

Arthur Segura Ortiz Novello
Luca Ezellner Miraglia
Lucas Marques de Araujo

Inteligência para Transporte Público

São Caetano do Sul
2020

Arthur Segura Ortiz Novello
Luca Ezellner Miraglia
Lucas Marques de Araujo

Inteligência para Transporte Público

Trabalho de Conclusão de Curso apresentado à Escola de Engenharia Mauá do Centro Universitário do Instituto Mauá de Tecnologia como requisito parcial para a obtenção de título de Engenheiro de Computação.

Área de Concentração: Engenharia de Computação

Orientador Tiago Sanches da Silva

Novello, Arthur Segura

Inteligência para Transporte Público / Arthur Segura Ortiz Novello , Luca Ezellner
Miraglia , Lucas Marques de Araujo . - São Caetano do Sul: CEUN-IMT, 2020.
34 p.

Trabalho de Conclusão de Curso - Escola de Engenharia Mauá do Centro
Universitário do Instituto Mauá de Tecnologia, São Caetano do Sul, SP, 2020.

Orientador: Prof. Me. Tiago Sanches da Silva

1. dados. 2. dashboard. 3. inteligência. 4. analytics. I. Miraglia, Luca Ezellner. II.
Araujo, Lucas Marques. III. Instituto Mauá de Tecnologia. Escola de Engenharia Mauá.
IV. Título.

Arthur Segura Ortiz Novello
Luca Ezellner Miraglia
Lucas Marques de Araujo

Inteligência para Transporte Público

Trabalho de Conclusão de Curso aprovado como requisito parcial para a obtenção de título de Engenheiro de Computação pela Escola de Engenharia Mauá do Centro Universitário do Instituto Mauá de Tecnologia.

Área de Concentração: Engenharia de Computação

Banca examinadora:

Prof. Me. Tiago Sanches da Silva
Orientador

Prof. Me. Murilo Zanini de Carvalho
Avaliador

Prof. Dr. Sergio Ribeiro Augusto
Avaliador

São Caetano do Sul, 29 de Abril de 2020.

Aos nossos pais, irmãos e irmãs, amigos e colegas de formação.

Agradecimentos

XXXXX

XXXXX

XXXXX

XXXXX

XXXXX

“XX”
("XX")
XX

Resumo

XXXXXXXXXXXXXXXX

Palavras-chaves: dados. dashboard. inteligência. analytics. decisão. ônibus.

Abstract

XXXXX

Key-words: data. dashboard. monitoring. training. transport. problem.

Lista de ilustrações

Figura 1 – Inteligência artificial, aprendizado de máquina e aprendizado profundo . . .	22
Figura 2 – Perceptron Elementar	23
Figura 3 – Rede Neural Multicamada	24
Figura 4 – Diagrama de blocos da solução	29
Figura 5 – Ilustração da solução	30

Lista de tabelas

Lista de abreviaturas e siglas

API	<i>Application Programming Interface</i>
CPTM	Companhia Paulista de Trens Metropolitanos
IBGE	Instituto Brasileiro de Geografia e Estatística
IA	Inteligência Artificial
OpenCV	<i>Open Source Computer Vision</i>
YOLO	<i>You Only Look Once</i>

Lista de símbolos

x

\mathbf{x}

Sumário

1	INTRODUÇÃO	15
1.1	Justificativa	15
1.2	Transporte público	15
1.3	Panorama econômico	15
1.4	Definição do problema	15
1.5	Mercados	16
1.6	Objetivos	16
1.7	Contribuições do trabalho	17
1.8	Oportunidades	17
1.9	Questão central da pesquisa	17
1.10	Impactos	17
2	REVISÃO BIBLIOGRÁFICA	19
2.1	Dashboard	19
2.1.1	Uma ferramenta de apoio à decisão	19
2.1.2	Benefícios do uso para as empresas	19
2.1.2.1	Auxilia na tomada de decisões	20
2.1.2.2	Transparência de informações	20
2.1.2.3	Otimização de tempo e recursos	20
2.1.2.4	Alinhamento estratégico	20
2.1.3	Antecipação de problemas	20
2.1.4	Experiência de uso	20
2.2	Inteligência Artificial, Machine Learning e Deep Learning	21
2.3	Deep Learning e Redes Neurais	22
2.3.1	História	22
2.3.2	Funcionamento das Redes Neurais	23
2.4	Assunto 3	25
2.4.1	SubAssunto 3	25
2.5	Análise de Dados	25
2.5.1	Primeira fase da análise de dados	25
2.5.2	Segunda fase da análise de dados	26
2.5.3	Terceira fase da análise de dados	26
2.5.4	Análise de dados no dia a dia	26
2.6	Tecnologia de Monitoramento de Ônibus	27
3	METODOLOGIA	29
3.1	Modelo da solução	29
3.1.1	Diagramas da solução	29

3.2	Python	30
3.3	Pandas	30
3.4	Numpy	30
3.5	OpenCV	31
3.6	Redes Neurais	31
3.7	YOLO	31
3.8	Django	31
3.9	Power BI	31
4	RESULTADOS OBTIDOS	32
4.1	Equações	32
4.2	Códigos Fonte de Programação	32
5	CONCLUSÕES	33
	REFERÊNCIAS	34

1 Introdução

Este capítulo apresenta conceitos iniciais do projeto, tais como justificativa, objetivos, panorama econômico, questão central da pesquisa e impactos.

1.1 Justificativa

O presente trabalho tem por motivação a grande quantidade de cidadãos de Santo André que apresentam adversidades quando se trata da utilização dos ônibus da cidade, como o elevado tempo de espera e as altas taxas de ocupação interna, que lideram as reclamações do Reclame Aqui. Além disso, a inexistência de um processo inteligente e automatizado que auxilie os gestores da SA-Trans no ajuste das frotas de ônibus justificam esse projeto.

