

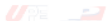
Defesa de Mestrado

Othon Luiz Teixeira de Oliveira

Escola Politécnica de Pernambuco – Poli — UPE

29 - Maio - 2017

PPGES



Programa de Pós-Graduação em Engenharia de Sistemas

DISSERTAÇÃO DE MESTRADO

Programa de Pós-Graduação em Engenharia de Sistemas

DISSERTAÇÃO DE MESTRADO

Modelo Preditivo para Sugestão de Roteamento de cargas considerando dados históricos, sócio-ambientais, e de redes sociais

Programa de Pós-Graduação em Engenharia de Sistemas

DISSERTAÇÃO DE MESTRADO

Modelo Preditivo para Sugestão de Roteamento de cargas considerando dados históricos, sócio-ambientais, e de redes sociais

Mestrando: Othon Luiz Teixeira de Oliveira

Orientador: Prof. Dr. Fernando Buarque de Lima Neto

Sumário

- 1 Introdução
 - Introdução
 - Objetivos
- 2 Background
- 3 Contribuição
- 4 Resultados
- 5 Conclusões

Resumo

O cenário

- O transporte de cargas no Brasil é feito principalmente pelas rodovias federais (BRs).

Resumo

O cenário

- O transporte de cargas no Brasil é feito principalmente pelas rodovias federais (BRs).
- Essas rodovias estão constantemente congestionadas nos perímetros urbanos.

Resumo

O cenário

- O transporte de cargas no Brasil é feito principalmente pelas rodovias federais (BRs).
- Essas rodovias estão constantemente congestionadas nos perímetros urbanos.
- Comunidades bloqueiam as rodovias para reivindicar, dos entes públicos, todo tipo de necessidades

Resumo

O cenário

- O transporte de cargas no Brasil é feito principalmente pelas rodovias federais (BRs).
- Essas rodovias estão constantemente congestionadas nos perímetros urbanos.
- Comunidades bloqueiam as rodovias para reivindicar, dos entes públicos, todo tipo de necessidades
- Em alguns trechos o traçado das rodovias está próximo a morros e florestas.

Proposição de um modelo preditivo

Proposição de um modelo preditivo

Objetivo Geral

Esta pesquisa teve como principal objetivo desenvolver um modelo de plataforma autoadaptável

Proposição de um modelo preditivo

Objetivo Geral

Esta pesquisa teve como principal objetivo desenvolver um modelo de plataforma autoadaptável que contemple predição do comportamento das rodovias federais pernambucanas,

Proposição de um modelo preditivo

Objetivo Geral

Esta pesquisa teve como principal objetivo desenvolver um modelo de plataforma autoadaptável que contemple predição do comportamento das rodovias federais pernambucanas, antecipando alguns eventos que nela possam ocorrer, apontando onde ocorrerão.

Objetivos Específicos

Objetivos Específicos

Quatro objetivos

- Caracterizar a problemática de cada rodovia;

Objetivos Específicos

Quatro objetivos

- Caracterizar a problemática de cada rodovia;
- Desenvolver um modelo preditivo dos fenômenos que envolvem as rodovias;

Objetivos Específicos

Quatro objetivos

- Caracterizar a problemática de cada rodovia;
- Desenvolver um modelo preditivo dos fenômenos que envolvem as rodovias;
- Desenvolver um ambiente de simulações interativas da estrutura viária em sua dinâmica

Objetivos Específicos

Quatro objetivos

- Caracterizar a problemática de cada rodovia;
- Desenvolver um modelo preditivo dos fenômenos que envolvem as rodovias;
- Desenvolver um ambiente de simulações interativas da estrutura viária em sua dinâmica
- Propor soluções para melhorar a experiência dos usuários que utilizam as rodovias pernambucanas

Sumário

1 Introdução

2 Background

- Dados a minerar
- Metodologias de Mineração Dados — CRISP – DM
- Minerando os dados do problema
- Árvore de Decisão
- Naïve Bayes
- Redes Neurais
- Tecnologias empregues na Mineração em Textos
- Análise de Desempenho aplicados à mineração

Dados originais da PRF

- Base de dados da Polícia Rodoviária Federal: acidente e interdições, entre 2007 a 2015.

Dados originais da PRF

- Base de dados da Polícia Rodoviária Federal: acidente e interdições, entre 2007 a 2015.
- Rede Social Twitter com 3200 tweets (limite permitido), de 2017 a 2014

Planilha com os dados originais

Planilha PRF

Figura: Planilha para Preprocessamento

Table "default" - Rows: 85209										Spec - Columns: 22		Properties		Flow Variables	
Row ID	↑ BR	↑ KM	↑ Latitude	↑ Longitude	↑ Condi...	↑ Restri...	↑ TipoAcid...	↑ CausaAcide...	↑ S...						
Row0	101	79	-7.620626749...	-34.9905194499...	Seca	Inexistente	Colisão late...	Falta de atenç...	De						
Row1	101	210	-7.66697808649...	-34.9329743899...	Seca	Inexistente	Colisão Tra...	Outras	De						
Row2	101	416	-7.83434453699...	-34.91253397799...	Seca	Inexistente	Colisão Tra...	Falta de atenç...	Cre						
Row3	101	416	-7.83434453699...	-34.91253397799...	Seca	Inexistente	Colisão Tra...	Falta de atenç...	Cre						
Row4	101	474	-7.86058770949...	-34.908337135499...	Seca	Inexistente	Colisão com...	Outras	Cre						
Row5	101	511	-7.90212986999...	-34.900132434499...	Seca	Inexistente	Atropelame...	Outras	Cre						
Row6	101	512	-7.91819106999...	-34.896062575499...	Seca	Inexistente	Colisão Tra...	Falta de atenç...	Cre						
Row7	101	512	-7.91819106999...	-34.896062575499...	Seca	Inexistente	Colisão Tra...	Falta de atenç...	Cre						
Row8	101	545	-7.94363545599...	-34.904202790499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre						
Row9	101	545	-7.94363545599...	-34.904202790499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre						
Row10	101	589	-7.97661785749...	-34.925137692499...	Seca	Inexistente	Colisão late...	Velocidade inc...	De						
Row11	101	680	-8.87953132499...	-35.628922247499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre						
Row12	101	680	-8.87953132499...	-35.628922247499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre						
Row13	101	680	-8.87953132499...	-35.628922247499...	Seca	Inexistente	Colisão late...	Ultrapassagem...	Cre						
Row14	101	680	-8.87953132499...	-35.628922247499...	Seca	Inexistente	Colisão late...	Ultrapassagem...	Cre						
Row15	101	680	-8.87953132499...	-35.628922247499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre						
Row16	101	680	-8.87953132499...	-35.628922247499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre						
Row17	101	680	-8.87953132499...	-35.628922247499...	Seca	Inexistente	Queda de m...	Outras	De						
Row18	101	680	-8.87953132499...	-35.628922247499...	Seca	Inexistente	Queda de m...	Velocidade inc...	Cre						
Row19	101	830	-8.17277505899...	-34.939988807999...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De						
Row20	101	830	-8.17277505899...	-34.939988807999...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De						
Row21	101	830	-8.17277505899...	-34.939988807999...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De						
Row22	101	830	-8.17277505899...	-34.939988807999...	Seca	Inexistente	Queda de m...	Outras	De						
Row23	101	865	-8.18644581449...	-34.971397555499...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De						
Row24	101	1018	-8.26668917249...	-35.041692694499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre						
Row25	101	1018	-8.26668917249...	-35.041692694499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre						
Row26	101	1020	-8.28279581549...	-35.057672119999...	Com bura...	Inexistente	Colisão late...	Outras	Cre						
Row27	101	1020	-8.28279581549...	-35.057672119999...	Com bura...	Inexistente	Colisão late...	Outras	Cre						

Entendendo a base de dados da PRF

Atributos e Instâncias – iniciais

- Dados de 2007 à 2015 – BRs de Pernambuco

Entendendo a base de dados da PRF

Atributos e Instâncias – iniciais

- Dados de 2007 à 2015 – BRs de Pernambuco
- 85.209 Instâncias – 27 Atributos

Entendendo a base de dados da PRF

Atributos e Instâncias – iniciais

- Dados de 2007 à 2015 – BRs de Pernambuco
- 85.209 Instâncias – 27 Atributos
- Dentre eles: Km, Latitude, Longitude, Condições da Pista, Causa do Acidente, Município, Data, Hora, Tipo de Veículo, Quantidade de Mortos, etc.

