

UNIVERSIDADE EVANGÉLICA DE GOIÁS - UNIEVANGÉLICA
ENGENHARIA DE COMPUTAÇÃO/ENGENHARIA DE SOFTWARE

Bianca Abreu Félix de Lima
João Vitor Sponchiado
Natal Junio Barbosa de Souza
Patrick Alves Freitas

**Mineração de dados na transação da Amazon
antes e depois da pandemia.**

Anápolis - GO
Dezembro, 2021

UNIVERSIDADE EVANGÉLICA DE GOIÁS - UNIEVANGÉLICA
ENGENHARIA DE COMPUTAÇÃO/ENGENHARIA DE SOFTWARE

Bianca Abreu Félix de Lima
João Vitor Sponchiado
Natal Junio Barbosa de Souza
Patrick Alves Freitas

**Mineração de dados na transação da Amazon
antes e depois da pandemia.**

Trabalho apresentado ao Curso de Engenharia de Software da Universidade Evangélica de Goiás – UniEVANGÉLICA, da cidade de Anápolis-GO como requisito parcial para obtenção do Grau de Bacharel em Engenharia de Software.

Orientador (a): Prof. Alexandre Moraes Tannus

Anápolis
Dezembro, 2021

UNIVERSIDADE EVANGÉLICA DE GOIÁS - UNIEVANGÉLICA
ENGENHARIA DE COMPUTAÇÃO/ENGENHARIA DE SOFTWARE

Bianca Abreu Félix de Lima
João Vitor Sponchiado
Natal Junio Barbosa de Souza
Patrick Alves Freitas

**Mineração de dados na transação da Amazon
antes e depois da pandemia.**

Monografia apresentada para Trabalho de Conclusão de Curso de Engenharia de Software da Universidade Evangélica de Goiás - UniEVANGÉLICA, da cidade de Anápolis-GO como requisito parcial para obtenção do grau de Engenheiro(a) de Software.

Aprovado por:

(ORIENTADOR)

(AVALIADOR)

Anápolis, Dezembro de 2021.

AGRADECIMENTOS

Agradeço aos meus pais, Leida Maria e Airton Félix, por todo amparo durante esses anos cursando Engenharia de Software e que sem eles, eu não teria chegado até aqui. Eles me guiaram e me orientaram para que eu pudesse progredir. Agradeço a todos que fizeram e que ainda vão continuar fazendo parte da minha vida, durante esses 4 anos estudando juntos. Conheci pessoas incríveis e com toda certeza, contribuíram para o meu aprendizado. Agradeço ao meu irmão, Diego Félix, por sempre acreditar em mim quando eu mesma duvidava e por todo apoio.

Bianca Abreu Félix de Lima.

Venho agradecer primeiramente a minha família que sempre me apoiou e me motivou em tudo o que sempre sonhei, meus pais por nunca deixarem me faltar oportunidades e principalmente meus avós os quais sempre me deram o suporte necessário para que eu realizasse esse desejo de me formar. Aos professores do curso pelos aprendizados tanto nas matérias quanto para a vida e o mercado de trabalho e também aos meus colegas de classe os quais sempre tiveram paciência e me ajudaram no que foi preciso, tanto nos trabalhos como na convivência.

João Vitor Sponchiado.

Agradeço a Deus pelos dons que me deu nesta existência que serviram na realização deste projeto. Sou grato aos meus pais por sempre me incentivarem e acreditarem que eu seria capaz de superar os obstáculos que a vida me apresentou. Agradeço ao meu orientador, Alexandre Moraes Tannus por sempre estar presente para indicar a direção correta que o trabalho deveria tomar. Também agradeço a meu amigo que sempre me ajudou com sua vasta experiência desde o início deste projeto de pesquisa. Também agradeço à Universidade UniEVANGÉLICA e aos seus docentes que nos incentivaram a percorrer o caminho da pesquisa.

Natal Junio Barbosa De Sousa.

Agradeço primeiramente a minha mãe que sempre esteve ao meu lado e sempre me motivou a finalizar o curso, aos meus avós que me influenciaram a começar e terminar o curso de engenharia de Software e me auxiliaram bastante tanto em motivação quanto financeiramente. Agradeço também aos meus amigos que sempre estiveram ao meu lado para realização das atividades e também para auxílio de realização de trabalhos e estudo em conjunto proporcionando melhor proveito de cada matéria.

Patrick Alves Freitas.

RESUMO

Neste trabalho iremos tratar sobre a mineração de dados, a qual está presente em todos os lugares atualmente. Quando realizamos uma pesquisa de preço ou até mesmo o acesso em algum site, implicitamente já estamos participando da mineração de dados e contribuindo para a realização da coleta de informações e conseqüentemente já nos trazendo promoções e propagandas referentes a nossa pesquisa requisitada. Ela é feita de forma tão rápida atualmente que no mesmo momento em que pesquisamos as informações que tem ligação com a pesquisa já nos são mostradas. Essa mineração envolve vários fatores como a IA (Inteligência Artificial), que envolve o próprio raciocínio da máquina com o mínimo de intervenção humana possível. Um dos tipos de IA que serão utilizados no decorrer desse trabalho será a Inteligência Artificial Limitada (ANI), pois é uma classe de IA considerada fraca, porém possui uma memória limitada mais avançada, capaz de armazenar dados nas escolhas anteriores do usuário e logo é utilizada para tomar decisões. A *Big Data* será apresentado um pouco sobre estatística onde é tratado sobre a melhoria de consulta de tabelas e análise de gráficos, utilizaremos a *Big Data* a favor da grande consulta de dados para otimizar e organizar os resultados obtidos, trabalharemos bastante com a ferramenta dos 5V's. Além destes requisitos este trabalho apresentará algumas ferramentas mais utilizadas para a mineração como a linguagem Python, por exemplo, que se trata de uma linguagem tipada no mundo da programação em que as variáveis são definidas em tempo de execução, e também o PostgreSQL que se trata de uma ferramenta de análise de dados fornecidos em tabelas. Os objetivos desse trabalho se resumem em trabalhar com a análise de dados obtendo conhecimento sobre a linguagem Python e agregando melhoria no conhecimento da utilização do PostgreSQL por se tratar de uma banco de dados mais utilizado hoje em dia e também por possuir uma grande base de conhecimento que podemos consultar a qualquer momento e fazer uma tratativa um pouco mais avançada com os resultados obtidos podendo estabelecer algumas teorias de melhoria ou até mesmo “prever” algumas situações do *E-commerce* e também tentar “prever” alguma melhoria futura na situação financeira.

