

USO DA CIÊNCIA DE DADOS NO COMBATE À “SÍNDROME DO BALDE FURADO” NAS ORGANIZAÇÕES

USE OF DATA SCIENCE TO FIGHT “LEAKY BUCKET SYNDROME” IN ORGANIZATIONS

WELLINGTON SOUSA AGUIAR
CENTRO UNIVERSITÁRIO ESTÁCIO DO CEARÁ

JOSÉ MÁRIO BEZERRIL FONTENELLE

JOSÉ EGNALDO DA SILVA CARVALHO

JOSÉ VALCLEMIR RODRIGUES DA SILVA

Nota de esclarecimento:

O X SINGEP e a 10ª Conferência Internacional do CIK (CYRUS Institute of Knowledge) foram realizados de forma remota, nos dias 26, 27 e 28 de outubro de 2022.

ANOS
SINGEP

USO DA CIÊNCIA DE DADOS NO COMBATE À “SÍNDROME DO BALDE FURADO” NAS ORGANIZAÇÕES

Objetivo do estudo

O objetivo principal do estudo é propor um modelo padronizado para gerar conhecimentos estratégicos usando Ciência de Dados e Inteligência Artificial (IA) no combate e redução dos impactos da “Síndrome do balde furado”, que afligem as organizações minando as suas estratégias.

Relevância/originalidade

As empresas buscam atrair e fidelizar os melhores clientes do mercado. Para tratar esses grandes desafios as organizações desenvolvem ações comerciais e de marketing, mas ainda não utilizam todas as possibilidades e ferramentas da Ciência de dados e IA disponíveis.

Metodologia/abordagem

O estudo proposto utiliza a metodologia qualitativa e exploratória através da descrição do modelo proposto. Foi utilizada ainda a pesquisa bibliográfica em artigos, teses e dissertações para dar sustentação ao estudo. O modelo proposto é o resultado de vários anos de pesquisas.

Principais resultados

O estudo aplicado em grandes organizações, que possuem bases de dados de qualidade, já nos mostrou que pode apoiar as estratégias organizacionais com a geração de conhecimentos estratégicos até então não observados ou desconhecidos.

Contribuições teóricas/metodológicas

A Inteligência Artificial e a Estatística permitem o desenvolvimento de modelos de dados capazes de fazer emergir conhecimentos até então não percebidos pelas organizações, capazes de apoiar na tomada de decisão e no desenho de estratégias vencedoras.

Contribuições sociais/para a gestão

O estudo apresenta um modelo para apoiar na tomada de decisão e na formulação de estratégias para a conquista e retenção de clientes nas organizações de todos os tamanhos e mercados.

Palavras-chave: Churn, Balde furado, Inteligência artificial, Cluster, Predição

USE OF DATA SCIENCE TO FIGHT “LEAKY BUCKET SYNDROME” IN ORGANIZATIONS

Study purpose

The main objective of the study is to propose a standardized model to generate strategic knowledge using Data Science and Artificial Intelligence (AI) to combat and reduce the impacts of the "Leaky Bucket Syndrome", which afflict organizations by undermining their strategies.

Relevance / originality

Companies seek to attract and retain the best customers in the market To address these great challenges, organizations develop commercial and marketing actions, but still do not use all the possibilities and tools of Data Science and AI.

Methodology / approach

The proposed study uses a qualitative and exploratory methodology through the description of the proposed model Bibliographic research was also used in articles, theses and dissertations to support the study The proposed model is the result of several years of research.

Main results

The study applied in large organizations, which have quality databases, has already shown us that it can support organizational strategies with the generation of strategic knowledge hitherto unobserved or unknown.

Theoretical / methodological contributions

Artificial Intelligence and Statistics allow the development of data models capable of bringing out knowledge hitherto not perceived by organizations, capable of supporting decision-making and the design of winning strategies.

Social / management contributions

The study presents a model to support decision-making and the formulation of strategies for winning and retaining customers in organizations of all sizes and markets.

Keywords: Churn, Leaky bucket, Artificial intelligence, Cluster, Prediction

1 Introdução

A evolução tecnológica computacional tem origem nas décadas de 50 e 60 com os Mainframes, disponíveis só para grandes organizações públicas e privadas. No Brasil, somente no início da década de 80 chegam a democratização da tecnologia com os PCs (*Personal Computers*) e a TV por assinatura no final dos anos 80. Na década de 90 chegam a internet, a telefonia móvel e a privatização das telecomunicações no Brasil que impulsionam a tecnologia. Todo esse caminho foi necessário para que a partir de 2010 pudéssemos assistir à popularização das tecnologias no Brasil com a chegada da computação em nuvem, smartphones modernos, mobilidade, e assim criar um ambiente tecnologicamente adequado para armazenar e processar grandes volumes de dados com capacidade para manipular textos, fotos, vídeos, TVs, links, tudo em alta velocidade e disponibilidade. É neste contexto moderno e atual que a Inteligência Artificial (IA) e a Ciência de Dados se desenvolvem, possibilitando seu desenvolvimento e aplicação, acessível a todos que desejam se beneficiar dos seus poderosos recursos de geração de conhecimento a partir de grandes volumes de dados (AGUIAR, 2021).

Historicamente a Inteligência Artificial (IA) teve uma evolução inicialmente lenta. Imaginava-se a IA semelhante ao que o cinema apresentava, com máquinas pensando e com emoções humanas, o interesse em investir nessas pesquisas por parte de empresas privadas não ocorreu nesse momento. Somente quando os sistemas baseados em IA tiveram por objetivo resolver problemas práticos e específicos empresariais foi que despertou o maior interesse pelo investimento em pesquisas nessa área.

Atualmente os termos *Data Science* (Ciência de Dados) e *Big Data* (Mega Dados) são utilizados em profusão nos meios acadêmicos e profissionais como se fossem conceitos novos, diferentes daqueles que os estatísticos já lidam há cerca de dois séculos. Em 1980, Jeff Wu já sugeria que se adotassem os rótulos *Statical Data Science*, ou simplesmente, *Data Science*, no lugar de *Statistics*, para dar maior visibilidade aos trabalhos dos estatísticos (MORETTIN; SINGER, 2022).

Em mercados competitivos e de larga escala, o tratamento, coleta e processamento adequado dos dados tem se firmado como uma ferramenta estratégica de alto poder de resposta e geração de conhecimentos.

Grandes corporações já se utilizam da Ciência de Dados e IA em suas operações, como por exemplo: a Accenture que usa *Machine Learning* (ML) para avaliar currículos de seus funcionários e candidatos à vagas; A Lennox que usa um software para prever defeitos nos equipamentos que dá manutenção com acerto de 90% em até quatro horas antes do defeito ocorrer; A American Travel que usa ML para ajudar seus clientes em viagens corporativas, disponibilizando melhores tarifas aéreas e de hotéis; A Amazon que faz com grande sucesso, a indicação de compras baseada no comportamento de outros consumidores; A Mastercard que usa ML para detectar anomalias que sugerem ataques de Hackers e tentativas de fraudes; A NFL que calcula e apresenta ao vivo durante suas transmissões as análises preditivas das jogadas, entre outras aplicações (AGUIAR, 2021).

