

Defesa da Dissertação 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
- 2 Objetivos
- 3 Metodologias de Mineração Dados
- 4 KDD e Mineração de Dados
- 5 Tecnologias empregues na Mineração Dados
- 6 Mineração em Textos
- 7 Trabalhos Futuros

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.

Sumário

- 1 Introdução
- 2 **Objetivos**
- 3 Metodologias de Mineração Dados
- 4 KDD e Mineração de Dados
- 5 Tecnologias empregues na Mineração Dados
- 6 Mineração em Textos
- 7 Trabalhos Futuros

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;
- 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

PRF

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

PRF

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

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...		
Row0	101	7	7.56206206749...	-34.990451894469...	Seca	Inexistente	Colisão late...	Falta de atenç...	De	
Row1	101	210	-7.66697808649...	-34.93329743899...	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...	-34.90833713549...	Seca	Inexistente	Colisão com...	Outras	Cre	
Row5	101	511	-7.90212986999...	-34.90013243449...	Seca	Inexistente	Atropelame...	Outras	Cre	
Row6	101	512	-7.91819106999...	-34.89606257549...	Seca	Inexistente	Colisão Tra...	Falta de atenç...	Cre	
Row7	101	512	-7.91819106999...	-34.89606257549...	Seca	Inexistente	Colisão Tra...	Falta de atenç...	Cre	
Row8	101	545	-7.94363545599...	-34.90420279049...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre	
Row9	101	545	-7.94363545599...	-34.90420279049...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre	
Row10	101	589	-7.97661785749...	-34.92513769249...	Seca	Inexistente	Colisão late...	Velocidade inc...	De	
Row11	101	680	-887953132499...	-35.62892224749...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre	
Row12	101	680	-887953132499...	-35.62892224749...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre	
Row13	101	680	-887953132499...	-35.62892224749...	Seca	Inexistente	Colisão late...	Ultrapassagem...	Cre	
Row14	101	680	-887953132499...	-35.62892224749...	Seca	Inexistente	Colisão late...	Ultrapassagem...	Cre	
Row15	101	680	-887953132499...	-35.62892224749...	Seca	Inexistente	Colisão tras...	Falta de atenç...	Cre	
Row16	101	680	-887953132499...	-35.62892224749...	Seca	Inexistente	Colisão tras...	Falta de atenç...	Cre	
Row17	101	680	-887953132499...	-35.62892224749...	Seca	Inexistente	Queda de m...	Outras	De	
Row18	101	680	-887953132499...	-35.62892224749...	Seca	Inexistente	Queda de m...	Velocidade inc...	Cre	
Row19	101	830	-817277505899...	-34.93998880799...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De	
Row20	101	830	-817277505899...	-34.93998880799...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De	
Row21	101	830	-817277505899...	-34.93998880799...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De	
Row22	101	830	-817277505899...	-34.93998880799...	Seca	Inexistente	Queda de m...	Outras	De	
Row23	101	865	-818644581449...	-34.97139755549...	Seca	Inexistente	Colisão tras...	Falta de atenç...	De	
Row24	101	1018	-826668917249...	-35.04169269449...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre	
Row25	101	1018	-826668917249...	-35.04169269449...	Seca	Inexistente	Colisão late...	Falta de atenç...	Cre	
Row26	101	1020	-828279581549...	-35.05767211999...	Com buca...	Inexistente	Colisão late...	Outras	Cre	
Row27	101	1020	-828279581549...	-35.05767211999...	Com buca...	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,

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,
- Qtd de Feridos Graves, Qtd Feridos Leves, Qtd Mortos, Qtd Pessoas.

O Tweeker

Acesso aos dados do Tweeker

Figura: Registro de App para acessar e baixar dados

The screenshot shows the Twitter Application Management page for an application named "dadosMestrado". The page has a blue header with the Twitter logo and the text "Application Management". Below the header, there are tabs for "Details", "Settings", "Keys and Access Tokens", and "Permissions". The "Settings" tab is currently selected. The "Application Settings" section includes a warning about the "Consumer Secret" being a secret key. Below this, there are fields for "Consumer Key (API Key)" with the value "hSEPFz9gCKJukb0tQFES0t1Z", "Consumer Secret (API Secret)" with the value "OIM5th6COPue8WWD9SRXeZX6Y9IMKuDT3iL33bsKP2Z7MLi9UI", "Access Level" set to "Read and write (modify app permissions)", "Owner" set to "otluix", and "Owner ID" set to "528603134". At the bottom, there is an "Application Actions" section with two buttons: "Regenerate Consumer Key and Secret" and "Change App Permissions".

