



MLP

Introdução

Aumento de casos

Até 2030 a depressão será a doença mais comum do mundo, superando problemas cardíacos e o câncer.

Multi Fator

Relação entre diversos fatores, sejam eles genéticos, ambientais, sociais ou neurobiológicos

Diagnósticos

O diagnóstico é realizado a partir de questionários e análise de sintomas

Auxilio de ML

Discussão grande na comunidade científica sobre o auxilio da tecnologia

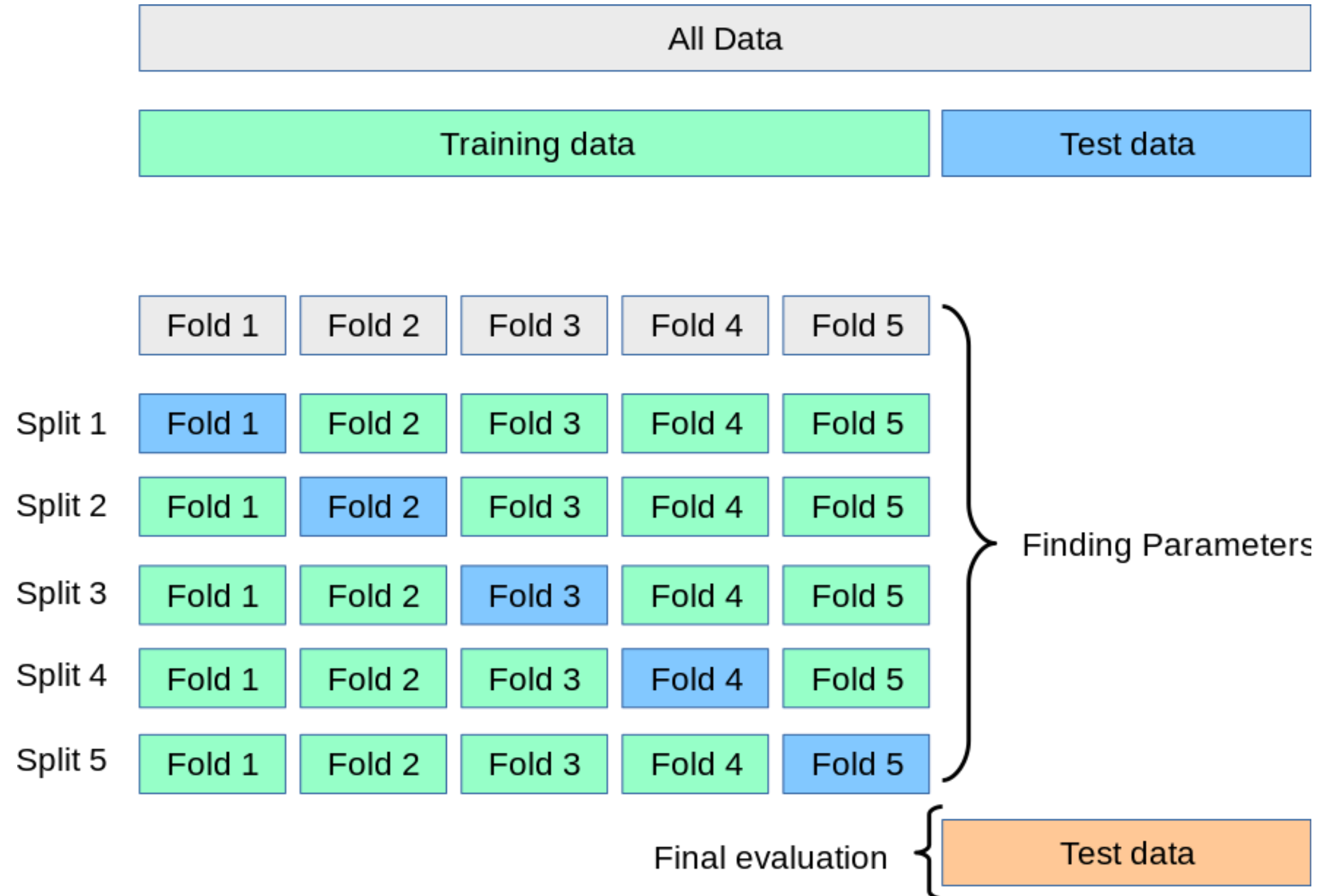
Descrição da Rede

- Rede neural do tipo MLP utilizando a biblioteca PyTorch.
- A arquitetura do modelo conta com uma camada de entrada compatível com o número de variáveis independentes (`input_dim`) e uma camada oculta com 64 neurônios e ativação ReLU,
- Seguida de uma camada de dropout com taxa de 50% para reduzir o overfitting.
- Por fim, uma camada de saída com dois neurônios foi utilizada, representando as duas classes da variável-alvo (`output_dim = 2`), com previsão tratada por meio de softmax.