A partir dos exemplos e possibilidades expostas acima, das experiências acadêmicas e profissionais que participamos, foi possível propor um modelo (ou processo) capaz de auxiliar e orientar as organizações nessa luta contra a “Síndrome do balde furado”, que invalida ou elimina todos os esforços e conquistas das áreas comerciais na atração de clientes, em contrapartida às saídas não esperadas ou não previstas de parte dos clientes.

A Figura 1 abaixo, apresenta uma analogia de um balde furado com uma organização, que capta clientes para o seu portfólio, mas os pequenos furos na sua estrutura ou nos seus

processos, não conseguem reter todos os clientes conquistados, invalidando o esforço realizado na captação.



Figura 1 – Analogia do Balde Furado.

Fonte: Elaborado pelo autor.

O objetivo principal desse estudo é a proposição de um modelo padronizado para gerar conhecimentos estratégicos usando Ciência de Dados e Inteligência Artificial (IA) no combate e redução dos impactos da “Síndrome do balde furado”, que tanto aflige as organizações minando as suas estratégias.

2 Referencial Teórico

2.1 Síndrome do Balde Furado

Reter clientes é um importante desafio das organizações através dos relacionamentos dos clientes com a empresa ou com os produtos ou serviços oferecidos. Kotler, Hayes e Bloom (2002) fazem analogia com a “teoria do balde furado”, para destacar a importância da retenção, corroborando com Kotler (2000, p.69):

“Não basta dominar as técnicas para atrair novos clientes; a empresa deve retê-los. Muitas empresas possuem um alto índice de rotatividade de clientes, ou seja, conquistam novos clientes e perdem muitos deles. É como sistematicamente adicionar água a uma panela que possui um pequeno furo. As empresas de hoje devem dar mais atenção a seu índice de abandono de clientes”.

Essa analogia ilustra o desafio de encher um balde de água com diversos furos no fundo. Como se a água fossem os clientes e o balde a empresa, ou seja, por mais que haja captação contínua de clientes (entrada de água) é inevitável que, enquanto a empresa não se atentar em “tapar os furos”, terá um trabalho considerável para manter o nível de água ou aumentá-lo. Assim, ainda de acordo com Kotler, Hayes e Bloom (2002), a companhia necessita entender e traçar planos de ação coerentes para vedar e entender por que os clientes estão indo embora e, em seguida, dedicar tempo e energia para mantê-los.

O estudo da “Síndrome do Balde Furado” nas organizações nos remete ao desafio de entender algumas questões sobre a estratégia das empresas: [a] Será que conhecemos o ambiente de onde coletamos nossos clientes ?; [b] Será que estamos enchendo nosso balde com

os clientes adequados ?; [c] Como podemos tapar os buracos do balde para os clientes não saírem ?; [d] O que podemos aprender com os clientes que escorreram pelos furos do nosso balde ?

2.2 Estatística

O uso de dados estatísticos vem dos censos realizados na antiga Babilônia no Egito e posteriormente no Império Romano, quando se coletavam dados sobre nascimentos e óbitos para uso do Estado. A palavra “Estatística” é derivada da palavra latina “Status”, que significa “Estado”. Em resumo, Estatística é a ciência que coleta, organiza, analisa e interpreta dados para a tomada de decisões (LARSON; FARBER, 2010).

A Estatística é um conjunto de técnicas úteis para a tomada de decisão sobre um processo ou população, baseada na análise da informação contida em uma amostra desta população. Esta é uma ciência que oferece uma coleção de métodos para planejar experimentos e levantamentos para obter dados, organizar, resumir, analisar, interpretar e deles extrair conhecimento. A Estatística faz parte da área de estudo de pesquisa operacional que oferece ferramentas para identificar problemas através de seus sintomas e procurar solucioná-los para tomar a decisão correta (ANDRADE, 2009).

2.3 Ciência de dados

Trata-se de um termo cada vez mais utilizado para uma área de conhecimento voltada para o estudo e análise voltado para os dados, em busca da extração de novos conhecimentos e criação de novas informações. Trata-se de uma área multidisciplinar envolvendo duas grandes áreas: Ciência da Computação e Estatística, podendo ser aplicada em outras grandes áreas como: Medicina, Biologia, Economia, Comunicação, Ciências Políticas etc. Apesar de não ser uma área nova, ganhou muito impulso com a capacidade computacional crescente e os grandes volumes de dados para a tomada de decisão (McDONNELL; GUERRA; OLIVEIRA, 2018).

A Figura 2 abaixo, apresenta um *Workflow* (Fluxo) básico dos dados na Ciência de Dados com suas principais atividades, finalizando com a comunicação dos resultados encontrados para apoiar a tomada de decisão.



Figura 2 – Fluxo Básico de Ciência de Dados.

Fonte: Elaborado pelo autor a partir de McDonnell, Guerra e Oliveira (2018).

A Ciência de dados se utiliza de várias linguagens e ferramentas para execução das suas atividades, entre as mais conhecidas e com maior potencial, estão : Python, R, SAS, SQL, Matlab, PowerBI, Stata etc. Essas ferramentas se utilizam de bibliotecas e funções disponibilizadas para facilitar as atividades dos programadores, estatísticos e analistas de dados no uso dessas ferramentas.

2.4 Inteligência Artificial

A inteligência Artificial começou a ser estudada no pós-guerra na década de 50, mas ficou adormecida por várias décadas à espera das condições ideais para o seu desenvolvimento, que encontramos agora, como: computação em nuvem, algoritmos complexos, Big Data e alto poder dos computadores atuais em armazenar e processar grandes volumes de dados em curto espaço de tempo e com alta disponibilidade (AGUIAR, 2019).

Segundo Luger (2013), a Inteligência Artificial (IA) revela um campo de estudo jovem e promissor, onde o principal interesse é encontrar um modo efetivo de entender e aplicar técnicas inteligentes para a solução de problemas e para o planejamento de uma gama de problemas práticos. São várias as áreas em que a IA pode auxiliar, entre elas podemos citar:

- O uso do computador para executar raciocínio, reconhecimento de padrões, aprendizado ou outras formas de inferência;
- Raciocínio que utiliza as características qualitativas significativas de uma situação;
- Respostas que não são nem exatas nem ótimas, mas que são “suficientes”. São casos que resultados exatos ou ótimos são caros demais ou mesmo impossíveis, entre outras áreas.

Seguem abaixo alguns conceitos de IA utilizados na academia:

“IA é o estudo de faculdades mentais através do uso de modelos computacionais” (CHARNIAK; McDERMOTT, 1987).

“IA é a arte de criar máquinas que executam funções que requerem inteligência quando executadas por pessoas” (KURZWEIL, 1990).