Twitter Application Management

dadosMestrado [Test OAuth](#)

[Details](#) [Settings](#) [Keys and Access Tokens](#) [Permissions](#)

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key) hSEPFz9gCKJukb0tQFES0t1Z

Consumer Secret (API Secret) OIM5th6COPue8WWD9SRXeZX6Y9IMKuDT3iL33bsKP2Z7MLi9UI

Access Level Read and write (modify app permissions)

Owner otluix

Owner ID 528603134

Application Actions

[Regenerate Consumer Key and Secret](#) [Change App Permissions](#)

Dados originais

Dados – Planilha Tweeter

Figura: Planilha para Preprocessamento

[illegible]

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

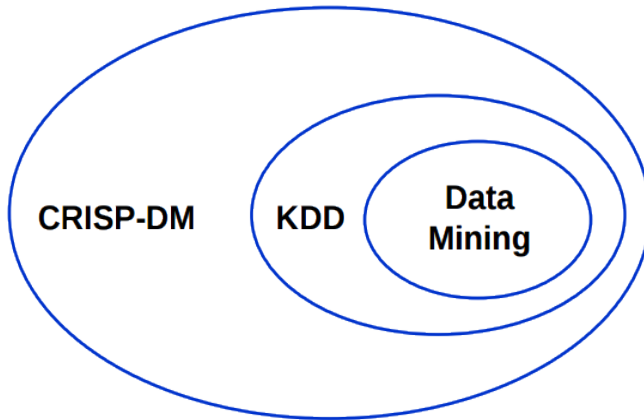
Atributos e Instâncias – iniciais

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

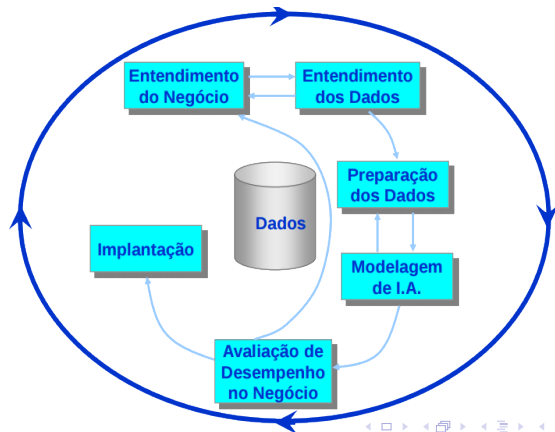
Sumário

- 1 Introdução
- 2 Objetivos
- 3 Metodologias de Mineração Dados**
- 4 KDD e Mineração de Dados
- 5 Tecnologias empregues na Mineração Dados
- 6 Mineração em Textos
- 7 Trabalhos Futuros

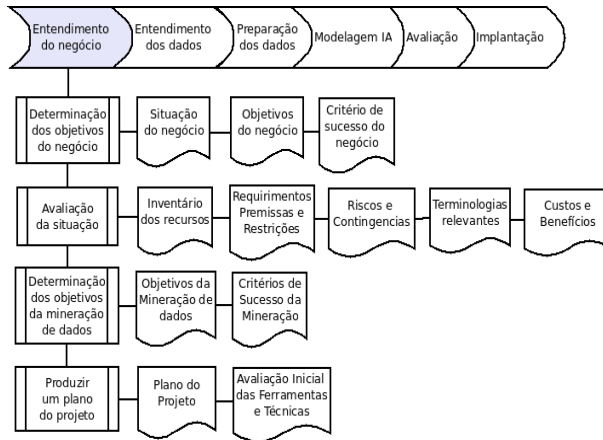
Domínio das técnicas de Mineração de Dados



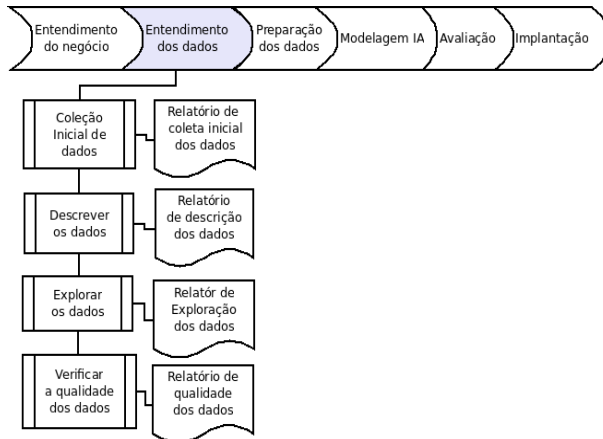
Cross Industry Standard Process for Data Maning – CRISP - DM