Palavras-chave: *Big data, mineração de dados, Inteligência artificial, PostgreSQL.*

ABSTRACT

In this work we will deal with data mining, which is present everywhere today. When we conduct a price survey or even access a website, we are implicitly already participating in data mining and contributing to the collection of information and consequently already bringing us promotions and advertisements related to our requested survey. It is done so quickly these days that at the same time we search, the information that is linked to the search is already shown to us. This mining involves several factors such as AI (Artificial Intelligence), which involves the machine's own reasoning with as little human intervention as possible. Then there is the support of Big Data, which is well recognized in data mining for its five V's that represent data queries in a more assertive and faster way. And for a Big Data add-on, a little bit about statistics will be presented, where it is treated about improving table query and graph analysis. In addition to these requirements, this work will present some commonly used tools for mining, such as the Python language, for example, which is a typed language in the world of programming in which variables are defined at runtime, and also PostgreSQL, which it is about. of an analysis tool for data provided in tables. The objectives of this work are summarized in working with data analysis, obtaining knowledge about the Python language and adding an improvement in the knowledge of the use of PostgreSQL and doing a little more advanced discussion with the results obtained, being able to establish some improvement theories or even " "predict" some E-commerce situations and also try to "predict" some future improvement in the financial situation.

Keywords: *Big data, Data mining, Artificial intelligence, PostgreSQL.*

LISTA DE ABREVIATURAS E SIGLAS

B2B	<i>Business to Business</i>
B2C	<i>Business to Consumer</i>
B2E	<i>Business to Employee</i>
B2G	<i>Business to Government</i>
C2B	<i>Consumer to Business</i>
C2C	<i>Consumer to Consumer</i>
C2G	<i>Citizen to Government</i>
EDI	<i>Electronic Data Interchange</i>
EFT	Transferência Eletrônica de Fundos
IA	Inteligência Artificial
KDD	<i>Knowledge Discovery in Databases</i>
SGBD	<i>Data base Management System</i>
SQL	<i>Structured Query Language</i>
VRM	<i>Visitor Relationship Management</i>

LISTA DE FIGURAS

Figura	1-	Etapas	Operacionais	do	Processo	
KDD.....						13

SUMÁRIO

INTRODUÇÃO.....	9
PROBLEMA DE PESQUISA.....	10
OBJETIVOS.....	11
OBJETIVOS ESPECÍFICOS.....	11
JUSTIFICATIVA.....	11
1. FUNDAMENTAÇÃO TEÓRICA.....	12
1.1. MINERAÇÃO DE DADOS E KNOWLEDGE DISCOVERY IN DATABASES - KDD.....	12
1.2. INTELIGÊNCIA ARTIFICIAL / MACHINE LEARNING.....	14
1.3. BIG DATA.....	16
1.4. PYTHON.....	18
1.5. AMAZON.....	19
2. METODOLOGIA DA PESQUISA.....	20
2.1 PROCEDIMENTOS METODOLÓGICOS (Referencial Teórico).....	21
3. RESULTADOS.....	22
REFERÊNCIAS BIBLIOGRÁFICAS.....	2

INTRODUÇÃO

O *E-commerce* nasceu em meados de 1970 nos Estados Unidos, como uma troca de arquivos e solicitações de pedidos. No Brasil o primeiro registro de *E-commerce* foi diretamente de uma grande livraria em 1996, porém, muitas pessoas acreditam que o mesmo teve início em 1999 com o site submarino, que de acordo com uma pesquisa do G1 em 2014, consta dentre os 50 maiores sites de *E-commerce* no ranking mundial com um faturamento de cerca US\$ 2,477 bilhões.

É importante ressaltar que o *E-commerce* vem sendo agregado com um crescimento contínuo desde 2011. E ao chegar no ano de 2019, mais especificamente a partir de novembro, as vendas por aplicativos mobile ultrapassaram as de desktop, pois, os sites de busca e redes sociais são os principais caminhos para as lojas e também foi comprovado por uma análise feita da E-bit | Nielsen em 2020 de que a *Black Friday* disparou na frente do natal que se consolidava a data mais importante do *E-commerce*.

Os impactos COVID-19, as transações online têm tido um grande aumento devido às modificações feitas no horário comercial e também às métricas aplicadas para quem vai fazer uma compra presencial. Podemos observar que ao longo da pandemia que teve início com grande impacto no Brasil a partir de março de 2020, teve como resultado a aceleração de uma tendência que estava sendo trabalhada a pouco mais de uma década (Alessandro Silveira,2021).

Este trabalho tem como objetivo analisar a implicação da COVID-19 nas transações online, visto que queremos tratar sobre o impacto dele perante a comercialização online, e o quanto foi afetado essa demanda no mercado brasileiro do período de 2018 até o ano de 2021. A análise será feita perante dados de vendas da Amazon, sendo livros, produtos eletrônicos, consumíveis, dentre outros que são normalmente fornecidos pela empresa Kaggle que se trata de um site que disponibiliza databases em XLS online.

Nesta pesquisa temos alguns objetivos, tratar sobre o uso da mineração de dados para aplicar uma análise perante transações online em geral voltada para compras e vendas e assim acabar tendo como resultado esperado uma análise do *E-commerce* brasileiro antes e depois da pandemia chegar ao Brasil.

4

1.1 Problema da pesquisa

A busca da competência e vantagem competitiva é premissa básica para as empresas que pretendem alcançar o sucesso. Sendo assim, acabar descobrindo formas como as empresas se desenvolvem e mantêm suas vantagens competitivas é um item central na teoria administrativa (Oliveira JR, 1999). Ao longo da pandemia, que iniciou com força no Brasil a partir de março de 2020, e sem dúvidas impactou em um grande aumento nos números de transações online, porém, o isolamento não foi a única explicação para o impacto que temos hoje, pois, na verdade apenas acelerou uma tendência que vinha sendo desenhada há pelo menos uma década (Alessandro Silveira, 2021).

