)$ (0 ou 1 de acordo com um limiar determinado), considerando w_1 a w_n como pesos; abordagem a qual pode ser usada para aprender um approximador de função não linear para classificação ou regressão. Perceptrons somente podem classificar conjuntos de instâncias linearmente separáveis, ou seja, se uma linha reta ou plano puder ser desenhado para separar as instâncias de entrada em suas categorias corretas, as instâncias de entrada serão linearmente separáveis e o perceptron encontrará a solução. Se as instâncias não forem linearmente separáveis, o aprendizado nunca chegará a um ponto em que todas as instâncias sejam classificadas corretamente, nesse contexto, Redes Neurais Artificiais (ANN — *Artificial Neural Networks*) as foram criadas para tentar resolver esse problema (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006; SINGH; THAKUR; SHARMA, 2016).

Dessa forma, os perceptrons podem ser utilizados para formar uma rede neural com multicamadas (MLP — *Multi-layer Perceptron*), que consiste em um grande número de unidades (neurônios) unidos em um padrão de conexões. Unidades nessa rede são geralmente segregadas em três classes: unidades de entrada, que recebem informações a serem processadas; unidades de saída, onde os resultados do processamento são encontrados; e unidades centrais conhecidas como unidades ocultas. *Feed-forward ANN*, como a MLP, permitem que os sinais percorram somente um caminho, da entrada à saída (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006).

Geralmente, determinar corretamente o tamanho da camada oculta é um problema porque uma subestimativa do número de neurônios pode levar a capacidades de aproximação e generalização ruins, enquanto nós excessivos podem resultar em superajuste e eventualmente tornar a busca pelo ótimo global mais difícil (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006).

A ANN depende de três aspectos fundamentais, funções de entrada e ativação da unidade, arquitetura de rede e o peso de cada conexão de entrada. Dentre os inúmeros algoritmos com os quais uma rede pode ser treinada, o algoritmo de aprendizado mais conhecido e amplamente utilizado para estimar os valores dos pesos é o algoritmo (BP — *Back Propagation*). No entanto, o BP tende a ser mais lento de treinar do que outros, o que pode ser problemático em redes muito grandes e com uma alta quantidade de dados. Além disso, outra desvantagem das ANN é o fato de ser difícil entender o aprendizado obtido pela rede (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006; SINGH; THAKUR; SHARMA, 2016).

Régressão Logística

A Régressão Logística (LR — *Logistic Regression*) é um método estatístico no qual uma curva logística é ajustada ao conjunto de dados, com o objetivo de predizer presença ou ausência de determinada característica. A LR é semelhante a um modelo de regressão linear, mas é mais adequada para modelos em que a variável dependente é dicotômica, apesar disso, essa metodologia também pode ser utilizada para previsões de múltiplas classes (SCHEIN; UNGAR, 2007; KURT; TURE; KURUM, 2008; SINGH; THAKUR; SHARMA, 2016).

Como a LR retorna a probabilidade de uma variável pertencer a determinada classe, os limites de classificação podem ser facilmente ajustados, no entanto, requer um tamanho de amostra grande para alcançar resultados estáveis, além de não lidar adequadamente com problemas não-lineares (KHEMPHILA; BOONJING, 2010; SINGH; THAKUR; SHARMA, 2016).

2.6.2 Validação dos modelos de aprendizado supervisionado

A validação dos modelos para tarefas de classificação pode ser realizada por meio de *validação cruzada*⁵ (nos experimentos desse trabalho utilizamos 10 *folds* — subconjuntos do conjunto de dados de treinamento — para validar a generalização dos modelos) e métricas tais como: *acurácia* (ACC — Accuracy, Eq. 1), *precisão PPV* — *Positive Predictive Value*, Eq. 2), *revocação* (TPR — *True Positive Rate*, Eq. 3) e *f₁score* (Eq. 4), que tem como entrada o número de casos reais positivos (P), negativos (N), verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN):

$$ACC = \frac{VP + VN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$PPV = \frac{VP}{VP + FP} \quad (2)$$

$$TPR = \frac{VP}{P} = \frac{VP}{VP + FN} \quad (3)$$

$$f_1 score = \frac{PPV * TPR}{PPV + TPR} = \frac{2VP}{2VP + FP + FN} \quad (4)$$

2.7 Term frequency–Inverse document frequency

Além dos processos de NLP para redução do espaço de *features* mencionados anteriormente, podemos utilizar abordagens, como a TF-IDF, que levam em consideração a frequência dos termos (*tokens*) existentes em um conjunto de documentos. TF-IDF é um algoritmo de ponderação de variáveis que combina as ponderações

⁵ <https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation>

frequência do termo (TF — *Term Frequency*) e inverso da frequência nos documentos (IDF — *Inverse Document Frequency*) para calcular os pesos dos termos linguísticos (variáveis) em um determinado corpus. Em outras palavras, o peso da variável é proporcional a frequência com a qual aparece nos documentos, e inversamente proporcional a quantidade de documentos que contém o termo linguístico em questão (WU; YUAN, 2018; YAHAV; SHEHORY; SCHWARTZ, 2018).

Dentre as variações de implementação da ponderação $W_{t,d}$ (TF-IDF) existentes, a abordagem tradicional considera uma coleção de termos $t \in T$ que aparecem em um conjunto de N documentos $d \in D$, posto isso, defini-se como o produto entre $tf_{i,j}$ e idf_i — onde $n_{i,j}$ é a frequência do termo t_i no documento d_j , $\sum_k n_{k,j}$ o somatório da frequência de todos os termos do documento d_j e n o número de documentos onde t_i aparece ($n + 1$, caso $n = 0$) — conforme a s seguinte equações (WU; YUAN, 2018):

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (5)$$

$$idf_i = \log \frac{N}{n + 1} \quad (6)$$

$$W_{t,d} = tf_{t,d} * idf_t \quad (7)$$

No contexto deste trabalho, entendemos documentos como as classes dos eventos de exceção. A *frequência dos termos* (TF — $tf_{t,d}$) é determinada por classe e a *frequência do termo - inverso da frequência nos documentos* (IDF — idf_t) como o inverso dos eventos de exceção, sendo N o tamanho do conjunto dos eventos de exceção, sob o qual df_t é definido. Os eventos de exceção são classificados em suas respectivas classes por meio dos modelos de aprendizado supervisionado, elencados na Seção 2.6.1.

2.8 Algoritmo Apriori

O algoritmo *Apriori*⁶ normalmente é utilizado em mineração de texto para identificar relações entre conjuntos de itens e padrões, por meio de comparações

⁶ Utilizamos para este trabalho a implementação do algoritmo *Apriori* feita pela biblioteca *Apyori* <<https://pypi.org/project/apyori>>. Acessado em 08 de janeiro de 2019

de conjuntos de itens frequentes, para assim determinar regras de associação com base em métricas, tais como a de *suporte (support)* — indicador da frequência de determinados registros no conjunto de dados; *confiância (confidence)* — frequência com que determinadas regras de associações entre registros são encontradas encontradas como verdadeiras e *lift* — probabilidade de ocorrência de um consequente B no conjunto de dados (*lift* > 1 indica que a regra de associação em questão pode ser utilizada para predição de um consequente B em conjuntos de dados futuros). Todas as métricas mencionadas anteriormente são detalhadas nas equações 8, 9, 10, 11 e 12 a seguir (PARK et al., 2018):

$$\text{support}(A) = \frac{\sum_{i=1}^n [A \in s_i]}{n} = P(A) \quad (8)$$

$$\text{support}(B) = \frac{\sum_{i=1}^n [B \in s_i]}{n} = P(B) \quad (9)$$

$$\text{support}(A \rightarrow B) = \frac{\sum_{i=1}^n [A \in s_i \wedge B \in s_i]}{n} = P(A \cap B) \quad (10)$$

$$\text{confidence}(A \rightarrow B) = \frac{\text{support}(A \rightarrow B)}{\text{support}(A)} = \frac{P(A \cap B)}{P(A)} = P(B|A) \quad (11)$$

$$\text{lift}(A \rightarrow B) = \frac{\text{confidence}(A \rightarrow B)}{\text{support}(B)} = \frac{P(B|A)}{P(B)} \quad (12)$$

Algoritmos para mineração de dados, como o *Apriori*, não tem tido seu potencial explorado no domínio de grandes volumes de dados relacionados ao transporte (PARK et al., 2018). Neste trabalho, aplicamos o algoritmo *Apriori* no conjunto de dados da *SPTrans* para identificarmos as regras de associação existentes, detalhadas no Capítulo 8, com o objetivo de contribuirmos para a gestão do transporte público por ônibus da cidade de São Paulo.

3 Revisão Sistemática

Conforme mencionado na Seção 1.3 o objetivo geral desse projeto de pesquisa é a caracterização de eventos de exceção e de seus respectivos impactos no sistema de transporte público por ônibus da cidade de São Paulo, por meio do cruzamento de dados AVL, da GTFS e tweets das contas oficiais responsáveis por reportar esse tipo de evento. Devido a isso este capítulo apresenta uma Revisão Sistemática (RS) com o objetivo de encontrar o estado da arte de trabalhos que visam melhorar sistemas de transporte público por meio do processamento de tweets em conjunto com outras fontes de dados.

Além disso, de uma forma mais ampla, busca-se também entender como os tweets têm sido utilizados na caracterização de problemas urbanos. Sendo assim, o capítulo é iniciado com a seção sobre o planejamento da Revisão Sistemática; seguida das questões de pesquisa utilizadas na formulação do problema da RS; do processo de coleta dos estudos primários; da avaliação dos dados coletados; da análise e interpretação dos estudos selecionados, concluindo com as considerações finais.

3.1 Planejamento da Revisão Sistemática

A presente Revisão Sistemática utiliza a metodologia proposta por BIOLCHINI et al. (2005), composta por cinco etapas. A primeira etapa está relacionada à formulação do problema, na qual é levantada uma questão central se referindo ao tipo de evidência que deverá estar contida na revisão. Em seguida, são construídas definições que permitem estabelecer uma distinção entre os estudos relevantes e irrelevantes para o propósito específico do que se está investigando (BIOLCHINI et al., 2005).

A segunda etapa da condução está relacionada à Coleta de Dados, na qual são definidos os procedimentos que serão utilizados para encontrar a evidência relevante que foi definida na etapa anterior. Nesta fase é extremamente importante determinar as fontes que podem fornecer estudos relevantes a serem incluídos na pesquisa (BIOLCHINI et al., 2005).

Na terceira etapa a Avaliação de Dados é definida, na qual são selecionadas as fontes primárias que deverão ser incluídas na revisão. Em seguida, são aplicados os critérios de qualidade para separar estudos que podem ser considerados válidos, e determinadas as diretrizes para o tipo de informação que deve ser extraída dos relatórios de pesquisas primárias (BIOLCHINI et al., 2005).

A quarta etapa da revisão é o processo de Análise e Interpretação, na qual os dados dos estudos primários válidos são sintetizados. E, na quinta etapa são realizados os processos de Conclusão e Apresentação (BIOLCHINI et al., 2005).

3.1.1 Justificativa da Revisão Sistemática

Esta Revisão Sistemática se justifica por não terem sido encontradas revisões sistemáticas com o foco em questões urbanas e de transporte público, abordando unicamente o processamento de *tweets*. Em (CHANIOTAKIS; ANTONIOU; PEREIRA, 2016), por exemplo, foi realizado um mapeamento de forma não sistemática dos trabalhos sobre o uso das mídias sociais em problemas relacionados ao transporte público; (STEIGER; ALBUQUERQUE; ZIPF, 2015), por outro lado, desenvolveram uma revisão sistemática sobre o uso do Twitter para questões espaço-temporais; e (JUNGHERR, 2016) no contexto político.

Devido a isso, a presente revisão sistemática se diferencia por ter como objetivo encontrar o estado da arte de trabalhos que visam melhorar sistemas de transporte público por meio do processamento de *tweets*, cruzando-os com outras fontes de dados. Além disso, de uma forma mais ampla, busca-se também entender como os *tweets* têm sido utilizados na caracterização de problemas urbanos.

3.2 Questões de Pesquisa

Nesta seção, são apresentadas as questões de pesquisa utilizadas para a formulação dos problemas abordados por essa Revisão Sistemática. Por meio das quais, busca-se atender os objetivos já mencionados na Seção 3.1.1.

1. Quais os tipos de problemas urbanos abordados utilizando processamentos de *tweets*?

O propósito da QP1 é identificar quais são as contribuições do processamento de *tweets* para a mitigação de problemas urbanos. A resposta a essa questão de pesquisa ajudará especialistas das áreas multidisciplinares relacionadas ao Urbanismo (como a de Análise de Redes Sociais e Políticas Públicas) a terem um panorama de como *tweets* podem ser utilizados para ajudar na solução de problemas urbanos.

Uma análise preliminar dos estudos primários permite elaborar a seguinte Hipótese de Pesquisa (HP1): alguns dos problemas urbanos abordados estão relacionados ao transporte, mobilidade urbana, turismo e desastres naturais.

2. Como *tweets* têm sido utilizados para abordar problemas relacionados ao transporte público?

O propósito da QP2 é identificar se *tweets* têm sido utilizados para solucionar problemas relacionados ao transporte público. A resposta a essa questão de pesquisa ajudará especialistas das áreas multidisciplinares relacionadas ao Urbanismo (como a de Análise de Redes Sociais e Políticas Públicas) a terem um panorama de como *tweets* podem ser utilizados para ajudar na solução de problemas referentes a mobilidade urbana.

Uma análise preliminar dos estudos primários permite elaborar a seguinte Hipótese de Pesquisa (HP2): *tweets* têm sido utilizados principalmente para questões relacionadas ao congestionamento, não tendo como foco o transporte público.

3. Quais as técnicas estatísticas utilizadas no processamento de *tweets*?

O propósito da QP3 é identificar quais as técnicas estatísticas utilizadas no processamento de *tweets*, principalmente no que se refere a validação do processo de aprendizado. A resposta a essa questão de pesquisa ajudará especialistas a terem um panorama de como validar tarefas de aprendizado que utilizam

dados oriundos de *tweets*.

Uma análise preliminar dos estudos primários permite elaborar a seguinte Hipótese de Pesquisa (HP3): F_1 score é a principal técnica utilizada para validação de classificação binária.

4. Quais os paradigmas de processamento têm sido utilizados ao lidar com *tweets*?

O propósito da QP4 é identificar os paradigmas utilizados para processamento de *tweets*. A resposta a essa questão de pesquisa ajudará especialistas a terem um panorama das técnicas de processamento utilizadas na análise de *tweets*.

Uma análise preliminar dos estudos primários permite elaborar a seguinte Hipótese de Pesquisa (HP4): o principal paradigma utilizado tem sido o processamento de *tweets* em lote (*batch — offline*), após um processo de armazenamento. Poucos são os estudos que constroem uma plataforma para processamento de dados em tempo real.

5. Quais são os eventos de exceção relacionados ao transporte público?

O propósito da QP5 é identificar os eventos de exceção relacionados ao transporte público. A resposta a essa questão de pesquisa ajudará especialistas no levantamento de eventos de exceção relacionados ao transporte público, os quais podem ser utilizados em algoritmos de classificação.

Uma análise preliminar dos estudos primários permite elaborar a seguinte Hipótese de Pesquisa (HP5): há poucos ou nenhum estudo que, ao tratar de problemáticas relacionadas ao transporte público, realizam um levantamento dos eventos de exceção desse contexto.

6. Quais as técnicas de aprendizado de máquina utilizadas no processamento de *tweets*?

O propósito da QP6 é identificar as técnicas de aprendizado de máquina utilizadas no processamento de *tweets*. A resposta a essa questão de pesquisa ajudará especialistas a terem um panorama das principais técnicas de aprendizado de máquina utilizadas no processamento de *tweets*.

Uma análise preliminar dos estudos primários permite elaborar a seguinte Hipótese de Pesquisa (HP6): a técnica *Support Vector Machine* tem sido utilizada na maioria dos estudos que aplicam aos *tweets* algum algoritmo de aprendizado de máquina.

3.3 Coleta de dados

Nesta Revisão Sistemática, os artigos foram coletados em quatro fontes de pesquisa, por meio da plataforma de indexação de trabalhos acadêmicos *Google Scholar*¹. Constam na Tabela 2 as bases pesquisadas no ano de 2017, quantidades de artigos coletados, descartados no processo de filtragem (Figura 2, descrito na Seção 3.4) e selecionados. Com base na QP1, a seguinte *string* de busca foi construída; restrita aos trabalhos publicados entre 2011 e 2016, escritos no idioma Inglês (devido ao fato das publicações relevantes, na área de Computação, estarem disponíveis nesse idioma):

String de busca: twitter urban planning city (analytics OR patterns OR tweets OR social OR media) AND (public transport)

Palavras-chave: twitter, urban, planning, city, analytics, patterns, tweets, social, media e public transport.

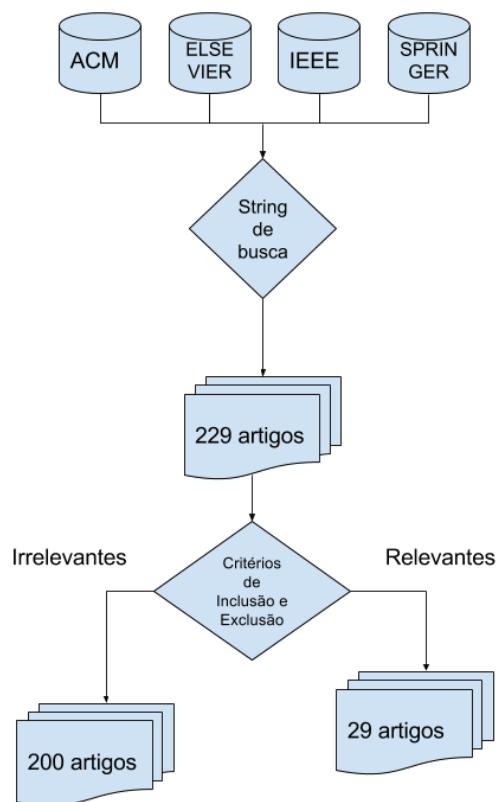
¹ <<https://scholar.google.com>>. Acessado em 29 de outubro de 2017.

Tabela 2 – Quantidades de artigos coletados e fontes de busca

| Fonte | Artigos coletados | Filtragem | Selecionados |
|--------------|--------------------------|------------------|---------------------|
| ACM | 44 | 34 | 10 |
| IEEE | 82 | 74 | 8 |
| Elsevier | 81 | 72 | 9 |
| Springer | 22 | 20 | 2 |
| Total | 229 | 200 | 29 |

Fonte: Elaborado pelo autor

Figura 2 – Processo de Filtragem



Fonte: Elaborado pelo autor

3.4 Avaliação de Dados

Visando selecionar os artigos relevantes para esta Revisão Sistemática, os seguintes critérios foram utilizados no processo de filtragem:

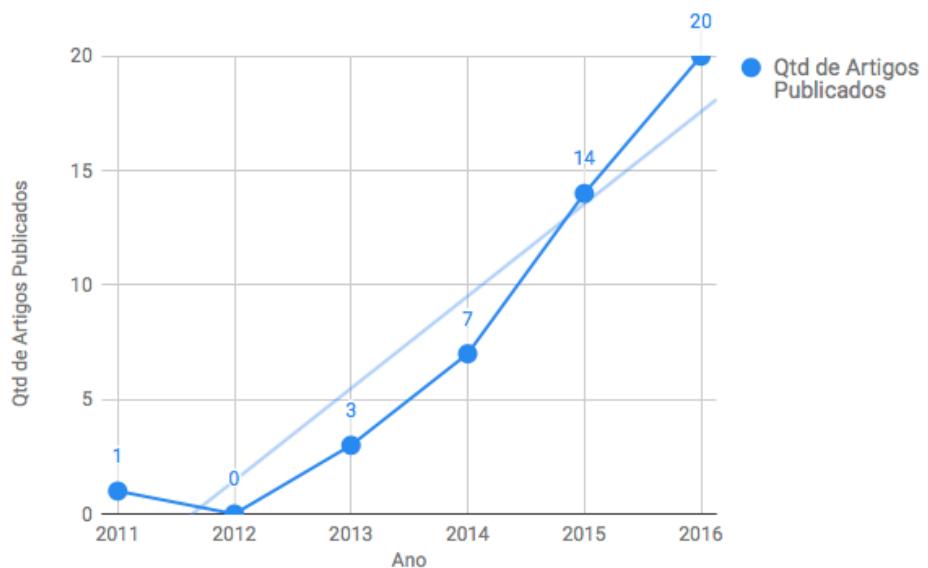
- Trabalho publicado (critério de qualidade).
- Trabalhos que utilizam tweets para abordar questões urbanas e de transporte público.

- Trabalhos duplicados.
- Trabalhos que estão fora do escopo da questão de pesquisa.

O processo de condução da Revisão Sistemática foi realizado utilizando os critérios acima mencionados, e está disponível em DIAS (2017) (não incluso no apêndice deste trabalho devido ao grande volume de páginas de cada documento), assim como seu respectivo protocolo (no qual contém o detalhamento dos critérios de inclusão e exclusão, dentre outros artefatos da condução). Após o processo de condução, alguns dos metadados dos artigos selecionados foram sintetizados.

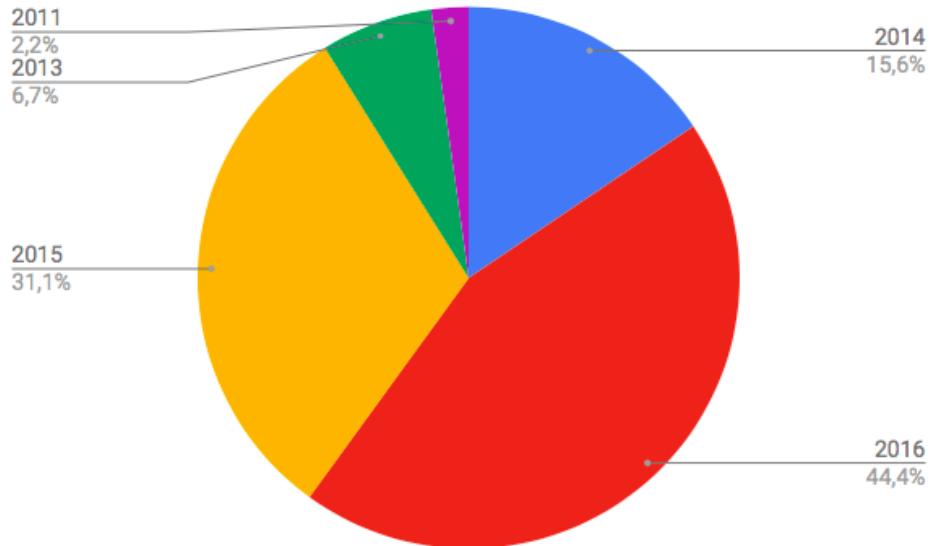
Sendo assim, a Figura 5 apresenta uma nuvem de *tags* (gerada com a biblioteca *wordcloud* (Andreas Mueller, 2018)) sintetizando as palavras chaves dos estudos primários selecionados; e a Figura 3 a quantidade de artigos publicados por ano, sendo possível analisar por meio dela a distribuição dos artigos entre 2011 e 2016, assim como sua respectiva porcentagem, ilustrada na Figura 4.

Figura 3 – Quantidade de artigos publicados por ano



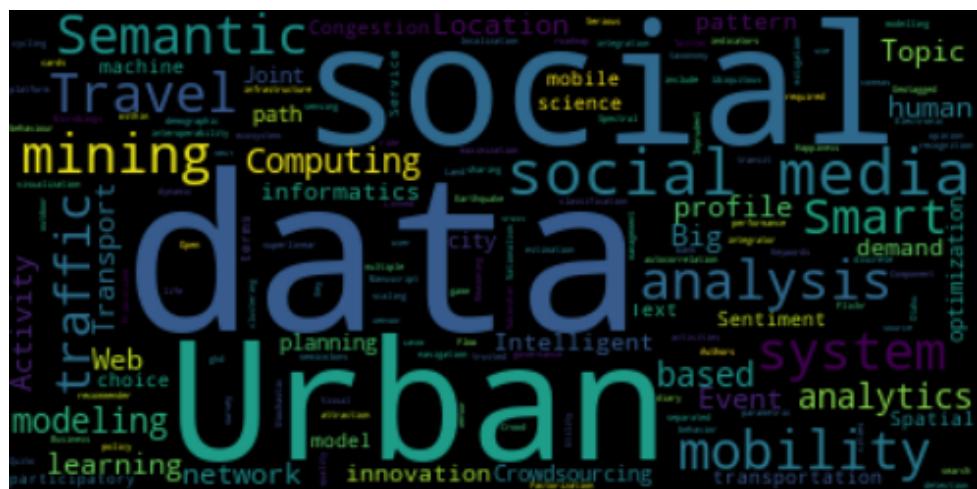
Fonte: Elaborado pelo autor

Figura 4 – Porcentagem dos artigos publicados por ano



Fonte: Elaborado pelo autor

Figura 5 – Nuvem de palavras das palavras chaves dos artigos selecionados



Fonte: Elaborado pelo autor

3.5 Análise e Interpretação

Nesta seção é realizada a análise e interpretação dos estudos primários selecionados pela Revisão Sistemática, sendo as seções divididas de acordo com as questões de pesquisa.

3.5.1 Tipos de problemas urbanos abordados utilizando o processamento *tweets* (QP1)

Os tipos de problemas urbanos abordados utilizando o processamento de *tweets* foram divididos nas seguintes categorias:

1. **e-Participation** (interação entre cidadãos e órgãos civis) (MUKHERJEE et al., 2015; SOOMRO; KHAN; HASHAM, 2016);
2. **detecção de zoneamento urbano** (FRIAS-MARTINEZ; FRIAS-MARTINEZ, 2014);
3. **identificação de pontos de interesse** (FARSEEV et al., 2015; GUTEV; NENKO, 2016; BENDLER et al., 2014; ABBASI et al., 2015; GKIOTSALITIS; STATHOPOULOS, 2015; GKIOTSALITIS; STATHOPOULOS, 2016; HASAN; UKKUSURI, 2014; MAGHREBI et al., 2015; DI LORENZO et al., 2013);
4. **mobilidade** (GUTEV; NENKO, 2016; CHEN et al., 2016; YOUSAF et al., 2014);
5. **padrões demográficos** (FARSEEV et al., 2015; GUTEV; NENKO, 2016; STEIGER et al., 2015; GUO et al., 2016);
6. **poluição** (ZAGAL; MATA; CLARAMUNT, 2016);
7. **segurança pública** (WEN; LIN; PELECHRINIS, 2016; MATA; CLARAMUNT, 2015);
8. **turismo** (THOMAZ et al., 2016; ABBASI et al., 2015; CHUA et al., 2016; SOBOLEVSKY et al., 2015);
9. **tráfego** (ANANTHARAM et al., 2015; LECUE et al., 2014).

Conforme os estudos primários analisados pela Revisão Sistemática, e enumerados nessa seção, é possível interpretar que *tweets* podem ser utilizados para auxiliar na mitigação de inúmeros problemas urbanos. Apesar disso, em (CHANIO-TAKIS; ANTONIOU, 2015) os autores observam que os *tweets* contendo informações sobre geolocalização são normalmente publicados em áreas relacionadas ao lazer, além de haver correlação entre regiões urbanas com maior renda *per capita* e o número de *tweets* postados. Tal evidência pode conduzir viés nas análises por representar somente algumas classes econômicas da população.

Considerando a observação anterior, um dos estudos analisados foi o realizado por (ZAGAL; MATA; CLARAMUNT, 2016), na Cidade do México. Nesse estudo, foram

mapeados os pontos da cidade referenciados em publicações relacionadas a doenças respiratórias e poluição, orientando tomadas de decisão no aspecto ambiental.

Além disso, há também exemplos de trabalhos relacionados a Segurança Pública, como o estudo de caso realizado por (WEN; LIN; PELECHRINIS, 2016), no qual foi enriquecido um conjunto de dados com *tweets* geolocalizados, visando analisar o impacto dos ataques terroristas (em Paris, em novembro de 2015) nos padrões de atividades urbanas (relacionadas ao uso de transporte público, serviços, realização de compras, e atividade noturna). Em um outro caso de aplicação, estimou-se por meio de *tweets*, a probabilidade de ocorrência de crimes e ameaças nas ruas da Cidade do México, sugerindo rotas seguras aos pedestres (MATA; CLARAMUNT, 2015).

Também, foram encontrados na literatura estudos que utilizaram *tweets* para inferir padrões demográficos. Por exemplo, em (FARSEEV et al., 2015; GKIOTSALITIS; STATHOPOULOS, 2015; GKIOTSALITIS; STATHOPOULOS, 2016), os autores processaram *tweets* para analisar a distribuição etária e de gênero da população, assim como seus respectivos pontos de interesse (como locais para entretenimento, residência, trabalho, recriação, compras, educação e serviços sociais) (HASAN; UKKUSURI, 2014; MAGHREBI et al., 2015).

Tais pontos de interesse podem ser utilizados em problemas relacionados ao transporte público (GUTEV; NENKO, 2016) e também ao turismo, como no estudo realizado por (ABBASI et al., 2015) para identificar a locomoção de visitantes e residentes em pontos turísticos de Sydney; por (CHUA et al., 2016), ao caracterizar aspectos espaciais, temporais e demográficos, dos turistas da cidade de Cilento, Itália; e por (THOMAZ et al., 2016) na cidade de Curitiba (Brasil), no contexto da Copa do Mundo de 2014.

Nesse mesmo contexto, (GUO et al., 2016) estudaram algumas questões demográficas por meio de análise de sentimento e encontraram correlação positiva entre oportunidades de emprego e sentimentos positivos, e negativa entre felicidade e número de crianças na população da Grande Londres. Outro caso de uso, foi o desenvolvido em (STEIGER et al., 2015), no qual os autores usaram *tweets* para identificar diferentes tipos de atividades em Londres, correlacionando-as com informações censitárias; e em (SOBOLEVSKY et al., 2015) ao estudar a atratividade da Espanha a turistas.

Um dos problemas relacionados à identificação de pontos de interesse se refere as incertezas espaço-temporais e de determinação de tópicos, o qual foi abordado pelo trabalho realizado por (BENDLER et al., 2014). Nele, os autores contribuíram com uma técnica para minimizar o problema ao processar *tweets*; analisando a causalidade entre o tempo e local das postagens realizadas, reduzindo assim os índices de incerteza, no contexto da cidade de São Francisco, EUA. Outro problema, relaciona-se com a questão da privacidade, pois as localizações dos usuários podem ser inferidas mesmo quando não disponibilizadas. Nesse cenário, (WANG; SINNOTT; NEPAL, 2016) propõem um Sistema de Calibração de Trajetórias Privadas (PTCS), por meio de mecanismos de Privacidade Diferencial e de *k-anonymity*, com isso é possível extrair informações sobre trajetórias sem exposição de informações sensíveis, testado na extração de localizações contidas em *tweets*.

Outro contexto na literatura revisada está relacionado ao processamento dos eventos que acontecem na cidade (idealmente em tempo real, como sugerem (SOMRO; KHAN; HASHAM, 2016)). Um dos estudos encontrados sobre esse assunto, foi o realizado por (ANANTHARAM et al., 2015), no qual os autores desenvolveram uma técnica para identificar os diferentes tipos de eventos do cotidiano urbano, rotulando-os sequencialmente, por meio da anotação de *tweets* e extração de eventos, considerando aspectos espaciais, temporais e temáticos. Para isso, utilizaram-se dos conhecimentos de domínio, tais como informações sobre os locais em uma cidade e possíveis termos para os eventos, identificando assim os relacionados ao tráfego da região da Baía de São Francisco, EUA.

Sobre a mesma temática, (DI LORENZO et al., 2013) desenvolveram uma ferramenta inteligente e interativa para exploração visual da dinâmica de eventos sociais ao longo das dimensões espacial, temporal e organizacional. O tráfego também foi objeto de estudo em (CHEN et al., 2016), ao relacionar eventos do trânsito com a demanda por bicicletas; e em (LECUE et al., 2014), ao demonstrar uma plataforma para análise inteligente do tráfego (em tempo real), com base em fontes heterogêneas de dados (incluindo *tweets* de agências oficiais de trânsito).

Em uma abordagem mais genérica, (MUKHERJEE et al., 2015) propuseram uma plataforma para processar (em *near real time*) questões urgentes da cidade, oriundas de diversas fontes (incluindo o *Twitter*), atuando como intermediadora entre cidadãos e agências civis. No que se refere a mobilidade urbana, mas não ao

uso de informações sobre pontos de interesse, (YOUSAF et al., 2014) inferiram a afinidade entre usuários por meio da análise de *retweets*, possibilitando que rotas de corridas sejam compartilhadas entre pessoas com interesses em comum, tornando a viagem mais agradável.

Por fim, em (FRIAS-MARTINEZ; FRIAS-MARTINEZ, 2014), os autores utilizaram apenas *tweets* geolocalizados para analisar suas respectivas distribuições no espaço urbano, com o objetivo de identificar a caracterização do uso da terra em zoneamentos urbanos industriais, residenciais, comerciais e de lazer. O trabalho foi realizado no contexto da cidade de Manhattan (EUA), Londres (Reino Unido) e Madrid (Espanha).

3.5.2 Casos de uso relacionados ao transporte público (QP2)

Nesta seção, são identificados os estudos primários que utilizam processamento de *tweets* com foco na mitigação dos problemas relacionados ao transporte público; enumerados a seguir:

1. Impacto de eventos no transporte público:

- a) impacto dos ataques terroristas em Paris no uso do transporte público (WEN; LIN; PELECHRINIS, 2016);
- b) impacto de eventos relacionados ao tráfego na demanda por bicicletas, em Nova Iorque e Washington D.C, EUA (CHEN et al., 2016);
- c) impacto dos pontos de interesse na demanda por transporte público (MAGHREBI et al., 2015);
- d) impacto dos eventos anormais nas tomadas de decisão dos passageiros do Metrô de Tokyo (ITOH et al., 2016);
- e) predição de fluxo de passageiros no Metrô de Nova Iorque (NI; HE; GAO, 2016).

2. Planejamento e gestão do transporte público:

- a) análise de sentimento relacionada ao acesso ao transporte público (GUO et al., 2016);
- b) coleta de informações relacionadas ao transporte público (GAL-TZUR et al., 2014);

- c) identificação de locais para estações de bicicletas, em St. Petersburg, Rússia (GUTEV; NENKO, 2016);
- d) identificação da disposição dos usuários para realizar viagens de lazer (GKIOTSALITIS; STATHOPOULOS, 2016);
- e) plataforma para notificação de problemas relacionados ao transporte público de Bangalore, Índia (MUKHERJEE et al., 2015).

Conforme os estudos primários analisados pela Revisão Sistemática, e enumerados nessa seção, é possível interpretar que os estudos estão classificados em análise de impacto de eventos, planejamento e gestão do transporte público. Por exemplo, (WEN; LIN; PELECHRINIS, 2016) utilizaram *tweets* para analisar o impacto dos ataques terroristas em Paris (2015) nos padrões de mobilidade referentes ao uso de transporte público. Semelhantemente, ITOH et al. (2016) desenvolveram uma ferramenta para analisar e explorar visualmente, com base em *tweets*, as tomadas de decisão dos passageiros do Metrô de Tokyo, ante a eventos anormais, tais como Tufões, Incêndios, Terremotos, dentre outros. Nesse mesmo contexto, (NI; HE; GAO, 2016) propuseram uma técnica de predição de fluxo de passageiros no Metrô de Nova Iorque, identificando eventos com base nas *hashtags* dos *tweets*. Enquanto que em (CHEN et al., 2016), analisaram a relação entre eventos do tráfego com a demanda por bicicletas.

No que se refere aos estudos focados no planejamento e gestão do transporte público, (MUKHERJEE et al., 2015) apresentam uma plataforma desenvolvida e utilizada pela Agência de Transporte Público de Bangalore, na Índia, a qual permite que usuários reportem questões relacionadas ao transporte público, o que possibilita a melhoria do planejamento de suas respectivas operações, assim como do serviço prestado para a população. Nessa mesma linha de estudo, em (GUTEV; NENKO, 2016), os autores usaram *tweets* para identificar a popularidade de determinados locais, pontos de interesse e distribuição etária, com o objetivo de determinar os melhores pontos para estações de bicicletas e incentivar assim o uso desse modal de transporte. Também relacionado aos pontos de interesse, (MAGHREBI et al., 2015) utilizaram *tweets* para identificar padrões das atividades humanas (em diferentes horários do dia) e seus respectivos impactos na demanda por transporte público.

Em (GAL-TZUR et al., 2014), por sua vez, utilizaram uma abordagem hierárquica para classificar *tweets* relacionados ao transporte. Com isso, demonstraram que é possível usar essas informações para fins de planejamento e gerenciamento do transporte. Tal técnica, foi aplicada em um estudo de caso associado a eventos esportivos, ocorridos no Reino Unido. A hierarquia é composta por três níveis, no primeiro, os *tweets* são classificados entre os que expressam a necessidade de serviços de transporte, opiniões e incidentes; o segundo, identifica a categoria do transporte; e último, relaciona *tweets* a tópicos.

Outro estudo que contribui com o planejamento do transporte público, é o realizado em (GKIOTSALITIS; STATHOPOULOS, 2015, 2016), no qual *tweets* foram processados para identificar a disposição dos usuários para realizar viagens relacionadas ao lazer (pontos de interesse), sugerindo a eles atividades com menor tempo de percurso e probabilidade de atrasos. Além do tempo de percurso, outro ponto relevante considerado foi o de bom nível de acesso ao transporte público, o qual quando existente impacta positivamente na felicidade das pessoas e se correlaciona com sentimentos positivos, segundo a análise de sentimentos realizada por (GUO et al., 2016), utilizando *tweets* publicados na Grande Londres.

3.5.3 Técnicas estatísticas utilizadas no processamento de *tweets* (QP3)

Nesta seção, são apresentadas as técnicas estatísticas utilizadas pelos estudos primários, no processamento de *tweets*, enumeradas a seguir:

1. **Análise de métricas relacionadas a desempenho** (erro de reconstrução relativo, qualidade dos componentes descritivos recuperados e qualidade dos componentes comuns recuperados) (WEN; LIN; PELECHRINIS, 2016);
2. **semelhança de cosseno** (Cosine similarity) (YOUSAF et al., 2014; FRIAS-MARTINEZ; FRIAS-MARTINEZ, 2014);
3. **f_1 score** (ANANTHARAM et al., 2015; CHEN et al., 2016);
4. **frequência do termo-inverso da frequência nos documentos** (TF-IDF — **Term frequency-inverse document frequency**) (MUKHERJEE et al., 2015);
5. **coeficiente de variação inversa** (BENDLER et al., 2014);
6. **método de reamostragem Jackknife** (BENDLER et al., 2014);

7. **indicadores locais de associação espacial (LISA — Local Indicators of Spatial Association)**) (STEIGER et al., 2015);
8. **local Moran's** (STEIGER et al., 2015);
9. **máxima verossimilhança** (MUKHERJEE et al., 2015);
10. **média móvel integrada autoregressiva sazonal (SARIMA — Seasonal Autoregressive Integrated Moving Average)** (NI; HE; GAO, 2016);
11. **otimização e previsão com função de perda híbrida** (NI; HE; GAO, 2016).

Em (NI; HE; GAO, 2016), os autores utilizaram a técnica SARIMA em conjunto com Regressão Linear, propondo uma abordagem baseada em otimização paramétrica e convexa, chamada *otimização e previsão com função de perda híbrida*, adequada para modelagem utilizando séries temporais. Com isso, tal técnica foi aplicada na predição de fluxo de passageiros com base em hashtags de tweets.

Referente aos problemas relacionados a ambiguidade e identificação de contextos, (ANANTHARAM et al., 2015); (CHEN et al., 2016; GAL-TZUR et al., 2014) aplicaram a técnica F_1 score validar o processo de classificação de tweets. Por outro lado, (MUKHERJEE et al., 2015) utilizaram a técnica de *máxima verossimilhança* para determinar a probabilidade de ocorrência de um evento, assim como a confiabilidade da informação.

No que se refere a agrupamento, (YOUSAF et al., 2014) agruparam usuários por meio de *semelhança de cossenos*, unindo pessoas com interesses em comum nos mesmos grupos. (FRIAS-MARTINEZ; FRIAS-MARTINEZ, 2014), por outro lado, usou a mesma técnica para agrupar tweets de acordo com suas semelhanças quanto aos tipos de zoneamento urbano.

De forma isolada, no trabalho realizado por (MUKHERJEE et al., 2015), utilizaram a técnica TF-IDF na fase de classificação para o definir o *score* de categorias de eventos, escolhendo a mais relevante a ser buscada em um dicionário de categorias. Também isoladamente, (STEIGER et al., 2015) usaram a técnica LISA na identificação de *clusters* espaciais e valores esporádicos espaciais, obtendo assim os locais com atividades sociais. Além disso, os autores também utilizaram a técnica *Local Moran's* para detectar diferentes padrões de atividade de acordo com o espaço geográfico.

Por último, (BENDLER et al., 2014) inovaram ao utilizar o *método de reamostragem Jackknife* como inspiração para o desenvolvimento de uma abordagem que visa analisar a estabilidade estatística de um conjunto de categorias. Além disso, usaram também a análise do *coeficiente de variação inversa* para verificar a dispersão negativa da distribuição de um conjunto de variáveis.

3.5.4 Paradigmas de processamento (QP4)

Nesta seção, encontram-se a seguir apenas os paradigmas de processamento extraídos dos estudos primários analisados:

1. **Processamento em lote (*batch, offline processing*)** (ANANTHARAM et al., 2015; WEN; LIN; PELECHRINIS, 2016; FARSEEV et al., 2015; GUTEV; NENKO, 2016; MATA; CLARAMUNT, 2015; CHEN et al., 2016; ABBASI et al., 2015; BENDLER et al., 2014; YOUSAF et al., 2014; FRIAS-MARTINEZ; FRIAS-MARTINEZ, 2014; STEIGER et al., 2015; GAL-TZUR et al., 2014; GKI-OTSALITIS; STATHOPOULOS, 2016; DI LORENZO et al., 2013; ITOH et al., 2016; CHANIOTAKIS; ANTONIOU, 2015);
2. **processamento em quase tempo real (*Near real time*)** (MUKHERJEE et al., 2015);
3. **processamento em tempo real (*Real time processing*)** (SOOMRO; KHAN; HASHAM, 2016; LECUE et al., 2014).

3.5.5 Eventos de exceção relacionados ao transporte público (QP5)

Nesta seção, encontram-se a seguir os eventos de exceção relacionados ao transporte público, extraídos dos estudos primários:

1. **Acidentes** (ITOH et al., 2016):
 - a) acidentes nas estações transporte;
 - b) incêndio.
2. **Espaço-temporais** (CHEN et al., 2016):
 - a) dia da semana;

- b) hora do dia.
3. **Eventos sociais** (CHEN et al., 2016; LECUE et al., 2014; GAL-TZUR et al., 2014; ITOH et al., 2016):
- a) feiras de rua;
 - b) festivais;
 - c) jogos esportivos;
 - d) passeatas e maratonas.
4. **Eventos urbanos** (CHEN et al., 2016; LECUE et al., 2014):
- a) relacionados ao tráfego.
5. **Desastres naturais** (ITOH et al., 2016):
- a) tempestades;
 - b) terremoto;
 - c) tufões.
6. **Metereológicos** (CHEN et al., 2016):
- a) dia claro, nublado, chuvoso, nevando, com neblina;
 - b) temperatura do ar.

3.5.6 Técnicas de Aprendizado de Máquina utilizadas no processamento de tweets (QP6)

Nesta seção, são apresentadas as técnicas de aprendizado de máquina utilizadas para processamento de tweets, extraídas dos estudos primários e enumeradas a seguir:

1. **Classificação bayesiana** (MATA; CLARAMUNT, 2015);
2. **algoritmo C5.0** (ZAGAL; MATA; CLARAMUNT, 2016);
3. **campo aleatório condicional com Regressão Logística (*Conditional Random Field (CRF) with Logistic Regression*)** (ANANTHARAM et al., 2015);
4. **alocação latente de Dirichlet (LDA — Latent Dirichlet Allocation)** (FAR-SEEV et al., 2015; ABBASI et al., 2015; HASAN; UKKUSURI, 2014; DI LORENZO et al., 2013; NI; HE; GAO, 2016);

5. **Regressão Linear** (GUTEV; NENKO, 2016; BENDLER et al., 2014; NI; HE; GAO, 2016; GUO et al., 2016);
6. **simulação de Monte Carlo** (CHEN et al., 2016);
7. **PairFac** (técnica inovadora que utiliza fatorização tensorial (*tensor factorization*)) (WEN; LIN; PELECHRINIS, 2016);
8. **Floresta Aleatória** (FARSEEV et al., 2015);
9. **Máquina de Vetores de Suporte (SVM — Support Vector Machine)** (MUKHERJEE et al., 2015), (GAL-TZUR et al., 2014);
10. **mapas auto-organizados (Self-Organizing Maps)** (FRIAS-MARTINEZ; FRIAS-MARTINEZ, 2014).

No contexto urbano, inúmeros eventos podem acontecer e impactar a população. O trabalho realizado por (WEN; LIN; PELECHRINIS, 2016), desenvolveu uma técnica que utiliza a análise de tensores discriminantes para aprender e de forma automatizada descobrir os impactos de um determinado evento no cotidiano da cidade. Numa abordagem mais simples, (CHEN et al., 2016) utilizou a técnica de *simulação de Monte Carlo* para treinar um modelo para predição de demanda por bicicletas, devido a dificuldade de encontrar exemplos suficientes para usar outras abordagens de treinamento.

Especificamente sobre as técnicas de classificação, (MUKHERJEE et al., 2015) utilizaram SVM para classificar os eventos recebidos de diversas fontes. Referente a essa abordagem, (GAL-TZUR et al., 2014) analisaram inúmeras técnicas de aprendizado de máquina, obtendo a melhor performance com o SVM, além disso, observaram como principal vantagem a sua capacidade de adaptação ao gênero e tarefas subjacentes.

Apesar disso, (GUO et al., 2016) utilizaram Processamento de Linguagem Natural (baseado em palavras chaves) para rotular sentimentos de tweets, devido a facilidade de escalar essa técnica (para processamento de milhões de tweets), em comparação a SVM. Outro caso de divergência é o do estudo realizado por (FARSEEV et al., 2015), no qual foi escolhido o algoritmo de *Floresta Aleatória* para treinamento do modelo de classificação, devido ao fato de ser mais adequado para classificação em espaço dimensional elevado, em vez das técnicas SVM e *Naive Bayes*, no que se refere a predição de idade e gênero usando tweets.

MATA; CLARAMUNT (2015), por sua vez, aplicou a técnica de classificação bayesiana em *tweets*, visando obter probabilidades relacionadas a crimes e ameaças em uma determinada localização. Por outro lado, (ZAGAL; MATA; CLARAMUNT, 2016) usaram o *algoritmo C5.0* devido ao melhor desempenho em relação a *Bayes*, dependendo do tópico que está sendo classificado.

Para anotação de eventos, (ANANTHARAM et al., 2015) treinaram um modelo CRF (usando anotações baseadas em dicionários) para determinar os locais da cidade e os termos relacionados aos eventos expressos em *tweets*. E, isoladamente (FRIAS-MARTINEZ; FRIAS-MARTINEZ, 2014) utilizaram a técnica *Self-Organizing Maps*, tendo como entrada os valores de latitude e longitude de *tweets*. Com isso, construíram um mapa segmentado em áreas urbanas, baseando-se nas regiões com diferentes concentrações de *tweets*.

Em relação a localidades, segundo (FARSEEV et al., 2015), a técnica LDA tem sido muito utilizada para identificação de pontos de interesses mencionados em *tweets*, sendo adequada para grandes bases de dados e agrupamento de *tweets* com tópicos similares, de acordo com (STEIGER et al., 2015). (ABBASI et al., 2015) exemplificou isso ao aplicar LDA para identificação de *tweets* relacionados ao turismo; (HASAN; UKKUSURI, 2014), para identificação de padrões de atividades humanas; e (DI LORENZO et al., 2013), para identificação de eventos sociais.

No entanto, (NI; HE; GAO, 2016) em vez de usarem LDA, extraíram hashtags de *tweets* para um vetor, utilizando-o para medir as atividades sociais e identificar seus respectivos contextos. Segundo (NI; HE; GAO, 2016), isso se justifica devido ao fato de que há uma grande chance do alto volume de *tweets* não indicar necessariamente eventos e atendimentos a eles. Além disso, afirmam que o método baseado em *hashtag* é capaz de indicar sobre o que é o evento, mesmo não utilizando o inglês formal.

Por sua vez, em (GUTEV; NENKO, 2016), os autores utilizaram Regressão Linear (RL) para analisar a demanda por bicicletas de acordo com as localizações extraídas dos *tweets*. Enquanto que (BENDLER et al., 2014) usaram RL para fornecer evidências de que as categorias dos pontos de interesse se relacionam com as variáveis referentes ao espectro espaço-temporal; e (GUO et al., 2016) para analisar a correlação entre sentimentos positivos com as oportunidades de trabalho, com a quantidade de crianças, e com o acesso a transporte.

3.6 Considerações finais sobre a revisão sistemática

Em uma análise quantitativa dos estudos primários selecionados, podemos concluir que a quantidade de artigos publicados sobre o uso de *tweets* na caracterização de problemas urbanos e relacionados ao transporte público tem crescido consideravelmente, entre 2011 e 2016. Provavelmente, devido ao fato da popularização das Redes Sociais e grande quantidade de dados disponíveis para processamento.

Tais estudos estão concentrados em maioria na identificação de pontos de interesse, utilizando-os em diferentes contextos, tais como o de turismo, mobilidade. Além disso, abordam também problemas relacionados ao transporte e desastres naturais, confirmando a primeira hipótese (HP1) dessa Revisão Sistemática. As temáticas não abordadas pela HP1 foram as relacionadas a *e-Participation*, detecção de zoneamento urbano, padrões demográficos e segurança pública, demonstrando a variedade de problemas urbanos explorados com o processamento de *tweets*.

