

SUMÁRIO

1 – INTRODUÇÃO	1
1.1 DESCRIÇÃO DO PROBLEMA E MOTIVAÇÃO	2
1.2 OBJETIVOS	2
1.2.1 Objetivos gerais	2
1.2.2 Objetivos específicos	2
1.3 ORGANIZAÇÃO DO TRABALHO	2
2 – DESCOBERTA DE CONHECIMENTO	4
2.1 CONCEITOS FUNDAMENTAIS DO KDD	4
2.2 ETAPAS DO KDD	4
2.2.1 Pré-processamento	4
2.2.2 Mineração de dados	4
2.2.2.1 Tarefas da mineração de dados	4
2.2.3 Pós-processamento	4
3 – REDES SOCIAIS	5
3.1 CONCEITOS FUNDAMENTAIS	5
3.2 CLASSIFICAÇÃO E RELEVÂNCIA	5
3.3 FACEBOOK	5
4 – REVISÃO SISTEMÁTICA DA LITERATURA	6
4.1 MÉTODO DE REVISÃO SISTEMÁTICA	6
4.2 APLICAÇÃO DO MÉTODO	6
5 – METODOLOGIA	7
5.1 PRÉ-PROCESSAMENTO	7
5.2 MINERAÇÃO DE DADOS	7
5.3 PÓS-PROCESSAMENTO	7
6 – ANÁLISE E DISCUSSÃO DOS RESULTADOS	8
6.1 RESULTADOS DO PROCESSO DE KDD	8
6.2 ANÁLISE DOS RESULTADOS	8
6.2.1 Comparativo entre dados reais coletados	8
6.2.2 Comparativo entre outros trabalhos relacionados	8
7 – CONCLUSÃO	9
7.1 TRABALHOS FUTUROS	9

7.2 CONSIDERAÇÕES FINAIS	9
------------------------------------	---

1 INTRODUÇÃO

Com a crescente presença de serviços digitais na vida cotidiana da população, como uso de redes sociais, serviços de streaming, lojas e compras online uma quantidade enorme de dados está sendo armazenada a todo momento em todos os lugares do mundo. Acumular dados não é suficiente por si só, tão importante quanto é extrair informações destes. Porém, a maioria dos métodos tradicionais de análise ainda requerem interpretação manual, sendo muitas vezes ineficientes, demorados, custosos além de propensos à subjetividade do analista.

(Fayyad 1996) 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.

O processo de Descoberta de Conhecimento em Bases de Dados (Knowledge Discovery in Databases - KDD), termo cunhado por (Shapiro 1989) na realização do primeiro workshop a respeito do tema, foi uma resposta aos problemas causados por esse grande acúmulo de dados.

O KDD é um processo iterativo não trivial dividido em 9 etapas dependentes de interações do usuário. A Figura 1 apresenta a sequência das etapas do processo de KDD de forma simplificada. O processo tem por objetivo identificar informações válidas, novas, potencialmente úteis e compreensíveis em um grupo de dados, segundo (Fayyad 1996). Essas informações extraídas também precisam ser, até certo grau de certeza, válidas para um novo conjunto de dados. O processo de KDD envolve desde a compreensão dos dados armazenados na base, sua seleção e pré-processamento, passando por etapas de transformação e aplicação de técnicas de mineração de dados, terminando nas etapas de identificação dos padrões gerados que se enquadram nos requisitos de conhecimento do processo.

Entre as etapas do processo de KDD, a mineração de dados é uma das que possui grande ênfase tanto nas áreas acadêmicas quanto nas áreas práticas devido as grandes quantidades de métodos e resultados disponíveis e testados. O processo de mineração de dados envolve a descoberta de padrões a partir de dados e a adaptação de modelos para melhor acomodar os dados existentes, se utilizando de técnicas de muitas áreas diversas, como aprendizagem de máquina, reconhecimento de padrões e estatística. Dentro dos diversos métodos de mineração de dados, sendo os mais comuns os de classificação, clusterização e regressão, existem uma extensa gama de algoritmos, porém todos com funcionamentos e fundamentos parecidos. (Fayyad 1996) diz que os métodos de mineração de dados podem ser compostos de três algoritmos diferentes: modelo de representação, modelo de avaliação, e busca.

Como foi citado anteriormente, somente ter acesso a determinados dados não garante a completa compreensão do problema, porém, com a aplicação de processos como o KDD, pode-se tirar conclusões úteis a respeito dos dados tratados. Este trabalho tem como objetivo utilizar o processo do KDD na base de dados descrita acima, a fim de encontrar uma modelagem de previsão de sucesso da postagem com uma taxa de erro aceitável, fazendo uso de diferentes

técnicas de mineração e análise de dados.

1.1 DESCRIÇÃO DO PROBLEMA E MOTIVAÇÃO

O grande aumento do volume de dados acumulados por diversos tipos de aplicações pode gerar um grande problema, uma vez que nem sempre somente tê-los significa ter algum conhecimento a respeito do problema em questão. Erros podem ser cometidos sem uma devida análise rigorosa e metódica a fim de se extrair informações relevantes, levando a conclusões erradas e na base de achismos. Aplicar o processo de KDD se torna cada vez mais interessante nesse cenário superpopuloso de dados uma vez que estabelece um processo que pode ser seguido e replicado em diversos tipos de situações diferentes. A base de dados utilizada no trabalho foi escolhida por possuir dados de uma situação real, além de serem relativamente recentes e com poucos trabalhos explorando os diversos tratamentos que os dados podem receber, permitindo assim uma grande gama de abordagens diferentes possíveis.

