

Projeto de Inteligência Artificial – Variáveis Socioeconômicas que Influenciam a Felicidade da População

Luiz Henrique Marcucci – 10395609

João Victor Moraes de Souza – 10375333

1. Resumo

O Projeto consiste em utilizar métodos de Machine Learning para analisar e prever a felicidade da população de diferentes países com base em variáveis socioeconômicas. Em vez de tratar a felicidade como um valor contínuo, classificamos os países em três categorias: "Muito Baixa, ""Baixa", "Média", "Alta" "Muito Alta" felicidade. Assim, verificamos a correlação das variáveis para identificar quais variáveis têm maior impacto na felicidade e prever a categoria de um país com base em suas características.

2. Introdução

a. Contextualização

A felicidade da população é um tema de grande relevância para governos, economistas, sociólogos e instituições internacionais, pois está diretamente ligada ao desenvolvimento de políticas públicas e produtos solucionadores. Diversos estudos indicam que fatores como economia, saúde, educação, segurança e liberdade individual desempenham papéis fundamentais na satisfação e qualidade de vida de uma população.

O uso de algoritmos de classificação permite classificar e prever a felicidade da população de um país com base em suas condições socioeconômicas e entender quais fatores mais influenciam esse indicador.

b. Justificativa

Como já possuímos o *Score de Felicidade* dos países, transformamos essa métrica em categorias e utilizamos algoritmos de classificação para modelar a faixa de felicidade de uma nação. Isso pode auxiliar na identificação dos principais determinantes da felicidade e na formulação de políticas públicas mais eficazes com foco em “aumentar” a felicidade da população.

c. Objetivo

Temos como principais objetivos entender quais são as principais variáveis socioeconômicas que influenciam na felicidade da população de cada país. Posteriormente, queremos evoluir para a construção de um modelo preditivo para ler as variáveis mais influentes e classificar o provável score de felicidade dessa população.

d. Opção do Projeto

Utilizamos o framework scikit-learn para análise exploratória, pré-processamento dos dados, classificação dos dados e modelagem preditiva. Após a Classificação dos primeiros anos, utilizaremos algoritmos preditivos para prever a categoria de felicidade dos países nos anos seguintes

3. Descrição do Problema

Os indicadores de felicidades de população podem ser importantes para a realização de políticas públicas, como por exemplo, prevenção ao suicídio. A felicidade da população é influenciada por diversos fatores socioeconômicos: renda, expectativa de vida, nível educacional, percepção de corrupção e suporte social. Com base no Score de Felicidade já disponível no dataset, o problema que buscamos resolver é a classificação dos países em diferentes categorias de felicidade com base nesses fatores, verificando quais são os maiores influenciadores

para a felicidade da população e posteriormente classificando os anos seguintes de acordo com as variáveis forte.

4. Aspectos Éticos e Responsabilidade no Uso da IA

Pesquisa sobre felicidade da população pode envolver algumas preocupações relacionadas aos princípios éticos no uso do machine learning. Os insights poderiam enviesar novas políticas, causando mal interpretação dos dados e aplicando ações que não visam o bem-estar social. É importante que tenha transparência na aplicação dos modelos e tratamento dos dados.

5. Dataset e Análise Exploratória

O dataset foi retirado do Kaggle.

O dataset utilizado contém informações sobre diversos países ao longo dos anos, incluindo as seguintes variáveis:

- **Country:** Nome do país.
- **Year:** Ano do registro.
- **Happiness_Score:** Índice de felicidade atribuído ao país.
- **GDP_per_Capita:** PIB per capita.
- **Social_Support:** Índice de suporte social.
- **Healthy_Life_Expectancy:** Expectativa de vida saudável.
- **Freedom:** Percepção de liberdade.
- **Generosity:** Nível de generosidade.
- **Corruption_Perception:** Percepção de corrupção.
- **Unemployment_Rate:** Taxa de desemprego.
- **Education_Index:** Índice de qualidade da educação.
- **Urbanization_Rate:** Percentual da população vivendo em áreas urbanas.
- **Work_Life_Balance:** Índice de equilíbrio entre vida pessoal e profissional.
- **Crime_Rate:** Nível de criminalidade.
- **Political_Stability:** Estabilidade política.
- **Employment_Rate:** Taxa de emprego.

A Análise Exploratória realiza o tratamento dos dados, classificação, histograma das variáveis, verificação da correlação das variáveis.

Link Github: