Defesa de Mestrado

Othon Luiz Teixeira de Oliveira

Escola Politécnica de Pernambuco – Poli — UPE

29 - Maio - 2017

PPGES


UPE
UNIVERSIDADE
DE PERNAMBUCO


POLI
ESCOLA POLITÉCNICA
DE PERNAMBUCO


ICB
INSTITUTO
DE CIÊNCIAS
BIOLÓGICAS


UPE
Pós-graduação
em Engenharia de Sistemas

DISSERTAÇÃO DE MESTRADO

DISSERTAÇÃO DE MESTRADO

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

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

Organização

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.

Organização

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

Organização

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	78	-7,56206206749...	-34,990451894499...	Seca	Inexistente	Colisão late...	Falta de atenç...	De
Row1	101	210	-7,66697808649...	-34,933297438999...	Seca	Inexistente	Colisão Tra...	Outras	De
Row2	101	416	-7,83434453699...	-34,912533977999...	Seca	Inexistente	Colisão Tra...	Falta de atenç...	Cre
Row3	101	416	-7,83434453699...	-34,912533977999...	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...	-34904202790499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre
Row9	101	545	-7,94363545599...	-34904202790499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre
Row10	101	589	-797661785749...	-34925137692499...	Seca	Inexistente	Colisão late...	Velocidade inc...	De
Row11	101	680	-887953132499...	-35628922247499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre
Row12	101	680	-887953132499...	-35628922247499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre
Row13	101	680	-887953132499...	-35628922247499...	Seca	Inexistente	Colisão late...	Ultrapassage...	Cre
Row14	101	680	-887953132499...	-35628922247499...	Seca	Inexistente	Colisão late...	Ultrapassage...	Cre
Row15	101	680	-887953132499...	-35628922247499...	Seca	Inexistente	Colisão tras...	Falta de atenç...	Cre
Row16	101	680	-887953132499...	-35628922247499...	Seca	Inexistente	Colisão tras...	Falta de atenç...	Cre
Row17	101	680	-887953132499...	-35628922247499...	Seca	Inexistente	Queda de m...	Outras	De
Row18	101	680	-887953132499...	-35628922247499...	Seca	Inexistente	Queda de m...	Velocidade inc...	Cre
Row19	101	830	-817277505899...	-34939988807999...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De
Row20	101	830	-817277505899...	-34939988807999...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De
Row21	101	830	-817277505899...	-34939988807999...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De
Row22	101	830	-817277505899...	-34939988807999...	Seca	Inexistente	Queda de m...	Outras	De
Row23	101	865	-818644581449...	-34971397555499...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De
Row24	101	1018	-826668917249...	-35041692694499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre
Row25	101	1018	-826668917249...	-35041692694499...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre
Row26	101	1020	-828279581549...	-35057672119999...	Com bura...	Inexistente	Colisão late...	Outras	Cre
Row27	101	1020	-828279581549...	-35057672119999...	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

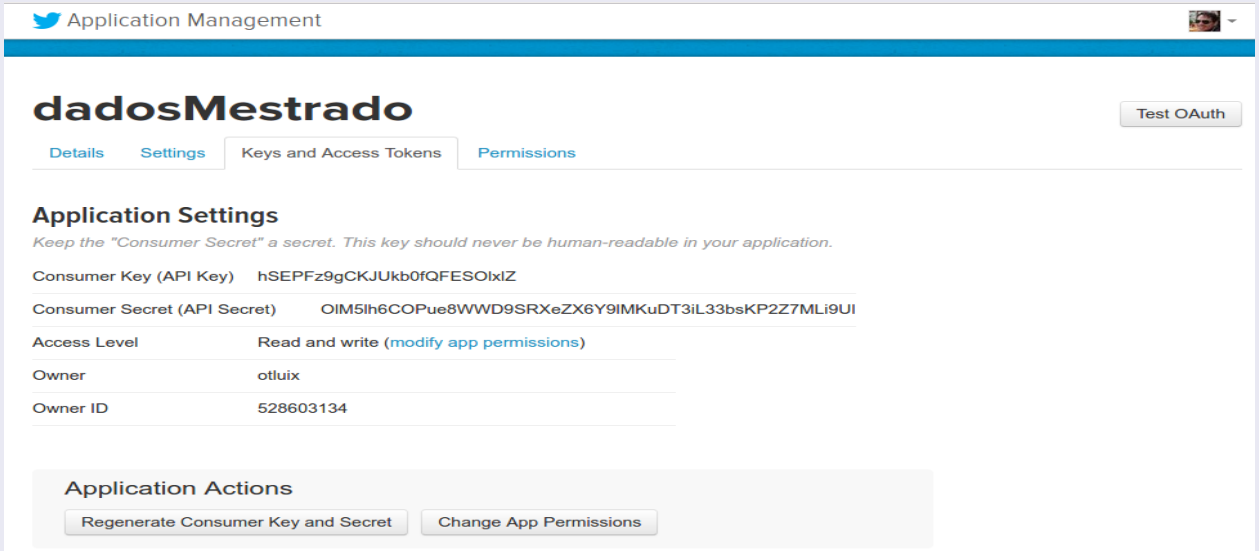
Atributos e Instâncias – iniciais

- 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 Tweeter

Acesso aos dados do Tweeter

Figura: Registro de App para para acessar e baixar dados



Dados originais

Dados – Planilha Tweeter

Figura: Planilha para Preprocessamento

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	text	favorited	favori	reply	created	trunc	replyT	id	repl	statusSource	screenNam	retweetC	isRetw	retwee	lon	latitude
2	00 MORT 175 TEST DE ALCOLE 01 PRISAO DE ALCOOLEMIA 722 PESS FISC 631 VEIC FISC	FALSE	8	NA	04/03/17 09:20	FALS	NA	83795601	NA	<a href="http://twitte	PRF191PE	2	FALSE	FALSE	NA	NA
3	https://t.co/0xHrbTnzGI	FALSE	5	NA	04/03/17 03:36	FALS	NA	83786925	NA	<a href="http://twitte	PRF191PE	1	FALSE	FALSE	NA	NA
4	BALANCO DIA 02/03/17: 09ACID; 12VEIC; 12FERIDOS; 00MOR	FALSE	6	NA	03/03/17 03:41	FALS	NA	83750809	NA	<a href="http://twitte	PRF191PE	0	FALSE	FALSE	NA	NA
5	PRF registra redução de 73% no número de mortes durante Ope Mais de 1800 imagens por excesso... https://t.co/z7vjDc63aT	FALSE	5	NA	02/03/17 14:29	FALS	NA	83730901	NA	<a href="http://www.t	PRF191PE	0	FALSE	FALSE	NA	NA
6	PROTESTO: Em Agrestina/PE, na BR 104, km 88, ocorre uma i	FALSE	12	NA	25/02/17 18:35	FALS	NA	83555880	NA	<a href="http://twitte	PRF191PE	3	FALSE	FALSE	NA	NA
7	Cinco suspeitos de furto são detidos pela PRF na BR 101 Quatro homens e uma mulher suspeitos de integrarem uma... htt	FALSE	2	NA	25/02/17 17:32	FALS	NA	83554306	NA	<a href="http://www.t	PRF191PE	2	FALSE	FALSE	NA	NA
8	https://t.co/jHHVsDI32O	FALSE	2	NA	25/02/17 16:57	FALS	NA	83553417	NA	<a href="http://www.t	PRF191PE	0	FALSE	FALSE	NA	NA
9	https://t.co/yfDjxuF1ZY	FALSE	32	NA	25/02/17 02:44	FALS	NA	83531957	NA	<a href="http://twitte	PRF191PE	15	FALSE	FALSE	NA	NA
10	Caminhão frigorífico é recuperado logo após roubo na BR 232 Um caminhão frigorífico carregado com carnes foi... https://t.co/l	FALSE	8	NA	24/02/17 17:35	FALS	NA	83518149	NA	<a href="http://www.t	PRF191PE	0	FALSE	FALSE	NA	NA
11	https://t.co/1T1kXzsZtJ	FALSE	2	NA	24/02/17 17:33	FALS	NA	83518084	NA	<a href="http://www.t	PRF191PE	0	FALSE	FALSE	NA	NA
12	https://t.co/OpOGQKoMkL	FALSE	10	NA	24/02/17 04:55	FALS	NA	83499021	NA	<a href="http://twitte	PRF191PE	2	FALSE	FALSE	NA	NA
13	https://t.co/2yHi2zrQEj	FALSE	9	NA	23/02/17 03:48	FALS	NA	83461081	NA	<a href="http://twitte	PRF191PE	0	FALSE	FALSE	NA	NA
14	ACID. FERIDOS:colisão frontal entre 2 micro-ônibus em Ponte d	FALSE	7	NA	22/02/17 13:37	FALS	NA	83439670	NA	<a href="http://twitte	PRF191PE	0	FALSE	FALSE	NA	NA
15	https://t.co/w7aQ1Hrdjf	FALSE	9	NA	22/02/17 08:28	FALS	NA	83431901	NA	<a href="http://twitte	PRF191PE	1	FALSE	FALSE	NA	NA
16	Av. Recife bloqueada nas proximidades da empresa Brinks,nº 49	FALSE	23	NA	21/02/17 09:37	FALS	NA	83397390	NA	<a href="http://twitte	PRF191PE	21	FALSE	FALSE	NA	NA
17	ACESSO À AV. RECIFE PARA QUEM VEM PELA BR-101 EST	FALSE	85	NA	21/02/17 08:34	FALS	NA	83395803	NA	<a href="http://twitte	PRF191PE	69	FALSE	FALSE	NA	NA
18	https://t.co/cYmV24bQZY	FALSE	8	NA	21/02/17 07:50	FALS	NA	83394688	NA	<a href="http://twitte	PRF191PE	1	FALSE	FALSE	NA	NA
	Após tentar fugir da PRF, suspeito de assalto é preso em Caruar															

Entendendo a base de dados do Tweeter

Atributos e Instâncias – iniciais

- Dados de 2017 – 2014 — canal @PRF191PE

Entendendo a base de dados do Tweeter

Atributos e Instâncias – iniciais

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

Entendendo a base de dados do Tweeter

Atributos e Instâncias – iniciais

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

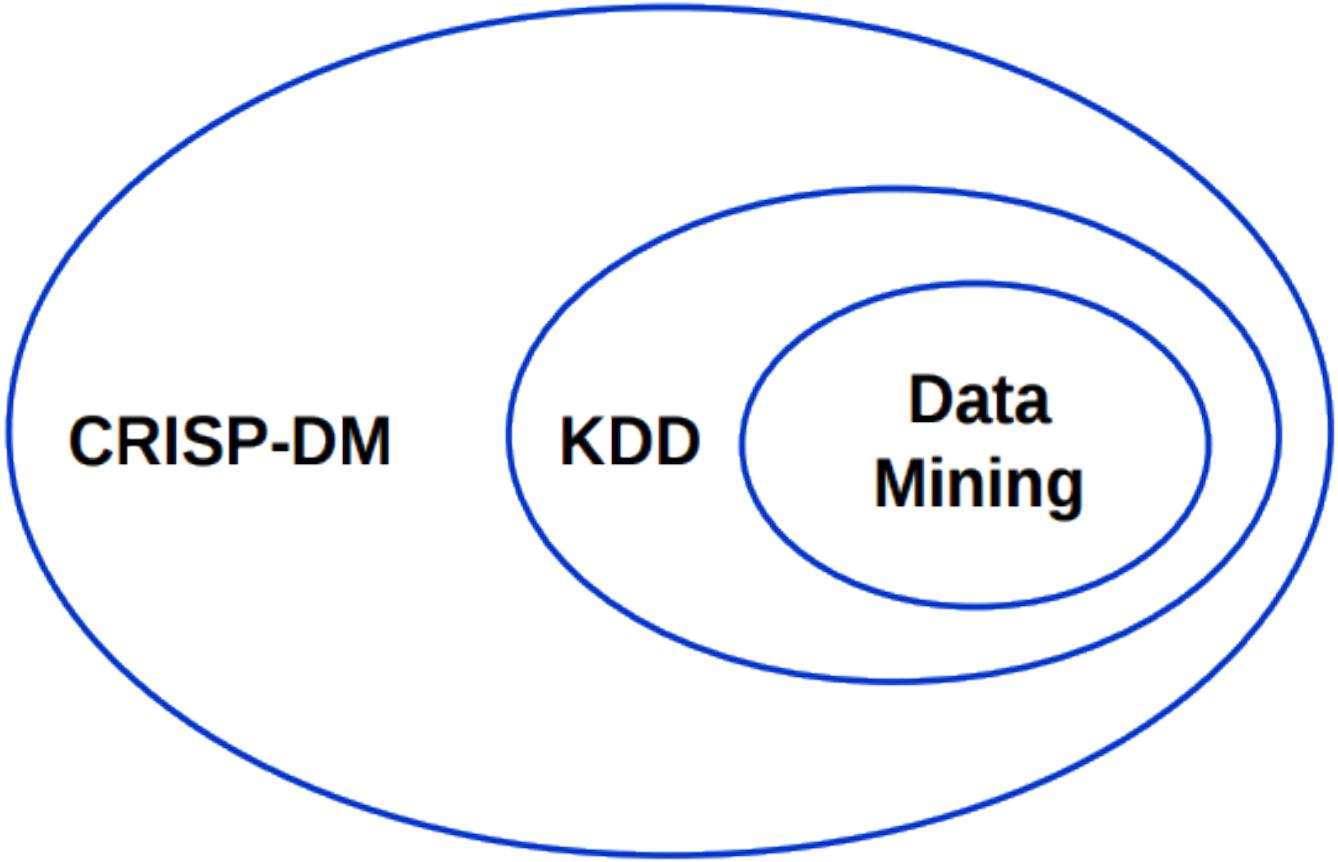
Entendendo a base de dados do Tweeter

Atributos e Instâncias – iniciais

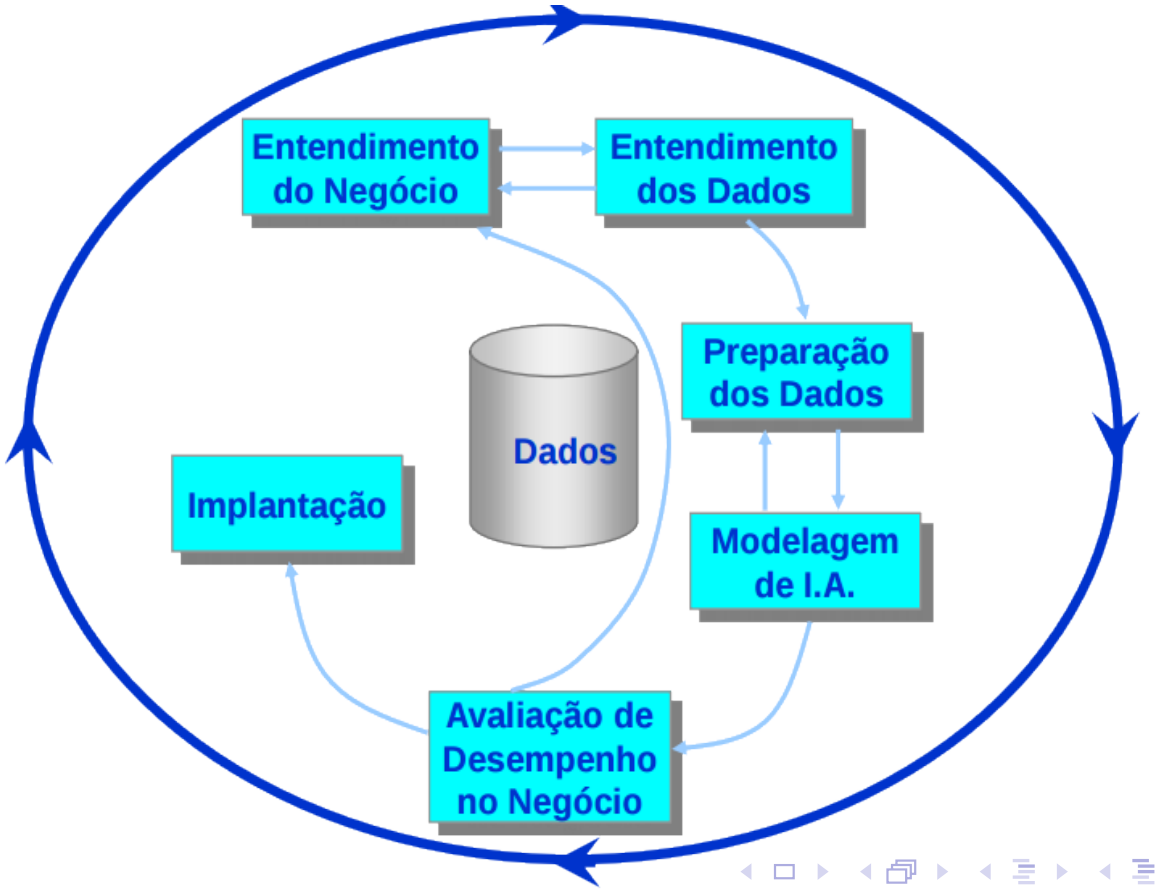
- Dados de 2017 – 2014 — canal @PRF191PE
- 2864 Instâncias – 16 Atributos
- Dentre eles: 'text, 'favorited', 'favoriteConunt', 'created', 'ID', 'statusSource', 'screenName', 'retweetCount',
- 'isRetweet', 'retweeted', dentre outros.

Organização

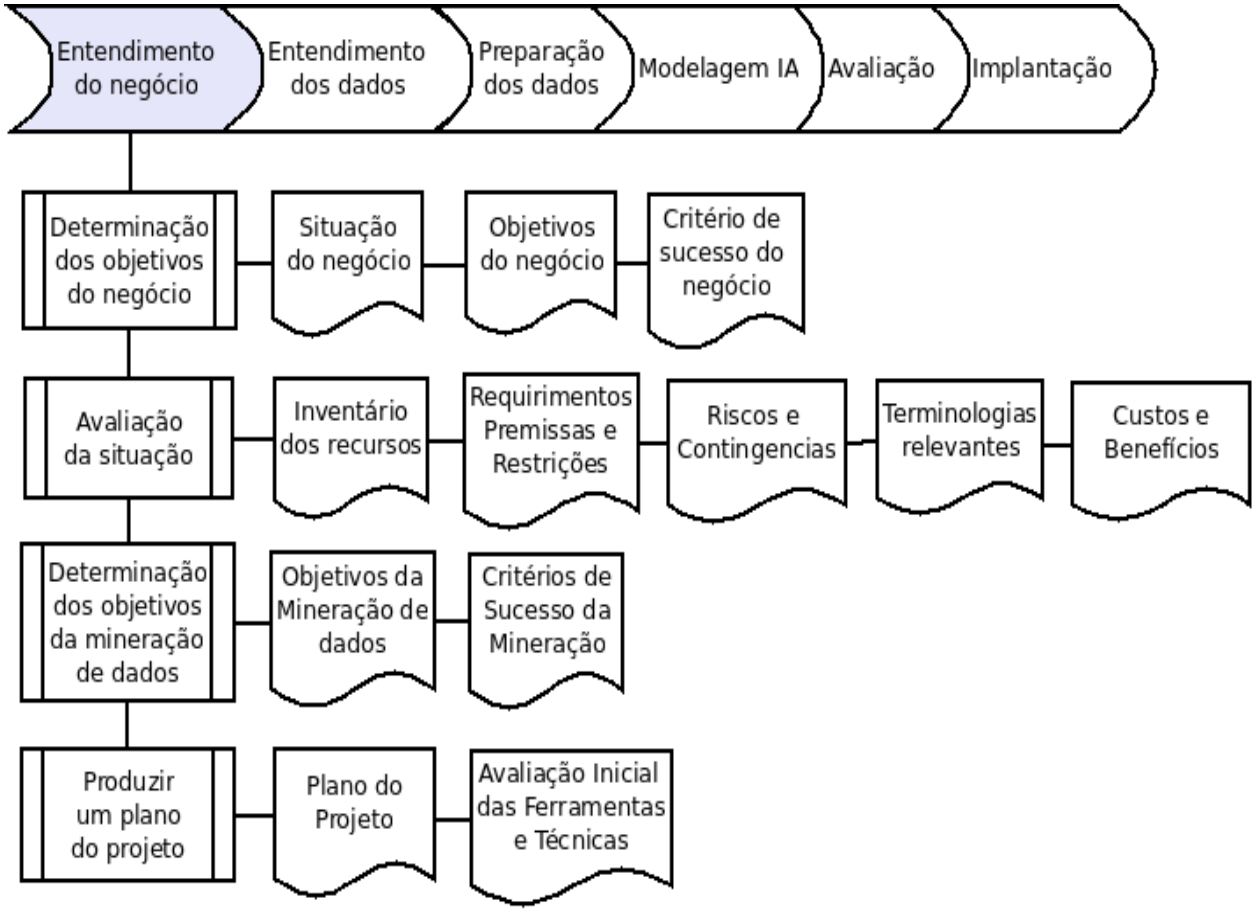
Domínio das técnicas de Mineração de Dados



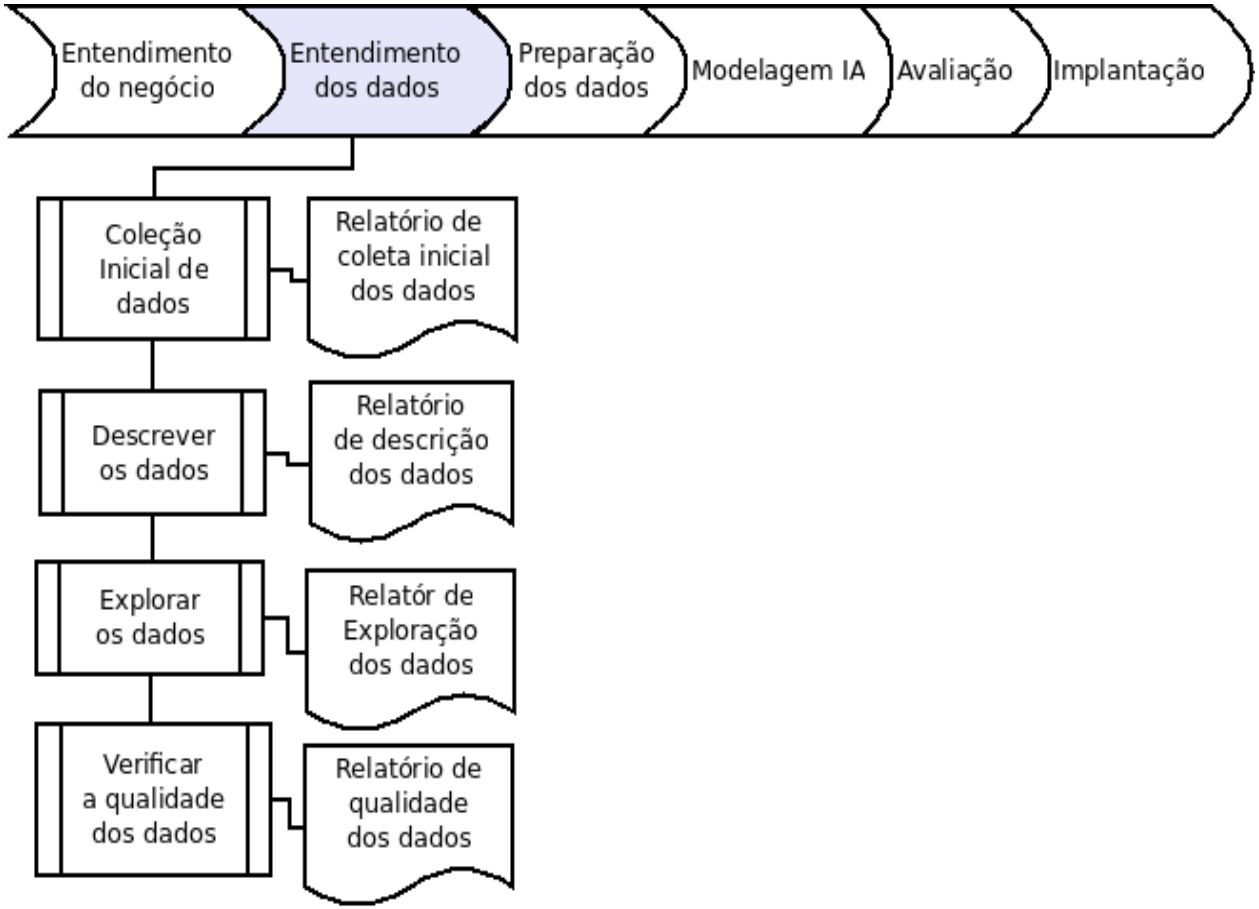
Cross Industry Standard Process for Data Maning – CRISP - DM



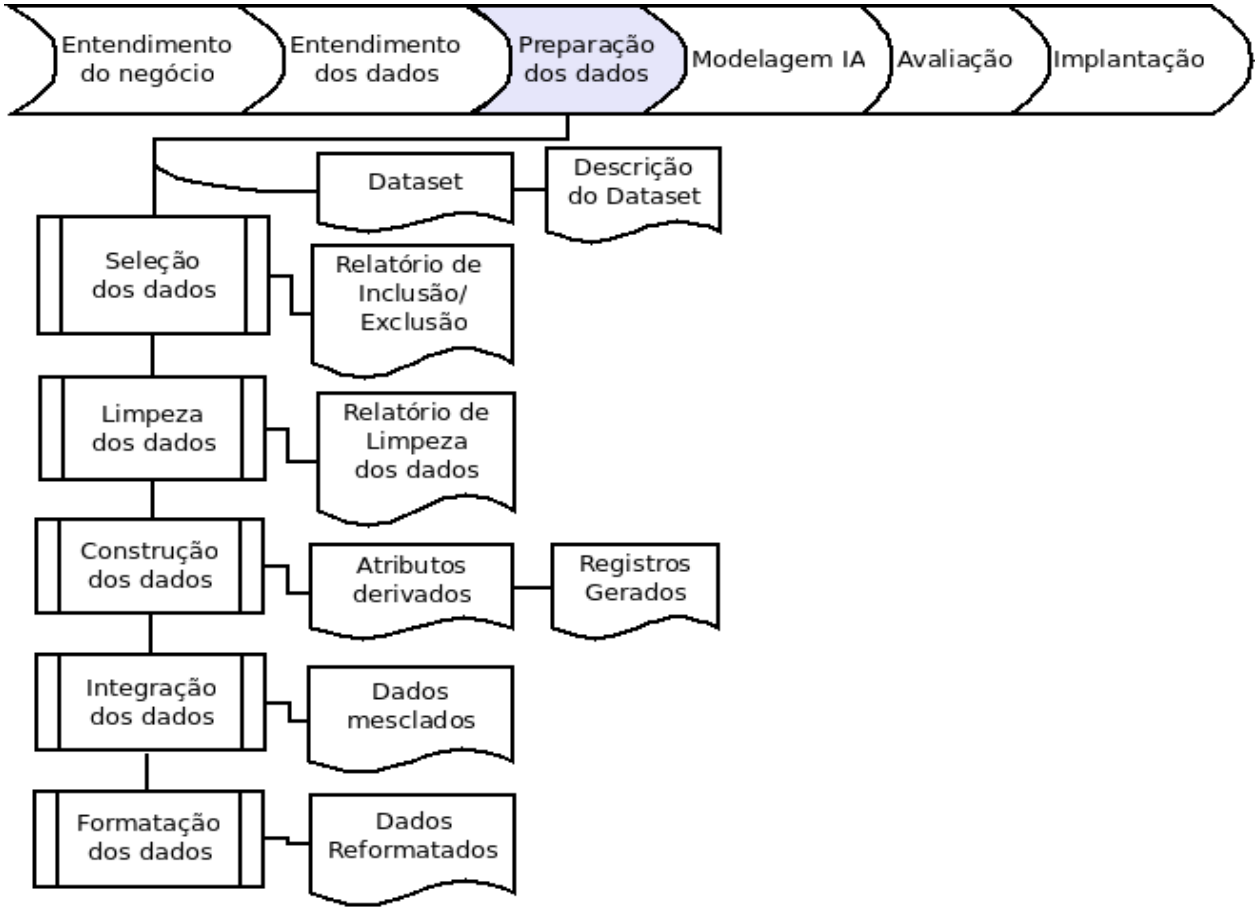
Etapas CRISP-DM resumo



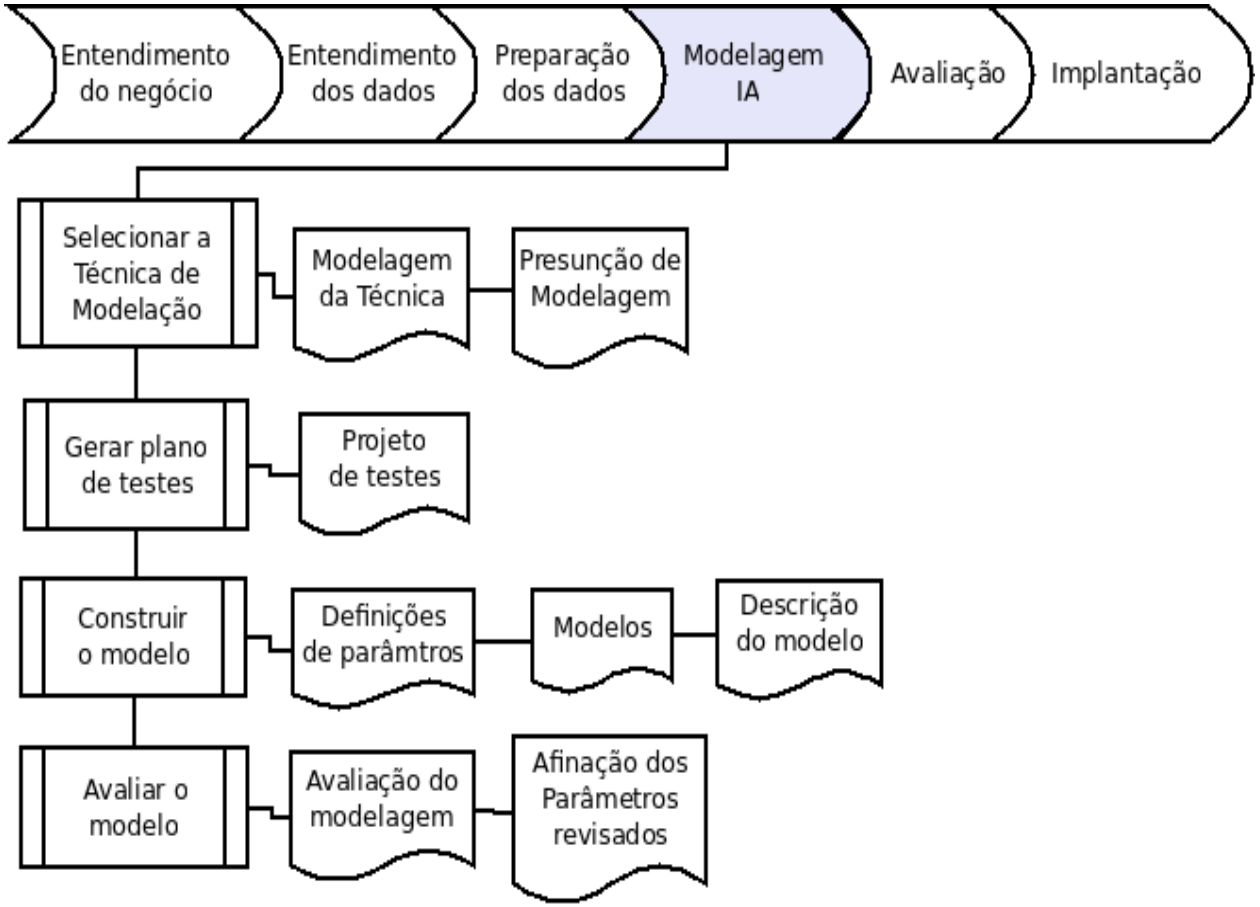
Etapas CRISP-DM resumo



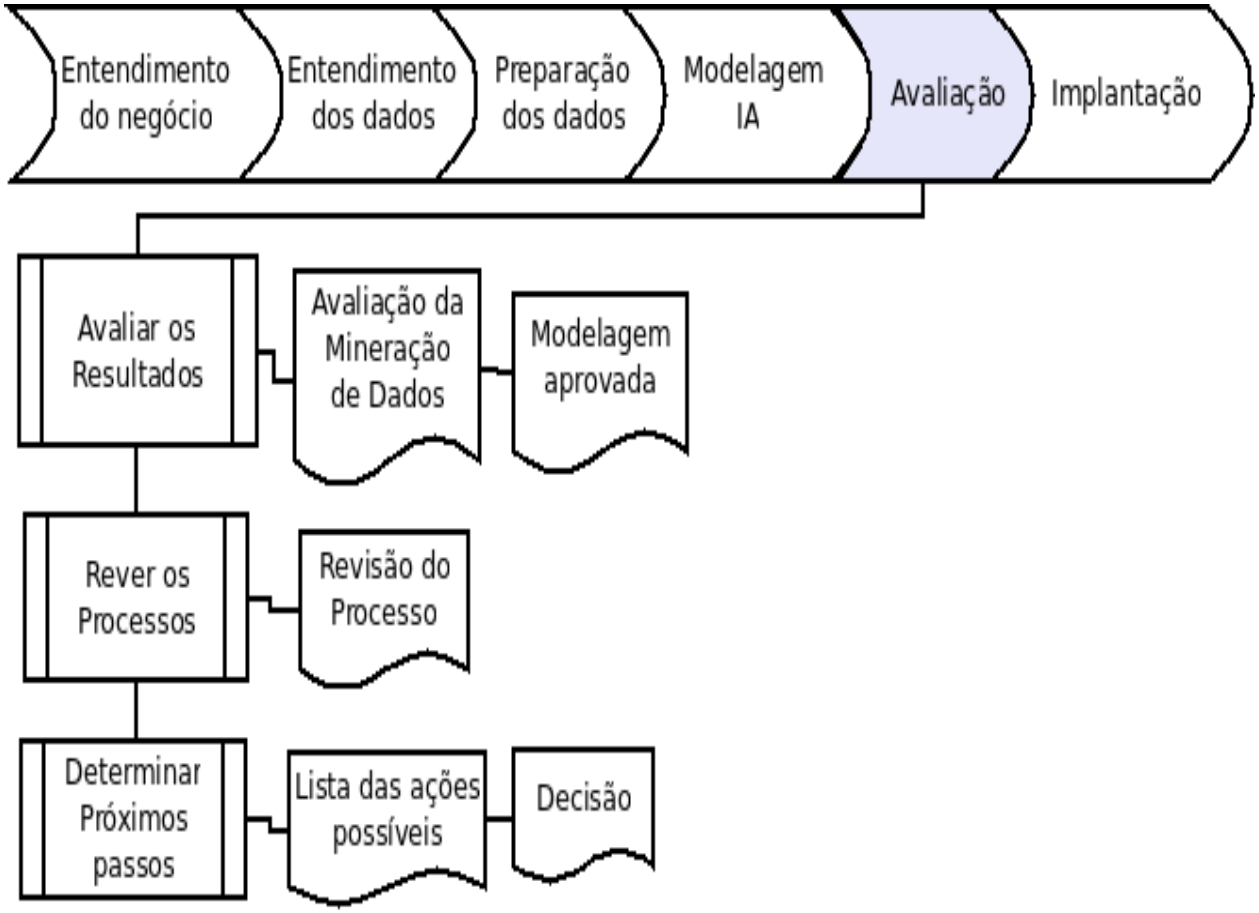
Etapas CRISP-DM resumo



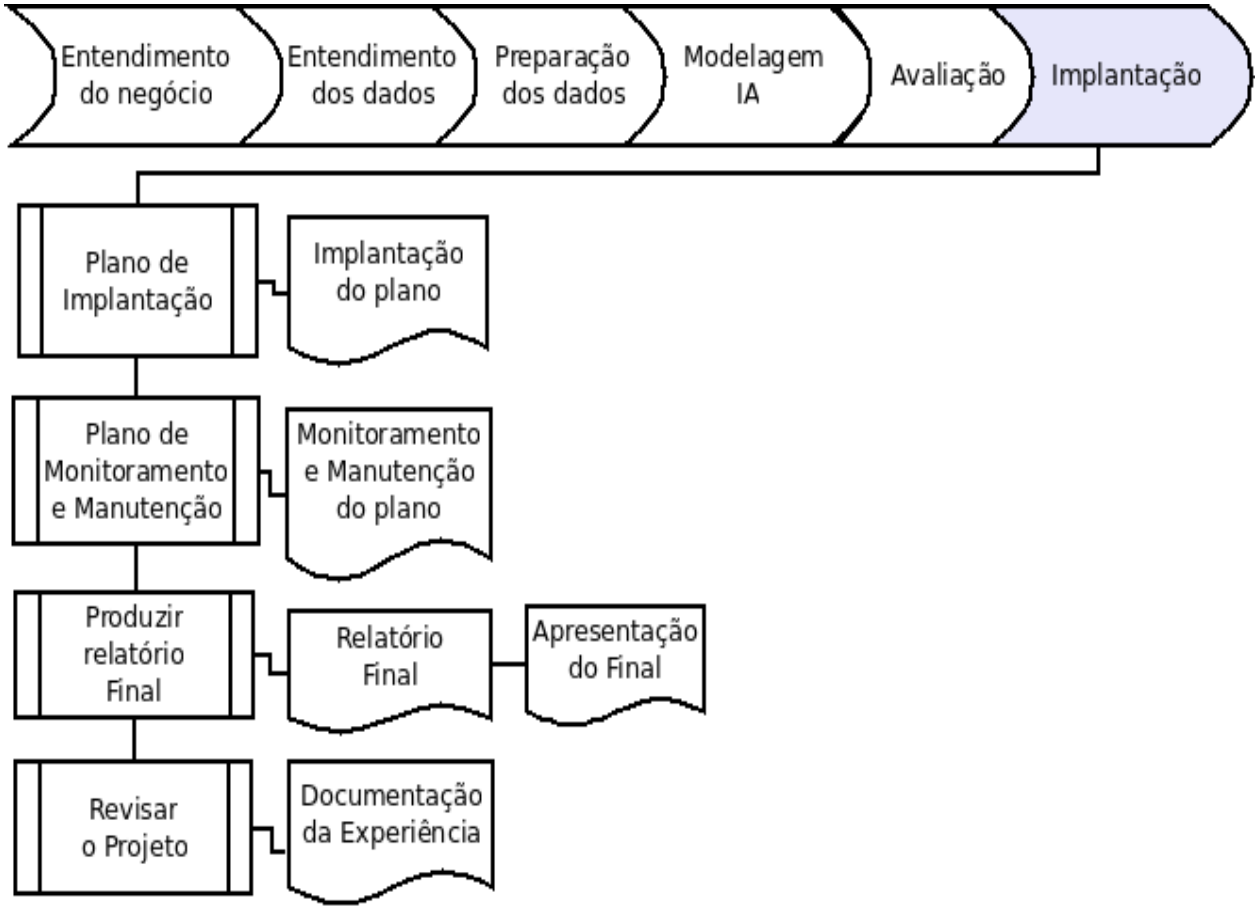
Etapas CRISP-DM resumo



Etapas CRISP-DM resumo

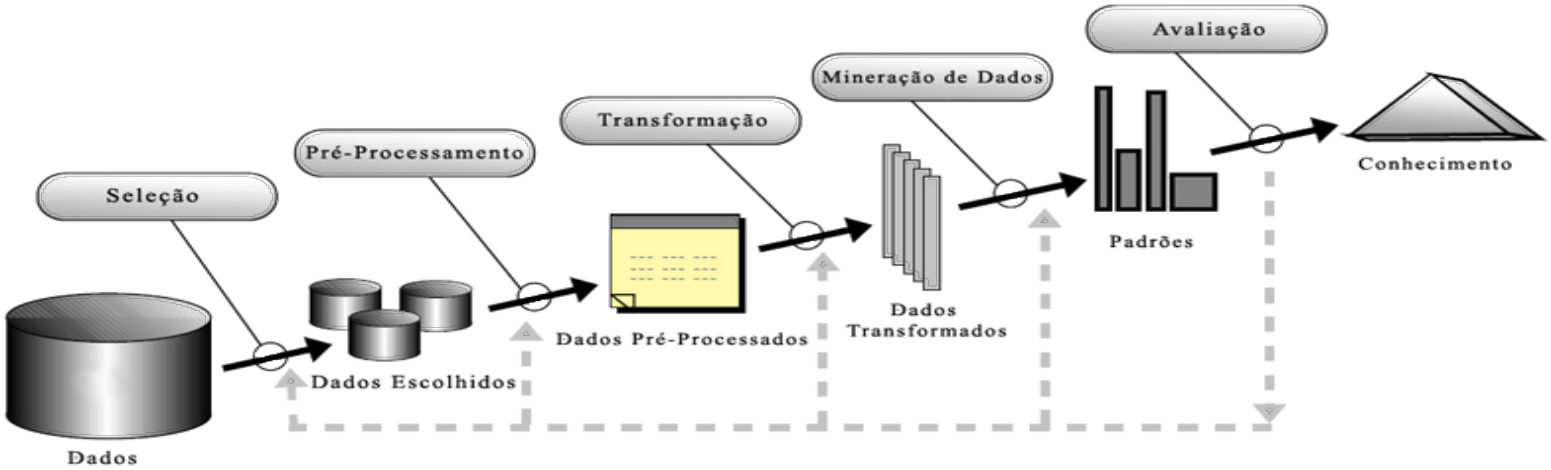


Etapas CRISP-DM resumo

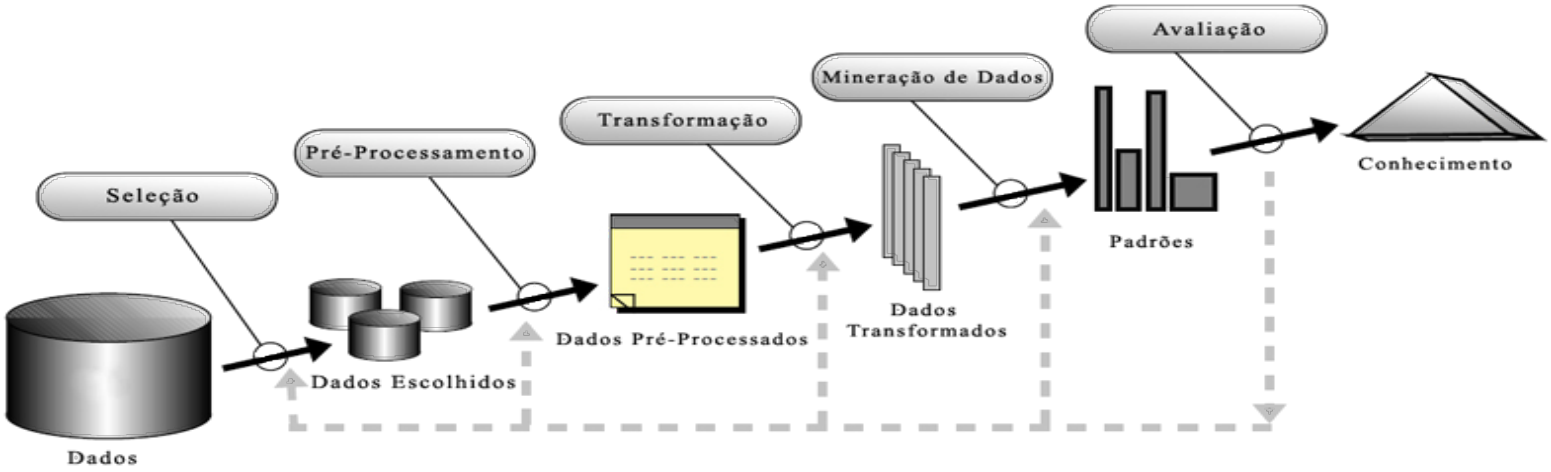


Organização

A descoberta de conhecimento (KDD) nas bases da PRF