Entendendo a base de dados da PRF

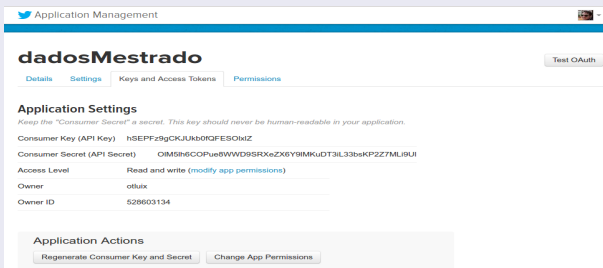
Atributos e Instâncias – iniciais

- Dados de 2007 à 2015 – BRs de Pernambuco
- 85.209 Instâncias – 27 Atributos
- Dentre eles: Km, Latitude, Longitude, Condições da Pista, Causa do Acidente, Município, Data, Hora, Tipo de Veículo, Quantidade de Mortos, etc.
- 40% missing data – 70% do tempo total para tratar

O Twitter

Acesso aos dados do Twitter

Figura: Registro de App para para acessar e baixar dados



Dados originais

Dados – Planilha Twitter

Figura: Planilha para Préprocessamento

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	text	favorited	favor	reply	created	trund	reply	id	repl	statusSource	screenNam	retweetC	isRetw	retwee	lor	latitude
2	00 MORT 175 TEST DE ALCOLE 01 PRISA DE ALCOOLEMIA 722 PESS FISC 631 VEIC FISC															
3	https://t.co/0xH4TrzGi	FALSE		8	NA	04/03/17 09:20	FALS	NA	83795601	NA	<a href="http://twitter	PRF191PE	2	FALSE	FALSE	NA
4	BALANÇO DIA 02/03/17: 09ACID; 12VEIC; 12FERIDOS; 00MOR	FALSE		5	NA	04/03/17 03:36	FALS	NA	83786925	NA	<a href="http://twitter	PRF191PE	1	FALSE	FALSE	NA
5	PRF registra redução de 73% no número de mortes durante Ope	FALSE		6	NA	03/03/17 03:41	FALS	NA	83750809	NA	<a href="http://twitter	PRF191PE	0	FALSE	FALSE	NA
6	Mais de 1800 imagens por excesso... https://t.co/z7yDc63aT	FALSE		5	NA	02/03/17 14:29	FALS	NA	83730901	NA	<a href="http://www.t	PRF191PE	0	FALSE	FALSE	NA
7	PROTESTO: Em Agressina/PE, na BR 104, km 88, ocorre uma	FALSE		12	NA	25/02/17 18:35	FALS	NA	83555880	NA	<a href="http://twitter	PRF191PE	3	FALSE	FALSE	NA
8	Cinco suspeitos de furto são detidos pela PRF na BR 101															
9	Quatro homens e uma mulher suspeitos de integrarem uma... https://t.co/yDpxuF1ZY	FALSE		2	NA	25/02/17 17:32	FALS	NA	83554306	NA	<a href="http://www.t	PRF191PE	2	FALSE	FALSE	NA
10	https://t.co/yDpxuF1ZY	FALSE		2	NA	25/02/17 16:57	FALS	NA	83553417	NA	<a href="http://www.t	PRF191PE	0	FALSE	FALSE	NA
11	Caminhão frigorífico é recuperado logo após roubo na BR 232	FALSE		32	NA	25/02/17 02:44	FALS	NA	83531957	NA	<a href="http://twitter	PRF191PE	15	FALSE	FALSE	NA
12	Um caminhão frigorífico carregado com carnes foi... https://t.co/yDpxuF1ZY	FALSE		8	NA	24/02/17 17:35	FALS	NA	83518149	NA	<a href="http://www.t	PRF191PE	0	FALSE	FALSE	NA
13	https://t.co/yDpxuF1ZY	FALSE		2	NA	24/02/17 17:33	FALS	NA	83518084	NA	<a href="http://www.t	PRF191PE	0	FALSE	FALSE	NA
14	https://t.co/yDpxuF1ZY	FALSE		10	NA	24/02/17 04:55	FALS	NA	83499021	NA	<a href="http://twitter	PRF191PE	2	FALSE	FALSE	NA
15	ACID. FERIDOS:colisão frontal entre 2 micro-ônibus em Ponte	FALSE		9	NA	23/02/17 03:48	FALS	NA	83461081	NA	<a href="http://twitter	PRF191PE	0	FALSE	FALSE	NA
16	Av. Recife bloqueada nas proximidades da empresa Brinks, nº 49	FALSE		7	NA	22/02/17 13:37	FALS	NA	83439670	NA	<a href="http://twitter	PRF191PE	0	FALSE	FALSE	NA
17	ACESSO À AV. RECIFE PARA QUEM VEM PELA BR-101 ESTÁ	FALSE		9	NA	22/02/17 08:28	FALS	NA	83431901	NA	<a href="http://twitter	PRF191PE	1	FALSE	FALSE	NA
18	https://t.co/cYmV24bQZY	FALSE		23	NA	21/02/17 09:37	FALS	NA	83397390	NA	<a href="http://twitter	PRF191PE	21	FALSE	FALSE	NA
19	Após tentar fugir da PRF, suspeito de assalto é preso em Caruar	FALSE		85	NA	21/02/17 08:34	FALS	NA	83395803	NA	<a href="http://twitter	PRF191PE	69	FALSE	FALSE	NA
20		FALSE		8	NA	21/02/17 07:50	FALS	NA	83394688	NA	<a href="http://twitter	PRF191PE	1	FALSE	FALSE	NA

Entendendo a base de dados do Twitter

Atributos e Instâncias – iniciais

- Dados de 2017 – 2014 — canal @PRF191PE

Entendendo a base de dados do Twitter

Atributos e Instâncias – iniciais

- Dados de 2017 – 2014 — canal @PRF191PE
- 2864 Instâncias – 16 Atributos

Entendendo a base de dados do Twitter

Atributos e Instâncias – iniciais

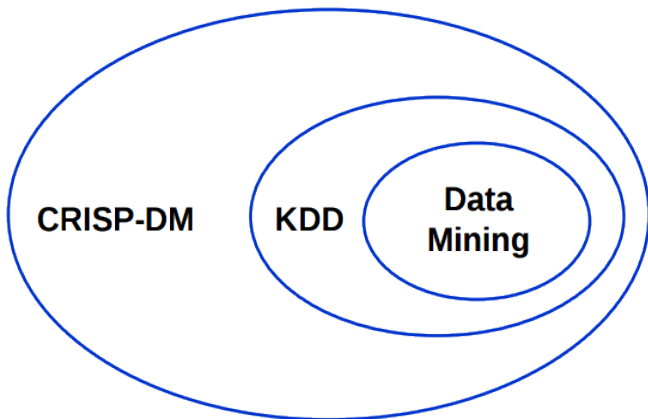
- Dados de 2017 – 2014 — canal @PRF191PE
- 2864 Instâncias – 16 Atributos
- Dentre eles: 'text', 'favorited', 'favoriteConunt', 'created', 'ID', 'statusSource', 'screenName', 'retweetCount',

Entendendo a base de dados do Twitter

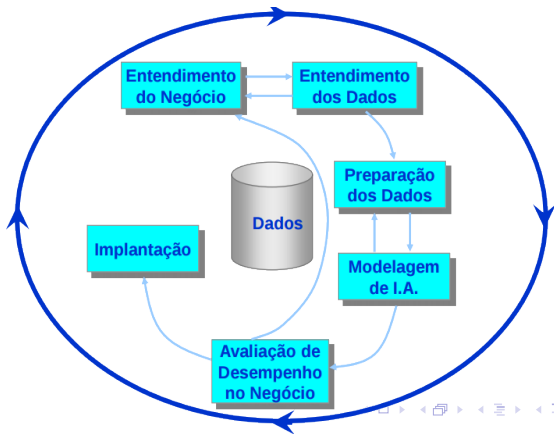
Atributos e Instâncias – iniciais

- Dados de 2017 – 2014 — canal @PRF191PE
- 2864 Instâncias – 16 Atributos
- Dentre eles: 'text', 'favorited', 'favoriteConunt', 'created', 'ID', 'statusSource', 'screenName', 'retweetCount',
- 'isRetweet', 'retweeted', dentre outros.

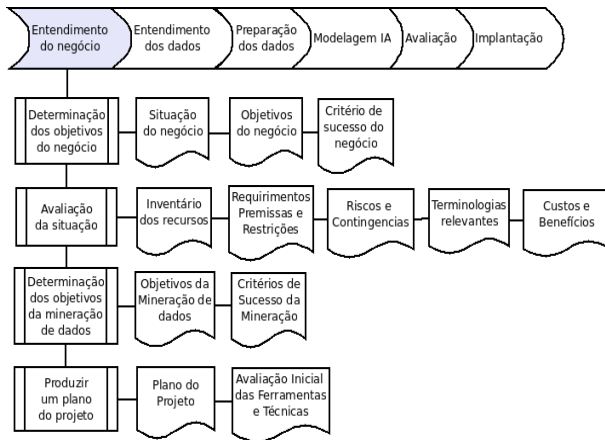
Domínio das técnicas de Mineração de Dados



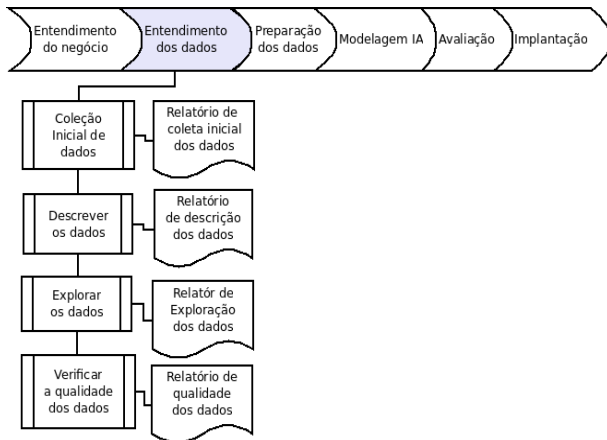
Cross Industry Standard Process for Data Mining – CRISP - DM



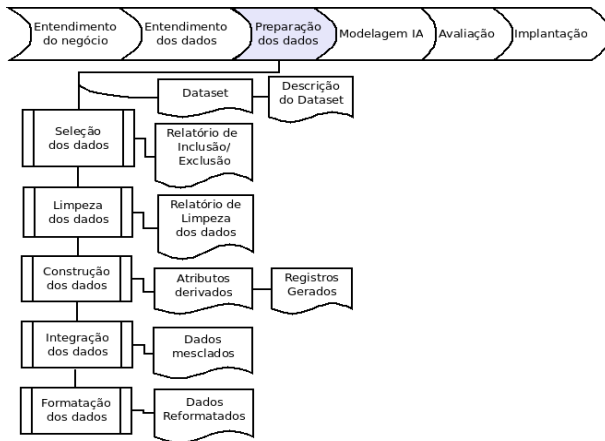
Etapas CRISP-DM resumo



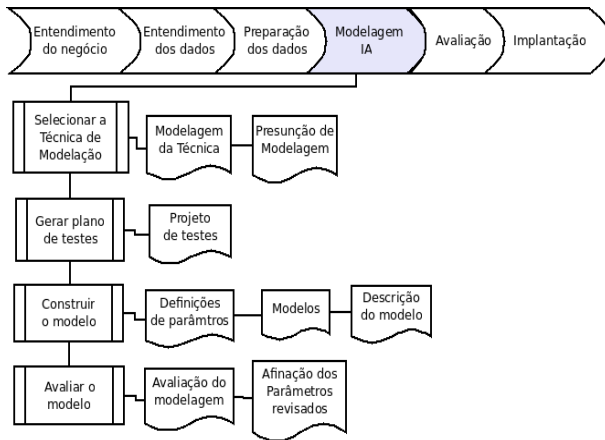
Etapas CRISP-DM resumo



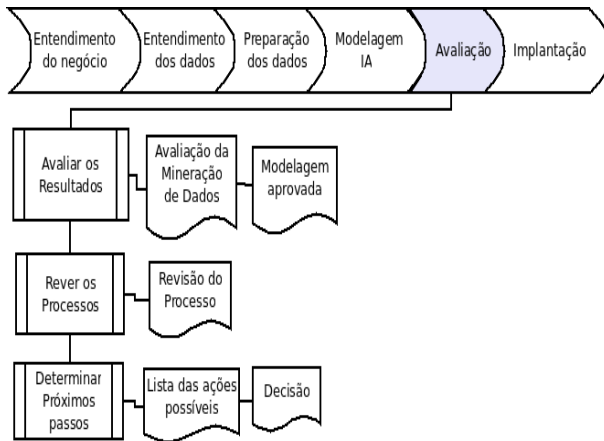
Etapas CRISP-DM resumo



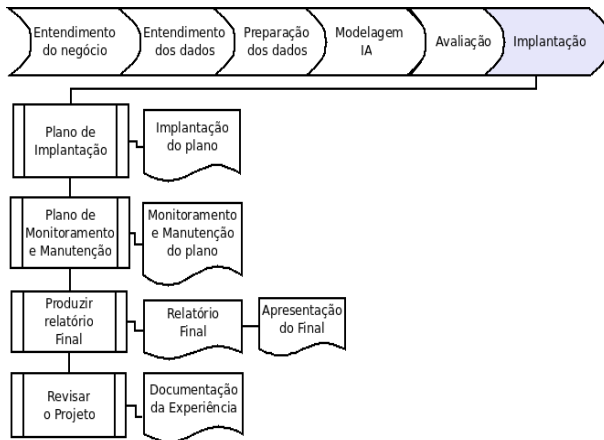
Etapas CRISP-DM resumo



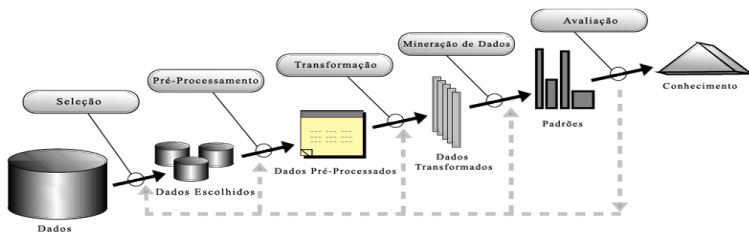
Etapas CRISP-DM resumo



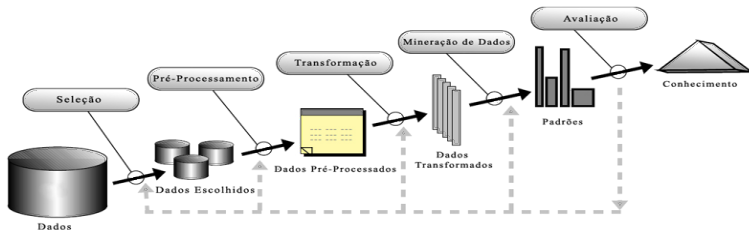
Etapas CRISP-DM resumo



A descoberta de conhecimento (KDD) nas bases da PRF



A descoberta de conhecimento (KDD) nas bases da PRF



Diferença CRISP-DM — KDD

O CRISP-DM difere do KDD principalmente pelas fases do entendimento do negócio (anterior ao KDD) e da implantação (posterior ao KDD)

A escolha

Pontos fortes

- Custo benefício – Uma árvore é gerada de maneira simples com resultados abrangente e facilidade de interpretação.

A escolha

Pontos fortes

- Custo benefício – Uma árvore é gerada de maneira simples com resultados abrangente e facilidade de interpretação.
- Permite fazer predição e classificação instantâneas

Pontos fortes

- Custo benefício – Uma árvore é gerada de maneira simples com resultados abrangente e facilidade de interpretação.
- Permite fazer predição e classificação instantâneas

Pontos fracos

- Nó raiz não é facilmente identificável

Pontos fortes

- Custo benefício – Uma árvore é gerada de maneira simples com resultados abrangente e facilidade de interpretação.
- Permite fazer predição e classificação instantâneas

Pontos fracos

- Nó raiz não é facilmente identificável
- Vários testes até se conseguir resultados satisfatórios

Cálculo da Entropia e Entropia Condicional

Equação da entropia

$$H_x = - \sum_{\forall x \in X} P(x) \log_2 P(x) \quad (1)$$

Equação da entropia condicional

$$H_{Y|X} = \sum_x P(x) H(Y|X = x) = - \sum_{\forall x \in X} P(x) \sum_{\forall y \in Y} P(y|x) \log_2 P(y|x) \quad (2)$$

Teorema de Bayes e Aprendizagem bayesiana

Teorema de Bayes: forma tradicional

$$p(C_k|x) = \frac{p(k)p(x|C_k)}{p(x)} \quad (3)$$

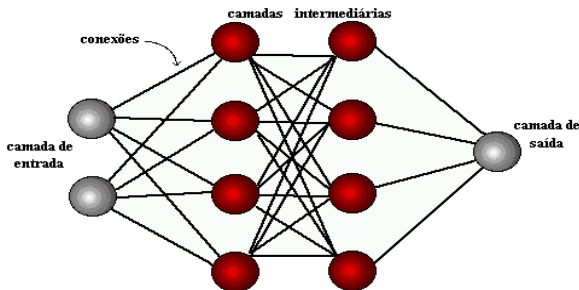
Teorema de Bayes: forma simplificada

$$p(\textit{posteriori}) = \frac{p(\textit{priori}) * \textit{verossimilhança}}{\textit{evidência}} \quad (4)$$

Classificação com Redes Neurais

Rede Neural com três camadas identificadas

- Camada de entrada: apresenta-se os padrões à rede
- Camada intermediária (ocultas): realiza a maior parte do processamento por conexões ponderadas
- Camada de saída: conclusão e apresentação do resultado.



TF – IDF

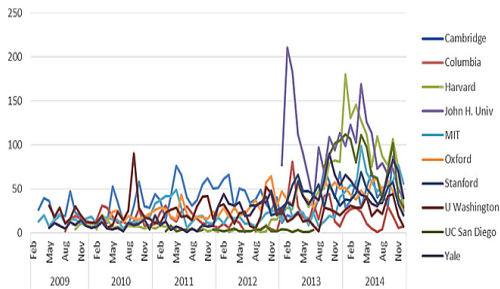
Text Frequency X Inverse Document Frequency: $TF * IDF$

$$idf(term) = \ln\left(\frac{n_{documents}}{n_{documents \text{ containing term}}}\right) \quad (5)$$

TF – IDF

Text Frequency X Inverse Document Frequency: $TF * IDF$

$$idf(term) = \ln\left(\frac{n_{documents}}{n_{documents \text{ containing term}}}\right) \quad (5)$$



AUC – Matriz Confusão

A área sob a curva AUC (Area under ROC – Receiver Operating Characteristic) é uma métrica que determina a qualidade do classificador. Ela é calculada cruzando-se verdadeiros positivos com os falso positivos. Esta área varia entre zero e um, quanto mais próximo de 1(um) o classificador conseguiu acertar mais vezes do que errar. A seguir a Matriz de Confusão que agrupa essas atributos.

Tabela: Matriz de Confusão

	Predito	
Real	TP FN	Positive – POS
Real	FP TN	Negative – NEG
—	PP PN	—

Análise de Desempenho aplicados à mineração

A Matriz Modelo de Confusão sintetiza a Matriz modelo de Confusão.

Tabela: Matriz modelo de Confusão

	Y	\bar{Y}	
X	$P(X, Y)$	$P(X, \bar{Y})$	Positive – POS
\bar{X}	$P(\bar{X}, Y)$	$P(\bar{X}, \bar{Y})$	Negative – NEG
—	$P(Y)$	$P(\bar{Y})$	—

Fonte: Bradley – 1997

Para construir a curva ROC utiliza-se as probabilidades condicionais cruzando-se a taxa de verdadeiros positivos ($tpr = P(Y|X)$) probabilidade de falsos alarmes ou taxa de falsos positivos será ($fpr = P(Y, \bar{X})$),

Sumário

1 Introdução

2 Background

3 Contribuição

- Nossa contribuição
- Modelo Proposto
- As Matrizes
- O Twitter
- Extrapolação para georreferenciamento

4 Resultados

Contribuição

Nossa contribuição:

- Do ponto de vista metodológico: Aplicação do processo CRISP-DM a uma abordagem que integre predição e classificação;

Contribuição

Nossa contribuição:

- Do ponto de vista metodológico: Aplicação do processo CRISP-DM a uma abordagem que integre predição e classificação;
- Do ponto de vista prático: Integração entre mineração de dados e mineração em textos.

Aplicação dos algoritmos de IA

- Para classificação foram empregues: Árvores de Decisão e Naïve Bayes;

Aplicação dos algoritmos de IA

- Para classificação foram empregues: Árvores de Decisão e Naïve Bayes;
- Para predição a priori utilizou-se Redes neurais.

Aplicação dos algoritmos de IA

- Para classificação foram empregues: Árvores de Decisão e Naïve Bayes;
- Para predição a priori utilizou-se Redes neurais.
- Os algoritmos escolhidos contemplam características especiais tais como: robustez, tolerância a “missing data”, aprendizagem e facilidade de interpretação.

Aplicação dos algoritmos de IA, continuação

- Propomos a integração de APIs de mapas de posicionamento global;

Aplicação dos algoritmos de IA, continuação

- Propomos a integração de APIs de mapas de posicionamento global;
- Integração de APIs de redes sociais formando um arco cibernético com essas informações melhorando a experiência do utilizador.

Aplicação dos algoritmos de IA, continuação

- Propomos a integração de APIs de mapas de posicionamento global;
- Integração de APIs de redes sociais formando um arco cibernético com essas informações melhorando a experiência do utilizador.
- Soluções conhecidas: Wase, Google Maps e outros não analisam dados históricos nem analisam outras redes sociais, também não integram tudo em uma única ferramenta.

O Modelo proposto

- A primeira fase da nossa metodologia contempla todas as fases do CRISP-DM, classificação, predição e descoberta de conhecimento nas bases históricas;

O Modelo proposto

- A primeira fase da nossa metodologia contempla todas as fases do CRISP-DM, classificação, predição e descoberta de conhecimento nas bases históricas;
- Para predição a priori utilizou-se Redes neurais;

O Modelo proposto

- A primeira fase da nossa metodologia contempla todas as fases do CRISP-DM, classificação, predição e descoberta de conhecimento nas bases históricas;
- Para predição a priori utilizou-se Redes neurais;
- Para extrapolção do modelo preditivo ocorre quando este se integra a APIs de mapas vetoriais, disparado pelo agente: utilizador.

Representação gráfica do modelo preditivo e da extrapolção

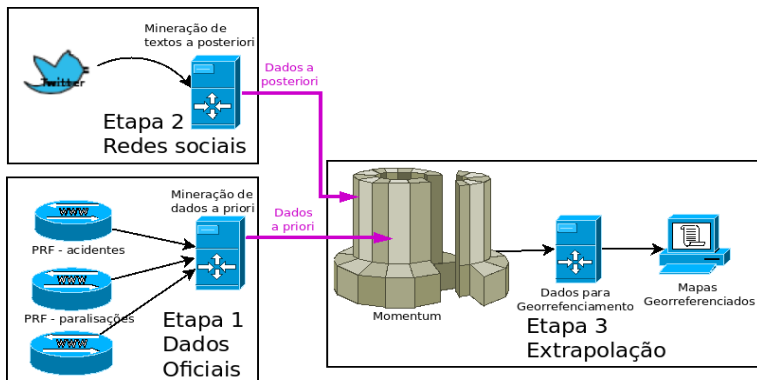


Figura: Etapas do modelo preditivo e extrapolção

Etapa 1 – Dados oficiais

A primeira etapa, *a priori*, contempla os dados da PRF. Esses dados foram tratados, variáveis foram transformadas em:

Tabela: Variáveis transformadas

KM	Numeração do quilômetro arredondada
BRajusta	Nome da BR literal
CondPista	Condição da pista: seca, molhada, ...
RestriVisibili	Restrição de visibilidade: inexistente,...
TipoAcid	Tipo de Acidente: atropelamento, colisão...
CausaAcid	A possível causa do acidente: Falta de...
TracadoVia	Tipo de traçado da via: reta, curva,...

Etapa 1 – continuação..

Tabela: Variáveis transformadas

Tipo veículo	Tipo de veículo envolvido no acidente
DiaSemana	Dia em que ocorreu o acidente
Hora	Horário do acidente/ocorrência: hh:mm
QtdMortos	Quantidade de mortos envolvidos
Gravidade	Quantidade de acidentes graves
Período	turno do dia em que se deu a ocorrência

Extração do conhecimento – KDD

- Optamos por coletar os dados diretamente dos órgãos oficiais, devido a não padronização dos dados. A PRF possui ao menos dois bancos de dados: de acidentes e de interdições.

Extração do conhecimento – KDD

- Optamos por coletar os dados diretamente dos órgãos oficiais, devido a não padronização dos dados. A PRF possui ao menos dois bancos de dados: de acidentes e de interdições.
- Após a integração dessas bases e seguidas as etapas de Mineração procedeu-se a extração do conhecimento.

Extração do conhecimento – KDD

- Optamos por coletar os dados diretamente dos órgãos oficiais, devido a não padronização dos dados. A PRF possui ao menos dois bancos de dados: de acidentes e de interdições.
- Após a integração dessas bases e seguidas as etapas de Mineração procedeu-se a extração do conhecimento.
- Para traçar um painel da diversidade das rodovias, foi feito, a priori, uma classificação com AD e comparado com NB.

Extração do conhecimento – KDD

- Optamos por coletar os dados diretamente dos órgãos oficiais, devido a não padronização dos dados. A PRF possui ao menos dois bancos de dados: de acidentes e de interdições.
- Após a integração dessas bases e seguidas as etapas de Mineração procedeu-se a extração do conhecimento.
- Para traçar um painel da diversidade das rodovias, foi feito, a priori, uma classificação com AD e comparado com NB.
- Foi escolhida a variável BRajustada porque esta continha alta de correlação linear e a entropia satisfatórias, produzindo altas taxas de classificação.