1.2 Transporte público

O transporte público caracteriza-se como uma opção amplamente utilizada por pessoas a fim de garantirem suas necessidades de locomoção. Por possuir um preço mais acessível e muitas vezes ser mais rápido e prático, 65% da população das capitais do Brasil utiliza essa forma de transporte, como aponta um estudo realizado pelo Instituto de Pesquisa Econômica Aplicada (Ipea) (PEDUZZI, 2011).

1.3 Panorama econômico

Foi direcionado o valor de R\$ 707 milhões no ano de 2019 pela União para a área de mobilidade urbana e trânsito, como indica a Lei Orçamentária Anual (LOA nº 13.808/2019). Desse montante, especificamente para o transporte público coletivo, foram separados apenas R\$ 348 milhões, uma fatia de 0,01% do orçamento total (NTURBANO, 2019).

De acordo com uma estimativa do BNDES, em 2015, seria necessário investir mais de R\$ 234 bilhões em transporte público para resolver os problemas da área nas principais regiões metropolitanas do país, portanto, caso mantido o nível de investimento atual, levaria mais de 600 anos para atingir o montante proposto pelo BNDES (SANTOS, 2015).

1.4 Definição do problema

Tendo em vista o baixo investimento frente a demanda, é esperado que o transporte público gere um elevado nível de insatisfação em seus usuários, o que o IPEA demonstrou em

outra pesquisa realizada em 2011 e 2012, na qual o transporte público foi avaliado por mais de 60% do público como "péssimo ou ruim"(SANTOS, 2015).

Para entender a situação em que se encontra o transporte, o primeiro passo é olhar como o país se urbanizou. No último século, o Brasil passou por um intenso processo de industrialização, o que gerou um forte êxodo rural e principalmente uma migração da população do Nordeste para o Sul e Sudeste, onde se concentrou a produção industrial do país.

Esse crescimento populacional nas metrópoles foi acompanhado por uma grande valorização dos terrenos e moradias nas áreas centrais das cidades. Com isso é observado um efeito de gentrificação, afastando a classe trabalhadora para regiões mais periféricas, longe de onde se concentra a maior parte da oferta de emprego, o que gerou uma, demanda por transporte, crescente até os dias atuais. Tendo em vista esse cenário, a população começou a consumir cada vez mais carros populares, que contam com incentivo do governo para serem produzidos. Com isso, o que se vê nos ambientes urbanos são ruas congestionadas e ônibus lotados (PENA, s.d.).

1.5 Mercados

A situação não é diferente em Santo André, onde os ônibus destacam-se como o principal meio de transporte coletivo fornecido pelo município, contando com cerca de cinco milhões de passageiros por mês segundo dados da SA-Trans. Em seguida, o transporte ferroviário, mais especificamente as estações Celso Daniel, Prefeito Saladino e Utinga da linha 10 Turquesa da CPTM, ligando a cidade a São Paulo, somam cerca de um milhão e oitocentos mil usuários em Fevereiro de 2020, segundo dados fornecidos pela CPTM, que administra a linha.

Hoje, a cidade captura em tempo real e armazena vários dados relacionados ao transporte público. Desde a velocidade dos carros, ocorrências, horários de chegadas e partidas, informação de bilhetagem, entre outros. Porém, nem o município, nem a SA-Trans usam essas informações, o que possibilita que esse trabalho sirva de fundamento para uma informatização do transporte público local, e que permita melhorias no setor que não seriam possíveis apenas com as técnicas utilizadas atualmente.

1.6 Objetivos

Tendo em vista o cenário apresentado, pretende-se criar um dashboard para centralizar informações e agilizar a tomada de decisão, tendo como base dados coletados durante os trajetos de ônibus, informações extraídas de APIs públicas e imagens de câmeras instaladas nos veículos. A ferramenta irá apoiar os gestores da SA-Trans no controle do fluxo de tráfego, analisando a disponibilidade de veículos e sugerindo uma inclusão ou remoção de carros em cada linha monitorada.

Além disso, o dashboard será atualizado em tempo real, exibindo gráficos e indicadores de maneira sucinta, com a possibilidade de aplicação de filtros nas consultas. Considerando o volume de passageiros para um determinado dia e horário somados a dados externos, a ferramenta irá utilizar algoritmos de inteligência artificial para identificar padrões históricos e prever futuras demandas nas frotas de ônibus, facilitando o controle do serviço e auxiliando no dia a dia do transporte urbano na cidade.

1.7 Contribuições do trabalho

Este trabalho tem como contribuição a melhoria na tomada de decisões da administradora de ônibus (SA-Trans) onde será possível, de forma centralizada, a visualização de dados e gráficos por meio de um dashboard. Além disso, pode-se citar como contribuição a diminuição do tempo de espera e ocupação dos transportes, resultando em uma melhor qualidade de vida para a população de Santo André.

1.8 Oportunidades

O parque tecnológico faz parte do desenvolvimento da prefeitura de Santo André e tem como missão promover a inovação e potencializar as estruturas já existentes na cidade e região. O objetivo é estimular a extensão tecnológica e atuar nas oportunidades econômicas do ABC. Desta forma, nós, membros do projeto, temos como oportunidade através deste, participar como contribuintes no parque tecnológico da cidade.

1.9 Questão central da pesquisa

Os benefícios esperados com esse trabalho de conclusão de curso são proporcionar uma melhor qualidade locomotiva dos ônibus de Santo André e proporcionar uma melhor gestão aos envolvidos da SA-Trans, otimizando e inovando a forma de ver os dados.

1.10 Impactos

O transporte público, mais especificamente, os ônibus, podem causar estresse. A sensação pode estar relacionada a um fator estressante, como às condições do transporte, o desconforto de estar entre muitas pessoas e o elevado tempo de espera.

Com os dados disponíveis e as aplicações de IA implementadas nas câmeras dos veículos de cada linha, os gestores e responsáveis poderão ver a situação de lotação e do tempo de espera através do dashboard. Numa situação crítica, eles poderão solicitar um aumento da frota e reduzir o desconforto de andar em um veículo com excesso de pessoas além de reduzir o tempo de

espera. Essas ações impactam diretamente na saúde das pessoas reduzindo o estresse gerado pelos transportes.

