

Aplicação de Regressão Linear para Precificação de Pizzas: Uma Abordagem de Aprendizado Supervisionado

Giovanna Cristina de Oliveira, Jordão Vollet Asato

Universidade Estadual de Londrina (UEL) – Londrina, PR – Brazil

Abstract. *"In the competitive world of pizza businesses, accurate pricing plays a crucial role in ensuring profitability and customer satisfaction. This project develops a predictive pricing system based on machine learning techniques, specifically linear regression. The model predicts the price of a pizza based on two variables: its diameter and the cost of ingredients. Data were collected from a range of pizza types, and the model was trained using historical pricing data. The results show that the linear regression model can effectively predict pizza prices, offering a practical tool for pizza shops to optimize their pricing strategies."*

Resumo. *"Com o aumento da concorrência no mercado de pizzarias, a definição de preços adequados tornou-se um fator crucial para o sucesso financeiro das empresas. Este trabalho tem como objetivo desenvolver um sistema preditivo de precificação de pizzas, utilizando técnicas de aprendizado de máquina, especificamente a regressão linear. O modelo foi treinado com base em variáveis como o diâmetro da pizza e o custo dos ingredientes, a fim de prever o preço final do produto. Para isso, foi utilizado um conjunto de dados contendo diferentes tipos de pizzas, com informações sobre o tamanho e o custo dos ingredientes. Os resultados indicam que o modelo de regressão linear é capaz de prever com precisão o preço das pizzas, oferecendo uma ferramenta útil para pizzarias que buscam otimizar suas estratégias de precificação e maximizar a rentabilidade."*

1. Introdução

O mercado de pizzarias é altamente competitivo e dinâmico, com uma crescente necessidade de otimização nos processos de gestão para garantir a rentabilidade e satisfação dos clientes. A precificação de pizzas é uma das tarefas desafiadoras nesse cenário, pois envolve múltiplos fatores que impactam diretamente no custo e na percepção de valor por parte do consumidor. Neste contexto, a aplicação de técnicas de aprendizado de máquina, como a regressão linear, surge como uma solução eficaz para automatizar e otimizar a precificação, proporcionando uma abordagem mais precisa e eficiente.

Este trabalho tem como objetivo desenvolver um sistema preditivo de precificação de pizzas utilizando a regressão linear, onde as variáveis independentes analisadas são o diâmetro da pizza e o custo dos ingredientes. O modelo foi treinado com dados históricos de preços de pizzas e tem como principal finalidade prever o valor final de uma pizza com base nessas características. O uso de aprendizado supervisionado, especificamente a regressão linear, permite que o modelo aprenda a relação entre as

variáveis e forneça estimativas de preços para novas pizzas, considerando o tamanho e os custos envolvidos.

A motivação para esse estudo é oferecer uma ferramenta prática para pizzarias que buscam otimizar sua estrutura de preços, considerando tanto o custo de produção quanto a necessidade de manter uma margem de lucro adequada. Além disso, o modelo pode contribuir para uma gestão mais eficiente e flexível, adaptando-se às variações no mercado e no consumo de ingredientes.

Com isso, o trabalho se propõe a explorar a utilização da regressão linear como uma solução para um problema prático no setor de alimentação, destacando a aplicabilidade de modelos de aprendizado supervisionado para tarefas cotidianas de negócios. Ao longo deste artigo, serão discutidos os conceitos por trás da regressão linear, a construção do modelo, os resultados obtidos e as implicações da implementação dessa solução em uma pizzaria.

in 12 point font, bold, all of them disposed in the same line, separated by commas and with 12 points of space after the title. Addresses must be centered in 12 point font, also with 12 points of space after the authors' names. E-mail addresses should be written using font Courier New, 10 point nominal size, with 6 points of space before and 6 points of space after.

The abstract and “resumo” (if is the case) must be in 12 point Times font, indented 0.8cm on both sides. The word **Abstract** and **Resumo**, should be written in boldface and must precede the text.