Extração do conhecimento – KDD

- Optamos por coletar os dados diretamente dos órgãos oficiais, devido a não padronização dos dados. A PRF possui ao menos dois bancos de dados: de acidentes e de interdições.
- Após a integração dessas bases e seguidas as etapas de Mineração procedeu-se a extração do conhecimento.
- Para traçar um painel da diversidade das rodovias, foi feito, a priori, uma classificação com AD e comparado com NB.
- Foi escolhida a variável BRajustada porque esta continha alta de correlação linear e a entropia satisfatórias, produzindo altas taxas de classificação.
- Por conseguinte partimos para a proposição de uma representação do conhecimento adquirido.

Matriz de Mortos

A Matriz de Mortos contém todos os óbitos registrado pela PRF, em cada trecho de cada rodovia, a cada hora do dia.

Matriz de Mortos

A Matriz de Mortos contém todos os óbitos registrado pela PRF, em cada trecho de cada rodovia, a cada hora do dia.

Hora (0 - 23)

Km (56 - 78)

	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	
0	0	0	10	14	10	1	0	0	3	4	3	9	11	19	17	19	20	4	7	6	8	9	0	8
1	0	1	2	1	0	0	0	0	0	7	8	1	3	29	11	4	3	7	0	2	1	2	2	2
2	5	3	15	4	0	8	21	2	12	1	0	1	16	20	4	0	4	8	3	0	0	0	0	8
3	0	11	0	0	0	8	0	4	0	2	1	7	6	16	16	8	0	4	2	0	4	0	0	0
4	0	8	2	8	0	0	0	0	10	10	4	5	13	9	25	4	1	2	2	2	4	6	1	1
5	0	11	6	11	0	6	3	8	7	20	34	11	25	4	32	13	11	9	15	0	2	13	10	10
6	2	17	24	26	31	13	0	23	38	26	44	50	106	30	45	45	14	38	21	39	7	52	22	22
7	8	49	38	29	39	50	5	63	51	67	106	75	182	116	119	111	64	72	75	54	34	22	64	50
8	4	25	45	42	18	48	5	41	53	90	114	73	244	109	163	109	36	88	60	24	39	51	50	50
9	9	34	37	22	24	21	7	48	55	41	87	62	203	114	100	66	67	50	79	29	14	54	53	53
10	1	40	42	27	28	13	2	37	49	33	52	64	145	66	55	80	25	45	25	10	19	23	37	37
11	2	70	27	19	8	5	4	48	60	27	56	62	156	77	71	95	26	51	29	38	30	31	76	76
12	1	29	25	6	14	7	0	15	27	26	73	122	259	64	51	66	26	25	24	12	29	9	25	25
13	0	10	21	11	19	17	2	36	14	13	62	34	157	65	66	61	25	33	21	16	26	39	56	56
14	0	24	57	4	13	13	3	45	53	41	53	63	205	59	94	82	24	29	19	19	11	26	36	36
15	7	33	37	16	14	32	2	20	33	29	84	84	213	115	67	38	46	36	34	25	33	22	44	44
16	2	16	51	16	21	11	20	43	48	35	43	79	176	63	99	54	21	22	32	17	9	54	18	18
17	0	23	22	21	9	8	5	25	32	34	67	81	191	158	110	63	31	15	69	13	12	18	35	35
18	0	42	36	34	10	14	8	44	50	29	120	98	287	133	128	85	32	54	31	19	24	11	27	27
19	4	12	23	2	5	10	1	16	21	35	69	58	194	120	76	37	26	29	6	27	0	10	12	12
20	2	12	18	20	20	1	2	9	9	14	23	37	89	24	19	21	21	23	10	12	5	11	13	13
21	4	16	7	8	3	7	2	18	9	14	16	23	52	42	43	28	8	11	15	2	16	26	9	9
22	0	8	13	3	1	1	0	6	9	13	10	27	87	22	52	17	10	9	10	5	8	8	12	12
23	0	8	5	12	3	2	4	2	2	6	13	8	73	11	46	25	9	7	6	7	7	4	1	1
somatório de acidentes / km																								
I	51	512	567	352	281	295	98	556	653	617	1141	1139	3127	1465	1504	1141	538	667	506	379	344	492	615	615

Matriz de Gravidade3D

A Matriz de Gravidade3D contempla uma maior variabilidade das ocorrência nos pontos que a Matriz de Mortos não detectou.

Matriz de Gravidade3D

A Matriz de Gravidade3D contempla uma maior variabilidade das ocorrência nos pontos que a Matriz de Mortos não detectou.

		Matriz Gravidade3D2														
HORA (0 - 23)		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	0	0	0	0	0	0	0	0	4,578	0	0	0,699	0	0	0	0
	1	1,518	0	0	1,888	0	0	0	0	0	2,06	0	0	0	0	0
	2	0	0	0	0	0	0	0	2,518	0	0	0	0	0	0	0
	3	0	0	0	0	2,518	0	0	2,518	0	2,524	0	0	0	0	0
	4	0	0	0	1,524	0	0	0	0	3,06	0	0	0	0	0	0
	5	0	0	0	0	0	0	1,416	0	0	0	5,12	0	0	2,06	0
	6	0	0	2,524	0,258	0	5,12	0	0	0	0	0	2,312	0	0	5,12
	7	0	0	3,06	3,501	1,983	3,06	2,518	3,06	4,584	0	0	1,518	3,06	0	0
	8	0	2,524	1,159	0,983	2,06	0	1,518	2,423	3,06	0	0	0	0	1,416	0
	9	0	0	0	1,524	0	3,06	0	0	0	0	0	3,578	0	0	0
	10	0	0	3,578	4,382	1,524	1,524	0	6,12	4,947	0	0	0	0	9,18	0
	11	0	0	3,06	0	2,678	0	2,06	3,06	7,156	0	0	0	3,941	2,06	0
	12	0	0	0	0	0	0	0	4,85	2,524	1,423	0	2,06	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	2,06	0	0	0	0
	14	0	0	0,152	0	0,888	3,06	0	0	7,456	2,518	0	2,06	0	2,06	0
	15	0	0	2,524	0	2,06	0	3,06	0	4,483	2,06	0	2,524	2,518	0	0
	16	0	0	0	0,983	1,524	0	0	3,06	4,94	0	1,524	0	0	2,06	0
	17	0	0	0	0	3,06	0	0	0	0	1,8	0	2,524	0	2,423	0
	18	0	0	0	0	0	0	2,06	3,06	0	1,778	0	0	3,06	1,518	0
	19	0	0	0	0	0	0	0	0	2,311	0	1,518	4,12	0	0	6,12
	20	0	0	0	0	0	2,423	0	0	2,06	0	0	0	0	0	3,06
	21	0	0	0	1,524	0	0	0	0	0	2,06	0	1,518	0	0	3,06
	22	0	0	3,06	0,983	2,518	0	0	0	3,06	0	0	0	0	0	0
	23	0	0	3,06	0,876	2,06	0	3,06	0	0	1,518	0	0	0	0	0
		Km (0 - 213)														

Matriz de Gravidade3D

Cruzando-se $\langle km, Hora \rangle$ de cada Matriz temos um ponto de ocorrência. Na Matriz de Gravidade3D esses pontos foram encontrado segundo a equação de probabilidades (BR 101):

Matriz de Gravidade3D

Cruzando-se $\langle km, Hora \rangle$ de cada Matriz temos um ponto de ocorrência. Na Matriz de Gravidade3D esses pontos foram encontrado segundo a equação de probabilidades (BR 101):

$$ProbAcid_{101} = (RestVisibi + CondPista + TracadoVia) * Erro_{101} + Gravidade \quad (6)$$

Matriz de Gravidade3D

Cruzando-se $\langle km, Hora \rangle$ de cada Matriz temos um ponto de ocorrência. Na Matriz de Gravidade3D esses pontos foram encontrado segundo a equação de probabilidades (BR 101):

$$ProbAcid_{101} = (RestVisibi + CondPista + TracadoVia) * Erro_{101} + Gravidade \quad (6)$$

Onde:

- RestVisibi - Restrição de Visibilidade;
- CondPista - Condição da Pista;
- TraçadoVia - Traçado da Via;
- Erro_{101} - Taxa de Erro da AD para a BR 101
- Gravidade - Ocorrência de óbito (variável binarizada)

Arco cibernético com dados do Twitter

Os dados do Twitter permitem buscar informações no momento em que o utilizador iniciar sua viagem.

Arco cibernético com dados do Twitter

Os dados do Twitter permitem buscar informações no momento em que o utilizador iniciar sua viagem. Verificamos que previsões como protestos, queda/remoção de barreiras e rochas são sazonais em Pernambuco, portanto não estão contemplados no modelo de previsão.

Arco cibernético com dados do Twitter

Os dados do Twitter permitem buscar informações no momento em que o utilizador iniciar sua viagem. Verificamos que previsões como protestos, queda/remoção de barreiras e rochas são sazonais em Pernambuco, portanto não estão contemplados no modelo de previsão.