Referente a segunda hipótese, os estudos exploraram principalmente o impacto de eventos no transporte público, confirmando-a parcialmente. Isso, devido ao fato de um dos trabalhos explorar como os eventos relacionados ao tráfego impactam na demanda por bicicletas; não havendo nenhum outro sobre processamento de *tweets* para mitigação dos problemas envolvendo tráfego. Outra temática não mencionada pela HP2 e sobre a qual há uma quantidade considerável de estudos, foi a do uso de *tweets* para o planejamento e gerenciamento do transporte público.

Independentemente dos problemas abordados por meio do processamento de *tweets*, dentre as 12 técnicas estatísticas elencadas, f_1 score foi a única referenciada como ferramenta para validação de classificação binária, confirmando a terceira hipótese (HP3). Apesar disso, a HP3 não considerou outras técnicas importantes (com propósitos distintos), como a de Regressão Linear, amplamente utilizada nos estudos analisados. Referente as técnicas de aprendizado de máquina, a mais utilizada foi a *Latent Dirichlet Allocation* (LDA), seguida da *Support Vector Machine* (SVM), confirmando parcialmente a sexta hipótese (HP6).

Por fim, apenas quatro dos vinte e nove estudos analisados, cerca de 14%, mencionaram *features* relacionadas ao transporte público, confirmando assim a quinta hipótese (HP5). Assim como a quantidade de trabalhos que realizam pro-

cessamento de *tweets* em tempo real, sendo apenas dois do total analisado, cerca de 6%, que utilizam esse paradigma de processamento, o que confirma a quarta hipótese (HP4). É importante ainda observar que, outros estudos que mencionaram processamento em tempo real, realizaram na verdade coleta de *tweets* em tempo real, para análises a posteriori via processamento em *batch* (offline), categoria na qual a maioria dos estudos foram enquadrados.

4 Dados abertos relacionados ao transporte público e eventos de exceção

Neste capítulo são apresentados o *corpus* da SPTrans e do *Twitter*, compostos por dados abertos relacionados ao transporte público e eventos de exceção, respectivamente. Os dados da SPTrans são divididos em dados AVL (enviados pelos módulos AVL instalados nos ônibus) e da GTFS (padrão utilizado para especificar os dados estáticos relacionados a operação dos ônibus da cidade de São Paulo). Os dados do *Twitter*, por sua vez, são compostos dos *tweets* coletados dos perfis governamentais responsáveis por reportar eventos de exceção da cidade de São Paulo.

4.1 Corpus SPTrans

Os dados AVL e da GTFS da SPTrans não são triviais de serem processados (grande volume de dados, dados sem tipo explicitamente definido — não tratados, dados separados em lotes de dados — um arquivo para cada hora de movimentação dos ônibus, dados fora do formato convencional — por exemplo, 24h em vez de 0h), devido a isso foram desenvolvidos *scripts* para um processo de ETL (*Extract, Transform and Load*).

4.1.1 Dados da General Transit Feed Specification da SPTrans

A GTFS (*General Transit Feed Specification*)¹, como o próprio nome sugere, é uma especificação de um formato comum (o que permite interoperabilidade) para troca de informações estáticas sobre transporte público. Um *feed* especificado na GTFS estática é composto por arquivos de texto (que seguem determinados requisitos semelhantes aos do formato *CSV*¹) compactados no formato *Zip*², e detalhados na Tabela 4. Cada arquivo modela diferentes perspectivas do transporte público, tais como paradas, trajetos, viagens e outros dados relativos a horário. As descrições dos arquivos da GTFS da SPTrans estão detalhadas na Tabela 3.

Além da GTFS estática existe a GTFS-realtime¹, que é uma extensão da GTFS estática, assim, para usar *feeds* em tempo real é necessário definir os arquivos

¹ <<https://developers.google.com/transit>>. Acessado em 29 de outubro de 2017.

² <<https://support.pkware.com/display/PKZIP/APPNOTE>>. Acessado em 29 de outubro de 2017.

Tabela 3 – Arquivos e número de registros especificados na GTFS pela SPTrans

| Nome do arquivo | Número de registros |
|----------------------------|----------------------------|
| <i>agency.txt</i> | 1 |
| <i>calendar.txt</i> | 6 |
| <i>fare_attributes.txt</i> | 6 |
| <i>fare_rules.txt</i> | 5.400 |
| <i>frequencies.txt</i> | 39.625 |
| <i>routes.txt</i> | 291.634 |
| <i>shapes.txt</i> | 800.767 |
| <i>stop_times.txt</i> | 95.134 |
| <i>stops.txt</i> | 19.933 |
| <i>trips.txt</i> | 2.273 |
| Total | 1.254.779 |

Fonte: Elaborado pelo autor

estáticos da GTFS, que são utilizados na GTFS-realtime para obter as informações do sistema de transporte público. A GTFS-realtime está fora do escopo desse trabalho.

Tabela 4 – Detalhamento dos arquivos da GTFS

| Nome do arquivo | Condisional | Contéudo ^a |
|----------------------------|--------------------|--|
| <i>agency.txt</i> | Obrigatório | Contém uma ou mais agências de transporte público como fonte dos dados. |
| <i>stops.txt</i> | Obrigatório | Contém os locais individuais em que os veículos peggam ou deixam passageiros. |
| <i>routes.txt</i> | Obrigatório | Contém os trajetos de um grupo de viagens exibidas aos passageiros como um único serviço. |
| <i>trips.txt</i> | Obrigatório | Contém as viagens de cada trajeto. Uma viagem é uma sequência de duas ou mais paradas que ocorrem em um horário específico. |
| <i>stop_times.txt</i> | Obrigatório | Contém os horários de partida e chegada dos veículos em paradas específicas em cada viagem. |
| <i>calendar.txt</i> | Obrigatório | Contém datas para IDs de serviço que usam uma programação semanal. Especificam quando o serviço começa e termina, bem como os dias da semana em que o serviço está disponível. |
| <i>calendar_dates.txt</i> | Opcional | Contém as exceções para IDs de serviço definidos no arquivo <i>calendar.txt</i> . Se o arquivo <i>calendar_dates.txt</i> inclui todas as datas de serviço, ele pode ser especificado no lugar do <i>calendar.txt</i> . |
| <i>fare_attributes.txt</i> | Opcional | Contém informações sobre tarifas dos trajetos de uma empresa de transporte público. |
| <i>fare_rules.txt</i> | Opcional | Contém regras para implementação das informações de tarifa dos trajetos de uma empresa de transporte público. |
| <i>shapes.txt</i> | Opcional | Contém regras para desenhar linhas em um mapa para representar os trajetos de uma empresa de transporte público. |
| <i>frequencies.txt</i> | Opcional | Contém os intervalos entre as viagens nos trajetos. |
| <i>transfers.txt</i> | Opcional | Contém regras para conexões em pontos de baldeação entre os trajetos. |
| <i>feed_info.txt</i> | Opcional | Contém informações adicionais sobre o <i>feed</i> , incluindo editor, versão e informações sobre validade. |

^a Os campos contidos em cada arquivo da especificação GTFS estão descritos no apêndice C, nas tabelas 17 à 29.

Transformações nos dados da GTFS da SPTrans

Conceituações à parte, realizamos alguns processos de transformação nos dados da GTFS estática, antes de inserí-los no *MongoDB*, para viabilizarmos a correlação com os dados AVL. Dessa forma, convertemos os dados originais de *string* para os seus respectivos tipos (*long*, *double*, *int* ou *string*) e padronizamos os valores referentes a hora para *POSIX timestamp*, e os referentes a latitude e longitude para o formato *legacy coordinate pairs*³. Além disso, para fosse possível realizarmos consultas geoespaciais, foram criados *índices geoespaciais*³ nas coleções que contém dados geolocalizados. Dessa forma, conseguimos usar consultas geoespaciais para identificarmos as linhas afetadas por um determinado evento de exceção, dentro de um raio ajustável, por exemplo.

4.1.2 Dados AVL da SPTrans

O conjunto de dados AVL da SPTrans é composto por dados obtidos do SIM, transferidos pelos módulos AVL instalados nos ônibus da cidade de São Paulo. Os dados AVL utilizados nesta análise são referentes aos movimentos de ônibus ocorridos entre janeiro e dezembro de 2017 (solicitados por meio da *Lei de Acesso à Informação*⁴). Os dados de movimentação referentes a 01/11, das 2 h às 5 h, e a 15/12, das 01 h às 09 h, não foram disponibilizados pela SPTrans, devido a períodos de indisponibilidade do sistema de monitoramento (protocolo e-SIC 33310).

4.1.3 Identificação de incosistências e indisponibilidade na base de dados AVL da SPTrans

Os períodos indisponíveis foram identificados por meio de um *script*⁵ desenvolvido por este trabalho para análise do total de arquivos e espaço em disco, por período. O funcionamento do *script* consiste em gerar os respectivos nomes dos arquivos de movimentação que deveriam existir em determinado período, confrontando-os

³ <<https://docs.mongodb.com/manual/geospatial-queries>>. Acessado em 29 de outubro de 2017.

⁴ <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acessado em 23 de junho de 2018.

⁵ <https://github.com/fcas/mobility-analysis/blob/master/scripts/data_set_analyser.py>. Acessado em setembro de 2018.

Tabela 5 – Metadados dos dados AVL da SPTrans

| Nome do campo | Descrição do campo |
|--------------------------------|---|
| <i>cd_evento_avl_movto</i> | Código sequencial identificador do evento |
| <i>cd_linha</i> | Código identificador da linha em operação |
| <i>dt_movto</i> | Data da gravação em banco de dados do evento gerado no AVL |
| <i>nr_identificador</i> | Código identificador do AVL |
| <i>nr_evento_linha</i> | Grupo de indicadores relacionados ao evento |
| <i>nr_ponto</i> | Código do ponto notável |
| <i>nr_velocidade</i> | Velocidade instantânea |
| <i>nr_voltagem</i> | Tensão de alimentação |
| <i>nr_temperatura_interna</i> | Temperatura do processador |
| <i>nr_evento_terminal_dado</i> | Código do evento relacionado no terminal de dados |
| <i>nr_evento_es_1</i> | Grupo de indicadores relacionados ao evento |
| <i>nr_latitude_grau</i> | Latitude da geolocalização do veículo |
| <i>nr_longitude_grau</i> | Longitude da geolocalização do veículo |
| <i>nr_indiceregistro</i> | Índice de geração do evento no AVL |
| <i>dt_avl</i> | Data da geração do evento no AVL |
| <i>nr_distancia</i> | Distância em metros do evento com relação ao evento anterior do mesmo AVL |
| <i>nr_tipo_veiculo_geo</i> | Código para identificação no software de mapeamento |
| <i>cd_avlconexao</i> | Código interno utilizado para identificar qual a conexão utilizada para transmissão do evento |
| <i>cd_prefixo</i> | Prefixo do veículo |

com os existentes na base obtida, além de sumarizar o espaço em disco e total de arquivos. Tais metadados estão especificados na Tabela 6. Além dos períodos indisponíveis, durante o processo de leitura encontramos arquivos com linhas divergentes do arquivo de metadados fornecido pela SPTrans, contidos na Tabela 5, tais registros foram ignorados para os experimentos deste trabalho.

Tabela 6 – Descrição do conjunto de dados AVL

| Mês | Intervalo (dias) | Total de arquivos AVL | Espaço em disco (GB) |
|----------------------|-------------------------|------------------------------|-----------------------------|
| Janeiro ^a | 1 - 31 | 744 | 102,44 |
| Fevereiro | 1 - 28 | 672 | 93,21 |
| Março | 1 - 31 | 744 | 102,64 |
| Abril | 1 - 30 | 720 | 97,04 |
| Maio | 1 - 31 | 744 | 101,46 |
| Junho | 1 - 30 | 720 | 97,13 |
| Julho | 1 - 31 | 744 | 104,95 |
| Agosto | 1 - 31 | 744 | 108,38 |
| Setembro | 1 - 30 | 720 | 109,89 |
| Outubro | 1 - 31 | 744 | 110,92 |
| Novembro | 1 - 30 | 717 | 108,16 |
| Dezembro | 1 - 31 | 738 | 110,89 |
| Total | — | 8.751 | 1.247,09 |

^a Arquivos Movto_201701111000_201701111100 com 35 campos na linha 60.025 e Movto_201701110900_201701111000 com 21 campos na linha 1.075.548, o esperado são 19 campos de acordo com os metadados fornecidos pela SPTrans.

4.2 Corpus Twitter

No Twitter as informações (*tweets*) são publicadas contendo no máximo 280 caracteres; cada publicação pode receber *retweets* (ser compartilhada por outros usuários), comentários (diretamente no *tweet* — *replies* — ou de forma privada via caixa de mensagens) e *likes* (indicador de quantos usuários gostaram da publicação). Além dessas funcionalidades, os *tweets* podem conter menções a outros usuários (@nome do perfil) e rótulos (#*hashtag*) indicando assuntos, categorias, etc.

Devido as características citadas nos parágrafos anteriores, o Twitter tem sido uma rede social importante para compartilhamento de informações e acontecimentos do cotidiano. Tais acontecimentos podem ser classificados como eventos sociais, capazes de descrever desde eventos rotineiros (*shows*, jogos esportivos, etc.) a situações de crise (eventos de exceção — desastres naturais, mobilizações sociais, dentre outros) (ZHOU; CHEN, 2014), (ATEFEH; KHREICH, 2015).

Portanto, o Twitter foi escolhido como fonte de dados para a construção do conjunto de dados relacionados aos eventos de exceção devido ao fato de conter dados abertos sobre o cotidiano da cidade, disponibilizados em tempo real pelos cidadãos e órgãos públicos. Tais características, fazem dos *tweets* uma rica fonte de dados, utilizada por inúmeros estudos que abordam problemas urbanos e de mobilidade urbana, conforme os analisados na revisão sistemática do Capítulo 3.

Neste trabalho, o conjunto de dados utilizado para a identificação dos eventos de exceção é composto por *tweets*, em português brasileiro, dos perfis contidos na Tabela 1. É importante observar que, para esse projeto de pesquisa, apenas os *tweets* publicados pelas contas selecionadas são considerados, descartando os relacionados às interações (*retweets* e *replies*) entre perfis governamentais e não governamentais. Ou seja, os dados utilizados estão relacionados ao canal unidirecional de comunicação, não utilizamos interações dos cidadãos com as publicações realizadas pelos perfis selecionados. Com essa restrição, evitamos problemas referentes a confiabilidade dos dados, o que nos permite focarmos na caracterização dos eventos de exceção e de seus respectivos impactos.

Sobre a seleção dos perfis, todos foram selecionados manualmente de acordo com os órgãos responsáveis por notificar eventos de exceção. Tais perfis são de caráter público, ou seja, o acesso aos *tweets* não envolve questões de privacidade.

4.2.1 Processo de coleta dos *tweets*

Apesar do acesso facilitado aos *tweets*, a API do Twitter limita a quantidade e frequência de requisições aos *endpoints*. Devido a isso, o artefato de *software* desenvolvido para coleta (na linguagem de programação Java), busca (utilizando o *plugin Twitter4J*⁶) os 3.200 *tweets* mais recentes (se disponíveis) de cada conta, através do *endpoint statuses/user_timeline*; o qual permite no máximo 180 requisições, em um intervalo de 15 minutos, com autenticação via conta de usuário⁷.

Durante a coleta dos *tweets*, eles são mapeados para a seguinte classe do modelo da aplicação: *TweetInfo*, que contém as informações respectivas ao *id*, texto da publicação, *timestamp*, endereço extraído, latitude e longitude. Em seguida, o modelo é persistido no banco de dados não relacional *MongoDB*⁸ e também no banco de dados de séries temporais *Druid*⁹ para exploração e visualização dos dados, processo explicado na Seção 5. Os detalhes sobre o intervalo de tempo e o número de *tweets* coletados constam na Tabela 7.

⁶ <twitter4j.org>. Acessado em 29 de outubro de 2017.

⁷ <<https://dev.twitter.com>>. Acessado em 29 de outubro de 2017.

⁸ <<https://www.mongodb.com>>. Acessado em 29 de outubro de 2017.

⁹ <<http://druid.io>>. Acessado em 29 de outubro de 2017.

Tabela 7 – Intervalo de tempo e número de tweets coletados

| Perfil no Twitter | Total de tweets^a | Timestamp 1^b | Timestamp 2^c |
|--------------------------|------------------------------------|--------------------------------|--------------------------------|
| @BombeirosPMESP | 6,632 | 2017-05-21 | 2017-12-01 |
| @CETSP_ | 5,735 | 2017-02-20 | 2017-12-01 |
| @CPTM_oficial | 6,301 | 2017-04-24 | 2017-12-01 |
| @governosp | 6,011 | 2017-05-10 | 2017-12-01 |
| @metrosp_oficial | 8,621 | 2017-06-07 | 2017-12-01 |
| @Policia_Civil | 3,417 | 2015-04-15 | 2017-11-30 |
| @PMESP | 4,365 | 2016-06-02 | 2017-11-30 |
| @saopaulo_agora | 3,960 | 2016-11-18 | 2017-11-30 |
| @smtsp_ | 1,316 | 2017-04-26 | 2017-12-01 |
| @SPCEDEC | 1,301 | 2015-06-09 | 2017-12-01 |
| @sptrans_ | 9,956 | 2017-06-13 | 2017-12-01 |
| @TurismoSaoPaulo | 3,369 | 2012-06-12 | 2017-11-29 |
| Total | 60,984 | — | — |

^a Número de tweets coletados.^b Timestamp mais antigo.^c Timestamp mais recente.

Fonte: Elaborado pelo autor

5 Exploração e visualização de grandes volumes de dados

Grandes volumes de dados como os relacionados ao transporte público possuem padrões complexos e demandam um sistema distribuído, apresentado neste capítulo, capaz de suportar atividades analíticas, como visualização e exploração de dados. Tais análises, são importantes para melhorar o gerenciamento, operação e planejamento do transporte público.

Dessa forma, apresentamos uma arquitetura para visualizar e explorar grandes volumes de dados (objetivo específico, apresentado na Seção 1.3), validada com o Corpus SPTrans. No demais, mencionamos na Seção 5.1 alguns trabalhos referentes a visualização de dados, encontrados por meio de uma revisão não sistemática da literatura; na 5.2 é descrita a arquitetura do banco de dados *Druid*, principal componente da arquitetura proposta; na 5.3 a arquitetura em questão para processamento e exploração dos dados AVL; na 5.4 os resultados obtidos no estudo de caso e, por fim, na 5.5 as considerações finais.

5.1 Trabalhos relacionados

Em (CHEN; GUO; WANG, 2015) são mencionados conceitos básicos e fluxos de visualização de dados de tráfego (dos dados brutos, pré-processamento ao mapeamento visual, construído com símbolos visuais), além de uma visão geral das técnicas e métodos de processamento de dados relacionados para processar e descrever propriedades temporais, espaciais, numéricas e categóricas de dados de tráfego.

Analogamente, em (ANDRIENKO et al., 2017) é descrita uma tipologia de dados de tráfego, capaz de abordar suas respectivas propriedades, problemas e transformações relevantes para a análise. Além disso, são apresentadas abordagens analíticas visuais para analisar dados de tráfego de veículos, pedestres, passageiros dentro de sistemas de transporte, etc.

Por fim, no trabalho desenvolvido em (SERAJ; MERATNIA; HAVINGA, 2017) é apresentado um novo algoritmo para mapeamento de medições coletivas para monitorar as infraestruturas de transporte terrestre e, aliviar o impacto de imprecisões

do GPS para monitoramento contínuo de infraestruturas de transporte por meio de *smart phones*.

Nenhum dos trabalhos mencionados anteriormente aborda o uso de software livre com suporte a computação distribuída, escalabilidade, tolerância a falhas, processamento em tempo real, baixa latência e visualização de grandes volumes de dados temporais. Tais requisitos, são explorados neste trabalho usando banco de dados *Druid*, descrito na seção seguinte e, o *Apache Superset* para analisar padrões complexos existentes nos dados AVL da SPTrans.

5.2 *Druid*

O *Druid* é um banco de dados para análises exploratórias em tempo real (latências abaixo de sub-segundos) em grandes conjuntos de dados. A arquitetura distribuída do *Druid* é composta por um *cluster* com diferentes tipos de nós (*real-time*, *historical*, *broker* e *coordinator nodes*), que operam independentemente uns dos outros e possuem interação mínima entre eles. Existem duas dependências externas: (I) Apache Zookeeper¹, responsável pela coordenação do cluster e (II) um sistema de gerenciamento de banco de dados relacional (RDBMS — *Relational Database Management Systems*), para armazenar parâmetros operacionais adicionais e configurações (YANG et al., 2014).

5.2.1 Real-time nodes

Real-time nodes são responsáveis por ingerir, indexar e consultar fluxos de eventos. Periodicamente, cada nó agenda uma tarefa em segundo plano para procurar todos os índices localmente persistentes, mesclando-os para construir *blocos imutáveis de dados com todos os eventos ingeridos em um período de tempo*, conhecidos como *segmentos imutáveis*, os quais podem posteriormente serem carregados para uma camada de sistema de arquivos (*deep storage*²) (YANG et al., 2014).

¹ <<https://zookeeper.apache.org>>. Acessado em 23 de junho de 2018.

² <<http://druid.io/docs/latest/dependencies/deep-storage.html>>. Acessado em 21 de janeiro de 2019.

Durante os processos mencionados anteriormente não há perda de dados. Além disso, a imutabilidade dos blocos permite a consistência de leitura e um modelo de paralelização simples: *historical nodes* podem simultaneamente examinar e agregar blocos imutáveis de forma não bloqueante (YANG et al., 2014).

5.2.2 Historical nodes

Os *historical nodes* são responsáveis por carregar, descartar e servir *segmentos* imutáveis por meio de uma arquitetura *shared-nothing* (sem um único ponto de contenção entre os nós) (YANG et al., 2014).

5.2.3 Broker nodes

Os *broker nodes* são responsáveis por receber consultas e mesclar resultados parciais dos *historicals* e *real-time nodes* antes de retornar um resultado final consolidado para o cliente (YANG et al., 2014).

5.2.4 Coordinator nodes

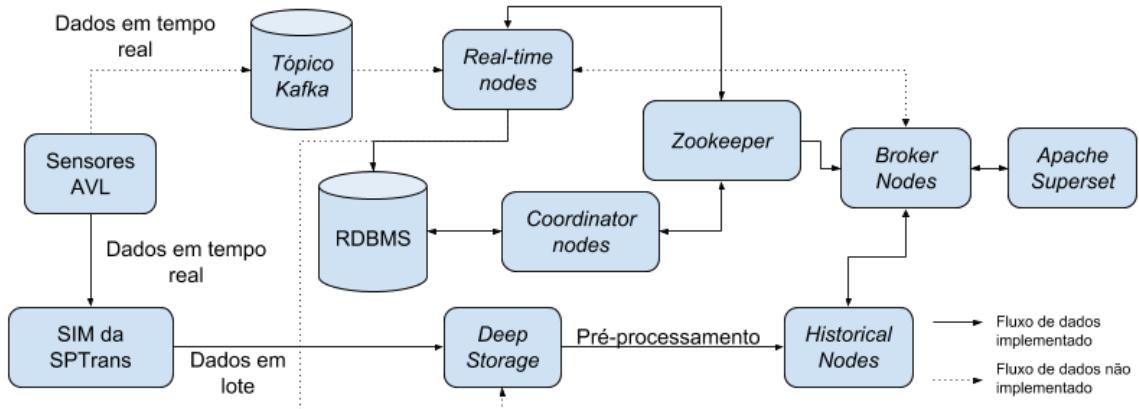
Os *coordinator nodes* são responsáveis pelo gerenciamento e distribuição dos dados nos *historical nodes*, exigindo destes o carregamento, descarte e replicação dos dados (YANG et al., 2014).

5.3 Arquitetura utilizada para visualização e exploração dos dados AVL da SPTrans

A Figura 6 mostra a arquitetura utilizada no estudo de caso deste capítulo, composta pelos componentes do *Druid* em conjunto com o módulo *Apache Superset*³ — software de código aberto para exploração e análise de dados, nativamente integrado ao *Druid*. Nesta arquitetura, dois fluxos para processamento de dados também são elencados: (I) em lote, para análises mais complexas (correlações, extrações de *features*, etc.) e (II) em tempo real, necessário devido ao requisito de análises tempestivas, normalmente envolvendo agregação e sumarização dos dados.

³ <<https://superset.incubator.apache.org>>. Acessado em 23 de junho de 2018.

Figura 6 – Arquitetura usada no estudo de caso para visualização e exploração dos dados AVL da SPTrans



O fluxo de processamento em lote é executado a partir dos dados extraídos do sistema de monitoramento da *SPTrans*, os quais são ingeridos nos *historical nodes* e disponibilizados para o *Apache Superset* por meio dos *broker nodes*. É importante observar que o fluxo de processamento em lote é o fluxo de dados implementado neste estudo de caso.

Na arquitetura ilustrada na Figura 6, o fluxo de dados em tempo real refere-se a uma proposta alvo para a *SPTrans*, a fim de permitir a exploração e visualização dos dados dos ônibus da cidade de São Paulo em tempo real. Nesta proposta, os tópicos do *Apache Kafka*⁴ (plataforma distribuída para processamento de fluxos de dados) desempenham o papel de receptores do fluxo de dados, a partir dos quais os dados podem seguir tanto o processamento em tempo real quanto em lote.

Por fim, é importante observar que em ambos os fluxos há um estágio de pré-processamento de dados, para adequar os dados AVL as especificações exigidas para a ingestão no *Druid* (o que adiciona atraso no fluxo de processamento).

5.4 Estudo de caso com os dados AVL da SPTrans

Grandes volumes de dados podem conter padrões complexos e difíceis de serem identificados. Devido a isso, é importante construir visualizações auxiliares para o processo de análise de dados. Com este propósito, usamos o *Apache Superset*⁵, com suporte nativo ao *Druid*, para exploração e visualização do *corpus* da *SPTrans*.

⁴ <<https://kafka.apache.org>>. Acessado em 21 de janeiro de 2019.

⁵ <<https://superset.incubator.apache.org>>. Acessado em 29 de junho de 2018

As figuras 7, 8, 9 e 10 são exemplos de algumas visualizações construídas a partir dos dados de janeiro das linhas de ônibus selecionadas aleatoriamente.

A Figura 7 ilustra uma série temporal referente à quantidade de dados enviados por ônibus selecionados aleatoriamente, referentes a janeiro de 2017. Com esta visualização é possível observar, por exemplo, a oscilação da quantidade de dados enviados, assim como os picos de maior e menor volume de envio de dados e janelas de tempo com dados ausentes. Tais oscilações podem indicar problemas relacionados a essas viagens, como eventos de exceção decorrentes de paralisação sindical (redução da frota de ônibus), atos de violência como os de 2006 que incendiaram noventa ônibus em São Paulo⁶ e os de 2019 com 21 ônibus incendiados no Ceará⁷.

Figura 7 – Quantidade de dados enviados por dia por ônibus (selecionados aleatoriamente) em janeiro de 2017



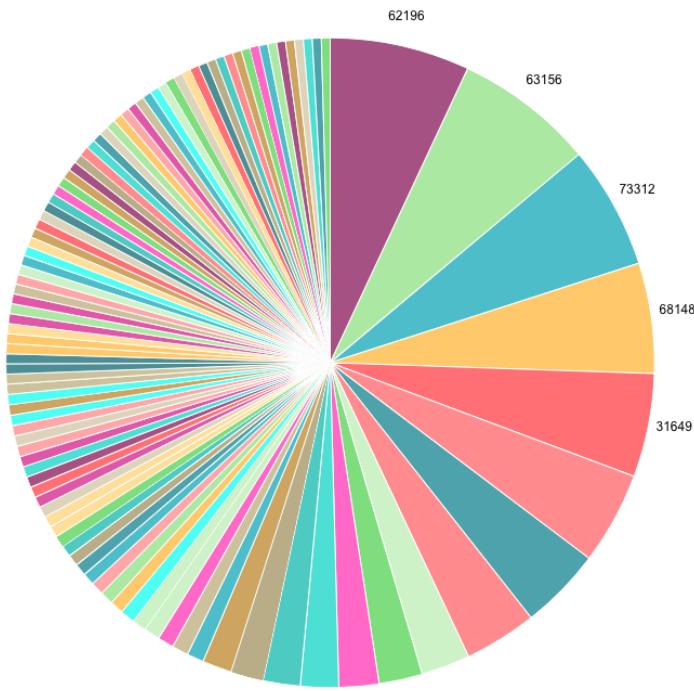
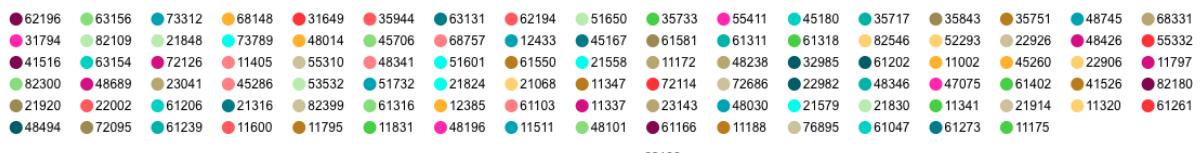
A Figura 8, representa a distribuição da quantidade de dados enviados em janeiro, a partir de uma amostra aleatória de linhas ônibus. Nessa figura é possível

⁶ <https://pt.wikipedia.org/wiki/Atos_de_viol%C3%Aancia_organizada_no_Brasil_em_2006>. Acessado em 21 de janeiro de 2019

⁷ <https://pt.wikipedia.org/wiki/Atentados_no_Cear%C3%A1_em_2019>. Acessado em 21 de janeiro de 2019.

analisar que a distribuição da quantidade de dados enviados não é normalizada, ou seja, existem ônibus que normalmente enviam mais dados do que os demais. Há muitas razões possíveis para isso, por exemplo: viagens de ônibus mais longas que outras, regiões com diferenças climáticas; módulos AVL desatualizados; maior quantidade de ônibus em uma determinada linha, etc.

Figura 8 – Distribuição da quantidade de dados enviados por ônibus (selecionados aleatoriamente) em janeiro de 2017



Finalmente, os mapas exibidos pelas figuras 10 e 9 ajudam a identificar a localização a partir da qual os dados estão sendo enviados, permitindo visualizar possíveis pontos de falhas durante a transmissão desses dados. O primeiro mapa, respectivamente, refere-se à rota de uma única linha de ônibus e o segundo de todas as rotas; em ambos os casos, referentes aos dados de janeiro. Além disso, na Figura 10, é possível observar a segregação urbana da cidade, devido ao fato de algumas regiões terem uma maior densidade de dados enviados, o que também indica regiões de maior tráfego, nas quais eventos de exceção teriam maior impacto.

Figura 9 – Localizações enviadas em Janeiro de 2017 de uma linha de ônibus selecionada aleatoriamente

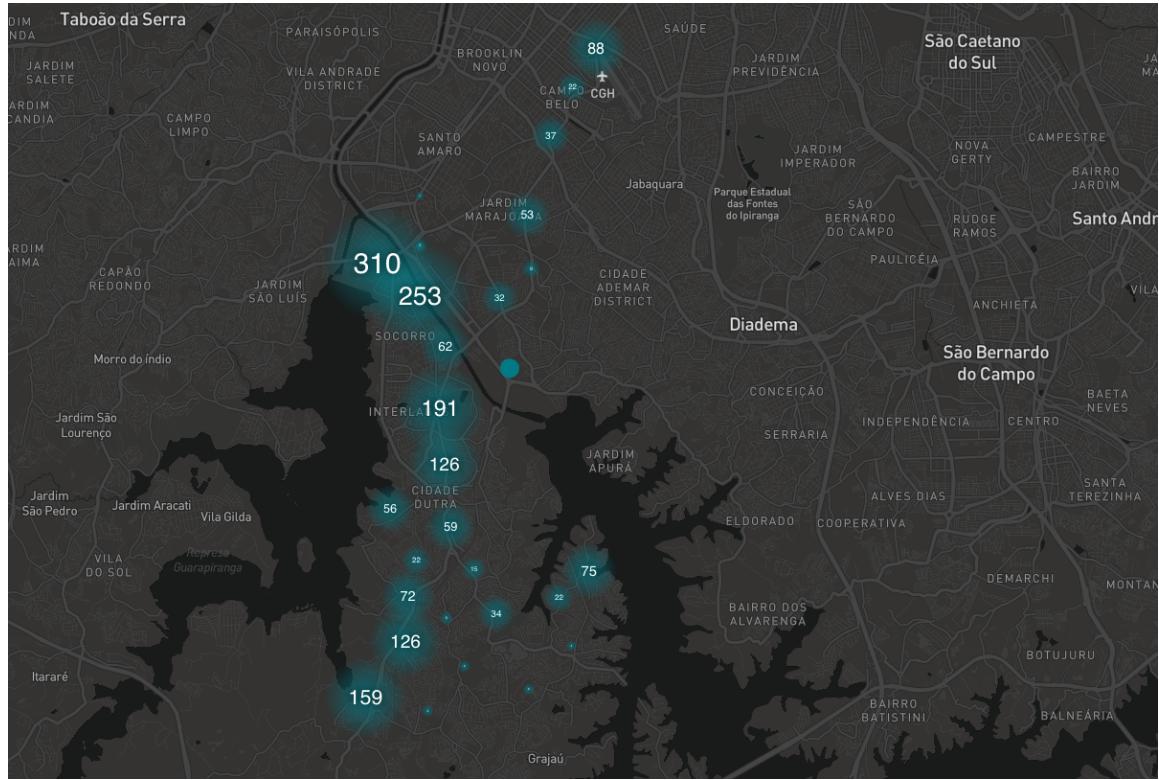
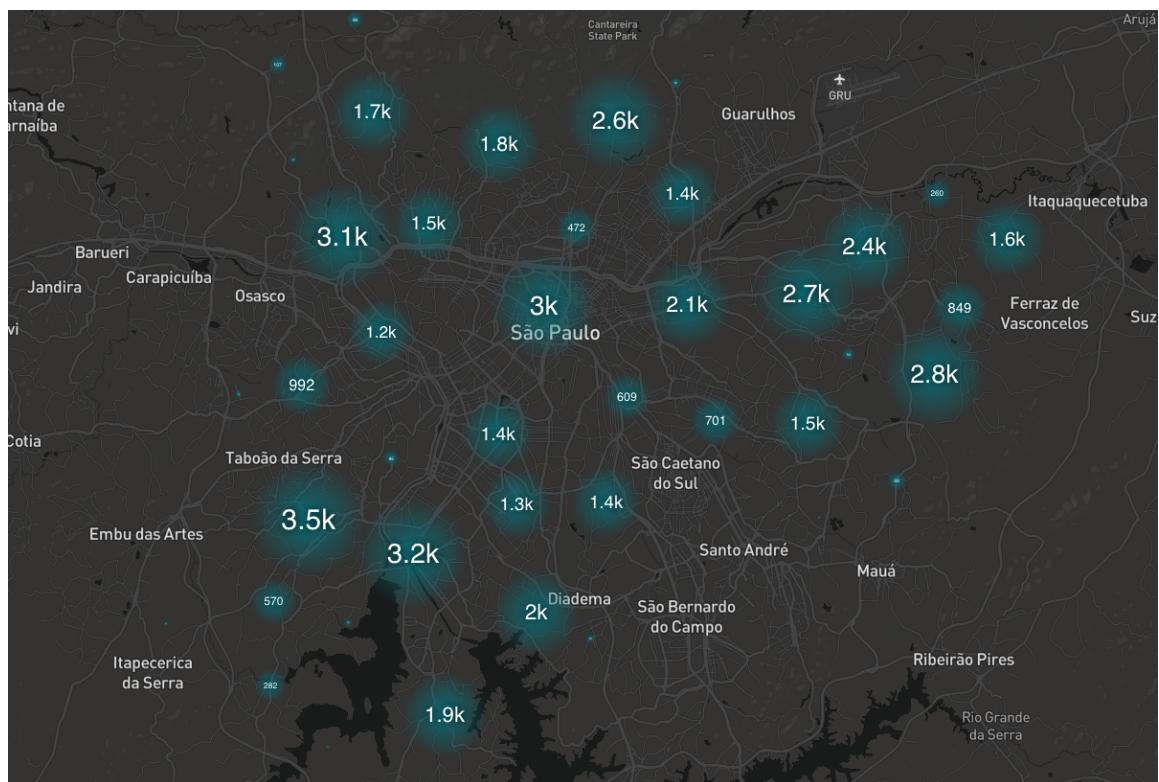


Figura 10 – Localizações dos ônibus referente a movimentação de Janeiro de 2017



5.5 Consideração sobre a arquitetura utilizada para exploração e visualização dos dados AVL da SPTrans

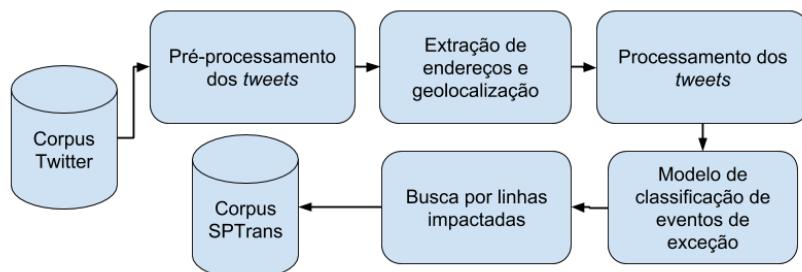
Este capítulo apresentou um estudo de caso relacionado à visualização de grandes conjuntos de dados, utilizando dados dos ônibus da cidade de São Paulo. Também, mostramos que é possível encontrar padrões complexos e incomuns e possíveis eventos de exceção em grandes conjuntos de dados por meio da visualização. O *Druid* e o *Apache Superset* demonstraram suporte a agregação, exploração e visualização de grandes conjuntos de dados.

6 Uma metodologia baseada em tweets para encontrar linhas de ônibus impactadas por eventos de exceção na cidade de São Paulo

O objetivo geral desse projeto de pesquisa é a caracterização de eventos de exceção e de seus respectivos impactos no sistema de transporte público por ônibus da cidade de São Paulo, conforme mencionado na Seção 1.3. Portanto, para alcançarmos o objetivo proposto precisamos de uma metodologia capaz de encontrar as linhas de ônibus que são impactadas por eventos de exceção, para então explorarmos as características desse impacto. Sendo assim, apresentamos neste capítulo uma metodologia baseada em tweets para identificar linhas de ônibus impactadas por eventos de exceção. De acordo com a Figura 11, a metodologia, explicada em detalhes nas seções seguintes, é composta por:

1. Uma base de dados de tweets — *Corpus Twitter*.
2. Pré-processamento dos tweets existentes no conjunto de dados.
3. Extração de localização e geolocalização.
4. Processamento dos tweets.
5. Criação de um modelo de classificação de tweets em classes de eventos de exceção.
6. Identificação das linhas impactadas — por meio de consultas a base GTFS existente no *Corpus SPTrans* — a partir de um raio de cada evento de exceção.

Figura 11 – Metodologia baseada em tweets para encontrar linhas de ônibus impactadas por eventos de exceção na cidade de São Paulo



6.1 Pré-processamento

Numa pré-análise do *Corpus Twitter*, podemos averiguar que os tweets publicados pelos perfis selecionados evitam o uso de gírias, abreviações, erros de

digitação; conforme consta nos *tweets* de exemplo contidos no trecho de código em *json*, no apêndice A. Isso diferencia tais *tweets* dos *tweets* publicados por usuários comuns do *Twitter*, que contém erros gramaticais, de sintáxe e que normalmente dependem de análise contextual para que possam ser interpretados.

Apesar disso, com base na literatura analisada ((STEIGER et al., 2015), (MIDDLETON; MIDDLETON; MODAFFERI, 2014), (KOBANI; SCHÜTZE; BURKOVSKI, 2010), (SETIAWAN; WIDYANTORO; SURENDRO, 2017), (ZAGAL; MATA; CLARAMUNT, 2016)), as seguintes etapas de pré-processamento são necessárias para remoção de ruído e redução da dimensão do espaço de *features* e foram realizadas para o *Corpus Twitter*:

- *Case folding*: processamento de normalização de todas as letras do texto (de a-z) para minúsculas.
- Remoção *URLs* e menções a outros *tweets*.
- Remoção de acentos, *emoticons* e pontuações substituídas por espaços vazios.
- *Stemming* — realizado neste trabalho na fase de processamento mencionada na Seção 6.3, com o objetivo de não afetar o processo de extração de endereços.

Além disso, é importante observar que (I) as informações referentes a data e hora mencionadas no conteúdo dos *tweets* (*stopwords* específicas do domínio) são removidas do texto original. As informações de data e hora consideradas para os eventos de exceção são as contidas nos metadados dos *tweets*, posto que ao analisarmos os *tweets* verificamos que as informações de data e hora contidas no texto normalmente são referentes a eventos futuros, os quais não são considerados por este trabalho; (II) os *retweets* não estão presentes no *Corpus Twitter*; (III) no pré-processamento não há transformação do conteúdo dos *tweets*, embora trabalhos como os relacionados a identificação de sentimentos usem esse meio para transformar *emoticons* nos sentimentos que eles representam (ZAGAL; MATA; CLARAMUNT, 2016); (IV) as *hashtags* não são removidas dos *tweets* originais, pois são importantes para a classificação dos eventos de exceção.

Uma atenção especial foi dada às *hashtags*, que são relevantes para a classificação de eventos de exceção, mas adicionam ruído à fase de extração de endereços. Para mitigar o problema, *hashtags* são identificadas e substituídas por espaços

vazios no processo de extração de endereço. Além disso, é importante notar que as *hashtags* não são removidas dos tweets originais.

6.2 Extração de endereço e geolocalização

Analizando o conteúdo dos tweets das contas selecionadas, é possível observar que os textos publicados seguem um determinado padrão e, portanto, são semi-estruturados. Ante a isso, usamos a seguinte expressão regular para extrair os endereços presentes no conteúdo dos tweets:

$$ER = \{L_1|S_1|L_2|S_2|\dots|L_n|S_n\}\{[a - z\AA - \ddot{y}] +\} \quad (13)$$

A expressão anterior é dividida em dois conjuntos, no primeiro ($\{L_1|S_1|L_2|S_2|\dots|L_n|S_n\}$), (L — logradouros) e (S — acrônimos de espaços públicos) são concatenados para especificar um filtro e identificar sequências inicializadas com espaços públicos ou seus respectivos acrônimos. No segundo conjunto ($\{[a - z\AA - \ddot{y}] +\}$), é especificado um filtro para identificar um conjunto de palavras após L ou S, que são candidatas a compor o endereço desejado.

Essas palavras são candidatas porque é difícil saber quantas palavras após L ou S pertencem ao endereço, no entanto, as contas selecionadas publicam padrões visíveis após os endereços. Como consequência, um método possível para encontrar o endereço desejado é a remoção desses padrões após o início do endereço.

Após a extração do endereço, é necessário geolocalizar o endereço encontrado — apenas 1,5 % de tweets têm geolocalização (NIU et al., 2016) — o que é possível, por exemplo, usando a API de geocodificação do Google Maps¹. Os parâmetros de URL utilizados neste trabalho para chamar a API mencionada anteriormente são: (I) *address* — o endereço desejado; (II) *bounds* — uma caixa delimitadora para o resultado retornado, a qual é especificada pelas coordenadas de latitude / longitude dos cantos sudoeste e nordeste de São Paulo; (III) *region* — código da região com dois caracteres, por exemplo, *br* para o Brasil e (IV) *token* — *token* usado na autenticação da API.

Em seguida, a resposta HTTP é processada para obter os valores da localização (que contém informações de latitude e longitude) e o endereço formatado.

¹ <<https://developers.google.com/maps/documentation/geocoding>>. Acessado em 11 de Abril de 2018.

É importante observar que os *tokens* do endereço extraído (*endereço não formattado*) são *stopwords* específicas do *corpus* em caso de alta frequência de eventos de exceção localizados neste endereço, devido ao fato de que nesse cenário elas são tratados como *features* relevantes para o modelo de classificação. Portanto, os *tokens* dos endereços extraídos são armazenados para serem removidos na fase de processamento dos *tweets*.

6.3 Processamento de tweets

Nesta fase, os *tweets* são preparados para serem usados para treinar um modelo de classificação de eventos de exceção; neste momento, todos os *tweets* já foram pré-processados. Conforme mencionado na seção anterior, nesta fase, os *tokens* dos endereços extraídos armazenados são removidos para redução de ruído e as *stopwords* do português brasileiro filtradas² e todos os demais *tokens* processados por um *stemmer* para o português brasileiro³ para reduzir a dimensão do espaço de *features*.

6.4 Classificação manual do Corpus Twitter

Encontrar eventos de exceção envolve a identificação de eventos relacionados a uma exceção, o que é possível por meio de classificação de *tweets* (manualmente ou de forma autônoma). De acordo com a revisão sistemática realizado no Capítulo 3, as seguintes classes podem ser usadas para classificar eventos de exceção:

1. **Acidentes.**

- a) Acidentes nas estações transporte (ITOH et al., 2016).
- b) Incêndio (ITOH et al., 2016).

2. **Espaço-temporais.**

- a) Dia da semana (CHEN et al., 2016).
- b) Hora do dia (CHEN et al., 2016).

² Stopwords do português brasileiro obtidas da NLTK — <<https://www.nltk.org>>. Acessado em 19 de Abril de 2018.

³ RSLP Stemmer — <http://www.nltk.org/_modules/nltk/stem/rslp.html>. Acessado em 19 de Abril de 2018.

3. Eventos sociais.

- a) Feiras de rua (CHEN et al., 2016).
- b) Festivais (CHEN et al., 2016), (LECUE et al., 2014).
- c) Jogos esportivos (CHEN et al., 2016), (GAL-TZUR et al., 2014).
- d) Passeatas e maratonas (CHEN et al., 2016), (ITOH et al., 2016).

4. Eventos urbanos.

- a) Relacionados ao tráfego (CHEN et al., 2016), (LECUE et al., 2014).

5. Desastres naturais.

- a) Tempestades (ITOH et al., 2016).
- b) Terremoto (ITOH et al., 2016).
- c) Tufões (ITOH et al., 2016).

6. Metereológicos.

- a) Dia claro, nublado, chuvoso, nevando, com neblina (CHEN et al., 2016).
- b) Temperatura do ar (CHEN et al., 2016).

Após o estudo do domínio do conhecimento, por meio da revisão sistemática para coletar as classes de exceção, o Corpus Twitter, contendo 60.984, foi classificado manualmente de acordo com suas respectivas classes. Tal conjunto foi usado para treinar o modelo de classificação de tweets em classes de eventos de exceção.

6.5 Modelo de classificação de tweets relacionados a eventos de exceção

O corpus obtido da fase de processamento de tweets é representado por meio de um *bag-of-words*, que contém vetores de *features* criados usando a medida *Term Frequency - Inverse Document Frequency* (TF-IDF). A bag-of-words é particionada aleatoriamente em conjuntos de treinamento (60%) e teste (40%), os quais são entradas para os algoritmos de classificação mencionados na Seção 2.6.1.

6.6 Encontrando linhas de ônibus afetadas por eventos de exceção

Para encontrar as linhas de ônibus afetadas por eventos de exceção, é necessário correlacionar latitude e longitude dos eventos de exceção com as *stops*

da GTFS da SPTrans. Como mencionado anteriormente, os dados referentes as *stops* contém os locais individuais em que os veículos pegam ou deixam passageiros, incluindo coordenadas de latitude e longitude.

De acordo com a Seção 4.1, todas as coordenadas são armazenadas em pares no formato *legacy* e em coleções com índices geoespaciais. Assim, é possível usar a função `$near` do MongoDB⁴ para encontrar as *stops* próximas às coordenadas do evento de exceção. Como consequência da GTFS, o *stop_id* faz parte dos atributos contidos no arquivo de *stops*, referindo-se a um código de parada de ônibus com o qual é possível correlacioná-lo com as bases *stop_times* e *lines* (por meio do atributo *trip_id* existente em *stops*) para obter mais detalhes sobre a direção da linha de ônibus, identificação, etc.

6.7 Resultados

A metodologia foi aplicada ao Corpus Twitter⁵. No final do pré-processamento e processamento dos *tweets*, o corpus obteve 414,637 palavras, com um vocabulário de 13,915 palavras. O comprimento máximo das sentenças do conjunto de dados é 19, sua respectiva variação é ilustrada pela Figura 12.