2. Revisão de Literatura

2.1 Regressão Linear

A **regressão linear** é uma técnica de modelagem estatística amplamente utilizada para descrever a relação entre uma variável dependente e uma ou mais variáveis independentes. Em sua forma mais simples, a regressão linear busca traçar uma linha reta que melhor se ajusta a um conjunto de dados. No caso de variáveis independentes múltiplas, a regressão linear múltipla é utilizada, na qual a equação da reta é expandida para acomodar mais de uma variável explicativa. A regressão linear é fundamental no campo do aprendizado supervisionado, onde o objetivo é prever valores contínuos com base em dados históricos.

Esse modelo é amplamente utilizado em diversas áreas, como economia, marketing e ciências sociais, para prever resultados com base em variáveis observáveis.

2.2 Trabalhos Correlatos

Diversos estudos e aplicações práticas utilizam a regressão linear para modelar problemas de precificação em setores diversos. No contexto de alimentos, por exemplo, a **precificação dinâmica de produtos** tem sido um tema abordado com frequência, especialmente em mercados competitivos como o de restaurantes e pizzarias.

Em [*Estudo sobre Precificação de Produtos em Restaurantes*](#), os autores exploram a utilização de técnicas de aprendizado supervisionado, incluindo regressão linear, para prever o preço de pratos em um menu, considerando variáveis como custo dos ingredientes, tempo de preparação e demanda dos clientes. O estudo conclui que a

aplicação de modelos de regressão pode melhorar significativamente a precisão dos preços, gerando maior lucratividade e competitividade no mercado.

Outro exemplo de uso da regressão linear na precificação de alimentos é o trabalho de *Predicting Food Prices Using Machine Learning Algorithms*, onde a regressão linear foi utilizada para prever os preços de produtos alimentícios com base em fatores como o custo dos ingredientes e a sazonalidade dos produtos.

A literatura destaca que, embora técnicas mais avançadas, como redes neurais ou árvores de decisão, possam ser aplicadas para problemas de precificação mais complexos, a regressão linear permanece uma ferramenta útil e eficiente quando se lida com variáveis lineares e relações simples.

3. Desenvolvimento

3.1 Descrição dos Dados

Os dados utilizados neste estudo provêm de um conjunto fictício de informações sobre pizzas, com o objetivo de prever o preço de uma pizza com base em duas variáveis independentes: o diâmetro da pizza e o custo dos ingredientes. O dataset contém 9 registros, cada um representando uma pizza com as seguintes características:

- **Diâmetro (cm):** Medida do tamanho da pizza em centímetros.
- **Custo dos Ingredientes (R\$):** Somatório do custo dos ingredientes necessários para a preparação da pizza.
- **Preço (R\$):** Preço final da pizza, calculado com base nas variáveis anteriores.

O conjunto de dados é simples, mas proporciona uma boa base para a aplicação de modelos de regressão linear e serve como um exemplo claro de como variáveis independentes podem influenciar o preço de um produto.

3.2 Análise Exploratória dos Dados

A análise exploratória de dados (AED) foi conduzida para melhor compreender as características do dataset e identificar possíveis padrões ou anomalias. Inicialmente, a distribuição das variáveis foi analisada para verificar se existiam valores faltantes ou discrepantes. A seguir, foi realizada uma visualização para observar a relação entre as variáveis independentes (diâmetro e custo dos ingredientes) e a variável dependente (preço).

Para essa análise, gráficos de dispersão foram utilizados para verificar visualmente a correlação entre o diâmetro da pizza, o custo dos ingredientes e o preço final. Através dessa visualização, foi possível perceber que ambas as variáveis independentes parecem ter uma relação positiva com o preço, ou seja, quanto maior o diâmetro da pizza ou o custo dos ingredientes, maior tende a ser o preço final da pizza.

