



Fundação Educacional do Município de Assis
Instituto Municipal de Ensino Superior de Assis
Campus "José Santilli Sobrinho"

JOÃO HENRIQUE DE OLIVEIRA

A IMPORTÂNCIA DAS ANÁLISES DE BIG DATA PARA AS EMPRESAS

ASSIS/SP

2019



Fundação Educacional do Município de Assis
Instituto Municipal de Ensino Superior de Assis
Campus "José Santilli Sobrinho"

JOÃO HENRIQUE DE OLIVEIRA

A IMPORTÂNCIA DAS ANÁLISES DE BIG DATA PARA AS EMPRESAS

Projeto de pesquisa apresentado ao Curso de Ciência da Computação do Instituto Municipal de Ensino Superior de Assis – IMESA e a Fundação Educacional do Município de Assis – FEMA, como requisito parcial à obtenção do Certificado de Conclusão.

Orientando: JOÃO HENRIQUE DE OLIVEIRA

Orientador: ALEX SANDRO R. DE SOUZA POLETTO

ASSIS/SP

2019

FICHA CATALOGRÁFICA

OLIVEIRA, João Henrique de.

A importância das análises de Big Data para as empresas. / João Henrique de Oliveira. Fundação Educacional do Município de Assis – Assis, 2019.

43p.

Trabalho de conclusão do curso (Ciência da Computação). – Fundação Educacional do Município de Assis-FEMA.

Orientador: Dr. Alex Sandro Romeo de Souza Poletto.

1. Big Data 2. Análises-Big Data. 3. Ciência De Dados

CDD: 005.74
Biblioteca da FEMA

A IMPORTÂNCIA DAS ANÁLISES DE BIG DATA PARA AS EMPRESAS

JOÃO HENRIQUE DE OLIVEIRA

Trabalho de Conclusão de Curso apresentado ao Instituto Municipal de Ensino Superior de Assis, como requisito do Curso de Graduação, avaliado pela seguinte comissão examinadora:

Orientador: _____ Alex Sandro Romeo de Souza Poletto _____

Examinador: _____ Luiz Ricardo Begosso _____

DEDICATÓRIA

Dedico este trabalho a todos familiares e amigos que sempre estiveram ao meu lado nesta trajetória, em especial dedico aos meus pais João Aparecido e Maria Ângela e as minhas irmãs Amanda Luiza e Aline Natalina pelo apoio e força em todos esses anos. Dedico também a minha companheira Michelle Teixeira por me ajudar muito na elaboração deste trabalho, por sempre estar me apoiando em todos os meus objetivos e me ajudando a superar qualquer obstáculo e mostrar que sou capaz de tudo. Agradeço a Deus por ter me dado pessoas especiais com quem pude contar na realização deste trabalho, e pela sabedoria e perseverança que vem me dado a cada dia.

AGRADECIMENTOS

Primeiramente agradeço a Deus por mais esta conquista, também agradeço a todo o corpo docente da FEMA por todo o conhecimento que me foi ensinado durante esses anos, por possuir professores qualificados para tais ensinamentos. Agradeço a banca de avaliação da qualificação pelas observações que ajudaram em uma melhor realização deste trabalho. Agradeço e também dedico este trabalho ao orientador Alex Poletto pela ajuda, pelo apoio, pelo tempo destinado a correções, observações, e sugestões que ajudaram muito na realização deste trabalho desde o começo da elaboração do pré-projeto até a finalização deste trabalho.

RESUMO

A quantidade de dados criados em todo mundo é muito grande. Extrair, analisar, e compreender os dados é fundamental para qualquer empresa nos dias atuais. Quando se fala em análises de *Big Data* nas empresas, muitos outros temas podem ser relacionados, tanto para empresas que desejam aumentar seus lucros, como para empresas que desejam conhecer seus clientes, fazer promoções específicas para determinados grupos, obter *insights* mais assertivos, ajudar a tomada de decisão, prever fraudes, e até mesmo saber como está a concorrência, tudo isso devido as poderosas análises de *Big Data*. Neste trabalho, serão apresentados conceitos fundamentais para entender o que é a tecnologia das análises de *Big Data*, bem como sua importância para as empresas.

Palavras-chave: *Big Data*, Análise de Dados, *Insights*, Tomada de Decisão, Ciência de Dados.

ABSTRACT

The amount of data created so far worldwide is very large. Extracting, analyzing, and understanding data is critical to any business these days. When talking about Big Data analyzes in companies many other issues can be related to both companies that want to increase their profits and to companies that want to know better their customers, to make specific promotions for certain groups, to obtain more assertive insights, to help making, predicting fraud, and even knowing how the competition is, all due to the powerful analysis of Big Data. In this paper we will present fundamental concepts to understand the technology of the Big Data analyzes, as well as their importance for companies.

Keywords: *Big Data, Analyzes, Data Science, Insights, Analytics.*

LISTA DE ILUSTRAÇÕES

Figura 1: Crescimento dos Dados	15
Figura 2: 5V's	17
Figura 3: Gráfico de Pizza.....	29
Figura 4: Mapa Dinâmico	30
Figura 5: Proposta de trabalho.....	31
Figura 6: Dados	32

LISTA DE ABREVIATURAS E SIGLAS

IBM (International Business Machines)	15
IDC (International Data Corporation)	15

SUMÁRIO

1. INTRODUÇÃO	12
2. CONCEITOS E NOÇÕES SOBRE BIG DATA	14
2.1 O QUE É BIG DATA?	14
2.2 OS 5V's DO BIG DATA.....	15
2.3 AS ANÁLISES DE BIG DATA	17
2.4 TIPOS DE ANÁLISE DE BIG DATA.....	18
3. BIG DATA NAS EMPRESAS	21
3.1 EMPRESAS E INDUSTRIA 4.0	21
3.2 BIG DATA NAS REDES FINANCEIRAS.....	22
3.3 CASOS DE SUCESSO: OUTRAS EMPRESAS QUE USAM BIG DATA	25
4. LINGUAGEM R	28
5. PROPOSTA DE TRABALHO	31
6. EXEMPLO PRÁTICO	312
6.1 INSTRUÇÕES DE CÓDIGO	32
6.2 CONCLUSÃO	38
REFERÊNCIAS	39

1. INTRODUÇÃO

Devido à variedade de dispositivos existentes no mundo, o volume e o fluxo de dados crescem gradativamente, é neste contexto que se encontra o *Big Data*. Celulares, Tablets, Computadores, dados de redes sociais como Facebook, LinkedIn, WhatsApp e Instagram, produzem um grande volume de dados. Basicamente, a maior parte de nossas informações está presente na Internet: músicas, produtos, fotos, vídeos, dentre outros. Essa variedade de informações produz um grande volume de dados que possibilita gerar *insights* para as empresas e auxiliar a Tomada de Decisão.

As empresas que desejam permanecer adiante devem se tornar orientadas à dados. Os dados são o combustível para o sucesso de qualquer organização. Os *insights* obtidos de seus dados ajudam as organizações a inovarem e tomarem decisões mais inteligentes e assertivas com base em fatos e evidências, ao invés de usar a intuição. FARINELLI (2018)

Segundo FARINELLI (2018), as empresas que buscam permanecer no mercado devem usar os dados a seu favor para obter *insights*, e com esses *insights*, conhecer seus clientes a fim de obter uma maior satisfação sobre algum produto e até mesmo buscar novos clientes para a empresa. Esses *insights* podem ser obtidos através de uma análise de dados, e não mais pela intuição, o que torna a tomada de decisão mais assertiva.

Com o crescimento da Web e o uso de dados em grande escala, a quantidade de dados gerados e disponibilizados tem crescido exponencialmente, o que gera um círculo virtuoso de oferta e demanda. (ÁVILA 2017 *apud* ALBUQUERQUE 2017)

Um dos desafios da análise de *Big Data* é examinar dados de forma adequada, para gerar *insights*, e determinar quais dados são consistentes e relevantes e qual a melhor forma de analisá-los, possibilitando que a empresa tenha mais benefícios, como lucros maiores e até mesmo saber o que acontece nas empresas concorrentes. Um exemplo de sucesso com a análise de *Big Data* é a empresa *Netflix*, uma das plataformas de *streaming* de vídeo que mais faz o uso das técnicas de análise de *Big Data* e que cresce de forma exponencial desde que começou a analisar os dados dos usuários, recomendando filmes e séries de acordo com o gosto e perfil de seus assinantes.

