

1. Etapa A

Pergunta a responder: *Este indivíduo foi diagnosticado com dengue com base nos dados disponíveis?*

1.1. Cenário 1

1.1.1. Pré-Processamento Livre;

1.1.2. Algoritmos com hiperparâmetros definidos;

1.1.2.1. K-Nearest Neighbors (KNN)

Número de Vizinhos (K): Comece com K=5, podendo ser ajustado por meio de validação cruzada.

Métrica de Distância: Euclidiana.

Peso das amostras: Uniforme.

1.1.2.2. Árvore de Decisão

Critério de Divisão: Entropia

Profundidade Máxima: Sem limite inicial, ajustável com validação cruzada.

Número Mínimo de Amostras para Divisão de um Nó: 2.

Número Mínimo de Amostras em um Nó Folha: 1

1.1.2.3. Rede Neural (Multilayer Perceptron - MLP)

Número de Camadas Ocultas: 2 camadas.

Número de Neurônios em Cada Camada Oculta: 100 neurônios.

Função de Ativação: ReLU.

Solver (Algoritmo de Otimização): Adam.

Taxa de Aprendizado: 0.001.

Número de Épocas: 200.

Tamanho do Lote: 32.

1.1.2.4. Logistic Regression (Regressão Logística)

Solver: lbfgs

Penalidade: L2

Máximo de Iterações: 100

1.1.2.5. Random Forest (Floresta Aleatória)

Número de Árvores na Floresta: 100

Critério de Divisão: Entropia

Entregáveis (Relatório no github):

04/06 - Estatística descritiva com distribuição dos dados, padrões e outliers;
Divisão dos conjuntos de treino, teste e validação; Validação dos algoritmos com métricas de P, R e F1. Deploy em um container e conexão com dados;

11/06 - Dashboard de visualização dos dados; Avaliação dos hiperparâmetros;
Monitoramento da performance do modelo;

1.2. Cenário 2

1.2.1. Pré-Processamento Pré-definido

1.2.2. Algoritmos fixos com hiperparâmetros livres

Entregáveis (Relatório no github):

18/06 - Estatística descritiva com distribuição dos dados, padrões e outliers para o pré-processamento definido; Divisão dos conjuntos de treino, teste e validação; Validação dos algoritmos com métricas de P, R e F1. Deploy em um container e conexão com dados;

25/06 - Dashboard de visualização dos dados; Avaliação dos hiperparâmetros e proposta de novos hiperparâmetros mais adequados aos dados; Monitoramento de desempenho do modelo;

2. Etapa B

- 2.1. Pergunta a responder: *Em quais períodos e regiões ocorrem os picos de incidência de dengue, e quais fatores contribuem para esses picos com base nos dados epidemiológicos e clínicos disponíveis?*

Entregáveis (Relatório no github):

09/07 -Estatística descritiva com distribuição dos dados e Dashboard de visualização dos dados temporal; Divisão dos conjuntos de treino, teste e validação; Validação do algoritmo de séries temporais. Deploy em um container e conexão com dados;