Descrição da Rede

- Os modelos foram treinados durante 100, 20 e 50 épocas , respectivamente, para cada fold, utilizando o otimizador Adam com taxa de aprendizado de 0,001 e a função de perda CrossEntropyLoss, adequada para tarefas de classificação.
- A cada época, os pesos da rede foram ajustados com base no erro entre as previsões do modelo e os valores reais.
- O modelo da base 03 também emprega uma rede MLP, porém com foco em uma tarefa de classificação multiclasse, com sete classes possíveis. Sua arquitetura é semelhante, com uma camada oculta de 64 neurônios, função ReLU e Dropout de 30%.

Validação cruzada



Primeira base

Proveniente de uma pesquisa realizada em 2014, cujo objetivo principal foi investigar as atitudes em relação à saúde mental e a frequência de transtornos mentais no ambiente de trabalho, especialmente entre profissionais da área de tecnologia.

Pré Processamento

- Foram removidas colunas não categóricas e consideradas irrelevantes para o modelo, tais como 'Timestamp', 'comments' e 'state'
- Em seguida, foi feito o tratamento de outliers na coluna "Age"
- Para lidar com variáveis categóricas, aplicou-se a técnica de codificação utilizando o LabelEncoder
- Além disso, os dados foram normalizados por meio do StandardScaler
- O alvo da modelagem indica se o indivíduo procurou ou não tratamento para uma condição de saúde mental.

Resultados

```
[ ] from sklearn.metrics import classification_report, roc_auc_score

model.eval()
with torch.no_grad():
    y_pred = model(X_val)
    _, y_pred_class = torch.max(y_pred, 1)

    print(classification_report(y_val, y_pred_class))
    print(f"AUC-ROC: {roc_auc_score(y_val, y_pred[:, 1]):.4f}")
```



	precision	recall	f1-score	support
0	0.73	0.74	0.74	124
1	0.74	0.73	0.74	126
accuracy			0.74	250
macro avg	0.74	0.74	0.74	250
weighted avg	0.74	0.74	0.74	250