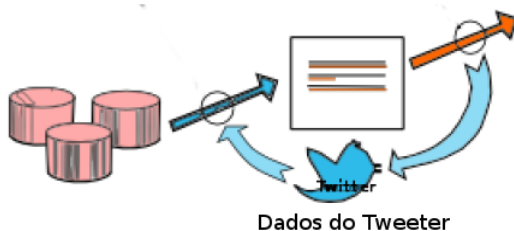


Figura: Arco cibernético com dados do Twitter

Localizando os pontos críticos

As matrizes apresentadas (Mortos e Gravidade) representam pontos críticos nas rodovias. Esses pontos podem ser extrapolados em coordenadas geográficas para mapas de georreferenciamento.

Localizando os pontos críticos

As matrizes apresentadas (Mortos e Gravidade) representam pontos críticos nas rodovias. Esses pontos podem ser extrapolados em coordenadas geográficas para mapas de georreferenciamento.

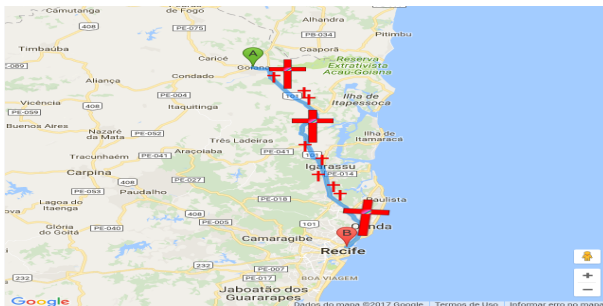


Figura: Etapas 3 – Georreferenciamento

Sumário

- 1 Introdução
- 2 Background
- 3 Contribuição
- 4 Resultados**
 - Execução do Modelo
 - Análise das Rodovias – Antes da MD
 - Mineração em Textos no Twitter
- 5 Conclusões

Visão geral do Modelo Proposto

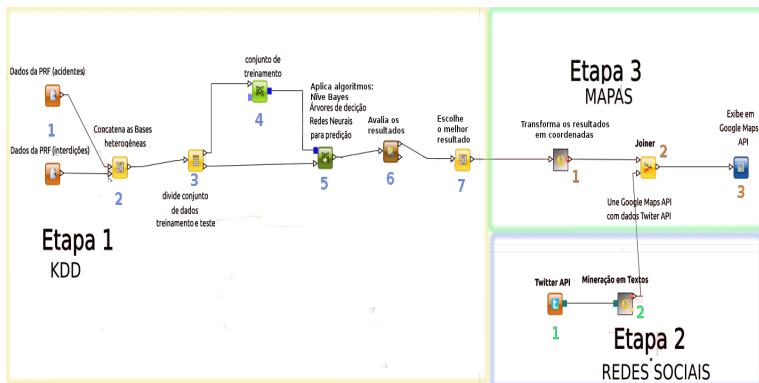


Figura: Etapas do modelo proposto

A Etapa 1 – KDD

O KDD contempla a fase de coleta da bases históricas, preparação e execução dos algoritmos e descoberta dos pontos críticos descrito a seguir:

A Etapa 1 – KDD

O KDD contempla a fase de coleta da bases históricas, preparação e execução dos algoritmos e descoberta dos pontos críticos descrito a seguir:

- 1 O modelo preditivo integra as bases de dados da PRF;

A Etapa 1 – KDD

O KDD contempla a fase de coleta da bases históricas, preparação e execução dos algoritmos e descoberta dos pontos críticos descrito a seguir:

- 1 O modelo preditivo integra as bases de dados da PRF;
- 2 As bases são concatenados de acordo com a lógica do negócio;

A Etapa 1 – KDD

O KDD contempla a fase de coleta da bases históricas, preparação e execução dos algoritmos e descoberta dos pontos críticos descrito a seguir:

- 1 O modelo preditivo integra as bases de dados da PRF;
- 2 As bases são concatenados de acordo com a lógica do negócio;
- 3 Tratamento e divisão dos dados em: conj. de treinamento e testes;

A Etapa 1 – KDD

O KDD contempla a fase de coleta da bases históricas, preparação e execução dos algoritmos e descoberta dos pontos críticos descrito a seguir:

- 1 O modelo preditivo integra as bases de dados da PRF;
- 2 As bases são concatenados de acordo com a lógica do negócio;
- 3 Tratamento e divisão dos dados em: conj. de treinamento e testes;
- 4 Treina e Executa os algoritmos de IA escolhidos;

A Etapa 1 – KDD

O KDD contempla a fase de coleta da bases históricas, preparação e execução dos algoritmos e descoberta dos pontos críticos descrito a seguir:

- 1 O modelo preditivo integra as bases de dados da PRF;
- 2 As bases são concatenados de acordo com a lógica do negócio;
- 3 Tratamento e divisão dos dados em: conj. de treinamento e testes;
- 4 Treina e Executa os algoritmos de IA escolhidos;
- 5 Caso os algoritmos não tenham produzidos os resultados esperados ocorre extração de conhecimento.

A Etapa 1 – KDD

O KDD contempla a fase de coleta da bases históricas, preparação e execução dos algoritmos e descoberta dos pontos críticos descrito a seguir:

- 1 O modelo preditivo integra as bases de dados da PRF;
- 2 As bases são concatenados de acordo com a lógica do negócio;
- 3 Tratamento e divisão dos dados em: conj. de treinamento e testes;
- 4 Treina e Executa os algoritmos de IA escolhidos;
- 5 Caso os algoritmos não tenham produzidos os resultados esperados ocorre extração de conhecimento.
- 6 Análise das métricas: AUC e Matriz de Confusão.

A Etapa 1 – KDD

O KDD contempla a fase de coleta da bases históricas, preparação e execução dos algoritmos e descoberta dos pontos críticos descrito a seguir:

- 1 O modelo preditivo integra as bases de dados da PRF;
- 2 As bases são concatenados de acordo com a lógica do negócio;
- 3 Tratamento e divisão dos dados em: conj. de treinamento e testes;
- 4 Treina e Executa os algoritmos de IA escolhidos;
- 5 Caso os algoritmos não tenham produzidos os resultados esperados ocorre extração de conhecimento.
- 6 Análise das métricas: AUC e Matriz de Confusão.
- 7 Escolha dos melhores: AUC, Inst. Classif. corretamente, Inst. classif. incorretamente e Erro médio absoluto.

A Etapa 2 – Redes Sociais

Esta etapa é um modulo dinâmico e analisa momentaneamente os feeds postados pelos usuários das redes sociais;

A Etapa 2 – Redes Sociais

Esta etapa é um modulo dinâmico e analisa momentaneamente os feeds postados pelos usuários das redes sociais;

- 1 São capturados “feeds” da rede social Twitter (para esta pesquisa). Essa técnica faz um arco cibernético mantendo o utilizador atualizado com as informações recentes.

A Etapa 2 – Redes Sociais

Esta etapa é um modulo dinâmico e analisa momentaneamente os feeds postados pelos usuários das redes sociais;

- 1 São capturados “feeds” da rede social Twitter (para esta pesquisa). Essa técnica faz um arco cibernético mantendo o utilizador atualizado com as informações recentes.
- 2 Após a captação dos dados do Twitter é feita a Mineração nos textos para localizar informações que permitam antever alguma paralização futura nas rodovias.

Etapa 3 – Integração de APIs

- 1 Conversão dos melhores resultados encontrados na Etapa 1
< *Km*, *Hora* > em coordenadas;

Etapa 3 – Integração de APIs

- 1 Conversão dos melhores resultados encontrados na Etapa 1 $\langle Km, Hora \rangle$ em coordenadas;
- 2 Localizações geográficas dos pontos críticos da malha viária, indicadas pelo Km, são agrupadas formando "clusters" de dados exibidos em mapas vetoriais. Une API do Google Mapas e API do Twitter;

Etapa 3 – Integração de APIs

- 1 Conversão dos melhores resultados encontrados na Etapa 1 $\langle Km, Hora \rangle$ em coordenadas;
- 2 Localizações geográficas dos pontos críticos da malha viária, indicadas pelo Km, são agrupadas formando "clusters" de dados exibidos em mapas vetoriais. Une API do Google Mapas e API do Twitter;
- 3 Exibe informações vindas do twitter sobre ocorrências na rodovia, encontrando sua geolocalização a ser transformado em marcos "milestone" para representação sobre mapas de bases vetoriais.

Dados encontrados para a BR 101

O dados revelaram que a grande maioria dos acidentes ocorre com pista seca, sem restrição de visibilidade e em linha reta. A seguir gráficos da BR 101, em vermelho são os pontos críticos, em azul com menos frequência de acidentes.

Dados encontrados para a BR 101

O dados revelaram que a grande maioria dos acidentes ocorre com pista seca, sem restrição de visibilidade e em linha reta. A seguir gráficos da BR 101, em vermelho são os pontos críticos, em azul com menos frequência de acidentes.