Para SCHERMANN (2017), as análises de Big Data abordam outras áreas como Matemática, Estatística, Teoria da Informação, Inteligência Artificial para auxiliar as técnicas de Big Data.

Segundo LAUDON e LAUDON (2014), há pouco tempo atrás a maioria dos dados coletados pelas empresas eram transacionais e poderiam facilmente ser relacionados e organizados em programa de gestão, porém com a explosão dos dados gerados por qualquer dispositivo e em qualquer formato como por exemplo um e-mail ou conteúdo de redes sociais, tornou essa tarefa um pouco mais complexa do que antes. E ainda os autores destacam que o *Big Data* não se refere a qualquer quantidade específica de informação, mas sim ao tamanho dos seus dados que estão próximos ou superiores a *petabytes* e *exabytes*, sendo de bilhões a trilhões de registros e das mais variadas fontes que conhecemos.

Esta pesquisa tem como objetivo entender como as análises de *Big Data* ajudam empresas a conhecer e melhorar o relacionamento com seus clientes e como as análises de *Big Data* geram valiosos *insights* para as empresas.

Também será demonstrada, por meio do algoritmo de *Machine Learning Naive Bayes*, uma classificação de solicitação de crédito para determinar se o cliente será um bom ou um mal pagador com base nos padrões de um conjunto de dados com 1000 registros.

2. CONCEITOS E NOÇÕES SOBRE BIG DATA

Nesta seção será abordado conceitos básicos e noções que fundamentam o que se sabe sobre *Big Data*, bem como os 5V's que fazem parte do contexto de *Big Data*.

2.1. O QUE É BIG DATA?

Para que essa pesquisa siga adiante é preciso fundamentar o que é esta tecnologia conhecida como *Big Data*. De forma geral o termo *Big Data* refere-se ao grande volume de dados presente no nosso dia a dia. A qualquer momento as pessoas estão produzindo dados, seja enviando uma mensagem, compartilhando publicações e fotos em redes sociais, ou assistindo um vídeo.

De acordo a IBM o termo *Big Data* se refere ao vasto volume de dados cuja capacidade está além da interpretação dos bancos de dados relacionais, e possui características como alto volume, alta velocidade, alta variedade.

Neste contexto o *Big Data* pode ajudar gestores, analistas de negócios, gerentes de marketing a tomar decisões melhores e mais corretas para determinado problema ou situação a partir dos dados que antes poderiam ser inutilizáveis, sendo assim, a partir destas informações as empresas conseguem extrair poderosos *insights* e valores desses dados.

Um estudo feito pela IDC mostra que de 2012 a 2020, o volume de dados armazenados na internet deverá dobrar a cada dois anos. Ainda segundo o instituto, existe uma previsão de que até 2020 tenha-se aproximadamente 40 trilhões de gigabytes de dados no mundo.

A figura 1 mostra o crescimento exponencial no volume dos dados.

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020

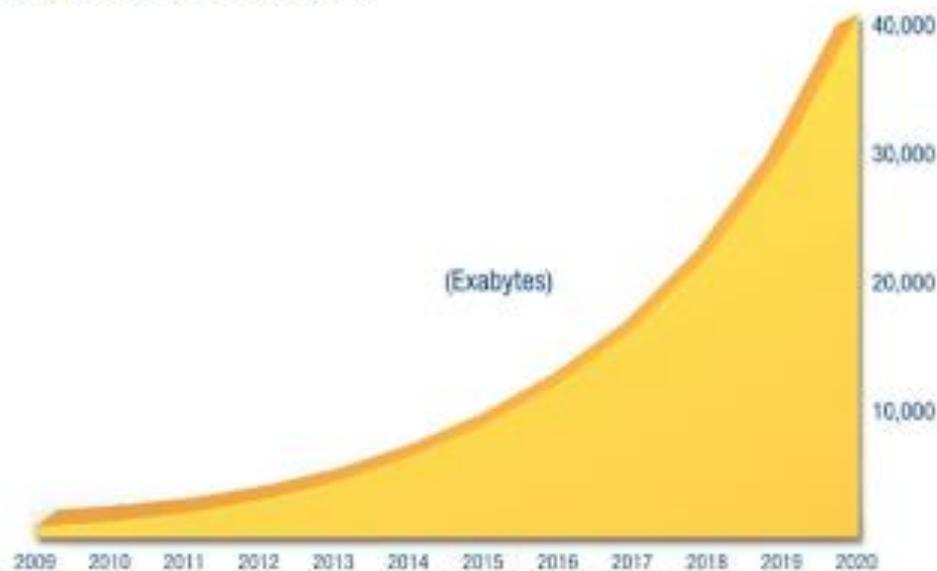


Figura 1: Crescimento dos Dados

Fonte: ÁVILA (2017)

2.2. OS 5V's DO BIG DATA

No contexto atual do *Big Data* consideram-se cinco importantes pilares que sintetizam o que é esta tecnologia. O **volume** refere-se ao tamanho e proporção dos dados; a **velocidade** é caracterizada pela rapidez na criação desses dados; a **variedade** pode ser entendida pela variação na forma como os dados se encontram ou são produzidos; a **veracidade** se dá pela verdade dos dados, ou seja, quais dados são verídicos de acordo com a realidade; o **valor** é capaz de gerar as informações certas nos momentos certos.

Volume: O volume trata-se do tamanho dos dados que se pode obter no contexto do *Big Data*.

Significa a alta capacidade de obtenção de dados ou o grande número de dados disponíveis em registro, por exemplo, pesquisas no Google ao upload de fotos no Facebook. CARVALHO, Letícia (2016)

Velocidade: Se refere a velocidade como os dados são criados, visto que a todo momento várias mensagens são enviadas nas redes sociais, um aplicativo recebe dados de geolocalização, redes bancárias efetuam várias transferências, entre outros.

A velocidade na qual os dados são coletados, analisados e utilizados. MARQUESONE (2016, p. 13)

Variedade: A variedade do *Big Data* se dá pelas diversas formas em que os dados são criados, ou seja, uma imagem possui uma estrutura diferente de um dado de localização.

No passado, a maior parte dos dados era estruturada e podia ser colocada em tabelas e relações. Hoje, 80% dos dados do mundo não se comportam dessa forma. Com o Big Data, mensagens, fotos, vídeos e sons, que são dados não-estruturados, podem ser administrados juntamente com dados tradicionais. CANALTECH

Veracidade: Em meio ao enorme volume e variedade de informações, é preciso destacar o que é rico em conteúdo importante para a empresa (E-SETORIAL, 2015). Nem tudo que se encontra na Internet é verídico, para isso precisa-se de boas análises para saber quais dados são verdadeiros ou e quais não são.

Veracidade é o que permite uma análise de dados mais consistente e adequada, baseada em informações relevantes e, mais que isso, potencialmente utilizáveis para melhorar as operações da empresa. TELIUM (2018)

Valor: O valor é considerado um dos pilares mais importantes do *Big Data* já que é através dele que o valor das análises é encontrado, o que também possui grande influência na tomada de decisão.

Os grandes dados devem ter valor. Investir na infraestrutura necessária para coletar e interpretar dados em grande escala, garantindo que os insights que são gerados sejam baseados

em dados precisos e levem a melhorias mensuráveis. MELO (2018)

A Figura 1 ilustra os 5V's e sua relação com o *Big Data*.



Figura 2: 5V's

Fonte: MITSUICHI (2017)

2.3. AS ANÁLISES DE BIG DATA

As análises de *Big Data*, também conhecida como *Big Data Analytics*, são a base para gerar *insights* e auxiliar a tomada de decisões. Essas análises buscam compreender, relacionar, e extrair valor de grandes volumes de dados.