“IA é o estudo de como fazer os computadores realizarem coisas que, no momento, as pessoas fazem melhor” (RICH; KNIGHT, 1990).

Em resumo, combinando todos os conceitos na literatura sobre IA, chegamos ao conceito mais usado na academia: “IA é o campo da Ciência da Computação que está preocupado com a automação do comportamento inteligente”.

O *Machine Learning* (ML) ou aprendizado de máquina, pode ser supervisionado ou não supervisionado. Existem algoritmos específicos para cada tipo de aprendizado. Os principais algoritmos de aprendizado supervisionado são: K-NN, Árvore de decisão, Naive Bayes, Perceptron/Adaline e MLP (Multi-Layer Perceptron), já os principais algoritmos de aprendizado não supervisionado são: K-means, Métodos hierárquicos e SOM (AGUIAR, 2019).

Diz-se que um algoritmo usa aprendizado supervisionado quando ele utiliza características prévias para prever o futuro, o modelo mais conhecido é a regressão linear que minimiza os erros quadráticos entre a variável predita e o valor real da variável. Já o aprendizado não supervisionado busca agrupar objetos ou fatos com características semelhantes ou próximas, a técnica mais conhecida é a clusterização ou agrupamento (AGUIAR, 2019).

A Figura 3 abaixo, apresenta os dois principais grupos de técnicas de aprendizado de máquina (*Machine Learning* – ML): Aprendizado Supervisionado (Modelos Preditivos) e o Aprendizado Não Supervisionado (Modelos Descritivos). Existe ainda um terceiro grupo que vem ganhando força nos últimos anos, mas que não abordaremos nesse trabalho, é o Aprendizado por Reforço (*Reinforcement Learning* - RL).

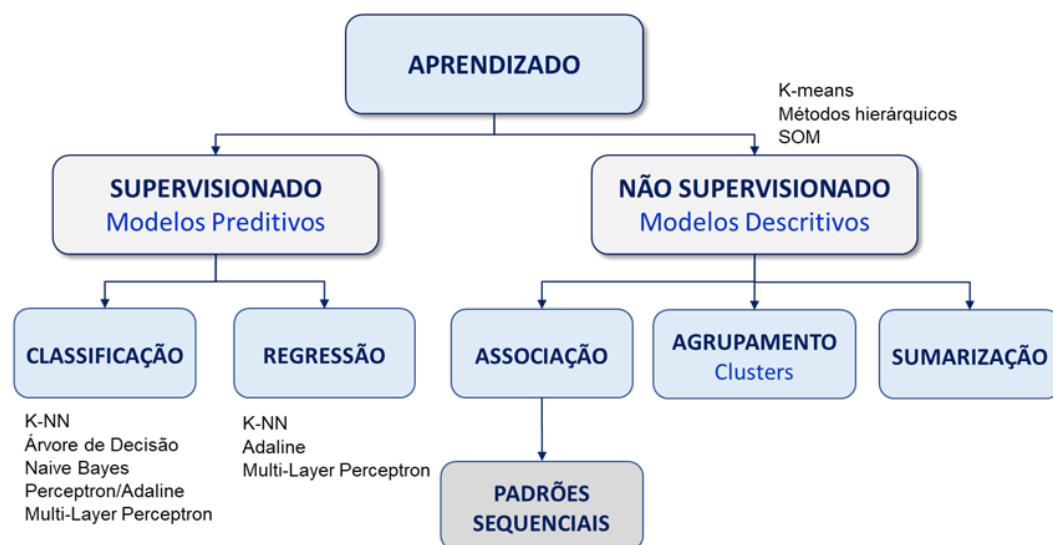


Figura 3 – Principais técnicas e algoritmos de aprendizado de máquina.

Fonte: Elaborado pelo autor.

2.5 Inteligência Empresarial

A Inteligência Empresarial pode ser definida como a capacidade de uma organização em capturar, selecionar, analisar e gerenciar as informações relevantes para a gestão do negócio e definição de suas estratégias organizacionais.

O conceito de Inteligência Empresarial parte do princípio que não basta uma empresa investir em bons sistemas de informação para alcançar uma gestão bem-sucedida. É fundamental que a empresa desenvolva as capacidades necessárias para utilizar adequadamente as informações geradas por estes sistemas para agregar valor ao negócio.

Segundo Teixeira (2009), os principais objetivos da Inteligência Empresarial são:

- Inovar e criar conhecimento até então desconhecido;
- Reduzir riscos na tomada de decisão e evitar surpresas;
- Direcionar, assertivamente, os planos de negócio e a implementação de ações;
- Criar oportunidades de negócios;
- Apoiar o desenvolvimento de produtos/serviços com uma base de informação confiável, eficiente e ágil;
- Monitorar analisar e prever questões relacionadas ao “coração” do negócio;
- Gerar valor aos negócios.

A Inteligência Empresarial busca entender os fatores e processos humanos e organizacionais envolvidos na busca por informações, principalmente dados externos à organização que possibilita e apoia a tomada de decisões.

3 Metodologia

Esta pesquisa utiliza uma abordagem de natureza qualitativa, por ser a mais adequada para investigar um setor específico da economia e para o tratamento das evidências obtidas através de aprofundamento do assunto.

Segundo Creswell (2010), a pesquisa qualitativa é um meio para explorar e para entender o significado que os indivíduos ou grupos atribuem a um problema social ou humano.

A abordagem qualitativa pesquisa detalhadamente os fenômenos do ambiente estudado, o pesquisador vive e conhece a realidade deste grupo ou ambiente. Na pesquisa qualitativa, o pesquisador participa, compreende e interpreta (MICHEL, 2009).

Quanto aos fins, esta pesquisa tem caráter descritivo, pois se propõe a verificar e explicar problemas, fatos ou fenômenos da vida real, com a maior precisão possível, observando e fazendo relações e conexões. Seu objetivo é explicar os fenômenos, relacionando-os com o ambiente. Está relacionada diretamente com a pesquisa qualitativa, pois levanta, interpreta e discute fatos e situações (MICHEL, 2009).

Ainda quanto aos fins é uma pesquisa aplicada, pois tem finalidade prática quando propõe um modelo de combate à “Síndrome do Balde Furado” que aflige às organizações. Segundo Vergara (2007), a pesquisa aplicada é motivada pela necessidade de resolver problemas concretos, mais imediatos, ou não.

Quanto aos meios, para investigação e fundamentação teórica, esta pesquisa utilizou a pesquisa bibliográfica, documental e de campo para dar embasamento teórico ao estudo.

Nesta pesquisa serão apresentados resultados reais da aplicação em uma organização real, entretanto a organização não será identificada e nem o ramo de atividade por questão éticas de pesquisa.