1.2 OBJETIVOS

Esta Seção apresenta o objetivo geral e os objetivos específicos deste trabalho. Na Subseção 1.2.1 se encontra o objetivo geral e na Subseção 1.2.2 se encontram os objetivos específicos.

1.2.1 Objetivos gerais

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.

1.2.2 Objetivos específicos

Como objetivos específicos têm-se:

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

1.3 ORGANIZAÇÃO DO TRABALHO

Este trabalho encontra-se dividido nos seguintes sete capítulos:

- Capítulo 1: Capítulo introdutório de contextualização do trabalho, apresentando em linhas gerais a situação atual, a motivação que levou a idealização deste trabalho e os objetivos a serem alcançados.

- Capítulo 2: Este capítulo aborda os conceitos necessários para compreender o processo de KDD. Contém uma descrição das etapas que o constituem como um todo e entra mais a fundo em conceitos importantes da parte de mineração de dados.
- Capítulo 3: O capítulo apresenta conceitos e definições a respeito de redes sociais, classificando-as e descrevendo sua relevância fora e dentro da área acadêmica. Também contém uma explicação a respeito de termos da rede social Facebook, sendo esta o local onde os dados utilizados neste trabalho foram retirados.
- Capítulo 4: Neste capítulo encontra-se uma explicação a respeito do método de revisão metodológica, ressaltando os pontos importantes do processo e a aplicação deste em bases de artigos acadêmicos, permitindo a identificação de trabalhos correlatos que serviram de referência para a elaboração deste.
- Capítulo 5: O capítulo descreve a base de dados utilizada para a realização dos experimentos deste trabalho, bem como as técnicas e métodos a serem aplicados na mesma, seguindo o processo definido no capítulo 2.
- Capítulo 6: Este capítulo contém a descrição e análise dos resultados obtidos após a aplicação dos métodos e técnicas descritos no capítulo 5 sobre a base de dados deste trabalho.
- Capítulo 7: O capítulo contém possibilidades de continuções deste trabalho, bem como as considerações finais a respeito dos resultados obtidos e do cumprimento dos objetivos propostos.

2 DESCOBERTA DE CONHECIMENTO

Capítulo sobre o processo de descoberta do conhecimento e suas etapas

2.1 CONCEITOS FUNDAMENTAIS DO KDD

Conceitos gerais e relevância do estudo do kdd

2.2 ETAPAS DO KDD

As etapas do kdd

2.2.1 Pré-processamento

2.2.2 Mineração de dados

2.2.2.1 Tarefas da mineração de dados

2.2.3 Pós-processamento

3 REDES SOCIAIS

Este capítulo falará sobre redes sociais e sua influência/importância

3.1 CONCEITOS FUNDAMENTAIS

Conceitos comuns.

3.2 CLASSIFICAÇÃO E RELEVÂNCIA

Exemplos e porque estudar redes sociais e quais beneficios podemos tirar de suas análises.

3.3 FACEBOOK

4 REVISÃO SISTEMÁTICA DA LITERATURA

Revisão dos trabalhos relacionados da área

4.1 MÉTODO DE REVISÃO SISTEMÁTICA

Explicação do método de revisão sistemática

4.2 APLICAÇÃO DO MÉTODO

Resultados da aplicação da revisão sistemática.

Quadro de bases aplicadas.

Quadro de resultados por base.

Tabela 1 – Bases de dados utilizadas.

Base de dados	Sites
<i>arXiv.org</i>	< https://arxiv.org/ >
<i>Emerald Insight</i>	< https://www.emeraldinsight.com/ >
<i>IEEEExplore</i>	< https://ieeexplore.ieee.org >
<i>Science Direct</i>	< https://www.sciencedirect.com/ >
<i>Springer</i>	< https://link.springer.com/ >

Tabela 2 – Quantidade de resultados por base.

Base de dados	Resultados
<i>arXiv.org</i>	21
<i>Emerald Insight</i>	4
<i>IEEEExplore</i>	752
<i>Science Direct</i>	266
<i>Springer</i>	902

5 METODOLOGIA

Metodologia, ferramenta e base utilizada na pesquisa.

5.1 PRÉ-PROCESSAMENTO

Tudo que foi aplicado na base de dados.

5.2 MINERAÇÃO DE DADOS

Tudo que foi aplicado para mineração, svm e regressão.

5.3 PÓS-PROCESSAMENTO

Tudo que foi feito após o proc de datamining.

6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

6.1 RESULTADOS DO PROCESSO DE KDD

Conhecimento retirado do processo de kdd dos dados

6.2 ANÁLISE DOS RESULTADOS

Comparação dos resultados com os dados reais da base e de outros trabalhos

6.2.1 Comparativo entre dados reais coletados

6.2.2 Comparativo entre outros trabalhos relacionados

7 CONCLUSÃO

Resultados obtidos da regressão

7.1 TRABALHOS FUTUROS

Continuações do trabalho

7.2 CONSIDERAÇÕES FINAIS

Considerações finais