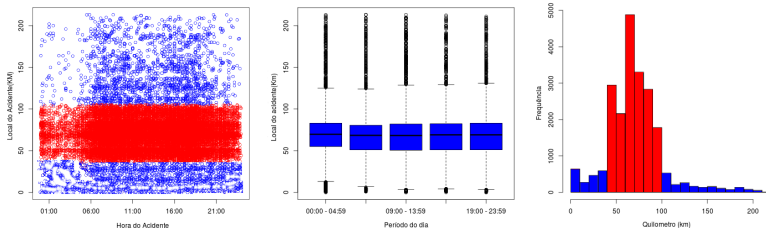


Figura: BR 101: Hora do acidente (1) Concentração em torno da hora (2) e Frequência (3)

Dados encontrados para a BR 104

Conhecida como rota da sulanca esta rodovia apresenta concentração de acidentes entre o km 60 e 70

Dados encontrados para a BR 104

Conhecida como rota da sulanca esta rodovia apresenta concentração de acidentes entre o km 60 e 70

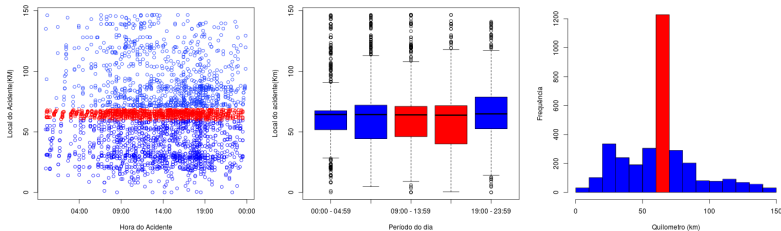


Figura: BR 104: Hora do acidente (1) Concentração em torno da hora (2) e Frequência (3)

Dados encontrados após a MD

Resultados dos Classificadores: As variáveis “TipoAcidente”, “Gravidade” e “BRajustada” obtiveram a melhor relação ganho de informação. A métrica para avaliar os resultados dos classificadores foram:

Dados encontrados após a MD

Resultados dos Classificadores: As variáveis “TipoAcidente”, “Gravidade” e “BRajustada” obtiveram a melhor relação ganho de informação. A métrica para avaliar os resultados dos classificadores foram:

- TP: True Positive;
- FP: False Positive;
- Prec.: $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$;
- Recall = $\text{TP} / (\text{TP} + \text{FN})$;
- F-Me: F-measure ou f-score = $2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$;
- AUC: Area Under Curve (Roc);

Comparando as métricas para a variável BRajustada

Resultado obtido pela AD:

Descrição	Valores	Percentual
Instâncias Corretamente Classificadas	13507	80.5522%
Instâncias Incorretamente Classificadas	3261	19.4478%
Erro médio quadratico	0.1656	—

Comparando as métricas para a variável BRajustada

Resultado obtido pela AD:

Descrição	Valores	Percentual
Instâncias Corretamente Classificadas	13507	80.5522%
Instâncias Incorretamente Classificadas	3261	19.4478%
Erro médio quadrático	0.1656	—

Resultado obtido pelo NB:

Descrição	Valores	Percentual
Instâncias Corretamente Classificadas	9232	73,3339%
Instâncias Incorretamente Classificadas	3357	26,6661%
Erro médio quadrático	0.1908	—

Matriz de Confusão para a variável BRajustada encontrada pela AD

a	b	c	d	e	f	g	h	Classificadas
6960	0	0	625	0	0	0	0	BR101
0	1071	0	156	0	0	0	0	BR104
0	0	0	625	0	0	26	11	BR110
0	0	85	0	90	11	0	0	BR116
970	9	0	3185	1	0	1	0	BR232
0	0	27	11	377	7	0	0	BR316
0	0	0	0	0	95	0	0	BR407
643	0	0	66	0	0	0	0	BR408

Matriz de Confusão para a variável BRajustada encontrada pelo NB

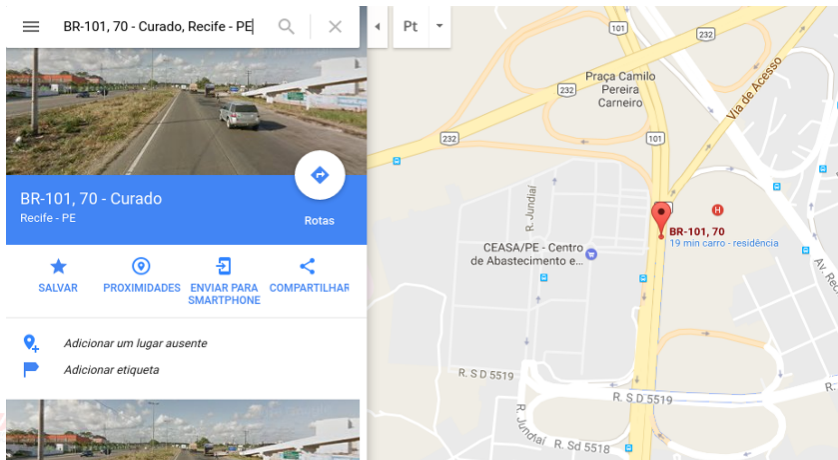
a	b	c	d	e	f	g	h	Classificadas
5317	0	0	0	198	0	0	102	a = BR101
0	761	0	0	88	0	0	0	b = BR104
0	0	7	0	12	0	0	0	c = BR110
0	0	0	130	1	24	0	0	d = BR116
1605	69	0	0	1424	159	59	0	e = BR232
0	0	0	94	47	206	0	0	f = BR316
0	0	0	0	1	1	346	0	g = BR407
323	0	0	0	1	0	0	64	h = BR408

Alguns resultados encontrados pela AD

“Tipo do Acidente” foram: (a) “Tipo de Acidente” [região metropolitana]: [Atropelamento de pessoa], [pista seca], [período: noite], [Br < 116 (101, 104)] , [Dia da semana: terça-feira]: [Gravidade = N (sem morte)], [Km <= 69] => falta de atenção. [Gravidade = S (com morte)] => outras.

Tipo de Acidente: [Atropelamento de pessoa], [pista seca], [período: noite], [Br < 116 (101, 104)] , [Dia da semana: sexta-feira]: [Gravidade = N (sem morte)], [Km != 58] => falta de atenção. [Gravidade = S (com morte)] => [Km > 58] [Km <= 67] => falta de atenção.

Trecho na BR 101 descrito pela AD



Sugestão de Roteamento: Goiana - Recife, a partir da Matriz de Gravidade 3D para 2ª-feira

km 0 a 20

1ª Hora		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	SOMA-1
	5	0	0	0	0	0	0	1,416	0	0	0	5,12	0	0	2,06	0	0	0	2,06	0	0	0	10,656
	6	0	0	2,524	0,258	0	5,12	0	0	0	0	0	2,312	0	0	5,12	0	0	0	0	3,06	0	18,394
	7	0	0	3,06	3,501	1,983	3,06	2,518	3,06	4,584	0	0	1,518	3,06	0	0	1,518	0	0	2,06	0	0	29,922

km 21 a 40

2ª Hora		21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	SOMA-2
	6	0	0	1,423	0	0	0	0	0	0	1,518	0	0	0	0	0	0	0	1,983	0	2,518	7,442
	7	0	0	0	0	0	0	0	0	0	0	1,524	0	0	0	0	0	0	0	4,12	0	5,644
	8	2,518	0	5,12	3,94	0	0	2,524	0	0	0	4,12	0	3,06	1,416	0	2,518	6,001	0	3,097	3,06	37,374

km 41 a 60

3ª Hora		41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	SOMA-3
	7	3,06	4,584	4,12	1,431	2,06	0	5,12	2,518	2,06	14,764	2,06	8,644	5,578	6,638	10,698	6,096	10,698	4,85	14,53	10,216	119,725
	8	2,423	5,584	7,156	4,12	8,18	9,18	5,036	8,18	7,007	25,462	9,18	2,431	8,18	3,043	0	8,698	23,752	10,434	12,572	15,276	175,894
	9	11,698	3,06	0	2,06	3,06	3,06	5,12	0	5,638	20,791	10,887	1,796	2,518	2,06	3,06	6,644	7,434	5,578	16,156	9,627	120,247

Aplicação da equação 3.1 para roteamentos

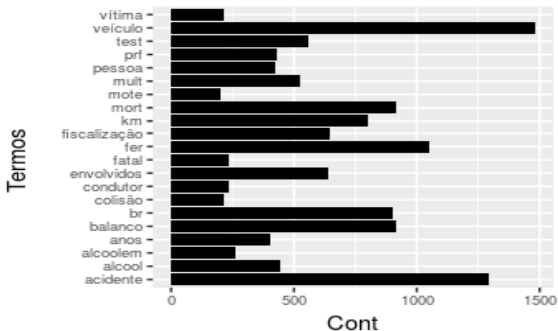
Três veículos partem do km 0 (zero) da BR 101 em horários diferente, segunda-feira, com destino ao Recife. O Veículo-x parte às 5h da manhã, o Veículo-y às 6h e o Veículo-z às 7h, eles percorrem com uma velocidade média de 20km/h. Resultados encontrado para os três veículos na mesma rota:

- Veículo-x: $10,656 + 7,442 + 119,725 = 137,823$
- Veículo-y: $18,394 + 5,644 + 175,894 = 199,932$
- Veículo-z: $29,922 + 37,374 + 120,247 = 187.543$

Segundo os resultados acima, o Veículo-x chegara primeiro ao seu destino, seguido do Veículo-z e por último o Veículo-y.