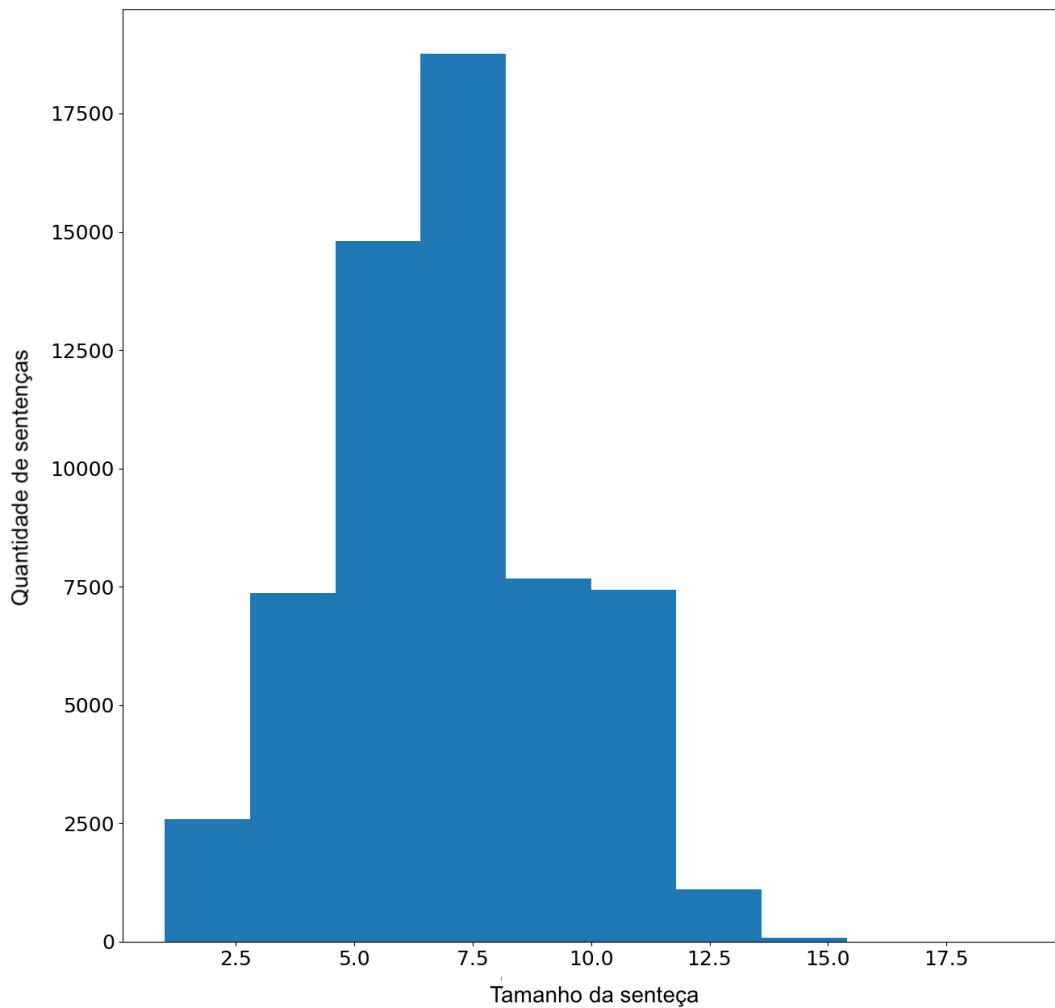
Todos os *tweets* existentes no *Corpus Twitter* foram classificados manualmente de acordo com os eventos de exceção identificados. Este conjunto de dados é composto pelas seguintes classes: Acidente, Irrelevante — quando o *tweet* não é um evento de exceção, Desastre Natural, Evento Social e Evento Urbano. A Figura 13 ilustra a distribuição das classes de eventos de exceção em cada conta selecionada.

Esse conjunto de dados rotulado foi usado para treinar modelos de classificação de eventos de exceção, com base em uma *bag-of-words*, descrita na Seção 6.5. De acordo com a Tabela 8, o modelo que usa o algoritmo *Multi-layer Perceptron* obteve maior acurácia para a tarefa de classificar os *tweets* em eventos de exceção. A matriz de confusão relacionada ao algoritmo de *Multi-layer Perceptron* é

⁴ <<https://docs.mongodb.com/manual/reference/operator/query/near/>>. Acessado em 18 de Maio de 2018.

⁵ Conjunto de dados disponível em: <<https://drive.google.com/drive/folders/16NIevLsBR0A45UHdPDvv2lZZx6gF4R0p?usp=sharing>>. Acessado em 8 de Setembro de 2018.

Figura 12 – Histograma da variação dos tamanhos das sentenças dos tweets existentes no *Corpus Twitter*



ilustrada pela Figura 14, as matrizes de confusão dos demais algoritmos podem ser consultadas no apêndice E.

Dos 60.984 tweets 10.027 foram classificados manualmente em eventos de exceção e desse subconjunto encontrados 8.112 endereços, de acordo com a Tabela 9 — desconsiderando o tipo de localidade APPROXIMATE — (o que representa 80,90% do total dos tweets classificados como eventos de exceção, sem considerar a classe *Irrelevant*). A quantidade de endereços extraídos por classe está descrita na Tabela 9, as razões para tweets sem endereço extraído são:

Figura 13 – Distribuição das classes dos eventos de exceção do Corpus Twitter

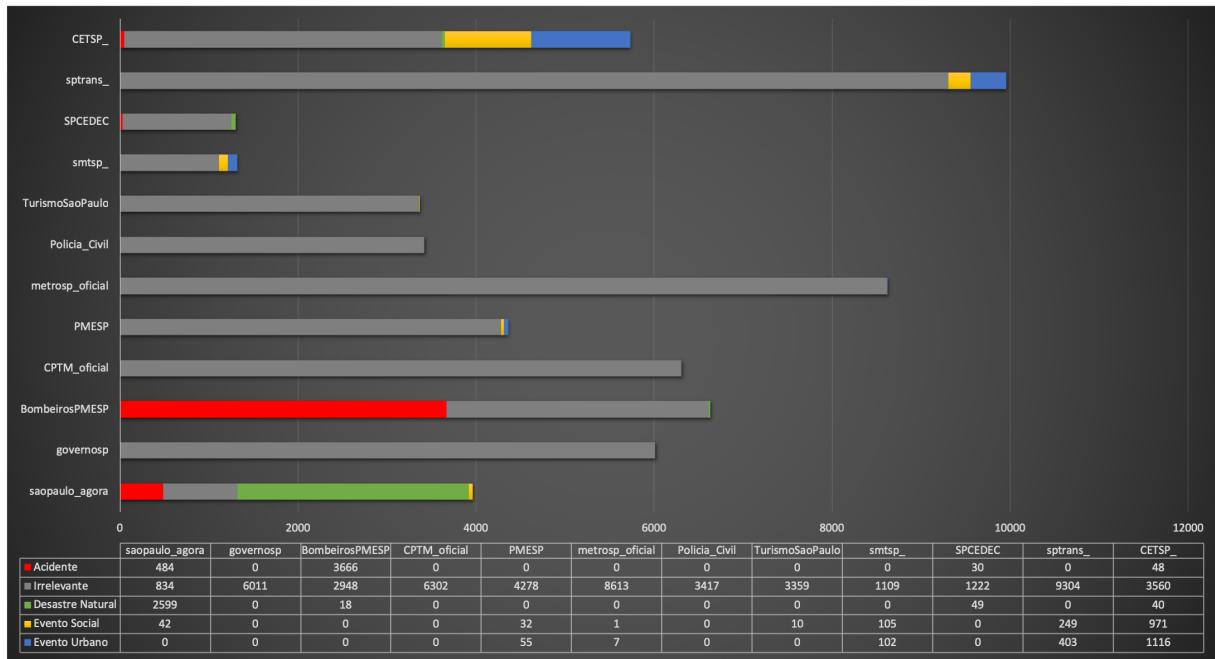


Tabela 8 – Métricas das avaliações dos algoritmos utilizados para classificação dos tweets em eventos de exceção

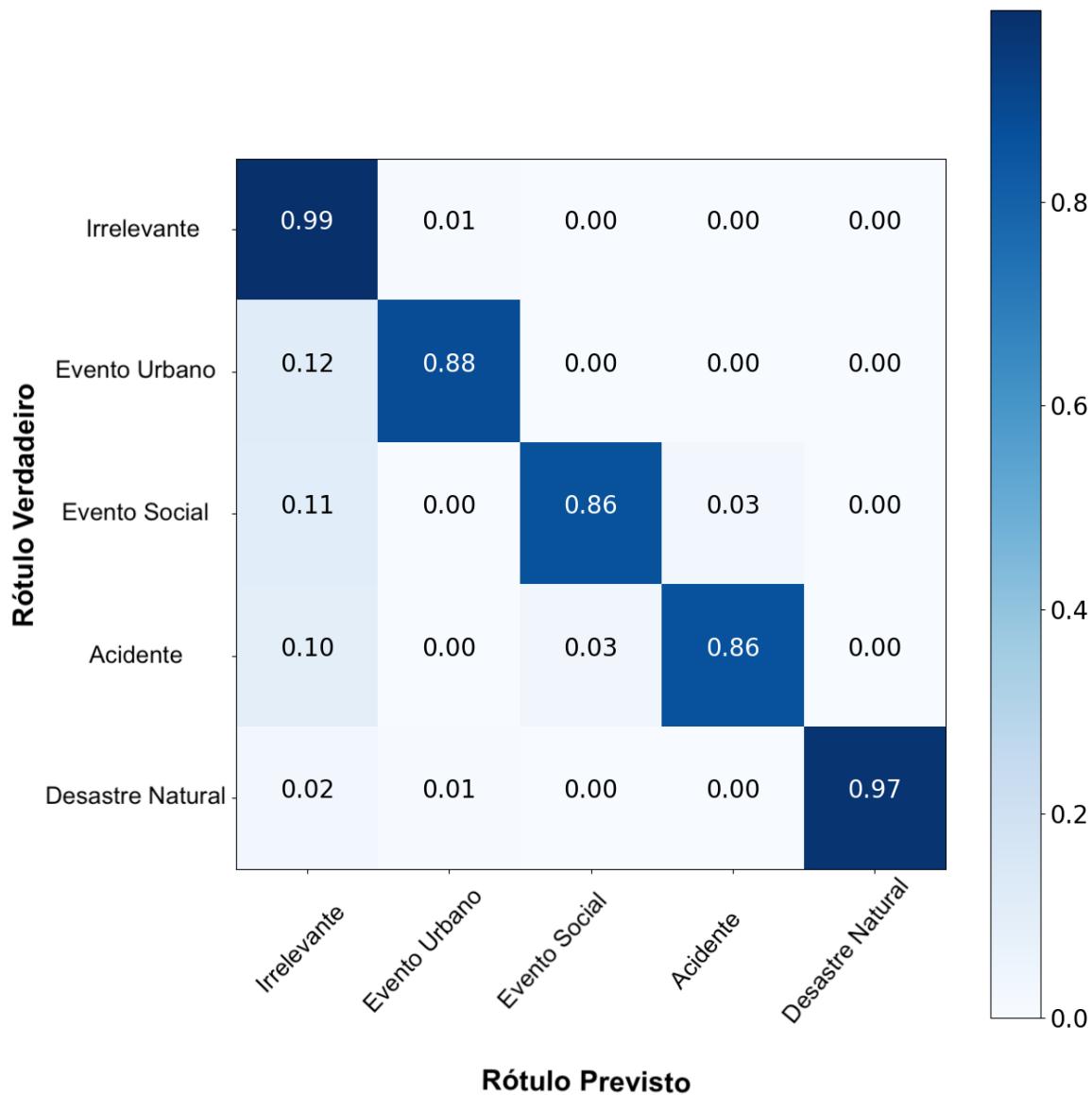
| Algoritmo | ACC | PPV | TPR | f1-score |
|-------------------------|-------|-------|-------|----------|
| Complement Naive Bayes | 0,941 | 0,949 | 0,941 | 0,944 |
| Decision Tree | 0,965 | 0,965 | 0,965 | 0,965 |
| K-Nearest Neighbors | 0,970 | 0,971 | 0,970 | 0,970 |
| Logistic Regression | 0,969 | 0,968 | 0,969 | 0,968 |
| Multi-layer Perceptron | 0,973 | 0,972 | 0,973 | 0,972 |
| Multinomial Naive Bayes | 0,953 | 0,952 | 0,953 | 0,949 |
| Random Forest | 0,970 | 0,970 | 0,970 | 0,970 |
| Support Vector Machine | 0,833 | 0,694 | 0,833 | 0,757 |

1. Tweets apenas com o ponto de interesse, ou seja, não consta explicitamente o endereço.
2. Tweets sem informação de endereço.
3. Tweets com nome de logradouro incomum (por exemplo *passagem*, *complexo viário*, *ligação sentido*).
4. Tweets com endereços com palavras concatenadas (por exemplo *avenidapaulista*).

Os tipos de localidades⁶ podem ser classificados em:

⁶ Disponível em <<https://developers.google.com/maps/documentation/geocoding>>. Acessado em 16 de setembro de 2018.

Figura 14 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo *Multi-layer Perceptron*



1. *ROOFTOP* — Indica que o resultado retornado há informações de localização com precisão a nível do endereço de rua.
2. *RANGE_INTERPOLATED* — Indica que o resultado retornado reflete uma aproximação interpolada entre dois pontos precisos (como interseções). Geralmente, os resultados interpolados são retornados quando os códigos geográficos do *rooftop* não estão disponíveis para um endereço de rua.
3. *GEOMETRIC_CENTER* — Indica que o resultado retornado é o centro geométrico de um resultado.

4. APPROXIMATE — Indica que o resultado retornado é aproximado.

Neste estudo de caso, desconsideramos os endereços com classificação *APPROXIMATE*, devido ao fato de poderem comprometer a confiabilidade das análises realizadas.

Tabela 9 – Quantidade de eventos extraídos por classe

| Classe | #endereços extraídos ^a | #APP ^b | #GEO ^c | #RANGE ^d | #ROOF ^e |
|------------------|-----------------------------------|-------------------|-------------------|---------------------|--------------------|
| Acidente | 3.439 | 7 | 805 | 1.130 | 1.497 |
| Irrelevante | 451 | 13 | 292 | 6 | 140 |
| Desastre Natural | 2.464 | 9 | 340 | 719 | 1.396 |
| Evento Social | 793 | 4 | 761 | 2 | 26 |
| Evento Urbano | 1.002 | 4 | 942 | 10 | 46 |
| Total | 8.149 | 37 | 3.140 | 1.867 | 3.105 |

^a Total de endereços extraídos

^b Total de endereços extraídos com o tipo de localidade *APPROXIMATE*

^c Total de endereços extraídos com o tipo de localidade *GEOMETRIC_CENTER*

^d Total de endereços extraídos com o tipo de localidade *RANGE_INTERPOLATED*

^e Total de endereços extraídos com o tipo de localidade *ROOFTOP*

A Figura 15 ilustra os endereços⁷ mais afetados por eventos de exceção e a Figura 16 parte da distribuição desses eventos na região central de São Paulo. É importante ressaltar que os eventos de exceção encontrados estão concentrados em endereços e regiões onde normalmente ocorrem em São Paulo, o que valida a metodologia desenvolvida.

Consideramos que uma linha de ônibus é afetada por um evento de exceção se uma *stop* estiver dentro de um raio de 1000 metros de distância do evento. Utilizando este critério, o total de 992 linhas de ônibus foram afetadas por eventos de exceção durante este período, sendo “33389” o código de linha de ônibus mais impactado. Essa linha específica foi impactada por 1.301 eventos de exceção. A Tabela 10 lista as linhas de ônibus que foram impactadas por mais de 600 eventos de exceção.

⁷ Lista completa está disponível em <<https://docs.google.com/spreadsheets/d/1gn1cTDifUJEPdgcU67SC45GdYHRKmIHtAfJwRBm088s/edit?usp=sharing>>. Acessado em 09 de setembro de 2018.

Figura 15 – Endereços mais impactados por eventos de exceção

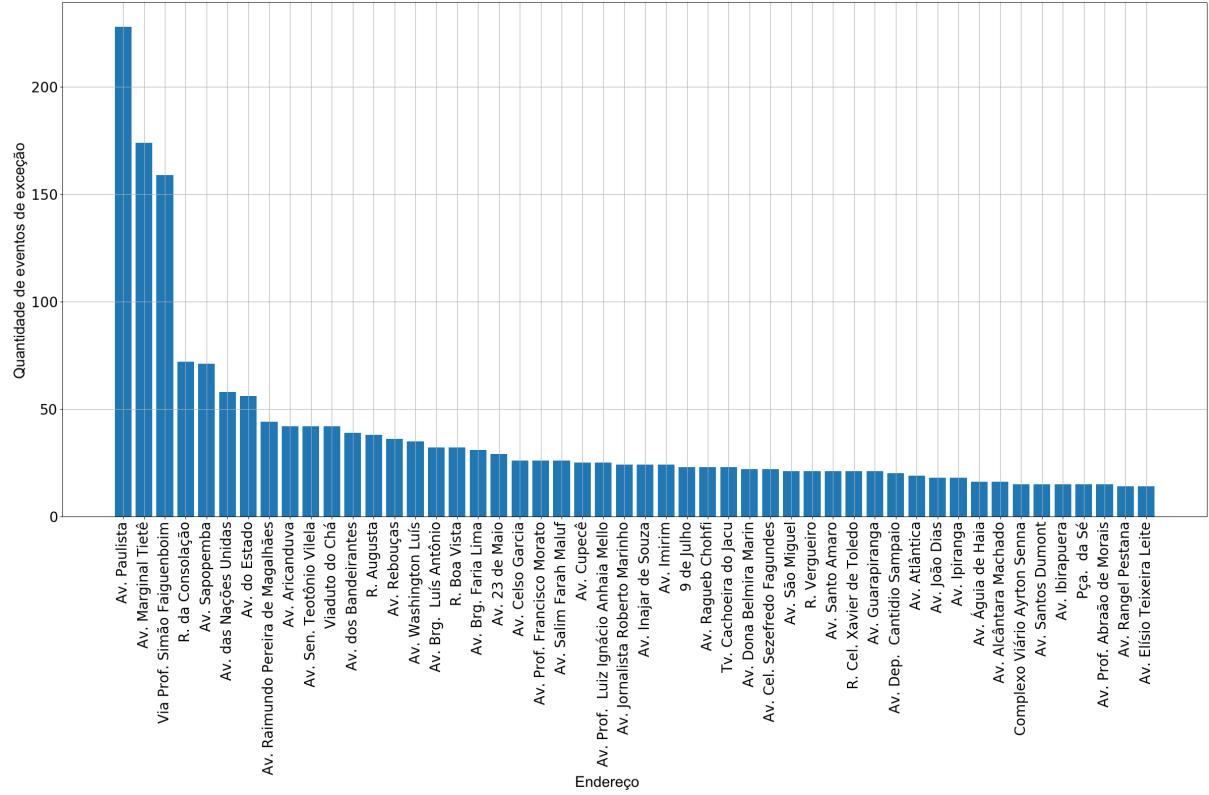


Figura 16 – Distribuição dos eventos de exceção na região central de São Paulo

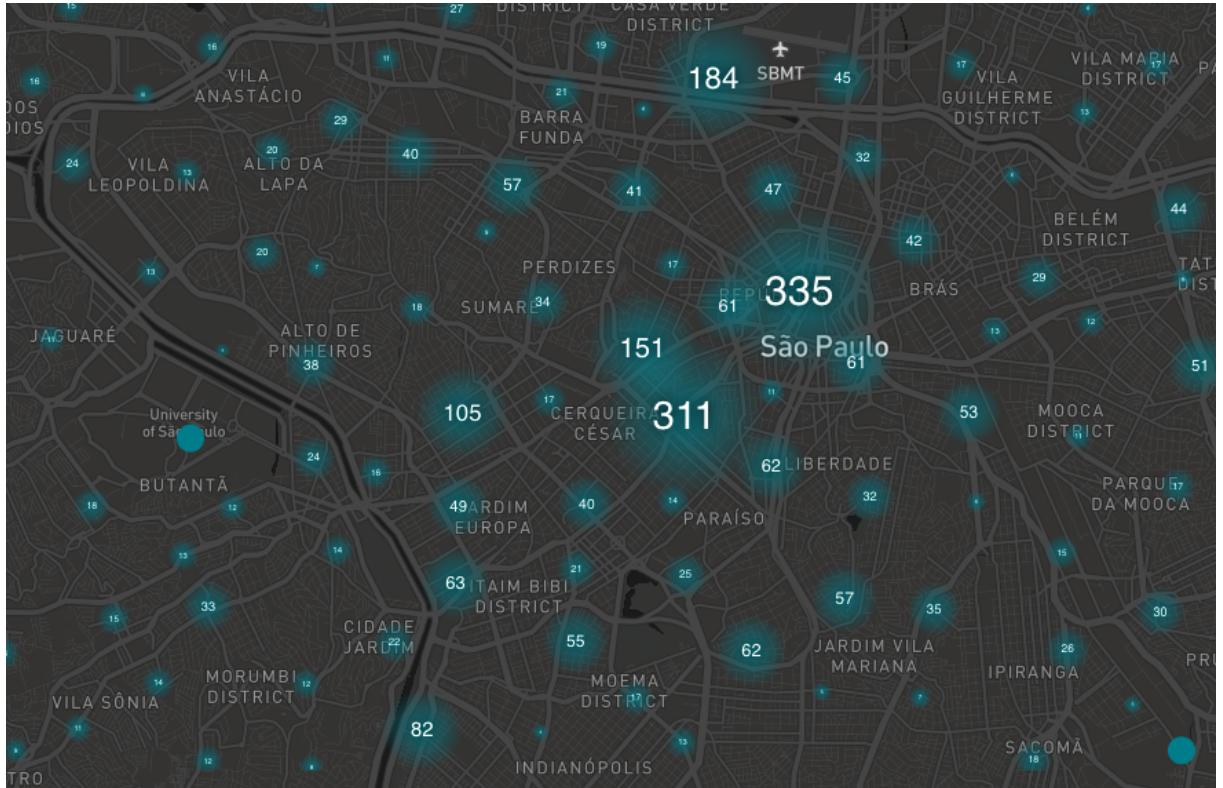


Tabela 10 – Linhas de ônibus mais impactadas por eventos de exceção^a

| Código da linha | # eventos de exceção | Leteiro |
|-----------------|----------------------|--|
| 33389 | 1301 | TERM. PINHEIROS / METRÔ TUCURUVI |
| 33284 | 1176 | ITAIM BIBI / METRÔ SANTANA |
| 33121 | 1023 | TERM. PRINC. ISABEL / TERM. STO. AMARO |
| 32805 | 1006 | TERM. PRINC. ISABEL / CHÁC. SANTANA |
| 33112 | 933 | TERM. PQ. D. PEDRO II / JD. SÃO SAVÉRIO |
| 33111 | 857 | TERM. AMARAL GURGEL / JD. DA SAÚDE |
| 35229 | 841 | TURISMO / CIRCULAR |
| 33443 | 816 | ANA ROSA / METRÔ SANTANA |
| 32897 | 805 | LUZ / TERM. A. E. CARVALHO |
| 35072 | 767 | METRÔ BARRA FUNDA / CONEXÃO PETRÔNIO PORTELA |
| 32772 | 759 | TERM. PRINC. ISABEL / TERM. STO. AMARO |
| 33253 | 754 | METRÔ BELÉM / JD. BONFIGLIOLI |
| 33391 | 748 | METRÔ JABAQUARA / METRÔ SANTANA |
| 32813 | 746 | PÇA. DA SÉ / CHÁC. SANTANA |
| 32829 | 746 | TERM. BANDEIRA / TERM. CAPELINHA |
| 34048 | 719 | LGO. SÃO FRANCISCO / JD. SELMA |
| 33486 | 715 | TERM. PQ. D. PEDRO II / TERM. SÃO MATEUS |
| 33236 | 708 | TERM. BANDEIRA / JD. JAQUELINE |
| 33336 | 697 | PINHEIROS / IMIRIM |
| 32816 | 693 | TERM. PQ. D. PEDRO II / TERM. STO. AMARO |
| 33534 | 690 | CARDOSO DE ALMEIDA / MACHADO DE ASSIS |
| 32838 | 647 | PÇA. DA SÉ / PQ. RES. COCAIA |
| 33398 | 639 | CID. UNIVERSITÁRIA / METRÔ SANTANA |
| 32769 | 638 | LGO. SÃO FRANCISCO / TERM. CAPELINHA |
| 33114 | 637 | TERM. PINHEIROS / SACOMÃ |
| 34210 | 637 | LGO. SÃO FRANCISCO / TERM. VARGINHA |
| 33116 | 625 | RIO PEQUENO / IPIRANGA |
| 33126 | 614 | TERM. BANDEIRA / INOCOOP CAMPO LIMPO |

^a Tabela completa no apêndice D.

6.8 Considerações finais sobre a metodologia desenvolvida

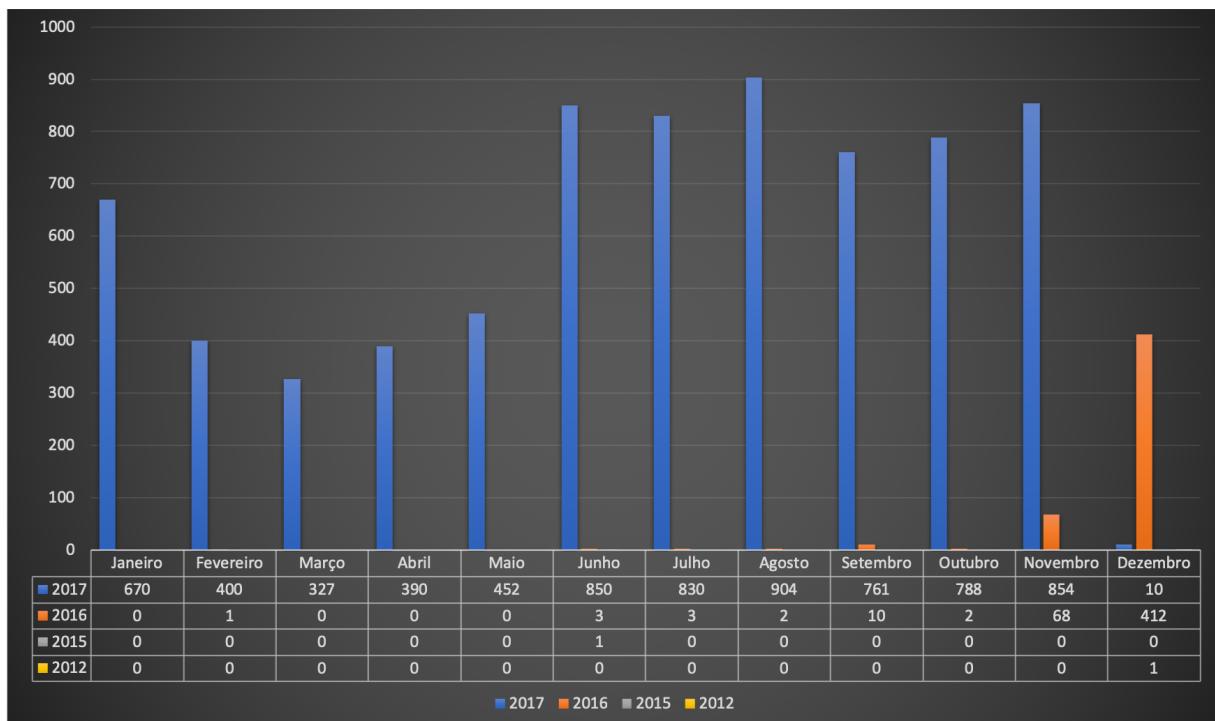
Este experimento apresenta uma nova metodologia para classificação de eventos de exceção e analisa seus respectivos impactos no sistema de transporte coletivo por ônibus da cidade de São Paulo. De acordo com os experimentos realizados, o algoritmo com maior acurácia para classificação de tweets em eventos de exceção foi *Multi-layer Perceptron*. Também, mostramos que é possível extrair endereços de tweets semi-estruturados usando apenas expressões regulares. A classificação desses eventos é o primeiro passo para entender melhor como os eventos de exceção afetam a rede de transporte público.

Embora o método tenha sido validado usando perfis selecionados do Twitter escritos em português brasileiro, o mesmo pode ser generalizado para diferentes idiomas e cidades. A GTFS é um formato ubíquo para o transporte público e ferramentas como a NLTK suporta vários idiomas.

7 Correlação dos eventos de exceção com os dados AVL da SPTrans

Dado que os eventos de exceção podem ser identificados utilizando *tweets* dos perfis contidos na Tabela 1, há também a possibilidade de caracterizarmos seus respectivos impactos em relação as velocidades medianas dos ônibus, por meio da base histórica dos dados AVL da SPTrans. Neste estudo consideramos os eventos de exceção geolocalizados do ano de 2017 do *Corpus Twitter*, classificados manualmente. A distribuição desses eventos ao longo dos meses pode ser observada na Figura 17, assim como de suas respectivas classes na Figura 18.

Figura 17 – Distribuição do número de eventos de exceção geolocalizados



Conforme descrito no Capítulo 6 e ilustrado na Figura 19, identificamos as linhas afetadas por eventos de exceção filtrando as paradas de ônibus (contidas na coleção *stops* e *shapes* da GTFS da SPTrans) dentro de um raio de 1.000 metros das geolocalizações extraídas dos *tweets*.

A partir disso, selecionamos os dados de movimentação que serão analisados, por exemplo, se o evento de exceção aconteceu em 17/08/2017 (quinta-feira), todas as demais quintas-feiras do mês de agosto (3, 10, 24 e 31/08/2017) serão consideradas na análise. Isso, porque os dias da semana possuem padrões diferentes de

movimentação (sazonalidade), por exemplo, nas sextas feiras ocorrem inúmeros eventos sociais que normalmente acarretam em um trânsito mais congestionado. Além da diferenciação dos dias da semana, os meses também possuem características distintas — festas no final do ano, férias, início de períodos letivos, etc. — conforme Figura 18, devido a isso os dias selecionados são restritos ao mês de ocorrência do evento.

Figura 18 – Distribuição das classes de eventos de exceção geolocalizados ao longo dos meses do ano de 2017

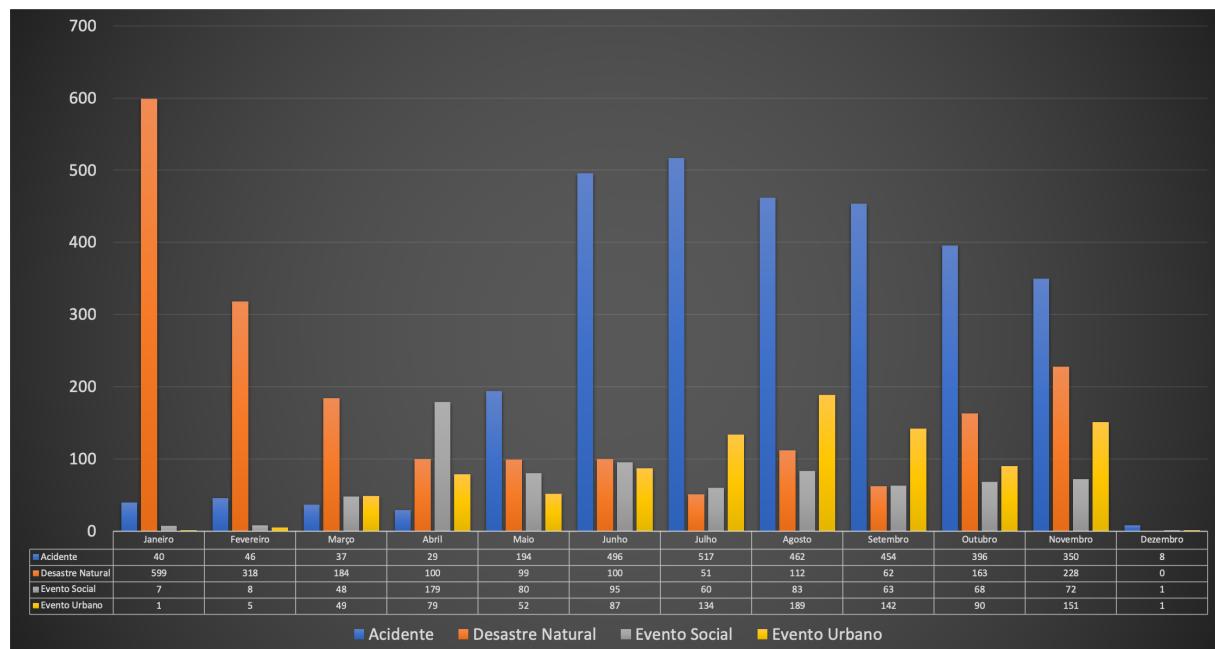
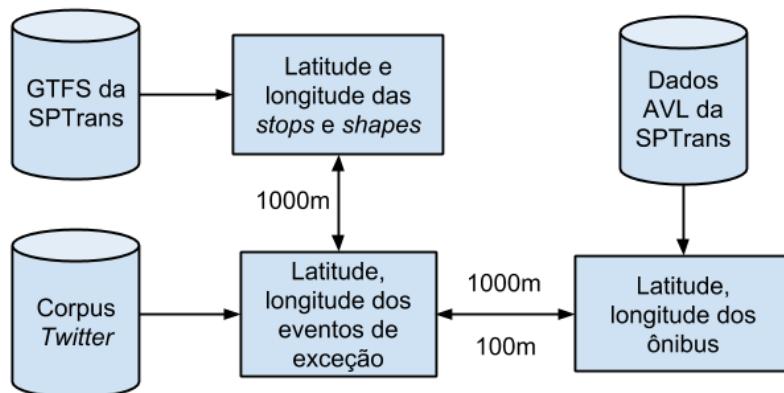


Figura 19 – Processo para correlação entre os dados AVL, GTFS e tweets para análise do impacto dos eventos de exceção



Seleções de dias e meses à parte, também filtramos os dados relacionados as linhas impactadas a um raio de distância de 100 e 1.000 metros do evento de

exceção em questão, além de considerarmos a mesma faixa de horário do *tweet*. Por exemplo, se o horário do *tweet* é às 17h15min, consideramos os dados AVL com horário entre 17 e 18h. É importante observar que esse trabalho não considera início e término exato dos eventos de exceção, mas uma faixa de horário de uma hora a partir da hora contida no *timestamp* do *tweet*.

Em seguida, agregamos os dados selecionados para analisarmos de forma descritiva a velocidade instantânea de cada linha de ônibus, com isso extraímos dados sobre a velocidade máxima, mínima, média, mediana, variância, desvio padrão e porcentagem de dados iguais e diferentes de zero. Após isso, comparamos a velocidade média das medianas da faixa de horário da ocorrência com a velocidade média das medianas dos dias que não fazem referência ao evento de exceção, para cada conjunto de linhas afetadas pelo evento de exceção e para cada linha.

Consideramos que a linha foi impactada se a média das velocidades medianas dos dias analisados for maior ou igual a velocidade mediana do dia referente ao evento de exceção. Com base nisso, consideramos que o conjunto de linhas foi impactado se a quantidade de linhas impactadas for maior ou igual a 50%.

Resultados

Análise da redução da velocidade mediana dos ônibus a partir das informações de latitude e longitude dos pontos de parada

Utilizando a metodologia anteriormente descrita, podemos observar na Tabela 11 que os eventos de exceção relacionados a eventos sociais possuem em média 87,04% de impacto na velocidade mediana nos grupos de linhas de ônibus afetadas a um raio de 1.000 metros de distância e 100% a um raio de 100 metros, isso provavelmente devido ao grande número de pessoas envolvidas neste tipo de evento, quantidade de avenidas com fluxo do trânsito modificado ou interrompido.

Os eventos urbanos, por sua vez, impactam em 70,11% a 1.000 metros e 98,86% a 100 metros, mesmo sendo realizados com planejamento de rotas alternativas e sinalizações nas vias públicas. A terceira e quarta classe mais afetadas são as de acidentes e desastres naturais, respectivamente, 66,51% e 59,77% a 1.000 metros e 98,39% e 99,80% a 100 metros, as quais normalmente resultam em bloqueios ou desvios em vias públicas utilizadas pelos ônibus.

Além disso, janeiro, fevereiro e março foram os três meses mais afetados por eventos de exceção relacionados a destrases naturais, período de grandes volumes de precipitação de chuva em São Paulo, no qual normalmente ocorre deslizamentos de terra, quedas de árvores e inundações. Em relação aos eventos sociais, o ano de 2017 foi marcado com inúmeras manifestações políticas, neste contexto, o mês de maio foi o mais impactado por esse tipo de evento de exceção, principalmente devido aos protestos contra o governo Temer¹. Os eventos relacionados a acidentes normalmente ocorrem em maior concentração nos períodos de festas e feriados, o que pode ser observado nos meses de janeiro e abril (único mês de 2017 com dois feriados prolongados), com média de impacto de 83,33% e 87,50% nas velocidades médias, respectivamente. Os impactos relacionados a eventos urbanos ocorrem normalmente durante todos os meses, devido a isso são mais uniformes.

Os meses próximos a 100% de impacto nas velocidades médias são justificados devido ao pouco volume de eventos para uma determinada classe em um determinados mês, conforme Figura 18, o que também acontece para cenários com dados geolocalizados muito próximos aos eventos de exceção. Analogamente, os meses e classes sem dados de impacto são meses com pouco dados para a classe em questão.

Análise da redução da velocidade mediana dos ônibus a partir das informações de latitude e longitude das rotas das linhas

As análises dos dados da Tabela 12 são análogas as da Tabela 11 da Seção 7, para o raio de 1.000 m os eventos que mais reduzem as velocidades medianas são os relacionados a eventos sociais, urbanos, acidentes e desastres naturais, nesta ordem. Apesar disso, é importante observar que o percentual de redução da velocidade mediana é reduzido quando consideramos as latitudes e longitudes das rotas como referência para encontrar as linhas impactadas.

Quando as rotas são consideradas obtemos um conjunto maior de linhas impactadas, o que aumenta a margem de erro, adicionando ruído a análise realizada, diferentemente da análise que considera as latitudes e longitudes dos pontos de

¹ <http://www1.folha.uol.com.br/poder/2017/05/1884977-manifestacao-anti-temer-reune-centenas-de-pessoas-na-av-paulista.shtml>. Acessado em 02 de dezembro de 2018

Tabela 11 – Porcentagem de ônibus dos grupos de linhas afetadas por eventos de exceção, a 1.000 m e 100 m de distância, respectivamente, que tiveram a velocidade mediana reduzida nos meses do ano de 2017

| Mês | Acidente | Desastre Natural | Evento Social | Evento Urbano |
|-----------|----------|------------------|---------------|---------------|
| Janeiro | 83,33 | 100 | 64,23 | 98,00 |
| Fevereiro | 70,58 | 100 | 66,25 | 100 |
| Março | 50,00 | — | 66,66 | 100 |
| Abril | 87,50 | 100 | 61,11 | 100 |
| Maio | 65,13 | 100 | 58,82 | 100 |
| Junho | 54,46 | 100 | 61,53 | 100 |
| Julho | 61,48 | 98,41 | 66,66 | 100 |
| Agosto | 57,86 | 87,17 | 55,35 | 100 |
| Setembro | 64,21 | 100 | 42,10 | 100 |
| Outubro | 70,49 | — | 56,81 | — |
| Novembro | 66,66 | 100 | 57,99 | 100 |
| Dezembro | — | — | — | — |
| Total | 66,51 | 98,39 | 59,77 | 99,80 |
| | | | 87,04 | 100 |
| | | | 70,11 | 98,86 |

parada dos ônibus como referência para encontrar as linhas afetadas pelos eventos de exceção. Outra diferença observada é em relação aos percentuais de redução de velocidades medianas para o raio de 100 m. Nesta abordagem, as ordem dos eventos que mais impactam as velocidades medianas são os relacionados a acidentes, desastres naturais, eventos sociais e urbanos, respectivamente.

Em relação às sazonalidades, os meses de março, abril, maio e outubro foram mais significativos para a redução das velocidades medianas, no raio de 1.000 m devido as inúmeras manifestações^{2,3,4,5,6} que ocorreram no Brasil. Sobre os desastres naturais, os impactos foram relevantes para os meses de janeiro a março, a distância de 100 m, conforme esperado devido ao período de chuvas.

Por fim, a quantidade de latitudes e longitudes relacionadas as paradas de ônibus é menor do que o que contém informações sobre as geolocalizações que formam as rotas. Consequentemente, de acordo com os experimentos, a abordagem

² <<https://g1.globo.com/politica/noticia/cidades-pelo-pais-tem-manifestacoes-a-favor-da-lava-jato-neste-domingo.ghtml>>. Acessado em 14 de janeiro de 2019.

³ <<https://g1.globo.com/resumo-do-dia/noticia/quarta-feira-15-de-marco-de-2017.ghtml>>. Acessado em 14 de janeiro de 2019.

⁴ <<https://www1.folha.uol.com.br/poder/2017/03/1866022-manifestacao-por-intervencao-militar-bloqueia-via-em-sp.shtml>>. Acessado em 14 de janeiro de 2019.

⁵ <<https://oglobo.globo.com/brasil/ato-de-artistas-no-rio-contra-temer-termina-com-bombas-de-efeito-moral-spray-de-pimenta-21987385>>. Acessado em 14 de janeiro de 2019.

⁶ <https://pt.wikipedia.org/wiki/Greve_geral_no_Brasil_em_2017>. Acessado em 14 de janeiro de 2019.

que utiliza o primeiro conjunto consome menos recursos computacionais, com menor ruído (falsos positivos) e resultados semelhantes ao da segunda abordagem.

Tabela 12 – Porcentagem de impacto na velocidade média dos grupos de linhas afetadas por eventos de exceção a 1.000 m e 100 m de distância, respectivamente, nos meses do ano de 2017

| Mês | Acidente | Desastre Natural | Evento Social | Evento Urbano |
|------------|-----------------|-------------------------|----------------------|----------------------|
| Janeiro | 66,66 | 100 | 47,68 | 78,49 |
| Fevereiro | 35,29 | 100 | 49,09 | 81,25 |
| Março | 66,66 | 100 | 42,85 | 62,5 |
| Abril | 62,50 | 60,00 | 47,05 | 100 |
| Maio | 49,09 | 77,77 | 64,70 | 100 |
| Junho | 47,78 | 79,76 | 46,15 | 70,00 |
| Julho | 44,85 | 75,55 | 66,66 | 83,33 |
| Agosto | 49,49 | 75,36 | 44,44 | 71,42 |
| Setembro | 49,47 | 79,16 | 36,84 | 54,54 |
| Outubro | 56,06 | 78,26 | 58,69 | 90,00 |
| Novembro | 54,32 | 66,66 | 44,00 | 74,07 |
| Dezembro | — | — | — | — |
| Total | 52,92 | 81,13 | 49,83 | 78,69 |
| | | | 79,51 | 77,95 |
| | | | 68,14 | 69,76 |

8 Identificação de padrões de velocidade média dos dados AVL

Neste capítulo, é apresentado um processo para identificação de padrões de velocidade média dos dados AVL, por meio do algoritmo *Apriori*. De acordo com (XIE et al., 2008), o algoritmo é ineficiente para grandes volumes de dados, devido a quantidade elevada agregações necessárias para calcular as métricas explicadas no Capítulo 2.8 e aos inúmeros acessos ao banco de dados. O foco deste trabalho não é melhorar o desempenho do algoritmo *Apriori* ou implementar as melhorias existentes na literatura (XIE et al., 2008; ZHANG et al., 2014), embora tenhamos como objetivo encontrar os padrões de velocidades médias existentes nos mais de um milhão de registros por hora, volume característico dos dados AVL.

Além da quantidade de registros total, o volume de dados para pequenos intervalos de tempo também é considerável, pois os módulos AVL enviam dados dos ônibus a todo instante. Dessa forma, para viabilizarmos o uso do algoritmo *Apriori*, agrupamos os dados por intervalos de tempo de cinco minutos e calculamos a velocidade média para cada intervalo. Com isso, executamos esse processo para cada mês do conjunto de dados AVL para determinarmos as velocidades médias e identificarmos os padrões existentes nos intervalos definidos.

Analogamente, o mesmo procedimento foi aplicado para os conjuntos de dados AVL correlacionados aos eventos de exceção (Acidente, *Urban Event*, *Natural Disaster* e *Social Event*), os dados anuais foram sintetizados nas tabelas 14 e 15 e os mensais nas seções G.2, G.3, G.4 e G.5. As regras de associação encontradas nestes conjuntos de dados estão disponíveis em DIAS (2017) (não inclusas no texto deste trabalho devido ao grande volume de dados).

8.1 Trabalhos relacionados

No trabalho realizado em (ZHAO et al., 2019), foram utilizados o algoritmo *Apriori* e a análise de *cluster* para encontrar padrões relacionados a transferência (entre metrô e ônibus), por meio dos dados dos cartões inteligentes usados no transporte público da China. Nesse estudo, encontraram que 85% dos resultados de reconhecimento de transferência são bastante estáveis durante toda a semana, e

o tempo médio de transferência entre o metrô e o ônibus é inferior a 20 minutos. O método proposto neste estudo pode ser usado para identificar os pontos de transferência mais movimentados e obter tempos médios de transferência, o que facilita uma rede de transporte público mais inteligente e eficiente.

Ainda relacionado a mobilidade urbana, o trabalho realizado em (ZENG et al., 2017) buscou compreender, por meio do algoritmo *Apriori*, os padrões existentes nos conjuntos de dados relacionados a movimentação diária no transporte público de Singapura e no *MIT Reality Mining* (dados sobre comunicação, proximidade, localização, etc. coletados entre Setembro de 2004 e Junho de 2005, dos celulares de voluntários do projeto *MIT Reality Mining Data*). O sistema desenvolvido é capaz de identificar e apresentar visualmente padrões de movimentação humana, em relação a espaço e tempo. Analogamente, o estudo realizado em (YU, 2018), é capaz de indentificar padrões de rotas de táxi, na cidade de Pequim, China.

Por fim, no trabalho realizado pro (CRUZ et al., 2018), propos uma metodologia para identificar e classificar as anomalias no comportamento do trânsito, por meio de agregações espaço-temporais usando o algoritmo *Apriori*, aplicadas aos dados de transporte rodoviário da cidade do Rio de Janeiro. A metodologia proposta foi capaz identificar características das principais anomalias e classificá-las como esperadas ou inesperadas. A proposta desse experimento se diferencia das demais por encontrar os padrões de velocidade média existentes nos dados do transporte público por ônibus da cidade de São Paulo, considerando ainda a correlação com eventos de exceção extraídos de Redes Sociais.

8.2 Resultados

A Tabela 13 é referente aos padrões encontrados com valores de *Lift* > 1, métrica que indica correlações entre dois valores. Mais de um padrão de associações entre velocidades médias foi encontrado para a maioria dos meses, exceto para os meses de janeiro {11 → 12} (aceleração), julho {11 → 12} (aceleração) e dezembro {12 → 11} (desaceleração), meses nos quais normalmente o trânsito é menos congestionado e mais estável, devido as férias escolares. Apesar disso, em setembro {12 → 11} (desaceleração) foi identificado apenas um padrão. Tais padrões indicam

correlações de velocidades médias nesses meses a cada cinco minutos entre 11 e 12Km/h.

Por sua vez, os meses com menores velocidades médias no intervalo de cinco minutos foram fevereiro $\{7 \rightarrow 8\}$ (aceleração), abril $\{7 \rightarrow 8\}$ (aceleração), maio $\{7 \rightarrow 8\}$ (aceleração), outubro $\{8 \rightarrow 7\}$ (desaceleração) e novembro $\{8 \rightarrow 7\}$ (desaceleração). Ou seja, nesses meses as correlações de velocidades médias no intervalo de estudo foram entre 7 e 8Km/h.

Referente ao demais meses, junho teve médias entre $\{11 \rightarrow 12\}$ e $\{12 \rightarrow 13\}$ (aceleração); agosto $\{11 \rightarrow 12\}$ (aceleração) e $\{13 \rightarrow 12\}$ (desaceleração); outubro $\{11 \rightarrow 12\}$ (aceleração) e $\{13 \rightarrow 12\}$ (desaceleração); novembro $\{12 \rightarrow 11\}$ (desaceleração). Tais padrões indicam correlações de velocidades médias a cada cinco minutos entre 11 e 13Km/h.

Os valores de *Support* indicados na Tabela 13 representam uma baixa frequência dos padrões encontrados, apesar das correlações existentes entre eles. Com os eventos de exceção em consideração, encontramos 585.804 regras de associação — correlacionando com os eventos de exceção a 100 metros de distância dos pontos de parada de ônibus — e 9.348.802 — correlacionando com os eventos de exceção a 1.000 metros de distância dos pontos de parada de ônibus — detalhadas na Tabela 14 e distribuídas graficamente (as regras de associação inesperadas) na Figura 20.

Analogamente, encontramos 7.857.504 regras de associação — correlacionando com os eventos de exceção a 100 metros de distância dos pontos de parada de ônibus — e 6.296.140 — correlacionando com os eventos de exceção a 1.000 metros de distância dos pontos de parada de ônibus — detalhadas na Tabela 15 e distribuídas graficamente (as regras de associação inesperadas) na Figura 21. A quantidade de regras de associação inesperadas em relação a sazonalidade é equivalente as análises realizadas no Capítulo 7.

Tabela 13 – Análise *Apriori*^a aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans

| Mês | Regra de associação | Support | Confidence | Lift |
|-----------|---------------------|---------|------------|-------|
| Fevereiro | 7 → 8 | 0,101 | 0,496 | 3,586 |
| Abril | 7 → 8 | 0,108 | 0,456 | 3,188 |
| Maio | 7 → 8 | 0,108 | 0,570 | 4,375 |
| Outubro | 8 → 7 | 0,100 | 0,595 | 3,433 |
| Novembro | 8 → 7 | 0,104 | 0,446 | 3,369 |
| Janeiro | 11 → 12 | 0,137 | 0,476 | 1,729 |
| Junho | 11 → 12 | 0,129 | 0,632 | 1,656 |
| Julho | 11 → 12 | 0,204 | 0,694 | 1,934 |
| Agosto | 11 → 12 | 0,169 | 0,670 | 1,662 |
| Outubro | 11 → 12 | 0,119 | 0,601 | 1,669 |
| Fevereiro | 12 → 11 | 0,126 | 0,582 | 1,770 |
| Março | 12 → 11 | 0,134 | 0,621 | 1,627 |
| Abril | 12 → 11 | 0,123 | 0,601 | 2,013 |
| Maio | 12 → 11 | 0,137 | 0,645 | 1,703 |
| Setembro | 12 → 11 | 0,163 | 0,608 | 1,863 |
| Novembro | 12 → 11 | 0,154 | 0,531 | 1,875 |
| Dezembro | 12 → 11 | 0,143 | 0,432 | 2,073 |
| Fevereiro | 12 → 13 | 0,123 | 0,375 | 1,956 |
| Março | 12 → 13 | 0,158 | 0,415 | 1,766 |
| Junho | 12 → 13 | 0,141 | 0,370 | 1,907 |
| Abril | 13 → 12 | 0,109 | 0,367 | 2,280 |
| Maio | 13 → 12 | 0,161 | 0,425 | 1,942 |
| Agosto | 13 → 12 | 0,147 | 0,366 | 1,830 |
| Outubro | 13 → 12 | 0,150 | 0,417 | 1,737 |

^a Tabela completa na Seção G.1.

