

Aplicação do Processo de Descoberta de Conhecimento em Bases de Dados de Domínios Públicos

Gabriel Levis Zawalski

Orientadora: Prof^a Helyane Bronoski Borges

Coorientadora: Thissiany Beatriz Almeida

Universidade Tecnológica Federal do Paraná

gabrielzawalski@gmail.com

<https://github.com/glzawalski/>

21 de Maio de 2018

Introdução

- Contexto

- Processo de KDD

Objetivos

- Objetivos Gerais

- Objetivos Específicos

Justificativa

Metodologia

- Etapas

- Ferramentas

- Base de dados

Cronograma

Referências

- ▶ Aumento da presença de serviços digitais na vida cotidiana;
- ▶ Acumular dados não é o suficiente;
- ▶ Métodos tradicionais de análise estatística requerem interpretação manual;
- ▶ Fayyad 1996 [1] diz que neste cenário há uma grande necessidade de novas teorias e ferramentas computacionais a fim de auxiliar na extração de conhecimento dos grandes e cada vez maiores volumes de dados.

- ▶ Knowledge Discovery in Databases - cunhado por Shapiro em 1989;
- ▶ Processo iterativo e não trivial dependente de interações do usuário;
- ▶ Tem como objetivo identificar informações válidas, novas, potencialmente úteis e compreensíveis em um grupo de dados.

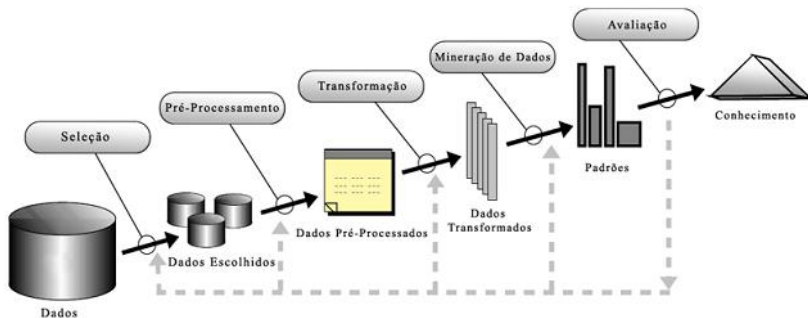


Figura: Fluxograma do processo de KDD. Adaptado de Fayyad 1996 [1].

- ▶ A mineração de dados é uma das etapas do KDD que mais recebe destaque;
- ▶ Grande quantidade de técnicas e resultados disponíveis para teste;
- ▶ Descoberta e adaptação de padrões e modelos;
- ▶ Classificação, Clusterização e Regressão.

- ▶ O objetivo geral deste trabalho é aplicar o processo de KDD numa base de dados de domínio público a fim de criar um modelo de previsão de sucesso de postagem.

1. Compreender o funcionamento do processo de KDD;
2. Analisar as etapas do KDD, identificando técnicas que podem ser aplicadas;
3. Aplicar o processo de KDD na base de dados;
4. Realizar experimentos e analisar os resultados obtidos através de comparação estatística com outros trabalhos da área.

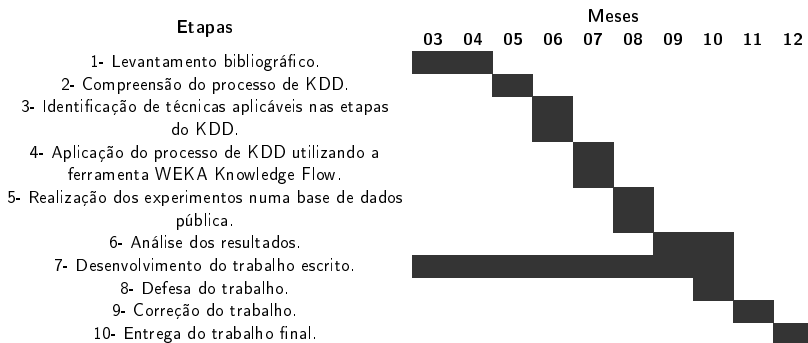
- ▶ Possuir dados não significa ter algum conhecimento;
- ▶ Sem análise rigorosa conclusões erradas podem ser tomadas;
- ▶ KDD se torna interessante pois estabelece um processo de fácil replicação;
- ▶ A base de dados escolhida possui dados reais e recentes.

1. Levantamento bibliográfico;
2. Compreensão do processo de KDD;
3. Identificação de técnicas aplicáveis nas etapas do KDD;
4. Aplicação do processo de KDD utilizando a ferramenta WEKA Knowledge Flow;
5. Realização dos experimentos numa base de dados pública;
6. Análise dos resultados.

- ▶ Aplicação de experimentos com apoio da WEKA (Waikato Enviroment for Knowledge Analysis);
- ▶ Ferramenta de auxílio desenvolvida na Nova Zelândia;
- ▶ WEKA Knowledge Flow permite a vizualização de processos como data-flow;
- ▶ 5 técnicas de regressão:
 1. Regressão linear;
 2. K vizinhos mais próximos;
 3. Árvores de decisão;
 4. Máquinas de vetores de suporte;
 5. Perceptron multicamadas.

- ▶ Inicialmente descrita por Moro 2016 [2];
- ▶ A base contém 500 postagens publicadas durante o ano de 2014 na página do Facebook de uma marca de cosméticos;
- ▶ 19 atributos, sendo 7 referentes a postagem e 12 referentes a performance e impacto medidas pelo Facebook;
- ▶ Análise original utilizou máquinas de vetores de suporte do pacote rminer do software de análise estatística R;
- ▶ Melhores previsões de erros foram ao entorno de 27%.

Tabela: Cronograma





Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth.

From data mining to knowledge discovery in databases.

AI Magazine, 1996.



Moro S, Rita P, and Vala B.

Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach.

Journal of Business Research, 2016.

Aplicação do Processo de Descoberta de Conhecimento em Bases de Dados de Domínios Públicos

Gabriel Levis Zawalski

Orientadora: Prof^a Helyane Bronoski Borges

Coorientadora: Thissiany Beatriz Almeida

Universidade Tecnológica Federal do Paraná

gabrielzawalski@gmail.com

<https://github.com/glzawalski/>

21 de Maio de 2018