

Aplicação de *machine learning* na previsão de faturamento de uma franquia de perfumaria e cosméticos: um estudo de caso

Armando Filipe Marques da Cruz

Trabalho de Conclusão de Curso
MBA em Inteligência Artificial e Big Data

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Aplicação de *machine learning* na
previsão de faturamento de uma
franquia de perfumaria e
cosméticos: um estudo de caso

Armando Filipe Marques da Cruz

ARMANDO FILIPE MARQUES DA CRUZ

APLICAÇÃO DE *MACHINE LEARNING* NA PREVISÃO DE FATURAMENTO DE
UMA FRANQUIA DE PERFUMARIA E COSMÉTICOS: UM ESTUDO DE CASO

Trabalho de conclusão de curso apresentado ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Adenilso da Silva Simão

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

M357a MARQUES DA CRUZ, ARMANDO FILIPE
Aplicação de machine learning na previsão de
faturamento de uma franquia de perfumaria e
cosméticos: um estudo de caso / ARMANDO FILIPE
MARQUES DA CRUZ; orientador Adenilso da Silva
Simão. -- São Carlos, 2024.
28 p.

Trabalho de conclusão de curso (MBA em
Inteligência Artificial e Big Data) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2024.

1. Inteligência artificial. 2. Machine learning.
3. Previsão de faturamento. 4. Modelo preditor. I.
Simão, Adenilso da Silva, orient. II. Título.

AGRADECIMENTOS

Neste momento especial da conclusão deste trabalho, gostaria de expressar minha profunda gratidão a todas as pessoas que contribuíram de maneira significativa para esta minha jornada acadêmica até aqui.

Primeiramente, agradeço a Deus, por Sua guia constante, força e inspiração ao longo de todo o percurso. Suas bênçãos e orientações foram fundamentais para que eu chegasse a este ponto.

Ao meu pai, meu grande mentor e fonte de apoio, dedico um agradecimento especial. Suas palavras de encorajamento serviram como combustível para eu seguir firme até o fim.

À minha querida noiva, que esteve ao meu lado durante as horas mais importantes curso, me ajudando sempre, expresso minha profunda gratidão.

RESUMO

Marques da Cruz, Armando Filipe. **Aplicação de *machine learning* na previsão de faturamento de uma franquia de perfumaria e cosméticos: um estudo de caso.** 2024. 28 f. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Este documento apresenta a pesquisa de aplicação de *machine learning* na previsão de faturamento de uma franquia de perfumaria e cosméticos. O estudo científico tem por objetivo realizar a previsão de faturamento de uma loja franqueada do segmento de perfumaria e cosméticos na cidade de Manaus/Amazonas. Realizando a aplicação do modelo de *machine learning* chamado “regressão linear simples”. São apresentados, de forma estruturada, os princípios teóricos, do campo de estudo, na fundamentação teórica, apresentando tópicos como: inteligência artificial, conceitos introdutórios sobre tipos, técnicas, modelos e algoritmos de *machine learning*, além da contextualização da empresa alvo do estudo científico. Com os resultados obtidos, destaca-se a capacidade do modelo em prever os faturamentos com base no número de funcionários em que o usuário adiciona ao modelo preditor. Finalmente, é possível notar também que há possíveis melhorias futuras, como: Levar em consideração o limite do número de funcionários por unidade da loja, considerar vendedores mais experientes e com alto desempenho e também, considerar datas comemorativas como o natal.

Palavras-chave: Inteligência artificial, *Machine learning*, Previsão de faturamento, Modelo preditor.

ABSTRACT

Marques da Cruz, Armando Filipe. **Machine Learning Application in Revenue Forecasting for a Perfumery and Cosmetics Franchise: A Case Study**. 2024. 28 p. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

This document presents the research on the application of machine learning in revenue forecasting for a franchise specializing in perfumery and cosmetics. The scientific study aims to forecast the revenue of a franchised store in the perfumery and cosmetics sector in the city of Manaus, Amazonas. The application of the machine learning model called "simple linear regression" is carried out. The theoretical principles of the field of study are systematically presented in the theoretical framework, covering topics such as artificial intelligence, introductory concepts about types, techniques, models, and algorithms of machine learning, along with contextualization of the company under scientific study. With the obtained results, the model's ability to predict revenues based on the number of employees added to the predictive model is highlighted. Finally, it is also possible to observe potential future improvements, such as taking into account the limit of the number of employees per store unit, considering more experienced and high-performing salespersons, and also factoring in festive dates such as Christmas.

Keywords: Artificial Intelligence, Machine Learning, Revenue Forecasting, Predictive Model.

SUMÁRIO

1	INTRODUÇÃO	6
1.1	TEMA.....	6
1.2	CONTEXTUALIZAÇÃO DO PROBLEMA	6
1.3	HIPÓTESE	7
1.4	OBJETIVO	7
1.5	JUSTIFICATIVA.....	8
2	REFERENCIAL TEÓRICO	8
2.1	EMPRESA ALVO DO ESTUDO DE CASO: DELABOM.....	8
2.2	INTELIGÊNCIA ARTIFICIAL	9
2.2.1	Técnicas de inteligência artificial.....	11
2.3	MACHINE LEARNING	12
2.3.1	Tipos de machine learning	12
2.3.2	Técnicas de <i>machine learning</i>.....	13
2.3.3	Modelos de <i>machine learning</i>.....	14
2.3.4	Algoritmos de <i>machine learning</i>	15
2.4	APROFUNDAMENTO TEÓRICO	16
2.4.1	Regressão linear.....	16
2.4.2	Ordinary Least Squares (OLS).....	18
2.4.3	Avaliadores de desempenho	19
3	METODOLOGIA.....	21
4	RESULTADOS	26
5	CONCLUSÃO.....	27
	REFERÊNCIAS.....	28

1 INTRODUÇÃO

1.1 TEMA

Aplicação de *machine learning*¹ na previsão de faturamento de uma franquia de perfumaria e cosméticos: um estudo de caso.