Figura 20 – Velocidades inesperadas dos ônibus impactados por eventos de exceção a 100 m e 1.000 m dos pontos de parada, ao longo dos meses do ano de 2017

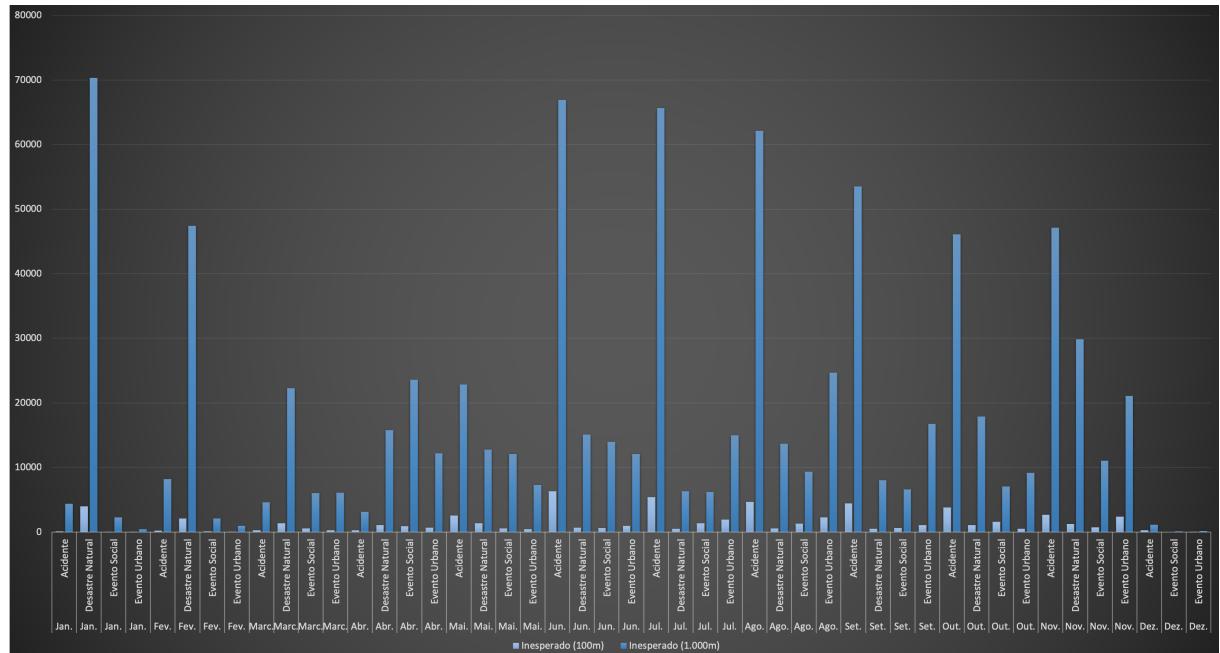


Figura 21 – Velocidades inesperadas dos ônibus impactados por eventos de exceção a 100 m e 1.000 m dos pontos de rota, ao longo dos meses do ano de 2017

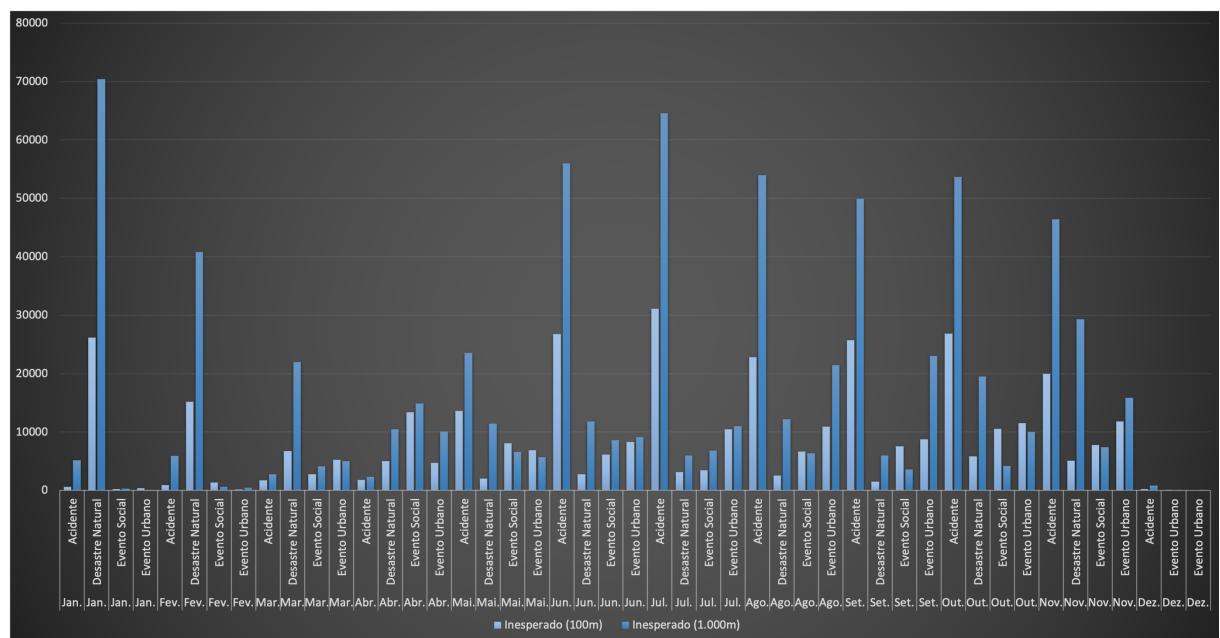


Tabela 14 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados aos eventos de exceção (a distância de 100 m^f e 1.000 m^g, respectivamente, dos pontos de parada de ônibus) dos meses do ano de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 1.677 | 315.063 | 278.493 | 30.804 | 5.766 |
| Desastre Natural | 912 | 115.301 | 99.206 | 14.282 | 1.813 |
| Evento Social | 506 | 61.927 | 52.403 | 8.245 | 1.279 |
| Evento Urbano | 596 | 93.513 | 81.261 | 10.480 | 1.772 |
| — | 3.691 | 585.804 | 511.363 | 63.811 | 10.603 |

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 3.029 | 3.980.542 | 3.415.780 | 385.728 | 179.034 |
| Desastre Natural | 2.016 | 2.624.415 | 2.253.123 | 259.285 | 112.007 |
| Evento Social | 764 | 1.262.805 | 1.118.546 | 100.224 | 44.035 |
| Evento Urbano | 980 | 1.481.040 | 1.296.476 | 125.803 | 58.761 |
| — | 6.789 | 9.348.802 | 8.083.925 | 871.040 | 393.837 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras esperadas ($Lift > 1$, $Support > 0,05$)

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 3.545 eventos de exceção não atingiram linhas de ônibus no raio de 100 m.

^g 447 eventos de exceção não atingiram linhas de ônibus no raio de 1.000 m.

Tabela 15 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados aos eventos de exceção (a distância de 100 m^g e 1.000 m^h, respectivamente, dos pontos de rota dos ônibus) dos meses do ano de 2017

| Classe do Evento | Total de Eventos^b | Qtd. Regras de Associação^c | Esperadas^d | Não Esperadas^e | Parcialmente inesperadas^f |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 2.367 | 3.390.690 | 3.164.726 | 171.860 | 54.104 |
| Desastre Natural | 1.307 | 1.342.048 | 1.247.219 | 75.981 | 18.848 |
| Evento Social | 704 | 1.522.423 | 1.433.700 | 67.835 | 20.888 |
| Evento Urbano | 825 | 1.602.343 | 1.499.305 | 79.155 | 23.883 |
| Total | 5.203 | 7.857.504 | 7.344.950 | 394.831 | 117.723 |

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 3.035 | 2.772.368 | 2.259.806 | 365.234 | 147.328 |
| Desastre Natural | 2017 | 1.876.843 | 1.545.172 | 239.897 | 91.774 |
| Evento Social | 764 | 683.037 | 588.385 | 63.549 | 31.103 |
| Evento Urbano | 980 | 963.892 | 805.901 | 111.898 | 46.093 |
| Total | 6.796 | 6.296.140 | 5.199.264 | 780.578 | 316.298 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras esperadas ($Lift > 1$, $Support > 0,05$)

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 2.033 eventos de exceção não atingiram linhas de ônibus no raio de 100 m.

^g 440 eventos de exceção não atingiram linhas de ônibus no raio de 1.000 m.

9 Conclusão

Neste capítulo, são apresentadas as contribuições e resultados esperados com o projeto de pesquisa, as limitações a ameaças à validade do estudo.

9.1 Contribuições

A principal contribuição deste projeto é o estudo realizado para caracterização de eventos de exceção e de seus respectivos impactos no sistema de transporte público por ônibus da cidade de São Paulo, por meio de tweets, dados históricos dos módulos AVL do SIM e da GTFS. Também, validamos uma metodologia para extração e geolocalização dos endereços contidos nas publicações dos órgãos responsáveis por reportar eventos de exceção da cidade de São Paulo. Além disso, propomos uma arquitetura distribuída para exploração e visualização de dados AVL.

9.2 Trabalhos publicados

DIAS, F. C. A.; Daniel Cordeiro. *Visualizing large datasets: A case study with data of the buses of São Paulo city*. In: *1st Workshop on the Distributed Smart City (WDSC'2018)*, 2018, Salvador, BA. *Proceedings of the 37th IEEE International Symposium on Reliable Distributed Systems*, 2018. p. 10-13.

9.3 Trabalhos submetidos

DIAS, F.C.A; Daniel Cordeiro. *Characterization of exception events and their respective impacts on the public transport system by bus of São Paulo*. Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC), 2019.

9.4 Trabalhos futuros

Como trabalho futuro, pretendemos implementar o fluxo de processamento de dados em *streaming* mencionado na Figura 6, em um cenário de exploração

e visualização de dados quase em tempo real. Além disso, há a necessidade de estabelecermos uma cooperação entre a Acadêmia e a SPTrans para aplicação cotidiana dos experimentos realizados por esse trabalho e outros relacionados a análise de grandes volumes de dados de transportes públicos. Outra possibilidade futura é a de aplicar os experimentos realizados por este trabalho a publicações de usuários que representam a sociedade civil.

Referências

- ABBASI, A. et al. Utilising Location Based Social Media in Travel Survey Methods: bringing Twitter data into the play. *Proc. 8th ACM SIGSPATIAL Int. Work. Locat. Soc. Networks - LBSN'15*, p. 1–9, 2015. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2830657.2830660>>. Citado 5 vezes nas páginas 62, 63, 69, 70 e 72.
- AHVENNIEMI, H. et al. What are the differences between sustainable and smart cities? *Cities*, Elsevier B.V., v. 60, p. 234–245, 2017. ISSN 02642751. Disponível em: <<http://dx.doi.org/10.1016/j.cities.2016.09.009>>. Citado 3 vezes nas páginas 33, 34 e 35.
- ALBINO, V.; BERARDI, U.; DANGELICO, R. M. Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, Taylor & Francis, v. 22, n. 1, p. 3–21, 2015. Citado na página 35.
- ANANTHARAM, P. et al. Extracting City Traffic Events from Social Streams. *ACM Trans. Intell. Syst. Technol.*, v. 6, n. 4, p. 1–27, 2015. ISSN 21576904. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2801030.2717317>>. Citado 7 vezes nas páginas 62, 64, 67, 68, 69, 70 e 72.
- Andreas Mueller. 2018. <<https://pypi.python.org/pypi/wordcloud>>. Acesso em Fevereiro, 13 de 2018. Citado na página 60.
- ANDRIENKO, G. et al. Visual analytics of mobility and transportation: State of the art and further research directions. *IEEE Transactions on Intelligent Transportation Systems*, IEEE, v. 18, n. 8, p. 2232–2249, 2017. Citado na página 84.
- ANG, L.-M. et al. Big Sensor Data Systems for Smart Cities. *IEEE Internet Things J.*, v. 4, n. 5, p. 1–1, 2017. ISSN 2327-4662. Disponível em: <<http://ieeexplore.ieee.org/document/7903653/>>. Citado 2 vezes nas páginas 34 e 35.
- ANTTIROIKO, A. V. U-cities reshaping our future: Reflections on ubiquitous infrastructure as an enabler of smart urban development. *AI Soc.*, v. 28, n. 4, p. 491–507, 2013. ISSN 09515666. Citado na página 28.
- ATEFEH, F.; KHREICH, W. A survey of techniques for event detection in twitter. *Computational Intelligence*, Wiley Online Library, v. 31, n. 1, p. 132–164, 2015. Citado na página 81.
- BARTH, J. et al. Informational urbanism . A conceptual framework of smart cities. *Proc. 50th Hawaii Int. Conf. Syst. Sci.*, p. 2814–2823, 2017. Citado 2 vezes nas páginas 34 e 35.
- BENDLER, J. et al. Taming Uncertainty in Big Data. *Bus. Inf. Syst. Eng.*, v. 6, n. 5, p. 279–288, 2014. ISSN 1867-0202. Disponível em: <<http://link.springer.com/10.1007/s12599-014-0342-4>>. Citado 6 vezes nas páginas 62, 64, 67, 69, 71 e 72.
- BIOLOCHINI, J. et al. Techincal report rt-es 679/05: Systematic review in software engineering. *COPPE/UFRJ, 2005*Rio de Janeiro, 2005. Citado 2 vezes nas páginas 54 e 55.

- CHANIOTAKIS, E.; ANTONIOU, C. Use of Geotagged Social Media in Urban Settings: Empirical Evidence on Its Potential from Twitter. *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, v. 2015-Octob, n. 1, p. 214–219, 2015. Citado 2 vezes nas páginas 62 e 69.
- CHANIOTAKIS, E.; ANTONIOU, C.; PEREIRA, F. Mapping Social media for transportation studies. *IEEE Intell. Syst.*, v. 31, n. 6, p. 64–70, 2016. ISSN 15411672. Citado na página 55.
- CHEN, L. et al. Dynamic Cluster-Based Over-Demand Prediction in Bike Sharing Systems. *UBICOMP*, p. 841–852, 2016. Citado 13 vezes nas páginas 31, 32, 62, 64, 65, 66, 67, 68, 69, 70, 71, 95 e 96.
- CHEN, W.; GUO, F.; WANG, F.-Y. A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems*, IEEE, v. 16, n. 6, p. 2970–2984, 2015. Citado na página 84.
- CHUA, A. et al. Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tour. Manag.*, Elsevier Ltd, v. 57, p. 295–310, 2016. ISSN 02615177. Disponível em: <<http://dx.doi.org/10.1016/j.tourman.2016.06.013>>. Citado 2 vezes nas páginas 62 e 63.
- COLLOBERT, R. et al. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, v. 12, n. Aug, p. 2493–2537, 2011. Citado na página 41.
- CONSULO, M. et al. An evaluation of the proposed ITS system for the city of São Paulo based on the 2015 tender. In: EDP SCIENCES. *MATEC Web of Conferences*. [S.l.], 2016. v. 76, p. 03004. Citado 2 vezes nas páginas 28 e 29.
- CRUZ, A. et al. Detecção de anomalias frequentes no transporte rodoviário urbano. In: SBC. *SBBD: Brazilian Symposium on Databases*. [S.l.], 2018. p. 271–276. Citado na página 111.
- DI LORENZO, G. et al. EXSED: An intelligent tool for exploration of social events dynamics from augmented trajectories. *Proc. - IEEE Int. Conf. Mob. Data Manag.*, v. 1, p. 323–330, 2013. ISSN 15516245. Citado 5 vezes nas páginas 62, 64, 69, 70 e 72.
- DIAS, F. *Repositório contendo os artefatos da Revisão Sistemática*. 2017. Disponível em: <<https://github.com/fcas/dissertacao>>. Citado 2 vezes nas páginas 60 e 110.
- DOGRU, N.; SUBASI, A. Traffic accident detection using random forest classifier. In: IEEE. *Learning and Technology Conference (L&T), 2018 15th*. [S.l.], 2018. p. 40–45. Citado 2 vezes nas páginas 46 e 47.
- DWIVEDI, S. K.; ARYA, C. Automatic text classification in information retrieval: A survey. In: ACM. *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*. [S.l.], 2016. p. 131. Citado 2 vezes nas páginas 44 e 46.

FARSEEV, A. et al. Harvesting Multiple Sources for User Profile Learning. *Proc. 5th ACM Int. Conf. Multimed. Retr. - ICMR '15*, p. 235–242, 2015. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2671188.2749381>>. Citado 6 vezes nas páginas 62, 63, 69, 70, 71 e 72.

FIGUEIREDO, L. et al. Towards the development of intelligent transportation systems. In: IEEE. *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*. [S.l.], 2001. p. 1206–1211. Citado 3 vezes nas páginas 35, 36 e 37.

FINGER, M.; RAZAGHI, M. Conceptualizing “Smart Cities”. *Informatik-Spektrum*, v. 40, n. 1, p. 6–13, 2017. ISSN 1432122X. Citado 3 vezes nas páginas 33, 34 e 35.

FRIAS-MARTINEZ, V.; FRIAS-MARTINEZ, E. Spectral clustering for sensing urban land use using Twitter activity. *Eng. Appl. Artif. Intell.*, Elsevier, v. 35, p. 237–245, 2014. ISSN 09521976. Disponível em: <<http://dx.doi.org/10.1016/j.engappai.2014.06.019>>. Citado 7 vezes nas páginas 62, 65, 67, 68, 69, 71 e 72.

GAL-TZUR, A. et al. The potential of social media in delivering transport policy goals. *Transp. Policy*, v. 32, p. 115–123, 2014. ISSN 0967070X. Citado 8 vezes nas páginas 31, 65, 67, 68, 69, 70, 71 e 96.

GKIOTSALITIS, K.; STATHOPOULOS, A. A utility-maximization model for retrieving users’ willingness to travel for participating in activities from big-data. *Transp. Res. Part C Emerg. Technol.*, Elsevier Ltd, v. 58, p. 265–277, 2015. ISSN 0968090X. Disponível em: <<http://dx.doi.org/10.1016/j.trc.2014.12.006>>. Citado 3 vezes nas páginas 62, 63 e 67.

GKIOTSALITIS, K.; STATHOPOULOS, A. Joint leisure travel optimization with user-generated data via perceived utility maximization. *Transp. Res. Part C Emerg. Technol.*, Elsevier Ltd, v. 68, p. 532–548, 2016. ISSN 0968090X. Disponível em: <<http://dx.doi.org/10.1016/j.trc.2016.05.009>>. Citado 5 vezes nas páginas 62, 63, 66, 67 e 69.

GUO, W. et al. Understanding happiness in cities using twitter: Jobs, children, and transport. *IEEE 2nd Int. Smart Cities Conf. Improv. Citizens Qual. Life, ISC2 2016 - Proc.*, 2016. Citado 6 vezes nas páginas 62, 63, 65, 67, 71 e 72.

GUTEV, A.; NENKO, A. Better Cycling - Better Life: Social Media Based Parametric Modeling Advancing Governance of Public Transportation System in St. Petersburg. *Proc. Int. Conf. Electron. Gov. Open Soc. Challenges Eurasia*, p. 242–247, 2016. Disponível em: <<http://doi.acm.org/10.1145/3014087.3014123>>. Citado 6 vezes nas páginas 62, 63, 66, 69, 71 e 72.

GUYON, I.; ELISSEEFF A. An introduction to feature extraction. *Feature extraction*, Springer, p. 1–25, 2006. Citado na página 42.

HASAN, S.; UKKUSURI, S. V. Urban activity pattern classification using topic models from online geo-location data. *Transp. Res. Part C Emerg. Technol.*, Elsevier Ltd, v. 44, p. 363–381, 2014. ISSN 0968090X. Disponível em: <<http://dx.doi.org/10.1016/j.trc.2014.04.003>>. Citado 4 vezes nas páginas 62, 63, 70 e 72.

- ITOH, M. et al. Visual Exploration of Changes in Passenger Flows and Tweets on Mega-City Metro Network. *IEEE Trans. Big Data*, v. 2, n. 1, p. 85–99, 2016. ISSN 2332-7790. Disponível em: <<http://ieeexplore.ieee.org/document/7445832/>>. Citado 7 vezes nas páginas 31, 65, 66, 69, 70, 95 e 96.
- JUNGHERR, A. Twitter use in election campaigns: A systematic literature review. *Journal of information technology & politics*, Taylor & Francis, v. 13, n. 1, p. 72–91, 2016. Citado na página 55.
- KHEMPHILA, A.; BOONJING, V. Comparing performances of logistic regression, decision trees, and neural networks for classifying heart disease patients. In: IEEE. *Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on*. [S.l.], 2010. p. 193–198. Citado na página 51.
- KIBANOV, M. et al. Adaptive knn using expected accuracy for classification of geo-spatial data. In: ACM. *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. [S.l.], 2018. p. 857–865. Citado na página 47.
- KOBDAKI, H.; SCHÜTZE, H.; BURKOVSKI, A. Relational feature engineering of natural language processing. *Proc. 19th . . . , n. ii*, p. 1705–1708, 2010. Disponível em: <<http://dl.acm.org/citation.cfm?id=1871709>>. Citado na página 93.
- KORENIUS, T. et al. Stemming and lemmatization in the clustering of finnish text documents. In: *Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2004. (CIKM '04), p. 625–633. ISBN 1-58113-874-1. Disponível em: <<http://doi.acm.org/10.1145/1031171.1031285>>. Citado na página 41.
- KOTSIANTIS, S. B.; ZAHARAKIS, I.; PINTELAS, P. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, v. 160, p. 3–24, 2007. Citado 3 vezes nas páginas 44, 45 e 46.
- KOTSIANTIS, S. B.; ZAHARAKIS, I. D.; PINTELAS, P. E. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, Springer, v. 26, n. 3, p. 159–190, 2006. Citado 4 vezes nas páginas 47, 48, 49 e 50.
- KUFLIK, T. et al. Automating a framework to extract and analyse transport related social media content: The potential and the challenges. *Transportation Research Part C: Emerging Technologies*, Elsevier, v. 77, p. 275–291, 2017. Citado na página 29.
- KUMMITHA, R. K. R.; CRUTZEN, N. How do we understand smart cities? An evolutionary perspective. *Cities*, Elsevier, v. 67, n. July 2016, p. 43–52, 2017. ISSN 02642751. Disponível em: <<http://dx.doi.org/10.1016/j.cities.2017.04.010>>. Citado 3 vezes nas páginas 33, 34 e 35.
- KURT, I.; TURE, M.; KURUM, A. T. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert systems with applications*, Elsevier, v. 34, n. 1, p. 366–374, 2008. Citado na página 50.

- LECUE, F. et al. Smart traffic analytics in the semantic web with STAR-CITY: Scenarios, system and lessons learned in Dublin City. *J. Web Semant.*, Elsevier B.V., v. 27, p. 26–33, 2014. ISSN 15708268. Disponível em: <<http://dx.doi.org/10.1016/j.websem.2014.07.002>>. Citado 6 vezes nas páginas 31, 62, 64, 69, 70 e 96.
- LIU, D.; LI, Y.; THOMAS, M. A. A roadmap for natural language processing research in information systems. In: *Proceedings of the 50th Hawaii International Conference on System Sciences*. [S.l.: s.n.], 2017. Citado na página 40.
- MAGHREBI, M. et al. Complementing Travel Diary Surveys with Twitter Data: Application of Text Mining Techniques on Activity Location, Type and Time. *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, v. 2015-Octob, p. 208–213, 2015. Citado 4 vezes nas páginas 62, 63, 65 e 66.
- MATA, F.; CLARAMUNT, C. A Mobile Trusted Path System Based on Social Network Data. *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst.*, p. 101:1—101:4, 2015. Disponível em: <<http://doi.acm.org/10.1145/2820783.2820799>>. Citado 5 vezes nas páginas 62, 63, 69, 70 e 72.
- MCDONALD, A. D. et al. Steering in a random forest: Ensemble learning for detecting drowsiness-related lane departures. *Human factors*, Sage Publications Sage CA: Los Angeles, CA, v. 56, n. 5, p. 986–998, 2014. Citado na página 46.
- MENUAR, H. et al. Uav-enabled intelligent transportation systems for the smart city: Applications and challenges. *IEEE Communications Magazine*, IEEE, v. 55, n. 3, p. 22–28, 2017. Citado 2 vezes nas páginas 35 e 37.
- MIDDLETON, S. E.; MIDDLETON, L.; MODAFFERI, S. Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*, v. 29, n. 2, p. 9–17, 2014. ISSN 15411672. Citado na página 93.
- MORENO, M. V. et al. Applicability of Big Data Techniques to Smart Cities Deployments. *IEEE Trans. Ind. Informatics*, v. 13, n. 2, p. 800–809, 2017. ISSN 15513203. Citado 2 vezes nas páginas 34 e 35.
- MOTODA, H.; LIU, H. Feature selection, extraction and construction. *Communication of IICM (Institute of Information and Computing Machinery, Taiwan) Vol*, v. 5, p. 67–72, 2002. Citado na página 43.
- MUKHERJEE, T. et al. Janayuja: A People-centric Platform to Generate Reliable and Actionable Insights for Civic Agencies. *Acm Dev 2015*, p. 137–145, 2015. Citado 7 vezes nas páginas 62, 64, 66, 67, 68, 69 e 71.
- MYERS, S. A. et al. Information network or social network?: the structure of the twitter follow graph. In: ACM. *Proceedings of the 23rd International Conference on World Wide Web*. [S.l.], 2014. p. 493–498. Nenhuma citação no texto.
- NADKARNI, P. M.; OHNO-MACHADO, L.; CHAPMAN, W. W. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, BMJ Group BMA House, Tavistock Square, London, WC1H 9JR, v. 18, n. 5, p. 544–551, 2011. Citado 3 vezes nas páginas 40, 41 e 42.

NARAYANAN, U. et al. A survey on various supervised classification algorithms. In: IEEE. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*. [S.l.], 2017. p. 2118–2124. Citado na página 44.

NELSON, J. D.; MULLEY, C. The impact of the application of new technology on public transport service provision and the passenger experience: A focus on implementation in Australia. *Res. Transp. Econ.*, Elsevier Ltd, v. 39, n. 1, p. 300–308, 2013. ISSN 07398859. Disponível em: <<http://dx.doi.org/10.1016/j.retrec.2012.06.028>>. Citado na página 29.

NI, M.; HE, Q.; GAO, J. Forecasting the Subway Passenger Flow Under Event Occurrences With Social Media. *IEEE Trans. Intell. Transp. Syst.*, v. 18, n. 6, p. 1623–1632, 2016. ISSN 15249050. Citado 6 vezes nas páginas 65, 66, 68, 70, 71 e 72.

NIU, W. et al. Community-based geospatial tag estimation. In: IEEE. *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*. [S.l.], 2016. p. 279–286. Citado na página 94.

NOI, P. T.; KAPPAS, M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 1, p. 18, 2018. Citado na página 47.

PARK, S. H. et al. Apriori-based text mining method for the advancement of the transportation management plan in expressway work zones. *The Journal of Supercomputing*, Springer, v. 74, n. 3, p. 1283–1298, 2018. Citado na página 53.

ROY, A.; MAJUMDER, A. G.; NATH, A. Understanding natural language processing and its primary aspects. *International Journal*, v. 5, n. 8, 2017. Citado na página 40.

SANTOS, H. et al. Contextual data collection for smart cities. *CoRR*, abs/1704.01802, 2017. Disponível em: <<http://arxiv.org/abs/1704.01802>>. Citado 2 vezes nas páginas 34 e 35.

SCHEIN, A. I.; UNGAR, L. H. Active learning for logistic regression: an evaluation. *Machine Learning*, Springer, v. 68, n. 3, p. 235–265, 2007. Citado na página 50.

SERAJ, F.; MERATNIA, N.; HAVINGA, P. J. An aggregation and visualization technique for crowd-sourced continuous monitoring of transport infrastructures. In: IEEE. *Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on*. [S.l.], 2017. p. 219–224. Citado na página 84.

SETIAWAN, E. B.; WIDYANTORO, D. H.; SURENDRO, K. Feature expansion using word embedding for tweet topic classification. *Proceeding 2016 10th Int. Conf. Telecommun. Syst. Serv. Appl. TSSA 2016 Spec. Issue Radar Technol.*, n. 2011, 2017. Citado 3 vezes nas páginas 40, 41 e 93.

SINGH, A.; THAKUR, N.; SHARMA, A. A review of supervised machine learning algorithms. In: IEEE. *Computing for Sustainable Global Development (INDIACoM), 2016 3rd International Conference on*. [S.l.], 2016. p. 1310–1315. Citado 5 vezes nas páginas 47, 48, 49, 50 e 51.

SOBOLEVSKY, S. et al. Scaling of City Attractiveness for Foreign Visitors through Big Data of Human Economical and Social Media Activity. *Proc. - 2015 IEEE Int. Congr. Big Data, BigData Congr. 2015*, p. 600–607, 2015. ISSN 2379-7703. Citado 2 vezes nas páginas 62 e 63.

SOOMRO, K.; KHAN, Z.; HASHAM, K. Towards Provisioning of Real-time Smart City Services Using Clouds. *ACM 9th Int. Conf. Util. Cloud Comput. Towar.*, v. 1691, p. 50–59, 2016. ISSN 16130073. Citado 3 vezes nas páginas 62, 64 e 69.

STEIGER, E.; ALBUQUERQUE, J. P.; ZIPF, A. An advanced systematic literature review on spatiotemporal analyses of twitter data. *Transactions in GIS*, Wiley Online Library, v. 19, n. 6, p. 809–834, 2015. Citado na página 55.

STEIGER, E. et al. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Comput. Environ. Urban Syst.*, Elsevier Ltd, v. 54, p. 255–265, 2015. ISSN 01989715. Disponível em: <<http://dx.doi.org/10.1016/j.compenvurbssys.2015.09.007>>. Citado 6 vezes nas páginas 62, 63, 68, 69, 72 e 93.

SÁ, T. H. et al. Health impact modelling of different travel patterns on physical activity, air pollution and road injuries for são paulo, brazil. *Environment International*, v. 108, n. Supplement C, p. 22 – 31, 2017. ISSN 0160-4120. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0160412017305974>>. Citado na página 27.

TALARI, S. et al. A Review of Smart Cities Based on the Internet of Things Concept. *Energies*, v. 10, n. 4, p. 421, 2017. ISSN 1996-1073. Disponível em: <<http://www.mdpi.com/1996-1073/10/4/421>>. Citado 2 vezes nas páginas 34 e 35.

THOMAZ, G. M. et al. Content mining framework in social media: A FIFA world cup 2014 case analysis. *Inf. Manag.*, Elsevier B.V., 2016. ISSN 03787206. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0378720616303354>>. Citado 2 vezes nas páginas 62 e 63.

United States Department of Transportation. *ITS Strategic Plan 2015-2019*. 2017. <<https://www.its.dot.gov/strategicplan.pdf>>. Acesso em Setembro, 17 de 2017. Citado na página 29.

WANG, S.; SINNOTT, R.; NEPAL, S. Privacy-protected social media user trajectories calibration. *Proc. 2016 IEEE 12th Int. Conf. e-Science, e-Science 2016*, p. 293–302, 2016. Citado 2 vezes nas páginas 33 e 64.

WEN, X.; LIN, Y.-R.; PELECHRINIS, K. Pairfac: Event analytics through discriminant tensor factorization. In: ACM. *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*. [S.l.], 2016. p. 519–528. Citado 7 vezes nas páginas 62, 63, 65, 66, 67, 69 e 71.

WU, H.; YUAN, N. An improved tf-idf algorithm based on word frequency distribution information and category distribution information. In: ACM. *Proceedings of the 3rd International Conference on Intelligent Information Processing*. [S.l.], 2018. p. 211–215. Citado na página 52.

- XIAO, Z.; LIM, H. B.; PONNAMBALAM, L. Participatory Sensing for Smart Cities: A Case Study on Transport Trip Quality Measurement. *IEEE Trans. Ind. Informatics*, v. 13, n. 2, p. 759–770, 2017. ISSN 1551-3203. Citado 2 vezes nas páginas 34 e 35.
- XIE, Y. et al. The optimization and improvement of the apriori algorithm. In: IEEE. *Education Technology and Training, 2008. and 2008 International Workshop on Geoscience and Remote Sensing. ETT and GRS 2008. International Workshop on*. [S.l.], 2008. v. 2, p. 663–665. Citado na página 110.
- YAHAV, I.; SHEHORY, O.; SCHWARTZ, D. Comments mining with tf-idf: The inherent bias and its removal. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, 2018. Citado na página 52.
- YANG, F. et al. Druid: A real-time analytical data store. In: ACM. *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. [S.l.], 2014. p. 157–168. Citado 2 vezes nas páginas 85 e 86.
- YOUAF, J. et al. Generalized multipath planning model for ride-sharing systems. *Front. Comput. Sci.*, v. 8, n. 1, p. 100–118, 2014. ISSN 20952228. Citado 5 vezes nas páginas 62, 65, 67, 68 e 69.
- YU, W. Discovering frequent movement paths from taxi trajectory data using spatially embedded networks and association rules. *IEEE Transactions on Intelligent Transportation Systems*, IEEE, 2018. Citado na página 111.
- ZAGAL, R.; MATA, F.; CLARAMUNT, C. Geographical Knowledge Discovery applied to the Social Perception of Pollution in the City of Mexico. *LBSN*, 2016. Citado 4 vezes nas páginas 62, 70, 72 e 93.
- ZENG, W. et al. A visual analytics design for studying rhythm patterns from human daily movement data. *Visual Informatics*, Elsevier, v. 1, n. 2, p. 81–91, 2017. Citado na página 111.
- ZHANG, K. et al. A method to optimize apriori algorithm for frequent items mining. In: IEEE. *Computational Intelligence and Design (ISCID), 2014 Seventh International Symposium on*. [S.l.], 2014. v. 1, p. 71–75. Citado na página 110.
- ZHAO, D. et al. Recognizing metro-bus transfers from smart card data. *Transportation Planning and Technology*, Taylor & Francis, v. 42, n. 1, p. 70–83, 2019. Citado na página 110.
- ZHOU, X.; CHEN, L. Event detection over twitter social media streams. *The VLDB journal*, Springer, v. 23, n. 3, p. 381–400, 2014. Citado na página 81.

Apêndices

Apêndice A – Exemplos de tweets

Exemplos de tweets dos perfis selecionados citados na Tabela 1

```

1  {
2      "tweet_id" : 895060642952077314,
3      "tweet_account": "BombeirosPMESP",
4      "text" : "19h58 Colisão de Carro x Caminhão, Estrada Sta Isabel,
5          5950 Itaquaquecetuba. 2 Vítimas, 1 Vtr. Aguardando maiores
6          informes"
7
8  }
9  {
10     "tweet_id" : 894707930217447427,
11     "tweet_account": "CETSP_",
12     "text" : "Referente manifestação Rua Augusta, pista liberada.#ZC"
13
14  }
15  {
16     "tweet_id" : 894147793060716544,
17     "tweet_account": "CPTM_oficial",
18     "text" : "#L11 Hoje, das 8h à meia-noite, circulação interrompida
19         entre Luz e Brás. P/ seguir viagem, use a L7-Rubi q prestará
20         serviço até a Est. Brás"
21
22     "tweet_id" : 895054721026838530,
23     "tweet_account": "governosp",
24     "text" : "@SANROGE Lamentamos o ocorrido, Rogerio. Estamos
25         trabalhando continuamente para melhorar a segurança na região.
26         Entre maio e junho, [+][1]"
27
28  }
29  {
30     "tweet_id" : 895000711284621312,

```

```
23     "tweet_account": "metrosp_oficial",
24     "text" : "08/08/2017 16:16: #metrosp : Linha 5-Lilás: Velocidade
25       Reduzida. Mais informações em https://t.co/CaeqD26iJR"
26   }
27   {
28     "tweet_id" : 884039273493803008,
29     "tweet_account": "PMESP",
30     "text" : "AGORA: Desfile Cívico-Militar de 9 de Julho no Obelisco
31       - Ibirapuera SP, transmissão ao vivo na página oficial Facebook
32       da Polícia Militar.",
33     "dateTime" : "2017-07-09 10:19:22"
34   }
35   {
36     "tweet_id" : 887315002117500932,
37     "tweet_account": "Policia_Civil",
38     "text" : "Policia Civil realiza operação para combater a prática
39       do Jogo conhecido como "Baleia Azul"... https://t.co/kh2HW6UZvT
40       ",
41   }
42   {
43     "tweet_id" : 895004079910518788,
44     "tweet_account": "saopaulo_agora",
45     "text" : "#ItaimPaulista Incêndio na Rua Mateus Barbosa de Resende
46       nº 235. Defesa Civil Regional acionada para o local. (CCOI) #
       spagora"
47   }
48   {
49     "tweet_id" : 894694704989732864,
50     "tweet_account": "smtpsp_",
51     "text" : "A @sptrans_ irá modificar 14 linhas na Zona Leste para
52       obras no Monotrilho Saiba mais: https://t.co/fCA0T7WCSY"
53   }
```

```
47 {  
48     "tweet_id" : 902953598857949184,  
49     "tweet_account": "SPCEDEC",  
50     "text" : "30-08-2017 - Acidente com produto perigoso em com 36 ,  
51         deixa 21 vítimas feridas e 02 ."  
52 {  
53     "tweet_id" : 895065137484320769,  
54     "tweet_account": "sptrans_ ",  
55     "text" : "Obras do Monotrilho desviam itinerários de 14 linhas que  
56         atendem a Av. Sapopemba entre 5 e 11/08, das 23h às 5h: https:  
57             //t.co/jH4LFgrSKZ"  
58 }  
59 {  
60     "tweet_id" : 895042604068458497,  
61     "tweet_account": "TurismoSaoPaulo",  
62     "text" : "Veganos, vegetarianos e simpatizantes: vem aí o Vegan  
63         Club, em 12/08, no Centro de SP! #crueltyfree #veganfood...  
64             https://t.co/7f7ggr4vn4"  
65 }
```

Apêndice B – Logradouros utilizados

Tabela 16 – Tabela de logradouros com abreviaturas

| Abreviatura | Logradouro |
|--------------------|-------------------|
| ACAMP | Acampamento |
| AC | Acesso |
| AD | Adro |
| ERA | Aeroporto |
| AL | Alameda |
| AT | Alto |
| A | Area |
| AE | Area especial |
| ART | Arteria |
| ATL | Atalho |
| AV | Avenida |
| AV-CONT | Avenida contorno |
| BX | Baixa |
| BLO | Balao |
| BAL | Balneario |
| BC | Beco |
| BELV | Belvedere |
| BL | Bloco |
| BSQ | Bosque |
| BVD | Boulevard |
| BCO | Buraco |
| C | Cais |
| CALC | Calcada |
| CAM | Caminho |
| CPO | Campo |
| CAN | Canal |
| CHAP | Chacara |

Continua na próxima página

Tabela 16 – continuação da página anterior

| Abreviatura | Logradouro |
|--------------------|--------------------|
| CHAP | Chapadao |
| CIRC | Circular |
| COL | Colonia |
| CMP-VR | Complexo viario |
| COND | Condominio |
| CJ | Conjunto |
| COR | Corredor |
| CRG | Corrego |
| DSC | Descida |
| DSV | Desvio |
| DT | Distrito |
| EVD | Elevada |
| ENT-PART | Entrada particular |
| EQ | Entre quadra |
| ESC | Escada |
| ESP | Esplanada |
| ETC | Estacao |
| ESTC | Estacionamento |
| ETD | Estadio |
| ETN | Estancia |
| EST | Estrada |
| EST-MUN | Estrada municipal |
| FAV | Favela |
| FAZ | Fazenda |
| FRA | Feira |
| FER | Ferrovia |
| FNT | Fonte |
| FTE | Forte |
| GAL | Galeria |

Continua na próxima página

Tabela 16 – continuação da página anterior

| Abreviatura | Logradouro |
|--------------------|----------------------|
| GJA | Granja |
| HAB | Habitacional |
| IA | Ilha |
| JD | Jardim |
| JDE | Jardinete |
| LD | Ladeira |
| LG | Lago |
| LGA | Lagoa |
| LRG | Largo |
| LOT | Loteamento |
| MNA | Marina |
| MOD | Modulo |
| TEM | Monte |
| MRO | Morro |
| NUC | Nucleo |
| PDA | Parada |
| PDO | Paradouro |
| PAR | Paralela |
| PRQ | Parque |
| PSG | Passagem |
| PSC-SUB | Passagem subterranea |
| PSA | Passarela |
| PAS | Passeio |
| PAT | Patio |
| PNT | Ponta |
| PTE | Ponte |
| PTO | Porto |
| PC | Praca |
| PC-ESP | Praça de esportes |

Continua na próxima página

Tabela 16 – continuação da página anterior

| Abreviatura | Logradouro |
|--------------------|---------------------|
| PR | Praia |
| PRL | Prolongamento |
| Q | Quadra |
| QTA | Quinta |
| QTAS | Quinta |
| RAM | Rama |
| RMP | Rampa |
| REC | Recanto |
| RES | Residencial |
| RET | Reta |
| RER | Retiro |
| RTN | Retorno |
| ROD-AN | RodoAnel |
| ROD | Rodovia |
| RTT | Rotatoria |
| ROT | Rotula |
| R | Rua |
| R-LIG | Rua de ligação |
| R-PED | Rua de pedestre |
| SRV | Servidao |
| ST | Setor |
| SIT | Sitio |
| SUB | Subida |
| TER | Terminal |
| TV | Travessa |
| TV-PART | Travessa particular |
| TRV | Trecho |
| TRV | Trevo |
| TCH | Trincheira |

Continua na próxima página

Tabela 16 – continuação da página anterior

| Abreviatura | Logradouro |
|--------------------|-------------------|
| TUN | Tunel |
| UNID | Unidade |
| VAL | Vala |
| VLE | Vale |
| VRTE | Variante |
| VER | Vereda |
| V | Via |
| V-AC | Via de acesso |
| V-PED | Via de pedestre |
| V-EVD | Via elevado |
| V-EXP | Via expressa |
| VD | Viaduto |
| VLA | Viela |
| VL | Vila |
| ZIG-ZAG | Zigue-zague |

Fonte: MS/SAS/DRAC/CGSI - Coordenação Geral dos Sistemas de Informação
 (adaptada)¹

¹ <http://www.pmf.sc.gov.br/arquivos/arquivos/pdf/04_01_2010_10.27.25.2b615e6755138defe1bdb00f1c86031f.PDF>. Acessado em 29 de outubro de 2017.

Apêndice C – Detalhamento dos campos da GTFS

Tabela 17 – Detalhamento dos campos do arquivo *agency.txt* da GTFS

| Nome do campo | Condisional | Descrição |
|------------------------|-------------|---|
| <i>agency_id</i> | Opcional | Identifica uma agência de transporte público. Um <i>feed</i> de transporte público pode representar dados de mais de uma agência. Este campo é opcional para <i>feeds</i> de transporte público que contenham somente dados de uma única agência. |
| <i>agency_name</i> | Obrigatório | Contém o nome completo da agência de transporte público. |
| <i>agency_url</i> | Obrigatório | Contém o <i>URL</i> da agência de transporte público. |
| <i>agency_timezone</i> | Obrigatório | Contém o fuso horário de onde a agência de transporte público está localizada. |
| <i>agency_lang</i> | Opcional | Contém um código <i>ISO 639-1</i> de duas letras para o idioma principal usado por essa agência de transporte público. |
| <i>agency_phone</i> | Opcional | Contém um único número de telefone da agência especificada. |
| <i>agency_fare_url</i> | Opcional | Especifica o <i>URL</i> de uma página da Web que permite que um passageiro compre passagens ou outros instrumentos de tarifas dessa agência <i>on-line</i> . |

Fonte: Google Transit (adaptada)¹

¹ <<https://developers.google.com/transit>>. Acessado em 29 de outubro de 2017.