Baseando-se em dados divulgados pela empresa Ebit Nilsen, as datas comemorativas tiveram grande impacto no *E-commerce* no ano de 2018, foram os dias da famosa *Black Friday* e em seguida do Natal. Apesar de não ser nenhuma grande surpresa, porém, são épocas que de forma tradicional, acabam agregando maior movimentação de compras e vendas. De acordo com uma pesquisa realizada pela Ebit Nilsen – plataforma de opinião de consumidores do Brasil, foi revelado que o faturamento da *Black Friday* de 2018 ultrapassou o esperado. Devido ao relatório gerado, os pedidos feitos até as 17h da sexta-feira tinham somado um valor de cerca de R\$2,1 bilhões, o que representa um aumento de até 27% nas transações do ano anterior (Ebit;Nilsen, 2018).

Com base em um relatório gerado pelo Movimento Compre e Confie, tivemos um crescimento de vendas de produtos no destaque de saúde atingindo cerca de 47,64% apenas em Gel Antisséptico, mais conhecido como Álcool em gel.

Baseando na atual situação da pandemia, esta pesquisa tende a tratar sobre a seguinte questão: Considerando os dados do *E-commerce* da Amazon de 2018, qual foi o aumento da demanda deste mercado com a chegada da pandemia no Brasil?

Objetivos

Objetivo Geral: Aplicar a mineração de dados para obter resultados futuros de transações online da Amazon e através disso buscar e analisar a diferença dessas transações no período antes e depois da pandemia.

Objetivos Específicos:

- Selecionar as ferramentas necessárias para a mineração dos dados.
- Selecionar as bases de dados que serão utilizadas para o estudo.
- Documentar o processo e o desenvolvimento da mineração e a linguagem utilizada.
- Aplicar o conhecimento adquirido no mercado de trabalho.
- Realizar a comparação do impacto causado durante o período pesquisado.

Justificativa

Com a modernização das tecnologias e as suas evoluções cada vez mais avançadas, visto que é fundamental a importância que as organizações do conhecimento disponham de técnicas e ferramentas para análise de dados e de informações, criadas para suportar as decisões estratégicas, táticas e operacionais. Nesse aspecto, a mineração de dados contribui com as descobertas de conhecimentos, pois através de técnicas e ferramentas, ajudam a buscar correlações importantes entre os dados (FAYYAD et al., 1996).

As técnicas de mineração de dados não podem substituir o papel significativo dos especialistas em domínio e seu conhecimento comercial. Porém, pode-se obter resultados úteis combinando com essas técnicas. Como por exemplo, combinar experiência pessoal no

campo ou informações de negócios com um modelo de mineração de dados para gerar resultados mais bem-sucedidos. Além disso, esses resultados devem ser sempre avaliados por especialistas. Assim, os conhecimentos do negócio podem ajudar e enriquecer os resultados da mineração de dados (ZIAFAT; SHAKERI, 2014).

No início do ano de 2020 a projeção do Ministério da Economia para o PIB (Produto Interno Bruto) de 2020, estimou um crescimento de 2,4%, esta informação se encontra presente no Boletim Macro Fiscal que foi divulgado pela secretaria Econômica, levando em conta que esta estimativa foi realizada antes de imaginar os impactos causados pelo COVID (Ministério da Economia, 2020).

Baseando em uma pesquisa feita pela empresa CIELO, o impacto do COVID-19 no varejo brasileiro, teve principal efeito na comercialização de produtos vendidos em drogarias/farmácias estimando um crescimento de até 16,7% e também em setores de supermercados, tendo um crescimento de até 13,6%, já em questão de outros meios, como transportes e turismo, acabou sofrendo uma queda de até 41,0% entre os meses de fevereiro e março de 2020 (Cielo, 2020).

O Movimento Compre e Confie, também gerou um relatório baseando as vendas de alguns produtos da categoria saúde, e estimou um crescimento de até 210% no produto Gel Antisséptico, comparando nos meses de fevereiro e março de 2019 para 2020 (Movimento Compre e Confie, 2020).

A partir destes fatos, a aplicação deste estudo pretende contribuir para examinar quão impactante foi a diferença de transações do site da Amazon brasileira no ano de 2018 e 2019 com a chegada da pandemia do COVID-19 no Brasil no ano de 2020, e com os resultados obtidos analisarmos o crescimento nas transações online da amazon e o quanto a pandemia teve influência perante ao mesmo, todavia com os resultados em mãos elaborar algumas sugestões que poderiam influenciar na melhoria de transações online e como compensação atribuir na melhoria da economia.

1. FUNDAMENTAÇÃO TEÓRICA

1.1 Mineração de dados e Knowledge Discovery in Databases - KDD

O termo *Knowledge Discovery in Databases - KDD* foi formalizado em 1989 em referência ao amplo conceito de procurar conhecimento a partir de bases de dados. Uma das definições mais populares foi proposta em 1996, por um grupo de pesquisadores (FAYYAD et

al., 1996). É um processo não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados.

O KDD é constituído de três passos básicos que são o pré-processamento, a mineração propriamente dita e o pós-processamento ou interpretação dos resultados, A figura 1 ilustra uma configuração resumida de como funciona esse processo.

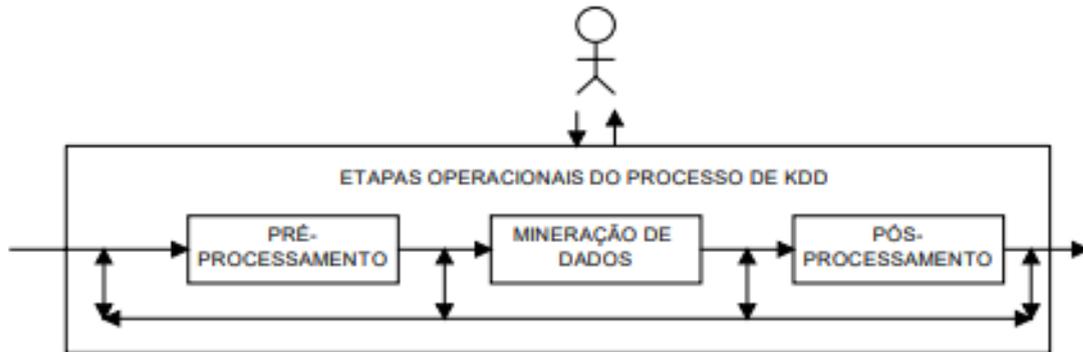


Figura 1: Etapas Operacionais do Processo KDD

Fonte: Simpósio sobre uma metodologia de suporte ao processo de descoberta de conhecimento em base de dados (BOENTE, A. N. P.; GOLDSCHMIDT, R. R.; ESTRELA, V. V.)