1.2 CONTEXTUALIZAÇÃO DO PROBLEMA

No mundo atual altamente dinâmico e competitivo, a previsão de vendas desempenha um papel muito importante e valorizado para as empresas. A capacidade de antecipar com precisão as tendências de demanda e ajustar estratégias de negócios de acordo se torna fundamental.

A previsão de vendas desempenha um papel fundamental no planejamento e na tomada de decisões das empresas, uma vez que afeta diretamente o desempenho financeiro e a eficiência operacional. No entanto, a natureza complexa e dinâmica do mercado torna desafiador obter previsões precisas e confiáveis.

Previsões precisas também permitem esforços de marketing mais eficazes, como promoções direcionadas e estratégias de preços. Isso pode resultar em aumento de vendas, fidelidade do cliente e reconhecimento da marca, levando ao aumento da receita. (OLIVEIRA, ANA., 2023)

¹ *Machine Learning*: Processo de treinamento de algoritmos para aprender padrões e tomar decisões sem programação explícita.

Assim, se faz necessário uma ferramenta de previsão de faturamento aplicando *machine learning*, uma abordagem promissora para lidar com essa complexidade e melhorar a precisão das previsões de vendas.

1.3 HIPÓTESE

Para tal aplicação, de previsão de faturamento, é possível realizar uma pesquisa, onde será aplicado uma das técnicas de inteligência artificial, a técnica de *machine learning*, utilizando como modelo a regressão linear simples, tendo como a variável independente, o número de funcionários, e a variável dependente, o valor do faturamento (em R\$).

1.4 OBJETIVO

Este estudo de caso tem como objetivo realizar a previsão de faturamento de uma loja franqueada do segmento de perfumaria e cosméticos na cidade de Manaus/Amazonas. Realizando a aplicação do modelo de *machine learning*: regressão linear simples espera-se obter os benefícios dessa abordagem em um contexto empresarial específico.

Portanto, este estudo busca contribuir para o avanço do conhecimento e fornecer *insights* valiosos para aprimorar as estratégias da empresa, beneficiando o desempenho e a competitividade. Porém, este estudo não tem a intenção de apresentar um modelo de previsão de faturamento perfeito, capaz de acertar em 100% das ocasiões. Seu propósito é fornecer uma orientação para os tomadores de decisões da empresa.

1.5 JUSTIFICATIVA

A justificativa para este estudo reside na importância da previsão de faturamento para o sucesso das empresas, mais especificamente, na importância de se preparar financeiramente diante à uma previsão de faturamento. Com essa previsão em mãos, a empresa pode otimizar o planejamento estratégico, a gestão de estoque e a alocação de recursos.

A aplicação de *machine learning* oferece um potencial significativo para aprimorar a precisão das previsões, considerando a complexidade dos fatores que afetam as vendas. Com uma previsão mais confiável, as empresas podem tomar decisões embasadas, evitando estoques excessivos ou escassez, reduzindo custos e maximizando os lucros.

2 REFERENCIAL TEÓRICO

Dentro deste referencial, serão apresentados os aspectos teóricos do tema relevantes para a pesquisa. Inicialmente, será contextualizado sobre a empresa alvo do estudo de caso e a inteligência artificial. Em seguida, serão abordados os conceitos introdutórios sobre tipos, técnicas, modelos e algoritmos de *machine learning*.

2.1 EMPRESA ALVO DO ESTUDO DE CASO: DELABOM

A DELABOM é uma empresa manauara franqueada da *Yes! Cosmetics*, cuja a atuação no ramo de cosméticos verbera há mais de 20 anos. Fundada em 2000, a *Yes! Cosmetics* se destaca pela oferta de uma ampla gama de produtos de

beleza, incluindo itens de cuidados com a pele, maquiagem, perfumes e produtos para o cabelo.

[...] a *Yes! Cosmetics* atua no segmento da beleza, com itens de perfumaria, pessoal, maquiagem e acessórios, além da linha assinada por Sabrina Sato. Investindo em alta qualidade e preços acessíveis, a marca acompanha as principais tendências mundiais. Em 2016, a *Yes!* entrou no *franchising*² e já tem mais de 100 unidades em todo o Brasil. (*YES COSMETICS, s.d.*)

Com uma filosofia voltada para a valorização da autoestima e o bem-estar dos seus clientes, a *Yes! Cosmetics* busca proporcionar produtos de qualidade e inovação, aliados a preços acessíveis. A empresa busca constantemente acompanhar as tendências do mercado e as necessidades dos consumidores, buscando desenvolver produtos que atendam às suas expectativas e preferências.