AUC-ROC: 0.8546

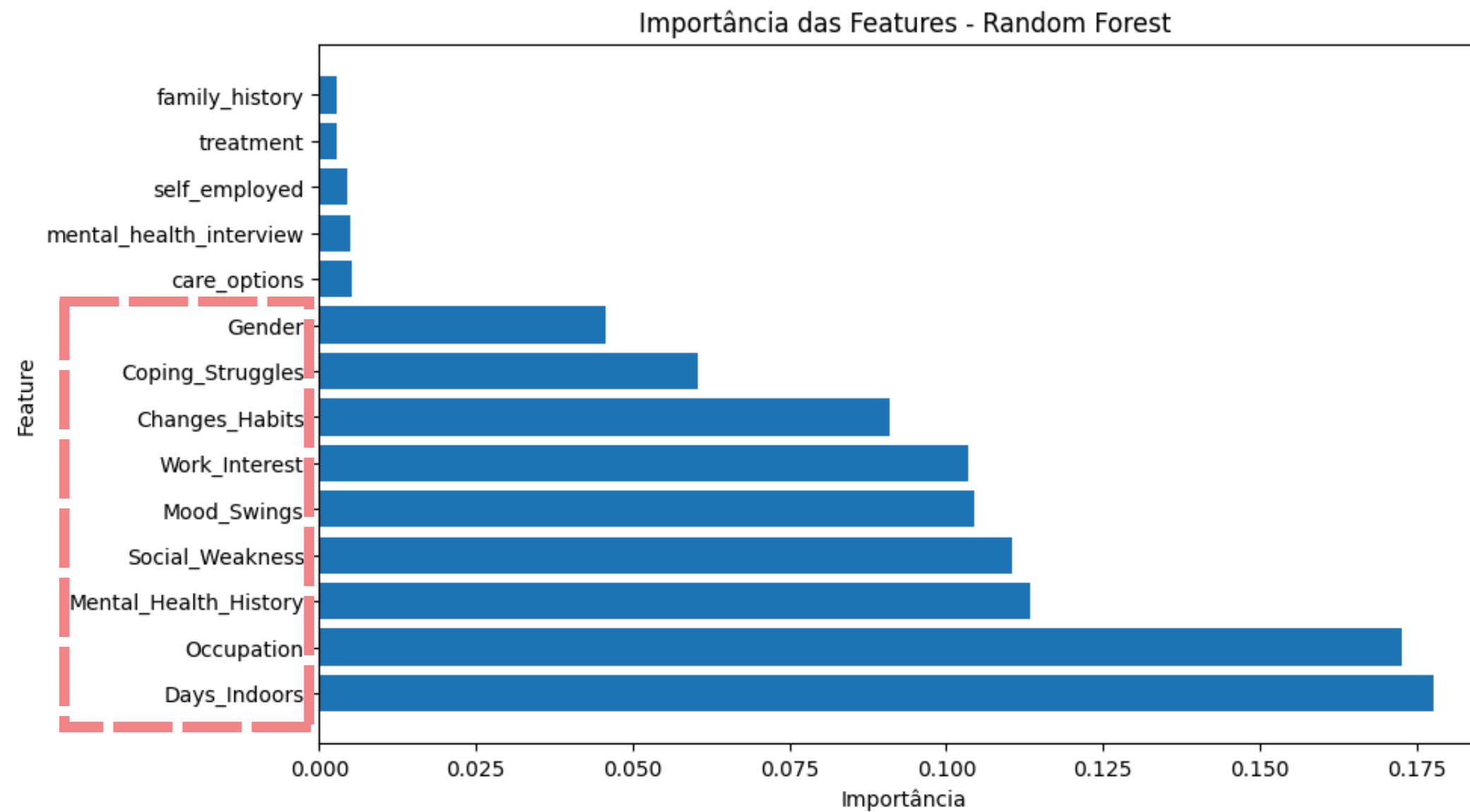
Segunda base

Este dataset é oriundo de uma pesquisa global realizada com o objetivo de acompanhar tendências relacionadas à saúde mental. Os dados abrangem uma variedade de variáveis, incluindo níveis de estresse, depressão, ansiedade, bem-estar subjetivo e uso de serviços de saúde mental.

Pré Processamento

- Constatou-se que apenas a coluna "self_employed" continha valores nulos, os quais foram preenchidos com o valor categórico 'None'.
- Em seguida, optou-se por transformar o problema em uma classificação binária, removendo todos os registros em que a variável "Growing_Stress" apresentava o valor "maybe".
- Foram excluídas colunas consideradas não informativas para a modelagem, como "timestamp" e "country".
- As variáveis categóricas foram então transformadas em variáveis numéricas utilizando o LabelEncoder.
- Registros duplicados foram removidos

Seleção de características



Resultados

```
▶ from sklearn.metrics import classification_report, roc_auc_score

model.eval()
with torch.no_grad():
    y_pred = model(X_val)
    _, y_pred_class = torch.max(y_pred, 1)

    print(classification_report(y_val, y_pred_class))
    print(f"AUC-ROC: {roc_auc_score(y_val, y_pred[:, 1]):.4f}")
```

```
⇒
```

	precision	recall	f1-score	support
0	0.82	0.72	0.77	3412
1	0.77	0.86	0.81	3774
accuracy			0.79	7186
macro avg	0.80	0.79	0.79	7186
weighted avg	0.79	0.79	0.79	7186

AUC-ROC: 0.8683

Terceira base

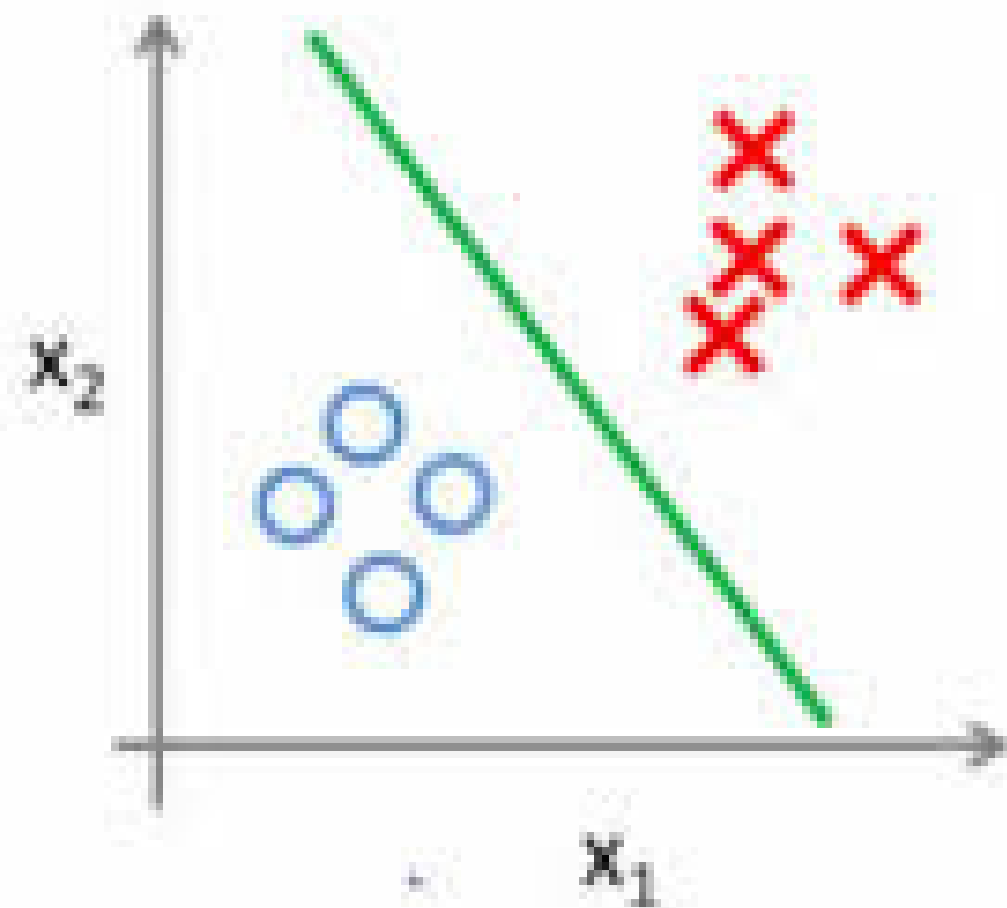
Este dataset de EEG contém aproximadamente 1.000 atributos extraídos de sinais eletroencefalográficos, com o objetivo de auxiliar na identificação de transtornos psiquiátricos.