A mineração de dados surgiu por volta de 1990 com base em estatísticas e inteligência artificial. Após a sua origem, toda empresa grande e que tem em vista o desenvolvimento de seus trabalhos começou a utilizar desta ferramenta. A técnica da mineração de dados – do Inglês *Data Mining* – é o processo de descoberta de padrões interessantes, inovadores e desconhecidos assim como de modelos descritivos, compreensíveis e preditivos a partir de dados em grande escala (ZAKY; MEIRA JUNIOR. 2014).

O uso da informação de maneira eficaz e eficiente se torna um elemento essencial para o sucesso das organizações, sendo até incorporado a seu patrimônio. Saber que a informação é um dos principais recursos estratégicos que a organização dispõe, requer que estas informações estejam estruturadas, disponíveis e sejam íntegras, condições que apenas se fazem possíveis com o uso de tecnologias computacionais, frequentemente conhecidas como Tecnologia da Informação e Comunicação, ou Sistemas de Informação (DALFOVO, 2007).

Com o avanço da tecnologia uma das consequências é a quantidade e variedade de informações que são geradas em uma gigantesca escala. São várias as aplicações. Atualmente todas as pesquisas realizadas em sites da web geram dados, principalmente sites de comércio eletrônico, tendo em vista essa quantidade de dados, foi desenvolvido o método de mineração

de dados para que pudesse ser extraído e retirado informações as quais as empresas poderiam se beneficiar e verificar os principais pontos que a empresa tem a melhorar e a focar para o crescimento dela.

A seguir, são relacionadas as principais áreas de interesse na utilização de mineração de dados:

- Marketing: Técnicas de mineração de dados são aplicadas para descobrir preferências do consumidor e padrões de compra, com o objetivo de realizar marketing direto de produtos e ofertas promocionais, de acordo com o perfil do consumidor.
- Detecção de fraudes: Muitas fraudes óbvias (tais como, a compensação de cheque por pessoas falecidas) podem ser encontradas sem mineração de dados, mas padrões mais sutis de fraude podem ser difíceis de ser detectados, por exemplo, o desenvolvimento de modelos que predizem quem será um bom cliente ou aquele que poderá se tornar inadimplente em seus pagamentos.
- Medicina: Caracterizar comportamento de paciente para prever visitas, identificar terapias médicas de sucesso para diferentes doenças, busca por padrões de novas doenças.
- Instituições governamentais: Descoberta de padrões para melhorar as coletas de taxas ou descobrir fraudes.
- Ciência: Técnicas de mineração de dados podem ajudar cientistas em suas pesquisas, por exemplo, encontrar padrões em estruturas moleculares, dados genéticos, mudanças globais de clima, oferecendo conclusões valiosas rapidamente.
- Controle de processos e controle de qualidade: Auxiliar no planejamento estratégico de linhas de produção e buscar por padrões de condições físicas na embalagem e armazenamento de produtos.

O funcionamento da mineração de dados se divide em três etapas. A primeira etapa é redução da duplicidade de informações, dessa forma fazer a coleta de vários dados, muitas vezes não somente dos dados da própria empresa, mas de outras também, mas junto disso realizar um trabalho minucioso para não haver dados duplicados dentro da sua pesquisa para não gerar conflito. Na segunda etapa temos a realização da mineração dos dados, para esta etapa existem várias técnicas para fazer a realização da mineração dos dados, a maioria utiliza os critérios de manipulação. Alguns dos métodos são: O método da classificação e a organização dos dados tendo base em um critério específico. A agregação é o método que

utiliza um conjunto de informações diferentes e adiciona outros dados, para assim compor um resultado específico e complementar. O *smoothing* ou suavização faz a eliminação dos dados que não são necessários para alcançar o objetivo estabelecido. A *clustering* faz a identificação dos problemas que se assemelham. E por último temos a generalização que é uma técnica mais genérica, para estatísticas não muito minuciosas e mais simples de serem feitas. Na terceira e última etapa temos a definição do modelo que será apresentado, desta forma é feita a modelagem dos dados coletados para que seja apresentado da melhor forma. Essa definição deve estar de acordo com a feita no início do projeto para que não haja incoerências no projeto apresentado e buscando a aprovação das pessoas envolvidas para que assim sejam adotados novos trajetos para as empresas e para os negócios.

(**Metodologia de Pesquisa para Ciência da Computação**.)-

1.2 Inteligência Artificial

Aprendizado de máquina (em inglês, *machine learning*), foi criado com a fundamentação para automatizar a construção de modelos analíticos de dados. É um ramo que se trata da inteligência artificial (IA) baseando-se na ideia onde sistemas podem aprender com dados e acabar identificando padrões e também tomar decisões com o mínimo de intervenção humanizada possível. “Inteligência computacional é o estudo do projeto de agentes inteligentes.” (POOLE OOLE et al., 1998).

O famoso teste de Turing, proposto por Alan Turing, foi projetado para fornecer uma definição operacional satisfatória de inteligência, e um dos grandes fatores de evolução da inteligência das máquinas devido utilizar o processamento de linguagem natural, que permite que a máquina se comunique com sucesso em um idioma natural; também permitiu a capacidade de raciocínio automatizado para armazenar as informações com finalidade de responder as perguntas e tirar novas conclusões; e por fim o aprendizado de máquina, para se adaptar a novas circunstâncias e para detectar e extrapolar padrões. (Norvig, Peter).

Ainda sobre o teste de Turing, foi uma máquina que evitou deliberadamente a intenção física direta entre o interrogador e o computador porque a simulação física de uma pessoa é desnecessária para a inteligência. Entretanto tivemos também o teste de Turing total, onde teve a inclusão de um sinal de vídeo, de forma que o interrogador possa testar as habilidades de percepção do indivíduo, além de oferecer ao interrogador a oportunidade de repassar objetos físicos “pela janela” para ser aprovado no teste de Turing total, o computador acaba precisando de visão computacional para perceber objetos e também a robótica para conseguir manipular tais objetos e movimentar-se. (Norvig, Peter; Russel, Stuart).