Possuindo uma presença significativa no mercado nacional, contando com uma extensa rede de revendedores espalhados por todo o país. Essa estratégia de venda direta, baseada no relacionamento próximo com os clientes, tem sido uma das chaves do sucesso da empresa. Além disso, a marca também possui lojas físicas e uma loja virtual, oferecendo opções de compra convenientes para os consumidores.

2.2 INTELIGÊNCIA ARTIFICIAL

Inteligência artificial, para Russell e Norvig (2020), se tratava do estudo de como fazer os computadores realizarem tarefas que, até o momento, requerem inteligência humana, ou seja, refere-se a busca por criar sistemas computacionais capazes de realizar tarefas complexas que normalmente exigiriam o uso de

² *Franchising*: Também conhecido como franquia, é um modelo de negócio em que uma empresa (franqueador) concede a outra pessoa ou empresa (franqueado) o direito de utilizar sua marca comercial, métodos de operação e know-how, em troca de pagamentos e cumprimento de determinadas obrigações.

inteligência humana, como reconhecimento de padrões, tomada de decisões, aprendizado e resolução de problemas.

A inteligência artificial pode ser aplicada em diversos campos, como finanças, saúde, manufatura, transporte e segurança. Ela pode ser utilizada para automatizar processos, analisar grandes quantidades de dados, identificar padrões, prever comportamentos e tomar decisões com base em informações precisas e confiáveis. Em resumo, a inteligência artificial tem o potencial de transformar a maneira como vivemos, trabalhamos e interagimos uns com os outros.

A inteligência artificial pode ser categorizada em três tipos:

- **Artificial Narrow Intelligence (ANI):** Em tradução livre, “Inteligência Artificial Limitada”, este tipo é projetado para realizar uma tarefa específica ou um conjunto simples de tarefas, como: reconhecimento de fala/imagem e processamento de linguagem natural (NLP). A ANI é atualmente o tipo mais comum e dominante de inteligência artificial.
- **Artificial General Intelligence (AGI):** A AGI (ou Inteligência Artificial Geral) foca em imitar uma inteligência semelhante à dos humanos, no qual seja possível realizar uma gama de tarefas em diferentes domínios. Sendo esse tipo de inteligência um conceito hipotético e não está totalmente realizada.
- **Artificial Superintelligence (ASI):** Este tipo refere-se a uma inteligência artificial bastante hipotética que supera a inteligência humana em todas as áreas, incluindo: resolução de problemas, criatividade, e habilidades sociais. A ASI (ou Superinteligência Artificial) também é um conceito que ainda não é concretizada.

Para esta pesquisa, será trabalhada dentro dos limites da inteligência artificial do tipo ANI.

2.2.1 Técnicas de inteligência artificial

As técnicas de inteligência artificial podem ser amplamente classificadas, sendo uma das categorias as seguintes:

- **Machine Learning:** *machine learning* envolve treinar um sistema de computador para aprender com dados e melhorar seu desempenho ao longo do tempo.
- **Rede Neural Artificial (RNA):** RNAs são modelos inspirados na biologia que são projetados para imitar a estrutura e função do cérebro humano. São usados para tarefas como reconhecimento de imagens e fala.
- **Natural Language Processing (PLN):** Trata-se de um subcampo da inteligência artificial que lida com a interação entre computadores e humanos usando linguagem natural (linguagem que se fala naturalmente pelos humanos).
- **Robótica:** A robótica envolve o uso de inteligência artificial para controlar e automatizar robôs para tarefas como manufatura, saúde e exploração espacial.
- **Sistemas baseados em regras:** Esses sistemas usam um conjunto de regras predefinidas para realizar uma tarefa ou tomar uma decisão.
- **Métodos de busca e otimização:** Essas técnicas são usadas para encontrar a melhor solução para um problema, buscando em um grande espaço de possíveis soluções.
- **Sistemas especialistas:** Esses são sistemas que usam uma base de dados e um conjunto de regras para simular as habilidades de tomada de decisão de um especialista humano em um domínio específico.

2.3 MACHINE LEARNING

Como mencionado anteriormente no texto, *Machine Learning* é um subcampo da inteligência artificial que, segundo a XP Educação, (2022): “[...] usa máquinas que têm a capacidade de aprender padrões a partir de uma análise de dados e conseguem chegar a um resultado, servindo de solução para um problema [...]”. Ou seja, se concentra em treinar algoritmos para aprender a partir de dados e tomar decisões ou fazer previsões sem serem explicitamente programados para realizar uma tarefa específica.

Machine Learning nada mais é que um ramo dentro da tecnologia, mais especificamente da inteligência artificial, que usa máquinas que têm a capacidade de aprender padrões a partir de uma análise de dados e conseguem chegar a um resultado, servindo de solução para um problema que você tenha. (XP EDUCAÇÃO, 2022).

A seguir, será apresentada uma visão panorâmica do *machine learning*, desde sua perspectiva mais abrangente até sua granularidade mais detalhada, semelhante a uma boneca russa³, passando camada por camada.

2.3.1 Tipos de *machine learning*

Refere-se às categorias gerais que descrevem diferentes abordagens do aprendizado de máquina com base na natureza do problema e no tipo de resposta desejada. Existem três tipos principais de *machine learning*:

³ Boneca russa: Também conhecida como *matrioska* ou *babushka*. É um conjunto de bonecas oco empilhadas uma dentro da outra, formando uma sequência de bonecas aninhadas.