2 Revisão bibliográfica

Este capítulo detalha toda revisão literária aplicada ao projeto, abordando assim conceitos teóricos, tecnologias e contexto histórico.

2.1 Dashboard

O Dashboard, também chamado de painel de controle, é uma ferramenta que auxilia os gestores a terem uma visão mais sistemática das principais informações do negócio. Em outras palavras, é um recurso que visa consolidar os dados de maior relevância em um painel, facilitando o processo de análise e a tomada de decisão.

O uso de planilhas e relatórios já são ultrapassados para análises de dados, não sendo suficientes para suprir as necessidades mais urgentes. Conforme a tecnologia foi evoluindo no mundo corporativo, surgiram os dashboards que evitam esforços desnecessários e ter uma visão mais ampla de todo o cenário corporativo para, assim, tomar decisões estratégicas e assertivas.

A visualização de dados através de dashboards já é uma realidade em softwares de gestão empresarial, integrando painéis de controle com inteligência artificial e fornecendo informações atualizadas automaticamente. É possível personalizá-los e comparar dados através de filtros, que facilitam análises de indicadores.

2.1.1 Uma ferramenta de apoio à decisão

Segundo um estudo feito pela UNINDU (The International Congress on University Industry), 83

O Business Intelligence (BI), por exemplo, é uma área que exige precisão na coleta e controle das informações para gerar insights em base desta ferramenta. Os dados são agrupados em conjuntos de registros e disponibilizadas por meio de dashboards para mensurar o desempenho atual e futuro da empresa de acordo com o cenário.

Além disso, os dashboards monitoram os dados, com o intuito de melhorar todos os processos. Eles permitem que o usuário personalize painéis e filtre informações para a visualização dos resultados, como quantidade, tempo e outras opções.

2.1.2 Benefícios do uso para as empresas

A aplicação dos dashboards na gestão empresarial traz muitos benefícios para a tomada de decisão e a visão estratégica do seu negócio.

2.1.2.1 Auxilia na tomada de decisões

O processo de tomada de decisões fica cada vez mais fácil através dos dashboards, que centralizam informações de fácil visualização e compreensão, possibilitando uma visão ampla do seu negócio.

2.1.2.2 Transparência de informações

Na gestão, é importante que todas as equipes tenham acesso aos indicadores da empresa, mantendo a transparência das informações. As ferramentas de dashboard tem o objetivo de facilitar a comunicação interna entre todos os profissionais.

2.1.2.3 Otimização de tempo e recursos

A visualização por dashboards otimiza o tempo para tomarem decisões e evitarem trabalhos manuais e complexos com a organização de dados, passando a priorizar outras atividades mais relevantes.

2.1.2.4 Alinhamento estratégico

Com as informações consolidadas em um único painel, a gestão se torna mais ágil e efetiva, possibilitando o alinhamento de estratégias e decisões para o negócio.

2.1.3 Antecipação de problemas

Como o dashboard trabalha com atualizações constantes e análises mais específicas, é mais fácil prever problemas e tendências negativas que podem vir a acontecer. Estes problemas ficam mais explícitos com o uso da tecnologia de inteligência artificial que os identifica de maneira mais fácil.

Qualquer mudança é detectada com mais simplicidade e o tempo para pensar nas possíveis soluções se torna muito maior. Isso melhora o processo de tomada de decisão e evita possíveis prejuízos.

2.1.4 Experiência de uso

Uma boa interface e uma boa experiência de uso se dá pela arquitetura das informações do dashboard. É fundamental que seja organizado, coerente e intuitivo. O objetivo é tornar o mais fácil possível encontrar o que se procura. Através de menus, cores e símbolos é possível saber quais são as opções e deixar claro as consequências que cada ação irá gerar. Dessa forma, a experiência do usuário ao usar o dashboard será rápida e efetiva, atendendo suas expectativas.

2.2 Inteligência Artificial, Machine Learning e Deep Learning

A Inteligência Artificial ganhou destaque mundialmente nos últimos anos como um importante e rentável campo da computação. A situação não é diferente no Brasil. Uma pesquisa realizada pela Microsoft, em 2019, revela que em um cenário de máxima utilização de inteligência artificial no Brasil, a taxa composta anual de crescimento (CAGR) do Produto Interno Bruto (PIB) pode aumentar para 7,1% ao ano, até 2030.

Andreas Kaplan e Michael Haenlein definem a inteligência artificial como “uma capacidade do sistema para interpretar corretamente dados externos, aprender a partir desses dados e utilizar essas aprendizagens para atingir objetivos e tarefas específicos através de adaptação flexível”.

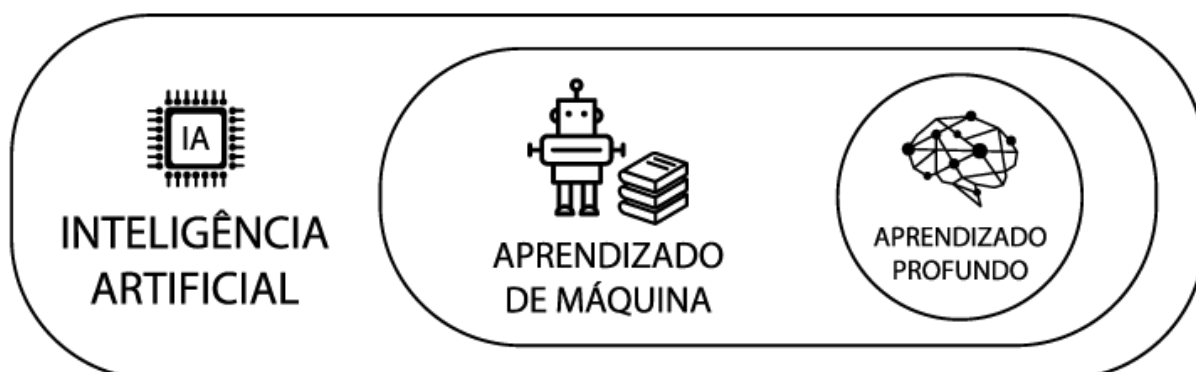
Um dos campos da Inteligência Artificial é o Machine Learning (Aprendizado de Máquina), que por sua vez pode ser definido como a capacidade de uma máquina, baseando-se em algoritmos e nos dados que estão sendo analisados possa tomar decisões capazes de resolver problemas ou prever comportamentos.