O *Big Data Analytics* é o trabalho analítico e inteligente de grandes volumes de dados, estruturados ou não-estruturados, que são coletados, armazenados e interpretados por softwares de altíssimo desempenho. Trata-se do

cruzamento de uma infinidade de dados do ambiente interno e externo, gerando uma espécie de bússola gerencial para tomadores de decisão. Tudo isso, é claro, em um tempo de processamento extremamente reduzido. HEKIMA (2017)

Para a INTEL (2013) a análise de *Big Data* é:

- Uma estratégia baseada em tecnologia que permite a coleta de insights mais profundos e relevantes dos clientes, parceiros e sobre o negócio — ganhando assim uma vantagem competitiva.
- Trabalhar com conjuntos de dados cujo porte e variedade estão além da habilidade de captura, armazenamento e análise de softwares de banco de dados típicos.
- Processamento de um fluxo contínuo de dados em tempo real, possibilitando tomada de decisões sensíveis ao tempo mais rápido do que em qualquer outra época.
- Distribuído na natureza. O processamento de análise vai aonde estão os dados para maior velocidade e eficiência.
- Um novo paradigma no qual a TI colabora com usuários empresariais e “cientistas de dados” para identificar e implementar análises que ampliam a eficiência operacional e resolvem novos problemas empresariais.
- Transferir a tomada de decisão dentro da empresa e permitir com que as pessoas tomem decisões melhores, mais rápidas e em tempo real.

2.4. TIPOS DE ANÁLISE DE BIG DATA

Nesta seção serão estudados os tipos de análise de *Big Data*, tal como o funcionamento de cada uma delas.

Quando se fala de extrair valor de grandes volumes de dados através das análises de *Big Data*, encontramos quatro principais pilares que ajudam a encontrar valor desses dados.

A **análise preditiva** é uma das mais importantes análises quando se trata de *Big Data*, pois é através dela que é possível prever eventos futuros através de um grande conjunto de dados. Essa análise ajuda empresas a obterem uma tomada de decisão mais assertiva sobre seus dados e não usar apenas a intuição para tomar decisões.

A partir da identificação de padrões passados em sua base dados, esse tipo de análise permite aos gestores o mapeamento de possíveis futuros em seus campos de atuação. A ideia é deixar de tomar decisões baseadas unicamente na intuição, conseguindo estabelecer um prognóstico mais sólido para cada ação HEKIMA (2016).

O uso da análise preditiva já está revolucionando o modo como interagimos com o nosso ambiente. À medida que a quantidade de dados aumenta e melhora a qualidade, auxiliada pela disponibilidade de energia de processamento de custo reduzido, a análise preditiva tende a ser ainda mais difundida do que é hoje. Se já identificou alguns problemas que pretende resolver com a análise preditiva, você concordará que esta não foi uma previsão difícil de ser feita. GUAZELLI (2012)

Este tipo de análise é um pilar fundamental para a tomada de decisão, produção de importantes *insights*, e previsão de eventos futuros a partir desses dados.

Na **análise prescritiva** é estudado as possíveis consequências de cada operação, ou seja, de forma geral a análise prescritiva busca examinar cada efeito de uma tomada de decisão.

Para HEKIMA (2016) análise descritiva se caracteriza pelo estudo das consequências para tomadas de decisões específicas. Enquanto a análise preditiva busca prever acontecimentos futuros, a análise descritiva busca analisar os efeitos de cada ação. Ainda segundo HEKIMA, quando se trata de uma empresa ou departamento, a análise prescritiva tem sua valia através da capacidade de numerar e filtrar padrões.

A **análise descritiva** se caracteriza pelo resultado imediato de uma tomada de decisão. Segundo HEKIMA (2016) é a mineração de dados na base da cadeia de *Big Data*.

Pode-se compreender a análise descritiva como uma classificação de análise de crédito feita por redes bancárias, onde elas são responsáveis por analisar as informações de clientes, de empresas, ou até grupos específicos para determinar, por exemplo, taxas de juros com base nos dados dos clientes como movimentações em contas, verificar se o cliente possui débitos com outras redes bancárias, entre outros fatores que possibilitam as instituições financeiras a elaborar modelos mais específicos para tomada de decisão. HEKIMA (2016)

A **análise diagnóstica** é a capacidade de entender os resultados de cada tomada de decisão da empresa.

A análise diagnóstica tem como finalidade compreender o impacto de cada ação dentro das estratégias da empresa. Dessa forma, é possível saber quem comprou, onde comprou, quando comprou, porque comprou ou, ainda, porque deixou de comprar PLUGAR (2016)

Enquanto a análise descritiva busca detalhar uma base de dados, a análise diagnóstica tem como objetivo compreender de maneira causal (Quem, Quando, Como, Onde e Por que) todas as suas possibilidades. Se uma empresa executa uma ação de marketing, por exemplo, a análise diagnóstica é o caminho mais curto e eficiente para que os profissionais avaliem os impactos e o alcance dessa ação após sua realização. HEKIMA (2016).

Diferentemente da análise prescritiva, a análise diagnóstica estudará as causas após a sua realização, ou seja, estudará os possíveis efeitos após a tomada de decisão, enquanto a análise prescritiva compreenderá as consequências antes da tomada de decisão.

3. BIG DATA NAS EMPRESAS

Neste capítulo será apresentado de forma mais específica a relação do *Big Data Analytics* e as empresas. Serão estudados casos de sucesso de corporações que fazem o uso das análises de Big Data, bem como o uso dos tipos de análise apresentadas no capítulo anterior.

3.1. EMPRESAS E INDÚSTRIA 4.0

Nesta seção será apresentada uma breve introdução sobre o que é empresas e indústrias 4.0 e sua relação com as análises de Big Data. Como visto no capítulo anterior, o *Big Data* refere-se ao grande volume de dados que temos ao redor do mundo. Nesta seção será enfatizado como as empresas e indústrias utilizam técnicas de análise de *Big Data* afim de extrair valor para gerar *insights* e obter uma tomada de decisão mais assertiva.

Para as empresas as análises de *Big Data* são importantes para conhecer melhor os seus clientes, desenvolver promoções específicas baseadas no perfil de seus clientes, saber como estão seus lucros, prever evasão de clientes, obter informações sobre seus concorrentes, e até mesmo tomar conhecimento sobre possíveis fraudes e como evitá-las.

Segundo HEKIMA (2016) a indústria 4.0 se fundamenta no uso inteligente de informações através de análises e automações em linhas de processo de produtos, com o objetivo de afetar de forma positiva a indústria da transformação. Para Hekima, a indústria 4.0 tem sua sustentação em três importantes fundamentos, a integração de cadeias, a digitalização de produtos e serviços, e os modelos de negócio. Como visto nesta seção as análises de Big Data são importantes para empresas que desejam administrar melhor seus recursos e conhecer melhor seus clientes, bem como nas indústrias 4.0 onde as tecnologias da informação estão mais difundidas em seu dia a dia.

3.2. BIG DATA NAS REDES FINANCEIRAS

Esta seção enfatizará o modo como o *Big Data* influencia as redes bancárias, que possuem incontáveis transações a cada segundo, ou seja, como as técnicas de *Big Data* são aplicadas aos bancos com base em altos valores, alto fluxo de transferência, débitos em tempo real, análises de crédito e como os bancos usam os dados de seus clientes para oferecer promoções, novos cartões e limites, entre outros.

Antes de dar continuidade neste capítulo, é necessário conhecer um pouco sobre as redes bancárias que serão estudadas, tal como a influência e uso de cada uma delas sobre as análises de *Big Data*.

A primeira rede bancária que será estudada é a Nubank. A Nubank é uma *startup* brasileira que veio com a ideia de acabar com a complexidade dos bancos comuns, ou seja, grande parte das funcionalidades da Nubank podem ser feitas pelo celular a qualquer momento e a qualquer hora, bem como a solicitação do cartão de crédito que é feita na hora com a resposta em poucos minutos. Para DESIDÉRIO (2018) a Nubank está entre os cinco maiores emissores de cartão de crédito do Brasil. No serviço de conta digital, a *fintech* alcançou a marca de 2,5 milhões de clientes.

Como visto nas análises de *Big Data*, a Nubank utiliza as análises descritivas para determinar a liberação de um cartão de crédito para um novo cliente, ou seja, a partir de uma nova solicitação as informações do cliente são submetidas a uma análise com base em seus dados que determinam a aprovação de crédito.