- ***Supervised Learning*** (Aprendizado Supervisionado) – Treinado com um conjunto de dados rotulados e é capaz de prever novas saídas a partir de novas entradas.
- ***Unsupervised Learning*** (Aprendizado Não-Supervisionado) – Já este, é treinado em um conjunto de dados não rotulados e tenta encontrar padrões ou estruturas significativas nos dados.
- ***Reinforcement Learning*** (Aprendizado por Reforço) – Finalmente, este aprende a tomar decisões em um ambiente específico através de tentativa e erro, recebendo recompensas ou penalidades para suas ações.

Cada tipo de *machine learning* é adequado para diferentes tipos de tarefas e problemas, e a escolha do tipo de algoritmo depende do conjunto de dados e do objetivo final do projeto de inteligência artificial. Para esta pesquisa, o tipo de *machine learning* será o aprendizado supervisionado.

2.3.2 Técnicas de *machine learning*

As técnicas de *machine learning* referem-se aos métodos e procedimentos específicos utilizados para implementar a abordagem que irá ser adotada. Essas técnicas são aplicadas dentro de um determinado tipo de aprendizado e definem como os dados são processados e modelados. Como a pesquisa será focada no tipo de aprendizagem supervisionada, serão listados alguns exemplos de técnicas deste tipo:

- **Regressão Linear** – Busca estabelecer uma relação linear entre uma variável dependente (ou alvo) e uma ou mais variáveis independentes (ou características). É uma das técnicas mais amplamente utilizadas em *machine learning* e análise estatística.
- **Árvores de Decisão** – Também conhecido como “*Decision Trees*”, utiliza uma estrutura de árvore para tomar decisões com base em condições e atributos

dos dados. Essa técnica divide os dados em diferentes caminhos de decisão com base em regras e características específicas (como ramos de galhos de uma árvore), permitindo a classificação ou previsão de novos dados com base nas decisões tomadas ao longo da árvore.

- **Redes Neurais** – Esta é inspirada no funcionamento do cérebro humano. Ela consiste em modelos matemáticos compostos por camadas de neurônios artificiais interconectados. Basicamente, cada neurônio processa informações e passa para o próximo, permitindo o aprendizado e a capacidade de lidar com problemas complexos e não lineares. Utilizadas para tarefas como: classificação, previsão e reconhecimento de padrões.

Para esta pesquisa, a técnica de *machine learning* utilizada será a de regressão linear.

2.3.3 Modelos de *machine learning*

O modelo é uma estrutura matemática que captura a relação entre as entradas e saídas dos dados de treinamento. A seguir, será mostrado dois exemplos presentes dentro de cada técnica apresentada anteriormente:

- **Regressão Linear Simples** – Trata-se um modelo que envolve uma única variável preditora e uma variável de resposta.
- **Regressão Linear Polinomial** – Modelo que utiliza uma relação polinomial entre as variáveis preditoras e a variável de resposta, permitindo capturar relações não-lineares.
- **Árvore de Decisão** – É o modelo básico da técnica de árvore de decisão, onde os dados são divididos em nós de decisão com base em condições e características específicas.

- **Random Forest** – Modelo que combina várias árvores de decisão para realizar previsões. Cada árvore é treinada em uma amostra aleatória dos dados e suas previsões são agregadas para obter uma previsão final.
- **Multilayer Perceptron (MLP)** – Modelo de rede neural *feedforward*⁴ com várias camadas ocultas de neurônios interconectados.
- **Convolutional Neural Network (CNN)** – É um modelo de rede neural projetado especialmente para processamento de imagens, utilizando camadas convolucionais⁵ para extrair características das imagens.

Para esta pesquisa, o modelo a ser utilizado será o de regressão linear simples.

2.3.4 Algoritmos de *machine learning*

Os algoritmos em *machine learning* referem-se às sequências de passos computacionais usados para treinar o modelo. Estes algoritmos implementam as técnicas de *machine learning*, definindo como os dados são processados e como o modelo é ajustado aos dados de treinamento. Tendo alguns exemplos como:

- **Ordinary Least Squares (OLS)** – Algoritmo que busca minimizar a soma dos quadrados dos resíduos entre as previsões do modelo e os valores reais da variável de resposta (será visto em mais detalhes posteriormente).
- **Gradient Descent** – Trata-se de um algoritmo iterativo que busca encontrar os coeficientes do modelo de regressão linear minimizando a função de perda.

⁴ *Feedforward*: Arquitetura de rede neural em que os dados fluem em uma única direção, sem retroalimentação.

⁵ Convolução: É uma operação matemática que combina funções locais para processar dados multidimensionais e extrair características relevantes.

- **Iterative Dichotomiser 3 (ID3)** – Clássico para construção de árvores de decisão, que utiliza o critério de ganho de informação para escolher as melhores divisões nos nós de decisão.
- **Perceptron** – Sendo um algoritmo de aprendizagem de rede neural de uma única camada, tem como função a ativação linear, utilizado para classificação binária.

O algoritmo a ser utilizado nesta pesquisa será o OLS.

De forma resumida, quanto os conceitos apresentados anteriormente, veja o Quadro 1:

Quadro 1 - Resumo da taxonomia de *machine learning*.

TIPOS	Definem as categorias gerais de abordagens.
TÉCNICAS	São os métodos específicos usados dentro dessas abordagens.
MODELOS	São as representações aprendidas a partir dos dados.
ALGORITMOS	São as implementações computacionais que treinam os modelos utilizando as técnicas específicas.