Existem duas maneiras de uma máquina aprender, criando duas categorias do Machine Learning, sendo elas aprendizagem supervisionada e aprendizagem não supervisionada. Na aprendizagem supervisionada, o cientista de dados é o responsável por monitorar o algoritmo, que irá aprender baseado em conhecimentos e experiências prévias. Resumidamente, que nessa categoria, sabe-se a saída que o algoritmo deve chegar tendo como base uma entrada. Como exemplo, pode-se citar a classificação de imagens e detecção de objetos.

Por outro lado, na aprendizagem não supervisionada, não existe um resultado específico esperado e o algoritmo irá aprender com base nos próprios dados que estão sendo processados. Em geral, necessita-se de um maior volume dados. Como exemplo, pode-se citar o processamento de linguagem natural para criação de legendas automáticas.

Por fim, dentre outras abordagens do Machine Learning, encontra-se o Deep Learning (Aprendizado Profundo), objeto de estudo deste trabalho, que é capaz de resolver problemas do mundo real por meio de redes neurais artificiais certas vezes com mais precisão e velocidade do que humanos.

Figura 1 – Inteligência artificial, aprendizado de máquina e aprendizado profundo



Fonte: Arquivo dos autores (2020)

2.3 Deep Learning e Redes Neurais

2.3.1 História

O início da história do deep learning data de 1943, quando o matemático Walter Pitts e o neurofisiologista Warren McCulloch, baseando-se em pesquisas do cérebro humano, modelaram o primeiro modelo computacional para uma rede neural, baseando-se em neurônios simples, criados a partir de circuitos elétricos. Walter e Warren utilizaram uma combinação de matemática e algoritmos, denominada de lógica de limiar (threshold logic), que é utilizada até os dias atuais como base das redes neurais.

Em 1960, Henry J. Kelley desenvolveu o conceito básico de um modelo de retropropagação (backpropagation), que também é utilizado até os dias atuais para o recálculo dos pesos das redes neurais no processo de aprendizagem. Apesar das descobertas e estudos nas décadas entre 40 a 80, devido à limitação computacional existente na época, o aprimoramento das redes neurais foi severamente impactado até 1981 e o assunto foi subestimado e criticado, além de ter sofrido grande redução no financiamento de pesquisas relacionadas à área.

Entretanto, em 1982, John Hopfield apresentou uma abordagem prática para redes neurais, demonstrando como elas poderiam atuar em problemas reais, o que fez com que o assunto voltasse a ter sua devida atenção. Em 1987 ocorreu a primeira Conferência Internacional sobre Redes Neurais do Institute of Electrical and Electronic Engineer's (IEEE). Ao mesmo tempo que o assunto vinha novamente ganhando destaque, o poder computacional da época ainda não conseguia acompanhar o desenvolvimento dos estudos, impossibilitando a criação de grandes aplicações com deep learning, o que gerou frustração e novamente uma redução de investimentos na área.

Ainda assim, algumas pessoas continuaram estudando o assunto e em 1995, Dana Cortes e Vladimir Vapnik desenvolveram a máquina de vetores de suporte, um sistema capaz de reconhecer padrões com aprendizado supervisionado. Nos anos seguintes, a partir de 1999, os computadores começaram a ganhar poder de processamento, viabilizando a implementação de redes neurais mais eficientes, que começaram a competir com as máquinas de vetores de suporte. Apesar de possuírem um maior tempo de processamento, em geral, as redes neurais se mostraram mais assertivas utilizando os mesmos dados e voltaram a ganhar visibilidade no mercado.

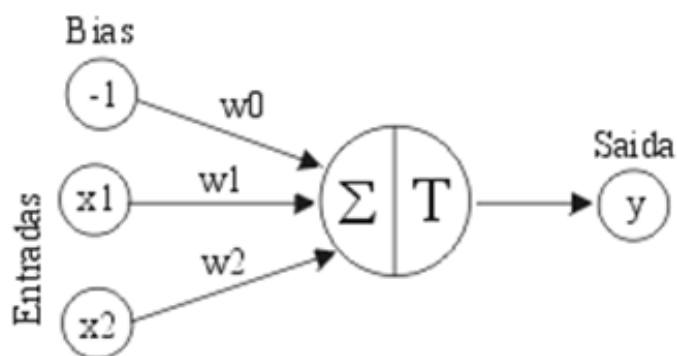
A partir de então diversos estudos e pesquisas realizados na área e à medida que os computadores foram evoluindo, as aplicações envolvendo redes neurais também o fizeram. Nos anos subsequentes, o algoritmo de aprendizagem profunda do Google foi capaz de identificar gatos em 2012 e em 2014 o Facebook implementou a DeepFace, tecnologia capaz de marcar automaticamente os rostos dos usuários da rede social em fotografias. Dois anos depois, o algoritmo do Google AlphaGo mapeou o jogo de tabuleiro Go e ganhou de Lee Sedol em um torneio em Seul, que à época tinha sido campeão mundial de Go 18 vezes.

Com isso, o campo da inteligência artificial e do deep learning em especial vem avançando em pesquisas até os dias atuais, além de novamente ter ganhado visibilidade e estar inserido em diversas aplicações corporativas que requerem uma análise de dados mais profunda, com reconhecimento de padrões, previsões ou classificações, por exemplo.

