

Relatório de Análise de Componentes

Gabriel Medina, Jonatas Fernandes

Junho 2023

1 Introdução

Neste relatório, será descrita uma análise de componentes principais (PCA) realizada em um conjunto de dados relacionados aos valores nutricionais de produtos do Starbucks. O PCA é uma técnica estatística que permite identificar padrões e estruturas nos dados, reduzindo sua dimensionalidade e facilitando a interpretação.

2 Base de Dados

O conjunto de dados utilizado neste estudo contém informações sobre alimentos do Starbucks e foi obtido através do site Kaggle.com. O arquivo CSV contém informações sobre calorias, gorduras e fibras dos alimentos.

3 Codificação

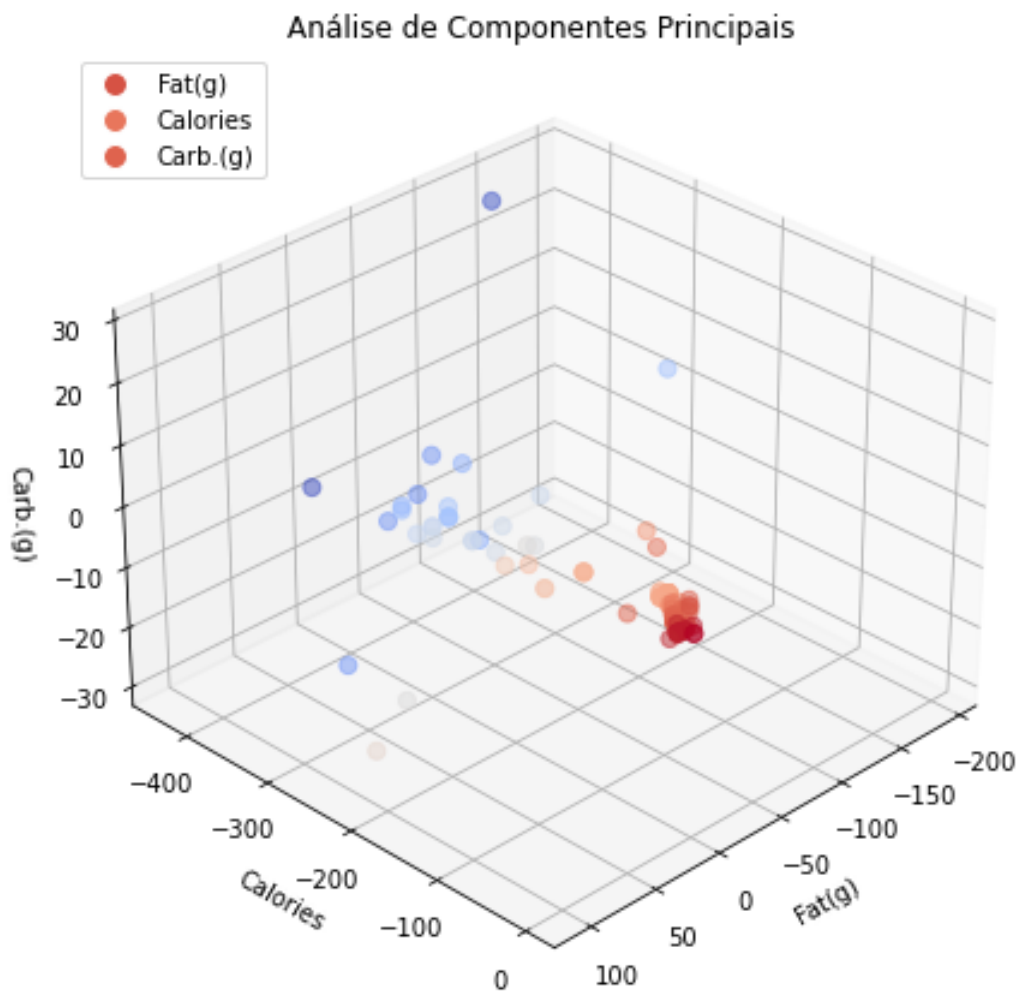
A codificação do algoritmo de Análise de Componentes Principais (PCA) foi implementada utilizando a linguagem de programação Python e as seguintes bibliotecas: NumPy para manipulação de arrays e realização de cálculos matemáticos, pandas para leitura e manipulação dos dados, e Matplotlib para a visualização gráfica.

Primeiramente, os dados são lidos de um arquivo CSV utilizando a biblioteca Pandas. Em seguida, são extraídas as colunas relevantes para a análise (Calories, Fat(g), Carb.(g), Fiber(g), Protein, Sodium). O algoritmo de PCA é aplicado utilizando a função "perform_pca", que utiliza as bibliotecas NumPy para manipulação de arrays e cálculos matemáticos. A função retorna os dados projetados nos componentes principais.

A visualização dos dados é realizada utilizando a biblioteca Matplotlib. Os pontos projetados nos componentes principais são plotados em um gráfico tridimensional, onde a cor de cada ponto representa a escala de valores da variável correspondente ao componente principal. O gráfico é configurado com rótulos nos eixos, título e uma legenda explicativa, por fim, o gráfico é exibido utilizando a função `plt.show()`, permitindo a visualização dos resultados da análise de componentes principais no conjunto de dados do Starbucks.

4 Resultado

A figura abaixo, gerada pelo código, apresenta uma representação tridimensional dos dados projetados nos três primeiros componentes principais.



Cada ponto no gráfico representa um produto do Starbucks, e a cor do ponto indica a escala de valores da variável correspondente ao componente principal, azul indica valores baixos e vermelho indica valores altos.