O volume de dados é um dos grandes empecilhos para que se possa analisar os dados em tempo hábil e encontrar uma solução válida. Apesar da popularidade dos sistemas especialistas a partir da década de 1970, em que eram mescladas as habilidades de programadores e indivíduos amplamente conhecidos, saber na área do problema que se pretendia solucionar, já não era possível acompanhar a resolução de problemas modernos devido a quantidade de demandas oferecidas por um curto período de tempo (Faceli et al., 2011).

De acordo com um levantamento feito por Frank Witten em 2005, estima-se que o volume de dados digitais dobra a cada vinte meses. Além disso, a possibilidade de aprender de forma contínua permitiu o desenvolvimento de novas aplicações que auxiliam desde projetos de pesquisa médica na busca por padrões em diagnósticos diversos até o controle de veículos autônomos bem mais complexos que os sistemas que orientam a navegação do usuário.

A máquina precisa ser programada, de forma que deva proceder com as informações que recebe. Para tanto, dispõe de um conjunto rígido de instruções para que quando organizados de forma lógica possam efetuar o trabalho esperado. Enquanto o homem toma decisões e analisa o mundo ao seu redor a partir de regras nem sempre tão fáceis de identificar, a máquina já pode processar quantidades elevadas de informações e ela depende diretamente apenas da qualidade do algoritmo desenvolvido e das ferramentas associadas a ela. Logo, podemos chegar em uma breve conclusão sobre a diferença entre a máquina eficiente e outra ineficiente, pois, tudo vai depender da habilidade do desenvolvedor (Faceli et al., 2011).

A inteligência artificial visa à implementação de rotinas em máquinas que possam aproximá-las do comportamento utilizado por agentes inteligentes na tomada de decisões. Assim, o aprendizado de máquina acaba sendo um ramo da IA cujo objetivo é habilitar um programa para melhorar seu desempenho na realização de determinada tarefa a partir da experiência obtida nas execuções anteriores (Mitchell, 1997).

1.3 Big Data

A definição de *Big Data*, são dados com maior variedade que chegam em volumes crescentes e com velocidade cada vez maior. Simplificando, *Big Data* é um conjunto de dados maior e mais complexo, especialmente de novas fontes de dados. Esses conjuntos de dados são tão volumosos que o *software* tradicional de processamento de dados simplesmente não consegue gerenciá-los. (Oracle, 2020).

Embora o conceito de *Big Data* em si seja realmente novo, as origens de grandes conjuntos de dados remontam às décadas de 1960 e 1970, quando o mundo dos dados estava apenas começando, com os primeiros *data centers* e o desenvolvimento do banco de dados relacional. Por volta de 2005, as pessoas começaram a perceber a quantidade de usuários de dados gerados pelo *Facebook*, *Youtube* e outros serviços online. O *Hadoop* (uma estrutura de código aberto criada especificamente para armazenar e analisar grandes conjuntos de dados) foi desenvolvido no mesmo ano. O NoSQL (termo genérico que representa os bancos de dados não relacionais) também começou a ganhar popularidade durante esse período. (Oracle, 2020).

A *Big Data* é estruturada seguindo os conceitos dos 5 Vs, apresentados com base em Barbieri (2011).

O primeiro V refere-se ao termo volume que se refere ao conceito principal da *Big Data*, em que a quantidade de dados a ser coletada e tratada representa um grande volume. Esses dados são advindos de diversas fontes, como redes sociais, motores de busca da internet, *E-commerce*, entre outras. Este talvez tenha sido o grande diferencial em relação aos outros bancos de dados: o tratamento de grandes volumes de informações de dados, com formas diferentes (estruturados, não estruturados e semiestruturados).

Já o segundo V, diz respeito a velocidade, como os dados são gerados em grandes volumes, o seu processamento deve efetuar o tratamento da mesma forma, caso contrário, ocorrem filas e a consequente degradação do serviço. Para tanto, o servidor (*hardware*) deve possuir uma capacidade de processamento adequada às necessidades. O SGBD (Sistema de Gerenciamento de Banco de Dados) também deve permitir que grande quantidade de dados seja recebida e processada. SQL -> citar os dados - >

Em seguida, o terceiro v se refere à variedade, os bancos de dados que trabalham dentro do conceito de *Big Data* devem possuir técnicas que possibilitem o tratamento dos mais diferentes tipos de dados, como números, tags, imagens e textos.

O quarto v, é voltado para a veracidade, onde acontece a necessidade de constante análise em tempo real, isso significa de dados que condizem com a realidade daquele momento, pois dados passados não podem ser considerados dados verídicos para o momento em que é analisado.

E por fim temos o valor, que mostra a significância dos dados coletados e tratados para a organização, de forma que possam trazer informações relevantes, que possam proporcionar um diferencial de mercado ou, ainda, auxiliar os gestores na tomada de decisão. Esse último ponto das características da *Big Data* é exatamente o que as empresas buscam e

precisam, pois os dados de valor podem ser convertidos em informações importantes para a companhia.

Mas, de que forma que os 5 Vs, de fato, são utilizados na aplicação da *Big Data*? Inicialmente, deve-se analisar o cenário no qual será aplicado e utilizado a *Big Data*. Com isso, será possível compreender o volume e a velocidade com que esses dados são gerados, sendo necessário observar a variedade dos dados e utilizar filtros para se garantir a veracidade desses dados coletados. Por fim, deve-se garantir que esses dados sejam de grande significância e valor para a empresa. (Barbieri, 2011).

A *Big Data* surgiu como uma tecnologia inovadora e com alto potencial para auxílio no crescimento das empresas para que possam garantir seu espaço no mercado que vem com a tendência de ficar cada vez mais competitivo. Existem alguns segmentos nos quais o cenário da *Big Data* pode ser aplicado para gerar mais lucro e assertividade no ramo empresarial, onde temos os comportamentos e tendências, a *Big Data* permite compreender o comportamento dos consumidores e a tendência de consumo, apontando quais produtos ou serviços foram mais relevantes em um período de tempo. (Dumbill,2013).