Fonte: Autoria própria.

2.4 APROFUNDAMENTO TEÓRICO

2.4.1 Regressão linear

Na regressão linear, o objetivo é encontrar uma equação linear que melhor se ajuste aos dados fornecidos. Essa equação linear é usada para fazer previsões ou estimativas da variável dependente com base nas variáveis independentes. A equação linear geralmente é representada como:

$$y = b_0 + a_0x_0 + a_1x_1 + \dots + a_nx_n$$

Onde:

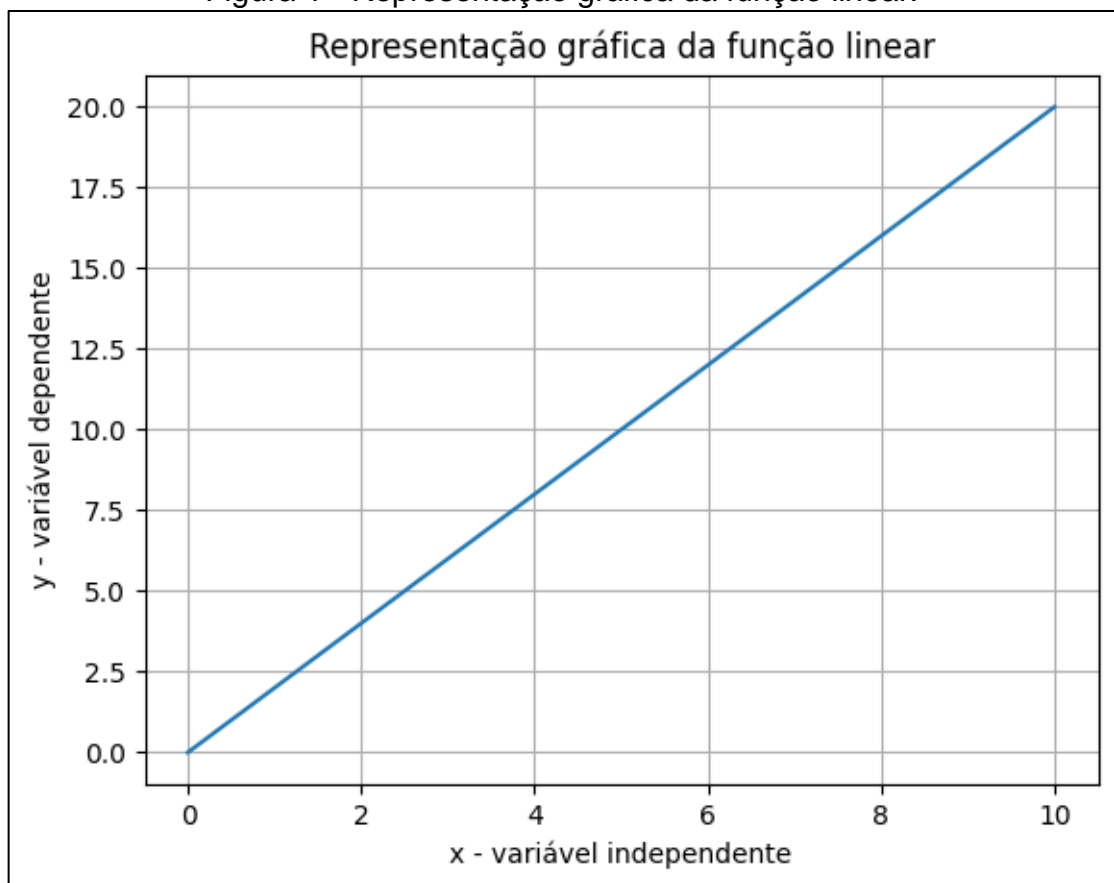
y : Variável dependente (alvo), também conhecido como $f(x)$.

b_0 : Coeficiente de intercepção, também chamado de coeficiente linear.

$a_0x_0 + a_1x_1 + \dots + a_nx_n$: Coeficientes das variáveis independentes x_1, x_2, \dots, x_n , também conhecido como coeficiente angular.

A Figura 1 a seguir representa como esta função se representa graficamente. Neste gráfico presente na figura, os pontos (x e y) formam uma linha reta, indicando que a cada aumento constante em uma variável, a outra também aumenta de forma constante.

Figura 1 - Representação gráfica da função linear.



Fonte: Autoria própria.

O objetivo de utilizar a regressão linear em um modelo de *machine learning*, é encontrar os melhores valores para os coeficientes linear e angular de forma que a

soma dos erros quadráticos⁶ entre os valores previstos e os valores reais seja minimizada. Essa minimização é geralmente feita utilizando-se métodos como o método OLS ou técnicas de otimização, como *Gradient Descent*.

Uma vez que o modelo de regressão linear é treinado, ele pode ser usado para fazer previsões ou inferências com base em novos dados.

2.4.2 Ordinary Least Squares (OLS)

OLS ou em português brasileiro, Mínimos Quadrados Ordinários, é um algoritmo utilizado para ajustar um modelo de regressão linear aos dados observados. Ele tem como objetivo encontrar os coeficientes que minimizam a soma dos quadrados dos resíduos entre os valores previstos pelo modelo e os valores reais dos dados.