A Nubank possui um recurso de não cobrar taxas de anuidade sobre o uso de seu cartão, ou seja, uma economia para o cliente, que acaba sendo um dos diferenciais dessa rede; outros benefícios da Nubank é a facilidade para pagamento de faturas, limites ajustáveis ao cliente, e descontos no pagamento de faturas pagas antecipadamente.

No caso comum de liberação de crédito a análise usada é a descritiva, como usado no exemplo anterior e isso não se restringe apenas a Nubank, mas a

grande parte das redes bancárias que processam novas solicitações de crédito, empréstimo, entre outros que fazem o uso desse tipo de análise para que seja tomada uma rápida decisão sobre a liberação dos benefícios ao cliente.

Outra empresa que usa o *Big Data* e que também é considerado um caso de sucesso como a Nubank é o Banco do Brasil.

O Banco do Brasil a cerca de 2 anos atrás passou a utilizar identificações biométricas com a finalidade de melhorar segurança de seus clientes para efetuar transações, e para obter melhores oportunidade de negócios. O desenvolvimento e investimento dessa solução não seria possibilitado sem o uso do Big Data para obter a identificação biométrica de seus correntistas. Além disso, o Banco do Brasil possui vários sistemas legados e antigos que foram desenvolvidos a mais de 10 anos, e o grande desafio é sem dúvidas, unir o grande volume de dados armazenados no banco que foi gerado ao longo dos anos. Entretanto, o Big Data é um forte aliado ao Banco do Brasil para possibilitar a realização dessas e de outras soluções que são desenvolvidas pelo banco. HEKIMA (2015)

O Banco do Brasil é um grande exemplo de como o *Big Data* auxilia na melhora de recursos para seus clientes. Desenvolvendo novas facilidades e serviços, o Banco do Brasil trabalha hoje com um grande aliado que é o *Big Data*. Como abordado no parágrafo anterior o *Big Data* está ajudando de forma significativa o Banco do Brasil a conectar, analisar, e extrair valores dos seus dados implementados ao longo das décadas.

A mais antiga instituição financeira dos Estados Unidos o *The Bank of New York Mellon*, é um destaque entre as redes bancárias por ser o maior banco de depósitos do mundo. O *The Bank of New York Mellon* assim com a Nubank e Banco do Brasil, utilizam-se do *Big Data* para desenvolver soluções sob medida a seus clientes e seu mercado.

De acordo com o artigo publicado pela HEKIMA em 2015, o *Big Data* está revolucionando os serviços financeiros com seis maneiras diferentes, mas que são indispensáveis para qualquer rede bancária.

O primeiro tópico da revolução do Big Data nos serviços financeiros é a detecção de fraudes. De acordo com HEKIMA (2015), as fraudes que ocorrem

em empresas de setores financeiros em específico estão cada vez mais elaboradas por se tratar diretamente com fundos monetários. Relacionar perfis de usabilidade, consumo e comportamento financeiro de seus clientes é algo que está no cotidiano para as instituições financeiras, entretanto, fazer uma boa monitoria dos recursos para identificar mudanças que fogem do padrão de seus clientes é fundamental a fim de intervir no bloqueio de operações que fogem desses padrões, e isso é considerado um grande desafio.

Outro ponto destacado no texto é a análise de risco para prever a demanda de crédito para determinado cliente, buscando um menor risco de inadimplência. Para HEKIMA, relacionar o perfil de consumo do cliente com o restante da base de dados permite prever a demanda de crédito, endividamento excessivo ou falta de capacidade para pagamento das dívidas. Muitas vezes, é preciso levar em consideração fatores que fogem ao ambiente financeiro para gerar esse perfil.

A segmentação de clientes com base em seus dados, também é outro fator que influencia na disseminação do *Big Data* nas redes bancárias. Segundo HEKIMA, traçar o perfil de consumo dos clientes com o que restante da base de dados possibilita a previsão da demanda de crédito, endividamento, entre outros. É preciso também analisar fatores que não se enquadram no ambiente financeiro para geração do perfil de cliente. A segmentação de clientes baseado em suas informações, é um fato que dissemina ainda mais o uso do *Big Data* nas redes bancárias.

Ainda segundo HEKIMA acompanhar notícias, fotos, vídeos, possibilita aos bancos categorizar seu público através de comportamentos comuns, elaborando estratégias específicas para determinado grupo. Um exemplo dessa segmentação pode ser na maximização de retorno, ou seja, clientes que devolveram o dinheiro usado para o banco. Buscar mais clientes cuja idade esteja entre 25 e 35 anos, que possua carro próprio e renda mensal superior a 03 salários mínimos, possibilita prever que os clientes pertencentes a esse grupo irão aplicar sua renda na aquisição de um imóvel, realizarão viagens, entre outros. Chegar a essa idealização de público exige uma análise que foge de seus dados de renda mensal e fluxo de sua conta bancária.

Uma das maneiras em que o Big Data também ganha destaque é na criação de ofertas personalizadas de produtos. Segundo HEKIMA, se um de seus clientes por ventura descobrir que será pai, muito provavelmente ele irá comprar um carro novo ou comprar uma casa maior e mais confortável. Para isso será necessário um crédito ajustado para essa finalidade. Caso a contratação do crédito seja realizada um desconto ou bonificação será oferecido ao cliente.

A predição de comportamento de seus clientes também é um tópico muito importante para as instituições financeiras. Para HEKIMA, uma solução de *Big Data* pode ajudar a manter o turnover dentro das metas estabelecidas pela empresa, ao evitar demissões desnecessárias fornecendo indicadores de resultados ruins em tempo real, ou mesmo decisões equivocadas tomadas pela equipe antes que se tornem irreversíveis.

A última maneira que HEKIMA define como as mais importantes na área de Big Data é a auditoria e governança, uma solução capaz de olhar para toda a operação das empresas, detectando padrões e relacionando uma grande quantidade de dados de forma rápida e eficiente, permite uma auditoria mais eficiente e eficaz, garantindo uma governança mais simples e barata. Simplicidade na operação e um custo operacional menor são alguns dos principais objetivos de uma empresa.

Dessa forma fica evidente nesta seção a importância significativa do Big Data e suas aplicações nos ramos financeiros, bem como as redes bancárias que a partir desta técnica conseguem prever comportamentos específicos de seus clientes, e possibilita oferecer promoções e produtos mais ajustados com o perfil de cada um deles, sendo assim as análises de *Big Data* se tornam um diferencial para qualquer empresa que deseja seguir no mercado financeiro.

3.3. CASOS DE SUCESSO: OUTRAS EMPRESAS QUE USAM BIG DATA

Esta seção abordará alguns casos de sucesso do *Big Data* aliado a outras grandes empresas.

Como visto em grande parte deste trabalho, as análises de *Big Data* são uma importante base para as empresas que desejam continuar fortes no mercado atual, bem como prever acontecimentos futuros, conhecer melhor seus clientes, melhorar a tomada de decisão deixando de lado a intuição por parte dos gestores, e também o uso do *Big Data* possibilita novos e poderosos *insights*.

Os exemplos abaixo foram extraídos da SantoDigital, uma consultoria de negócios em tecnologia especializada em serviços de computação em nuvem.

Um caso de sucesso de *Big Data* segundo a SantoDigital é a Nike, uma empresa estadunidense de calçados, equipamentos esportivos, roupas e acessórios que investiu em aplicativos móveis que facilitam a experiência de seus clientes.

Ao investir em aplicativos e dispositivos vestíveis (*wearables*), a fabricante de materiais esportivos reuniu uma série de informações sobre os praticantes de corrida. São dados relacionados à distância percorrida, velocidade, locais e rotas preferidos para treino, entre outros. Com todos esses dados, a Nike conhece melhor os hábitos e comportamentos esportivos do seu público, criando produtos cada vez mais alinhados a essas características SANTODIGITAL (2017).

Ao estimular esse comportamento, a empresa multiplicou a quantidade de dados gerados diariamente. Essas informações preciosas foram usadas pela Nike para compreender seu público, aprimorar seus produtos e desenvolver novos modelos de tênis e roupas esportivas de acordo com o que descobriu por meio da iniciativa DIVEO (2018).