4 Levantamento e Análise dos resultados

A pesquisa proposta, apresenta um modelo ou processo de combate à “Síndrome do Balde Furado” nas organizações. Para tal, foi proposto o uso da Ciência de dados com passos para bem definidos para evitar essa síndrome nas organizações a partir das experiência exitosas dos pesquisadores. A Figura 4 abaixo, apresenta em exemplo de contexto para aplicação da Ciência de dados.

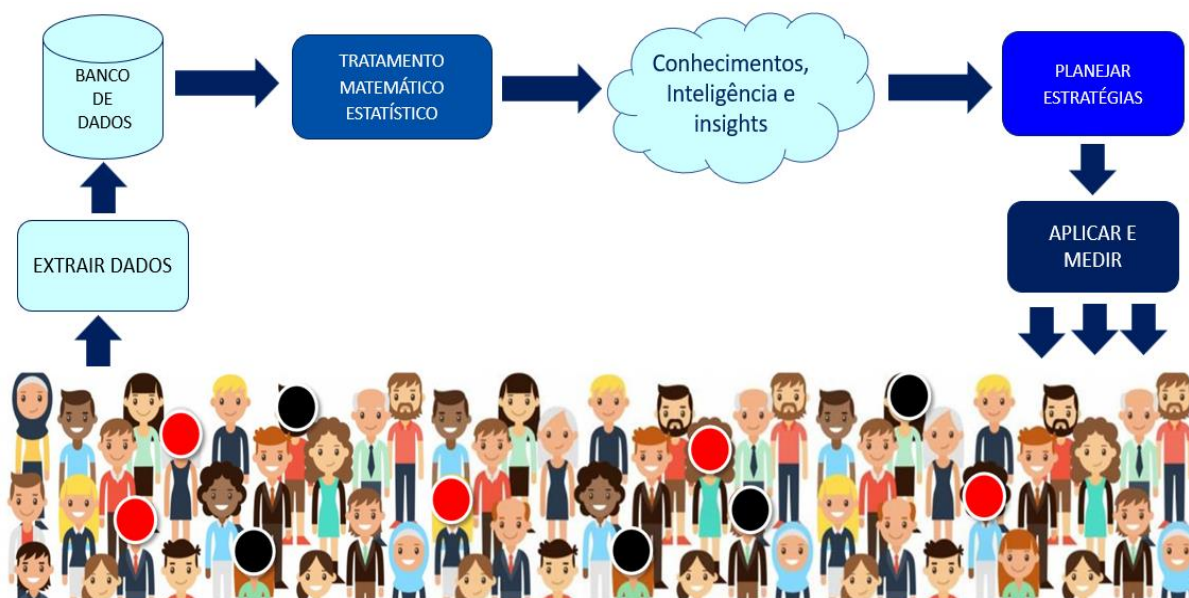


Figura 4 – Fluxo de atividades da Ciência de dados aplica ao contexto do mercado.

Fonte: Elaborado pelo autor.

Na figura acima, podemos observar alguns componentes de mercado e atividades do fluxo da Ciência de Dados para entender esse mercado proposto. Segue a descrição desses componentes e atividades.

- O mercado é formado pela população de uma determinada região, envolvendo pessoas que nunca serão clientes e de pessoas que são possíveis clientes. Nessa população temos ainda, pessoas que se tornaram clientes e permanecem clientes (representados pelas bolas pretas) e de pessoas que são ex-clientes, isto é, abandonaram a nossa corporação (representados pelas bolas vermelhas);

- Quando extraímos os dados (variáveis) que descrevem as características dessas categorias de pessoas do mercado, estamos compondo a nossa base de dados. Estes dados podem ser estruturados ou não-estruturados;

- Com os dados coletados, se faz necessário armazená-los em um banco de dados para futuro processamento;

- O próximo passo é a construção do modelo matemático e da estatística descritiva para fazer emergir conhecimentos até então não detectados;

- A partir daqui, surgem novos conhecimentos, inteligência de negócio e *insights* para aplicação no mercado proposto;

- Com novos conhecimentos e *insights* é possível propor e planejar novas estratégias com embasamento matemático e estatístico;

- A Inteligência Empresarial vai agora aplicar as novas estratégias, medir e acompanhar os resultados alcançados ou não alcançados no mercado proposto;

- Após a avaliação dos resultados, repete-se todo o fluxo para fazer emergir novos conhecimentos e readequar as estratégias.

A Figura 5 abaixo, apresenta de forma reduzida as principais atividades de combate à “Síndrome do Balde Furado” propostas nesse estudo.



Figura 5 – Principais atividades propostas para combater a “Síndrome”.

Fonte: Elaborado pelo autor.

A Figura acima, apresenta os três grupos de atividades que o estudo propõe:

- A estatística descritiva para entender e conhecer o mercado e grupos de clientes;

- A Clusterização para utilizar um recursos de aprendizado não supervisionado da IA para conhecer os grupos e “Personas” de agrupamentos de clientes;

- Por fim, o uso do aprendizado supervisionado da IA para prever os clientes que provavelmente vão cancelar, antes que cancelem e abandonem a nossa organização, usando Redes Neurais Artificiais (RNA).

O Quadro 1 abaixo, apresenta em detalhes as fases propostas no combate a “Síndrome do Balde Furado”: Entrada de clientes para encher o balde; Retenção de clientes para não sair pelos furos do balde; Predição dos clientes que vão sair pelos buracos do balde e Saída de clientes que saíram pelos furos do balde.

FASE	GRUPO	ATIVIDADE	SÍNDROME DO BALDE FURADO
F1	ESTATÍSTICA	Uso da Estatística Descritiva para conhecer o mercado em que desejamos atuar e atrair clientes	ENTRADA DE CLIENTES
F2	CLUSTERIZAÇÃO	Uso da IA, Aprendizado não supervisionado, para conhecer os agrupamentos (Clusters) dos futuros clientes presentes no mercado em que desejamos atuar e atrair clientes	ENTRADA DE CLIENTES
F3	ESTATÍSTICA	Uso da Estatística Descritiva para conhecer os clientes que ENTRARAM na base de clientes (portfólio) nos últimos meses	ENTRADA DE CLIENTES
F4	CLUSTERIZAÇÃO	Uso da IA, Aprendizado não supervisionado, para conhecer os agrupamentos (Clusters) dos clientes que ENTRARAM na base de clientes (portfólio) nos últimos meses	ENTRADA DE CLIENTES
F5	ESTATÍSTICA	Uso da Estatística Descritiva para conhecer os clientes que PERMANECEM na base de clientes (portfólio) por um longo período	RETENÇÃO DE CLIENTES
F6	CLUSTERIZAÇÃO	Uso da IA, Aprendizado não supervisionado, para conhecer os agrupamentos (Clusters) dos clientes que PERMANECEM na base de clientes (portfólio) por um longo período	RETENÇÃO DE CLIENTES
F7	PREDIÇÃO	Uso da IA, Aprendizado supervisionado, para PREDIZER os clientes que vão ABANDONAR a base de clientes (portfólio)	PREDIÇÃO DE CANCELAMENTO
F8	ESTATÍSTICA	Uso da Estatística Descritiva para conhecer os clientes que ABANDONAM a base de clientes (portfólio) nos últimos meses	SAÍDA DE CLIENTES
F9	CLUSTERIZAÇÃO	Uso da IA, Aprendizado não supervisionado, para conhecer os agrupamentos (Clusters) dos clientes que ABANDONARAM a base de clientes (portfólio) nos últimos meses	SAÍDA DE CLIENTES

Quadro 1 – Fases para combater a “Síndrome”.