A descoberta de conhecimento (KDD) nas bases da PRF



Diferencça 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)

Organização

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 intantâneas

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 intantâneas

Pontos fracos

- Nó raiz não é facilmente identificável

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 intantâneas

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

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

Organização

Teorema de Bayes e Aprendizagem bayesiana

Teorema de Bayes: forma tradicional

$$p(C_k|x) = \frac{p(_k)p(x|C_k)}{p(x)} \tag{3}$$

Teorema de Bayes: forma simplificada

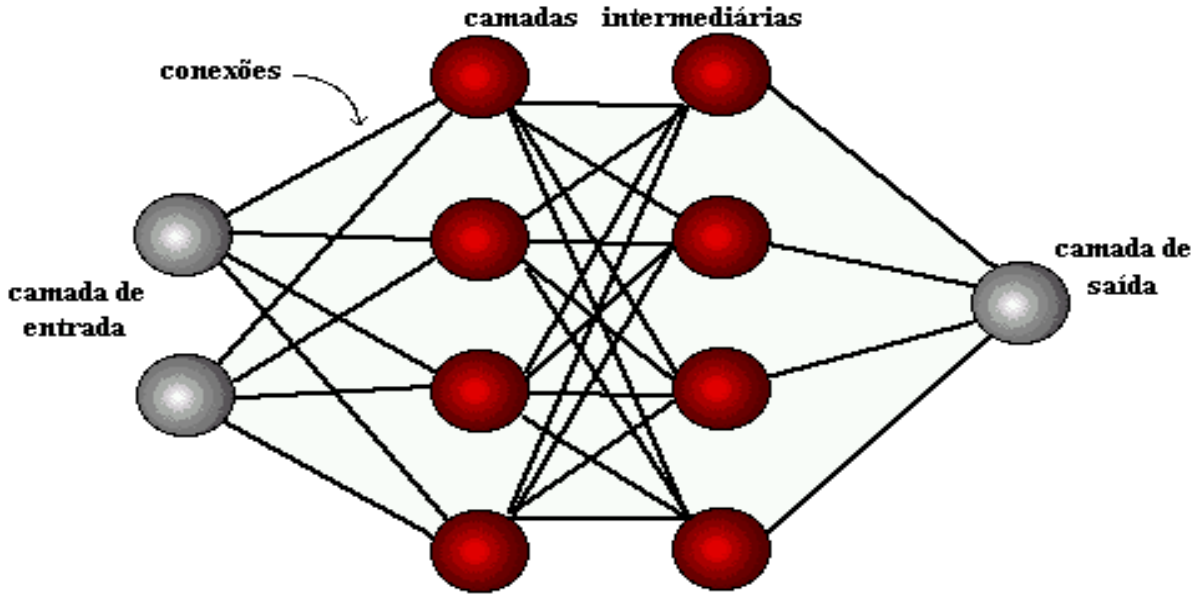
$$p(posteriori) = \frac{p(priori) * \text{verossimilhança}}{\text{evidência}} \tag{4}$$

Organização

Classificação com Redes Neurais

Rede Nerual 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.