Outro exemplo que será apresentado é a *streaming* de vídeo Netflix, que é uma provedora global de filmes e séries de televisão via *streaming*, atualmente com mais de 100 milhões de assinantes, e hoje é uma empresa que usa de forma frequente as análises de *Big Data* para recomendar filmes e séries baseados no perfil de seus assinantes.

A Netflix tem como base do seu trabalho diário a análise de dados. Seus algoritmos e especialistas possuem habilidades analíticas para personalizar o atendimento, distribuir conteúdo, analisar melhores dispositivos e conhecer os

hábitos dos clientes. Com isso, a empresa pode prever e recomendar ao usuário o estilo que mais combina com ele se baseando em seu comportamento na rede HEKIMA (2016).

HEKIMA *apud* MAGNUSSON (2016) diz que a Netflix trabalha com as seguintes premissas em relação aos seus dados: eles devem ser acessíveis, de fácil processamento, de fácil visualização e quanto mais tempo você demora para encontrá-los, menos valiosos eles se tornam.

A General Motors também conhecida por GM, também é um bom exemplo de *Big Data*, a GM disponibiliza um serviço de monitoramento em tempo real que também é aliada a Internet das Coisas (IoT).

Segundo HEKIMA (2017), a General Motors é um ótimo exemplo de fabricante de veículos que aproveita muito bem essa parceria entre Big Data e Internet das Coisas. A empresa utiliza a Onstar (subsidiária da GM) para oferecer a assinatura de serviços de segurança, comunicação e diagnósticos remotos para os usuários de seus carros. Esses diagnósticos em tempo real auxiliam na verificação da “saúde” dos sistemas mais importantes dos veículos, mesmo durante a condução. Os dados gerados desse monitoramento são enviados de volta às montadoras, que os usa também para melhorar seus carros.

4. LINGUAGEM R

A Linguagem R é um software livre e gratuito, e uma poderosa linguagem de programação que auxilia Matemáticos, Estatísticos, Cientistas de Dados e outras áreas relacionadas a dados como contabilidade, administração e marketing, por sua facilidade de gerar poderosos e sofisticados gráficos.

Neste capítulo serão apresentados alguns conceitos da Linguagem R e sua relação com as áreas de Ciência de Dados e *Big Data*, e como esta ferramenta auxilia a criação de insights e também a Tomada de Decisão.

O R surgiu em meados de 1993, foi desenvolvida por Ross Ihaka e Robert Gentleman, o nome R provém das iniciais de seus criadores, como também de um jogo figurado junto a Linguagem S que foi desenvolvida pela *Bell Laboratories*.

A Linguagem R é a ferramenta mais completa para análise de dados, visualização de gráficos e *dashboards* em comparação a outras linguagens que também podem ser utilizadas para analisar dados, como o Python. O R pode ser considerado mais completo devido a sua forte comunidade presente na Internet, bem como a sua facilidade de uso e implementação, instalação de vários pacotes que facilitam seu uso, e além disso, o R possui dados já instalados para que os desenvolvedores possam fazer análises e estudar conceitos dessa linguagem, como fazer cálculos estatísticos, previsões, entre outros.

Vale lembrar que este capítulo não tem o objetivo de demonstrar que uma linguagem é melhor que outra, porém o foco de estudo, será a linguagem R e sua contribuição para a área de *Big Data* e Ciência de Dados.

A Figura 2 apresenta um gráfico de pizza elaborado pela Linguagem R. Os dados para criação deste gráfico são do próprio R, como dito anteriormente o R possui dados já instalados em sua interface. Os dados apresentados no gráfico provêm de uma base de dados com o número de prisões que ocorreram em cidades nos Estados Unidos, onde a população urbana é superior a 80.000 habitantes.

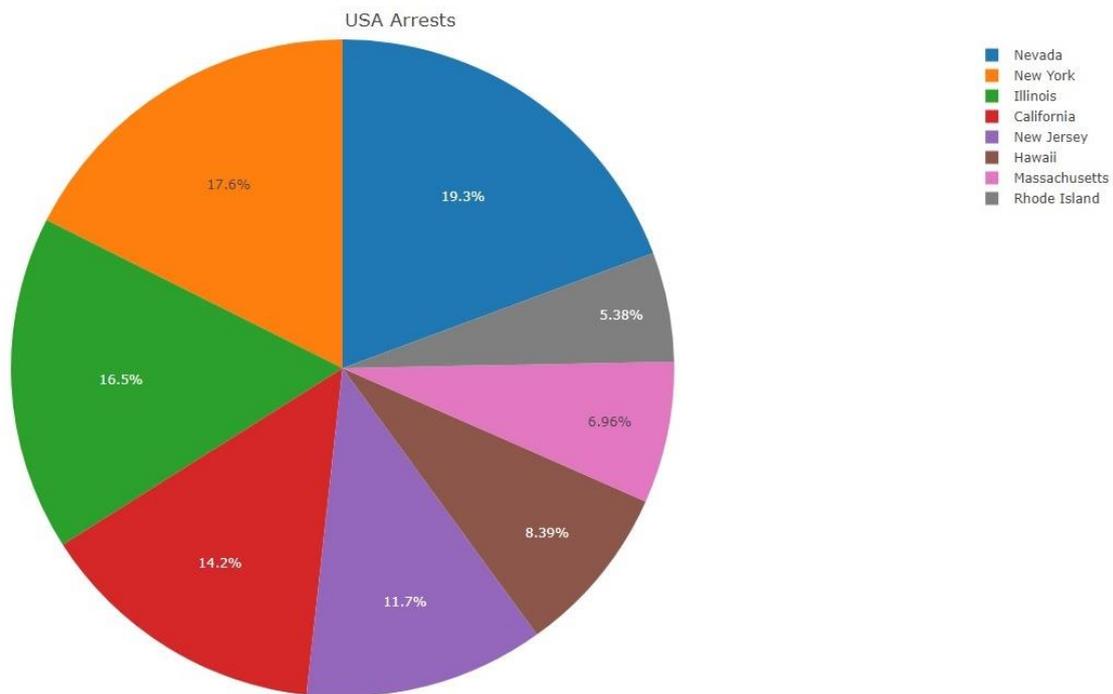


Figura 3: Gráfico de Pizza

Fonte: Elaborado pelo Autor

A Figura 3 apresenta outro gráfico sofisticado que a Linguagem R pode gerar para seus desenvolvedores, este gráfico é conhecido como mapa dinâmico.

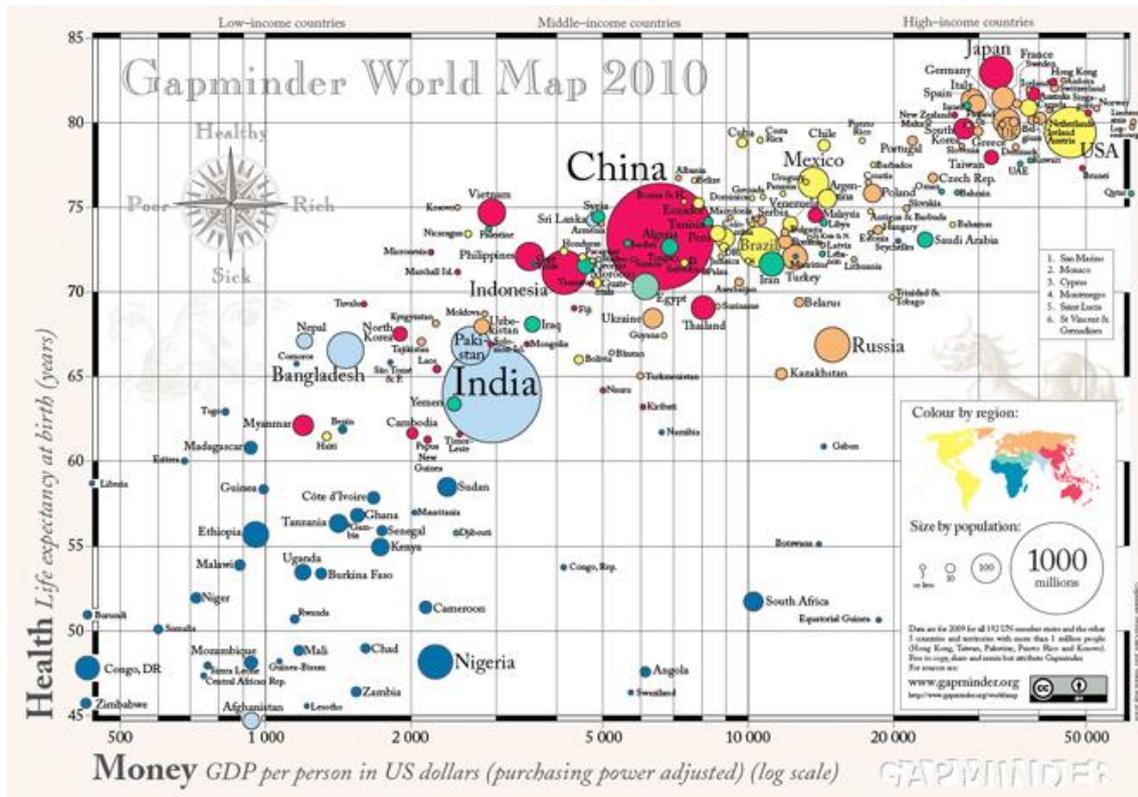


Figura 4: Mapa Dinâmico

Fonte: RAMOS (2015)