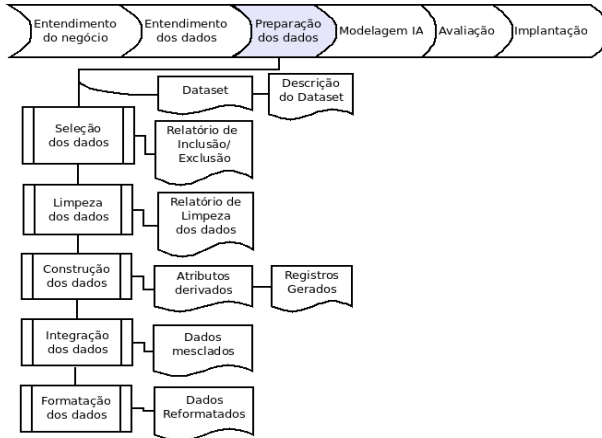
Etapas CRISP-DM resumo



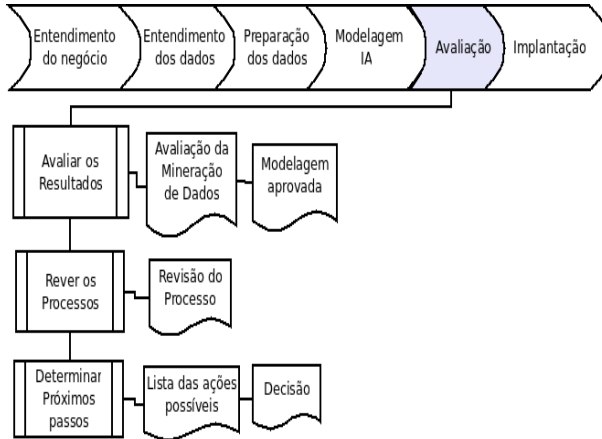
Etapas CRISP-DM resumo



Etapas CRISP-DM resumo



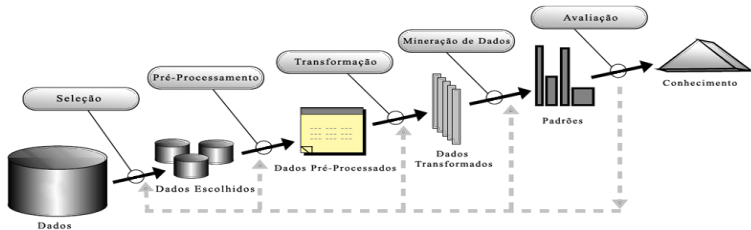
Etapas CRISP-DM resumo



Sumário

- 1 Introdução
- 2 Objetivos
- 3 Metodologias de Mineração Dados
- 4 KDD e Mineração de Dados**
- 5 Tecnologias empregues na Mineração Dados
- 6 Mineração em Textos
- 7 Trabalhos Futuros

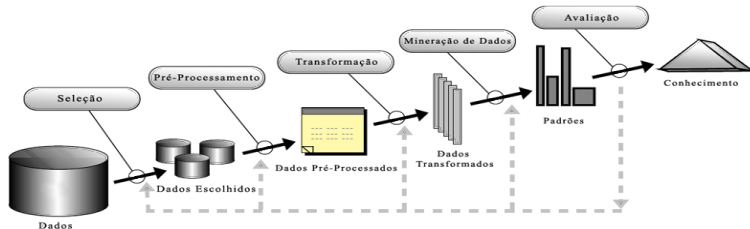
A descoberta de conhecimento (KDD) nas bases da PRF