Fonte: Elaborado pelo autor.

Todas as fases propostas acima geram informações e conhecimentos úteis para o desenvolvimento de estratégias para o combate à “Síndrome do balde furado”. Cada organização deve usar as fases acima de acordo com o seu contexto e ramo de atividades.

A seguir apresentamos uma série de resultados da aplicação com dados reais de algumas fases propostas em uma organização que combate a “Síndrome do balde furado”. Nesta apresentação não será identificada a organização e nem o ramo de atividade por questão ética de pesquisa.

Grupo: Estatística Descritiva

A Fase-1 que usa a Estatística descritiva para conhecer o mercado em que desejamos atuar e atrair clientes para encher o nosso “balde de clientes”, utiliza gráficos como o apresentado abaixo no Gráfico 1, que apresenta um resumo do mercado em esta organização deseja atuar, classificando a população de uma região por “sexo” e “faixa etária”, permitindo entender que público a organização vai encontrar nessa região.

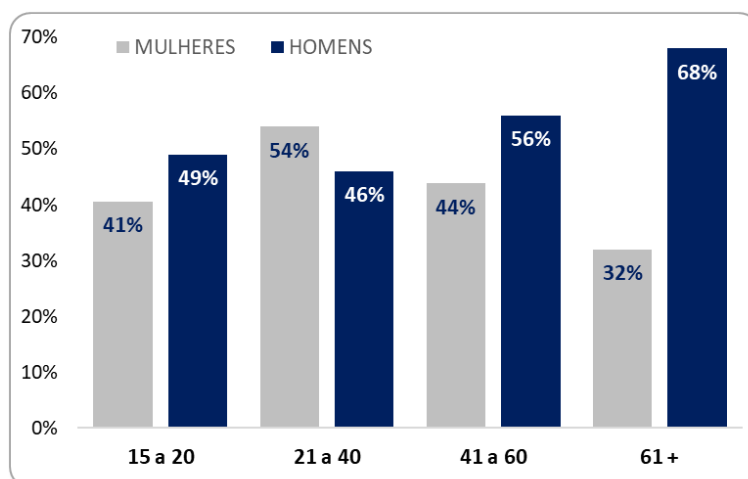


Gráfico 1 – Mercado alvo por SEXO e FAIXA ETÁRIA.

Fonte: Elaborado pelo autor.

A Fase-5 que usa a Estatística descritiva para conhecer os clientes que estão no portfólio de clientes mantendo o “Balde com clientes”, utiliza gráficos como o apresentado abaixo no Gráfico 2, que apresenta um resumo do portfólio de clientes na organização, classificando o portfólio “Faixa de tempo de contrato” versus a “Taxa de reclamação dos clientes”, permitindo avaliar o nível de satisfação dos clientes no seu tempo de vida como cliente.

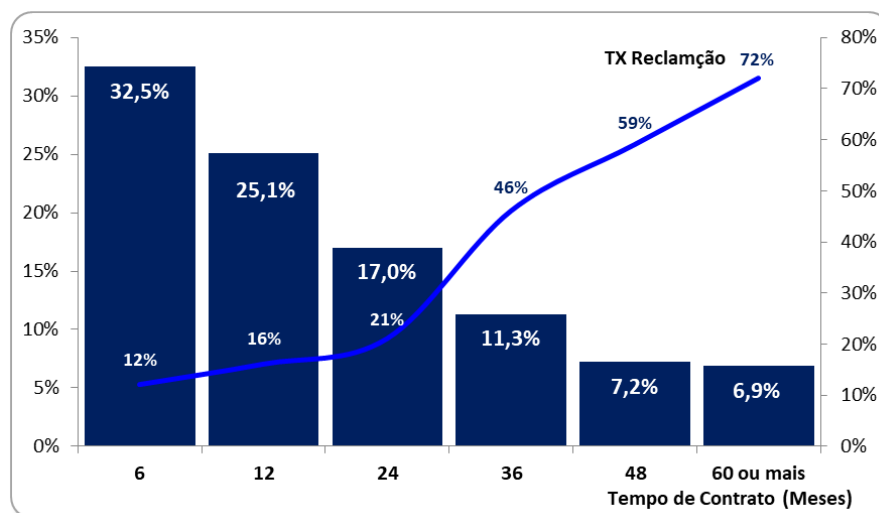


Gráfico 2 – Portfólio de clientes ATIVOS.

Fonte: Elaborado pelo autor.

Observem que a Estatística descritiva exemplificada em dois exemplos acima, nas Fases 1e 5, também deve ser utilizada para gerar conhecimentos nas Fases 3 e 8. Os gráficos 1 e 2 são exemplos simples de sua aplicação, mas existe uma grande variedade de opções de uso para cada tipo de conjunto de variáveis disponíveis e atividades de negócio.

Grupo: Clusterização

As Fases 2, 4, 6 e 9, utilizam a técnica da “Clusterização” ou agrupamento, que utiliza o aprendizado não supervisionado da Inteligência Artificial, isto é, os dados dos clientes são processados por um algoritmo que faz o agrupamentos dos semelhantes utilizando técnicas matemáticas, como a Distância Euclidiana, sem a interferência ou a opinião determinística das opiniões de um ser humano.

A distância euclidiana é uma das medidas de dissimilaridade entre comunidades (agrupamentos) mais utilizadas na prática (GAUCH, 1982). Quanto menor o valor da distância euclidiana entre duas comunidades, mais próximas elas se apresentam em termos de parâmetros quantitativos por classe, logo, quanto menor a distância euclidiana, maior a eficiência do procedimento. A Figura 6 abaixo, apresenta a fórmula de cálculo.

$$DE(x, y) = \sqrt{\sum_i^p (x_i - y_i)^2}$$

Figura 6 – Fórmula do cálculo da distância euclidiana.

Fonte: Prof. Ricardo J. G. B. Campello, 2012.

Na fórmula acima podemos identificar como é realizado o cálculo matemático, onde DE representa Distância Euclidiana dos pontos cartesianos X e Y, em seguida calculamos a raiz quadrada onde dentro da racionalização possui um somatório e por fim entre parênteses os pontos X e Y elevado ao quadrado.

Para o processamento que gera a clusterização utilizamos um script na linguagem Python que inicia com a redução dos dados (variáveis) que queremos estudar em uma par de valores (x, y) para aplicação em um plano cartesiano. O Gráfico 3 abaixo, representa um plano cartesiano com as vendas dos três últimos meses de uma organização, representando a Fase 4.

