



UMA VISÃO GERAL DE COMO A MINERAÇÃO DE DADOS FUNCIONA E DE COMO VEM INTERFERINDO NO MUNDO DOS NEGÓCIOS E EM NOSSA VIDA PESSOAL

Eduardo Gabriel Thomas¹

Professor Orientador: Elio Ribeiro Faria Junior²
Modalidade de Apresentação: Comunicação Oral

INTRODUÇÃO

Com o avanço da tecnologia, diversos processos foram otimizados para agilizar em tempo e em eficiência. Isso levou todos os tipos de comércios serem obrigados a se informatizar, assim, transferências, processos, ou até dados brutos que fizessem qualquer interação com algum sistema, começaram a ser armazenados em enormes períodos de tempo e com uma precisão do que ocorreu de fato, muito maior, isso levou à estimativa do ano de 2013, que afirmava que, nos dois próximos anos, a quantidade de dados criada seria equivalente a 90% dos dados já criados no mundo, estimativa realizada pela SINTEF, que é considerada a maior organização de pesquisa da Europa. Essa enorme quantidade de dados armazenados possibilitou o uso de técnicas de Mineração de Dados (*Data Mining*). A Mineração de Dados utiliza-se de estatística aplicada, com o objetivo de encontrar padrões que, às vezes, são invisíveis somente à percepção humana. A mineração de dados pode ser comparada ao processo de se fazer uma previsão do tempo, em que se tem o foco em dois grandes pontos: Simplificar os padrões climáticos para algo mais simples de se entender, e ter uma previsão de algum dia específico. O fundamental para uma mineração bem-sucedida é uma grande quantidade de dados, isso faz com que as associações sejam mais seguras. Existem várias formas de fazer a mineração de dados, cada forma específica para algum tipo de resultado. Os algoritmos principais são: classificação, que visa determinar se algo/alguém se encaixa em alguma categoria, como se a pessoa está doente ou não. Regressão, esse algoritmo é muito bem implementado como complemento da classificação, e tem como objetivo prever quando algo vai acontecer. No caso de uma empresa de formaturas, a empresa usaria a classificação para descobrir quem está fazendo faculdade e, então, usaria a regressão para tentar definir quando a pessoa vai se formar. *Clustering*, é o algoritmo capaz de agrupar dados semelhantes dentro de um grupo especificado de pessoas ou itens. *Anomaly Detection*, serve para identificar ações destoantes do usual de algum processo, é bastante usado com cartões de crédito para identificar compras que fogem do padrão. *Association Learning*, compara tudo em um contexto geral, a fim de identificar padrões existentes, usado com variáveis externas, por exemplo, fazendo a associação de que quando chove as pessoas pegam mais táxis.

¹ Acadêmico do 6º semestre do curso de Sistemas de Informação da UNIUV. E-mail: si.eduardo.thomas@uniuv.edu.br

² Professor da UNIUV e pesquisador na área de Mineração de Processos. E-mail: prof.elio@uniuv.edu.br



METODOLOGIA

Como metodologia deste estudo, realizou-se uma pesquisa bibliográfica na área de estudo, juntamente com vídeos pesquisados.

REFERENCIAL TEÓRICO, RESULTADOS E CONSIDERAÇÕES FINAIS

Para o mundo dos negócios, a mineração de dados foi um avanço tecnológico, o qual possibilitou estudos que geraram muito impacto. A precisão de anúncios chegou a um lugar assustador, onde podem ser oferecidos produtos que você ainda não sabe que precisa. Segundo a pesquisa, conduzida pela Hsiang-Yang Chen (2011), na área médica, a mineração de dados contribuiu para o diagnóstico de doenças em bebês prematuros, o que mostra que é realmente uma ferramenta que deve ser explorada e pode ser usada para algo que contribua significativamente de maneira positiva. Entretanto, quanto mais um usuário usa algum serviço *online*, mais informações dele são armazenadas em algum lugar, o que pode chegar a um ponto em que os usuários não tenham mais controle de nada, por exemplo, o usuário pode ter seu pedido de crédito em um banco negado antes mesmo de ir à agência, ou ser preso por um crime que ainda não cometeu. Quanto mais o tempo passa, mais processos do dia a dia se tornam digitais, tudo o que fazemos, seja no trabalho ou em casa, é interagindo com algo que está conectado. Essa questão deveria ser vista mais de perto, tanto pelo lado da lei, quanto também pelo lado do usuário, que deixa uma pegada digital que, às vezes, ele não quer deixar, não por questões ilícitas, mas, sim, na questão de privacidade mesmo, algo que a cada dia que passa está se tornando mais precioso.

REFERÊNCIAS

- CHEN, H. Y. **Exploring the risk factors of preterm birth using data mining**, 2011. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417410011619?via%3Dihub>>. Acesso em: 13 set. 2019.
- HARRIS, R. **More data will be created in 2017 than the previous 5,000 years of humanity**, 2016. Disponível em: <<https://appdeveloperomagazine.com/more-data-will-be-created-in-2017-than-the-previous-5,000-years-of-humanity/>>. Acesso em: 13 set. 2019.
- SCIHOW. **Data Mining: How You're Revealing More Than You Think**, 2018. Disponível em: <<https://www.youtube.com/watch?v=EH3bp5335IU>>. Acesso em: 13 set. 2019.
- SINTEF. **Big Data, for better or worse: 90% of world's data generated over last two years**, 2013. Disponível em: <<https://www.sciencedaily.com/releases/2013/05/130522085217.htm>>. Acesso em: 13 set. 2019.