Tabela 18 – Detalhamento dos campos do arquivo
stops.txt da GTFS

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|---|
| <i>stop_id</i> | Obrigatório | Contém um ID que identifica uma parada ou uma estação. Diversos trajetos podem usar a mesma parada. |
| <i>stop_code</i> | Opcional | Contém um pequeno texto ou um número que identifica a parada para os passageiros. Os códigos das paradas são usados muitas vezes em sistemas de informações sobre transporte público por telefone ou impressos em sinalizações nas paradas para que os passageiros possam obter informações sobre o horário das paradas com mais facilidade ou sobre chegadas de uma parada específica em tempo real. O campo <i>stop_code</i> só deve ser usado para códigos de parada exibidos aos passageiros. Para os códigos internos, use <i>stop_id</i> . Este campo deve ser deixado em branco para as paradas que não têm um código. |
| <i>stop_name</i> | Obrigatório | Contém o nome de uma parada ou estação. Use um nome compreensível para as pessoas locais e linguagem turística. |
| <i>stop_desc</i> | Opcional | Contém uma descrição de uma parada. Forneça informações úteis e de qualidade. Não basta repetir o nome da parada. |
| <i>stop_lat</i> | Obrigatório | Contém a latitude de uma parada ou estação. O valor do campo deve ser uma latitude WGS 84 válida. |

Continua na próxima página

Tabela 18 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|--|
| <i>stop_lon</i> | Obrigatório | Contém a longitude de uma parada ou estação. O valor do campo deve ser uma latitude WGS 84 válida entre -180 e 180. |
| <i>zone_id</i> | Opcional | Define a zona tarifária do ID de uma parada. Os IDs de zonas são obrigatórios para fornecer informações sobre tarifas usando <i>fare_rules.txt</i> . Se esse ID de parada representa uma estação, o ID de zona é ignorado. |
| <i>stop_url</i> | Opcional | Contém o URL de uma página da Web sobre uma parada específica. Ele deve ser diferente dos campos <i>agency_url</i> e <i>route_url</i> . |
| <i>location_type</i> | Opcional | Identifica se este ID de parada representa uma parada ou uma estação. Se nenhum tipo de local for especificado ou se o campo <i>location_type</i> estiver em branco, os IDs de parada serão tratados como paradas. As estações podem ter propriedades diferentes das paradas quando são representadas em um mapa ou usadas em planejamento de viagens. O campo de tipo de local pode ter os seguintes valores: 0 ou em branco (para parada) e 1 (estação). |

Continua na próxima página

Tabela 18 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|-----------------------|--------------------|---|
| <i>parent_station</i> | Opcional | Para paradas que estejam fisicamente localizadas dentro de estações, o campo <i>parent_station</i> identifica a estação associada à parada. Para usar este campo, o arquivo <i>stops.txt</i> também deve conter uma linha em que esse ID de parada tenha o tipo de localização=1. |

Continua na próxima página

Tabela 18 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|---|
| <i>stop_timezone</i> | Opcional | <p>Contém o fuso horário em que a parada ou estação está localizada. Se omitido, assume-se que a parada está localizada no fuso horário especificado por <i>agency_timezone</i> no arquivo <i>agency.txt</i>.</p> <p>Quando uma parada tem uma estação principal, considera-se que a parada esteja no fuso horário especificado pelo valor <i>stop_timezone</i> da estação principal. Se uma parada específica possui um valor <i>parent_station</i>, qualquer valor <i>stop_timezone</i> especificado para essa parada deve ser ignorado. Mesmo que os valores de <i>stop_timezone</i> sejam fornecidos no arquivo <i>stops.txt</i>, os horários em <i>stop_times.txt</i> devem continuar a ser especificados como horários desde a meia-noite no fuso horário especificado por <i>agency_timezone</i> em <i>agency.txt</i>. Isso garante que os valores de tempo em uma viagem sempre aumentam durante uma viagem, independentemente dos fusos horários pelos quais uma viagem passa.</p> |

Continua na próxima página

Tabela 18 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------------|--------------------|--|
| <i>wheelchair_boarding</i> | Opcional | <p>Identifica se é possível o embarque de passageiros em cadeira de rodas na parada ou estação especificada. O campo pode ter os seguintes valores: 0 (ou vazio) - indica que não há informações sobre acessibilidade para a parada; 1 - indica que, pelo menos, alguns veículos nesta parada possibilitam o embarque de passageiros em cadeira de rodas; 2 - o embarque de pessoas em cadeiras de roda não é possível nesta parada. Quando uma parada faz parte de um complexo de estações maiores, como indicado por uma para com um valor <i>parent_station</i>, o campo <i>wheelchair_boarding</i> da parada possui a seguinte semântica adicional: 0 (ou vazio) - a parada herdará o valor para <i>wheelchair_boarding</i> da estação principal, se especificado; 1 - existem vias de acesso na parte externa da estação para a parada/plataforma específica; 2 - não há vias de acesso na parte externa da estação para a parada/plataforma específica</p> |

Fonte: Google Transit (adaptada)¹

Tabela 19 – Detalhamento dos campos do arquivo *routes.txt* da GTFS

| Nome do campo | Condisional | Descrição |
|-------------------------|--------------------|---|
| <i>route_id</i> | Obrigatório | Contém um ID que identifica um trajeto. |
| <i>agency_id</i> | Opcional | Define uma agência para o trajeto especificado. Este valor é indicado no arquivo <i>agency.txt</i> . Campo destinado para quando for fornecido dados para trajetos de mais de uma agência. |
| <i>route_short_name</i> | Obrigatório | Contém o nome abreviado de um trajeto. Geralmente, será um identificador pequeno e abstrato, como, por exemplo "32", "100X" ou "Verde", que os passageiros usam para identificar um trajeto, mas que não fornece nenhuma identificação de quais lugares são atendidos pelo trajeto. Se o trajeto não tem um nome abreviado, especifique um <i>route_long_name</i> e use uma sequência vazia como o valor deste campo. |
| <i>route_long_name</i> | Obrigatório | Contém o nome completo de um trajeto. Em geral, esse nome é mais descritivo que <i>route_short_name</i> e incluirá o destino ou a parada do trajeto. Se o trajeto não tem um nome completo, especifique um <i>route_short_name</i> e use uma sequência vazia como o valor deste campo. |
| <i>route_desc</i> | Opcional | Contém uma descrição de um trajeto. Não basta repetir o nome do trajeto. |

Continua na próxima página

Tabela 19 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|-------------------------|--------------------|--|
| <i>route_type</i> | Obrigatório | Descreve o tipo de transporte usado em um trajeto. Os valores válidos deste campo são: 0 - Bonde, ônibus elétrico, veículo leve sobre trilhos; 1 - Metrô, trem subterrâneo; 2 - Via férrea; 3 - Ônibus; 4 - Balsa; 5 - Teleférico; 6 - Gôndola, teleférico suspenso; 7 - Funicular. |
| <i>route_url</i> | Opcional | Contém o URL de uma página da Web sobre esse trajeto específico. Ele deve ser diferente de <i>agency_url</i> . |
| <i>route_color</i> | Opcional | Define uma cor que corresponda ao trajeto. A cor deve ser informada como um número hexadecimal de seis caracteres. Se nenhuma cor é especificada, a cor padrão de trajetos é branca (FFFFFF). A diferença de cores entre <i>route_color</i> e <i>route_text_color</i> deve fornecer contraste suficiente quando visualizado em uma tela em preto e branco. |
| <i>route_text_color</i> | Opcional | Usado para especificar uma cor legível para usar em desenho de texto contra um plano de fundo de <i>route_color</i> . |

Fonte: Google Transit (adaptada)¹

Tabela 20 – Detalhamento dos campos do arquivo
trips.txt da GTFS

| Nome do campo | Condisional | Descrição |
|------------------------|--------------------|---|
| <i>route_id</i> | Obrigatório | Contém um ID que identifica um trajeto. Este valor é indicado no arquivo <i>agency.txt</i> . |
| <i>service_id</i> | Obrigatório | Contém um ID que identifica um conjunto de datas em que o serviço está disponível para um ou mais trajetos. Este valor é indicado no arquivo <i>calendar.txt</i> ou <i>calendar_dates.txt</i> . |
| <i>trip_id</i> | Obrigatório | Contém um ID que identifica uma viagem. |
| <i>trip_headsign</i> | Opcional | Contém o texto que aparece em uma sinalização que identifica o destino da viagem para os passageiros. Use este campo para distinguir diferentes padrões de serviço no mesmo trajeto. Se a placa muda durante uma viagem, você pode substituir o campo <i>trip_headsign</i> , especificando valores para o campo <i>stop_headsign</i> em <i>stop_times.txt</i> . |
| <i>trip_short_name</i> | Opcional | Contém o texto que aparece em programações e placas de sinalização para identificar a viagem para os passageiros, por exemplo, para identificar números de trens para viagens de trens suburbanos. Se os passageiros não recorrem normalmente aos nomes da viagem, deixe este campo em branco. Um valor de <i>trip_short_name</i> , se possível, deve identificar, com exclusividade, uma viagem em um dia de serviço; ele não deve ser usado para nomes de destino ou designações limitadas/expressas. |

Continua na próxima página

Tabela 20 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|--|
| <i>direction_id</i> | Opcional | Contém um valor binário que indica a direção de uma viagem. Use este campo para distinguir viagens bidirecionais com o mesmo <i>route_id</i> . Este campo não é usado na criação de trajetos; ele fornece uma maneira de separar viagens por direção durante a publicação de tabelas de horário. Você pode especificar nomes para cada direção com o campo <i>trip_headsign</i> . 0 - viagem em uma única direção (por exemplo, só ida); 1 - viagem na direção oposta (por exemplo, de volta), os campos <i>trip_headsign</i> e <i>direction_id</i> podem ser usados juntos para atribuir um nome a uma viagem em cada direção "1234". |
| <i>block_id</i> | Opcional | Identifica o quadro a que a viagem pertence. Um bloco consiste em duas ou mais viagens sequenciais feitas usando o mesmo veículo, em que um passageiro pode passar de uma viagem para a próxima permanecendo no veículo. O campo <i>block_id</i> deve ser indicado por duas ou mais viagens no arquivo <i>trips.txt</i> . |
| <i>shape_id</i> | Opcional | Contém um ID que define a forma da viagem. Este valor é indicado no arquivo <i>shapes.txt</i> . O arquivo <i>shapes.txt</i> permite definir como será traçada uma linha no mapa para representar uma viagem. |

Continua na próxima página

Tabela 20 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|-----------------------|--------------------|---|
| wheelchair_accessible | Opcional | 0 (ou vazio) - indica que não há informações sobre acessibilidade para a viagem; 1 - indica que o veículo que está sendo usado nesta viagem específica pode acomodar, pelo menos, um passageiro em cadeira de rodas; 2 - indica que não é possível acomodar passageiros em cadeiras de rodas nesta viagem |

Fonte: Google Transit (adaptada)¹

Tabela 21 – Detalhamento dos campos do arquivo
stop_times.txt da GTFS

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|--|
| <i>trip_id</i> | Obrigatório | Contém um ID que identifica uma viagem. Este valor é indicado no arquivo <i>trips.txt</i> . |

Continua na próxima página

Tabela 21 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|--|
| <i>arrival_time</i> | Obrigatório | <p>Especifica o horário de chegada em uma parada específica de uma viagem específica de um trajeto. No caso de horários que ocorram após a meia-noite na data do serviço, digite o horário como um valor maior que 24:00:00 em horário local HH:MM:SS para o dia em que começa a programação da viagem. Se não há horários separados para chegada e partida em uma parada, insira o mesmo valor para <i>arrival_time</i> e <i>departure_time</i>. É necessário especificar os horários de chegada para a primeira e a última paradas de uma viagem.</p> <p>Se essa parada não for programada, use uma sequência vazia para os campos <i>arrival_time</i> e <i>departure_time</i>. As paradas sem horário de chegada são programadas conforme a parada programada anterior mais próxima. Para garantir trajetos precisos, forneça horários de chegada e de partida para todas as paradas programadas.</p> <p>Não intercale as paradas, ou, preencha os horários com espaços. Observação: as viagens que abrangem várias datas terão horários de parada maiores que 24:00:00. Por exemplo, se uma viagem começa às 10:30:00 p.m e termina às 2:15:00 a.m. do dia seguinte, os horários de parada seriam 22:30:00 e 26:15:00. A inclusão desses horários de parada como 22:30:00 e 02:15:00 não produzem os resultados desejados.</p> |

Tabela 21 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|-----------------------|--------------------|---|
| <i>departure_time</i> | Obrigatório | <p>Especifica o horário de partida de uma parada específica para uma viagem específica de um trajeto. O horário é medido de "meio-dia menos 12h"(efetivamente meia-noite, exceto para dias do horário de verão), no início da data do serviço. No caso de horários que ocorram após a meia-noite na data do serviço, digite o horário como um valor maior que 24:00:00 em horário local HH:MM:SS para o dia em que começa a programação da viagem. Se não há horários diferentes para a chegada e a saída em uma parada, insira o mesmo valor para <i>arrival_time</i> e <i>departure_time</i>. É necessário especificar os horários de partida da primeira e da última paradas em uma viagem. Se essa parada não for programada, use uma sequência vazia para os campos <i>arrival_time</i> e <i>departure_time</i>. As paradas sem horário de chegada são programadas conforme a parada programada anterior mais próxima. Para garantir trajetos precisos, forneça horários de chegada e de partida para todas as paradas programadas. Não intercale as paradas. Os horários devem ter oito dígitos no formato HH:MM:SS (o formato H:MM:SS também é aceito, se a hora iniciar com 0). Não preencha os horários com espaços.</p> |

Continua na próxima página

Tabela 21 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|--|
| <i>stop_id</i> | Obrigatório | Contém um ID que identifica uma parada. Diversos trajetos podem usar a mesma parada. O campo <i>stop_id</i> é indicado no arquivo <i>stops.txt</i> . Se <i>location_type</i> é usado no arquivo <i>stops.txt</i> , todas as paradas indicadas em <i>stop_times.txt</i> deverão ter <i>location_type</i> igual a 0. Onde possível, os valores de <i>stop_id</i> devem permanecer consistentes entre as atualizações de feed. Se uma parada não está programada, digite valores em branco para <i>arrival_time</i> e <i>departure_time</i> . |
| <i>stop_sequence</i> | Obrigatório | Identifica a ordem das paradas de uma viagem específica. Os valores de <i>stop_sequence</i> devem ser números inteiros positivos e devem aumentar ao longo da viagem. |
| <i>stop_headsign</i> | Opcional | Contém o texto que aparece em uma sinalização que identifica o destino da viagem para os passageiros. Use este campo para substituir o <i>trip_headsign</i> padrão quando as placas mudarem durante as viagens. Se esta placa está associada a uma viagem inteira, use <i>trip_headsign</i> no lugar. |

Continua na próxima página

Tabela 21 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|--|
| <i>pickup_type</i> | Opcional | Indica se os passageiros são embarcados em uma parada como parte da programação normal ou se não há embarque disponível na parada. Este campo também permite que a agência de transporte público indique se os passageiros devem ligar para a agência ou notificar o motorista para agendar um embarque em uma parada específica. Os valores válidos deste campo são: 0 - Embarque no horário normal; 1 - Sem embarque disponível; 2 - Deve ligar para a agência a fim de agendar o embarque; 3 - Deve combinar com o motorista para agendar o embarque. O valor padrão deste campo é 0. |

Continua na próxima página

Tabela 21 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|---|
| <i>drop_off_type</i> | Opcional | Indica se há desembarque de passageiros em uma parada, como parte da programação normal ou se não há desembarques na parada. Este campo também permite que a agência de transporte público indique se os passageiros devem ligar para a agência ou notificar o motorista para agendar um desembarque em uma determinada parada. Os valores válidos deste campo são: 0 - Desembarque no horário normal; 1 - Desembarque não disponível; 2 - Deve telefonar para agendar o desembarque; 3 - Deve combinar com o motorista para agendar o desembarque. O valor padrão deste campo é 0. |

Continua na próxima página

Tabela 21 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------------|--------------------|--|
| <i>shape_dist_traveled</i> | Opcional | <p>Quando usado no arquivo <i>stop_times.txt</i>, o campo <i>shape_dist_traveled</i> posiciona uma parada como uma distância a partir do primeiro ponto de forma. O campo <i>shape_dist_traveled</i> representa uma distância real percorrida ao longo do trajeto em unidades como, por exemplo, pés ou quilômetros. Essas informações permitem que o planejador da viagem determine o quanto da forma deve ser desenhado ao exibir parte de uma viagem no mapa. Os valores usados para <i>shape_dist_traveled</i> devem aumentar juntamente com <i>stop_sequence</i>. As unidades usadas para <i>shape_dist_traveled</i> no arquivo <i>stop_times.txt</i> devem corresponder às unidades usadas para este campo no arquivo <i>shapes.txt</i>.</p> |

Fonte: Google Transit (adaptada)¹

Tabela 22 – Detalhamento dos campos do arquivo *calendar.txt* da GTFS

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|---|
| <i>service_id</i> | Obrigatório | Contém um ID que identifica um conjunto de datas em que o serviço está disponível para um ou mais trajetos. Cada valor de <i>service_id</i> pode aparecer, no máximo, uma vez em um arquivo <i>calendar.txt</i> . Este valor é um conjunto de dados exclusivo. Ele é indicado pelo arquivo <i>trips.txt</i> . |
| <i>monday</i> | Obrigatório | Contém um valor binário que indica se o serviço é válido para todas as segundas-feiras. O valor 1 indica que o serviço está disponível todas as segundas-feiras durante o período. O período é especificado utilizando-se os campos <i>start_date</i> e <i>end_date</i> . O valor 0 indica que o serviço não está disponível às segundas-feiras no período. Observação: você pode listar exceções para datas específicas, como, por exemplo, feriados, no arquivo <i>calendar_dates.txt</i> . |
| <i>tuesday</i> | Obrigatório | Contém um valor binário que indica se o serviço é válido para todas as terças-feiras. O valor 1 indica que o serviço está disponível todas as terças-feiras durante o período. O período é especificado utilizando-se os campos <i>start_date</i> e <i>end_date</i> . O valor 0 indica que o serviço não está disponível às terças-feiras no período. |

Continua na próxima página

Tabela 22 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|---|
| wednesday | Obrigatório | <p>Contém um valor binário que indica se o serviço é válido para todas as quartas-feiras.</p> <p>O valor 1 indica que o serviço está disponível todas as quartas-feiras durante o período. O período é especificado utilizando-se os campos <i>start_date</i> e <i>end_date</i>. O valor 0 indica que o serviço não está disponível às quartas-feiras no período.</p> |
| thursday | Obrigatório | <p>Contém um valor binário que indica se o serviço é válido para todas as quintas-feiras.</p> <p>O valor 1 indica que o serviço está disponível todas as quintas-feiras durante o período. O período é especificado utilizando-se os campos <i>start_date</i> e <i>end_date</i>. O valor 0 indica que o serviço não está disponível às quintas-feiras no período.</p> |
| friday | Obrigatório | <p>Contém um valor binário que indica se o serviço é válido para todas as sextas-feiras.</p> <p>O valor 1 indica que o serviço está disponível todas as sextas-feiras durante o período. O período é especificado utilizando-se os campos <i>start_date</i> e <i>end_date</i>. O valor 0 indica que o serviço não está disponível às sextas-feiras no período.</p> |

Continua na próxima página

Tabela 22 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|--|
| <i>saturday</i> | Obrigatório | Contém um valor binário que indica se o serviço é válido para todos os sábados. O valor 1 indica que o serviço está disponível todos os sábados durante o período. O período é especificado utilizando-se os campos <i>start_date</i> e <i>end_date</i> . O valor 0 indica que o serviço não está disponível aos sábados no período. |
| <i>sunday</i> | Obrigatório | Contém um valor binário que indica se o serviço é válido para todos os domingos. O valor 1 indica que o serviço está disponível todos os domingos durante o período. O período é especificado utilizando-se os campos <i>start_date</i> e <i>end_date</i> . O valor 0 indica que o serviço não está disponível aos sábados no período. |
| <i>start_date</i> | Obrigatório | O campo <i>start_date</i> contém a data de início do serviço. O valor do campo <i>start_date</i> deve estar no formato YYYYMMDD. |
| <i>end_date</i> | Obrigatório | O campo <i>end_date</i> contém a data final do serviço. Essa data está incluída no intervalo do serviço. O valor do campo <i>end_date</i> deve estar no formato AAAAMMDD. |

Fonte: Google Transit (adaptada)¹

Tabela 23 – Detalhamento dos campos do arquivo *calendar_dates.txt* da GTFS

| | | |
|-----------------------|-------------|---|
| <i>service_id</i> | Obrigatório | Contém um ID que identifica um conjunto de datas em que uma exceção ao serviço está disponível para um ou mais trajetos. Cada par (<i>service_id</i> , <i>date</i>) pode aparecer somente uma vez em <i>calendar_dates.txt</i> . Se um valor de <i>service_id</i> aparece nos arquivos <i>calendar.txt</i> e <i>calendar_dates.txt</i> , as informações contidas em <i>calendar_dates.txt</i> modifica as informações de serviço especificadas em <i>calendar.txt</i> . Este campo é indicado pelo arquivo <i>trips.txt</i> . |
| <i>date</i> | Obrigatório | Especifica uma data específica em que a disponibilidade do serviço é diferente do normal. Você pode usar o campo <i>exception_type</i> para indicar se o serviço está disponível na data especificada. O valor do campo <i>date</i> deve estar no formato AAAAMMDD. |
| <i>exception_type</i> | Obrigatório | Indica se o serviço está disponível na data especificada no arquivo <i>date</i> . O valor 1 indica que o serviço foi adicionado para a data especificada. O valor 2 indica que o serviço foi removido para a data especificada. |

Fonte: Google Transit (adaptada)¹

Tabela 24 – Detalhamento dos campos do arquivo *fare_attributes.txt* da GTFS

| | | |
|--------------------------|-------------|--|
| <i>fare_id</i> | Obrigatório | Contém um ID que identifica uma classe de tarifas. |
| <i>price</i> | Obrigatório | Contém o preço da tarifa, na unidade especificada por <i>currency_type</i> . |
| <i>currency_type</i> | Obrigatório | Define a moeda usada para pagar a tarifa. Use os códigos de moeda em ordem alfabética ISO 4217. |
| <i>payment_method</i> | Obrigatório | Indica quando a tarifa deve ser paga. Os valores válidos deste campo são: 0 - A tarifa é paga a bordo; 1 - A tarifa deve ser paga antes do embarque. |
| <i>transfers</i> | Obrigatório | O campo <i>transfers</i> especifica o número de baldeações permitidas nesta tarifa. Os valores válidos deste campo são: 0 - Não são permitidas baldeações nesta tarifa; 1 - Os passageiros só podem fazer uma baldeação; 2 - Os passageiros podem fazer duas baldeações; (empty) - Se o campo estiver vazio, não há limites para o número de baldeações. |
| <i>transfer_duration</i> | Opcional | Especifica a duração, em segundos, antes da expiração da baldeação. Quando usado com um valor 0 para <i>transfers</i> , o campo <i>transfer_duration</i> indica por quanto tempo uma passagem é válida para uma tarifa quando as baldeações não são permitidas. A menos que você pretenda usar este campo para indicar a validade da passagem, <i>transfer_duration</i> deve ser omitido ou deve ficar em branco, quando <i>transfers</i> é definido como 0. |

Fonte: Google Transit (adaptada)¹

Tabela 25 – Detalhamento dos campos do arquivo *fare_rules.txt* da GTFS

| | | |
|-----------------------|-------------|---|
| <i>fare_id</i> | Obrigatório | Contém um ID que identifica uma classe de tarifas. Este valor é indicado no arquivo <i>fare_attributes.txt</i> . |
| <i>route_id</i> | Opcional | Associa o ID da tarifa a um trajeto. Os IDs de trajetos são indicados no arquivo <i>routes.txt</i> . Se você tem diversos trajetos com os mesmos atributos de tarifa, crie uma linha no arquivo <i>fare_rules.txt</i> para cada trajeto. |
| <i>origin_id</i> | Opcional | Associa o ID da tarifa a um ID de zona de origens. Os IDs de zona são indicados no arquivo <i>stops.txt</i> . Se há vários IDs de origem com os mesmos atributos, crie uma linha no arquivo <i>fare_rules.txt</i> para cada ID de origem. |
| <i>destination_id</i> | Opcional | Associa o ID da tarifa a um ID de zona de destino. IDs de zona são indicados no arquivo <i>stops.txt</i> . Se há vários IDs de destino com os mesmos atributos de tarifa, cria-se uma linha no arquivo <i>fare_rules.txt</i> para cada ID de destino. |
| <i>contains_id</i> | Opcional | Associa o ID da tarifa a um ID de zona ID, indicado no arquivo <i>stops.txt</i> . O ID da tarifa é, então, associado a itinerários que transmitem cada zona de <i>contains_id</i> . |

Fonte: Google Transit (adaptada)¹

Tabela 26 – Detalhamento dos campos do arquivo *shapes.txt* da GTFS

| | | |
|----------------------------|-------------|--|
| <i>shape_id</i> | Obrigatório | Contém um ID que identifica uma forma. |
| <i>shape_pt_lat</i> | Obrigatório | Associa a latitude de um ponto de forma ao ID de uma forma. O valor do campo deve ser uma latitude WGS 84 válida. Cada linha do arquivo <i>shapes.txt</i> representa um ponto de forma em sua definição de formas. |
| <i>shape_pt_lon</i> | Obrigatório | Associa a longitude de um ponto de forma ao ID de uma forma. O valor do campo deve ser uma longitude WGS 84 de valor de -180 a 180. Cada linha do arquivo <i>shapes.txt</i> representa um ponto de forma em sua definição de formas. |
| <i>shape_pt_sequence</i> | Obrigatório | Associa a latitude e a longitude de uma forma de um ponto de formas com sua ordem sequencial juntamente com a forma. Os valores de <i>shape_pt_sequence</i> devem ser números inteiros positivos e devem aumentar com a viagem. |
| <i>shape_dist_traveled</i> | Opcional | Quando usado no arquivo <i>shapes.txt</i> , o campo <i>shape_dist_traveled</i> posiciona um ponto de forma como uma distância percorrida juntamente com uma forma a partir do primeiro ponto de forma. O campo <i>shape_dist_traveled</i> representa uma distância real percorrida ao longo do trajeto em unidades como, por exemplo, pés ou quilômetros. Esta informação permite que o planejador de viagens determine o quanto da forma deve ser desenhado ao mostrar parte de uma viagem no mapa. Os valores usados para <i>shape_dist_traveled</i> devem aumentar juntamente com <i>shape_pt_sequence</i> . As unidades usadas para <i>shape_dist_traveled</i> no arquivo <i>shapes.txt</i> devem corresponder às unidades usadas para este campo no arquivo <i>stop_times.txt</i> . |

Fonte: Google Transit (adaptada)¹

Tabela 27 – Detalhamento dos campos do arquivo *frequencies.txt* da GTFS

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|---|
| <i>trip_id</i> | Obrigatório | Contém um ID que identifica uma viagem à qual a frequência especificada de serviço se aplica. Os IDs de viagem são indicados no arquivo <i>trips.txt</i> . |
| <i>start_time</i> | Obrigatório | Especifica o horário em que o serviço começa com a freqüência especificada. Para horários após a meia-noite, insira-os como um valor maior que 24:00:00 no horário local HH:MM:SS para o dia em que a programação das viagens começa. |
| <i>end_time</i> | Obrigatório | Especifica o horário em que o serviço muda para uma frequência diferente (ou é interrompido), na primeira parada da viagem. Para horários após a meia-noite, insira-os como um valor maior que 24:00:00 no horário local HH:MM:SS para o dia em que a programação das viagens começa. |

Tabela 27 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|---|
| <i>headway_secs</i> | Obrigatório | <p>Indica o horário entre as saídas da mesma parada (intervalo entre as viagens) deste tipo de viagem, durante o intervalo de tempo especificado por <i>start_time</i> e <i>end_time</i>. O valor do intervalo de tempo entre duas viagens deve ser inserido em segundos.</p> <p>Períodos em que intervalos entre as viagens são definidos (as linhas no arquivo <i>frequencies.txt</i>) não devem ser sobrepostos para a mesma viagem, uma vez que é difícil determinar o que deve ser inferido de dois intervalos de viagem sobrepostos. No entanto, um período de intervalo entre viagens pode começar exatamente no mesmo horário em que outro termina.</p> |

Tabela 27 - continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|--|
| <i>exact_times</i> | Opcional | <p>Determina se viagens baseadas em frequência devem ser programadas com exatidão com base nas informações especificadas dos intervalos entre as viagens.</p> <p>Os valores válidos deste campo são: 0 ou (vazio) - Viagens baseadas em frequência não são programadas com exatidão. Este é o comportamento padrão; 1 - Viagens baseadas em frequência são programadas com exatidão. Para uma linha no <i>frequencies.txt</i>, as viagens são programadas com início com <i>trip_start_time =start_time + x * headway_secs</i> para todos x em (0, 1, 2, ...), em que <i>trip_start_time < end_time</i>. O valor de <i>exact_times</i> deve ser o mesmo para todas as linhas de <i>frequencies.txt</i> com o mesmo <i>trip_id</i>. Se <i>exact_times</i> for igual a 1, e uma linha de <i>frequencies.txt</i> tiver um <i>start_time</i> igual a <i>end_time</i>, nenhuma viagem deverá ser programada. Quando <i>exact_times</i> é 1, deve-se escolher um valor <i>end_time</i> que seja maior que o último horário de início da viagem programada, mas menor que o último horário de início da viagem desejada + <i>headway_secs</i>.</p> |

Fonte: Google Transit (adaptada)¹

Tabela 28 – Detalhamento dos campos do arquivo
transfer.txt da GTFS

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|--|
| <i>from_stop_id</i> | Obrigatório | Contém um ID que identifica uma parada ou uma estação onde começa uma conexão entre trajetos. Os IDs de paradas são indicados no arquivo <i>stops.txt</i> . Se a ID de parada se refere a uma estação que contém várias paradas, essa regra de baldeação se aplica a todas as paradas nesta estação. |
| <i>to_stop_id</i> | Obrigatório | Contém um ID que identifica uma parada ou uma estação onde termina uma conexão entre trajetos. Os IDs de paradas são indicados no arquivo <i>stops.txt</i> . Se a ID de parada se refere a uma estação que contém várias paradas, essa regra de baldeação se aplica a todas as paradas nesta estação. |
| <i>transfer_type</i> | Obrigatório | Especifica o tipo de conexão para o par (<i>from_stop_id, to_stop_id</i>) especificado. Os valores válidos deste campo são: 0 ou (vazio) <ul style="list-style-type: none"> - Este é um ponto de baldeação recomendado entre dois trajetos; 1 - Este é um ponto de baldeação programado entre dois trajetos; 2 - Essa baldeação exige um tempo mínimo entre a chegada e a partida para garantir uma conexão. O tempo necessário para a baldeação é especificado por <i>min_transfer_time</i>; 3 - Não é possível fazer baldeações entre trajetos neste local. |

Continua na próxima página

Tabela 28 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|--------------------------|--------------------|---|
| <i>min_transfer_time</i> | Opcional | Quando uma conexão entre trajetos exige um tempo entre a chegada e a partida (<i>transfer_type=2</i>), o campo <i>min_transfer_time</i> define o período de tempo que deve estar disponível em um itinerário para permitir uma baldeação entre trajetos nestas paradas. O <i>min_transfer_time</i> deve ser suficiente para que um passageiro típico se desloque entre as duas paradas, incluindo um tempo extra para variação na programação em cada trajeto. O valor de <i>min_transfer_time</i> deve ser inserido em segundos e deve ser um número inteiro positivo. |

Fonte: Google Transit (adaptada)¹

Tabela 29 – Detalhamento dos campos do arquivo
feed_info.txt da GTFS

| Nome do campo | Condisional | Descrição |
|----------------------------|--------------------|---|
| <i>feed_publisher_name</i> | Obrigatório | Contém o nome completo da organização que publica o <i>feed</i> . Pode ser o mesmo que aquele definido pelos valores de <i>agency_name</i> no arquivo <i>agency.txt</i> . Aplicativos que utilizam GTFS podem exibir este nome ao concederem atribuições relacionadas aos dados de um <i>feed</i> específico. |
| <i>feed_publisher_url</i> | Obrigatório | Contém o URL do website da organização que está publicando o <i>feed</i> . Pode ser o mesmo que um dos valores de <i>agency_url</i> no arquivo <i>agency.txt</i> . |
| <i>feed_lang</i> | Obrigatório | Contém um código de idiomas IETF BCP 47 que especifica o idioma padrão usado para o texto neste <i>feed</i> . Esta configuração ajuda os consumidores de GTFS a escolherem regras para o uso de letras maiúsculas e minúsculas e outras configurações específicas do idioma para o <i>feed</i> . |

Continua na próxima página

Tabela 29 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|--|--------------------|---|
| <i>feed_start_date / feed_end_date</i> | Opcional | <p>O <i>feed</i> fornece informações completas e confiáveis sobre a programação de um serviço, no período entre o início do dia <i>feed_start_date</i> e o final do dia <i>feed_end_date</i>. As datas nos dois dias estão no formato AAAAMMDD, assim como no arquivo <i>calendar.txt</i>, ou são deixadas em branco se não estiverem disponíveis. A data <i>feed_end_date</i> não deve preceder a data <i>feed_start_date</i>, se ambas forem fornecidas.</p> <p>Os provedores de feeds são encorajados a oferecerem dados de programação fora desse período a fim de informarem sobre possíveis serviços no futuro, mas os consumidores de <i>feed</i> devem estar conscientes de seu status não autorizado. Se <i>feed_start_date</i> ou <i>feed_end_date</i> se estendem além das datas do calendário ativo definidas nos arquivos <i>calendar.txt</i> e <i>calendar_dates.txt</i>, o <i>feed</i> se torna uma afirmação explícita de que não há serviços para as datas entre <i>feed_start_date</i> ou <i>feed_end_date</i> que não estão incluídas nas datas do calendário ativo.</p> |

Continua na próxima página

Tabela 29 – continuação da página anterior

| Nome do campo | Condisional | Descrição |
|----------------------|--------------------|--|
| <i>feed_version</i> | Opcional | O editor de <i>feeds</i> pode especificar uma sequência que indique a versão atual do <i>feed</i> GTFS. Os aplicativos que utilizam GTFS podem exibir este valor para ajudar os editores de <i>feed</i> a determinar se foi incorporada a versão mais recente do <i>feed</i> . |