Ainda no ramo empresarial contamos com a estratégia de *marketing*, onde grande parte da aplicação comercial da *Big Data* está ligada às estratégias de *marketing*, pois ele permite realizar análises de dados e direcionar as ofertas de produtos e serviços de forma mais assertiva.

Melhoria de produtos e serviços, com a estratégia de *marketing* os *feedbacks* dos consumidores tendem a serem mais fornecidos em redes sociais, melhorando assim o retorno dos produtos oferecidos pela empresa.

A *Big Data* também oferece melhoria na área da saúde, como medicina de precisão que visa na correção das deficiências na prescrição de medicamentos que funcionam bem para determinados pacientes e para outros não. Além disso, já vem sendo utilizada a fusão entre *Big Data*, internet das coisas e aprendizagem de máquina para redução de acidentes vasculares, diabetes dentre outras várias doenças. Prontuários eletrônicos, esse tipo de prontuário já é utilizado em grande parte do sistema de saúde, no entanto, é bem pouco explorado pela *Big Data*. Essa ferramenta computacional auxilia os profissionais da saúde na prescrição de medicamentos e diagnósticos baseados em históricos baseados em diferentes documentos em vários centros de saúde. (Chiavegatto Filho, 2015).

1.4 Python

O Python foi criado por Guido Van Rossum em 1991, tendo como origem de seu nome a série humorística britânica *Monty Python's Flying Circus*, e atualmente tem conquistado bastante espaço entre as outras ferramentas de programação, por ter uma interface “amigável” de fácil aprendizagem e pela sua grande aplicabilidade. Atualmente o Python é utilizado por grandes empresas de tecnologia, como a *Google*, *Microsoft*, *Instagram*, *Spotify* e várias outras. (SANTANA; GALES, 2010). A linguagem de programação Python é uma linguagem de altíssimo nível, de tipagem dinâmica e forte, iterativa e interpretada e orientada a objeto. Ela possui uma sintaxe clara e concisa, favorecendo a legibilidade do código fonte e com isso faz com que a linguagem seja mais produtiva (BORGES, 2010). Existem muitas ferramentas de desenvolvimento para Python, como IDEs, editores e *shells* (que aproveitam da capacidade interativa do Python).

Com o Colab, é possível importar um conjunto de dados de imagem, treinar um classificador de imagens dentro dele e avaliar o modelo, tudo com apenas algumas linhas de código. É uma ferramenta que permite que você misture código fonte (geralmente em *python*) e texto rico (geralmente em *markdown*) com imagens e o resultado desse código. É uma técnica conhecida como *notebook* (“caderno”). Tudo isso em um ambiente colaborativo, que você pode compartilhar com seus colegas, permitindo que outros rodem seu código e até modifiquem criando suas próprias versões. Os *notebooks* foram fortemente inspirados nos *notebooks* do Jupyter e utilizam essa ferramenta por trás dos panos. Ela trabalha em especial com a linguagem Python, mas com alguns ajustes é possível rodar código em R, Julia, *Swift* e outras linguagens. Os *notebooks* do Colab executam código dos servidores em nuvem do *Google*. Isso significa que você pode tirar proveito da potência de *hardware* do *Google*, como GPUs e TPUs, independentemente da potência da sua máquina. Você só precisa de um navegador. É possível importar para os notebooks do Colab os dados da sua conta do *Google Drive*, como planilhas. Também é possível importar do *GitHub* e de muitas outras fontes (GOOGLE RESEARCH, 2014).

O Python possui uma sintaxe clara e concisa, que favorece a legibilidade do código fonte, tornando a linguagem mais produtiva. A linguagem inclui diversas estruturas de alto nível (listas, tuplas, dicionários, data / hora, complexos e outras) e uma vasta coleção de módulos prontos para uso, além de *frameworks* de terceiros que podem ser adicionados. Também possui recursos encontrados em outras linguagens modernas, tais como: geradores, introspecção, persistência, meta classes e unidades de teste. Multiparadigma, a linguagem suporta programação modular e funcional, além da orientação a objetos. Mesmo os tipos básicos no Python são objetos. A linguagem é interpretada através de *bytecode* pela máquina

virtual Python, tornando o código portátil. Com isso é possível compilar aplicações em uma plataforma e rodar em outras ou executar direto do código fonte (LUIZ E. BORGES, 2009).

Essa linguagem é muito utilizada como linguagem *script* em vários *softwares*, permitindo automatizar tarefas e adicionar novas funcionalidades, entre eles: *BrOffice.org*, PostgreSQL, *Blender* e GIMP. Também é possível integrar o Python a outras linguagens, como a Linguagem C. Em termos gerais, o Python apresenta muitas similaridades com outras linguagens dinâmicas, como *Perl* e *Ruby*. Em SQL o processo de desenvolvimento ocorre em fases. Se cada fase for abordada de forma consciente e os resultados forem devidamente gerados, você desenvolverá um ótimo site de *E-commerce* de maneira mais eficaz possível. Se, por outro lado, você realizar tarefas aleatoriamente, apressar o processo, pular passos e omitir outros, o procedimento como um todo exigirá muito mais tempo e o resultado apresentará mais *bugs*. No final do processo de desenvolvimento, espera-se que o melhor site possível de *E-commerce* tenha sido criado, porém esse site, sem dúvida, terá de ser alterado na próxima semana (como os clientes sempre querem), no próximo mês ou no próximo ano. Se a primeira meta é ter um processo tranquilo e otimizado, a segunda consiste em obter um resultado (LARRY UIMAN, 2014).

1.5 Amazon

A Amazon.com, criada nos EUA por Jeff Bezos em 1994, nos primórdios do *E-commerce*. Naquela época, muita gente duvidava que o comércio pela Internet pudesse ser rentável. Tudo começou a partir de um estudo sobre o hábito de compras dos norte-americanos. Jeff concluiu que os livros figuravam como o segundo item mais comprado pelos consumidores, mas que, curiosamente, não eram vendidos pelo correio porque não havia como enviar catálogos que abrangessem os mais variados interesses das pessoas. Um catálogo desses seria tão pesado que tornaria inviável seu envio pelo correio. Assim, ele teve a ideia de comercializar esses produtos pela *Internet*, considerando que, além de não haver limitação de peso ou tamanho no ambiente digital, o catálogo de livros poderia ficar disponível para os possíveis compradores 24 horas por dia. E foi assim que começou a maior empresa de *E-commerce* do mundo. A Amazon é um exemplo de varejo puramente virtual. (TURCHI, S. 2018).