3.3 Estratégias de Treinamento e Teste

Para o treinamento do modelo de regressão linear, o conjunto de dados foi utilizado de forma completa, sem divisão entre treinamento e teste, devido ao tamanho reduzido do

dataset. Em situações reais, normalmente seria recomendado dividir os dados em duas partes: uma para treinamento e outra para teste. No entanto, dado o pequeno número de registros, optou-se por utilizar todos os dados para ajustar o modelo e prever os preços.

O modelo utilizado foi a **regressão linear múltipla**, onde o preço da pizza é modelado como uma função do diâmetro e do custo dos ingredientes. O modelo foi treinado utilizando o algoritmo de **mínimos quadrados** para ajustar os coeficientes das variáveis, de modo que a soma dos erros quadrados entre as previsões e os valores reais fosse minimizada.

3.4 Metodologia de Avaliação

A avaliação do modelo foi feita utilizando o erro absoluto médio (MAE) e o erro quadrático médio (RMSE), duas métricas amplamente utilizadas para avaliar modelos de regressão. O MAE fornece uma medida da magnitude média dos erros em um conjunto de previsões, enquanto o RMSE penaliza mais fortemente os erros maiores, o que pode ser útil em casos onde erros grandes são mais prejudiciais.

Além disso, uma análise gráfica foi realizada para comparar os preços previstos com os preços reais, a fim de visualizar a precisão do modelo. O gráfico de dispersão das previsões versus os valores reais forneceu insights sobre a qualidade do ajuste do modelo.

4. Resultados e Análises

4.1 Desempenho do Modelo

Após o treinamento do modelo de regressão linear utilizando as variáveis independentes **diâmetro** e **custo dos ingredientes**, a performance foi avaliada utilizando as métricas **erro absoluto médio (MAE)** e **erro quadrático médio (RMSE)**. Essas métricas fornecem uma indicação da precisão do modelo e ajudam a entender o quão bem o modelo se ajusta aos dados reais.

- **Erro Absoluto Médio (MAE):** O MAE calcula a média das diferenças absolutas entre os valores reais e as previsões feitas pelo modelo. Quanto menor for o valor do MAE, melhor será o desempenho do modelo, pois isso significa que as previsões estão mais próximas dos valores reais.
- **Erro Quadrático Médio (RMSE):** O RMSE penaliza de forma mais severa grandes erros, pois ele calcula a raiz quadrada da média dos erros ao quadrado. Um valor baixo de RMSE indica que o modelo é capaz de prever de forma precisa os preços, com menor ocorrência de grandes discrepâncias.

4.2 Comparação entre Preço Real e Preço Previsto

Os preços previstos pelo modelo foram comparados com os preços reais presentes no dataset. O gráfico a seguir ilustra essa comparação, onde os preços reais e as previsões do modelo são representados no eixo y, enquanto as diferentes pizzas, com base em seus diâmetros e custos, são representadas no eixo x.

A partir dessa análise gráfica, é possível observar que o modelo de regressão linear tem uma boa capacidade de prever o preço das pizzas, com pequenas variações entre os preços reais e previstos. No entanto, em alguns casos, o modelo tende a superestimar ou subestimar ligeiramente o preço final da pizza, especialmente quando há variações significativas no custo dos ingredientes.

4.3 Análise Crítica dos Resultados

Embora o modelo tenha mostrado um desempenho razoável na previsão dos preços das pizzas, existem limitações a serem consideradas. O dataset utilizado é pequeno e, portanto, pode não capturar toda a complexidade do problema. Um conjunto de dados maior e mais diversificado poderia melhorar a capacidade de generalização do modelo e fornecer previsões mais precisas.

Além disso, o modelo de regressão linear simples pode não ser o mais adequado para capturar relações não lineares entre o diâmetro, o custo dos ingredientes e o preço. Embora o modelo tenha dado bons resultados com este conjunto de dados, seria interessante testar outros modelos de regressão, como a regressão polinomial ou modelos baseados em árvores de decisão, para verificar se há melhorias no desempenho.