Estes são alguns dos sofisticados gráficos que esta linguagem proporciona, além de possibilitar a geração de diversos outros gráficos, como o de barras, o de dispersão, como também visualizações de histogramas, dentre outros. A Linguagem R também pode implementar modelos de *Machine Learning* e *Artificial Intelligence* para obter dados mais assertivos e coerentes para as análises, bem como ser uma ferramenta de auxílio a Tomada de Decisão e *insights*.

5. PROPOSTA DE TRABALHO

A proposta de trabalho visa estudar como as análises de Big Data influenciam empresas na Tomada de Decisão, e como geram resultados a partir de seus dados, tais conceitos foram abordados nos capítulos anteriores.

Para exemplificar a abordagem deste trabalho, será utilizada a Linguagem R e algumas de suas funcionalidades, a fim de demonstrar a extração de resultados, bem como o modelo de análise descritiva para a tomada de decisão.

De forma geral, o estudo de caso apresentado por este trabalho, demonstrará a classificação de uma solicitação de crédito para determinar se o cliente será um bom ou um mal pagador com base no histórico de classificação dos dados de outras solicitações de crédito.

A figura 4 mostra como é dividido este trabalho.

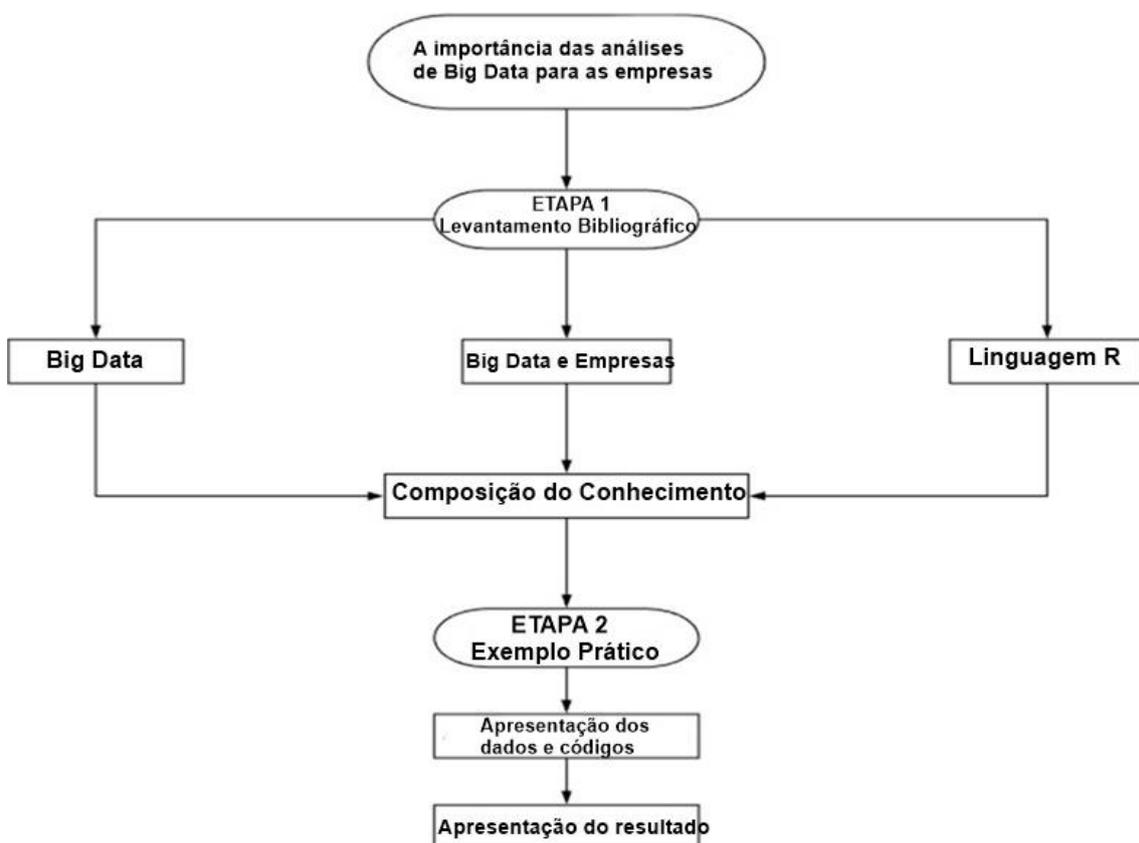


Figura 5: Proposta de trabalho

Fonte: Elaborado pelo Autor

6. EXEMPLO PRÁTICO

Para realização do exemplo prático foi utilizado um arquivo com 1000 registros contendo dados de clientes, como renda, cheque especial, histórico de crédito, propósito da solicitação, entre outros. Essas informações são importantes para treinar o modelo de classificação, ou seja, com o modelo treinado pode ser entendido que, pessoas que possuem cheque especial, histórico de crédito em estado crítico, e renda de 2000 reais, não são bons pagadores.

A Figura 6 mostra exemplos de alguns desses dados.

	A	B	C	D	E
1	CHEQUEESPECIAL	USO_CREDITO	HISTORICO_CREDITO	PROPOSITO	BALANCO_ATUAL
2	<0	6	crítica / outros de crédito	Computador/TV	1169
3	0<=X<200	48	existente e pago	Computador/TV	5951
4	sem cheque	12	crítica / outros de crédito	educação	2096
5	<0	42	existente e pago	mobilia/equipamento	7882
6	<0	24	atrasos anteriores	carro novo	4870

Figura 6: Dados

Fonte: AMARAL (2018), GRANATYR (2018)

Os dados para realizar este exemplo prático são da Formação Cientista de Dados ministrado por Fernando Amaral (2018) e Jones Granatyr (2018).

Para treinar o modelo foi usado o algoritmo de classificação Naive Bayes. Segundo CANDIAGO (2017) o algoritmo Naive Bayes é comumente usado em Machine Learning por ter um bom desempenho e por precisar de poucos dados para determinar as classificações com uma boa precisão.

6.1. INSTRUÇÕES DE CÓDIGO

A primeira instrução que será usada para o exemplo prático é a de leitura de arquivos do R, como o arquivo é um .csv onde seus valores são separados por vírgula ou ponto e vírgula, é utilizado o `read.csv`. O primeiro parâmetro é o `file.choose()` que usado para ler um arquivo presente em um diretório do computador; o segundo parâmetro `sep = “,”` é utilizado para dizer que o

separador do arquivo csv é o ponto e vírgula (;); o último parâmetro é “header=T”, este parâmetro diz ao nosso código que há um cabeçalho no arquivo. O arquivo csv que contém todo o histórico de credito ficará armazenado em uma variável chamada credito.

```
credito = read.csv(file.choose(), sep = ";", header = T)
```

Após a leitura do arquivo será criado uma amostra. As amostras em estatística podem ser entendidas como uma seleção de valores presentes em um conjunto de dados. A seguir o código para criar a amostra usada no exemplo prático.

```
amostra = sample (2, 1000, replace = T, prob = c(0.7,0.3))
```

A função `sample` é usada para criar as amostras. O primeiro parâmetro do exemplo acima significa que será gerado dois valores (1 e 2), em uma quantidade igual a 1000 pois é o tamanho do conjunto de dados; o `replace` é usado para dizer a função `sample` que haverá reposição deste valores; e o `prob` é um vetor que diz a função `sample` que existirá uma probabilidade de 70% para gerar o valor 1 e de 30% para gerar o valor 2, a divisão destes valores será entendido nos próximos códigos.