No momento atual, a Amazon trabalha com inúmeros produtos, contendo eletrônicos, vestuários, acessórios e entre outros. E também oferece serviços como a *Amazon Web Services* (serviços de armazenamento em nuvem), *Amazon Prime (streaming)* e *Alexa*

(assistente virtual). Segundo a companhia, o essencial é entender a demanda de seus usuários. O diferencial é ser voltada para o cliente, para o longo prazo, e inventar é inovar. A maioria das empresas não são assim. Elas focam no concorrente, não no cliente. E preferem seguir as tendências de perto em vez de inventá-las, porque é mais seguro. (BRIAN DUMAINE, 2020)

Uma das principais estratégias que fizeram o sucesso da Amazon foi o uso inteligente dos dados capturados dos clientes e visitantes do site para fazer recomendações de produtos de acordo com o perfil de navegação de cada um, além do registro de compras, uma estratégia denominada *VRM – Visitor Relationship Management*. (TURCHI, S. 2018).

2. METODOLOGIA DA PESQUISA

O fundamento deste tema foi realizado na área de Banco de Dados. Para considerar a Mineração de Dados para Verificação de Transações do *E-Commerce*, foi utilizada a pesquisa exploratória que será utilizada para buscar mais informações sobre o tema e fazer um levantamento de dados para tornar o nosso tema mais compreensível, e a pesquisa descritiva que será descrito o que será observado, ou seja, os dados que serão analisados conforme a disponibilidade das tabelas oferecidas.

É proposto um estudo categorizado como um estudo de caso, cujo objeto é a Verificação de Transações do *E-Commerce*. Segundo Gilberto de Andrade, o estudo de caso deve apresentar indicadores de confiabilidade dos instrumentos de coleta de dados utilizados. Uma grande utilidade dos estudos de caso é verificada nas pesquisas exploratórias. Por sua flexibilidade, é recomendável nas fases iniciais de uma investigação sobre temas complexos, para a construção de hipóteses ou reformulação do problema. (GIL 2008)

Tendo isso, utilizaremos a ferramenta Jupyter para a linguagem Python por ser programação estruturada, codificação orientada a objetos, além de ser uma linguagem de programação funcional e ter outras funções.

Simplicidade com robustez: Python permite o desenvolvimento de grandes projetos que podem ser constituídos por diversos módulos, que acessem bancos de dados, que enviem e recebam dados por meio de redes, trabalhem com recursos multimídia, entre outros. (Banin, Sérgio L, 2018).

Grande aplicabilidade: o Python pode ser utilizado em um grande número de áreas do desenvolvimento de software, das quais se destacam: ferramentas para administração e interface com sistemas operacionais; aplicações que trabalhem com grandes volumes de dados armazenados em sistemas gerenciadores de bancos de dados. (Banin, Sérgio L, 2018).

A ferramenta PostgreSQL também será utilizada por ser um sistema que lida bem com altos volumes de solicitações e com cargas de trabalho grandes e funciona muito bem para sites com intensidade de acesso. Além de ter a função de gerenciar os dados desses bancos de maneira ordenada e eficiente, rodando e gravando todas as informações que ficam registradas nesses compartimentos.

O PostgreSQL, tem forte apelo para aplicações WEB de pequeno e médio porte. Grande parte dos sites e sistemas do tipo WEB utilizam um banco de dados PostgreSQL para armazenamento dos dados. (Barboza, F.F. M., & Freitas, P.H. C, 2018).

2.1 Procedimentos metodológicos

Com o desenvolvimento da *Internet*, o ambiente digital passou a ser visto pelo setor corporativo como um terreno fértil a ser explorado e como forma de imprimir maior agilidade aos negócios. Na década de 1980, as grandes corporações, principalmente dos setores industrial e varejista, já realizavam transações comerciais pela via eletrônica, com o uso de tecnologias específicas, como o *Electronic Data Interchange* (EDI – Troca Eletrônica de Documentos) e *Electronic Funds Transfer* (EFT – Transferência Eletrônica de Fundos) para envio de documentos (ordens de compra e de pagamentos), visando agilizar as operações de logística e da cadeia de suprimentos. (TURCHI, S. 2018).

A partir do final dos anos 1990 e início de 2000, com a evolução da *Internet* e dos mecanismos de segurança, essas operações começaram a ser feitas também pela *web*. Estavam lançadas as bases para o *E-commerce*, operações de compra e venda de produtos e de serviços pela *Internet*. (TURCHI, S. 2018).

Atualmente, o pequeno empresário pode iniciar um negócio com investimento mínimo, que pode ir aumentando no mesmo passo em que aumentam a sua base de clientes e o seu faturamento. Este é o grande diferencial do comércio eletrônico: uma microempresa ou mesmo um empreendedor individual pode começar as atividades com pouquíssimo recurso material, atendendo de imediato a clientes em qualquer parte do mundo. Assim, o canal de vendas deixa de ser um limitador para o negócio. (A. Wagner, 2018).

Por meio dessas interações entre a empresa e seus clientes, a *internet* se tornou a ferramenta mais utilizada para transações, como compras e divulgações de marcas, que podem ser realizadas a qualquer hora, lugar, sem restrição de data, hora e local. No caso de compra, os obstáculos são bastante reduzidos devido à falta de tempo e as compras *online* costumam ser a opção mais viável.

Segundo Wagner da Silveira, o *E-commerce* se subdivide em sete tipos: O primeiro tipo é *Business to Business (B2B)* empresas que vendem para outras empresas. O segundo tipo é *Business to Consumer (B2C)* envolve vendas diretas entre fabricantes e distribuidores ao consumidor final. O terceiro tipo é *Consumer to Consumer (C2C)* pessoa física exerce o papel de ofertante e outra pessoa representa o consumidor final. O quarto tipo é *Consumer to Business (C2B)* consumidor fornece seu produto para a empresa. O quinto tipo é *Business to Government (B2G)* são sites de pregão eletrônico, licitações públicas e leilão reverso. O sexto tipo é *Citizen to Government (C2G)* transações eletrônicas entre o cidadão e o governo, que são oferecidos eletronicamente alguns serviços ao usuário do serviço público. E por último, o sétimo tipo é *Business to Employee (B2E)* modelo de venda direta, em que as empresas vendem serviços ou produtos a seus funcionários.