2.3.2 Funcionamento das Redes Neurais

O funcionamento das redes neurais atuais pode ser explicado tendo como base o perceptron elementar, o modelo mais simples e didático de rede neural de camada única.

Figura 2 – Perceptron Elementar



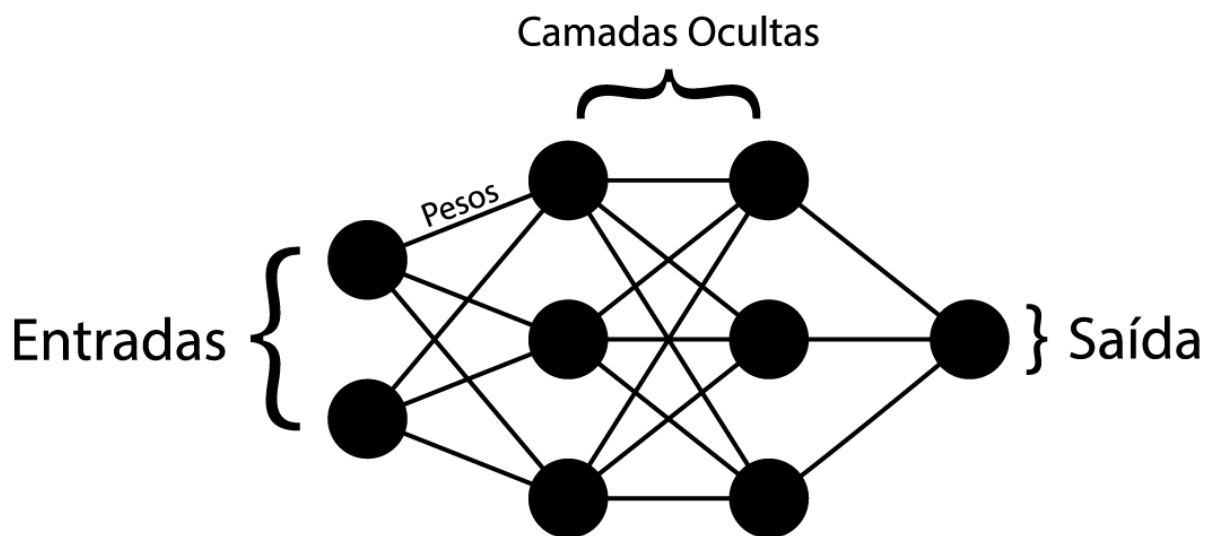
Fonte: UFSC, s.d. (<https://www.gsigma.ufsc.br/~popov/aulas/rna/neuronio_implementacao/>)

No modelo apresentado (Figura 2), constituído por apenas um único neurônio artificial, duas entradas, uma unidade de bias e uma saída, cada entrada possui pesos w_0 , w_1 e w_2 associados a elas, assim como a unidade de bias, que também funciona como uma entrada, mas possui seu valor fixado normalmente em 1 ou -1, com a finalidade de aumentar o grau de liberdade do ajuste dos pesos. Por padrão, em bibliotecas que possibilitam a implementação de redes neurais, a unidade de bias já é definida automaticamente.

Assim sendo, o algoritmo que simula o neurônio irá multiplicar cada entrada por seu respectivo peso para todas as entradas e realizar a somatória de todos esses resultados. Após a obtenção do valor da somatória, ele deve ser submetido a uma função de ativação ou transferência T , gerando a saída y . Dentre outras funções de transferência, pode-se citar a função sigmoide, tangente hiperbólica, unidade linear retificada (ReLU) e unidade linear exponencial (ELU).

Durante os estudos e desenvolvimento de algoritmos para redes neurais, ficou claro que para problemas mais complexos, seria necessário expandir a quantidade de neurônios e de camadas, tornando a saída de um neurônio, a entrada para o próximo. A quantidade de neurônios e de camadas deve ser estudada e testada individualmente para cada problema que se pretende solucionar, variando caso a caso.

Figura 3 – Rede Neural Multicamada



Fonte: Arquivo dos autores (2020)

As redes neurais artificiais multicamadas são divididas basicamente em três principais partes: entradas, camadas ocultas e saída, podendo ser uma ou várias dependendo do tipo de problema. Nesse tipo de rede, uma série de neurônios são interligados como mostrado e quando uma saída é obtida, ela é comparada com uma saída considerada correta, gerando um valor de

erro. Com base nesse erro, os pesos são então recalculados utilizando o conceito de gradiente para a minimização do erro, até que seu valor seja considerado satisfatório. Essa etapa em que os pesos são recalculados e a rede neural está sofrendo alterações é denominada de treinamento.

Após essa fase, a rede deve ser testada com dados diferentes dos que foram usados durante o treinamento para garantir sua assertividade. Essa etapa recebe o nome de teste. Caso comprovada sua eficiência e as taxas de erro permanecerem baixas, a rede neural está apta para ser utilizada com dados novos obtidos de um contexto real.

2.4 Assunto 3

XXXXXX

XXXXXX

2.4.1 SubAssunto 3

XXXXXX

XXXXXX

2.5 Análise de Dados

Apesar de estar cada vez mais em destaque nestes últimos anos, podemos traçar o uso de estatística para tomada de decisão desde do Egito Antigo, de acordo com um artigo publicado por Keith D. Foote (FOOTE, 2018) os egípcios já usavam cálculos estatísticos para construir as pirâmides. Já na década de 1880, o governo americano levou pelo menos 7 anos para completar o censo da população, processo que foi reduzido para um ano e meio na década seguinte por causa do desenvolvimento de uma máquina de tabulação, por Herman Hollerith, que processava os dados de forma sistêmica em cartões perfurados.

2.5.1 Primeira fase da análise de dados

Trazendo para os dias atuais, a Universidade de Villanova separa a análise de dados em três estágios (VILLANOVA, s.d.). O primeiro deles seria o início do que chamamos de Business Intelligence, que surgiu por volta de 1950 como uma forma de processar pequenas quantidades de informações estruturadas. Esse estágio, que podemos chamar de Analytics 1.0, durou até cerca de 2009, quando foi consolidado o termo "big-data".