A próxima instrução que será mostrada é a divisão da amostra em duas variáveis, são elas `creditotreino` (será usada para treinar o modelo), e `creditoteste` (usada para testar o modelo). O código a seguir mostra a divisão da amostra.

```
creditotreino = credito[amostra==1,]  
creditoteste = credito[amostra==2,]
```

A variável “creditotreino” receberá os dados onde a amostra de credito for igual a 1, como dito no código anterior a amostra foi gerada com 70% para o número 1 e 30% para o número 2, essa divisão é importante para que não ocorra um sobre ajuste e que o modelo fique melhor treinado para os dados de credito e

obtenha bons resultados quando colocado em produção. Já a variável “creditoteste” recebe os valores de credito onde a amostra é igual a 2, ou seja, os 30% da divisão da amostra.

O próximo passo a ser feito neste exemplo prático é utilizar de fato o algoritmo naiveBayes para realizar a classificação, para isso será utilizado o comando *library* do R, que é usado para ler as funções de uma biblioteca em específico, neste caso a biblioteca usada será a e1071, responsável por conter os algoritmos de *Machine Learning* como o naiveBayes.

```
library(e1071)
```

Um dado em especial no conjunto de dados é a CLASSE, nela contém as classificações dos clientes, ou seja, se o cliente foi um bom ou mal pagador, é através dessa variável que o modelo estuda as resultantes dessa classificação, como renda, o propósito da solicitação, histórico de credito, entre outros.

O próximo código é de fato o treinamento do modelo de predição, para isso tem-se o parâmetro CLASSE já citada acima, e o creditotreino. Esses parâmetros são passados a função naiveBayes que treinará o modelo para a classificação. A seguir o código para criação do modelo.

```
modelo = naiveBayes(CLASSE ~ . , creditotreino)
```

Depois da criação do modelo de classificação, será criada uma variável denominada como “predicao”, nesta variável será passado como parâmetros juntamente a função *predict* o modelo, que foi criado através da CLASSE e do creditotreino, e também a variável creditoteste. O código abaixo mostra o uso da função *predict* e dos parâmetros para criação da variável de “predicao”. O próximo para realizar o modelo de predição para este exemplo prático é a criação de uma matriz de confusão para medir as taxas de erro e acerto da predição criada nos comandos anteriores.

```
predicao = predict(modelo , creditoteste)
```

Uma matriz de confusão é uma tabela que mostra as ocorrências para o modelo criado. No caso deste exemplo prático para medir a taxa de acerto precisa-se visualizar os dados onde as linhas e as colunas encontram-se, já a taxa de erro é medida quando as colunas não são as mesmas que as linhas, a imagem abaixo mostra um exemplo de uma matriz de confusão gerada em uma execução com os dados deste exemplo prático. A seguir pode ser visualizado o código para gerar a matriz de confusão.

```
confusao = table(creditoteste$CLASSE, predicao)
```

É armazenado na variável confusão uma tabela criada a partir de creditoteste e a coluna CLASSE, vale ressaltar que esta coluna é uma das mais importantes dos dados, pois nela contém as classificações dos outros clientes. A imagem a seguir mostra os números gerados para a matriz de confusão.

```
> confusao
      predicao
      bom ruim
bom  193  23
ruim  53  44
```

A imagem acima ilustra que foi dito nos parágrafos anteriores sobre a matriz de confusão. Os dados contornados em vermelho mostram os valores que serão usados para calcular a taxa de acerto do modelo, ou seja, onde encontra-se bom nas linhas e colunas, o mesmo vale para as linhas e colunas de ruim. Já a taxa de erro será medida através dos dados contornados em azul, onde não se tem o encontro das colunas com as linhas, dessa forma, a taxa de erro do modelo é calculada onde tem-se linha bom e coluna ruim, e linha ruim e coluna bom.

Para entender os próximos comandos, é preciso visualizar a matriz de confusão da seguinte forma. Onde temos os encontros das colunas com as linhas será calculado de fato as taxas de erro e de acerto do modelo, e para isso é necessário que se compreenda a imagem a seguir.

```
> confusao
      predicao
      bom ruim
bom    1    2
ruim   3    4
```

Para calcular as taxas de acerto será preciso que seja efetuado uma soma dos valores em vermelho, ou seja, onde as colunas coincidem com as linhas, de forma mais simples onde tem se os valores 1 e 4, vale lembrar que esses valores são apenas para entender o código abaixo, e não os reais valores da matriz de confusão, os valores da matriz estão na imagem anterior. Dessa forma tem-se o comando seguinte.

```
acerto = (confusao [1] + confusao [4]) / sum(confusao)
```

Sendo assim, a variável “acerto” recebeu a soma de confusão onde as colunas coincidem com as linhas e após a soma, recebeu uma divisão por uma também soma de toda a matriz de confusão.

Seguindo o mesmo conceito do cálculo da taxa de acerto, a taxa de erro é calculada com a soma dos valores 2 e 3 da matriz de confusao, ou seja, onde não se tem uma combinação das linhas com as colunas, dessa forma o seguinte código realiza o cálculo da taxa de erro do modelo de predição.

```
erro = (confusao [2] + confusao [3]) / sum(confusao)
```

Agora que o modelo de classificação já foi treinado e testado e já se têm os valores das taxas de acerto e de erro, o próximo passo é submeter esse modelo a uma nova classificação. Essa classificação vai de fato mostrar se o cliente que solicita um novo limite de crédito, será um bom ou mal pagador.

Para fazer essa classificação será necessário mais um arquivo, nele contém os dados do cliente que solicita o limite de crédito. Os dados desse cliente seguem a mesma estrutura do modelo que foi treinado e testado com os 1000 registros. Abaixo pode ser visualizado um pouco sobre os dados desse cliente.

L	M	N	O	P	Q	R	S	T
TIPOBENS	IDADE	OUTROSPLANOSPGTO	RESIDENCIA	CREDITOS	EMPREGO	DEPENDENTES	TRABAESTRANGEIRO	CLASSE
seguro de vida	53	nenhum	própria	2	especializado	1	sim	

Vale destacar que para classificar o cliente é necessário que a coluna CLASSE esteja vazia como na imagem acima, pois não teria sentido realizar uma classificação de uma solicitação que já esteja classificada.

Abaixo pode ser visualizado o comando de leitura de arquivos do R, que já foi visto neste exemplo prático para fazer a leitura do novo propósito de crédito.

```
novocredito = read.csv(file.choose(), sep = ";", header = T)
```

O comando acima fará a leitura do arquivo para realizar a classificação desse novo cliente.

O próximo comando também já foi visto neste exemplo, e é o comando que faz o uso da função predict, e junto a ele se tem os parâmetros do modelo de classificação já criado e a nova solicitação de crédito.

```
predicaocredito = predict(modelo , novocredito)
```

Após a execução do código acima a classificação já foi realizada, e para visualizar o resultado basta digitar o nome da variável, neste caso o nome da variável é predicaocredito. Veja abaixo o resultado obtido desta classificação.

```
> predicaocredito
[1] bom
Levels: bom ruim
```

A imagem acima mostra o resultado da classificação, e pode ser notado que o cliente foi classificado como um bom pagador, e no caso de uma rede bancária por exemplo, o limite de crédito pode ser concedido a este cliente.

Abaixo da classificação pode ser visualizado os *Levels*, esses *levels* são os níveis que classificação pode atingir para este modelo, ou seja, ou cliente será um bom pagador ou será um mal pagador.

6.2. CONCLUSÃO

Antigamente os dados eram apenas informações de pouco uso e conhecimento para as empresas, hoje através das análises de Big Data esses dados ajudam a conhecer melhor seus clientes, o que torna um bem valioso para as empresas que buscam seguir no mercado.