Tomando em conta as considerações anteriores, compreendemos que o comércio eletrônico nada mais é do que uma extensão do comércio convencional, o qual existe há milênios. Num primeiro momento realizado por meio de trocas in natura (produtos não industrializados); mais tarde com a aquisição de produtos utilizando outros como se fossem moedas, especialmente o sal e o gado. Posteriormente pela aquisição de bens pela utilização de moedas metálicas e em papel, sem prejuízo da utilização do crédito para obtenção de produtos mediante pagamento ulterior, cujo uso foi massificado no comércio com a criação da letra de câmbio no século XI. (A LGPD E O *E-COMMERCE*; Teixeira, Tarcísio).

3. RESULTADOS ESPERADOS

Além do levantamento bibliográfico, esperamos obter resultados da frequência que as pessoas compraram durante a pandemia, utilizando a linguagem python e as bibliotecas utilizadas no programa. Assim, terá a facilidade de visualizar os resultados gráficos obtidos de forma extensa. Juntamente com a ferramenta PostgreSQL para ajudar a organizar e compreender as relações entre os dados.

Referências Bibliográficas

ASSUNÇÃO, W. Comércio Eletrônico. Grupo A, 2018.

AFFONSO, A. Como a Mineração de Dados Pode te Ajudar a Alcançar Melhores Resultados. Voitto, 2021. Disponível em: < <https://www.voitto.com.br/blog/artigo/mineracao-de-dados> >. Acesso em: 20/08/2021.

BOENTE, A. N. P.; GOLDSCHMIDT, R. R.; ESTRELA, V. V. Uma metodologia de suporte ao processo de descoberta de conhecimento em bases de dados. Disponível em: < <http://boente.eti.br/publica/seget2008kdd.pdf> > Acesso em: 10/12/2021.

Colab research. Colaboratory; O Colaboratory ou "Colab". Colab research, 2014. < https://colab.research.google.com/?utm_source=scs-index > Acesso em: 09/11/2021

DALFOVO, O. Modelo de integração de um sistema de inteligência competitiva com um sistema de gestão da informação e de conhecimento. UFSC; 2007.

DUMAINE, B. Bezonomics: Como a Amazon Está Mudando Nossas Vidas e o que as Melhores Empresas do Mundo Estão Aprendendo com ela. Alta Cult, 2020.

Ebit Nielsen. Plataforma de opinião de consumidores do Brasil. Disponível em:
< <https://www.ebit.com.br/> >.

E-commerce Brasil. E-commerce Brasileiro cresce 73,88% em 2020, revela índice MCC-ENET. E-commerce Brasil, 2021. Disponível em:
<<https://www.ecommercebrasil.com.br/artigos/comercio-eletronico-antes-e-depois-da-pandemia-do-coronavirus/>>

FERNANDO, A. Aprenda Mineração de dados: Teoria e Prática. Rio de Janeiro: Alta Books Editora, 2016.

GIBBS, G. Análise de dados Qualitativos. Porto Alegre: Artmed Editora S.A., 2009.

KUNIYOSHI, M. S. Comércio Eletrônico: A Revolução em Tempos Digitais. Revistas PUCSP, 2000. < <https://revistas.pucsp.br/index.php/rad/article/view/1689/1083> > . Acesso em: 01/10/2021.

LARRY, U. E-commerce com PHP e MySQL, Novatec Editora, 2014.

LUIZ, E. B. Introdução à Programação com Python: Algoritmos e Lógica de Programação Para Iniciantes, Novatec Editora, 2009.

MCC-ENET. Referência em métricas e indicadores do consumo online no Brasil. Disponível em
< <https://www.mccenet.com.br/> >.

MOREIRA, P. Comércio Eletrônico: Antes e Depois da Pandemia do Coronavírus. E-commerce Brasil, 2020. <<https://www.ecommercebrasil.com.br/artigos/comercio-eletronico-antes-e-depois-da-pandemia-do-coronavirus/>> . Acesso em: 25/09/2021.

NOLETO, C. Mineração de dados: O que é e como funciona o Data Mining. Blog da Trybe, 2021. < <https://blog.betrybe.com/tecnologia/mineracao-de-dados/> > Acesso em: 28/09/2021.

SAS Insights. Mineração de dados, o que é e qual sua importância? Disponível em:
<https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html#dmhistory%E2%80%AF>

TURCHI, S. R. Estratégia de Marketing Digital e E-Commerce, 2ª edição. Grupo GEN, 2018.

VIVEROS, M.S. et al. Applying data mining techniques to a health insurance information system. In: VLDB CONFERENCE, 22., 1996, Bombay. Proceedings... Bombay: IIT Bombay, 1996. p. 286-295.

ZAKI, M.; MEIRA JR.W. Data mining and analysis: fundamental concepts and algorithms. New York: Cambridge University Press, 2014.

ZIAFAT, H; SHAKERI, M. Using Data Mining Techniques in Customer Segmentation. Journal of Engineering Research and Applications, Volume 4, Fascículo 9, Página 70-79. Setembro 2014.

< https://www.ijera.com/papers/Vol4_issue9/Version%203/K49037079.pdf > .

Livro com até três autores:

BANIN, SÉRGIO L. *Python 3 - Conceitos e Aplicações - Uma abordagem didática*. Editora Saraiva, 2018.

OTAVIANO, A. M. H. VANIN R. V. Título: Tratamento Estatístico de Dados: Em Física Experimental. 2º Edição. São Paulo, 1991

Livro com quatro autores ou mais:

BARBOZA, FABRÍCIO FELIPE, M. e PEDRO HENRIQUE CHAGAS FREITAS. *Modelagem e desenvolvimento de banco de dados*, 2018.

VETORAZZO, Adriana de Souza et al; revisão técnica: MACHADO, Jeferson F. L. de Souza]. *Estrutura de Dados*. Porto Alegre: SAGAH, 2018.