O fim do primeiro estágio se deu pelo aumento exponencial de dados sendo produzidos diariamente, podendo vir de qualquer lugar e forma, desde de informações simples e estruturadas, como quais produtos alguém comprou em um determinado site, ou coisas mais complexas, como

quais sites fizeram essa pessoa chegar até uma determinada página e qual a posição geográfica dessa pessoa quando acessou a página. Se convencionou a chamar esses dados em grande quantidade e pouco estruturados de "big-data".

2.5.2 Segunda fase da análise de dados

Pela dificuldade de armazenar e processar essa massiva quantidade de dados, se fez necessário o desenvolvimento de novos métodos e ferramentas para o processamento deles, o que deu início ao Analytics 2.0, que trouxe alternativas como Hadoop, que pode processar essas grandes massas de dados, e o NoSQL, que permite armazenar toda a informação de forma mais eficiente que os bancos de dados relacionais faziam até então. Além disso, cada vez mais se fez necessário um conhecimento tecnológico para fazer essas análises, quando até então bastava o conhecimento de métodos estatísticos.

2.5.3 Terceira fase da análise de dados

Hoje em dia, muitos especialistas dizem que chegamos a um terceiro estágio, o Analytics 3.0, onde esses dados que são produzidos a qualquer momento podem ser usados como moeda de troca entre consumidor e fornecedor. Essa moeda pode ser analisada imediatamente pelo fornecedor, que passa a entregar uma experiência personalizada para quem está consumindo o seu produto.

2.5.4 Análise de dados no dia a dia

Com a disponibilidade de dados que temos, as empresas estão se dedicando cada vez mais em coletar e auxiliar as suas decisões nas análises feitas com eles. O Diretor de Estratégia e Marketing de Precisão da Coca-Cola, Justin De Graaf, afirmou em entrevista para ADMA, Association for Data-Driven Marketing & Advertising, que cada vez mais a empresa usa informações coletadas diretamente dos consumidores, como por meio de telefone, redes sociais ou e-mail, para criar desde campanhas publicitárias até novos produtos (TAN, 2017). Outra marca consolidada usando análise de dados para a tomada de decisão é a rede de hotéis Marriott, que, de acordo com o seu chief development officer, Eric Jacobs, tem usado esse tipo de informação para decidir como identificar, atrair e manter os seus cliente mais lucrativos (EISEN, 2018). Para isso eles coletam desde dados como padrão social dos clientes até se eles consomem mais jeans Levi's ou Gap.

Apesar disso, alguns resultados dessas análises podem ter um efeito contrário do que era pretendido no início. Caso a análise não seja feita com atenção, verificando sempre quem está sendo atingido, ela pode levar a alguns atritos e desgastar a imagem de quem usa essas informações.

Em 2012, o New York Time publicou uma história de como um pai descobriu a gravidez da filha através de uma promoção oferecida pela rede varejista Target, nos Estados Unidos

(DUHIGG, 2012). Andrew Pole, um estatístico da rede, foi designado para desenvolver um "preditor de gravidez", no fim do estudo ele conseguiu chegar a uma lista de 25 produtos, que geram uma probabilidade da cliente estar grávida de acordo com o seu preditor. Esses produtos contêm itens como manteiga de cacau, bolsas grandes o suficiente para caber pacotes de fraldas, e suplementos como magnésio e zinco.

A Target vinculava essa informação a um identificador do cliente, e então ofereceria um desconto na próxima visita que essa mesma pessoa fizesse a loja. Um ano após a aplicação deste preditor, o pai de uma adolescente foi a uma das lojas reclamar que sua filha tinha recebido um desconto relacionado a esse programa específico para grávidas, com isso, o gerente da loja se desculpou pelo erro e alguns dias depois ligou para se desculpar novamente. Entretanto, para a surpresa do gerente, durante essa ligação o cliente disse que a filha não tinha contado para ele que estava grávida, e que a Target soube antes dele do acontecido. Após esse caso, o departamento de marketing decidiu desacelerar o programa de análises de dados, e passar mais tempo avaliando qual impacto cada iniciativa pode ter.

2.6 Tecnologia de Monitoramento de Ônibus

De acordo com a informação exposta em uma publicação feita em 2017 pela Escola de Negócios da Universidade de Indiana, era previsto um crescimento anual de mais de 23% no mercado de Big Data durante o período de 2014 a 2019, com um custo de \$48,6 bilhões no último ano. Isso inclui um crescimento de 30% entre 2014 e 2015 de aparelhos conectados e dispositivos de IoT. Estes aparelhos geram uma quantidade enorme de dados valiosos para quem tiver interesse de processá-los (LEE, 2017).

Esse cenário hoje não é diferente para o setor de transporte público, com a prefeitura de Santo André capturando dados em tempo real da sua frota de ônibus, informações como bilhetagem, velocidade e paradas dos veículos, porém não utilizando a informação para a tomada de decisão.

Um exemplo de uso prático desses dados foi publicado em um artigo da IEE, para a Conferência Internacional da Logística e Transporte Avançado. Em uma colaboração entre a IBM e o Conselho da Cidade de Dublin foi realizado um projeto de cidade inteligente entre 2010 e 2013 (Ben Ayed; Ben Halima; ALIMI, 2015). A IBM passou a processar os dados gerados pela frota de ônibus, além de outras fontes, com intenção de reduzir o trânsito na cidade sem precisar alterar a sua estrutura atual, que conta com vários pontos históricos.

O início do processo se dava com informações coletadas do ônibus, como dados de GPS, velocidade, paradas e bilhetagem, e depois eram adicionadas novas informações vindas de sistemas de semáforos, CCTV, sistemas meteorológicos, entre outros. Todos esses dados então eram processados em um servidor da IBM e disponibilizados em mapa em tempo real do

transporte público de Dublin.