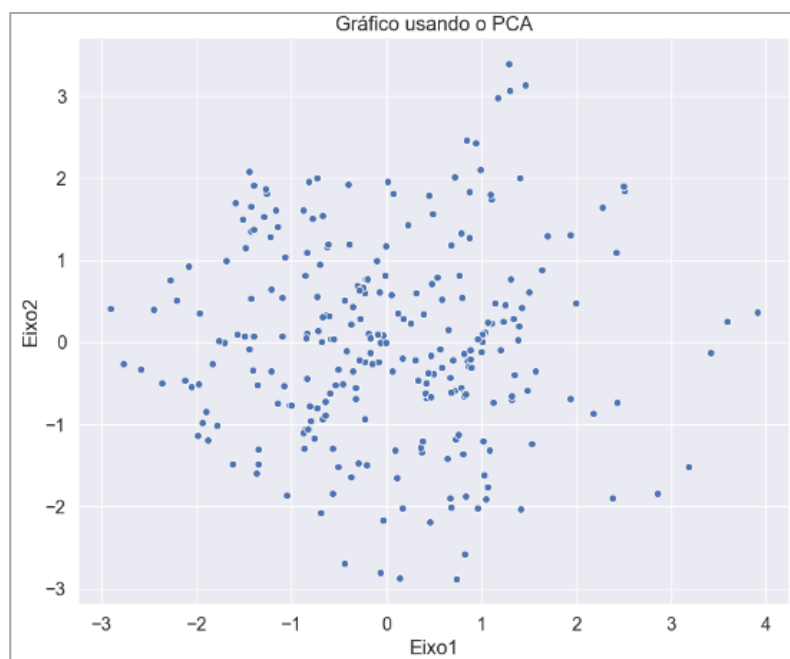


Gráfico 3 – Plano cartesiano com as vendas reais de uma organização.

Fonte: Elaborado pelo autor usando script em Python.

No exemplo a seguir, vamos executar o script do K-means para “clusterizar” os dados das vendas. Agora precisamos definir em quantos grupos iremos dividir as vendas. O Gráfico 4 conhecido como Curva de cotovelo é utilizado como guia para definir a quantidade ideal de agrupamentos para aquele tipo de dados. Esse gráfico indica que quanto mais próximo do vértice (x, y), deve-se colher o valor de K, que será a quantidade ideal de Cluster para aqueles dados estudados. No caso desta pesquisa, a “curva” se encontrou entre 5 e 6 do plano cartesiano.

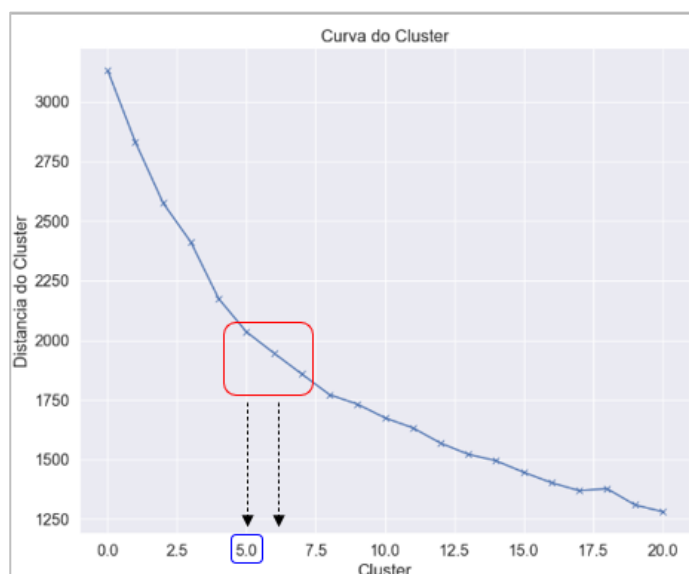


Gráfico 4 – Curva do cotovelo.

Fonte: Elaborado pelo autor usando script em Python.

Com a quantidade de grupos já definido pelo gráfico de cotovelo, escolhemos $K = 5$. Nessa etapa, os dados foram processados pelo algoritmo K-means que é o responsável por calcular a distância Euclidiana entre os pontos e assim definir os 5 (cinco) grupos.

Em um primeiro o momento, o algoritmo foi executado e selecionou 5 cluster aleatórios para utilizar como centroide (ponto referencial) e agrupou os demais pontos. O K-means utilizou cálculos matemáticos baseado em Distância Euclidiana para calcular as semelhanças entre as vendas efetuadas. Para conseguirmos chegar ao resultado mais assertivo possível. O Algoritmo K-means executa *loops* diversas vezes, pois a cada execução do laço, os pontos são recalculados e redirecionados para outros centroides. O algoritmo realizou uma serie de repetição até chegarmos em uma classificação ideal. O Gráfico 5 abaixo nos apresenta a classificação.

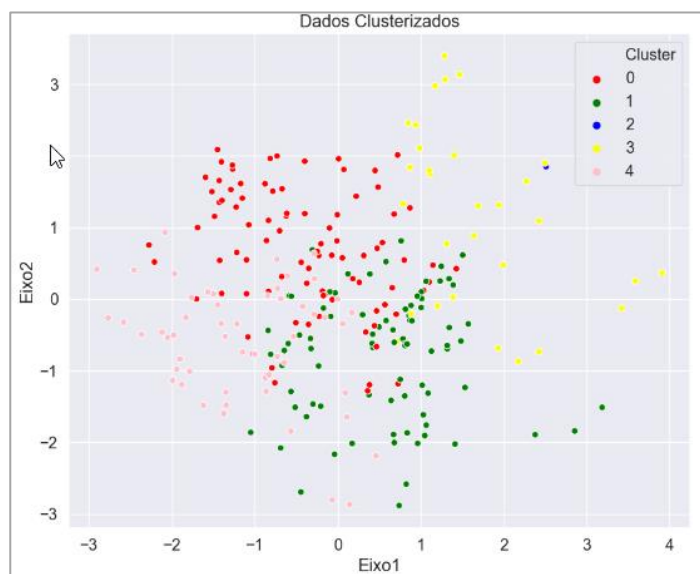


Gráfico 4 – Curva do cotovelo.

Fonte: Elaborado pelo autor usando script em Python.

Agora, vamos usar esse conhecimento gerado. Cada venda receberá um *label* (identificação) do grupo que pertence. Nesse momento deve-se fazer o estudo estatístico de cada grupo separadamente, podendo criar “personas” ou estudos estatísticos para entender as características intrínsecas de cada grupo e tomar decisões baseadas nessas características e novos conhecimentos que surgirão de cada agrupamento.