Organização

TF – IDF

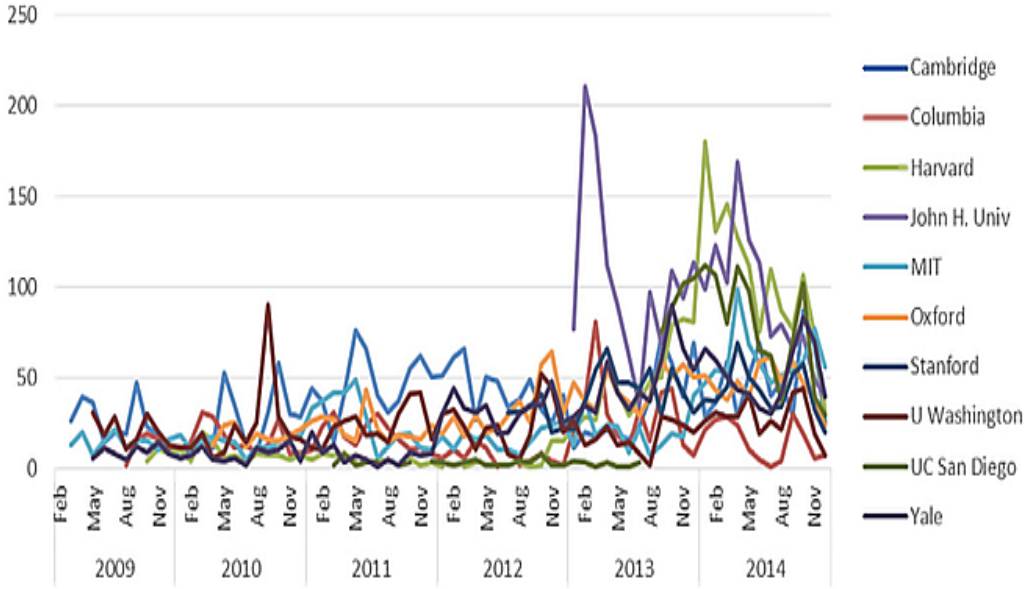
Text Frequence X Inverse Document Frequence: $TF * IDF$

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

TF – IDF

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

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



Análise de Desempenho aplicados à mineração

A área sob a curva AUC (Area under ROC – Reciver Operating Characeteristic) 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 Contingência (ou confusão) que agrupa essas atributos.

Tabela: Matriz de Confusão

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

Fonte: Bradley – 1997

Análise de Desempenho aplicados à mineração

A Matriz Modelo de Confusão sintetiza a Matriz 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

Sumário

Sumário

Sumário

Organização

Aplicativo

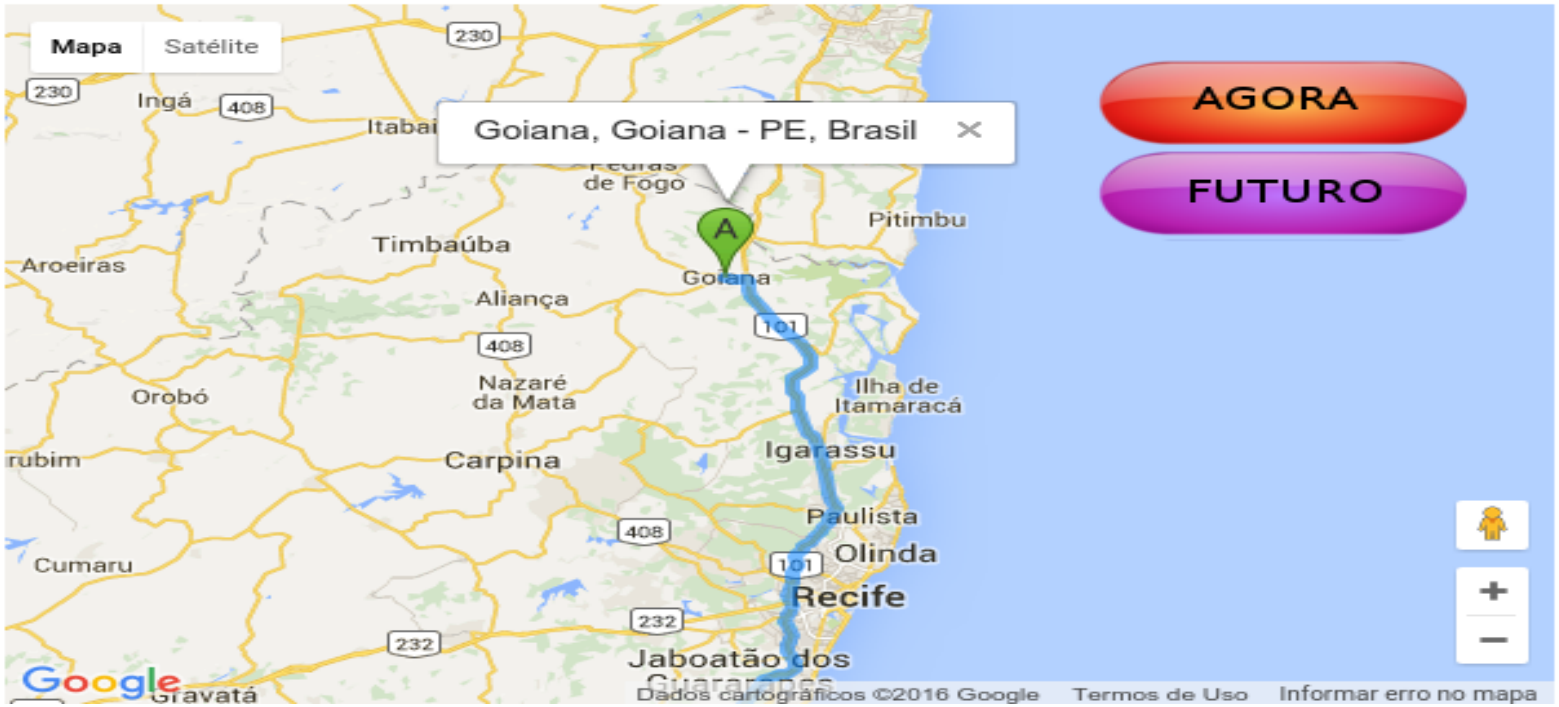


Figura: Interface Gráfica Proposta

Organização

Artigo aprovado em evento internacional

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

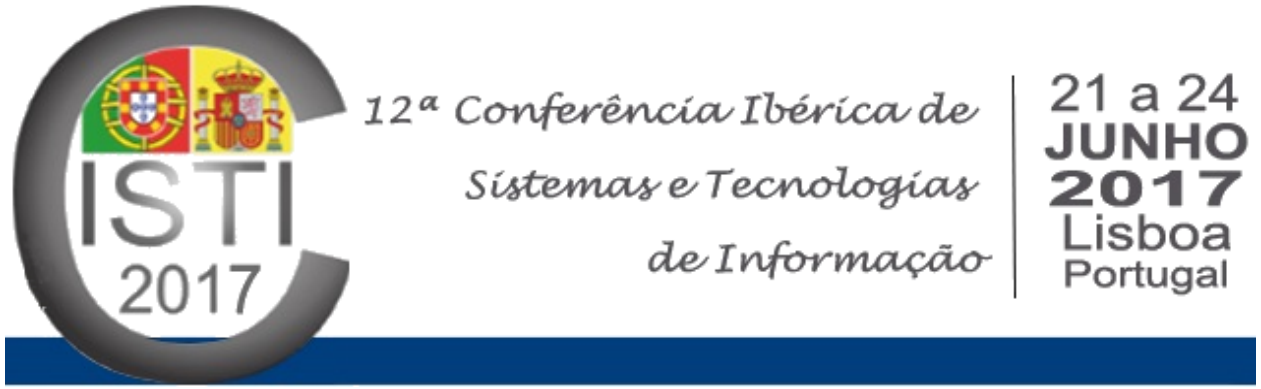


Figura: CISTI – Conf. Ibérica de Sistemas e Tecnologia da Informação