Com toda essa informação processada, a cidade teve uma maior capacidade de monitorar o seu sistema de transporte público, diminuindo o tempo para uma tomada de decisão.

Outro exemplo de aplicação de tecnologia no monitoramento de transporte coletivo é do USapiens, um sistema desenvolvido por um time de pesquisa da IBM do Brasil (VIEIRA et al., 2015). Essa equipe usou dados de transporte coletivo da cidade do Rio de Janeiro para desenvolver um sistema que processa os dados recebidos pelo GPS dos ônibus e depois disso analisa-se por diversos modelos esses dados.

Para isso eles integraram os dados obtidos pelo GPS com informações disponíveis de GTFS, sigla para General Transit Feed Specification, que contém dados mais gerais das rotas de ônibus, como paradas e horários esperados. Feita essa integração, os dados são limpos para prevenir problemas como latitude/longitude imprecisas ou dados com intervalo de tempo muito grande. Com os dados prontos, é feita uma comparação com a rota do GTFS, se descobre a direção do veículo e com isso tem seu trajeto normalizado em uma escala de distância e tempo acumulativas.

Com a informação normalizada, o sistema pode ajudar a responder três perguntas principais. Através de uma análise descritiva histórica podemos responder "O que aconteceu e porque?", analisando os dados em tempo real se responde "O que está acontecendo e porque?" e uma análise preditiva responde "O que vai acontecer e porquê?".

Os pesquisadores da IBM depois aplicaram esse sistema a 5 casos de estudo. O primeiro foi uma Análise de Uniformidade dos Ônibus, para evitar um agrupamento de veículos na linha, o segundo caso foi uma verificação na rotas dos ônibus, para avaliar a aderência do veículo a sua rota pré-definida, o terceiro uma análise de fluxo no trânsito, o quarto a variância do tempo de viagem do veículo, que permite avaliar a consistência da rota analisada, e para o quinto ele usaram a análise preditiva para prever o tempo de chegada do ônibus.

3 Metodologia

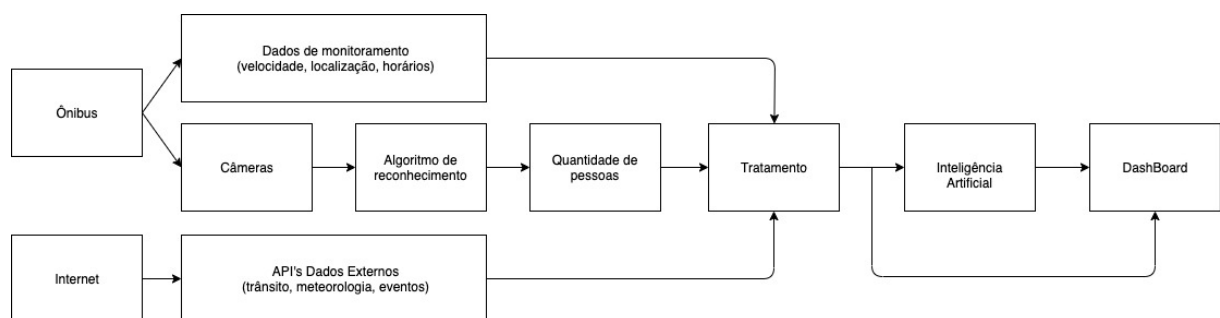
Este capítulo aborda os materiais e métodos utilizados no projeto, contendo maiores detalhes sobre algoritmos, tecnologias e estratégias empregadas na solução.

3.1 Modelo da solução

A fim de solucionar o problema identificado, foram definidos métodos e estratégias com o objetivo de criar um fluxo de trabalho. As figuras 4 e 5 fornecem um diagrama ilustrativo e simplificado da solução proposta.

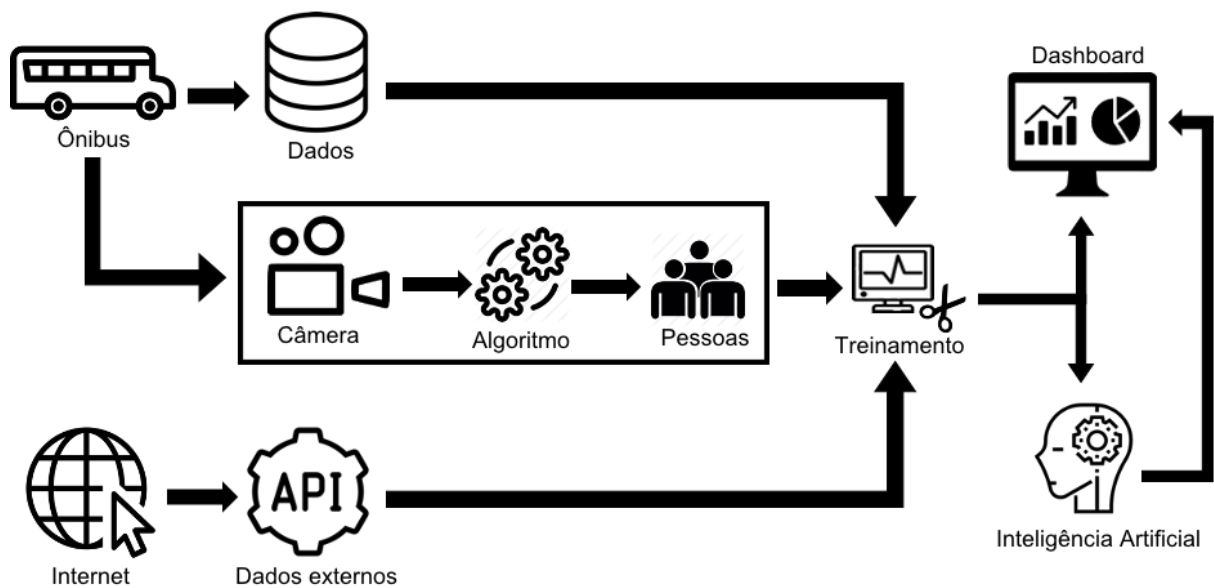
3.1.1 Diagramas da solução

Figura 4 – Diagrama de blocos da solução



Fonte: Arquivo dos autores (2020)

Figura 5 – Ilustração da solução



Fonte: Arquivo dos autores (2020)

3.2 Python

Linguagem de programação de alto nível lançada em 1991. Atualmente possui um modelo de desenvolvimento open source e gerenciado pela Python Software Foundation. A linguagem prioriza a legibilidade de código e possui poderosos recursos advindos de suas bibliotecas padrão combinados com bibliotecas de terceiros.