Fonte: Google Transit (adaptada)¹

Apêndice D – Linhas de ônibus impactadas por eventos de exceção

Tabela 30 – Linhas de ônibus impactadas por eventos de exceção

| Código da linha | Total de eventos de exceção | Letreiro |
|-----------------|-----------------------------|--|
| 33121 | 1623 | TERM. PRINC. ISABEL / TERM. STO. AMARO |
| 32826 | 1502 | TERM. PQ. D. PEDRO II / TERM. JOÃO DIAS |
| 32805 | 1490 | TERM. PRINC. ISABEL / CHÁC. SANTANA |
| 34085 | 1464 | TERM. BANDEIRA / JD. VAZ DE LIMA |
| 34233 | 1418 | TERM. BANDEIRA / TERM. VARGINHA |
| 33123 | 1408 | TERM. BANDEIRA / TERM. STO. AMARO |
| 32829 | 1405 | TERM. BANDEIRA / TERM. CAPELINHA |
| 35174 | 1388 | TERM. PQ. D. PEDRO II / TERM. STO. AMARO |
| 32827 | 1378 | TERM. BANDEIRA / TERM. CAPELINHA |
| 33128 | 1373 | TERM. BANDEIRA / SOCORRO |
| 33129 | 1366 | TERM. BANDEIRA / VL. CRUZEIRO |
| 33389 | 1342 | TERM. PINHEIROS / METRÔ TUCURUVI |
| 32772 | 1324 | TERM. PRINC. ISABEL / TERM. STO. AMARO |
| 33377 | 1310 | PERDIZES / AEROPORTO |
| 33336 | 1308 | PINHEIROS / IMIRIM |
| 33126 | 1306 | TERM. BANDEIRA / INOCOOP CAMPO LIMPO |
| 34861 | 1305 | METRÔ STA. CECÍLIA / TERM. STO. AMARO |
| 34062 | 1291 | TERM. BANDEIRA / JD. LUSO |
| 34789 | 1287 | METRÔ ARMÊNIA / SHOP. MORUMBI |
| 34218 | 1276 | TERM. BANDEIRA / TERM. GUARAPIRANGA |
| 32825 | 1263 | TERM. BANDEIRA / TERM. JOÃO DIAS |
| 34061 | 1255 | PQ. IBIRAPUERA / JD. MIRIAM |
| 32814 | 1230 | TERM. BANDEIRA / TERM. STO. AMARO |
| 34050 | 1230 | PQ. D. PEDRO II / CID. ADEMAR |
| 32816 | 1220 | TERM. PQ. D. PEDRO II / TERM. STO. AMARO |
| 34831 | 1217 | TERM. BANDEIRA / JD. PAULO VI |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 35109 | 1202 | TERM. PINHEIROS / TERM. PQ. D. PEDRO II |
| 33284 | 1199 | ITAIM BIBI / METRÔ SANTANA |
| 34139 | 1196 | TERM. BANDEIRA / CEASA |
| 33236 | 1194 | TERM. BANDEIRA / JD. JAQUELINE |
| 35229 | 1193 | TURISMO / CIRCULAR |
| 34884 | 1181 | BUTANTÃ / TERM. PQ. D. PEDRO II |
| 34048 | 1177 | LGO. SÃO FRANCISCO / JD. SELMA |
| 32885 | 1174 | ACLIMAÇÃO / TERM. PRINC. ISABEL |
| 34883 | 1173 | TERM. PINHEIROS / TERM. PQ. D. PEDRO II |
| 34064 | 1170 | PQ. IBIRAPUERA / JD. MIRIAM |
| 34076 | 1164 | TERM. PQ. D. PEDRO II / TERM. GUARAPI-RANGA |
| 32813 | 1144 | PÇA. DA SÉ / CHÁC. SANTANA |
| 32892 | 1140 | ACLIMAÇÃO / TERM. PRINC. ISABEL |
| 34685 | 1138 | TERM. BANDEIRA / TERM. CAMPO LIMPO |
| 32769 | 1135 | LGO. SÃO FRANCISCO / TERM. CAPELI-NHA |
| 33258 | 1131 | LGO. DA PÓLVORA / JD. MARIA LUIZA |
| 34100 | 1121 | TERM. PRINC. ISABEL / CID. UNIVERSITÁRIA |
| 33363 | 1110 | PÇA. JOÃO MENDES / JD. MIRIAM |
| 34210 | 1099 | LGO. SÃO FRANCISCO / TERM. VARGINHA |
| 32838 | 1096 | PÇA. DA SÉ / PQ. RES. COCAIA |
| 34138 | 1082 | TERM. PQ. D. PEDRO II / TERM. PINHEIROS |
| 33253 | 1077 | METRÔ BELÉM / JD. BONFIGLIOLI |
| 32837 | 1074 | PÇA. DO CORREIO / SESC/ORION |
| 35197 | 1069 | TERM. PQ. D. PEDRO II / TERM. PINHEIROS |
| 33075 | 1062 | LAPA / IPIRANGA |
| 32849 | 1058 | LGO. SÃO FRANCISCO / VL. SÃO JOSÉ |
| 32846 | 1056 | METRÔ BRÁS / TERM. GRAJAÚ |
| 33112 | 1056 | TERM. PQ. D. PEDRO II / JD. SÃO SAVÉRIO |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 35208 | 1052 | STA. CECÍLIA / TERM. VL. MARIANA |
| 34134 | 1048 | METRÔ ANA ROSA / MORRO GRANDE |
| 34045 | 1047 | TERM. PRINC. ISABEL / JD. MIRIAM |
| 33443 | 1046 | ANA ROSA / METRÔ SANTANA |
| 33357 | 1030 | METRÔ ANA ROSA / VL. BRASILÂNDIA |
| 33457 | 1026 | METRÔ VL. MADALENA / PQ. EDÚ CHAVES |
| 35160 | 1019 | TERM. PQ. D. PEDRO II / TERM. GRAJAÚ |
| 33425 | 1015 | CID. UNIVERSITÁRIA / METRÔ SANTANA |
| 33117 | 1012 | POMPÉIA ATÉ VL. ROMANA / SACOMÃ |
| 34660 | 1008 | ACLIMAÇÃO / TERM. CAMPO LIMPO |
| 35207 | 988 | STA. CECÍLIA / TERM. VL. MARIANA |
| 33264 | 983 | EST. DA LUZ / JD. BOA VISTA |
| 32939 | 976 | LGO. SÃO FRANCISCO / JD. ÂNGELA |
| 32831 | 972 | LGO. SÃO FRANCISCO / TERM. CAPELINA |
| 33131 | 964 | HOSP. DAS CLÍNICAS / TERM. STO. AMARO |
| 34098 | 960 | TERM. PQ. D. PEDRO II / CID. UNIVERSITÁRIA |
| 35276 | 955 | PÇA. RAMOS DE AZEVEDO / TERM. CAMPO LIMPO |
| 33538 | 953 | PAULISTA / PARAISÓPOLIS |
| 33111 | 947 | TERM. AMARAL GURGEL / JD. DA SAÚDE |
| 33391 | 943 | METRÔ JABAQUARA / METRÔ SANTANA |
| 33122 | 939 | TERM. PQ. D. PEDRO II / TERM. STO. AMARO |
| 35175 | 939 | TERM. PQ. D. PEDRO II / TERM. STO. AMARO |
| 33328 | 935 | HOSP. DAS CLÍNICAS / LAUZANE PAULISTA |
| 33114 | 929 | TERM. PINHEIROS / SACOMÃ |
| 32897 | 924 | LUZ / TERM. A. E. CARVALHO |
| 33280 | 914 | PÇA. RAMOS DE AZEVEDO / JD. JOÃO XXII-I/EDUC. |
| 33272 | 912 | PÇA. RAMOS DE AZEVEDO / JD. JOÃO XXIII |
| 34832 | 906 | TERM. PRINC. ISABEL / RIO PEQUENO |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 33275 | 902 | METRÔ ANA ROSA / JD. GUARAÚ |
| 34144 | 902 | PÇA. DA SÉ / CID. UNIVERSITÁRIA |
| 35280 | 902 | TERM. PQ. D. PEDRO II / TERM. PINHEIROS |
| 35196 | 894 | TERM. PQ. D. PEDRO II / METRÔ BUTANTÃ |
| 34694 | 890 | PARAÍSO / TERM. CAMPO LIMPO |
| 33398 | 884 | CID. UNIVERSITÁRIA / METRÔ SANTANA |
| 33042 | 879 | PÇA. DA SÉ / JD. IV CENTENÁRIO |
| 33277 | 870 | TERM. PRINC. ISABEL / COHAB RAPOSO TAVARES |
| 34840 | 870 | ANHANGABAÚ / SHOP. CONTINENTAL |
| 34149 | 869 | METRÔ PARAÍSO / VL. ANASTÁCIO |
| 33224 | 861 | METRÔ VL. MARIANA / TERM. PIRITUBA |
| 33116 | 860 | RIO PEQUENO / IPIRANGA |
| 34196 | 855 | SOCORRO / LAPA |
| 34108 | 832 | METRÔ VL. MARIANA / TERM. LAPA |
| 33361 | 831 | PÇA. DA SÉ / BALN. SÃO FRANCISCO |
| 32884 | 826 | TERM. PQ. D. PEDRO II / TERM. CASA VERDE |
| 33366 | 820 | PÇA. JOÃO MENDES / ELDORADO |
| 33239 | 819 | PÇA. RAMOS DE AZEVEDO / PQ. CONTINENTAL |
| 34101 | 819 | PÇA. RAMOS DE AZEVEDO / MERCADO DA LAPA |
| 33343 | 812 | PÇA. DO CORREIO / JD. GUARANI |
| 34283 | 810 | PÇA. JOÃO MENDES / ELDORADO |
| 35148 | 804 | METRÔ VL. MADALENA / TERM. SACOMÃ |
| 35072 | 800 | METRÔ BARRA FUNDA / CONEXÃO PETRÔNIO PORTELA |
| 35085 | 793 | TERM. PQ. D. PEDRO II / TERM. CASA VERDE |
| 33534 | 791 | CARDOSO DE ALMEIDA / MACHADO DE ASSIS |
| 33198 | 790 | PÇA. DO CORREIO / CID. D'ABRIL 3 ^a GLEBA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 35050 | 786 | TERM. PQ. D. PEDRO II / TERM. LAPA |
| 33486 | 782 | TERM. PQ. D. PEDRO II / TERM. SÃO MARTEUS |
| 33342 | 776 | PÇA. DO CORREIO / JD. PAULISTANO |
| 33090 | 771 | PÇA. DA REPÚBLICA / SHOP. PLAZA SUL |
| 33356 | 770 | PÇA. DO CORREIO / PEDRA BRANCA |
| 33763 | 767 | PÇA. JOÃO MENDES / JD. VL. FORMOSA |
| 34102 | 765 | PÇA. RAMOS DE AZEVEDO / LAPA |
| 34109 | 757 | METRÔ ANA ROSA / METRÔ BARRA FUNDA |
| 32869 | 753 | PINHEIROS / GRAJAÚ |
| 34107 | 753 | TERM. PQ. D. PEDRO II / PQ. DA LAPA |
| 33348 | 748 | PÇA. DO CORREIO / TAIPAS |
| 34393 | 743 | PÇA. DO CORREIO / TERM. SAPOPEMBA |
| 34200 | 741 | LGO. DO PAISSANDÚ / TERM. PIRITUBA |
| 33200 | 737 | PÇA. RAMOS DE AZEVEDO / CID. D'ABRIL |
| 33230 | 732 | LGO. DO PAISSANDÚ / TERM. CACHOEIRINHA |
| 33476 | 726 | PÇA. DO CORREIO / TERM. CACHOEIRINHA |
| 34195 | 724 | PÇA. RAMOS DE AZEVEDO / APIACÁS |
| 33130 | 723 | METRÔ ANA ROSA / TERM. STO. AMARO |
| 34127 | 716 | PÇA. DO CORREIO / FREGUESIA DO Ó |
| 35104 | 716 | TERM. PQ. D. PEDRO II / TERM. A. E. CARVALHO |
| 32934 | 714 | TERM. PQ. D. PEDRO II / JD. SÃO PAULO |
| 33077 | 714 | BOM RETIRO / PQ. SÃO LUCAS |
| 33170 | 714 | TERM. PQ. D. PEDRO II / ITAIM PAULISTA |
| 33211 | 714 | LGO. DO PAISSANDÚ / JD. LÍBANO |
| 33206 | 707 | PÇA. RAMOS DE AZEVEDO / MORRO DOCE |
| 35011 | 704 | METRÔ - TRIANON - MASP / VL. GOMES |
| 34128 | 703 | PÇA. DO CORREIO / BRASILÂNDIA |
| 33214 | 702 | LGO. DO PAISSANDÚ / MANGALOT |
| 32834 | 700 | TERM. PINHEIROS / TERM. CAPELINHA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 33229 | 699 | PÇA. DO CORREIO / TERM. CACHOEIRINHA |
| 32871 | 696 | PINHEIROS / VL. SÃO JOSÉ |
| 35051 | 696 | TERM. PQ. D. PEDRO II / TERM. LAPA |
| 35163 | 691 | TERM. PQ. D. PEDRO II / METRÔ JABAQUARA |
| 32953 | 689 | TERM. PINHEIROS / TERM. JD. ÂNGELA |
| 34033 | 682 | PÇA. RAMOS DE AZEVEDO / TERM. PIRITUBA |
| 34942 | 680 | TERM. PQ. D. PEDRO II / INÁCIO MONTEIRO |
| 32900 | 672 | PÇA. DO CORREIO / SÃO MIGUEL |
| 33089 | 672 | TERM. PQ. D. PEDRO II / VL. GUMERCINDO |
| 33966 | 662 | METRÔ VL. MARIANA / TERM. PARELHAIROS |
| 33365 | 660 | PÇA. JOÃO MENDES / DIV. DIADEMA |
| 34941 | 659 | TERM. PQ. D. PEDRO II / TERM. CID. TIRADENTES |
| 33506 | 655 | TERM. PQ. D. PEDRO II / SÃO MATEUS |
| 33536 | 651 | PÇA. DA REPÚBLICA / GENTIL DE MOURA |
| 35143 | 651 | TERM. PQ. D. PEDRO II / TERM. SÃO MATEUS |
| 35145 | 649 | TERM. PQ. D. PEDRO II / TERM. SÃO MATEUS |
| 33502 | 647 | TERM. PQ. D. PEDRO II / SÃO MATEUS |
| 34940 | 643 | TERM. PQ. D. PEDRO II / JD. MARÍLIA |
| 34396 | 642 | TERM. PQ. D. PEDRO II / TERM. SAOPEMBA |
| 34928 | 641 | TERM. PQ. D. PEDRO II / E.T. ITAQUERA |
| 33088 | 640 | PÇA. DA REPÚBLICA / VL. MONUMENTO |
| 33245 | 640 | METRÔ - TRIANON - MASP / PQ. CONTINENTAL |
| 33058 | 638 | TERM. PQ. D. PEDRO II / PQ. STA. MADALENA |
| 33151 | 634 | TERM. PQ. D. PEDRO II / OLIVEIRINHA |
| 35162 | 634 | TERM. PINHEIROS / METRÔ JABAQUARA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 34761 | 632 | TERM. PINHEIROS / EST. STO. AMARO/-GUIDO CALOI |
| 34939 | 632 | TERM. PQ. D. PEDRO II / TERM. SÃO MATEUS |
| 34394 | 630 | TERM. PQ. D. PEDRO II / TERM. SAPOPEMBA |
| 32833 | 629 | HOSP. DAS CLÍNICAS / TERM. JOÃO DIAS |
| 34086 | 628 | METRÔ SÃO JUDAS / PQ. STO. ANTONIO |
| 33226 | 627 | PÇA. DO CORREIO / TERM. CASA VERDE |
| 33237 | 624 | METRÔ BARRA FUNDA / RIO PEQUENO |
| 32879 | 623 | METRÔ VL. MARIANA / TERM. GRAJAÚ |
| 33146 | 623 | TERM. PQ. D. PEDRO II / JD. CAMARGO VELHO |
| 34938 | 623 | TERM. PQ. D. PEDRO II / TERM. CID. TIRADENTES |
| 33144 | 619 | TERM. PQ. D. PEDRO II / JD. NAZARÉ |
| 33232 | 615 | ITAIM BIBI / COHAB TAIPAS |
| 33535 | 613 | PÇA. DA REPÚBLICA / STA. MARGARIDA MARIA |
| 33078 | 612 | PÇA. ALMEIDA JR. / PQ. STA. MADALENA |
| 34140 | 611 | TERM. PRINC. ISABEL / TERM. PINHEIROS |
| 35081 | 609 | TERM. PQ. D. PEDRO II / METRÔ TUCURUVI |
| 35274 | 608 | PÇA. RAMOS DE AZEVEDO / TERM. LAPA |
| 35110 | 606 | TERM. PQ. D. PEDRO II / METRÔ ITAQUERA |
| 35178 | 605 | TERM. PINHEIROS / TERM. STO. AMARO |
| 35246 | 605 | TERM. PINHEIROS / METRÔ SANTANA |
| 32910 | 597 | TERM. PQ. D. PEDRO II / VL. MARA |
| 34977 | 597 | TERM. MERCADO / TERM. SÃO MATEUS |
| 33093 | 596 | TERM. PQ. D. PEDRO II / JD. PLANALTO |
| 33448 | 594 | METRÔ BARRA FUNDA / JD. FONTÁLIS |
| 34409 | 594 | PÇA. ALMEIDA JR. / TERM. SAPOPEMBA |
| 34443 | 594 | TERM. PQ. D. PEDRO II / JD. CELESTE |
| 33142 | 593 | TERM. PQ. D. PEDRO II / VL. NOVA CURUÇÁ |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 33072 | 592 | TERM. STO. AMARO / IPIRANGA |
| 33461 | 590 | LIBERDADE / PQ. EDÚ CHAVES |
| 34427 | 586 | PÇA. DO CORREIO / TERM. SACOMÃ |
| 33462 | 585 | PÇA. DO CORREIO / PQ. EDÚ CHAVES |
| 34386 | 585 | TERM. PQ. D. PEDRO II / TERM. SÃO MIGUEL |
| 35150 | 585 | TERM. PQ. D. PEDRO II / TERM. SACOMÃ |
| 34804 | 581 | E.T. ÁGUA ESPRAIADA / TERM. GRAJAÚ |
| 35068 | 579 | METRÔ BARRA FUNDA / TERM. PQ. D. PEDRO II |
| 35080 | 579 | TERM. PINHEIROS / METRÔ SANTANA |
| 33441 | 575 | MUSEU DO IPIRANGA / VL. SABRINA |
| 33468 | 574 | PÇA. DO CORREIO / JD. BRASIL |
| 35103 | 574 | TERM. PQ. D. PEDRO II / TERM. A. E. CARVALHO |
| 33439 | 568 | TERM. AMARAL GURGEL / VL. SABRINA |
| 33326 | 566 | LAPA / METRÔ SANTANA |
| 33095 | 565 | TERM. PQ. D. PEDRO II / ZOOLÓGICO |
| 32975 | 564 | TERM. PQ. D. PEDRO II / TERM. A. E. CARVALHO |
| 33372 | 564 | PINHEIROS / VL. CLARA |
| 33482 | 561 | PÇA. DA SÉ / PÇA. SILVIO ROMERO |
| 33481 | 559 | PÇA. DA SÉ / TERM. VL. CARRÃO |
| 33460 | 557 | LIBERDADE / VL. MEDEIROS |
| 32815 | 556 | TERM. PINHEIROS / TERM. STO. AMARO |
| 33000 | 554 | METRÔ VL. MARIANA / PENHA |
| 35278 | 554 | METRÔ STA. CRUZ / TERM. LAPA |
| 35079 | 553 | TERM. PQ. D. PEDRO II / METRÔ TUCURUVI |
| 33680 | 550 | PQ. D. PEDRO II / UNIÃO DE VL. NOVA |
| 32909 | 548 | TERM. PQ. D. PEDRO II / TERM. A. E. CARVALHO |
| 33514 | 546 | TERM. PQ. D. PEDRO II / VL. DALILA |
| 33610 | 545 | CORREIO / PQ. VL. MARIA |
| 33879 | 545 | IBIRAPUERA / JD. ELBA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 33359 | 543 | TERM. PRINC. ISABEL / VOITH |
| 33427 | 540 | PÇA. DO CORREIO / VL. SABRINA |
| 34007 | 540 | ITAIM BIBI / TERM. JD. ÂNGELA |
| 35125 | 539 | TERM. PQ. D. PEDRO II / TERM. VL. CARRÃO |
| 33578 | 536 | BOM RETIRO / JD. ELISA MARIA |
| 33287 | 532 | TERM. AMARAL GURGEL / JD. PERY ALTO |
| 33079 | 530 | PÇA. ALMEIDA JR. / VL. EMA |
| 32903 | 528 | TERM. PQ. D. PEDRO II / JD. DANFER |
| 32776 | 527 | METRÔ ANA ROSA / TERM. CAPELINHA |
| 35230 | 526 | TERM. PINHEIROS / TERM. STO. AMARO |
| 33354 | 525 | TERM. PRINC. ISABEL / COHAB TAIPAS |
| 34788 | 525 | ITAIM BIBI / TERM. GUARAPIRANGA |
| 34090 | 523 | METRÔ VL. MARIANA / TERM. CAPELINHA |
| 34943 | 522 | TERM. PQ. D. PEDRO II / TERM. VL. CARRÃO |
| 34693 | 521 | METRÔ STA. CRUZ / TERM. CAMPO LIMPO |
| 35146 | 512 | TERM. PQ. D. PEDRO II / TERM. SACOMÃ |
| 33852 | 509 | TERM. PQ. D. PEDRO II / JD. COLORADO |
| 34758 | 505 | METRÔ PÇA. DA ÁRVORE / JD. ÂNGELA |
| 35082 | 505 | TERM. PQ. D. PEDRO II / METRÔ TUCURUVI |
| 33191 | 504 | ITAIM BIBI / TERM. PIRITUBA |
| 33255 | 504 | PAULISTA / COHAB EDUCANDÁRIO |
| 33479 | 500 | TERM. BANDEIRA / TERM. PQ. D. PEDRO II |
| 33276 | 494 | METRÔ BARRA FUNDA / JD. ARPOADOR |
| 33034 | 490 | PÇA. D. GASTÃO / JD. MIRIAM |
| 35144 | 489 | TERM. PQ. D. PEDRO II / TERM. SACOMÃ |
| 34395 | 488 | TERM. PRINC. ISABEL / TERM. SAOPEMBA |
| 34650 | 485 | TERM. PQ. D. PEDRO II / TERM. PENHA |
| 34669 | 483 | METRÔ CONCEIÇÃO / TERM. CAMPO LIMPO |
| 34008 | 478 | MORUMBI SHOP. / JD. GUARUJÁ |
| 32966 | 476 | METRÔ STA. CRUZ / TERM. JD. ÂNGELA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 35069 | 467 | TERM. PINHEIROS / CACHOEIRINHA |
| 35147 | 467 | TERM. PINHEIROS / TERM. SACOMÃ |
| 33564 | 457 | HOSP. DAS CLÍNICAS / JD. DAS PALMAS |
| 34903 | 455 | TERM. PINHEIROS / CONEXÃO VL. IÓRIO |
| 32893 | 454 | TERM. PQ. D. PEDRO II / TERM. VL. PRUDENTE |
| 34084 | 454 | TERM. PINHEIROS / COHAB ADVENTISTA |
| 33032 | 448 | PQ. IBIRAPUERA / JD. SELMA |
| 33904 | 448 | SHOP. MORUMBI / METRÔ CONCEIÇÃO |
| 33473 | 447 | PQ. D. PEDRO II / PQ. NOVO MUNDO |
| 35023 | 447 | TERM. PQ. D. PEDRO II / METRÔ SANTANA |
| 33337 | 445 | METRÔ SANTANA / HOSP. CACHOEIRINHA |
| 33628 | 445 | MOOCA / CEM. PQ. DOS PINHEIROS |
| 34083 | 444 | PINHEIROS / VALO VELHO |
| 32836 | 443 | METRÔ SÃO JUDAS / TERM. JOÃO DIAS |
| 34860 | 440 | METRÔ ANA ROSA / E.T. ÁGUA ESPRAIADA |
| 33585 | 437 | METRÔ SANTANA / JD. ALMANARA |
| 33375 | 434 | METRÔ VERGUEIRO / ELDORADO |
| 33558 | 434 | STO. AMARO / REAL PQ. |
| 33539 | 433 | BROOKLIN NOVO / REAL PQ. |
| 34745 | 426 | ITAIM BIBI / JD. MIRIAM |
| 33233 | 423 | ITAIM BIBI / JD. NARDINI |
| 33561 | 423 | E.T. ÁGUA ESPRAIADA / JD. PAULO VI |
| 35252 | 418 | TERM. STO. AMARO / E.T. ÁGUA ESPRAIADA |
| 35083 | 409 | TERM. PINHEIROS / TERM. CACHOEIRINHA |
| 35206 | 408 | METRÔ VL. MARIANA / METRÔ BUTANTÃ |
| 34619 | 399 | TERM. MERCADO / TERM. VL. PRUDENTE |
| 33555 | 398 | CAMPO BELO / PARAISÓPOLIS |
| 34419 | 396 | TERM. MERCADO / TERM. SACOMÃ |
| 33302 | 395 | METRÔ BARRA FUNDA / PEDRA BRANCA |
| 34246 | 393 | METRÔ STA. CRUZ / TERM. STO. AMARO |
| 34527 | 390 | E.T. ÁGUA ESPRAIADA / METRÔ CONCEIÇÃO |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 33251 | 387 | METRÔ BARRA FUNDA / PINHEIROS/VILA IDA |
| 33234 | 386 | TERM. PRINC. ISABEL / TERM. CACHOEIRINHA |
| 33548 | 378 | SHOP. MORUMBI / JD. INGÁ |
| 34684 | 368 | SHOP. MORUMBI / TERM. CAMPO LIMPO |
| 33994 | 366 | STO. AMARO / JD. UNIVERSAL |
| 35084 | 365 | METRÔ VL. MADALENA / METRÔ SANTANA |
| 34051 | 363 | PQ. IBIRAPUERA / VL. STA. CATARINA |
| 33333 | 361 | CEASA / METRÔ SANTANA |
| 33370 | 360 | LGO. CAMBUCI / AMERICANÓPOLIS |
| 32874 | 357 | METRÔ JABAQUARA / PQ. RES. COCAIA |
| 33274 | 356 | HOSP. DAS CLÍNICAS / JD. JOÃO XXIII |
| 34059 | 355 | METRÔ ANA ROSA / JD. MIRIAM |
| 33675 | 354 | ITAIM PAULISTA / VL. CALIFÓRNIA |
| 33516 | 349 | METRÔ BRESSER / CID. TIRADENTES |
| 35067 | 348 | MORRO GRANDE / METRÔ BARRA FUNDA |
| 33450 | 347 | TERM. PRINC. ISABEL / PQ. VL. MARIA |
| 33540 | 346 | HOSP. DAS CLÍNICAS / JD. ROSA MARIA |
| 33455 | 345 | TERM. VL. CARRÃO / JAÇANÃ |
| 33581 | 344 | METRÔ BARRA FUNDA / JD. VISTA ALEGRE |
| 34209 | 344 | METRÔ JABAQUARA / TERM. VARGINHA |
| 35015 | 343 | LGO. DA CONCÓRDIA / JD. FILHOS DA TERRA |
| 33037 | 342 | PQ. IBIRAPUERA / JD. APURÁ |
| 34966 | 341 | METRÔ TATUAPÉ / JD. SOARES |
| 35149 | 341 | METRÔ SANTANA / TERM. SACOMÃ |
| 33429 | 340 | TERM. PRINC. ISABEL / PQ. EDÚ CHAVES |
| 34049 | 340 | TERM. GUARAPIRANGA / JD. MIRIAM |
| 34867 | 340 | STO. AMARO / PARAÍSÓPOLIS |
| 35166 | 340 | MORUMBI SHOP. / METRÔ JABAQUARA |
| 34014 | 338 | SHOP. ARICANDUVA / HOSP. IPIRANGA |
| 34494 | 338 | SHOP. MORUMBI / BUTANTÃ |
| 34856 | 338 | ITAIM BIBI / TERM. LAPA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---------------------------------------|
| 35191 | 338 | TERM. PINHEIROS / TERM. JOÃO DIAS |
| 32855 | 337 | TERM. STO. AMARO / JD. ICARAÍ |
| 32877 | 334 | METRÔ JABAQUARA / GRAJAÚ |
| 34826 | 334 | SHOP. MORUMBI / TERM. CAMPO LIMPO |
| 34043 | 331 | METRÔ STA. CRUZ / CPTM AUTÓDROMO |
| 33472 | 328 | LUZ / CANGAÍBA |
| 33190 | 327 | TERM. PINHEIROS / VL. PIAUÍ |
| 33243 | 324 | ITAIM BIBI / RIO PEQUENO |
| 33657 | 323 | METRÔ BARRA FUNDA / JD. GUARANI |
| 34717 | 322 | LAPA / CAMPO LIMPO |
| 33635 | 319 | PINHEIROS / METRÔ BARRA FUNDA |
| 34425 | 318 | METRÔ VERGUEIRO / TERM. SACOMÃ |
| 34968 | 316 | METRÔ TATUAPÉ / TERM. CID. TIRADENTES |
| 34132 | 314 | METRÔ BARRA FUNDA / PENTEADO |
| 33625 | 307 | METRÔ TATUAPÉ / JD. TREMEMBÉ |
| 33897 | 307 | E.T. ÁGUA ESPRAIADA / JD. SELMA |
| 33474 | 302 | PENHA / METRÔ SANTANA |
| 34398 | 302 | METRÔ BRESSER / HOSP. SAPOPEMBA |
| 33269 | 301 | METRÔ BARRA FUNDA / JD. JOÃO XXIII |
| 33952 | 301 | AEROPORTO / CONJ. HAB. PALMARES |
| 33544 | 299 | PINHEIROS / PARAISÓPOLIS |
| 33656 | 299 | METRÔ BARRA FUNDA / JD. TEREZA |
| 34453 | 299 | TERM. VL. CARRÃO / METRÔ CONCEIÇÃO |
| 33176 | 298 | LAPA / JARAGUÁ |
| 33908 | 298 | TERM. STO. AMARO / TERM. PARELHEIROS |
| 34979 | 298 | MUSEU DO IPIRANGA / SÃO MATEUS |
| 33543 | 295 | PINHEIROS / PQ. ARARIBA |
| 33553 | 294 | STO. AMARO / JD. JAQUELINE |
| 34439 | 293 | JD. ITÁPOLIS / TERM. SACOMÃ |
| 35156 | 293 | JD. ITÁPOLIS / TERM. SACOMÃ |
| 33596 | 291 | METRÔ BARRA FUNDA / VL. TEREZINHA |
| 33015 | 290 | METRÔ TATUAPÉ / VL. SANTANA |
| 33182 | 290 | LAPA / PERUS |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 35013 | 288 | LAPA / JD. BOA VISTA |
| 33922 | 283 | TERM. STO. AMARO / JD. SÃO BERNARDO |
| 34110 | 282 | JAGUARÉ / CITY JARAGUÁ |
| 34990 | 282 | METRÔ BRESSER / CONJ. MANOEL DA NÓ-BREGA |
| 33387 | 281 | SHOP. CENTER NORTE / JD. VISTA ALEGRE |
| 33964 | 280 | TERM. STO. AMARO / JD. HERPLIN |
| 33056 | 276 | MOOCA / PQ. STA. MADALENA |
| 33477 | 276 | SHOP. CENTER NORTE / JD. DAMASCENO |
| 35201 | 276 | TERM. PINHEIROS / TERM. LAPA |
| 33241 | 275 | PINHEIROS / JD. ADALGIZA |
| 35179 | 275 | TERM. PINHEIROS / TERM. CAMPO LIMPO |
| 33859 | 274 | METRÔ BRESSER / JD. ITÁPOLIS |
| 34962 | 274 | LGO. DA CONCÓRDIA / SHOP. ARICANDUVA |
| 34659 | 273 | TERM. PINHEIROS / TERM. CAMPO LIMPO |
| 34397 | 271 | METRÔ BELÉM / JD. WALKIRIA |
| 34058 | 270 | TERM. STO. AMARO / METRÔ JABAQUARA |
| 34391 | 270 | METRÔ BELÉM / TERM. SAPOPEMBA |
| 34857 | 269 | TERM. PINHEIROS / LAPA |
| 33470 | 267 | METRÔ SANTANA / TERM. PENHA |
| 34211 | 267 | TERM. STO. AMARO / TERM. VARGINHA |
| 35157 | 267 | VL. PRUDENTE / METRÔ VL. MARIANA |
| 33943 | 266 | TERM. STO. AMARO / VARGEM GRANDE |
| 33919 | 265 | TERM. GRAJAÚ / JD. CASTRO ALVES |
| 33043 | 264 | METRÔ CONCEIÇÃO / SHOP. SP MARKET |
| 33001 | 263 | METRÔ PENHA / GUAIANAZES |
| 33049 | 261 | SHOP. METRÔ TATUAPÉ / JD. GUIARACÁ |
| 33165 | 261 | METRÔ TATUAPÉ / JD. ROMANO |
| 34191 | 261 | METRÔ BARRA FUNDA / VL. ZATT |
| 34414 | 261 | MOEMA / TERM. SACOMÃ |
| 33017 | 259 | CERET / JD. HELENA |
| 33876 | 259 | METRÔ BELÉM / PQ. BANCÁRIO |
| 33882 | 259 | STO. AMARO / JABAQUARA |
| 35165 | 259 | TERM. STO. AMARO / TERM. GRAJAÚ |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 33982 | 258 | STO. AMARO / JD. MACEDÔNIA |
| 33989 | 258 | TERM. STO. AMARO / JD. D. JOSÉ |
| 35022 | 258 | METRÔ BARRA FUNDA / CID. UNIVERSITÁRIA |
| 32882 | 255 | METRÔ JABAQUARA / JD. STA. BARBARA |
| 33240 | 255 | TERM. LAPA / RIO PEQUENO |
| 33266 | 254 | LAPA / JD. D'ABRIL |
| 33626 | 254 | METRÔ BELÉM / VL. ZILDA |
| 33382 | 253 | METRÔ SANTANA / CPTM JARAGUÁ |
| 33990 | 253 | STO. AMARO / VALO VELHO |
| 35151 | 253 | TERM. SACOMÃ / TERM. SAPOPEMBA |
| 33933 | 252 | TERM. STO. AMARO / JD. PROGRESSO |
| 35209 | 252 | COHAB RAPOSO TAVARES / TERM. PINHEIROS |
| 33668 | 250 | METRÔ BARRA FUNDA / JD. PERY ALTO |
| 33614 | 249 | TIETÊ / JOVA RURAL |
| 34273 | 249 | TERM. STO. AMARO / TERM. GRAJAU |
| 34964 | 249 | METRÔ CARRÃO / JD. NOVA VITÓRIA |
| 35161 | 249 | TERM. STO. AMARO / TERM. GRAJAU |
| 33985 | 248 | STO. AMARO / VALO VELHO |
| 33986 | 248 | STO. AMARO / JD. JANGADEIRO |
| 35014 | 246 | LAPA / COHAB RAPOSO TAVARES |
| 35271 | 246 | METRÔ JABAQUARA / TERM. GUARAPIRANGA |
| 33956 | 245 | TERM. STO. AMARO / JD. ICARAÍ |
| 33426 | 244 | SHOP. D / JD. PRIMAVERA |
| 33893 | 243 | HOSP. SÃO PAULO / JD. MIRIAM |
| 33611 | 242 | LAPA / JD. PERY ALTO |
| 35203 | 242 | PQ. CONTINENTAL / TERM. PINHEIROS |
| 33991 | 241 | STO. AMARO / JD. SÃO BENTO NOVO |
| 33067 | 240 | METRÔ VL. MARIANA / JD. MARIA ESTELA II |
| 33346 | 240 | TERM. LAPA / JD. DOS CUNHAS |
| 33595 | 240 | METRÔ BARRA FUNDA / JD. DOS FRANCOS |
| 33609 | 238 | LAPA / LAUZANE PAULISTA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---------------------------------------|
| 33924 | 238 | TERM. STO. AMARO / JD. ORION |
| 32987 | 237 | CONJ. JOSÉ BONIFÁCIO / PENHA |
| 33984 | 237 | STO. AMARO / JD. DAS ROSAS |
| 34976 | 237 | METRÔ CARRÃO / TERM. SAPOPEMBA |
| 34077 | 236 | STO. AMARO / VALO VELHO |
| 33380 | 235 | METRÔ SANTANA / VL. PENTEADO |
| 33632 | 235 | METRÔ TUCURUVI / JD. MARINA |
| 34423 | 235 | PQ. BELÉM / TERM. SACOMÃ |
| 33371 | 234 | STO. AMARO / METRÔ JABAQUARA |
| 33983 | 234 | STO. AMARO / JD. MITSUTANI |
| 35033 | 234 | LAPA / MANDAQUI |
| 33869 | 232 | METRÔ TAMANDUATEÍ / PQ. STA. MADALENA |
| 33910 | 232 | TERM. STO. AMARO / UNISA-CAMPUS 1 |
| 33987 | 232 | STO. AMARO / JD. TRÊS ESTRELAS |
| 34960 | 232 | TERM. PENHA / CPTM JOSÉ BONIFÁCIO |
| 32824 | 230 | STO. AMARO / CAPÃO REDONDO |
| 34834 | 230 | TERM. PINHEIROS / JD. COLOMBO |
| 34945 | 230 | TERM. VL. CARRÃO / GUAIANAZES |
| 32872 | 229 | TERM. STO. AMARO / PQ. AMÉRICA |
| 33106 | 229 | SHOP. IBIRAPUERA / VL. BRASILINA |
| 35164 | 229 | TERM. STO. AMARO / TERM. GRAJAÚ |
| 33339 | 227 | METRÔ SANTANA / COHAB BRASILÂNDIA |
| 33848 | 226 | METRÔ BELÉM / VL. INDUSTRIAL |
| 34935 | 226 | METRÔ BELÉM / TERM. SÃO MATEUS |
| 35200 | 226 | CEASA / TERM. PINHEIROS |
| 33157 | 224 | METRÔ PENHA / JD. ROMANO |
| 34872 | 224 | SHOP. D / PQ. EDU CHAVES |
| 33311 | 220 | SHOP. D / JD. PERY ALTO |
| 33819 | 220 | METRÔ CARRÃO / 3A. DIVISÃO |
| 33827 | 220 | METRÔ CARRÃO / RES. STA. BÁRBARA |
| 33867 | 220 | VL. PRUDENTE / SÃO MATEUS |
| 34983 | 220 | METRÔ CARRÃO / JD. STO. ANDRÉ |
| 35180 | 220 | TERM. STO. AMARO / TERM. CAPELINHA |
| 33550 | 218 | SHOP. SP MARKET / CAMPO LIMPO |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Leteiro |
|------------------------|------------------------------------|--|
| 33026 | 216 | TERM. VL. CARRÃO / GUAIANAZES |
| 33101 | 216 | METRÔ VL. MARIANA / JD. SÃO SAVÉRIO |
| 33299 | 216 | LAPA / COHAB ANTÁRTICA |
| 33325 | 215 | SHOP. CENTER NORTE / COHAB ANTÁRTICA |
| 33393 | 215 | METRÔ SANTANA / JD. CORISCO |
| 33432 | 215 | METRÔ BELÉM / SHOP. CENTER NORTE |
| 34560 | 215 | METRÔ SANTANA / PEDRA BRANCA |
| 33009 | 214 | METRÔ TATUAPÉ / CID. PEDRO JOSÉ NUNES |
| 34016 | 214 | LAPA / METRÔ BARRA FUNDA |
| 32858 | 213 | TERM. STO. AMARO / JD. GRAUNA |
| 34836 | 213 | TERM. PINHEIROS / COHAB EDUCANDÁRIO |
| 33936 | 212 | SHOP. INTERLAGOS / JD. LUCÉLIA |
| 33981 | 212 | STO. AMARO / VL. GILDA |
| 34668 | 212 | TERM. STO. AMARO / TERM. CAMPO LIMPO |
| 34967 | 212 | METRÔ GUILHERMINA/ESPERANÇA / BARRO BRANCO |
| 33878 | 211 | METRÔ CARRÃO / JD. VERA CRUZ |
| 32999 | 209 | METRÔ PENHA / PARADA XV DE NOVEMBRO |
| 33039 | 209 | TERM. STO. AMARO / VL. IMPÉRIO |
| 33136 | 209 | TERM. PENHA / JD. DAS OLIVEIRAS |
| 33158 | 209 | METRÔ VL. MATILDE / CID. KEMEL II |
| 33314 | 209 | SHOP. CENTER NORTE / VL.NOVA CACHOEIRINHA |
| 35153 | 209 | JD. PLANALTO / TERM. SACOMÃ |
| 33863 | 208 | METRÔ TATUAPÉ / VL. CALIFÓRNIA |
| 34837 | 208 | TERM. PINHEIROS / JD. D'ABRIL |
| 34904 | 208 | METRÔ SANTANA / VL. SABRINA |
| 33914 | 207 | SHOP. INTERLAGOS / JD. SÃO BERNARDO |
| 34406 | 207 | SHOP. METRÔ TATUAPÉ / DIV. SÃO CAETANO |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 32923 | 206 | CERET / TERM. A. E. CARVALHO |
| 32944 | 206 | TERM. STO. AMARO / TERM. CAPELINHA |
| 33244 | 206 | SESC POMPÉIA / PQ. CONTINENTAL |
| 33412 | 206 | METRÔ SANTANA / CACHOEIRA |
| 35167 | 206 | JD. LUSO / TERM. STO. AMARO |
| 33597 | 205 | METRÔ BARRA FUNDA / JD. PAULISTANO |
| 33770 | 205 | PENHA / JD. MARÍLIA |
| 33139 | 203 | TERM. ARICANDUVA / CID. KEMEL |
| 33992 | 203 | STO. AMARO / JD. LÍDIA |
| 35177 | 203 | TERM. STO. AMARO / TERM. CAPELINHA |
| 32820 | 202 | TERM. STO. AMARO / TERM. CAPELINHA |
| 33906 | 202 | METRÔ CONCEIÇÃO / PQ. PRIMAVERA |
| 35202 | 202 | JD. JOÃO XXIII / TERM. PINHEIROS |
| 33515 | 201 | TERM. PENHA / TERM. SÃO MATEUS |
| 33556 | 201 | STO. AMARO / PARAISÓPOLIS |
| 33653 | 201 | LAPA / VL. TEREZINHA |
| 35128 | 201 | TERM. PENHA / TERM. SÃO MATEUS |
| 34053 | 200 | TERM. STO. AMARO / JD. LUSO |
| 34936 | 200 | METRÔ CARRÃO / TERM. SÃO MATEUS |
| 34400 | 199 | METRÔ CARRÃO / TERM. SAPOPEMBA |
| 34105 | 198 | HOSP. DAS CLÍNICAS / LAPA |
| 34407 | 198 | DIV. DE SÃO CAETANO / SÃO MATEUS |
| 35095 | 198 | CEM. PQ. DOS PINHEIROS / METRÔ SANTANA |
| 33030 | 197 | LAR ESC. SÃO FRANCISCO / METRÔ VL. MARIANA |
| 33726 | 197 | SHOP. ARICANDUVA / COHAB JOSÉ BONIFÁCIO |
| 33934 | 197 | SHOP. INTERLAGOS / CANTINHO DO CÉU |
| 34483 | 197 | METRÔ TATUAPÉ / PQ. SÃO LUCAS |
| 34847 | 196 | BUTANTÃ / PQ. IPÊ |
| 33551 | 195 | STO. AMARO / JD. TABOÃO |
| 33972 | 195 | STO. AMARO / PQ. INDEPENDÊNCIA |
| 33074 | 194 | METRÔ VL. MARIANA / HELIÓPOLIS |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 33873 | 194 | TERM. NORTE METRÔ CARRÃO / VL. INDUSTRIAL |
| 33973 | 194 | STO. AMARO / VL. CALÚ |
| 33100 | 193 | METRÔ VL. MARIANA / JD. CLÍMAX |
| 33402 | 193 | METRÔ SANTANA / JD. FONTÁLIS |
| 33870 | 193 | OBJETIVO UNIP / VL. DAS MERCÊS |
| 32954 | 192 | TERM. STO. AMARO / JD. NAKAMURA |
| 32876 | 190 | METRÔ JABAQUARA / CENTRO SESC |
| 32994 | 190 | METRÔ ARTUR ALVIM / JD. ROBRU |
| 33166 | 190 | METRÔ PENHA / JD. NAZARÉ |
| 33741 | 190 | METRÔ BELÉM / JD. ITÁPOLIS |
| 33885 | 190 | STO. AMARO / JD. LUSO |
| 34035 | 190 | METRÔ BARRA FUNDA / TERM. PIRITUBA |
| 35091 | 190 | LGO. DO PERY / METRÔ TUCURUVI |
| 33011 | 189 | METRÔ VL. MATILDE / CPTM JOSÉ BONIFÁCIO |
| 33451 | 189 | METRÔ BELÉM / CENTER NORTE |
| 33630 | 189 | METRÔ BELÉM / VL. CONSTANÇA |
| 35034 | 189 | METRÔ CARANDIRU / JD. BRASIL |
| 33421 | 188 | METRÔ SANTANA / JD. FONTÁLIS |
| 33731 | 188 | SHOP. ARICANDUVA / VL. MINERVA |
| 34010 | 188 | STO. AMARO / JD. CAPELA |
| 34851 | 188 | BUTANTÃ / JD. INGÁ |
| 34937 | 188 | METRÔ PENHA / TERM. CID. TIRADENTES |
| 33411 | 187 | METRÔ SANTANA / VL. NOVA GALVÃO |
| 33761 | 187 | METRÔ TAMANDUATEÍ / SHOP. ARICANDUVA |
| 34137 | 187 | METRÔ BARRA FUNDA / TERM. CACHOEIRINHA |
| 34171 | 187 | METRÔ SANTANA / CEM. PQ. DOS PINHEIROS |
| 34557 | 187 | METRÔ SANTANA / JD. CABUÇU |
| 33566 | 186 | LAPA / PERUS |
| 33788 | 186 | METRÔ PENHA / COHAB JOSÉ BONIFÁCIO |
| 33874 | 186 | METRÔ STA. CRUZ / SACOMÃ |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 33975 | 186 | STO. AMARO / PQ. CEREJEIRA |
| 35097 | 186 | CACHOEIRA / METRÔ SANTANA |
| 32913 | 185 | METRÔ TATUAPÉ / VL. CISPER |
| 32932 | 185 | TERM. SÃO MATEUS / JD. HELENA |
| 33651 | 185 | LAPA / JD. PAULISTANO |
| 33086 | 184 | METRÔ VL. MARIANA / VL. MONUMENTO |
| 33577 | 183 | LAPA / CAPELA DA LAGOA |
| 33549 | 182 | STO. AMARO / JD. INGÁ |
| 33613 | 182 | CARANDIRU / JOVA RURAL |
| 33871 | 181 | NOVA CONQUISTA / JD. GUAIRACÁ |
| 33974 | 181 | STO. AMARO / JD. NAKAMURA |
| 34239 | 181 | PENHA / JD. VL. NOVA |
| 34812 | 181 | METRÔ BUTANTÃ / TERM. CAMPO LIMPO |
| 34835 | 181 | TERM. PINHEIROS / RIO PEQUENO |
| 35105 | 181 | TERM. ARICANDUVA / TERM. A. E. CARVALHO |
| 32912 | 180 | METRÔ TATUAPÉ / ERMELINO MATA-RAZZO |
| 33051 | 180 | METRÔ BELÉM / JD. IMPERADOR |
| 33104 | 180 | METRÔ STA. CRUZ / JD. CELESTE |
| 33406 | 180 | SHOP. CENTER NORTE / VL. ALBERTINA |
| 35099 | 180 | JD. CAMPO LIMPO / METRÔ SANTANA |
| 33929 | 179 | CPTM JURUBATUBA / JD. GAIOTAS |
| 33143 | 178 | TERM. ARICANDUVA / VL. CURUÇÁ |
| 33627 | 178 | METRÔ BELÉM / JAÇANÃ |
| 33794 | 178 | METRÔ CARRÃO / SHOP. ARICANDUVA |
| 34842 | 178 | BUTANTÃ / CDHU MUNCK |
| 33615 | 177 | SHOP. CENTER NORTE / JD. FONTÁLIS |
| 34291 | 177 | METRÔ SÃO JUDAS / JD. UBIRAJARA |
| 34852 | 177 | TERM. STO. AMARO / JD. CAIÇARA |
| 32956 | 176 | TERM. STO. AMARO / TERM. JD. JACIRA |
| 33988 | 176 | TERM. STO. AMARO / JD. CAPELINHA |
| 35066 | 176 | TERM. CASA VERDE / TERM. PIRITUBA |
| 32964 | 175 | TERM. STO. AMARO / JD. ARACATI |
| 33188 | 175 | LAPA / PQ. SÃO DOMINGOS |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 33739 | 175 | METRÔ TATUAPÉ / VL. GUARANI |
| 35134 | 175 | METRÔ BELÉM / TERM. VL. CARRÃO |
| 35173 | 175 | ELDORADO / TERM. STO. AMARO |
| 33044 | 174 | VL. PRUDENTE / PQ. BANCÁRIO |
| 33734 | 174 | METRÔ TATUAPÉ / JD. DAS ROSAS |
| 33405 | 173 | MANDAQUI / CEM. PQ. DOS PINHEIROS |
| 35131 | 173 | METRÔ BELÉM / TERM. VL. CARRÃO |
| 35199 | 173 | PQ. CONTINENTAL / TERM. LAPA |
| 35266 | 173 | METRÔ BELÉM / PQ. EDÚ CHAVES |
| 33631 | 172 | CANTAREIRA / JD. GUANÇÃ |
| 35012 | 171 | LAPA / VL. DALVA |
| 33708 | 170 | METRÔ PENHA / LIMOEIRO |
| 34260 | 170 | JABAQUARA / SHOP. INTERLAGOS |
| 34036 | 169 | TERM. LAPA / ITABERABA |
| 35090 | 169 | VL. SABRINA / METRÔ SANTANA |
| 33186 | 168 | TERM. LAPA / VL. PIAUÍ |
| 33307 | 168 | METRÔ SANTANA / VL. DIONISIA |
| 34463 | 167 | METRÔ SANTANA / TERM. CACHOEIRINHA |
| 34949 | 167 | TERM. VL. CARRÃO / COHAB JUSCELINO |
| 35087 | 167 | METRÔ SANTANA / TERM. CACHOEIRINHA |
| 34402 | 166 | METRÔ ALTO DO IPIRANGA / CONJ. HAB. HELIÓPOLIS |
| 34405 | 166 | VL. ALPINA / METRÔ BRESSER |
| 34882 | 166 | METRÔ ARTUR ALVIM / CONJ. ENCOSTA NORTE |
| 34969 | 166 | METRÔ BELÉM / TERM. VL. CARRÃO |
| 35035 | 166 | SANTANA / VL. NOVA GALVÃO |
| 33006 | 165 | METRÔ PATRIARCA / GUAIANAZES |
| 33624 | 165 | TATUAPÉ / JD. BRASIL |
| 33830 | 165 | METRÔ CARRÃO / JD. STA. TEREZINHA |
| 34056 | 165 | METRÔ CONCEIÇÃO / CID. JÚLIA |
| 34974 | 165 | METRÔ VL. PRUDENTE / PQ. SAVOY CITY |
| 33040 | 164 | TERM. STO. AMARO / VL. GUACURI |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 33174 | 164 | TERM. LAPA / SOL NASCENTE |
| 35141 | 164 | SAVOY/DALILA / TERM. VL. CARRÃO |
| 33160 | 163 | METRÔ ARTUR ALVIM / JD. DAS OLIVEIRAS |
| 34387 | 163 | TERM. SÃO MATEUS / TERM. SÃO MIGUEL |
| 33386 | 162 | METRÔ SANTANA / VL. STA. MARIA |
| 33883 | 162 | STO. AMARO / ELDORADO |
| 35106 | 162 | TERM. ARICANDUVA / TERM. SÃO MIGUEL |
| 32990 | 161 | METRÔ ARTUR ALVIM / PQ. D. JOÃO NERY |
| 33440 | 161 | METRÔ SANTANA / VL. CONSTANÇA |
| 32926 | 160 | TERM. SÃO MATEUS / TERM. A. E. CARVALHO |
| 33180 | 160 | LAPA / PQ. MORRO DOCE |
| 33565 | 160 | LAPA / PQ. CONTINENTAL |
| 33821 | 160 | METRÔ TATUAPÉ / JD. IVA |
| 35267 | 160 | TERM. ROD. TIETÊ / VL. SABRINA |
| 33173 | 159 | METRÔ ITAQUERA / JD. CAMARGO VELHO |
| 33822 | 159 | METRÔ TATUAPÉ / TERM. VL. CARRÃO |
| 33890 | 159 | STO. AMARO / MISSIONÁRIA |
| 33901 | 159 | STO. AMARO / JD. SELMA |
| 34021 | 159 | TERM. LAPA / REMÉDIOS |
| 34030 | 159 | TERM. LAPA / TERM. PIRITUBA |
| 34838 | 159 | BUTANTÃ / VL. SÔNIA |
| 35102 | 159 | JD. PERY ALTO / METRÔ SANTANA |
| 32943 | 158 | SHOP. INTERLAGOS / JD. HERCULANO |
| 33575 | 158 | CEM. VL. NOVA CACHOEIRINHA / PIRITUBA |
| 33599 | 158 | SHOP. CENTER NORTE / CEM. DO HORTO |
| 33673 | 158 | SHOP. PENHA / PQ. PAINEIRAS |
| 34193 | 158 | TERM. LAPA / MORRO GRANDE |
| 33571 | 157 | JD. PRIMAVERA / CPTM VL. AURORA |
| 33812 | 156 | TATUAPÉ / JD. IMPERADOR |
| 35055 | 156 | VL. PIAUÍ / TERM. LAPA |
| 35100 | 156 | JD. BRASIL / METRÔ SANTANA |
| 35126 | 156 | TERM. VL. CARRÃO / METRÔ ITAQUERA |
| 35176 | 156 | TERM. STO. AMARO / TERM. JD. ÂNGELA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 34498 | 155 | TERM. STO. AMARO / TERM. JD. ÂNGELA |
| 35192 | 155 | JD. PLANALTO / TERM. JOÃO DIAS |
| 33189 | 154 | LAPA / VL. CLARICE |
| 33242 | 154 | METRÔ VL. MADALENA / RIO PEQUENO |
| 35060 | 154 | MORRO DOCE / TERM. LAPA |
| 33896 | 153 | METRÔ CONCEIÇÃO / JD. APURÁ |
| 34022 | 153 | TERM. LAPA / STA. MÔNICA |
| 34136 | 153 | LAPA / TERM. CACHOEIRINHA |
| 34818 | 152 | METRÔ BUTANTÃ / JD. JOÃO XXIII |
| 34886 | 152 | TERM. STO. AMARO / JD. ÂNGELA |
| 35268 | 152 | TERM. LAPA / VL. SULINA |
| 33434 | 151 | METRÔ CARANDIRU / VL. SABRINA |
| 33600 | 151 | METRÔ SANTANA / VL. ROSA |
| 33968 | 151 | TERM. STO. AMARO / JD. PLANALTO |
| 33979 | 151 | TERM. STO. AMARO / RIVIERA |
| 34849 | 151 | BUTANTÃ / JD. GUARAÚ |
| 35054 | 151 | TERM. LAPA / TERM. CACHOEIRINHA |
| 32978 | 150 | TERM. ARICANDUVA / JD. COIMBRA |
| 34000 | 150 | TERM. STO. AMARO / JD. SÃO FRANCISCO |
| 34020 | 150 | TERM. LAPA / TERM. PIRITUBA |
| 32992 | 149 | METRÔ ARTUR ALVIM / CPTM JOSÉ BONIFÁCIO |
| 33167 | 149 | METRÔ ITAQUERA / JD. CAMARGO NOVO |
| 33175 | 149 | LAPA / HAB. TURÍSTICA |
| 33320 | 149 | METRÔ SANTANA / JD. ANTÁRTICA |
| 33899 | 149 | STO. AMARO / JD. APURÁ |
| 33784 | 148 | METRÔ TATUAPÉ / VL. STA. ISABEL |
| 35093 | 148 | JD. ANTÁRTICA / METRÔ SANTANA |
| 35049 | 147 | METRÔ TIETÊ / VL. MEDEIROS |
| 33629 | 146 | METRÔ TATUAPÉ / PQ. NOVO MUNDO |
| 33702 | 146 | METRÔ ITAQUERA / JD. NAZARÉ |
| 34944 | 146 | TERM. VL. CARRÃO / ITAQUERA |
| 33310 | 145 | METRÔ SANTANA / JD. ANTÁRTICA |
| 33321 | 145 | METRÔ SANTANA / JD. PERY |
| 34920 | 144 | TERM. STO. AMARO / JD. NAKAMURA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 34972 | 144 | METRÔ CARRÃO / JD. IV CENTENÁRIO |
| 33315 | 143 | METRÔ SANTANA / PEDRA BRANCA |
| 35092 | 142 | PEDRA BRANCA / METRÔ SANTANA |
| 35111 | 142 | OLIVEIRINHA / TERM. A. E. CARVALHO |
| 33646 | 141 | TERM. CACHOEIRINHA / CPTM JARAGUÁ |
| 34019 | 141 | TERM. LAPA / VL. PIAUÍ |
| 35032 | 141 | METRÔ SANTANA / LAUZANE PAULISTA |
| 33573 | 140 | TERM. CACHOEIRINHA / PERUS |
| 34055 | 140 | METRÔ CONCEIÇÃO / VL. MISSIONÁRIA |
| 34454 | 140 | VL. MATIAS / IPIRANGA |
| 35198 | 140 | PQ. DA LAPA / TERM. LAPA |
| 32781 | 139 | TERM. JOÃO DIAS / CAPÃO REDONDO |
| 33568 | 139 | LAPA / TERM. JD. BRITANIA |
| 33686 | 139 | METRÔ VL. MATILDE / CEM. DA SAUDADE |
| 35030 | 139 | METRÔ PARADA INGLESA / HORTO FLORESTAL |
| 34448 | 138 | METRÔ TATUAPÉ / MOOCA |
| 35170 | 138 | VL. MISSIONÁRIA / METRÔ JABAQUARA |
| 33679 | 136 | SHOP. PENHA / BURGO PAULISTA |
| 33817 | 136 | METRÔ ITAQUERA / INÁCIO MONTEIRO |
| 34757 | 136 | TERM. JOÃO DIAS / JD. MARACÁ |
| 34844 | 135 | BUTANTÃ / JD. ROSA MARIA |
| 35122 | 135 | JD. DANFER / TERM. PENHA |
| 35127 | 135 | METRÔ ITAQUERA / TERM. CID. TIRADENTES |
| 32922 | 134 | PENHA / VL. PARANAGUÁ |
| 34133 | 134 | TERM. CASA VERDE / VL. PENTEADO |
| 34237 | 133 | METRÔ PENHA / JD. DANFER |
| 33020 | 132 | COHAB II / JD. HELENA |
| 33140 | 132 | TERM. A. E. CARVALHO / CONJ. ENCOSTA NORTE |
| 33758 | 131 | SÃO MIGUEL PAULISTA / TERM. CID. TIRADENTES |
| 33868 | 131 | SHOP. ARICANDUVA / FAZENDA DA JUTA |
| 33459 | 130 | METRÔ TUCURUVI / PQ. NOVO MUNDO |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 35089 | 130 | PQ. NOVO MUNDO / METRÔ TUCURUVI |
| 35139 | 130 | HOSP. STA. MARCELINA / METRÔ ITAQUERA |
| 32991 | 129 | METRÔ ARTUR ALVIM / JD. HELENA |
| 33135 | 129 | TERM. A. E. CARVALHO / CID. KEMEL II |
| 33687 | 129 | METRÔ ITAQUERA / CHABILÂNDIA |
| 33970 | 129 | TERM. GUARAPIRANGA / CHÁC. STA. MARIA |
| 33094 | 128 | VL. PRUDENTE / VL. INDUSTRIAL |
| 33846 | 128 | SHOP. ARICANDUVA / JD. SÃO FRANCISCO |
| 34839 | 128 | METRÔ BUTANTÃ / PQ. CONTINENTAL |
| 34909 | 127 | CONEXÃO VL. IÓRIO / PERUS |
| 35053 | 127 | TERM. LAPA / TERM. PIRITUBA |
| 35275 | 127 | TERM. LAPA / TERM. PIRITUBA |
| 33787 | 126 | METRÔ ARTUR ALVIM / SHOP. ARICANDUVA |
| 34950 | 126 | TERM. VL. CARRÃO / JD. CIBELE |
| 35158 | 126 | JD. CELESTE / TERM. SACOMÃ |
| 34733 | 125 | METRÔ ITAQUERA / CPTM ERMELINO MATARAZZO |
| 34824 | 125 | METRÔ SÃO JUDAS / JD. MIRIAM |
| 34907 | 125 | SOCORRO / JD. APURÁ |
| 35098 | 125 | VL. ALBERTINA / METRÔ SANTANA |
| 32927 | 124 | METRÔ ITAQUERA / VL. MARA |
| 33108 | 124 | METRÔ PÇA. DA ÁRVORE / JD. CLÍMAX |
| 33417 | 124 | METRÔ SANTANA / VL. ALBERTINA |
| 33452 | 124 | METRÔ SANTANA / PQ. NOVO MUNDO |
| 33796 | 124 | METRÔ VL. MATILDE / SHOP. ARICANDUVA |
| 33884 | 124 | JABAQUARA / VL. GUACURI |
| 34399 | 124 | SHOP. ARICANDUVA / MASCARENHAS DE MORAIS |
| 34901 | 124 | METRÔ BARRA FUNDA / LIMÃO |
| 34951 | 124 | TERM. VL. CARRÃO / JD. NSA. SRA. DO CARMO |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 32907 | 123 | TERM. ARICANDUVA / BURGO PAULISTA |
| 32981 | 123 | TERM. ARICANDUVA / VL. SÃO FRANCISCO |
| 35052 | 123 | TERM. LAPA / TERM. PIRITUBA |
| 33105 | 122 | METRÔ JABAQUARA / SHOP. PLAZA SUL |
| 33488 | 122 | CIRCULAR / TERM. VL. CARRÃO |
| 34113 | 122 | TERM. PIRITUBA / CEM. DE PERUS |
| 34955 | 122 | TERM. VL. CARRÃO / JD. VL. CARRÃO |
| 35028 | 122 | METRÔ TUCURUVI / CACHOEIRA |
| 35114 | 122 | CPTM GUAIANAZES / TERM. A. E. CARVALHO |
| 35265 | 122 | CONEXÃO VL. IÓRIO / PERUS |
| 33619 | 121 | METRÔ TUCURUVI / JD. JOANA D'ARC |
| 33637 | 121 | TERM. PARADA INGLESA / JD. HEBRON |
| 33887 | 121 | METRÔ JABAQUARA / JD. SÃO JORGE |
| 32799 | 120 | TERM. JOÃO DIAS / TERM. CAPELINHA |
| 33728 | 120 | PQ. SÃO RAFAEL / SHOP. ARICANDUVA |
| 34031 | 120 | TERM. LAPA / TERM. PIRITUBA |
| 32880 | 119 | CPTM GRAJAÚ / JD. ALPINO |
| 33070 | 119 | METRÔ SAÚDE / VL. LIVIERO |
| 35031 | 119 | METRÔ TUCURUVI / VL. AYROSA |
| 35096 | 119 | VL. NOVA GALVÃO / METRÔ TUCURUVI |
| 33617 | 118 | METRÔ TUCURUVI / JD. SÃO JOÃO |
| 33806 | 118 | METRÔ ITAQUERA / COHAB JUSCELINO |
| 33837 | 118 | METRÔ ARTUR ALVIM / SHOP. ARICANDUVA |
| 33918 | 118 | TERM. GRAJAÚ / JD. MARILDA |
| 35058 | 118 | TAIPAS / TERM. CACHOEIRINHA |
| 33403 | 117 | METRÔ TUCURUVI / JD. FILHOS DA TERRA |
| 33645 | 117 | TERM. CACHOEIRINHA / COHAB BRASILÂNDIA |
| 33811 | 117 | METRÔ ITAQUERA / SÃO MATEUS |
| 33110 | 116 | METRÔ SÃO JUDAS / JD. CLÍMAX |
| 33312 | 116 | METRÔ SANTANA / LAUZANE PAULISTA |
| 33418 | 116 | METRÔ SANTANA / VL. MARIETA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 33652 | 116 | TERM. CASA VERDE / PQ. TIETÊ |
| 33684 | 116 | METRÔ ITAQUERA / JD. ROBRU |
| 34435 | 116 | TERM. SACOMÃ / ÁGUA FUNDA |
| 35010 | 116 | METRÔ SÃO JUDAS / AEROPORTO |
| 35155 | 116 | JD. CELESTE / TERM. SACOMÃ |
| 35219 | 116 | LAPA / VL. IÓRIO |
| 33771 | 115 | E.T. ITAQUERA / COHAB PRES. JUSCELINO KUBITSCHECK |
| 34433 | 115 | TERM. SACOMÃ / VL. BRASILINA |
| 34528 | 115 | METRÔ ITAQUERA / JD. CAMPOS |
| 34760 | 115 | EST. STO. AMARO/GUIDO CALOI / TERM. JD. JACIRA |
| 35205 | 115 | CPTM LEOPOLDINA / METRÔ VL. MADA- LENA |
| 33316 | 114 | METRÔ SANTANA - CIRCULAR / CONJ. DOS BANCÁRIOS |
| 33667 | 114 | CACHOEIRINHA / COHAB ANTÁRTICA |
| 33805 | 114 | METRÔ GUILHERMINA/ESPERANÇA / SHOP. ARICANDUVA |
| 33905 | 114 | METRÔ JABAQUARA / REFÚGIO STA. TE- REZINHA |
| 33713 | 113 | METRÔ ARTUR ALVIM / VL. JACUI |
| 34434 | 113 | TERM. SACOMÃ / JD. CELESTE |
| 35101 | 113 | JD. FLÔR DE MAIO / METRÔ TUCURUVI |
| 35129 | 113 | METRÔ ITAQUERA / TERM. VL. CARRÃO |
| 35169 | 113 | JD. NORONHA / TERM. GRAJAÚ |
| 35204 | 113 | CDHU BUTANTÃ / TERM. JOÃO DIAS |
| 35218 | 113 | LAPA / CONEXÃO VL. IÓRIO |
| 33396 | 112 | METRÔ TUCURUVI / PQ. EDÚ CHAVES |
| 33685 | 112 | METRÔ ITAQUERA / JD. FANGANIELO |
| 33831 | 112 | METRÔ ITAQUERA / COHAB BARRO BRANCO |
| 35038 | 112 | METRÔ TUCURUVI / JD. FONTÁLIS |
| 35094 | 112 | PQ. EDÚ CHAVES / METRÔ TUCURUVI |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 35140 | 112 | COHAB FAZENDA DO CARMO / METRÔ ITAQUERA |
| 33913 | 111 | TERM. GRAJAÚ / JD. ELLUS |
| 34853 | 111 | CID. UNIVERSITÁRIA / METRÔ BUTANTÃ |
| 34913 | 111 | METRÔ ITAQUERA / CPTM GUAIANAZES |
| 34965 | 111 | METRÔ ITAQUERA / JD. SÃO FRANCISCO |
| 35037 | 111 | PQ. NOVO MUNDO / JAÇANÃ |
| 35065 | 111 | JD. CAROMBÉ / TERM. CACHOEIRINHA |
| 33576 | 110 | TERM. PIRITUBA / RECANTO DOS HUMILDES |
| 33957 | 110 | TERM. GRAJAÚ / JD. LUCÉLIA |
| 34791 | 110 | CID. UNIVERSITÁRIA / METRÔ BUTANTÃ |
| 34841 | 110 | BUTANTÃ / JD. MARIA LUIZA |
| 35061 | 110 | PERUS / TERM. PIRITUBA |
| 35108 | 110 | METRÔ ITAQUERA / TERM. SÃO MIGUEL |
| 33643 | 109 | TERM. CACHOEIRINHA / PQ. DE TAIPAS |
| 33935 | 109 | TERM. GRAJAÚ / ILHA DO BORORÉ |
| 33945 | 109 | TERM. GRAJAÚ / JD. DAS PEDRAS |
| 33682 | 108 | METRÔ ARTUR ALVIM / VL. AMERICANA |
| 33807 | 108 | METRÔ ITAQUERA / RECANTO VERDE SOL |
| 33815 | 108 | METRÔ ITAQUERA / COHAB PRESTES MAIA |
| 33954 | 108 | TERM. GRAJAÚ / JD. PRAINHA |
| 35210 | 108 | PQ. ARARIBA / TERM. CAPELINHA |
| 32863 | 107 | TERM. GRAJAÚ / PQ. RES. COCAIA |
| 33633 | 107 | SANTANA / CENTER NORTE |
| 33689 | 107 | METRÔ ITAQUERA / JD. LAJEADO |
| 33953 | 107 | TERM. GRAJAÚ / CANTINHO DO CÉU |
| 33961 | 107 | TERM. GRAJAÚ / JD. GAIOTAS |
| 34501 | 107 | TERM. GRAJAÚ / JD. ELIANA |
| 35168 | 107 | JD. GAIOTAS / TERM. GRAJAÚ |
| 33045 | 106 | VL. PRUDENTE / VL. INDUSTRIAL |
| 33841 | 106 | METRÔ ITAQUERA / JD. SÃO CARLOS |
| 34027 | 106 | TERM. PIRITUBA / CID. D'ABRIL 3ª GLEBA |
| 33021 | 105 | TERM. SÃO MATEUS / GUAIANAZES |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 33103 | 105 | METRÔ SAÚDE / JD. MARIA ESTELA |
| 33705 | 105 | METRÔ ITAQUERA / UNIÃO DE VL. NOVA |
| 35112 | 105 | VL. CISPER (CPTM USP) / TERM. A. E. CARVALHO |
| 33711 | 104 | METRÔ PENHA / JD. DO CASTELO |
| 33810 | 104 | METRÔ ITAQUERA / JD. LARANJEIRA |
| 34550 | 104 | METRÔ TUCURUVI / JD. CABUÇU |
| 35088 | 104 | PEDRA BRANCA / TERM. CACHOEIRINHA |
| 32852 | 103 | TERM. GRAJAÚ / JD. NORONHA |
| 34930 | 103 | BUTANTÃ / VL. DALVA |
| 33877 | 102 | METRÔ SAÚDE / VL. MORAES |
| 35123 | 102 | METRÔ ITAQUERA / TERM. A. E. CARVALHO |
| 35270 | 102 | TERM. GUARAPIRANGA / JD. GUARUJÁ |
| 32774 | 101 | TERM. CAPELINA / SHOP. PORTAL |
| 32797 | 101 | TERM. JOÃO DIAS / JD. CAPELINA |
| 33683 | 101 | METRÔ ITAQUERA / JD. ETELVINA |
| 33835 | 101 | METRÔ ITAQUERA / CID. TIRADENTES |
| 34380 | 101 | METRÔ ITAQUERA / CID. TIRADENTES |
| 35062 | 101 | JD. PRINCESA / TERM. CACHOEIRINHA |
| 35193 | 101 | JD. VAZ DE LIMA / TERM. JOÃO DIAS |
| 33915 | 100 | TERM. GRAJAÚ / PQ. STA. CECÍLIA |
| 33993 | 100 | TERM. GRAJAÚ / PQ. COCAIA |
| 34680 | 100 | TERM. CAMPO LIMPO / PQ. DO LAGO |
| 35120 | 100 | JD. STO. ANTÔNIO / METRÔ ITAQUERA |
| 32902 | 99 | METRÔ ARTUR ALVIM / TERM. A. E. CARVALHO |
| 33777 | 99 | METRÔ ITAQUERA / JD. SÃO JOÃO |
| 33818 | 99 | METRÔ ITAQUERA / BARRO BRANCO |
| 34240 | 99 | TERM. VARGINHA / TERM. GRAJAÚ |
| 34332 | 99 | METRÔ TAMANDUATEÍ / JD. GUAIRACÁ |
| 33560 | 98 | HOSP. CAMPO LIMPO / JD. REBOUÇAS |
| 33699 | 98 | METRÔ PENHA / VL. SÍLVIA |
| 33700 | 98 | METRÔ VL. MATILDE / ERMELINO MATA-RAZZO |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--------------------------------------|
| 34656 | 98 | TERM. CAMPO LIMPO / JD. GUARUJÁ |
| 33736 | 97 | SÃO MATEUS / GUAIANAZES |
| 33823 | 97 | METRÔ ITAQUERA / COHAB II |
| 34431 | 97 | TERM. SACOMÃ / VL. LIVIERO |
| 35115 | 97 | ERMELINO MATARAZZO / TERM. PENHA |
| 33640 | 96 | TERM. CACHOEIRINHA / JD. PRINCESA |
| 33916 | 96 | TERM. GRAJAÚ / VL. NATAL |
| 33735 | 95 | METRÔ ITAQUERA / JD. ALTO PAULISTANO |
| 33743 | 95 | HOSP. SAPOPEMBA / JD. PALANQUE |
| 34848 | 95 | HOSP. CAMPO LIMPO / JD. DAS PALMAS |
| 35135 | 95 | JD. IV CENTENÁRIO / TERM. VL. CARRÃO |
| 32908 | 94 | METRÔ ITAQUERA / JD. STO. ANTÔNIO |
| 34353 | 94 | TERM. GRAJAÚ / VARGEM GRANDE |
| 34689 | 94 | METRÔ PENHA / JD. KERALUX |
| 35172 | 94 | VARGEM GRANDE / TERM. GRAJAÚ |
| 34355 | 93 | TERM. GRAJAÚ / DIVISA DE EMBU-GUAÇU |
| 34429 | 93 | TERM. SACOMÃ / PQ. BRISTOL |
| 34430 | 93 | TERM. SACOMÃ / VL. ARAPUÁ |
| 35119 | 93 | ARTUR ALVIM / METRÔ ITAQUERA |
| 35154 | 93 | VL. ARAPUÁ / TERM. SACOMÃ |
| 33027 | 92 | CPTM GUAIANAZES / TERM. SÃO MIGUEL |
| 34584 | 91 | METRÔ SANTANA / VL. AURORA |
| 34927 | 91 | E.T. ITAQUERA / INÁCIO MONTEIRO |
| 32860 | 90 | TERM. GRAJAÚ / JD. SÃO BERNARDO |
| 33714 | 90 | METRÔ ITAQUERA / VL. PROGRESSO |
| 33888 | 90 | JABAQUARA / VL. STA. MARGARIDA |
| 35171 | 90 | UNISA / TERM. GRAJAÚ |
| 33707 | 89 | JD. SÃO CARLOS / METRÔ ARTUR ALVIM |
| 34029 | 89 | TERM. PIRITUBA / CPTM VL. AURORA |
| 35152 | 89 | HOSP. SÃO MATEUS / TERM. SAPOPEMBA |
| 33018 | 88 | TERM. A. E. CARVALHO / VL. PROGRESSO |
| 33222 | 88 | COHAB TAIPAS / PERUS |
| 33648 | 88 | TERM. CACHOEIRINHA / JD. DAMASCENO |
| 35124 | 88 | VL. CISPER / TERM. PENHA |
| 33716 | 87 | METRÔ ITAQUERA / PQ. GUARANI |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 33793 | 87 | PQ. SAVOY CITY / METRÔ ARTUR ALVIM |
| 32802 | 86 | TERM. CAPELINHA / JD. GUARUJÁ |
| 33797 | 86 | METRÔ ITAQUERA / CPTM JOSÉ BONIFÁ-CIO |
| 33636 | 85 | METRÔ JD. SÃO PAULO / VL. AMÉLIA |
| 33639 | 85 | TERM. CACHOEIRINHA / JD. ELISA MARIA |
| 34024 | 85 | TERM. PIRITUBA / JD. DONÁRIA |
| 35059 | 85 | JD. DONÁRIA / TERM. PIRITUBA |
| 33795 | 84 | METRÔ ITAQUERA / JD. LIMOEIRO |
| 34114 | 84 | TERM. PIRITUBA / JD. RINCÃO |
| 34415 | 84 | METRÔ ITAQUERA / JD. SANTANA |
| 35221 | 84 | CONEXÃO VL. IÓRIO / COHAB BRASILÂN-DIA |
| 33057 | 83 | VL. PRUDENTE / VL. CALIFÓRNIA |
| 34437 | 83 | TERM. SACOMÃ / JD. MARIA ESTELA |
| 34926 | 83 | E.T. ITAQUERA / COHAB FAZENDA DO CARMO |
| 35116 | 83 | JD. CAMARGO VELHO / TERM. SÃO MI-GUEL |
| 32790 | 82 | TERM. CAPELINHA / JD. MACEDÔNIA |
| 35184 | 82 | JD. GUARUJÁ / TERM. CAPELINHA |
| 32773 | 81 | TERM. JOÃO DIAS / JD. IBIRAPUERA |
| 32780 | 81 | TERM. CAPELINHA / VALO VELHO |
| 32800 | 81 | TERM. JOÃO DIAS / JD. NOVO ORIENTE |
| 32969 | 81 | TERM. CAPELINHA / TERM. JD. JACIRA |
| 33697 | 81 | METRÔ PENHA / CHÁC. CRUZ. DO SUL |
| 33715 | 81 | VL. REGINA / METRÔ ARTUR ALVIM |
| 35117 | 81 | JD. CAMARGO VELHO / TERM. SÃO MI-GUEL |
| 32782 | 80 | TERM. CAPELINHA / VALO VELHO |
| 34657 | 80 | TERM. CAMPO LIMPO / VALO VELHO |
| 35183 | 80 | VALO VELHO / TERM. CAPELINHA |
| 35188 | 80 | JD. UNIVERSAL / TERM. CAPELINHA |
| 33809 | 79 | METRÔ ITAQUERA / CPTM D. BOSCO |
| 34880 | 79 | CONJ. CHAPARRAL / METRÔ PENHA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 35190 | 79 | VALO VELHO / TERM. CAPELINHA |
| 33028 | 78 | CPTM GUAIANAZES / SÃO MIGUEL |
| 33710 | 78 | CONJ. A. E. CARVALHO / METRÔ ARTUR ALVIM |
| 33804 | 78 | METRÔ ARTUR ALVIM / JD. NSA. SRA. DO CARMO |
| 34747 | 78 | HOSP. PEDREIRA / CID. DUTRA |
| 32798 | 77 | TERM. JOÃO DIAS / JD. INGÁ |
| 35070 | 77 | CONEXÃO PETRÔNIO PORTELA / JD. CARROMBÉ |
| 32784 | 76 | TERM. CAPELINHA / JD. D. JOSÉ |
| 33895 | 76 | CPTM JURUBATUBA / VL. GUACURI |
| 34238 | 76 | CPTM GUAIANAZES / CPTM JD. ROMANO |
| 35071 | 76 | CONEXÃO PETRÔNIO PORTELA / JD. CARROMBÉ |
| 35107 | 76 | CPTM GUAIANAZES / TERM. SÃO MIGUEL |
| 35194 | 76 | TERM. CAMPO LIMPO / TERM. CAPELINHA |
| 33701 | 75 | METRÔ GUILHERMINA/ESPERANÇA / JD. BELÉM |
| 34205 | 75 | TERM. PIRITUBA / PQ. DE TAIPAS |
| 34970 | 75 | METRÔ ITAQUERA / JD. REDIL |
| 32791 | 74 | TERM. CAPELINHA / PQ. FERNANDA |
| 33706 | 74 | CONJ. ARAUCÁRIA / METRÔ ARTUR ALVIM |
| 33790 | 74 | VL. DALILA / METRÔ VL. MATILDE |
| 34576 | 74 | SÃO MIGUEL / JD. MABEL |
| 32789 | 73 | TERM. CAPELINHA / JD. JANGADEIRO |
| 32795 | 73 | TERM. CAPELINHA / JD. SÃO BENTO |
| 33638 | 73 | TERM. CACHOEIRINHA / JD. PERY ALTO |
| 33642 | 73 | TERM. CACHOEIRINHA / VL. PENTEADO |
| 33842 | 73 | METRÔ ITAQUERA / GLEBA DO PESSEGO |
| 33847 | 73 | JD. DA CONQUISTA / HOSP. SÃO MATEUS |
| 34811 | 73 | TERM. GUARAPIRANGA / PQ. DO LAGO |
| 33670 | 72 | SÃO MIGUEL / JD. DAS OLIVEIRAS |
| 33678 | 72 | CPTM GUAIANAZES / HOSP. ITAIM |
| 33865 | 72 | JD. STO. ANDRÉ / HOSP. SÃO MATEUS |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 33601 | 70 | TERM. PIRITUBA / COHAB BRASILÂNDIA |
| 33703 | 70 | METRÔ GUILHERMINA/ESPERANÇA / JD. VERONIA |
| 34763 | 70 | JD. ÂNGELA / JD. HORIZONTE AZUL |
| 35113 | 70 | CID. KEMEL / TERM. SÃO MIGUEL |
| 33695 | 69 | VL. RUI BARBOSA / METRÔ VL. MATILDE |
| 34201 | 69 | TERM. PIRITUBA / JD. PAULISTANO |
| 35063 | 69 | JD. PAULISTANO / TERM. PIRITUBA |
| 33692 | 68 | BURGO PAULISTA / METRÔ PATRIARCA |
| 33803 | 68 | JD. SÃO JOÃO / METRÔ ARTUR ALVIM |
| 34766 | 68 | JD. ÂNGELA / JD. VERA CRUZ |
| 34924 | 68 | METRÔ ITAQUERA / COHAB FAZENDA DO CARMO |
| 32787 | 67 | TERM. CAPELINHA / JD. DAS ROSAS |
| 32988 | 67 | CHÁC. BELA VISTA / METRÔ PENHA |
| 33661 | 67 | SÃO MIGUEL / JD. ROMANO |
| 33690 | 67 | VL. UNIÃO / METRÔ PATRIARCA |
| 33866 | 67 | DIV. DE MAUÁ / HOSP. SÃO MATEUS |
| 34666 | 67 | TERM. CAMPO LIMPO / INOCOOP CAMPO LIMPO |
| 32793 | 66 | TERM. CAPELINHA / JD. COMERCIAL |
| 32796 | 66 | TERM. CAPELINHA / JD. VALE DAS VIRTUDES |
| 33694 | 66 | SHOP. METRÔ ITAQUERA / JD. SÃO NICOLAU |
| 34418 | 66 | TERM. SÃO MIGUEL / ITAIM PAULISTA |
| 34652 | 66 | TERM. CAMPO LIMPO / JD. DAS ROSAS |
| 34692 | 66 | TERM. CAMPO LIMPO / PQ. DO ENGENHO |
| 33475 | 65 | TERM. CACHOEIRINHA / JD. STA. CRUZ |
| 33802 | 65 | METRÔ ARTUR ALVIM / CID. LIDER |
| 35138 | 65 | TERM. CID. TIRADENTES / CPTM GUAIANAZES |
| 32794 | 63 | TERM. CAPELINHA / JD. MITSUTANI |
| 34959 | 63 | TERM. SÃO MATEUS / METALÚRGICOS |
| 33781 | 62 | NOVA AMÉRICA / METRÔ ARTUR ALVIM |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 34953 | 62 | TERM. SÃO MATEUS / JD. IGUATEMI |
| 35187 | 62 | PQ. DO LAGO / TERM. GUARAPIRANGA |
| 33783 | 61 | METRÔ BRESSER / UNIV. SÃO JUDAS TADEU |
| 34436 | 60 | TERM. SACOMÃ / VL. ARAPUÁ |
| 35121 | 59 | VL. CISPER (CPTM USP) / TERM. SÃO MIGUEL |
| 33698 | 58 | CANGAÍBA / METRÔ GUILHERMINA/ESPERANÇA |
| 34514 | 58 | TERM. SACOMÃ / VL. ARAPUÁ |
| 34404 | 57 | TERM. SAPOPEMBA / JD. ESTER |
| 34440 | 57 | TERM. SACOMÃ / JD. PATENTE |
| 35132 | 57 | JD. DA CONQUISTA / TERM. SÃO MATEUS |
| 35195 | 57 | JD. IRENE / TERM. CAMPO LIMPO |
| 34438 | 56 | TERM. SACOMÃ / HOSP. HELIÓPOLIS |
| 34828 | 56 | VL. NHOCUNÉ / METRÔ PATRIARCA |
| 35056 | 56 | PQ. SÃO DOMINGOS / TERM. PIRITUBA |
| 35074 | 56 | CONEXÃO PETRÔNIO PORTELA / VL. IARA |
| 35159 | 56 | HOSP. HELIÓPOLIS / TERM. SACOMÃ |
| 33671 | 55 | SÃO MIGUEL / JD. ROBRU |
| 33730 | 55 | CPTM GUAIANAZES / VL. IOLANDA II |
| 35130 | 55 | TERM. SÃO MATEUS / TERM. CID. TIRADENTES |
| 33696 | 53 | METRÔ PATRIARCA / VL. SÍLVIA |
| 34202 | 53 | TERM. PIRITUBA / VL. MIRANTE |
| 35057 | 53 | CID. D'ABRIL 3ª GLEBA / TERM. PIRITUBA |
| 35064 | 53 | VL. MIRANTE / TERM. PIRITUBA |
| 32906 | 52 | TERM. A. E. CARVALHO / CEM. DA SAUDADE |
| 33729 | 52 | CPTM GUAIANAZES / JD. WILMA FLOR |
| 34037 | 52 | TERM. PIRITUBA / VL. ZATT |
| 35214 | 52 | TERM. A. E. CARVALHO / ERMELINO MATARAZZO |
| 35263 | 52 | TERM. PIRITUBA / SOL NASCENTE |
| 32801 | 51 | TERM. CAPELINHA / JD. LÍDIA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 33693 | 51 | METRÔ PATRIARCA / PONTE RASA |
| 34914 | 51 | CPTM JOSÉ BONIFÁCIO / GUAIANAZES |
| 33824 | 50 | CPTM JOSÉ BONIFÁCIO / VL. YOLANDA |
| 33748 | 49 | CPTM GUAIANAZES / CID. TIRADENTES |
| 35220 | 49 | CONEXÃO VL. IÓRIO / VL. IARA |
| 34952 | 48 | TERM. SÃO MATEUS / JD. LIMOEIRO |
| 33672 | 46 | SÃO MIGUEL / JD. CAMPOS |
| 34667 | 46 | TERM. CAMPO LIMPO / JD. MACEDÔNIA |
| 34565 | 45 | CPTM VL. MARA/ITAIM / JD. SÃO MARTINHO |
| 34566 | 45 | CPTM VL. MARA/ITAIM / JD. SÃO MARTINHO |
| 35005 | 44 | TERM. SÃO MATEUS / PQ. BOA ESPERANÇA |
| 34026 | 43 | TERM. PIRITUBA / STA. MÔNICA |
| 33666 | 42 | CPTM GUAIANAZES / JD. ROBRU |
| 33912 | 42 | TERM. VARGINHA / JD. SETE DE SETEMBRO |
| 34653 | 42 | TERM. CAMPO LIMPO / JD. ROSANA |
| 35133 | 42 | JD. STO. ANDRÉ / TERM. SÃO MATEUS |
| 33920 | 41 | TERM. VARGINHA / JD. VARGINHA |
| 33937 | 41 | TERM. VARGINHA / JD. CHÁC. DO SOL |
| 34039 | 41 | TERM. PIRITUBA / VL. MIRANTE |
| 34654 | 41 | TERM. CAMPO LIMPO / JD. HELGA |
| 34912 | 41 | CPTM GUAIANAZES / JD. SÃO PAULO |
| 34954 | 40 | TERM. SÃO MATEUS / JD. STO. ANDRÉ |
| 33752 | 39 | TERM. SÃO MATEUS / JD. RECANTO VERDE SOL |
| 34258 | 39 | JD. ALFREDO / TERM. GUARAPIRANGA |
| 33569 | 38 | PERUS / MORRO DOCE |
| 33664 | 38 | CPTM ITAIM PAULISTA / CID. KEMEL II |
| 33911 | 38 | TERM. VARGINHA / JD. ITAJAÍ |
| 32809 | 37 | TERM. CAMPO LIMPO / JD. MACEDÔNIA |
| 33138 | 37 | CPTM ITAIM PAULISTA / JD. NÉLIA |
| 34025 | 37 | TERM. PIRITUBA / HAB. TURÍSTICA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|--|
| 34655 | 37 | TERM. CAMPO LIMPO / JD. MACEDÔNIA |
| 34719 | 37 | JD. ÂNGELA / JD. SÃO LOURENÇO |
| 34720 | 37 | JD. ÂNGELA / JD. DOS REIS |
| 33665 | 36 | CPTM ITAIM PAULISTA / JD. NSA. SRA. DO CAMINHO |
| 33674 | 36 | CPTM GUAIANAZES / JD. BANDEIRANTES |
| 33917 | 36 | TERM. VARGINHA / JD. NORONHA |
| 34665 | 36 | TERM. CAMPO LIMPO / JD. MITSUTANI |
| 34403 | 35 | JD. SÃO ROBERTO / CONJ. TEOTÔNIO VILELA |
| 34633 | 35 | CPTM GUAIANAZES / JD. FANGANIELO |
| 34958 | 35 | METALÚRGICOS / VL. YOLANDA |
| 35142 | 35 | VL. YOLANDA / TERM. CID. TIRADENTES |
| 33660 | 34 | CPTM GUAIANAZES / JD. NSA. SRA. DO CAMINHO |
| 34956 | 34 | JD. RODOLFO PIRANI / TERM. SÃO MATHEUS |
| 33663 | 33 | CPTM ITAIM PAULISTA / CID. KEMEL I |
| 34490 | 33 | JD. ÂNGELA / VL. GILDA |
| 34957 | 33 | JD. RODOLFO PIRANI / TERM. SÃO MATHEUS |
| 33677 | 29 | CPTM GUAIANAZES / DIV. DE FERRAZ |
| 34946 | 29 | BARRO BRANCO / TERM. CID. TIRADENTES |
| 34947 | 29 | BARRO BRANCO / TERM. CID. TIRADENTES |
| 35137 | 29 | BARRO BRANCO / TERM. CID. TIRADENTES |
| 35222 | 29 | JD. MABEL / JD. ROMANO |
| 35189 | 26 | JD. RIVIERA / TERM. JD. ÂNGELA |
| 35249 | 26 | JD. SÃO ROBERTO / TERM. SAPOPEMBA |
| 35215 | 25 | VL. SOLANGE / CPTM GUAIANAZES |
| 33927 | 22 | TERM. VARGINHA / PQ. FLORESTAL |
| 34356 | 22 | TERM. VARGINHA / JD. SÃO NICOLAU |
| 34948 | 22 | CIRCULAR / JD. NOVA VITÓRIA |