O funcionamento do OLS pode ser dividido em algumas etapas, descrita no Quadro 2:

Quadro 2 - Etapas do funcionamento do OLS.

ETAPAS	DESCRIÇÃO
Formulação do modelo	Antes de aplicar o OLS, é necessário formular o modelo de regressão linear que representa a relação entre as variáveis dependentes e independentes. O modelo geralmente é expresso como uma equação matemática, onde as variáveis independentes são multiplicadas pelos seus respectivos coeficientes e somadas a um termo constante.
Cálculo dos resíduos	Os resíduos são a diferença entre os valores observados e os valores previstos pelo modelo. Para cada observação dos dados, calcula-se o resíduo subtraindo o valor real do valor previsto pelo modelo.
Minimização da função de perda	Essa função é conhecida como a função de custo ou função de erro. O algoritmo utiliza métodos de otimização para encontrar os valores dos coeficientes que minimizam essa função.
Estimativa dos coeficientes	Utiliza-se a matriz dos valores das variáveis independentes e o vetor das variáveis dependentes para estimar os coeficientes do modelo. A estimativa dos coeficientes é calculada usando métodos matemáticos, como a inversão de matrizes.

⁶ Erro quadrático: São métricas usadas para quantificar a diferença entre os valores observados e os valores previstos em um modelo, elevando essa diferença ao quadrado para enfatizar erros maiores. O conceito será mais bem desenvolvido na seção 2.4.3.

Avaliação do modelo	Após estimar os coeficientes, é importante avaliar a qualidade do modelo ajustado. Isso pode ser feito através da análise dos resíduos, que devem ser aleatoriamente distribuídos ao redor de zero e não devem apresentar padrões.
Previsões	Uma vez que os coeficientes foram estimados e o modelo foi avaliado, é possível realizar previsões utilizando novos valores das variáveis independentes. Os coeficientes estimados são aplicados à fórmula do modelo para calcular os valores previstos correspondentes.

Fonte: Autoria própria.

Este algoritmo é amplamente utilizado devido à sua simplicidade e interpretabilidade. Ele fornece estimativas eficientes dos coeficientes do modelo que minimizam a soma dos quadrados dos resíduos, permitindo a análise da relação linear entre as variáveis independentes e dependentes.

2.4.3 Avaliadores de desempenho

Avaliadores de desempenho são métricas utilizadas para medir o quão bem um modelo de *machine learning* está realizando as suas previsões ou classificações. Eles fornecem uma medida objetiva para avaliar a precisão e a qualidade das previsões geradas pelo modelo. No caso desta pesquisa, trata-se de avaliadores de desempenho de regressão linear supervisionadas. Tendo como principais exemplos:

- **R²** – “O R-quadrado (também conhecido como coeficiente de determinação) é uma medida de quão bem o modelo de regressão se encaixa nas variáveis dependentes” (ZIEGER, K. *et al.*, p. 47). Sendo uma métrica que varia de 0 a 1 e indica o quanto as variações na variável dependente (alvo) podem ser explicadas pelas variáveis independentes (características) do modelo.

Quanto mais próximo de 1, melhor é o ajuste do modelo aos dados. Um valor de 0 significa que o modelo não é capaz de explicar nenhuma variação nos dados. Por exemplo, se o R² de um modelo de previsão de vendas for 0,8, isso significa que 80% das variações nas vendas são explicadas pelas variáveis

do modelo, enquanto os 20% restantes são atribuídos a outros fatores não incluídos no modelo.

- **Mean Absolute Error (MAE)** – Chamado de “erro médio absoluto”, trata-se de uma métrica que mede a média dos erros absolutos entre as previsões do modelo e os valores reais. Ele fornece uma medida direta do tamanho médio dos erros de previsão, sem considerar sua direção (positiva ou negativa). Por exemplo, se o MAE de um modelo de previsão de vendas for 100, isso significa que, em média, as previsões do modelo têm uma diferença absoluta de 100 unidades em relação às vendas reais.
- **Coeficiente de Correlação (Pearson)** – Trata-se de uma medida estatística que avalia a força e a direção de uma relação linear entre duas variáveis. Variando de “-1” a “1”. Por exemplo: “1” indica uma correlação positiva perfeita (à medida que uma variável aumenta, a outra também aumenta). Por outro lado, “-1” indica uma correlação negativa perfeita e “0” indica ausência de correlação linear (ou seja, as duas variáveis não têm uma relação linear clara).
- **Mean Squared Error (MSE)** – Segundo (ZIEGER, K. *et al.*, 2021):

O erro quadrático médio (MSE) é a média do quadrado das diferenças entre o valor real e o valor previsto. O valor do MSE é sempre positivo, e um valor mais próximo de zero sugere um melhor modelo de regressão. O quadrado da diferença garante que os erros positivos e negativos não se anulem.

Ao elevar os erros ao quadrado, o MSE penaliza mais fortemente os erros maiores, ampliando sua influência na métrica. Por exemplo, se o MSE de um modelo de previsão de vendas for 5.000, isso significa que, em média, as previsões do modelo têm uma diferença quadrática média de 5.000 unidades em relação às vendas reais.

3 METODOLOGIA

A metodologia para esta pesquisa está descrita, em forma de ficha técnica no Quadro 3 e Quadro 4, a seguir:

Quadro 3 - Tipo de pesquisa.

