

Projeto Extensionista - Tópicos de Big Data em Python

1. DIAGNÓSTICO E TEORIZAÇÃO

a. Identificação das partes interessadas e parceiros (Opcional)

Este projeto tem como partes interessadas as autoridades de saúde pública, especialmente no estado do Rio de Janeiro, onde os dados de dengue são monitorados. Os parceiros podem incluir órgãos governamentais e hospitais que lidam com o controle e prevenção da doença, como a Secretaria de Saúde do Estado e os Centros de Controle de Doenças (CDC).

b. Problemática e/ou problemas identificados (Obrigatório)

A dengue é uma doença endêmica que afeta milhões de pessoas em várias regiões do Brasil. O aumento no número de casos tem sido uma constante, e prever as possíveis epidemias pode ajudar no planejamento das ações de saúde pública. Este projeto visa usar Big Data para prever a evolução dos casos de dengue no estado do Rio de Janeiro, com base em dados históricos.

c. Justificativa (Obrigatório)

Este projeto visa aplicar técnicas de análise de dados e regressão linear para prever a quantidade de casos de dengue nos anos seguintes. Ao utilizar dados históricos de 2014 a 2024, é possível identificar padrões e tendências que podem ser usados para otimizar o controle da doença. A análise de Big Data permite uma resposta mais ágil e informada para reduzir a propagação da doença.

d. Objetivos/resultados/efeitos a serem alcançados (Obrigatório)

O objetivo deste projeto é:

1. Desenvolver uma análise preditiva sobre os casos de dengue utilizando Big Data e ferramentas como Python e bibliotecas de Machine Learning.
2. Aplicar modelos de regressão linear para prever os casos de dengue com base nos dados históricos disponíveis.
3. Visualizar os resultados em gráficos claros e informativos para facilitar a interpretação dos dados.
4. Fornecer insights para as autoridades de saúde pública, com previsões que possam ser usadas para otimizar o controle da doença.

e. Referencial teórico (subsídio teórico para propositura de ações da extensão) (Obrigatório)

O uso de Big Data na saúde pública tem sido uma estratégia crescente para melhorar a previsão e controle de doenças. O modelo de regressão linear é amplamente utilizado para

prever variáveis numéricas com base em dados históricos, e sua aplicação no contexto de saúde tem mostrado ser eficaz na previsão de surtos de doenças endêmicas como a dengue.

2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO

a. Plano de trabalho (usando ferramenta acordada com o docente) (Obrigatório)

O plano de trabalho foi estruturado para realizar uma análise de Big Data utilizando Python. As etapas principais incluem:

1. Coleta e limpeza de dados: Os dados históricos sobre casos de dengue de 2014 a 2024 foram coletados e preparados para análise.
2. Análise exploratória dos dados: A primeira análise envolveu a visualização da distribuição dos casos ao longo dos anos, usando gráficos de barras e linha.
3. Modelo preditivo: Foi aplicado o modelo de regressão linear simples para prever o número de casos de dengue nos anos seguintes.
4. Visualização de resultados: Foram gerados gráficos para exibir a evolução dos casos de dengue, a média, a mediana e o desvio padrão.

b. Descrição da forma de envolvimento do público participante na formulação do projeto. (Opcional)

Este projeto foi desenvolvido de forma individual, com a aplicação de técnicas de análise de dados para resolver a problemática da saúde pública.

c. Grupo de trabalho (descrição da responsabilidade de cada membro) (Obrigatório)

Este projeto foi desenvolvido individualmente. O único membro responsável pelo trabalho foi o Celso Muniz, que desempenhou as funções de:

1. Cientista de Dados: Responsável pela coleta, limpeza e preparação dos dados. Responsável pela análise exploratória dos dados, aplicação do modelo preditivo (regressão linear) e visualização dos resultados.

d. Metas, critérios ou indicadores de avaliação do projeto (Opcional)

As metas incluem:

1. Obter previsões precisas de casos de dengue para os próximos anos.
2. Visualizar as tendências de forma clara e intuitiva.
3. Avaliar a acurácia do modelo de regressão linear comparando as previsões com os dados reais.

f. Detalhamento técnico do projeto (Obrigatório)

i. Explicação do tema escolhido. Deve justificar.

O tema escolhido, análise de dados de dengue, foi selecionado devido à sua alta relevância social, já que a dengue continua a ser um problema de saúde pública grave no Brasil.

Utilizando Big Data, podemos melhorar as previsões e intervenções.

ii. Explicação do código desenvolvido

O código desenvolvido segue as etapas: coleta de dados, limpeza, análise exploratória e construção do modelo de regressão linear para prever o número de casos de dengue com base em dados históricos. A visualização inclui gráficos de barras, linhas de tendência e medidas como média, mediana e desvio padrão.

iii. Explicação dos resultados alcançados

Os resultados mostram uma tendência crescente de casos de dengue, com uma média de casos que pode ser utilizada como referência para ações futuras. A regressão linear demonstrou um bom ajuste aos dados, indicando uma alta previsibilidade com base nas informações disponíveis.

3. ENCERRAMENTO DO PROJETO

a. Relatório Coletivo (Obrigatório)

Fui o único responsável por todas as etapas deste projeto, que visou prever os casos de dengue no Rio de Janeiro utilizando **Big Data**. O projeto começou com a escolha do tema, dada a importância social da doença e a relevância de se aplicar **modelos preditivos** na saúde pública.

Desafios e Soluções:

1. **Coleta e Limpeza de Dados:** A primeira etapa envolveu a limpeza e preparação dos dados, o que exigiu um trabalho cuidadoso para garantir que estivessem prontos para análise.
2. **Análise Exploratória:** Utilizei gráficos para entender a distribuição dos casos e identificar padrões ao longo dos anos.
3. **Aplicação de Modelos Preditivos:** A implementação da **regressão linear** foi um desafio técnico importante, mas que resultou em previsões úteis sobre a evolução dos casos de dengue.
4. **Visualização de Resultados:** Criei gráficos para apresentar a média, mediana e desvio padrão, facilitando a interpretação dos resultados.

Aprendizado:

O projeto me permitiu aplicar **Big Data** em um problema real, aprimorando minhas habilidades em análise de dados e mostrando como essas ferramentas podem impactar positivamente a saúde pública.

Esse resumo destaca os principais pontos do seu relato. Se precisar de mais ajustes, estou à disposição!

4. CONSIDERAÇÕES FINAIS (Obrigatório)

O projeto demonstrou como a aplicação de Big Data e ferramentas de Python podem ser efetivas para melhorar o controle de doenças endêmicas como a dengue. A utilização de técnicas preditivas, como a regressão linear, provou ser uma abordagem viável para ajudar autoridades de saúde pública a tomar decisões mais rápidas e informadas.