Continua na próxima página

Tabela 30 – continuação da página anterior

| Código da linha | Total de eventos de exceção | Letreiro |
|------------------------|------------------------------------|---|
| 34313 | 21 | TERM. VARGINHA / JD. STA. FÉ |
| 35185 | 21 | JD. HORIZONTE AZUL / TERM. JD. ÂNGELA |
| 32945 | 20 | JD. NOVA ERA / TERM. VARGINHA |
| 33928 | 20 | TERM. VARGINHA / JD. STA. TEREZINHA |
| 35216 | 20 | TERM. CID. TIRADENTES / CID. TIRADENTES |
| 33930 | 19 | TERM. VARGINHA / JD. REC. CAMPO BELO |
| 33939 | 19 | TERM. VARGINHA / MARSILAC |
| 35136 | 19 | METALÚRGICOS / TERM. CID. TIRADENTES |
| 35181 | 19 | TERM. JD. JACIRA / TERM. JD. ÂNGELA |
| 33725 | 18 | VL. PAULISTA I / TERM. CID. TIRADENTES |
| 35018 | 18 | CIRCULAR / TERM. CID. TIRADENTES |
| 35182 | 18 | VL. GILDA / TERM. JD. ÂNGELA |
| 35186 | 18 | PQ. DO LAGO / TERM. JD. ÂNGELA |
| 33722 | 15 | METALÚRGICOS / TERM. CID. TIRADENTES |
| 35254 | 12 | JD. MONTE BELO / TERM. JD. BRITANIA |
| 34644 | 10 | SETOR IIB / TERM. CID. TIRADENTES |
| 34998 | 7 | CHÁC. MARIA TRINDADE / TERM. JD. BRITÂNIA |
| 35029 | 6 | CACHOEIRA / DIB |
| 33941 | 2 | TERM. PARELHEIROS / BARRAGEM |
| 35223 | 2 | PARELHEIROS / CHÁC. BOSQUE DO SOL |
| 33946 | 1 | TERM. PARELHEIROS / JD. EUCA LIPTOS |
| 33948 | 1 | TERM. PARELHEIROS / JD. ORIENTAL/FONTES |
| 34333 | 1 | TERM. PARELHEIROS / CIPÓ DO MEIO |

Apêndice E – Matrizes de confusão

Figura 22 – Matriz de confusão relacionada a classificação dos *tweets* em eventos de exceção por meio do algoritmo Árvore de Decisão

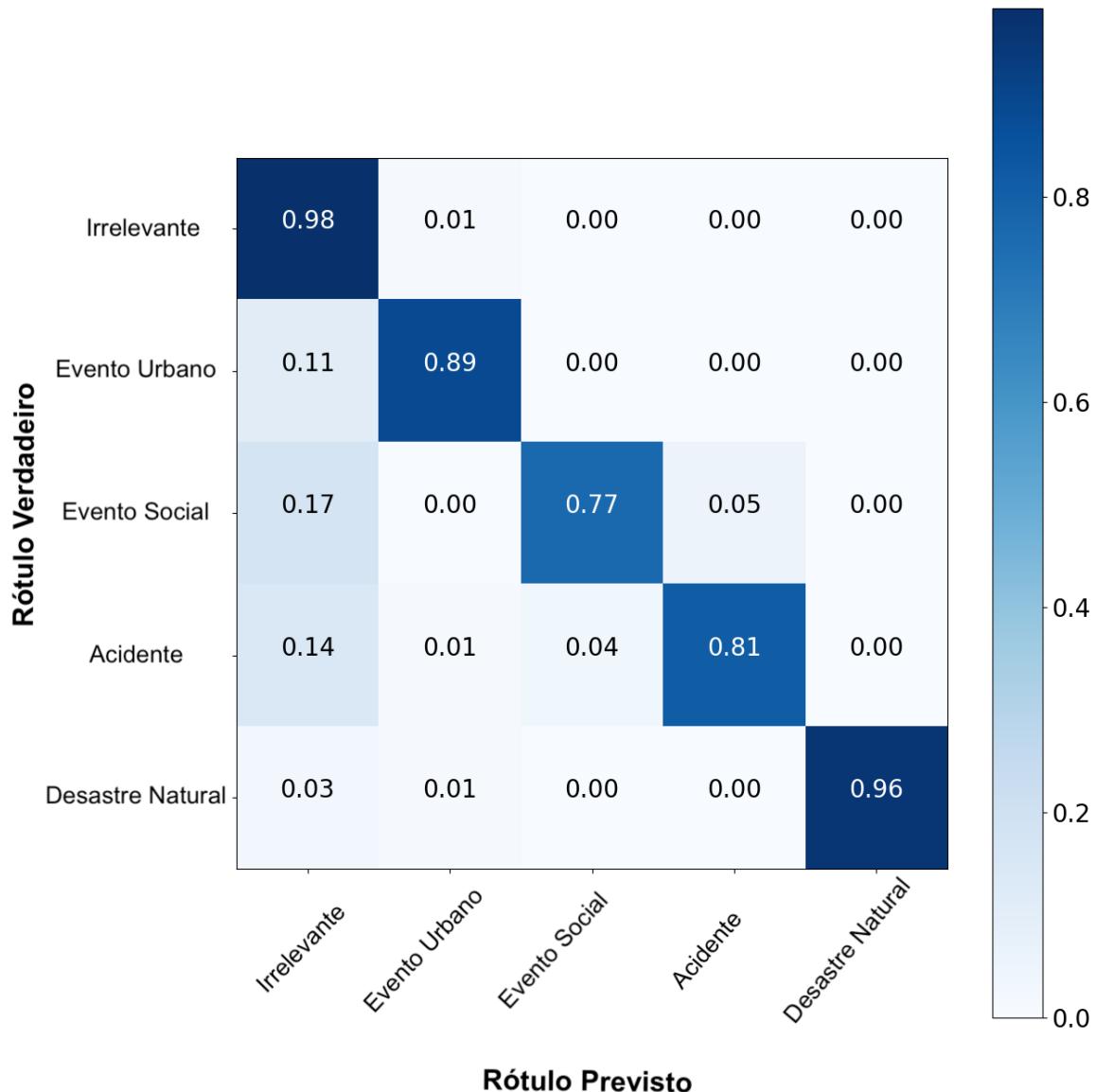


Figura 23 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo *Complement Naive Bayes*

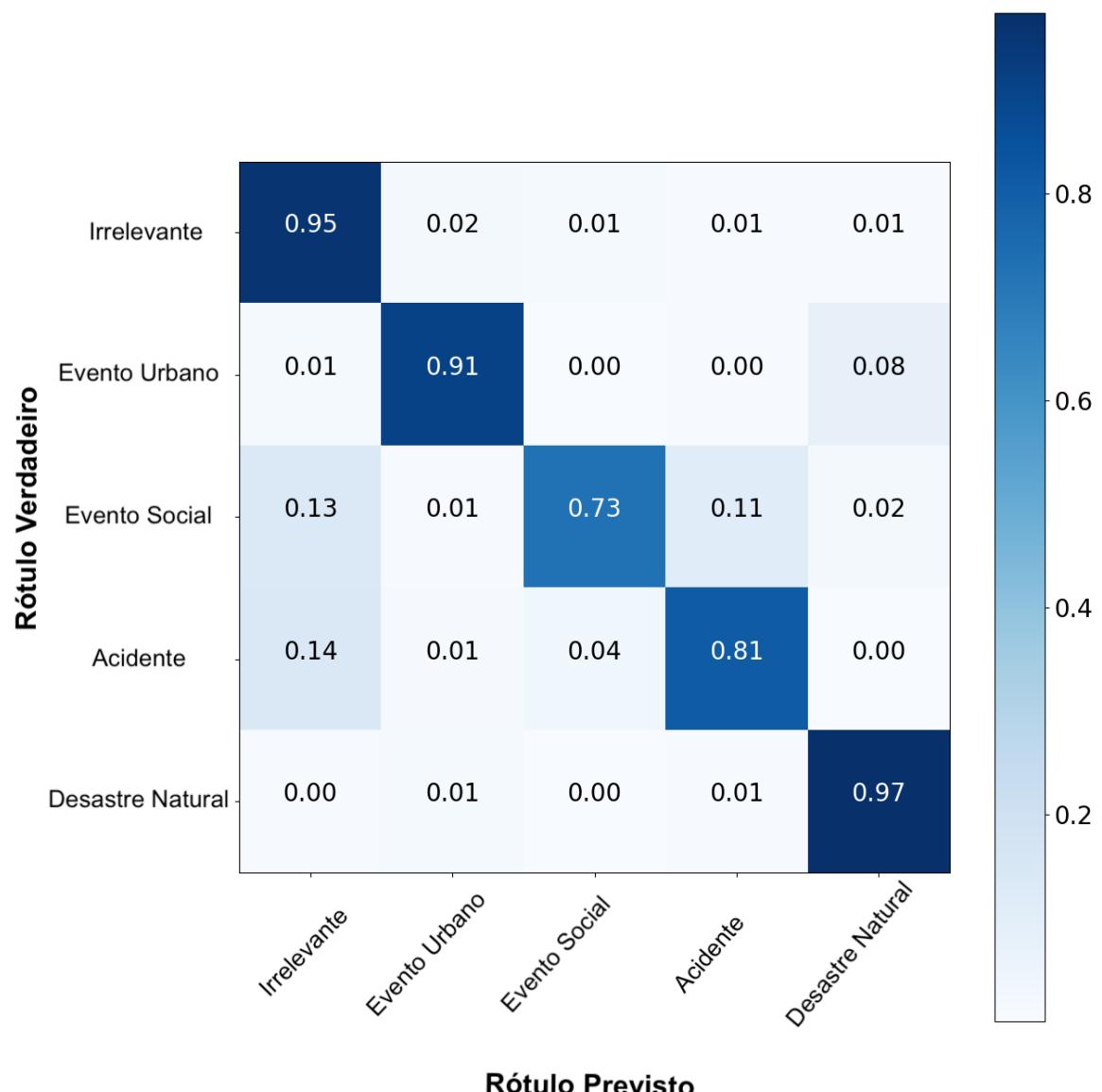


Figura 24 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo Florestas Aleatórias

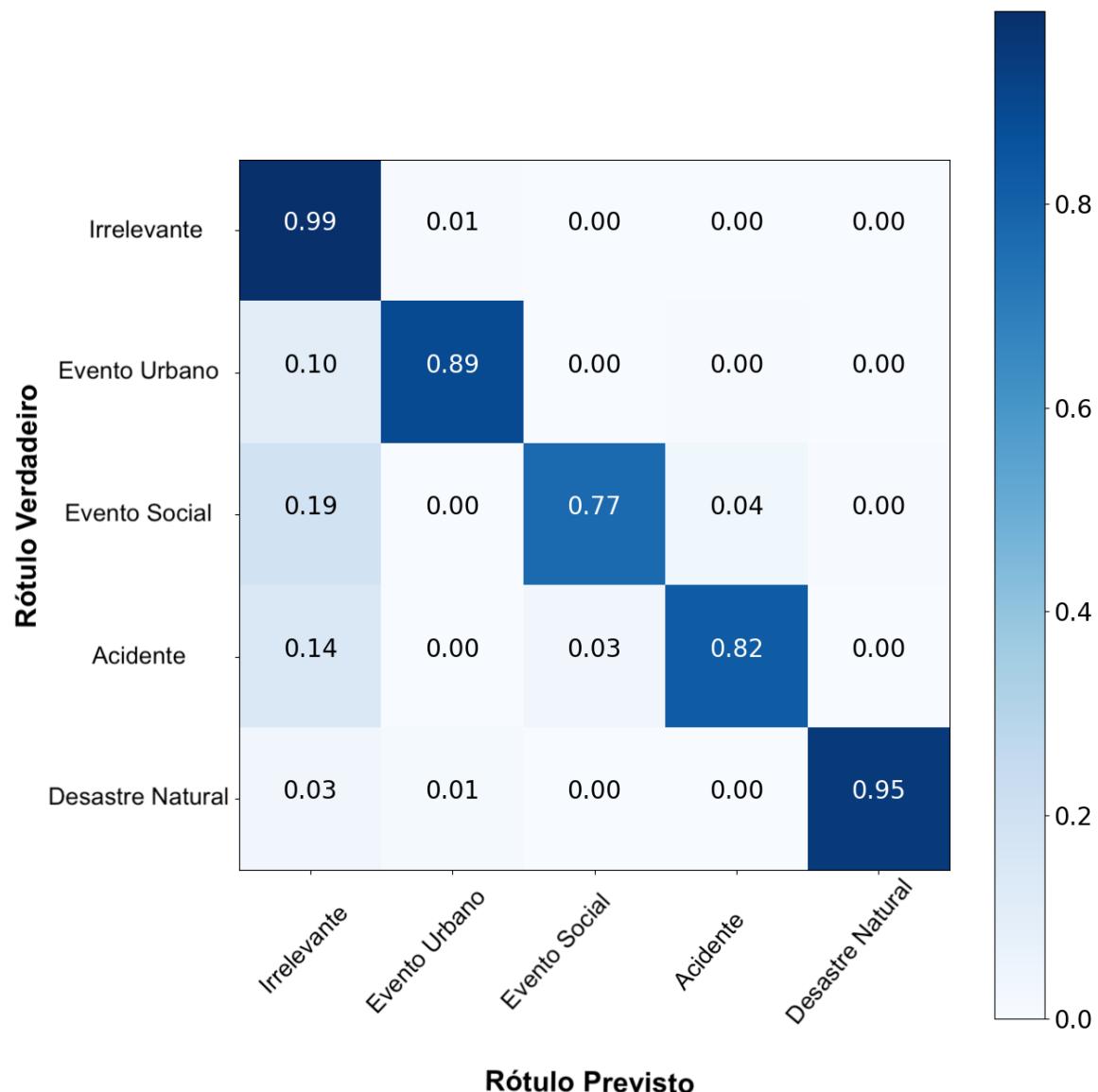


Figura 25 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo *Multinomial Naive Bayes*

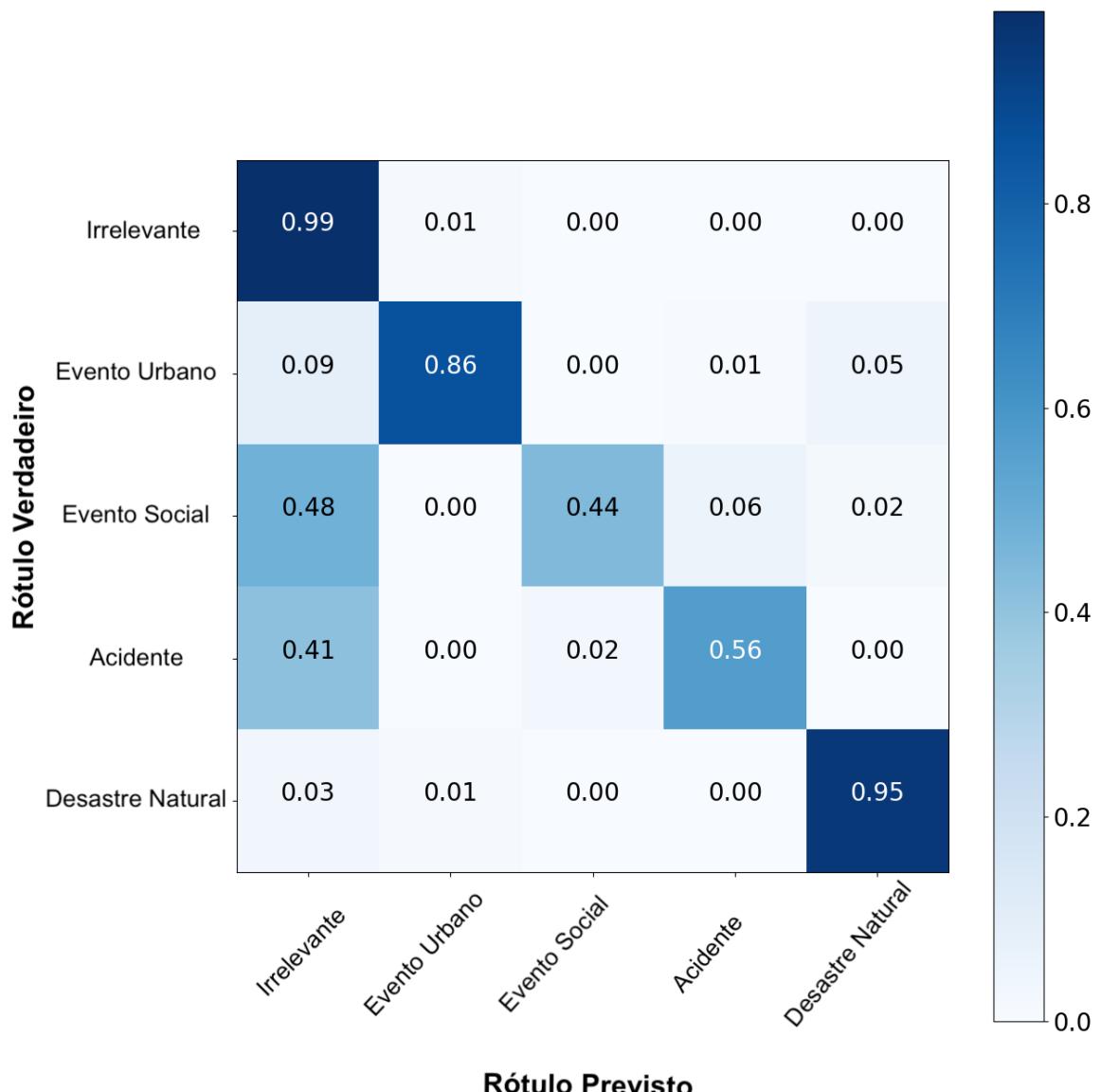


Figura 26 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo Regressão Logística

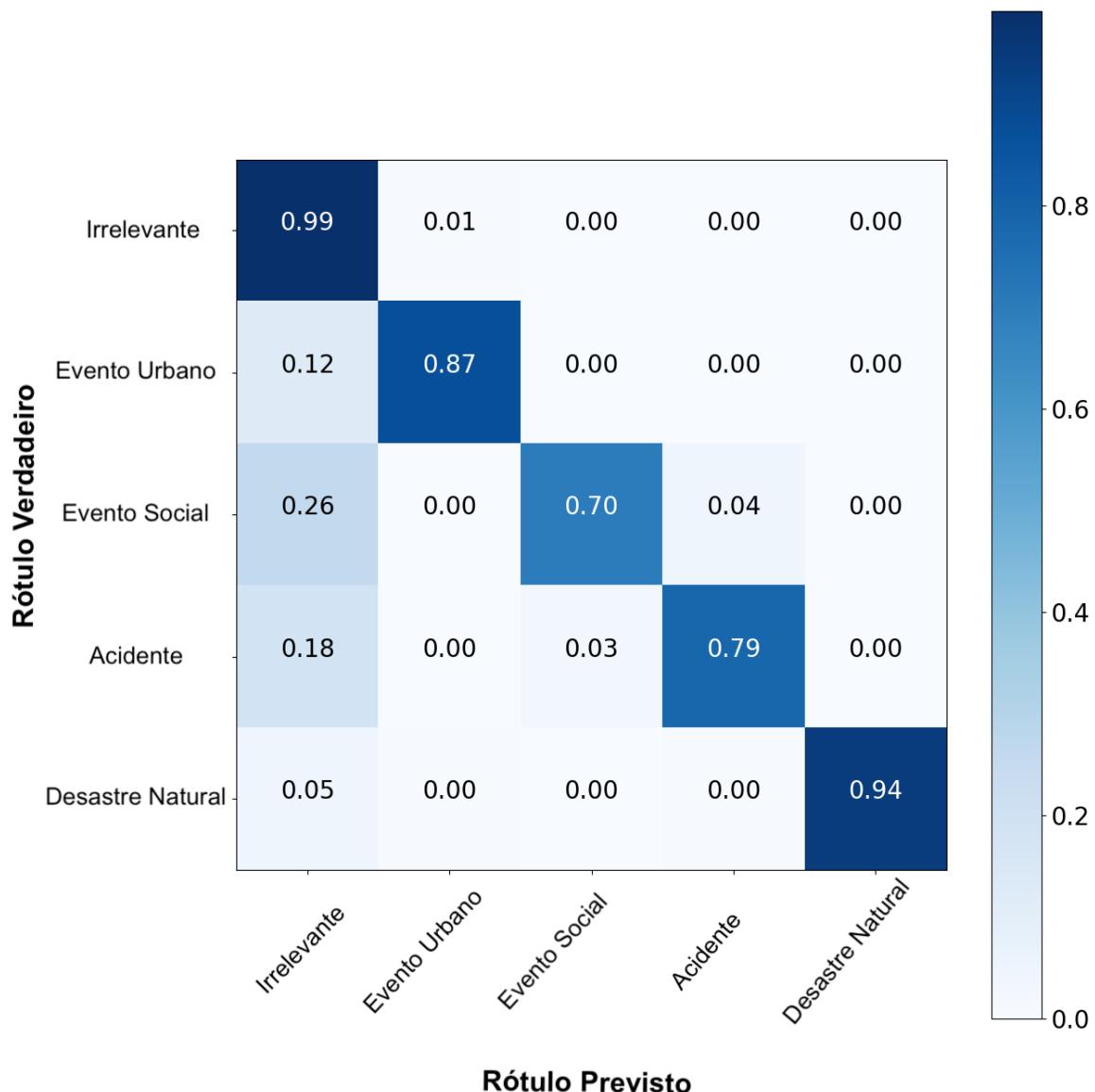
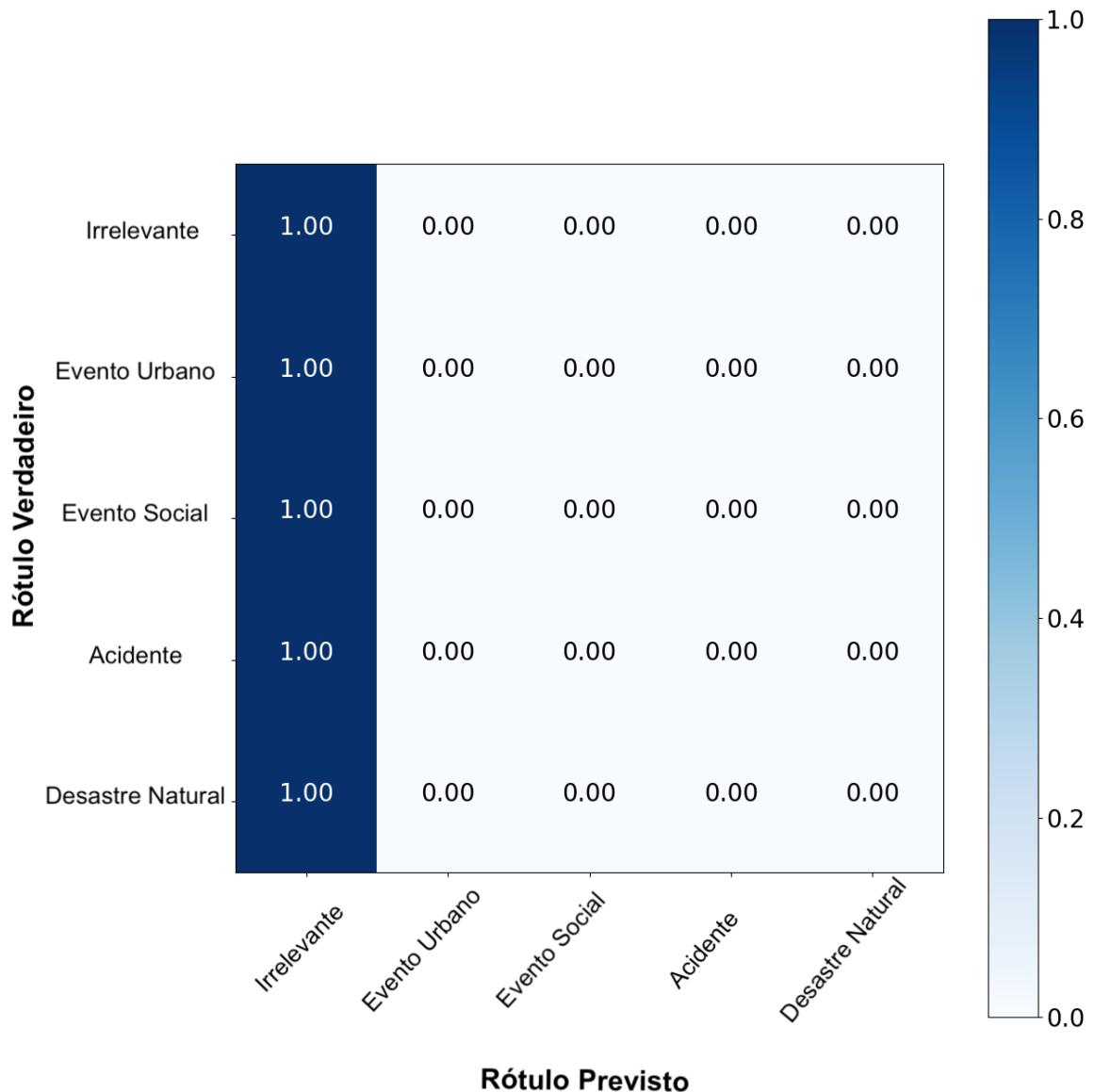


Figura 27 – Matriz de confusão relacionada a classificação dos tweets em eventos de exceção por meio do algoritmo Máquina de Vetores de Suporte



Apêndice F – Parametrizações dos algoritmos

F.1 Árvore de Decisão

1

- *criterion* — *string*, opcional (*default* = “*gini*”) — Parâmetro responsável por definir a função que mede a qualidade da divisão da árvore de decisão. Os valores suportados são *gini* para a *impureza Gini* e *entropy* para o *ganho de informação*.
- *splitter* — *string*, opcional (*default* = “*best*”) — Parâmetro responsável por definir a estratégia usada para escolher a divisão em cada nó. As estratégias suportadas são *best* para escolher a melhor divisão e *random* para escolher a melhor divisão aleatoriamente.
- *max_depth* — *int* ou *None*, opcional (*default* = *None*) — Parâmetro responsável por definir a profundidade máxima da árvore. Se definido como *None*, os nós são expandidos até que todas as folhas fiquem puras ou até que todas as folhas contenham menos amostras que *min_samples_split*.
- *min_samples_split* — *int*, *float*, opcional (*default* = 2) — Parâmetro responsável por definir o número mínimo de amostras necessárias para dividir um nó interno.
- *min_samples_leaf* — *int*, *float*, opcional (*default* = 1) — Parâmetro responsável por definir o número mínimo de amostras necessárias em um nó folha. Um ponto de divisão em qualquer profundidade só será considerado se deixar pelo menos *min_samples_leaf* amostras de treinamento em cada uma das ramificações esquerda e direita. Isso pode ter o efeito de suavizar o modelo, especialmente na regressão.
- *min_weight_fraction_leaf* — *float*, opcional (*default* = 0.) — Parâmetro responsável por definir a fração ponderada mínima da soma total de pesos (de todas as amostras de entrada) necessária para estar em um nó folha. As amostras têm peso igual quando *sample_weight* não é fornecido.

¹ Descrições das parametrização adaptadas com base em:<<http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>>. Acessado em 08 de outubro de 2018.

- *max_features* — *int, float, string ou None*, opcional (*default = None*) — Parâmetro responsável por definir o número de *features* (características) a serem consideradas ao procurar a melhor divisão. A procura por uma divisão não é interrompida até que pelo menos uma partição válida das amostras de nó seja localizada, mesmo que seja necessário inspecionar do que mais de *max_features* características.
- *random_state* — *int, RandomStateinstance ou None*, opcional (*default = None*) — Parâmetro responsável por determinar a estratégia de geração de número aleatórios. Se definido como *RandomState*, *random_state* será o gerador de números aleatórios; se *None* o gerador de números aleatórios é a instância *RandomState* usada por *np.random*.
- *max_leaf_nodes* — *int ou None*, opcional (*default = None*) — Parâmetro responsável por gerar uma árvore com o máximo número de nós folhas, usando a estratégia *best-first*. Os melhores (*best*) nós são os definidos como redução relativa a impureza. Caso o parâmetro seja definido como *None* então o número máximo de nós folhas será ilimitado.
- *min_impurity_decrease* — *float*, opcional (*default = 0.*) — Parâmetro responsável por definir que um nó será dividido se essa divisão induzir uma diminuição da impureza maior ou igual a esse valor.
- *class_weight* — *dict, list de dict, “balanced”, None, default = None* — Parâmetro responsável por associar ponderação as classes, no seguinte formato: *“class_label : weight”*. Caso não haja valores para esse parâmetro, supõem-se que todos as classes possuam o mesmo peso.
- *presort* — *bool*, opcional (*default = False*) — Se o valor desse parâmetro é igual a *true* é realizada uma pré-ordenação dos dados, o que acelera encontrar as melhores divisões das árvores de decisão no processo de ajuste. Ao habilitar esse parâmetro, a velocidade do processo de treinamento de um grande volume de dados é reduzida. Por outro lado, habilitar esse parâmetro em alguns casos pode acelerar o processo de treinamento, como quando há pequenos conjuntos de dados, ou, restrição quanto a profundidade da árvore de decisão.

F.2 Floresta Aleatória

2

- *n_estimators* — *integer*, opcional (*default* = 100) — Parâmetro responsável pelo número de árvores na floresta.
- *criterion* — *string*, opcional (*default* = “*gini*”) — Parâmetro responsável por definir a função que mede a qualidade da divisão da árvore de decisão. Os valores suportados são *gini* para a *impureza Gini* e *entropy* para o *ganho de informação*.
- *max_depth* — *int* ou *None*, opcional (*default* = *None*) — Parâmetro responsável por definir a profundidade máxima da árvore. Se definido como *None*, os nós são expandidos até que todas as folhas fiquem puras ou até que todas as folhas contenham menos amostras que *min_samples_split*.
- *min_samples_split* — *int*, *float*, opcional (*default* = 2) — Parâmetro responsável por definir o número mínimo de amostras necessárias para dividir um nó interno.
- *min_samples_leaf* — *int*, *float*, opcional (*default* = 1) — Parâmetro responsável por definir o número mínimo de amostras necessárias em um nó folha. Um ponto de divisão em qualquer profundidade só será considerado se deixar pelo menos *min_samples_leaf* amostras de treinamento em cada uma das ramificações esquerda e direita. Isso pode ter o efeito de suavizar o modelo, especialmente na regressão.
- *min_weight_fraction_leaf* — *float*, opcional (*default* = 0.) — Parâmetro responsável por definir a fração ponderada mínima da soma total de pesos (de todas as amostras de entrada) necessária para estar em um nó folha. As amostras têm peso igual quando *sample_weight* não é fornecido.
- *max_features* — *int*, *float*, *string* ou *None*, opcional (*default* = *None*) — Parâmetro responsável por definir o número de *features* (características) a serem consideradas ao procurar a melhor divisão. A procura por uma divisão não é interrompida até que pelo menos uma partição válida das amostras de

² Descrições das parametrização adaptadas com base em:<<http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>>. Acessado em 08 de outubro de 2018.

nó seja localizada, mesmo que seja necessário inspecionar do que mais de `max_features` características.