Mineração de dados textuais

Minerar dados do Twitter demonstrou ser viável. O utilizador faz referências ao que acontece nas rodovias, podendo impactar em sua experiência ao utilizá-la. A figura a seguir são os termo mais frequentes capturados no canal @PRF191PE em um dia normal.



Mineração de dados textuais

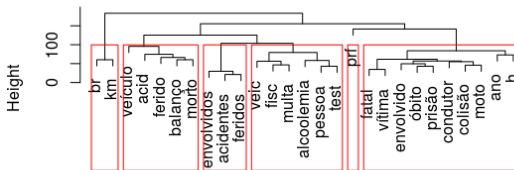
O gráfico tipo Nuvem de palavras é outra maneira de se apresentar as frequências de unigramas encontrados em outro dia comum no canal @PRF191PE.



Mineração de dados textuais

O dendograma é um gráfico que agrupa palavras de acordo com o assunto. No dendograma a seguir destacamos os agrupamentos sobre o assunto **colisão** com vítima fatal com um dos condutores estava em uma moto e houve prisão.

Cluster Dendrogram

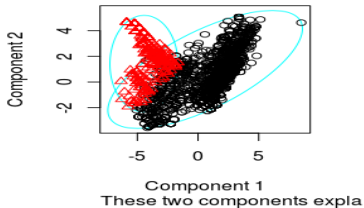


distMatrix
hclust (*, "ward.D")

Mineração de dados textuais

O segundo gráfico de agrupamento demonstra a utilização do algoritmo K-means. Uma análise mais minuciosa desses gráficos podem trazer consideráveis contribuições ao modelo de predição da Etapa 2 da figura 4.1.

$\kappa = \text{sdata}, k = k, \text{diss} = \text{diss}, \text{meti}$

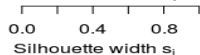


Silhouette plot of pam(

$n = 2873$ 2 clusters C_j
 $j : n_j \mid \text{ave}_{i \in C_j} s_i$

1 : 2514 | 0.48

2 : 359 | 0.43



Average silhouette width : 0.4

Sumário

- 1 Introdução
- 2 Background
- 3 Contribuição
- 4 Resultados
- 5 Conclusões**
- 6 Extras

Considerações finais – 1

O realização desta pesquisa sugere que é possível propor um modelo de predição que possibilite a gestão logística do ponto de vista do utilizador, em dias e horários mais convenientes, melhorando a experiência dos atores que fazem uso dessas rodovias. Resultados encontrados em outras pesquisas apontam a eficácia da utilização de I.A. para analisar e empregar soluções ao tráfego de veículos.

Considerações finais – 2

O dados encontrados sugerem que há um padrão de comportamento nas rodovias que pode ser analisado e predito pela I.A., de maneira a facilitar o tráfego de veículos e tornar as rodovias mais seguras. O registro pelo órgãos de trânsito tende a aumentar, oferecendo uma gama de opções para as pesquisas nessa área. Sentimos necessidade de informações de mais precisas sobre latitude e longitude, elas aparecem precariamente nos dados da PRF. Esta informação é de suma importância para pesquisadores e desenvolvedores que utilizam georreferenciamento. Por exemplo a BR 101, atravessa o país de norte a sul, em muitos trechos, em outros Estados o Km se repete.

Considerações finais – 3

Uma das contribuições desta pesquisa é de cunho metodológico-prático. Destaca-se inicialmente pela articulação entre os resultados envolvendo diferentes algoritmos. A pesquisa utilizou Naïve Bayes, TF-IDF, Árvores de Decisão, Redes Neurais para o trabalho com os dados da PRF e do Twitter. Observa-se que isso parece ser uma tendência nos trabalhos que envolvem análise do tráfego em rodovias. Outra contribuição dessa pesquisa é a de unir predição e classificação, dados históricos a priori e dados de redes sociais a posteriori.

Considerações finais – 4

Esta pesquisa também contribuiu para a compreensão das causas dos constantes constrangimentos e incidentes que paralisam as rodovias pernambucanas. Os dados revelam que o condutor é o principal causador e vítima dos danos registrados pela PRF e que a imprudência e desrespeito às leis são os maiores responsáveis pela grande quantidade de óbitos encontrados pois a esmagadora maioria dos acidentes ocorre em via reta, sem restrição de visibilidade com pista seca e em boas condições.

Em análise restrita a morte por atropelamento, os dados sugerem concluir que não há uma estrutura segura que evite esse tipo de acidentes nem obrigue o pedestre a utilizar as passarelas de forma segura e eficiente.

Considerações finais – 5

Outras condicionantes a acidentes nos perímetros urbanos, por exemplo, são falta de sinalização, passarelas, faixa de pedestres, barreiras de segurança, limitador de velocidades, dentre outros. A quantidade de acidentes que compreende o trecho: km 66 ao km 71 da BR 101, com mais de 9 500 mortos por atropelamento nos últimos nove anos, aponta como o local mais perigoso para um pedestre atravessar no Estado de Pernambuco.

O padrão de comportamento encontrado nesta pesquisa também foi encontrado em outras, neste país e em outros, já discutidas nesta dissertação, sugerindo que as variáveis relacionadas ao condutor são as mais fortes condicionantes de acidentes.

Considerações finais – 6

Ainda com relação à contribuição ao Estado da Arte descrito nesta Dissertação, esta pesquisa avança em relação ao que foi proposto em outros trabalhos ao identificar as ocorrências nas rodovias levando em conta o passado e o presente propondo um modelo que contemple o futuro possibilitando escolhas mais seguras e assertivas.

Sumário

- 1 Introdução
- 2 Background
- 3 Contribuição
- 4 Resultados
- 5 Conclusões
- 6 Extras

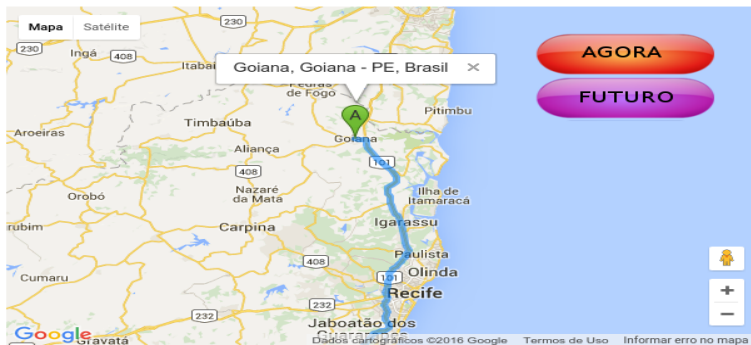
• Trabalhos futuros

Sistema de suporte a decisão

Trabalhos futuros sugerem ampliar as pesquisas a fim de propor o desenvolvimento de uma ferramenta que atenda a um gestor tomar decisões sobre a logística do transporte de cargas, utilizando os resultados encontrados, melhoramento modelo preditivo, incorporando novas técnicas para obter resultados melhores. Ampliar a Equação 3.1 incorporando variáveis que permita com que uma equação do Momentum seja eficaz incorporando situações espaço-temporais ao integrar passado e futuro.

Aplicativo

A fim de ampliar a utilização dessa pesquisa propomos aprimorar a etapa de extrapolação que culmine em um aplicativo para celulares. Uma breve sugestão de API é a figura a seguir.



Artigo aprovado em evento internacional

ANÁLISE DE COMPORTAMENTO RODOVIÁRIO NA REGIÃO METROPOLITANA DO RECIFE-BRASIL – UMA ABORDAGEM DE MINERAÇÃO DE DADOS

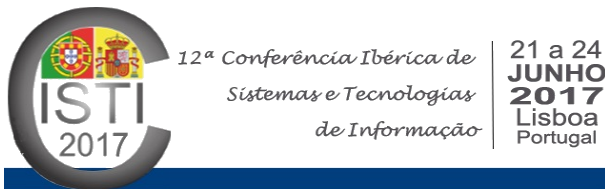


Figura: CISTI – Conf. Ibérica de Sistemas e Tecnologia da Informação