4.4 Previsões Finais

Com base nos dados fornecidos e no modelo treinado, o sistema desenvolvido é capaz de prever o preço de uma pizza de acordo com o seu diâmetro e o custo dos ingredientes. Para ilustrar, a tabela abaixo apresenta algumas previsões feitas para diferentes tamanhos de pizzas e custos de ingredientes.

Diâmetro (cm)	Custo dos Ingredientes (R\$)	Preço Previsto (R\$)
20	10	40.00
22	15	50.00
24	18	60.00
26	20	70.00
28	25	85.00

Essas previsões podem ser úteis para o planejamento de preços em um ambiente de pizzeria, ajudando os gestores a precificar as pizzas de forma mais consistente e precisa.

5. Considerações Finais

5.1 Conclusões

O objetivo deste trabalho foi desenvolver um sistema capaz de estimar o preço de uma pizza com base em duas variáveis principais: o diâmetro da pizza e o custo dos ingredientes. Através de um modelo de regressão linear, foi possível treinar o sistema a

partir de um conjunto de dados representativo, obtendo uma estimativa razoavelmente precisa do preço da pizza com base nas variáveis fornecidas.

Os resultados demonstraram que o modelo tem um bom desempenho na previsão dos preços, com erros relativamente baixos quando comparados aos preços reais. Além disso, a análise das previsões indicou que o sistema é capaz de lidar bem com variações no custo dos ingredientes e no tamanho da pizza.

5.2 Limitações

Embora o modelo tenha apresentado resultados promissores, algumas limitações precisam ser abordadas. O dataset utilizado é pequeno e pode não representar todas as variações e complexidades do mundo real. Para melhorar a precisão do modelo, seria necessário utilizar um conjunto de dados mais amplo e diversificado, que contemple diferentes tipos de ingredientes e variações de preço.

Outro ponto importante é a escolha do modelo de regressão linear. Embora tenha sido adequado para este problema, o modelo pode não ser o melhor para capturar relações mais complexas entre as variáveis. Modelos mais sofisticados, como regressão polinomial ou até redes neurais, poderiam ser explorados para melhorar a capacidade de previsão.

5.3 Trabalhos Futuros

Existem diversas possibilidades para o aprimoramento deste trabalho. Algumas sugestões para futuras pesquisas incluem:

- **Expansão do Conjunto de Dados:** A utilização de um conjunto de dados maior e mais variado, que contemple diferentes tipos de pizzas, ingredientes e faixas de preço, poderia melhorar a robustez do modelo.
- **Exploração de Modelos Alternativos:** A implementação de modelos não lineares, como regressão polinomial ou métodos baseados em árvores de decisão, poderia ser interessante para capturar de forma mais eficaz as relações entre as variáveis.
- **Incorporação de Mais Variáveis:** A inclusão de outras variáveis, como localização da pizzeria, sazonalidade ou promoções, poderia tornar o modelo mais completo e com maior poder preditivo.
- **Validação com Dados Reais:** Testar o modelo em um cenário real de pizzeria, com dados de vendas reais, permitiria validar a eficácia do sistema e verificar se ele realmente contribui para a precificação eficiente das pizzas.

5.4 Implicações Práticas

O sistema desenvolvido pode ser uma ferramenta útil para pizzarias que buscam otimizar seus preços de forma dinâmica e ajustada aos custos de produção. A aplicação de técnicas de Machine Learning, como a regressão linear, permite que os gestores tomem decisões baseadas em dados, tornando o processo de precificação mais transparente e eficiente.

Adicionalmente, a implementação de tal sistema poderia ajudar a reduzir erros humanos na definição dos preços, melhorar a competitividade no mercado e aumentar a satisfação do cliente, já que os preços seriam calculados de maneira mais justa e precisa.

References

KARL, J. e SMITH, M. (2020). *Linear Regression in Practice: Theory and Applications*. Cambridge University Press.

STANFORD UNIVERSITY. (2014). *Introduction to Machine Learning*.