3.3 Pandas

Biblioteca desenvolvida em Python que possui estruturas de dados que facilitam e agilizam a manipulação e análise de dados.

3.4 Numpy

Biblioteca desenvolvida em Python criada para facilitar o desenvolvimento de aplicações com fins matemáticos e complexidade computacional avançada. Possui suporte para vetores e matrizes multidimensionais e diversas funções matemáticas para interação com essas estruturas.

3.5 OpenCV

Se trata de uma iniciativa open source que teve sua primeira versão lançada nos anos 2000 e continua em expansão até os dias atuais. Atualmente, suporta uma ampla variedade de algoritmos relacionados a Visão Computacional e Machine Learning, além de estar disponível em diversas linguagens de programação como C++, Python, Java e diferentes plataformas como Windows, Linux, OS X, Android e IOS.

3.6 Redes Neurais

Sistemas de computação que tem como objetivo reconhecer e classificar padrões em dados brutos. Tais sistemas buscam agir como o sistema nervoso humano, aprendendo e melhorando continuamente.

3.7 YOLO

Método para detecção e classificação de objetos em uma imagem combinando OpenCV e redes neurais. Se tornou popular pela sua grande eficiência e agilidade quando comparado com outros frameworks desenvolvidos anteriormente.

3.8 Django

Framework de alto nível, gratuito e open source desenvolvido em Python para programação de aplicações web. Apoia o desenvolvimento rápido e limpo, possuindo muitas ferramentas e métodos previamente construídos para facilitar e apoiar o desenvolvedor na criação das aplicações.

3.9 Power BI

Serviço de análise de negócios lançado em 2015 pela Microsoft. Permite a criação e compartilhamento de dashboards para a visualização de dados de forma interativa, além de possuir integração com outras ferramentas, como o Microsoft Excel.

4 Resultados Obtidos

XXXXX

XXXXX

4.1 Equações

XXXXX

4.2 Códigos Fonte de Programação

XXXXX

5 Conclusões

XXXXX

XXXXX

Referências

- Ben Ayed, A.; Ben Halima, M.; ALIMI, A. M. Big data analytics for logistics and transportation. **2015 4th IEEE International Conference on Advanced Logistics and Transport, IEEE ICAALT 2015**, p. 311–316, 2015. 27
- DUHIGG, C. **How Companies Learn Your Secrets**. 2012. Disponível em: <<https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&\>>. Acesso em: 2020-04-26. 27
- EISEN, D. **Marriott bets on predictive analytics for brand growth**. 2018. Disponível em: <<https://www.hotelmanagement.net/tech/marriott-builds-its-brands-by-knowing-more-about-you>>. Acesso em: 2020-04-26. 26
- FOOTE, K. D. **A Brief History of Analytics**. 2018. Disponível em: <<https://www.dataversity.net/brief-history-analytics/>>. Acesso em: 2020-04-26. 25
- LEE, I. Big data: Dimensions, evolution, impacts, and challenges. **Business Horizons**, "Kelley School of Business, Indiana University", v. 60, n. 3, p. 293–303, 2017. ISSN 00076813. Disponível em: <<http://dx.doi.org/10.1016/j.bushor.2017.01.004>>. 27
- NTURBANO. Só 0,01% do Orçamento Federal para o transporte coletivo urbano. **Revista NTurbano**, 2019. Disponível em: <<https://www.ntu.org.br/novo/NoticiaCompleta.aspx?idArea=10&idNoticia=1>>. Acesso em: 2020-04-18. 15
- PEDUZZI, P. **Estudo do Ipea mostra que 65% da população usam transporte público nas capitais**. 2011. Disponível em: <<https://memoria.ebc.com.br/agenciabrasil/noticia/2011-05-04/estudo-do-ipea-mostra-que-65-da-populacao-usam-transporte-publico-nas-capitais>>. Acesso em: 2020-04-18. 15
- PENA, R. F. A. **Problemas no transporte público**. s.d. Disponível em: <https://www.reclameaqui.com.br/sa-trans-santo-andre-tranportes/demora-de-onibus_aY3RgrYFWE-gTSb-/>. Acesso em: 2020-04-18. 16
- SANTOS, G. **Transporte público precisa de R\$ 235 bi em investimento, estima BNDES**. 2015. Disponível em: <<https://www1.folha.uol.com.br/mercado/2015/09/1685379-transporte-publico-precisa-de-r-235-bi-em-investimento-estima-bndes.shtml>>. Acesso em: 2020-04-18. 15, 16
- TAN, A. **How Coca-Cola uses data to supercharge its superbrand status**. 2017. Disponível em: <<https://www.adma.com.au/resources/how-coca-cola-uses-data-to-supercharge-its-superbrand-status>>. Acesso em: 2020-04-26. 26
- VIEIRA, M. R.; BARBOSA, L.; KORMÁKSSON, M.; ZADROZNY, B. USapiens: A System for Urban Trajectory Data Analytics. **Proceedings - IEEE International Conference on Mobile Data Management**, v. 1, p. 255–262, 2015. ISSN 15516245. 28
- VILLANOVA. **The Evolution of Data Collection and Analytics**. s.d. Disponível em: <<https://taxandbusinessonline.villanova.edu/blog/the-evolution-of-data-collection-and-analytics/>>. Acesso em: 2020-04-26. 25