- `random_state` — `int`, `RandomState` instance ou `None`, opcional (`default = None`) — Parâmetro responsável por determinar a estratégia de geração de número aleatórios. Se definido como `RandomState`, `random_state` será o gerador de números aleatórios; se `None` o gerador de números aleatórios é a instância `RandomState` usada por `np.random`.
- `max_leaf_nodes` — `int` ou `None`, opcional (`default = None`) — Parâmetro responsável por gerar uma árvore com o máximo número de nós folhas, usando a estratégia `best-first`. Os melhores (`best`) nós são os definidos como redução relativa a impureza. Caso o parâmetro seja definido como `None` então o número máximo de nós folhas será ilimitado.
- `min_impurity_decrease` — `float`, opcional (`default = 0.`) — Parâmetro responsável por definir que um nó será dividido se essa divisão induzir uma diminuição da impureza maior ou igual a esse valor.
- `bootstrap` — `boolean`, opcional (`default = True`) — Parâmetro responsável por definir se amostras de `bootstrap` serão usadas ao construir árvores.
- `oob_score` — `boolean`, opcional (`default = False`) — Parâmetro responsável por definir o uso de amostras `out-of-bag` para estimar a precisão da generalização.
- `n_jobs` — `int` ou `None`, opcional (`default = None`) — Parâmetro responsável por definir o número de `jobs` a serem executados em paralelo durante os processos de `fit` e `predict`. `None` define 1 `job` a menos que esteja em um contexto `joblib.parallel_backend`; -1 define que todos os processadores sejam usados.
- `verbose` — `int`, opcional (`default = 0`) — Parâmetro responsável por controlar a verbosidade durante os processos de `fit` e `predict`.
- `warm_start` — `bool`, opcional (`default = False`) — Parâmetro que quando definido como `True` reutiliza a solução da chamada anterior no processo de `fit` e adiciona mais estimadores ao `ensemble`, caso contrário, apenas aplica o processo de `fit` a toda uma nova floresta.
- `class_weight` — `dict`, `list` de `dict`, “`balanced`”, `None`, `default = None` — Parâmetro responsável por associar ponderação as classes, no seguinte formato:

"class_label : weight". Caso não haja valores para esse parâmetro, supõem-se que todos as classes possuam o mesmo peso.

- *presort* — *bool*, opcional (*default = False*) — Se o valor desse parâmetro é igual a *true* é realizada uma pré-ordenação dos dados, o que acelera encontrar as melhores divisões das árvores de decisão no processo de ajuste. Ao habilitar esse parâmetro, a velocidade do processo de treinamento de um grande volume de dados é reduzida. Por outro lado, habilitar esse parâmetro em alguns casos pode acelerar o processo de treinamento, como quando há pequenos conjuntos de dados, ou, restrição quanto a profundidade da árvore de decisão.

F.3 K-ésimo Vizinho mais Próximo

3

- *n_neighbors* — *int*, opcional (*default = 5*) — Parâmetro responsável por definir o número padrão de *neighbors* usados pelas *kneighbors queries*.
- *weights* — *str* ou *callable*, opcional (*default = 'uniform'*) — Parâmetro usado para definir a função de peso usada no processo *predict*. Valores possíveis: (I) *uniform*: pesos uniformes; todos os pontos em cada vizinha são ponderados igualmente; (II) *distance*: pontos de ponderação pelo inverso da suas respectivas distâncias; nesse caso, os vizinhos mais próximos de um ponto de consulta terão uma influência maior do que os vizinhos mais distantes; (III) *callable*: uma função definida pelo usuário que aceita uma matriz de distâncias e retorna uma matriz da mesma forma, contendo contém os pesos.
- *algorithm* — *auto*, *ball_tree* (*BallTree*), *kd_tree* (*KDTree*), *brute* (pesquisa por força bruta), opcional (*default = 'auto'*) — Parâmetro responsável por definir algoritmo utilizado para calcular os vizinhos mais próximos. O valor padrão tentará decidir o algoritmo mais apropriado com base nos valores passados para o método *fit*. Em caso de dados esparsos no processo de ajuste esse parâmetro é ignorado e usado a opção *brute* por padrão.
- *leaf_size* — *int*, opcional (*default = 30*) — Parâmetro responsável por definir o tamanho da folha passado para o *BallTree* ou *KDTree*. Isso pode afetar a

³ Descrições das parametrização adaptadas com base em: <<http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>>. Acessado em 08 de outubro de 2018.

velocidade da construção e consulta, bem como a memória necessária para armazenar a árvore. O valor ideal depende da natureza do problema.

- p — *integer*, opcional ($default = 2$) — Parâmetro de potência para a métrica *Minkowski*. Quando $p = 1$, isso equivale a usar *manhattan_distance* ($l1$) e *euclidean_distance* ($l2$) para $p = 2$. Para p arbitrário, *minkowski_distance* (l_p) é usado.
- *metric* — *string* ou *callable*, opcional ($default = 'minkowski'$) — Parâmetro responsável por definir a distância métrica para usar na árvore. A métrica padrão é *minkowski* e com $p = 2$ é equivalente à métrica euclidiana padrão.
- *n_jobs* — *int* ou *None*, opcional ($default = None$) — Parâmetro responsável por definir o número de *jobs* a serem executados em paralelo durante os processos de *fit* e *predict*. *None* define 1 *job* a menos que esteja em um contexto *joblib.parallel_backend*; -1 define que todos os processadores sejam usados.

F.4 Máquina de Vetores de Suporte

4

- C — *float*, opcional ($default = 1.0$) — Parâmetro de *penalidade C* do termo de erro.
- *kernel* — *string*, opcional ($default = 'rbf'$) — Parâmetro responsável por especificar o tipo de *kernel* a ser usado no algoritmo. Pode ser *linear*, *poly*, *rbf*, *sigmoid*, *precomputed* ou *callable*.
- *degree* — *int*, opcional ($default = 3$) — Parâmetro responsável por definir a *polynomial kernel function (poly)*. Ignorado por todos os outros *kernels*.
- *gamma* — *float*, opcional ($default = 'auto'$) — Parâmetro responsável por definir o coeficiente de *Kernel* para *rbf*, *poly* e *sigmoid*.
- *coef0* — *float*, opcional ($default = 0.0$) — Parâmetro responsável por definir o termo independente na função *kernel*. É significativo apenas para *poly* e *sigmoid*.

⁴ Descrições das parametrização adaptadas com base em:<<http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC>>. Acessado em 08 de outubro de 2018.

- *shrinking* — *boolean*, opcional (*default = True*) — Parâmetro responsável por definir o uso da heurística *shrinking*.
- *probability* — *boolean*, opcional (*default = False*) — Parâmetro responsável por definir o uso de estimativas de probabilidade, o qual deve ser ativado antes do processo de *fit* (implica em perda de desempenho).
- *tol* — *float*, opcional (*default = 1e - 3*) — Parâmetro responsável por definir a tolerância ao critério de parada.
- *cache_size* — *float*, opcional — Parâmetro responsável por definir o tamanho do cache do *kernel* (em MB).
- *class_weight* — *dict, balanced, optional* (*default = None*) — Parâmetro responsável por definir o parâmetro C da classe *i* para *class_weight[i] * C* para o SVC.
- *verbose* — *bool*, (*default = False*) — Parâmetro responsável por habilitar a saída detalhada.
- *max_iter* — *int*, opcional (*default = -1*) — Parâmetro responsável por definir um limite rígido em iterações no *solver*, ou -1 para sem limite.
- *decision_function_shape* — *ovo, ovr, (default =' ovr')* — Parâmetro responsável por definir se deve retornar uma função de decisão *one-vs-rest* (*ovr*) ou a função de decisão original *one-vs-one*.
- *random_state* — *int, RandomState instance ou None, opcional (default = None)* — Parâmetro responsável por determinar a estratégia de geração de número aleatórios. Se definido como *RandomState*, *random_state* será o gerador de números aleatórios; se *None* o gerador de números aleatórios é a instância *RandomState* usada por *np.random*.

F.5 Naive Bayes

5

- *alpha* — *float*, opcional (*default = 1.0*) — Parâmetro de suavização (0 para não suavização) aditivo (Laplace / Lidstone).

⁵ Descrições das parametrização adaptadas com base em:<http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html#sklearn.naive_bayes.MultinomialNB> e <http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.ComplementNB.html#sklearn.naive_bayes.ComplementNB>. Acessado em 08 de outubro de 2018.

- *fit_prior* — *boolean*, opcional (*default = True*) — Parâmetro responsável por definir ou não o aprendizado das probabilidades anteriores da classe.
- *class_prior* — *array-like, size (n_classes,)*, opcional (*default = None*) — Parâmetro responsável por definir probabilidades anteriores das classes. Se especificado, os antecedentes não são ajustados de acordo com os dados.
- *norm* — *boolean*, opcional (*default = False*) — Parâmetro responsável por definir se uma segunda normalização dos pesos é executada ou não. Disponível somente na implementação *ComplementNB*.

F.6 Redes Neurais

6

- *hidden_layer_sizes* — *tuple, length = n_layers - 2*, opcional(*default = (100,)*) — Parâmetro responsável por definir o *ith* elemento que representa o número de neurônios na *ith* camada oculta.
- *activation* — *identity, logistic, tanh, relu*, opcional (*default = relu*) — Parâmetro responsável por definir a função de ativação para a camada oculta.
- *solver* — *lbfgs, sgd, adam*, opcional (*default = adam*) — Parâmetro responsável por definir o solucionador para otimização de peso.
- *alpha* — *float*, opcional (*default = 0.0001*) — Parâmetro de penalidade L2 (termo de regularização).
- *batch_size* — *int*, opcional (*default = auto*) — Parâmetro responsável pelo tamanho de *mini-batches* para otimizadores estocásticos. Se o solucionador for *lbfgs*, o classificador não usa *minibatch*. Quando definido como *auto*, *batch_size = min(200, n_samples)*.
- *learning_rate* — *constant, invscaling, adaptive*,opcional (*default = constant*) — Parâmetro responsável pela programação da taxa de aprendizado para atualizações de ponderações.

⁶ Descrições das parametrização adaptadas com base em:<http://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html#sklearn.neural_network.MLPClassifier>. Acessado em 08 de outubro de 2018.

- *learning_rate_init* — *double*, opcional (*default* = 0.001) — Parâmetro responsável por definir a taxa inicial de aprendizado utilizada, somente quando *solver* = *sgd* ou *adam*.
- *power_t* — *double*, opcional (*default* = 0.5) — Parâmetro responsável por definir o expoente para a taxa de aprendizado de escala inversa, quando a *learning_rate* é definida como *invscaling* e *solver* = *sgd*.
- *max_iter* — *int*, opcional (*default* = 200) — Parâmetro responsável por definir o número máximo de iterações. O *solver* itera até a convergência (determinada por *tol*) ou pelo *max_iter*. Para solvers estocásticos (*sgd*, *adam*), esse parâmetro determina o número de *epochs* (quantas vezes cada ponto de dados será usado), não o número de etapas do gradiente.
- *shuffle* — *bool*, opcional (*default* = *True*) — Parâmetro responsável por definir o embaralhamento das amostras em cada iteração. Usado somente quando *solver* = *sgd* ou *adam*.
- *random_state* — *int*, *RandomState* instance ou *None*, opcional (*default* = *None*) — Parâmetro responsável por determinar a estratégia de geração de número aleatórios. Se definido como *RandomState*, *random_state* será o gerador de números aleatórios; se *None* o gerador de números aleatórios é a instância *RandomState* usada por *np.random*.
- *verbose* — *bool*, (*default* = *False*) — Parâmetro responsável por habilitar a saída detalhada.
- *tol* — *float*, opcional, (*default* = $1e-4$) — Parâmetro responsável por definir a tolerância para a otimização.
- *warm_start* — *bool*, opcional (*default* = *False*) — Parâmetro responsável por definir a reutilização da solução da chamada anterior para o processo de *fit* como inicialização, caso contrário, a solução anterior é apagada.
- *momentum* — *float*, opcional (*default* = 0.9) — Parâmetro responsável por definir o *momentum* para a atualização de descida de gradiente. Deve estar entre 0 e 1. Apenas utilizado quando *solver* = *sgd*.
- *nesterovs_momentum* — *boolean*, (*default* = *True*) — Parâmetro responsável por definir o uso do *Nesterov's momentum*. Apenas utilizando quando *solver* = *sgd* e *momentum* > 0.

- *early_stopping* — *bool*, opcional (*default = False*) — Parâmetro responsável por definir parada antecipada para finalizar o treinamento quando a pontuação de validação não estiver melhorando. Se definido como verdadeiro, automaticamente 10% dos dados de treinamento são usados como validação, encerrando o treinamento quando a pontuação de validação não estiver melhorando em pelo menos *tol* para *n_iter_no_change epochs* consecutivos. Esse parâmetro somente é efetivo quando *solver = sgd* ou *adam*.
- *validation_fraction* — *float*, opcional, (*default = 0.1*) — Parâmetro responsável por definir a proporção de dados de treinamento a serem definidos como um conjunto de validação para interrupção antecipada. O valor deve estar entre 0 e 1. Apenas usado se *early_stopping = True*.
- *beta_1* — *float*, opcional (*default = 0.9*) — Parâmetro responsável por definir a taxa de decaimento exponencial (entre 0 e 1) para estimativas do primeiro momento vetorial em *adam*. Usado somente quando *solver = adam*.
- *beta_2* — *float*, opcional (*default = 0.9*) — Parâmetro responsável por definir a taxa de decaimento exponencial (entre 0 e 1) para estimativas do segundo momento vetorial em *adam*. Usado somente quando *solver = adam*.
- *epsilon* — *float*, opcional (*default = 1e-8*) — Parâmetro responsável por definir o valor para estabilidade numérica em *adam*. Usado somente quando *solver = adam*.
- *n_iter_no_change* — *int*, opcional (*default = 10*) — Parâmetro responsável por definir o número máximo de *epochs* para não atender a melhoria definida pelo parâmetro *tol*. Usado somente quando *solver = adam*.

F.7 Regressão Logística

7

- *penalty* — *str*, *l1'* ou *l2*, opcional (*default = l2*) — Usado para especificar a norma usada na penalização. Os solucionadores "newton-cg", "sag" e "lbfgs" apoiam apenas as penalidades *l2*.

⁷ Descrições das parametrização adaptadas com base em: <http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html>. Acessado em 08 de outubro de 2018.

- *dual* — *bool*, opcional (*default = False*) — Parâmetro responsável por definir formulação *dual* ou *primal*. Formulação *dual* é apenas implementada para penalidade *l2* com o *liblinear solver*. Preferível *dual = False* quando *n_samples > n_features*.
- *tol* — *float*, opcional (*default = 1e-4*) — Parâmetro responsável por definir a tolerância para o critério de parada.
- *C* — *float*, opcional (*default = 1.0*) — Parâmetro responsável por definir a inversão da força de regularização; deve ser um *float* positivo. Como nas máquinas de vetores de suporte, valores menores especificam uma regularização mais forte.
- *fit_intercept* — *bool*, opcional (*default = True*) — Parâmetro responsável por definir se uma constante (viés ou interceptação) deve ser adicionada a função de decisão.
- *intercept_scaling* — *float*, opcional (*default = 1*) — Parâmetro responsável por definir a escala de interceptação. Útil somente quando *solver = liblinear* e *self.fit_intercept = True*.
- *class_weight* — *dict*, *list* de *dict*, “balanced”, *None*, *default = None* — Parâmetro responsável por associar ponderação as classes, no seguinte formato: “*class_label : weight*”. Caso não haja valores para esse parâmetro, supõem-se que todos as classes possuam o mesmo peso.
- *random_state* — *int*, *RandomState instance* ou *None*, opcional (*default = None*) — Parâmetro responsável por determinar a estratégia de geração de número aleatórios. Se definido como *RandomState*, *random_state* será o gerador de números aleatórios; se *None* o gerador de números aleatórios é a instância *RandomState* usada por *np.random*.
- *solver* — *str*, *newton-cg*, *lbfgs*, *liblinear*, *sag*, *saga*, opcional (*default = liblinear*) — Parâmetro responsável por definir o algoritmo utilizado no problema de otimização.
- *max_iter* — *int*, opcional (*default = 100*) — Parâmetro utilizado com *solver = newton-cg*, *sag* e *lbfgs*. Número máximo de iterações tomadas para os solvers convergirem.
- *verbose* — *bool*, (*default = False*) — Parâmetro responsável por habilitar a saída detalhada.

- *multi_class* — *str, ovr, multinomial, auto*, opcional (*default = ovr*) — Parâmetro responsável por definir multi classes.
- *warm_start* — *bool*, opcional (*default = False*) — Parâmetro que quando definido como *True*, reutiliza a solução da chamada anterior para o processo de *fit* como inicialização, caso contrário, a solução anterior é apagada. Sem efeitos quando *solver = liblinear*.
- *n_jobs* — *int* ou *None*, opcional (*default = None*) — Parâmetro responsável por definir a quantidade de núcleos de CPU utilizados na paralelização sob as classes, quando *multi_class = ovr*. Esse parâmetro é ignorado quando *solver = liblinear*, independentemente de *multi_class* estar especificado ou não. *None* define 1 núcleo a menos que esteja em um contexto *joblib.parallel_backend*; -1 define o uso de todos os processadores.

Apêndice G - Análise Apriori

G.1 Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans, referentes aos meses do ano de 2017.

Tabela 31 – Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Janeiro

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,14 | 0,14 | 1 |
| 11 | 0,28 | 0,28 | 1 |
| 12 | 0,27 | 0,27 | 1 |
| 13 | 0,15 | 0,15 | 1 |
| 2 | 0,12 | 0,12 | 1 |
| 4 | 0,10 | 0,10 | 1 |
| 5 | 0,11 | 0,11 | 1 |
| 6 | 0,12 | 0,12 | 1 |
| 7 | 0,19 | 0,19 | 1 |
| 8 | 0,15 | 0,15 | 1 |
| 9 | 0,13 | 0,13 | 1 |
| 11 → 12 | 0,13 | 0,47 | 1,72 |

Tabela 32 – Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Fevereiro

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,17 | 0,17 | 1 |
| 11 | 0,21 | 0,21 | 1 |
| 12 | 0,32 | 0,32 | 1 |
| 13 | 0,19 | 0,19 | 1 |
| 14 | 0,10 | 0,10 | 1 |
| 2 | 0,12 | 0,12 | 1 |
| 5 | 0,12 | 0,12 | 1 |
| 6 | 0,10 | 0,10 | 1 |
| 7 | 0,20 | 0,20 | 1 |
| 8 | 0,13 | 0,13 | 1 |
| 9 | 0,18 | 0,18 | 1 |
| 12 → 11 | 0,12 | 0,58 | 1,77 |
| 12 → 13 | 0,12 | 0,37 | 1,95 |
| 7 → 8 | 0,10 | 0,49 | 3,58 |

Tabela 33 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Março

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,15 | 0,15 | 1 |
| 11 | 0,21 | 0,21 | 1 |
| 12 | 0,38 | 0,38 | 1 |
| 13 | 0,23 | 0,23 | 1 |
| 14 | 0,13 | 0,13 | 1 |
| 2 | 0,12 | 0,12 | 1 |
| 5 | 0,12 | 0,12 | 1 |
| 6 | 0,10 | 0,10 | 1 |
| 7 | 0,17 | 0,17 | 1 |
| 9 | 0,16 | 0,16 | 1 |
| 12 → 11 | 0,13 | 0,62 | 1,62 |
| 12 → 13 | 0,15 | 0,41 | 1,76 |

Tabela 34 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Abril

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,125 | 0,12 | 1 |
| 11 | 0,20 | 0,20 | 1 |
| 12 | 0,29 | 0,29 | 1 |
| 13 | 0,16 | 0,16 | 1 |
| 2 | 0,12 | 0,12 | 1 |
| 4 | 0,12 | 0,12 | 1 |
| 5 | 0,11 | 0,11 | 1 |
| 6 | 0,10 | 0,10 | 1 |
| 7 | 0,23 | 0,23 | 1 |
| 8 | 0,14 | 0,14 | 1 |
| 9 | 0,16 | 0,16 | 1 |
| 12 → 11 | 0,12 | 0,60 | 2,01 |
| 13 → 12 | 0,10 | 0,36 | 2,28 |
| 7 → 8 | 0,10 | 0,45 | 3,18 |

Tabela 35 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Maio

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,11 | 0,11 | 1 |
| 11 | 0,21 | 0,21 | 1 |
| 12 | 0,37 | 0,37 | 1 |
| 13 | 0,21 | 0,21 | 1 |
| 14 | 0,12 | 0,12 | 1 |
| 2 | 0,12 | 0,12 | 1 |
| 5 | 0,12 | 0,12 | 1 |
| 6 | 0,11 | 0,11 | 1 |
| 7 | 0,19 | 0,19 | 1 |
| 8 | 0,13 | 0,13 | 1 |
| 9 | 0,15 | 0,15 | 1 |
| 12 → 11 | 0,13 | 0,64 | 1,70 |
| 13 → 12 | 0,16 | 0,42 | 1,94 |
| 7 → 8 | 0,10 | 0,57 | 4,37 |

Tabela 36 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Junho

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,12 | 0,12 | 1 |
| 11 | 0,20 | 0,20 | 1 |
| 12 | 0,38 | 0,38 | 1 |
| 13 | 0,19 | 0,19 | 1 |
| 14 | 0,12 | 0,12 | 1 |
| 2 | 0,13 | 0,13 | 1 |
| 5 | 0,12 | 0,12 | 1 |
| 6 | 0,10 | 0,10 | 1 |
| 7 | 0,19 | 0,19 | 1 |
| 8 | 0,12 | 0,12 | 1 |
| 9 | 0,15 | 0,15 | 1 |
| 11 → 12 | 0,12 | 0,63 | 1,65 |
| 12 → 13 | 0,14 | 0,37 | 1,90 |

Tabela 37 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Julho

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,13 | 0,13 | 1 |
| 11 | 0,29 | 0,29 | 1 |
| 12 | 0,35 | 0,35 | 1 |
| 13 | 0,11 | 0,11 | 1 |
| 2 | 0,13 | 0,13 | 1 |
| 5 | 0,12 | 0,12 | 1 |
| 6 | 0,10 | 0,10 | 1 |
| 7 | 0,19 | 0,19 | 1 |
| 8 | 0,11 | 0,11 | 1 |
| 9 | 0,17 | 0,17 | 1 |
| 11 → 12 | 0,20 | 0,69 | 1,93 |

Tabela 38 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Agosto

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,12 | 0,12 | 1 |
| 11 | 0,25 | 0,25 | 1 |
| 12 | 0,40 | 0,40 | 1 |
| 13 | 0,20 | 0,20 | 1 |
| 14 | 0,13 | 0,13 | 1 |
| 2 | 0,12 | 0,12 | 1 |
| 5 | 0,12 | 0,12 | 1 |
| 6 | 0,10 | 0,10 | 1 |
| 7 | 0,17 | 0,17 | 1 |
| 8 | 0,10 | 0,10 | 1 |
| 9 | 0,16 | 0,16 | 1 |
| 11 → 12 | 0,16 | 0,67 | 1,66 |
| 13 → 12 | 0,14 | 0,36 | 1,83 |

Tabela 39 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Setembro

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,14 | 0,14 | 1 |
| 11 | 0,26 | 0,26 | 1 |
| 12 | 0,32 | 0,32 | 1 |
| 13 | 0,19 | 0,19 | 1 |
| 2 | 0,12 | 0,12 | 1 |
| 5 | 0,11 | 0,11 | 1 |
| 6 | 0,11 | 0,11 | 1 |
| 7 | 0,2 | 0,2 | 1 |
| 8 | 0,12 | 0,12 | 1 |
| 9 | 0,17 | 0,17 | 1 |
| 12 → 11 | 0,16 | 0,60 | 1,86 |

Tabela 40 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Outubro

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,13 | 0,13 | 1 |
| 11 | 0,19 | 0,19 | 1 |
| 12 | 0,36 | 0,36 | 1 |
| 13 | 0,24 | 0,24 | 1 |
| 14 | 0,11 | 0,11 | 1 |
| 3 | 0,11 | 0,11 | 1 |
| 5 | 0,10 | 0,10 | 1 |
| 6 | 0,11 | 0,11 | 1 |
| 7 | 0,16 | 0,16 | 1 |
| 8 | 0,17 | 0,17 | 1 |
| 9 | 0,15 | 0,15 | 1 |
| 11 → 12 | 0,11 | 0,60 | 1,66 |
| 13 → 12 | 0,15 | 0,41 | 1,73 |
| 8 → 7 | 0,10 | 0,59 | 3,43 |

Tabela 41 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Novembro

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,11 | 0,11 | 1,0 |
| 11 | 0,29 | 0,29 | 1 |
| 12 | 0,28 | 0,28 | 1 |
| 13 | 0,13 | 0,13 | 1 |
| 2 | 0,12 | 0,12 | 1 |
| 5 | 0,12 | 0,12 | 1 |
| 6 | 0,12 | 0,12 | 1 |
| 7 | 0,23 | 0,23 | 1 |
| 8 | 0,13 | 0,13 | 1 |
| 9 | 0,14 | 0,14 | 1 |
| 12 → 11 | 0,15 | 0,53 | 1,87 |
| 8 → 7 | 0,10 | 0,44 | 3,36 |

Tabela 42 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans — Referente ao mês de Dezembro

| Regra de associação | Support | Confidence | Lift |
|----------------------------|----------------|-------------------|-------------|
| 10 | 0,15 | 0,15 | 1,0 |
| 11 | 0,33 | 0,33 | 1 |
| 12 | 0,20 | 0,20 | 1 |
| 13 | 0,11 | 0,11 | 1 |
| 2 | 0,12 | 0,12 | 1 |
| 5 | 0,12 | 0,12 | 1 |
| 6 | 0,15 | 0,15 | 1 |
| 7 | 0,19 | 0,19 | 1 |
| 8 | 0,14 | 0,14 | 1 |
| 9 | 0,15 | 0,15 | 1 |
| 12 → 11 | 0,14 | 0,43 | 2,07 |

G.2 Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada), referentes aos meses do ano de 2017

Tabela 43 – Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de janeiro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 40 | 49.012 | 41.870 | 4.399 | 2.743 |
| Desastre Natural | 590 | 809.338 | 703.331 | 70.304 | 35.703 |
| Evento Social | 7 | 13.863 | 11.022 | 2.259 | 582 |
| Evento Urbano | 1 | 2.907 | 2.412 | 424 | 71 |
| Total | 638 | 875.120 | 758.635 | 77.386 | 39.099 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 23 eventos de exceção não atingiram linhas de ônibus.

Tabela 44 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de fevereiro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 45 | 49.452 | 39.294 | 8.213 | 1.945 |
| Desastre Natural | 316 | 336.685 | 278.368 | 47.390 | 10.927 |
| Evento Social | 8 | 7.972 | 5.590 | 2.077 | 305 |
| Evento Urbano | 5 | 7.750 | 6.391 | 960 | 399 |
| Total | 374 | 401.859 | 329.643 | 58.640 | 13.576 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 23 eventos de exceção não atingiram linhas de ônibus.

Tabela 45 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de março^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 36 | 37.421 | 30.253 | 4.604 | 2.564 |
| Desastre Natural | 184 | 279.140 | 243.777 | 22.276 | 13.087 |
| Evento Social | 48 | 56.600 | 46.990 | 6.006 | 3.604 |
| Evento Urbano | 49 | 55.893 | 46.005 | 6.076 | 3.812 |
| Total | 317 | 429.054 | 367.025 | 38.962 | 23.067 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 9 eventos de exceção não atingiram linhas de ônibus.

Tabela 46 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de abril^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 29 | 21.105 | 16.955 | 3.105 | 1.045 |
| Desastre Natural | 98 | 143.394 | 122.541 | 15.743 | 5.110 |
| Evento Social | 179 | 527.761 | 493.875 | 23.576 | 10.310 |
| Evento Urbano | 79 | 146.124 | 129.399 | 12.191 | 4.534 |
| Total | 385 | 838.384 | 762.770 | 54.615 | 20.999 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 3 eventos de exceção não atingiram linhas de ônibus.

Tabela 47 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de maio^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 194 | 263.940 | 228.631 | 22.846 | 12.463 |
| Desastre Natural | 99 | 130.660 | 111.003 | 12.778 | 6.879 |
| Evento Social | 80 | 87.786 | 70.995 | 12.082 | 4.709 |
| Evento Urbano | 51 | 108.674 | 98.464 | 7.283 | 2.927 |
| Total | 424 | 591.060 | 509.093 | 54.989 | 26.978 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 27 eventos de exceção não atingiram linhas de ônibus.

Tabela 48 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de junho^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 493 | 595.019 | 498.826 | 66.882 | 29.311 |
| Desastre Natural | 98 | 138.667 | 118.128 | 15.067 | 5.472 |
| Evento Social | 95 | 150.404 | 131.000 | 13.924 | 5.480 |
| Evento Urbano | 86 | 131.486 | 115.145 | 12.094 | 4.247 |
| Total | 772 | 1.015.576 | 863.099 | 107.967 | 44.510 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 72 eventos de exceção não atingiram linhas de ônibus.

Tabela 49 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de julho^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 515 | 696.754 | 596.893 | 65.641 | 34.220 |
| Desastre Natural | 51 | 52.739 | 44.165 | 6.319 | 2.255 |
| Evento Social | 58 | 97.610 | 87.833 | 6.184 | 3.593 |
| Evento Urbano | 133 | 138.378 | 116.202 | 14.998 | 7.178 |
| Total | 757 | 985.481 | 845.093 | 93.142 | 47.246 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 68 eventos de exceção não atingiram linhas de ônibus.

Tabela 50 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de agosto^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 459 | 677.882 | 585.663 | 62.135 | 30.084 |
| Desastre Natural | 112 | 150.161 | 128.268 | 13.640 | 8.253 |
| Evento Social | 83 | 63.040 | 49.253 | 9.327 | 4.460 |
| Evento Urbano | 186 | 286.852 | 249.817 | 24.670 | 12.365 |
| Total | 840 | 1.177.935 | 1.013.001 | 109.772 | 55.162 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 58 eventos de exceção não atingiram linhas de ônibus.

Tabela 51 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de setembro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 454 | 454.837 | 379.438 | 53.504 | 21.895 |
| Desastre Natural | 62 | 78.835 | 66.581 | 8.047 | 4.207 |
| Evento Social | 63 | 60.756 | 50.897 | 6.570 | 3.289 |
| Evento Urbano | 139 | 204.034 | 178.696 | 16.735 | 8.603 |
| Total | 718 | 798.462 | 675.612 | 84.856 | 37.994 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 40 eventos de exceção não atingiram linhas de ônibus.

Tabela 52 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de outubro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 391 | 649.238 | 578.523 | 46.116 | 24.599 |
| Desastre Natural | 162 | 246.022 | 218.846 | 17.877 | 9.299 |
| Evento Social | 68 | 68.507 | 57.616 | 7.069 | 3.822 |
| Evento Urbano | 90 | 140.985 | 125.946 | 9.175 | 5.864 |
| Total | 711 | 1.104.752 | 980.931 | 80.237 | 43.584 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 71 eventos de exceção não atingiram linhas de ônibus.

Tabela 53 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de novembro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 349 | 477.061 | 412.095 | 47.137 | 17.829 |
| Desastre Natural | 223 | 258.774 | 218.115 | 29.844 | 10.815 |
| Evento Social | 72 | 127.963 | 113.103 | 11.061 | 3.799 |
| Evento Urbano | 150 | 251.380 | 221.597 | 21.067 | 8.716 |
| Total | 794 | 1.115.178 | 964.910 | 109.109 | 41.159 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 53 eventos de exceção não atingiram linhas de ônibus.

Tabela 54 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de parada) aos eventos de exceção do mês de dezembro de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 8 | 8.821 | 7.339 | 1.146 | 336 |
| Desastre Natural | - | - | - | - | - |
| Evento Social | 1 | 543 | 372 | 89 | 82 |
| Evento Urbano | 1 | 6.577 | 6.402 | 130 | 45 |
| Total | 10 | 15.941 | 14.113 | 1.365 | 463 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

G.3 Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada), referentes aos meses do ano de 2017

Tabela 55 – Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de janeiro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 11 | 284 | 209 | 75 | 0 |
| Desastre Natural | 210 | 41.646 | 37.002 | 3.945 | 699 |
| Evento Social | 1 | 111 | 84 | 27 | 0 |
| Evento Urbano | 1 | 5 | 1 | 4 | 0 |
| Total | 223 | 42.046 | 37.296 | 4.051 | 699 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 383 eventos de exceção não atingiram linhas de ônibus.

Tabela 56 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de fevereiro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 14 | 663 | 477 | 182 | 4 |
| Desastre Natural | 123 | 12.595 | 10.346 | 2.091 | 158 |
| Evento Social | 5 | 706 | 617 | 85 | 4 |
| Evento Urbano | 4 | 139 | 92 | 43 | 4 |
| Total | 146 | 14.103 | 11.532 | 2.401 | 170 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 215 eventos de exceção não atingiram linhas de ônibus.

Tabela 57 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de março^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 17 | 2.368 | 2.044 | 271 | 53 |
| Desastre Natural | 76 | 11.188 | 9.664 | 1.378 | 146 |
| Evento Social | 29 | 10.072 | 9.206 | 527 | 339 |
| Evento Urbano | 22 | 823 | 575 | 248 | 0 |
| Total | 144 | 24.451 | 21.489 | 2.424 | 538 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 158 eventos de exceção não atingiram linhas de ônibus.

Tabela 58 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de abril^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 16 | 1.757 | 1.476 | 260 | 21 |
| Desastre Natural | 32 | 9.858 | 8.642 | 1.040 | 176 |
| Evento Social | 73 | 3.068 | 2.139 | 907 | 22 |
| Evento Urbano | 42 | 3.577 | 2.894 | 666 | 17 |
| Total | 163 | 18.260 | 15.151 | 2.873 | 236 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 171 eventos de exceção não atingiram linhas de ônibus.

Tabela 59 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de maio^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 102 | 18.064 | 15.179 | 2.523 | 362 |
| Desastre Natural | 28 | 5.338 | 3.908 | 1.330 | 100 |
| Evento Social | 42 | 7.118 | 6.396 | 576 | 146 |
| Evento Urbano | 29 | 3.027 | 2.567 | 435 | 25 |
| Total | 201 | 33.547 | 28.050 | 4.864 | 633 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 212 eventos de exceção não atingiram linhas de ônibus.

Tabela 60 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de junho^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 240 | 49.659 | 42.233 | 6.293 | 1.133 |
| Desastre Natural | 39 | 4.403 | 3.704 | 647 | 52 |
| Evento Social | 53 | 2.366 | 1.775 | 585 | 6 |
| Evento Urbano | 46 | 7.729 | 6.617 | 949 | 163 |
| Total | 378 | 64.157 | 54.329 | 8.474 | 1.354 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 397 eventos de exceção não atingiram linhas de ônibus.

Tabela 61 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de julho^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 233 | 60.501 | 53.828 | 5.382 | 1.291 |
| Desastre Natural | 20 | 3.681 | 3.104 | 521 | 56 |
| Evento Social | 33 | 10.965 | 9.338 | 1.359 | 268 |
| Evento Urbano | 73 | 25.140 | 22.954 | 1.947 | 239 |
| Total | 359 | 100.287 | 89.224 | 9.209 | 1.854 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 395 eventos de exceção não atingiram linhas de ônibus.

Tabela 62 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de agosto^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 211 | 68.926 | 62.995 | 4.666 | 1.265 |
| Desastre Natural | 36 | 3.318 | 2.712 | 555 | 51 |
| Evento Social | 57 | 10.380 | 8.835 | 1.288 | 257 |
| Evento Urbano | 96 | 22.585 | 19.837 | 2.262 | 486 |
| Total | 400 | 105.209 | 94.379 | 8.771 | 2.059 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 432 eventos de exceção não atingiram linhas de ônibus.

Tabela 63 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de setembro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 186 | 36.130 | 31.202 | 4.437 | 491 |
| Desastre Natural | 30 | 4.698 | 4.116 | 501 | 81 |
| Evento Social | 40 | 4.066 | 3.440 | 591 | 35 |
| Evento Urbano | 74 | 10.793 | 9.123 | 1.074 | 596 |
| Total | 330 | 55.687 | 47.881 | 6.603 | 1.203 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 371 eventos de exceção não atingiram linhas de ônibus.

Tabela 64 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de outubro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 189 | 41.610 | 36.983 | 3.793 | 834 |
| Desastre Natural | 68 | 9.356 | 8.132 | 1.064 | 160 |
| Evento Social | 43 | 7.948 | 6.256 | 1.581 | 111 |
| Evento Urbano | 39 | 1.574 | 1.089 | 470 | 15 |
| Total | 339 | 60.488 | 52.460 | 6.908 | 1.120 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 389 eventos de exceção não atingiram linhas de ônibus.

Tabela 65 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de novembro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 164 | 31.596 | 28.664 | 2.670 | 262 |
| Desastre Natural | 73 | 9.220 | 7.876 | 1.210 | 134 |
| Evento Social | 44 | 5.127 | 4.317 | 719 | 91 |
| Evento Urbano | 84 | 18.038 | 15.450 | 2.361 | 227 |
| Total | 365 | 63.981 | 56.307 | 6.960 | 714 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 415 eventos de exceção não atingiram linhas de ônibus.

Tabela 66 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de dezembro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 2 | 3.505 | 3.203 | 252 | 50 |
| Desastre Natural | — | — | — | — | — |
| Evento Social | — | — | — | — | — |
| Evento Urbano | 1 | 83 | 62 | 21 | 0 |
| Total | 3 | 3.588 | 3.265 | 273 | 50 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 7 eventos de exceção não atingiram linhas de ônibus.

G.4 Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota), referentes aos meses do ano de 2017

Tabela 67 – Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de janeiro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 40 | 40.098 | 32.671 | 5.171 | 2.256 |
| Desastre Natural | 596 | 631.546 | 529.593 | 70.416 | 31.537 |
| Evento Social | 7 | 2.398 | 1.768 | 317 | 313 |
| Evento Urbano | 1 | 191 | 77 | 106 | 8 |
| Total | 644 | 674.233 | 564.109 | 76.010 | 34.114 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 23 eventos de exceção não atingiram linhas de ônibus.

Tabela 68 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de fevereiro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 46 | 32.868 | 25.572 | 5.878 | 1.418 |
| Desastre Natural | 318 | 289.997 | 238.904 | 40.791 | 10.302 |
| Evento Social | 8 | 2.759 | 2.054 | 580 | 125 |
| Evento Urbano | 5 | 6.441 | 5.869 | 419 | 153 |
| Total | 377 | 332.065 | 272.399 | 47.668 | 11.998 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 23 eventos de exceção não atingiram linhas de ônibus.

Tabela 69 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de março^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 37 | 27.961 | 23.214 | 2.762 | 1.985 |
| Desastre Natural | 185 | 185.059 | 152.367 | 21.951 | 10.741 |
| Evento Social | 48 | 16.694 | 11.355 | 4.146 | 1.193 |
| Evento Urbano | 49 | 34.905 | 27.571 | 5.005 | 2.329 |
| Total | 319 | 264.619 | 214.507 | 33.864 | 16.248 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 8 eventos de exceção não atingiram linhas de ônibus.

Tabela 70 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de abril^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 29 | 13.217 | 9.687 | 2.346 | 1.184 |
| Desastre Natural | 100 | 109.838 | 95.618 | 10.437 | 3.783 |
| Evento Social | 178 | 327.216 | 299.992 | 14.896 | 12.328 |
| Evento Urbano | 79 | 110.442 | 97.346 | 10.111 | 2.985 |
| Total | 386 | 560.713 | 502.643 | 37.790 | 20.280 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 3 eventos de exceção não atingiram linhas de ônibus.

Tabela 71 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de maio^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 194 | 152.861 | 118.989 | 23.569 | 10.303 |
| Desastre Natural | 99 | 91.304 | 74.244 | 11.482 | 5.578 |
| Evento Social | 80 | 35.770 | 27.006 | 6.615 | 2.149 |
| Evento Urbano | 52 | 56.085 | 47.649 | 5.676 | 2.760 |
| Total | 425 | 336.020 | 267.888 | 47.342 | 20.790 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 27 eventos de exceção não atingiram linhas de ônibus.

Tabela 72 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de junho^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 496 | 439.434 | 360.398 | 56.034 | 23.002 |
| Desastre Natural | 99 | 81.202 | 64.798 | 11.847 | 4.557 |
| Evento Social | 95 | 74.447 | 61.934 | 8.605 | 3.908 |
| Evento Urbano | 87 | 68.764 | 56.225 | 9.140 | 3.399 |
| Total | 777 | 663.847 | 543.355 | 85.626 | 34.866 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 72 eventos de exceção não atingiram linhas de ônibus.

Tabela 73 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de julho^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 519 | 515.974 | 422.331 | 64.621 | 29.022 |
| Desastre Natural | 51 | 29.317 | 21.420 | 6.009 | 1.888 |
| Evento Social | 60 | 94.340 | 84.230 | 6.773 | 3.337 |
| Evento Urbano | 134 | 67.178 | 52.091 | 11.023 | 4.064 |
| Total | 764 | 706.809 | 580.072 | 88.426 | 38.311 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 64 eventos de exceção não atingiram linhas de ônibus.

Tabela 74 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de agosto^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 461 | 425.118 | 346.818 | 53.965 | 24.335 |
| Desastre Natural | 112 | 68.397 | 50.758 | 12.191 | 5.448 |
| Evento Social | 83 | 32.456 | 23.365 | 6.386 | 2.705 |
| Evento Urbano | 189 | 186.185 | 153.052 | 21.457 | 11.676 |
| Total | 845 | 712.156 | 573.993 | 93.999 | 44.164 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 58 eventos de exceção não atingiram linhas de ônibus.

Tabela 75 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de setembro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 454 | 326.108 | 256.377 | 49.967 | 19.764 |
| Desastre Natural | 62 | 46.819 | 37.428 | 5.974 | 3.417 |
| Evento Social | 63 | 16.752 | 11.645 | 3.603 | 1.504 |
| Evento Urbano | 142 | 176.393 | 146.187 | 23.019 | 7.187 |
| Total | 721 | 566.072 | 451.637 | 82.563 | 31.872 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 40 eventos de exceção não atingiram linhas de ônibus.

Tabela 76 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de outubro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 394 | 410.860 | 338.494 | 53.676 | 18.690 |
| Desastre Natural | 163 | 136.971 | 110.728 | 19.514 | 6.729 |
| Evento Social | 68 | 30.715 | 24.936 | 4.199 | 1.580 |
| Evento Urbano | 90 | 110.827 | 96.689 | 10.028 | 4.110 |
| Total | 715 | 689.373 | 570.847 | 87.417 | 31.109 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 69 eventos de exceção não atingiram linhas de ônibus.

Tabela 77 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de novembro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 350 | 382.405 | 320.812 | 46.428 | 15.165 |
| Desastre Natural | 227 | 206.393 | 169.314 | 29.285 | 7.794 |
| Evento Social | 72 | 49.443 | 40.079 | 7.405 | 1.959 |
| Evento Urbano | 151 | 141.936 | 118.720 | 15.840 | 7.376 |
| Total | 800 | 780.177 | 648.925 | 98.958 | 32.294 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 53 eventos de exceção não atingiram linhas de ônibus.

Tabela 78 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 1.000 m dos pontos de rota) aos eventos de exceção do mês de dezembro de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 8 | 5.464 | 4.443 | 817 | 204 |
| Desastre Natural | — | — | — | — | — |
| Evento Social | 1 | 47 | 21 | 24 | 2 |
| Evento Urbano | 1 | 4.545 | 4.425 | 74 | 46 |
| Total | 10 | 10.056 | 8.889 | 915 | 252 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

G.5 Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de rota), referentes aos meses do ano de 2017

Tabela 79 – Análise Apriori aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de rota) aos eventos de exceção do mês de janeiro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 23 | 6.231 | 5.490 | 572 | 169 |
| Desastre Natural | 350 | 451.142 | 417.741 | 26.136 | 7.265 |
| Evento Social | 6 | 13.059 | 12.683 | 216 | 160 |
| Evento Urbano | 1 | 2.517 | 2.133 | 357 | 27 |
| Total | 380 | 472.949 | 438.047 | 27.281 | 7.621 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 249 eventos de exceção não atingiram linhas de ônibus.

Tabela 80 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de rota) aos eventos de exceção do mês de fevereiro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 23 | 22.804 | 21.824 | 874 | 106 |
| Desastre Natural | 196 | 207.781 | 190.486 | 15.176 | 2.119 |
| Evento Social | 8 | 22.032 | 20.518 | 1.333 | 181 |
| Evento Urbano | 4 | 6.190 | 5.741 | 157 | 292 |
| Total | 231 | 258.807 | 238.569 | 17.540 | 2.698 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 148 eventos de exceção não atingiram linhas de ônibus.

Tabela 81 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de março^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 27 | 42.787 | 40.376 | 1.738 | 673 |
| Desastre Natural | 114 | 141.112 | 131.856 | 6.724 | 2.532 |
| Evento Social | 41 | 114.425 | 108.993 | 2.767 | 2.665 |
| Evento Urbano | 35 | 79.297 | 72.860 | 5.213 | 1.224 |
| Total | 217 | 377.621 | 354.085 | 16.442 | 7.094 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 95 eventos de exceção não atingiram linhas de ônibus.

Tabela 82 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de abril^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 21 | 24.651 | 22.449 | 1.798 | 404 |
| Desastre Natural | 56 | 68.582 | 62.947 | 5.049 | 586 |
| Evento Social | 134 | 259.328 | 243.509 | 13.394 | 2.425 |
| Evento Urbano | 65 | 114.482 | 108.400 | 4.752 | 1.330 |
| Total | 276 | 467.043 | 437.305 | 24.993 | 4.745 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 82 eventos de exceção não atingiram linhas de ônibus.

Tabela 83 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de maio^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 158 | 306.781 | 288.104 | 13.584 | 5.093 |
| Desastre Natural | 50 | 69.479 | 66.261 | 2.008 | 1.210 |
| Evento Social | 70 | 192.743 | 182.207 | 8.065 | 2.471 |
| Evento Urbano | 43 | 115.899 | 107.702 | 6.851 | 1.346 |
| Total | 321 | 684.902 | 644.274 | 30.508 | 10.120 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 119 eventos de exceção não atingiram linhas de ônibus.

Tabela 84 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de junho^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 349 | 581.205 | 545.257 | 26.753 | 9.195 |
| Desastre Natural | 59 | 67.471 | 64.213 | 2.805 | 453 |
| Evento Social | 88 | 190.986 | 183.260 | 6.111 | 1.615 |
| Evento Urbano | 75 | 150.374 | 139.968 | 8.316 | 2.090 |
| Total | 571 | 990.036 | 932.698 | 43.985 | 13.353 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 238 eventos de exceção não atingiram linhas de ônibus.

Tabela 85 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de julho^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 364 | 598.814 | 558.198 | 31.089 | 9.527 |
| Desastre Natural | 35 | 39.136 | 35.247 | 3.140 | 749 |
| Evento Social | 51 | 87.913 | 83.100 | 3.467 | 1.346 |
| Evento Urbano | 111 | 249.116 | 234.973 | 10.472 | 3.671 |
| Total | 561 | 974.979 | 911.518 | 48.168 | 15.293 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 216 eventos de exceção não atingiram linhas de ônibus.

Tabela 86 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de agosto^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 343 | 545.707 | 511.619 | 22.799 | 11.289 |
| Desastre Natural | 65 | 50.446 | 47.118 | 2.542 | 786 |
| Evento Social | 81 | 151.278 | 141.884 | 6.626 | 2.768 |
| Evento Urbano | 154 | 296.634 | 280.586 | 10.950 | 5.098 |
| Total | 643 | 1.044.065 | 981.207 | 42.917 | 19.941 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 227 eventos de exceção não atingiram linhas de ônibus.

Tabela 87 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de setembro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 326 | 535.819 | 502.231 | 25.708 | 7.880 |
| Desastre Natural | 41 | 45.235 | 42.651 | 1.499 | 1.085 |
| Evento Social | 56 | 162.314 | 152.317 | 7.528 | 2.469 |
| Evento Urbano | 113 | 253.610 | 240.924 | 8.764 | 3.922 |
| Total | 536 | 996.978 | 938.123 | 43.499 | 15.356 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 187 eventos de exceção não atingiram linhas de ônibus.

Tabela 88 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de outubro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 293 | 367.280 | 335.395 | 26.808 | 5.077 |
| Desastre Natural | 102 | 92.094 | 85.275 | 5.811 | 1.008 |
| Evento Social | 66 | 192.755 | 179.139 | 10.580 | 3.036 |
| Evento Urbano | 67 | 105.158 | 92.625 | 11.499 | 1.034 |
| Total | 528 | 757.287 | 692.434 | 54.698 | 10.155 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 225 eventos de exceção não atingiram linhas de ônibus.

Tabela 89 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de novembro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 266 | 351.118 | 326.702 | 19.933 | 4.483 |
| Desastre Natural | 116 | 109.570 | 103.424 | 5.091 | 1.055 |
| Evento Social | 64 | 135.590 | 126.090 | 7.748 | 1.752 |
| Evento Urbano | 125 | 228.973 | 213.329 | 11.799 | 3.845 |
| Total | 571 | 825.251 | 769.545 | 44.571 | 11.135 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 241 eventos de exceção não atingiram linhas de ônibus.

Tabela 90 – Análise *Apriori* aplicada as velocidades médias (intervalos de 5 minutos) ao conjunto de dados AVL da SPTrans correlacionados (a distância de 100 m dos pontos de parada) aos eventos de exceção do mês de dezembro^f de 2017

| Classe do evento | Total de eventos^a | Total de Regras de Associação^b | Esperadas^c | Não esperadas^d | Parcialmente inesperadas^e |
|-------------------------|-------------------------------------|--|------------------------------|----------------------------------|---|
| Acidente | 3 | 7.493 | 7.081 | 204 | 208 |
| Desastre Natural | — | — | — | — | — |
| Evento Social | — | — | — | — | — |
| Evento Urbano | 1 | 93 | 64 | 25 | 4 |
| Total | 4 | 7.586 | 7.145 | 229 | 212 |

^a Total de eventos de exceção.

^b Total de correlações de velocidade média.

^c Regras de associação esperadas ($Lift > 1$, $Support > 0,05$).

^d Regras de associação inesperadas ($Lift = 1$).

^e Regras de associação parcialmente inesperadas ($0 < Lift < 1$).

^f 6 eventos de exceção não atingiram linhas de ônibus.