Pode ser constatado através do estudo de alguns casos de sucesso como a Netflix, que as análises de Big Data quando visam a comodidade do cliente tem maior chance de obter sucesso. No caso da Netflix, ela combina os dados com algoritmos de Inteligência Artificial e Aprendizagem de Máquina para recomendar filmes e séries, e elaborar um catálogo específico para cada cliente de acordo com seu gosto.

A realização do exemplo prático foi importante para demonstrar o que foi teoricamente apresentado nessa pesquisa, a importância das análises de *Big Data* para as empresas se dá pelo conhecimento sobre seus dados, e como é importante conhecer e saber usar esses dados da melhor forma, já que com eles é possível conhecer novos mercados, detectar falhas na rede da empresa, fazer promoções para grupos específicos de clientes, entre outros.

O exemplo da rede bancária em específico, demonstrou que quando um limite de crédito é solicitado é preciso ter uma tomada de decisão rápida para dar *feedback* ao cliente, o modelo de predição usado no exemplo prático evidenciou em instantes se o cliente seria um bom ou mal pagador.

O estudo de caso prático pode evidenciar a importância dos dados para as empresas. Através de um grande conjunto de dados treinado com Machine Learning foi possível prever se um cliente será um bom ou mal pagador com base no histórico de outros clientes e em seu próprio histórico de dados.

No exemplo prático, o tipo de análise usada foi a análise preditiva, ou seja, ela tem o intuito de prever acontecimentos futuros com base na segmentação de grandes conjuntos de dados, e deixa que as tomadas de decisões não sejam mais feitas através do uso da intuição.

REFERÊNCIAS

ALBUQUERQUE, E. “O que faremos com os 40 trilhões de gigabytes de dados disponíveis em 2020?”. Open Knowledge Brasil. Disponível em: <<https://br.okfn.org/2017/09/29/o-que-faremos-com-os-40-trilhoesde-gigabytes-de-dados-disponiveis-em-2020> />. Acesso em: 21 fev. 2019.

AMARAL, F.; GRANATYR, J. “FORMAÇÃO CIENTISTA DE DADOS”. FERNANDO AMARAL; JONES GRANATYR. Disponível em: <<https://www.udemy.com/cientista-de-dados/>>

CANALTECH, R. “Big Data: Os cinco Vs que todo mundo deveria saber”. Canaltech. Disponível em: <https://canaltech.com.br/big-data/Big-Data-os-cinco-Vs-que-todo-mundo-deveria-saber/>. Acesso em: 27 fev. 2019

CARVALHO, L. “Big Data: oportunidades e desafios para os negócios”. 2016. 55. Gestão de Empresas. Universidade Estadual de Campinas. Campinas. 2016.

DIVEO, U. “Como grandes marcas estão usando o Big Data para entender o consumidor”. DIVEO UOL, 2018. Disponível em: <<https://computerworld.com.br/brandpost/como-grandes-marcas-estao-usando-o-big-data-para-entender-o-consumidor/>>. Acesso em: 02 mar.2019

DESIDÉRIO, M. “Nubank chega a 5 milhões de clientes e já é um dos maiores do mundo”. EXAME, 2018. Disponível em: <<https://exame.abril.com.br/negocios/nubank-chega-a-5-milhoes-de-clientes-no-cartao-de-credito/>> Acesso em: 01 mar. 2019

E-SETORIAL. “Volume + Variedade + Veracidade + Velocidade + Valor. Entenda os cinco Vs do Big Data.”. E-SETORIAL, 2015. Disponível em: <https://www.e-setorial.com.br/blog/170-volume-variedade-veracidade-velocidade-valor-entenda-os-5-vs-da-big-data>. Acesso em: 27 fev. 2019

FARINELLI, F. “O surgimento das organizações orientadas por dados”. IGTI BLOG. Disponível em: < <http://igti.com.br/blog/o-surgimento-das-organizacoes-orientadas-a-dados> />. Acesso em: 20 fev. 2019.

GUAZELLI, A. “O que é a análise preditiva?”. ALEX GUAZELLI, 2012. Disponível em: < <https://www.ibm.com/developerworks/br/industry/library/ba-predictive-analytics1/index.html>>. Acesso em: 28 fev. 2019

HEKIMA. “Tipos de Análise de Big Data: Você conhece todos os 4?”. HEKIMA, 2016. Disponível em: < <http://www.bigdatabusiness.com.br/conheca-os-4-tipos-de-analises-de-big-data-analytics/>> Acesso em: 28 fev. 2019

HEKIMA. “A importância de Big Data para a indústria 4.0”. HEKIMA, 2017. Disponível em: < <http://www.bigdatabusiness.com.br/big-data-na-industria-4-0/>>. Acesso em: 28 fev. 2019

HEKIMA. “Big Data em serviços financeiros: Conheça as aplicações e cases de sucesso”. 2016, 27. Belo Horizonte. 2016

HEKIMA. “Big Data e Netflix: Uma parceria de sucesso”. HEKIMA, 2016. Disponível em: <http://www.bigdatabusiness.com.br/netflix-e-big-data-uma-parceria-de-sucesso/>. Acesso em: 02 mar. 2019

HEKIMA. “Big Data na indústria da transformação: da revolução à indústria 4.0”. 2017, 41. Belo Horizonte. 2017

INTEL. “Guia de Planejamento: Saiba mais sobre Big Data”. INTEL, 2013. Disponível em: < www.intel.com.br/content/dam/www/public/lar/br/pt/documents/articles/90318386-1-por.pdf> . Acesso em: 28 fev. 2019

LAUDON, Kenneth C.; LAUDON, Jane P. Sistemas de Informação Gerenciais. Ed. 11. São Paulo, SP: Pearson Education do Brasil, 2014.

CANDIAGO, L. “ALGORITMO DE CLASSIFICAÇÃO NAIVE BAYES”. LORENZO CANDIAGO, 2017. Disponível em: <<https://www.organicadigital.com/seeds/algoritmo-de-classificacao-naive-bayes/>> . Acesso em: 12 Jul. 2019

MARQUESONE, R. “BIG DATA O NOVO DESAFIO DAS EMPRESAS E PROFISSIONAIS DO MERCADO”. 2016. 55. Gestão de Empresas. Universidade Estadual de Campinas. Campinas. 2016.

MITSUICHI, L. “Big Data: Conheça os 5V’s e sua aplicação prática para PMEs”. SEM RUSH, 2017. Disponível em: < <https://pt.semrush.com/blog/big-data->

[conheca-os-5-vs-e-sua-aplicacao-pratica-para-pmes/](#)>. Acesso em: 14 mar. 2019

MELO, E. “Os 5 Vs do BigData”. Ederson Melo, 2018. Disponível em: < <https://www.edersonmelo.com/os-5-vs-do-bigdata/>>. Acesso em: 27 fev. 2019

PLUGAR. “Quatro faces da análise de dados: análise diagnóstica”. PLUGAR, 2016. Disponível em: < <https://www.plugar.com.br/quatro-faces-da-analise-de-dados-analise-diagnostica/>>. Acesso em: 28 fev. 2019

RAMOS, R. “Por que devemos aprender a programar em R?”. Raniere Ramos, 2015. Disponível em: <https://oestatistico.com.br/por-que-devemos-aprender-a-programar-em-r/>>. Acesso em: 14 mar. 2019

SANTODIGITAL. “3 casos de sucesso de empresas que fizeram uso do Big Data”. SANTODIGITAL, 2017. Disponível em: < <http://www.santodigital.com.br/3-cases-de-sucesso-de-empresas-que-fizeram-uso-big-data> />. Acesso em: 02 mar. 2019

SCHERMANN, D. “Os 5 maiores desafios para quem quer utilizar o Big Data”. Opinion Box, 2015. Disponível em: < <https://blog.opinionbox.com/os-maiores-desafios-para-quem-quer-utilizar-o-big-data> />. Acesso em: 20 fev. 2019.

TELIUM. “Você conhece os 4v’s do Big Data?”. TELIUM, 2018. Disponível em: < <https://blog.telium.com.br/voce-conhece-os-4-vs-do-big-data/> >. Acesso em: 27 fev. 2019