TIPO DE PESQUISA	
Abordagem	Mista
Procedimento	Pesquisa bibliográfica e Estudo de caso

Fonte: Autoria própria.

Quadro 4 - Método de pesquisa.

MÉTODO DE PESQUISA	
Abordagem	Hipotético-dedutivo
Procedimento	Monográfico

Fonte: Autoria própria.

Em relação à coleta de dados, será utilizada a documentação indireta e a análise e interpretação de dados, ocorrerá de forma qualitativa de maneira global. Será realizada uma revisão de literatura nas áreas de inteligência artificial e *machine learning*.

Para realizar esta pesquisa, serão conduzidas observações e coleta de dados nas unidades DELABOM (franqueada Yes! Cosmetics) em Manaus/AM. A linguagem de programação escolhida para análise será o Python. Utilizaremos técnicas de inteligência artificial, mais especificamente, o *machine learning*. O tipo de aprendizado deste *machine learning* utilizado será o supervisionado, com a técnica de regressão linear. O modelo adotado será o de regressão linear simples, aplicando o algoritmo OLS.

A implementação a proposta da pesquisa será utilizando as seguintes etapas:

1. Definição do problema:

O objetivo deste trabalho é desenvolver um previsor de faturamento utilizando regressão linear como técnica de *machine learning* para estimar o faturamento com base no número de funcionários.

- Variável dependente: Faturamento.
- Variável independente: Número de funcionários.

2. Coleta de dados:

Coletar os dados históricos de faturamento mensal e número de funcionários (no caso do trabalho, estes dados são confidenciais e foram enviados pela empresa e contemplavam no intervalo de outubro de 2021 até dezembro de 2023).

3. Pré-processamento dos dados:

Realizar uma análise exploratória dos dados para identificar possíveis *outliers* (pontos de dados que se afastam significativamente do padrão geral do conjunto de dados), valores discrepantes ou padrões.

Tratar e limpar os dados, removendo *outliers*, preenchendo valores ausentes e realizando transformações necessárias.

4. Divisão dos dados:

Separar os dados em conjunto de treinamento e conjunto de teste.

Para este trabalho, foi utilizado a proporção regular e normalmente empregada de 70% para treinamento e 30% para teste.

5. Construção do modelo:

Utilizar a técnica de regressão linear, que se enquadra no aprendizado supervisionado, para construir o modelo de previsão de vendas.

Especificamente, utilizar o modelo de regressão linear simples, que envolve uma única variável independente (número de funcionários).

6. Treinamento do modelo:

Treinar o modelo utilizando o algoritmo OLS. O algoritmo OLS ajustará os coeficientes do modelo para minimizar a soma dos quadrados dos resíduos entre o faturamento real e o faturamento previsto.

7. Avaliação do modelo:

Avaliar o desempenho do modelo utilizando métricas apropriadas, como o coeficiente de determinação (R^2), erro médio absoluto (MAE) e Coeficiente de Correlação (*Pearson*).

8. Ajuste do modelo:

Se necessário, realizar ajustes no modelo, como a inclusão de variáveis adicionais, transformações dos dados ou utilização de outros métodos de regularização.

9. Previsões:

Utilizar o modelo ajustado para fazer previsões de faturamento com base no número de funcionários fornecidos pelo usuário. Comparar as previsões com os valores reais para avaliar a precisão do modelo.

A seguir, será apresentado as etapas da implementação deste trabalho em código. A plataforma utilizada foi o Google Colab (o código completo está disponível em: <https://shorturl.at/jwNQW>).

1. Importação dos módulos

Figura 2 - Importação dos módulos.

```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression

from sklearn.metrics import mean_absolute_error, r2_score
```

Fonte: Autoria própria.

2. Processamento e tratamento de dados

Figura 3 - Processamento e tratamento de dados.

```
# Baixando a base de dados
sheet_id = "1F7qCCY6A00kD3dxELuxzG_tZYfFqh8UP"
df_raw = pd.read_excel(f"https://docs.google.com/spreadsheets/export?id={sheet_id}&format=xlsx", header=2)
df_raw.head()

# Deletando linhas e colunas vazias
df_tratado = df_raw.dropna(how="all", axis=0)
df_tratado = df_raw.dropna(how="all", axis=1)
df_tratado.head(15)

# Adicionando colunas: "Ano" e "Mês"
df_tratado["Data"] = pd.to_datetime(df_tratado["Data"], format='%m/%d/%Y')
df_tratado['Ano'] = df_tratado['Data'].dt.year
df_tratado['Mês'] = df_tratado['Data'].dt.month
df_tratado.head(15)
```

Fonte: Autoria própria.

3. Machine learning

3.1. Verificação da correlação da variável dependente e independente

Figura 4 - Verificação da correlação da variável dependente e independente.

```
correlacao = df_tratado["Funcionários"].corr(df_tratado["Faturamento"])
print("Correlação: {:.2f}%".format(correlacao))
```

Fonte: Autoria própria.

3.2. Definição da variável independente "X" e a variável dependente "Y"

Figura 5 - Definição da variável independente "X" e a variável dependente "Y".

```
X = df_tratado["Funcionários"].values.reshape(-1, 1)
y = df_tratado["Faturamento"]
```

Fonte: Autoria própria.

3.3. Divisão de treino e teste da base de dados

Figura 6 - Divisão de treino e teste da base de dados.

```
X_treino, X_teste, y_treino, y_teste = train_test_split(X, y, test_size=0.3, random_state=40, shuffle=True)
```

Fonte: Autoria própria.