Grupo: Predição

A Fase-7 é uma das mais importantes para combater a saída de clientes pelos “furos do balde”. O uso das Redes Neurais Artificiais (RNA) nos ajuda a prever, isto é, prever os clientes com maior probabilidade de nos abandonar. O algoritmo RNA é capaz de identificar nas milhares de combinações de dados relacionadas aos clientes que nos abandonaram e nos que permaneceram, qual o padrão matemático e estatístico que mais se assemelha aos clientes que sairão pelos “furos do balde”. Dessa forma, podemos avaliar entre os clientes ativos, aqueles que estão se aproximando do padrão de características dos clientes que nos abandonam, e assim, preparar ações de retenção, antes que estes clientes saiam pelos “furos do balde”.

Para geração dos modelos preditivos foi utilizado um processo de preparação, tratamento, processamento e avaliação dos dados (treinamento e testes) relativos aos clientes, conforme descrito na Figura 7 abaixo.

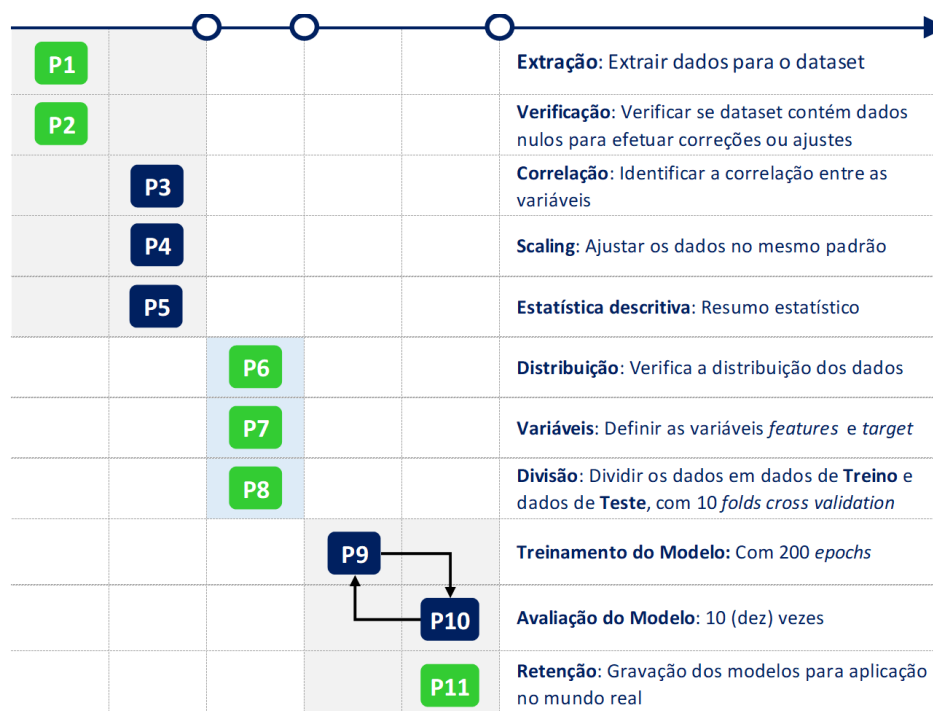


Figura 7 – Fases para geração dos modelos preditivos.

Fonte – Elaborado pelo autor.

Na fase de treinamento e testes, são inseridos no modelos as variáveis explicativas dos clientes: que permanecem e que nos abandonam, para o algoritmo “aprender” qual o padrão de quem sai e quem fica. Assim sendo, para o treinamento do modelo precisamos usar dados dos dois tipos de clientes, e assim, o algoritmo “aprende” sobre as característica de quem sai e quem fica no nosso portfólio.

A Figura 8 abaixo, nos apresenta um exemplo que foi utilizado em uma organização em que os principais dados utilizados no *dataset* foram: tempo de contrato; valor da fatura; status de pagamento; tipo de produto; faixa etária; sexo; região; grau de instrução e a variável *target*: ATIVO ou CANCELADO.

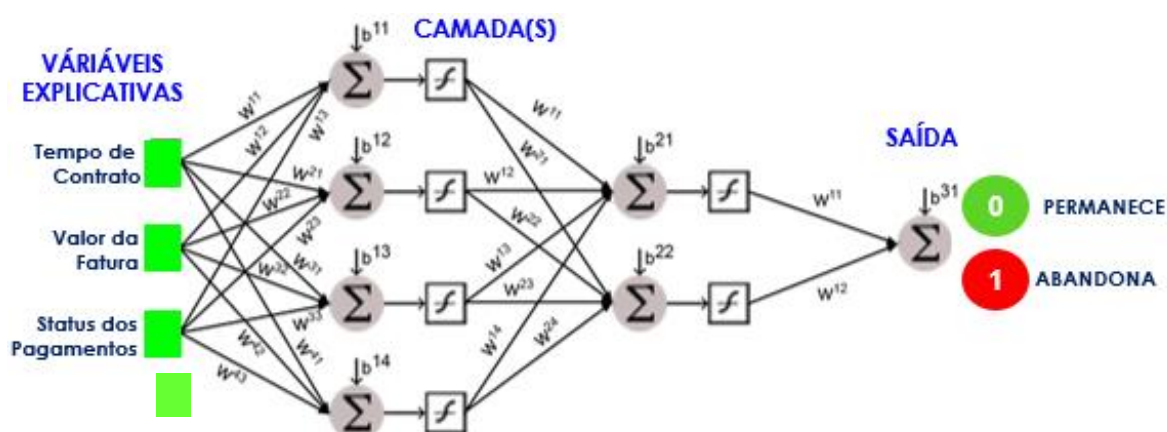


Figura 8 – Modelo de predição usando MLP.
 Fonte – Elaborado pelo autor.

Já a Figura 9 abaixo nos apresenta o processo de como o modelo de IA é preparado até a extração da melhor Matriz de Pesos que será utilizada na predição dos dados, baseado no modelo proposto por Aguiar (2019).