A base de dados original – PRF

40% missing data – 70% do tempo total para tratar

A descoberta de conhecimento (KDD) nas bases da PRF



A base de dados original – PRF

40% missing data – 70% do tempo total para tratar

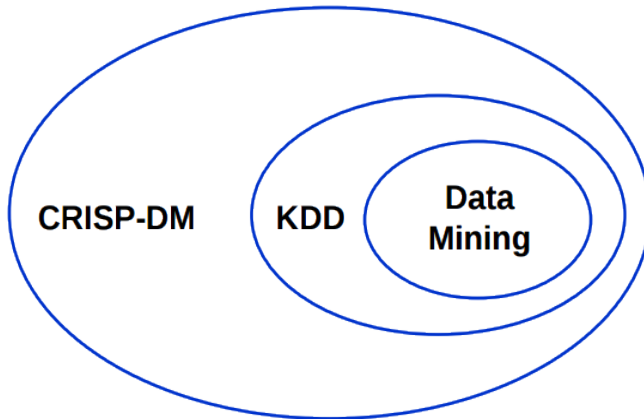
Diferença CRISP-DM — KDD

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

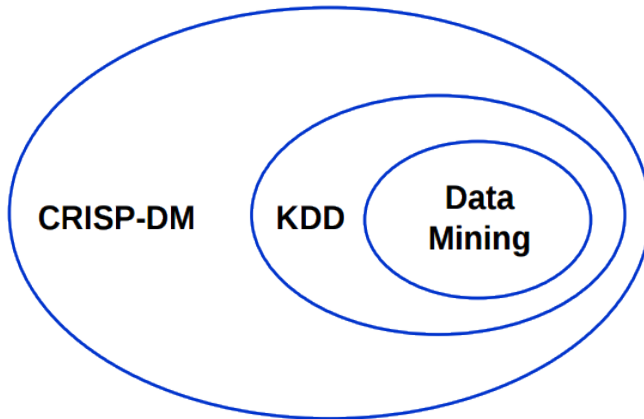
Sumário

- 1 Introdução
- 2 Objetivos
- 3 Metodologias de Mineração Dados
- 4 KDD e Mineração de Dados
- 5 Tecnologias empregues na Mineração Dados**
- 6 Mineração em Textos
- 7 Trabalhos Futuros

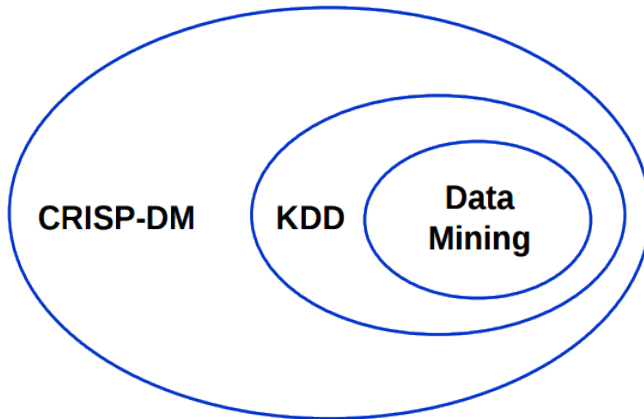
Domínio das técnicas de Mineração de Dados



Domínio das técnicas de Mineração de Dados



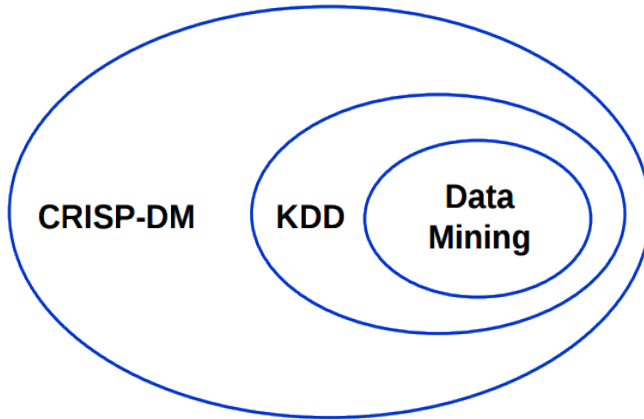
Domínio das técnicas de Mineração de Dados



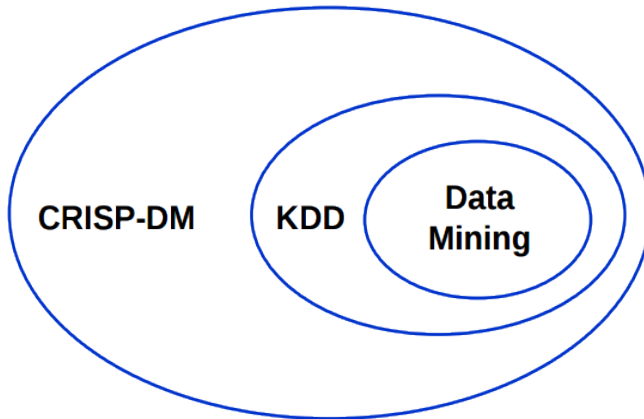
Sumário

- 1 Introdução
- 2 Objetivos
- 3 Metodologias de Mineração Dados
- 4 KDD e Mineração de Dados
- 5 Tecnologias empregues na Mineração Dados
- 6 Mineração em Textos**
- 7 Trabalhos Futuros

TF – IDF



K-means para Agrupamentos dos dados



Sumário

- 1 Introdução
- 2 Objetivos
- 3 Metodologias de Mineração Dados
- 4 KDD e Mineração de Dados
- 5 Tecnologias empregues na Mineração Dados
- 6 Mineração em Textos
- 7 **Trabalhos Futuros**

Trabalhos futuros

Figura: Interface Gráfica Proposta