3.4. Instanciação do modelo de Regressão Linear

Figura 7 - Instanciação do modelo de Regressão Linear.

```
regressao = LinearRegression()
```

Fonte: Autoria própria.

3.5. Treinamento do modelo

Figura 8 - Treinamento do modelo de *machine learning*.

```
regressao.fit(X_treino, y_treino)
```

Fonte: Autoria própria.

3.6. Realização de previsões com os dados inseridos

Figura 9 - Previsões com os dados inseridos.

```
# Inserção manual do número de funcionários
numero_de_funcionarios = 8

# Realização do reshape para garantir que seja uma matriz bidimensional (ao invés de uma Series)
entrada_predicao = np.array(numero_de_funcionarios).reshape(1, -1)

# Realização da predição
predicao_faturamento = regressao.predict(entrada_predicao)

if numero_de_funcionarios > 0:
    print("Previsão de Faturamento para {} funcionários: R$ {:.2f}".format(numero_de_funcionarios, predicao_faturamento[0]))
else:
    print("Previsão de Faturamento para {} funcionários: R$ {:.2f}".format(numero_de_funcionarios, 0))
```

Fonte: Autoria própria.

4. Avaliação dos resultados

4.1. Avaliação de métricas [MAE, R² e Coeficiente de Correlação (Pearson)]

Figura 10 - Avaliação dos resultados.

```
y_pred_teste = regressao.predict(X_teste)

# MAE
mae = mean_absolute_error(y_teste, y_pred_teste)
desvio_percentual = (mae / y_teste.mean()) * 100

# R²
r2_teste = r2_score(y_teste, y_pred_teste)

# Coeficiente de Correlação (Pearson) "p"
correlacao_pearson = np.corrcoef(y_teste, y_pred_teste)[0, 1]

print("\nErro Absoluto Médio (Mean Absolute Error)\n\033[1mMAE\033[0m = R$ {:.2f} [\033[1m{:.2f} %\033[0m]\n".format(mae, desvio_percentual))
print("\nCoeficiente de Determinação\n\033[1mR²\033[0m = {:.2f}\n".format(r2_teste))
print("\nCoeficiente de Correlação (Pearson)\n\033[1;3mp\033[0m = {:.2f}\n".format(correlacao_pearson))
```

Fonte: Autoria própria.

4 RESULTADOS

Com os resultados obtidos, destaca-se a capacidade do modelo em prever os faturamentos com base no número de funcionários em que o usuário adiciona ao modelo preditor. A forte correlação entre as variáveis (cerca de 94%), ou seja, ambas as variáveis possuíam uma correlação direta-positiva. A avaliação do modelo, como MAE sendo cerca de 8,7% de desvio (uma porcentagem dentro do esperado pela empresa onde aceitavam até 10% de desvio).

É possível notar também que há possíveis melhorias futuras, como: Levar em consideração o limite do número de funcionários por unidade da loja, considerar vendedores mais experientes e com alto desempenho e também, considerar datas comemorativas como o natal.

5 CONCLUSÃO

No desenvolvimento da presente pesquisa foram realizadas revisões acerca de assuntos como: inteligência artificial, *machine learning* e suas especificidades. Houve uma certa dificuldade em algumas partes da trajetória para a realização da pesquisa, como a não ter uma base dados maior disponível. Essa dificuldade foi entendida e contornada por seguir com a pesquisa utilizando apenas os dados históricos disponíveis (a partir de meados de 2021).

Quanto a hipótese defendida, a análise desses resultados demonstrou, que as previsões de faturamento foram entregues de forma satisfatória ($MAE < 10\%$) para a empresa alvo do estudo. Ao contar com essa ferramenta, a empresa tem previsões mais direcionadas, onde os proprietários e gerentes podem tomar decisões fundamentadas e estratégicas, evitando excessos ou falta de estoque, o que resulta na redução de custos e na maximização dos lucros.

No entanto, é importante considerar que há outros fatores a serem considerados, como a venda em datas comemorativas e a quantidade padrão de funcionários em cada unidade específica. Com isso, como sugestão para trabalhos futuros é possível apontar a pesquisa aplicando não apenas o número de funcionários x faturamento, mas considerar outros níveis de dados (ex.: Regressão linear múltipla ou Regressão logística).

REFERÊNCIAS

OLIVEIRA, A. **Aumentando sua receita com Previsões de Vendas precisas**. 2023. Disponível em: <https://blog.nectarcrm.com.br/previsoes-de-vendas/>. Acesso em: 16 Mai. 2023.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. ed. 04, Upper Saddle River. Pearson Education. 2020.

XP EDUCAÇÃO. **Quais são os tipos de Machine Learning?**. 2022. Disponível em: <https://blog.xpeducacao.com.br/tipos-de-machine-learning/>. Acesso em: 15 Mai. 2023.

YES COSMETICS. **Sobre a Yes!** [s.d]. Disponível em: <https://www.yescosmetics.com.br/institucional>. Acesso em: 19 Mai. 2023.

ZIEGER, K. *et al. Machine Learning (ML), métricas de desempenho funcional. In: Certified Tester Syllabus AI Testing*. [2021]. ed. 1.0. p. 44-49. [s.l]: [s.n]. Acesso em: 15 Mai. 2023.