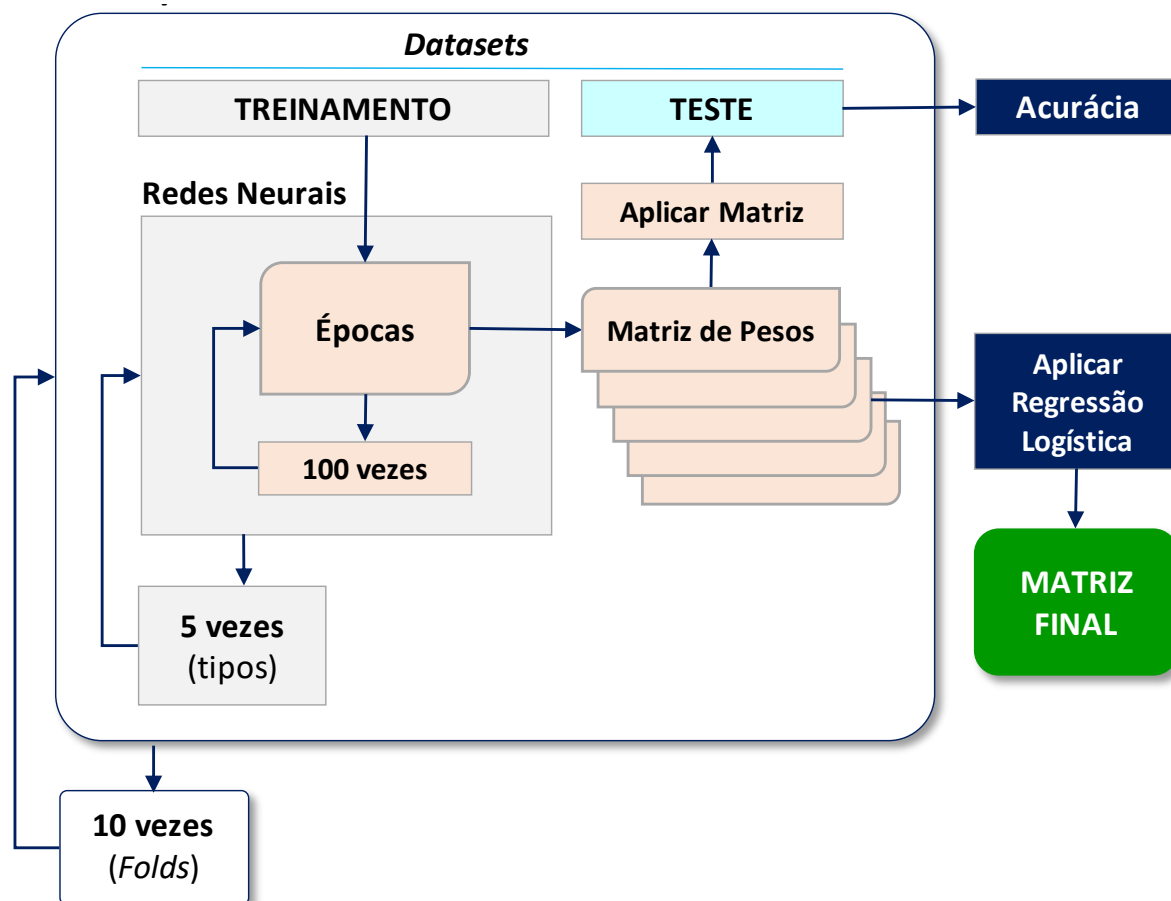


Figura 9 – Processo de Aprendizado e geração da Matriz Final.
 Fonte: AGUIAR (2019).

Com a Matriz de Pesos calculada e selecionada, podemos aplicá-la nos dados de um cliente qualquer da base, que a matriz vai nos responder se esse cliente tem o padrão de “Permanecer no balde” ou “Sair pelos furos do balde”.

5 Considerações finais

O modelo proposto vem se consolidando em vários tipos de negócios, podendo ser aplicado em vários tipos de negócios e organizações: pequenas, médias e grandes. O IA se populariza tornando-se uma tecnologia de baixo custo acessível à vários nichos de negócio. Recentemente aplicamos em uma loja de venda de instrumentos musicais e ao mesmo tempo em uma operadora de planos de saúde.

Os modelos de IA e suas técnicas vem se popularizando no meio acadêmico e profissional, ganhando espaço na geração de conhecimento estratégico. Entretanto, vale salientar que de nada adiantarão os recursos de IA se as áreas de negócio não participarem nos projetos de geração dos modelos, principalmente na definição das dores e na identificação das variáveis explicativas para os fenômenos estudados.

É importante ressaltar que os modelos de IA de nada servem se o estudo do ambiente, do contexto e do negócio não estiver bem-feito e adequado, isto é, algoritmos de IA de nada servirão em estudos onde os dados não explicam o contexto e as características estudadas.

Referências Bibliográficas

AGUIAR, Wellington. S.; Desenvolvimento de modelos preditivos de mortalidade infantil com base em inteligência artificial no estado do Ceará. Tese de Doutorado - Universidade Federal do Ceará, Faculdade de Medicina, Programa de Pós-graduação em Saúde Coletiva, Fortaleza, 2019.

AGUIAR, W. S.; FONTENELLE, J. M. B.; SILVA, J. V. R.; SANTOS, R. F. B. A Practical process for deploying low cost Artificial Intelligence (AI) projects. Anais do 18th CONTECSI FEA USP – INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGY MANAGEMENT. São Paulo-SP, 2021.

ANDRADE, E. L. Introdução à pesquisa operacional: métodos e modelos para análise de decisões. 4. ed. – Rio de Janeiro: LTC, 2009.

CAMPELLO, R. J. G. B. “SCC0173 – Mineração de Dados Biológicos - Preparação de Dados: Parte B”, disponível em http://wiki.icmc.usp.br/images/e/e1/Preparacao_Dados_II.pdf, acesso em 12 de novembro de 2012.

CHARNIAK, Eugene; McDERMOTT, Drew. Introduction to artificial intelligence. Addison-Wesley Logman Publishing Co., Boston, USA, 1987.

CRESWELL, J. W. Projeto de pesquisa: Métodos qualitativo, quantitativo e misto. 3ª ed. Porto Alegre: Ed. Artmed, 2010.

GAUCH, H.G. Multivariate analysis in community ecology. Cambridge: Cambridge University Press, 1982. 298p.

KOTLER, P.; HAYES, T.; BLOOM, P. N. Marketing de Serviços Profissionais: Kotler estratégias inovadoras para impulsionar sua atividade, sua imagem e seus lucros. 2a. ed. São Paulo: Manole, 2002.

KOTLER, Philip. Administração de marketing: a edição do novo milênio. Tradução Bazán Tecnologia e Linguística. 10ª edição. 2000.

KURZWEIL, Raymond. The age of Intelligent Machines. MIT Press, USA, 1990.

LARSON, Ron; FARBER, Betsy. Estatística aplicada. 4ª ed. São Paulo: Pearson Prentice Hall, 2010.

LUGER, George F. Inteligência Artificial. 6ª ed. São Paulo: Pearson Education do Brasil, 2013.

McDONNELL, Robert; GUERRA, Saulo; OLIVEIRA, Paulo Felipe. Ciência de dados em R. 1ª ed. Brasília: Editora IBPAD, 2018.

MICHEL, Maria Helena. Metodologia e Pesquisa Científica em Ciências Sociais: Um guia prático para acompanhamento da disciplina e elaboração de trabalhos monográficos. 2ª ed. São Paulo: Ed. Atlas, 2009.

MORETTIN, Pedro Alberto; SINGER, Júlio da Mota. Estatística e Ciência de dados. 1ª edição. – Rio de Janeiro: LTC, 2022.

RICH, Elaine; KNIGHT, Kavin. Artificial Intelligence. McGraw-Hill Higher Education, USA, 1990.

TEIXEIRA, D.R. Rede de Valor para Inteligência Empresarial. Revista da ESPM, vol. 16, n. 1, 2009.