Pré Processamento

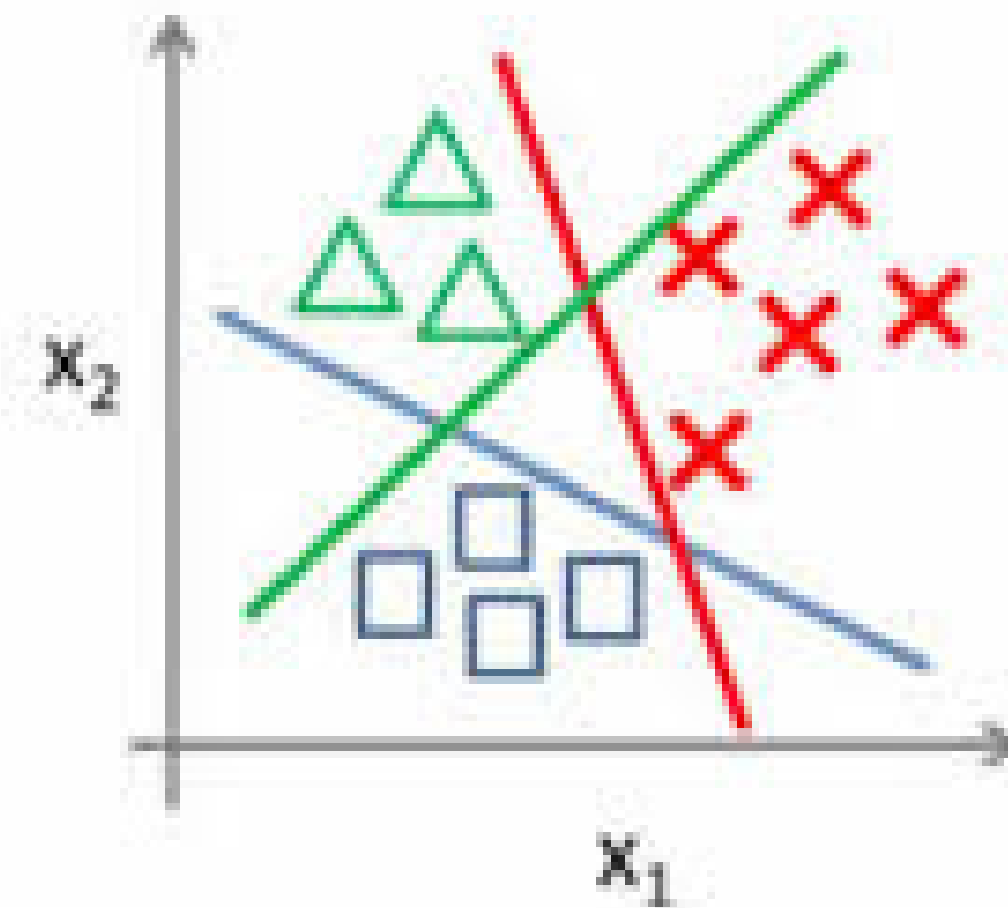
- remoção de colunas consideradas não interessantes para a tarefa de classificação,
- Foi realizado um filtro das colunas contendo os valores de EEG.
- Para garantir que todas as variáveis numéricas estivessem na mesma escala e contribuíssem de maneira uniforme no treinamento, os dados foram normalizados utilizando o StandardScaler.
- A variável alvo (main.disorder), que representa o transtorno associado ao indivíduo, foi transformada em valores numéricos por meio do Label Encoder.
- Como a distribuição das classes estava desequilibrada, o conjunto de dados foi submetido a uma técnica de balanceamento utilizando o SMOTE (Synthetic Minority Over-sampling Technique).

Multi-Classe

Binary classification:

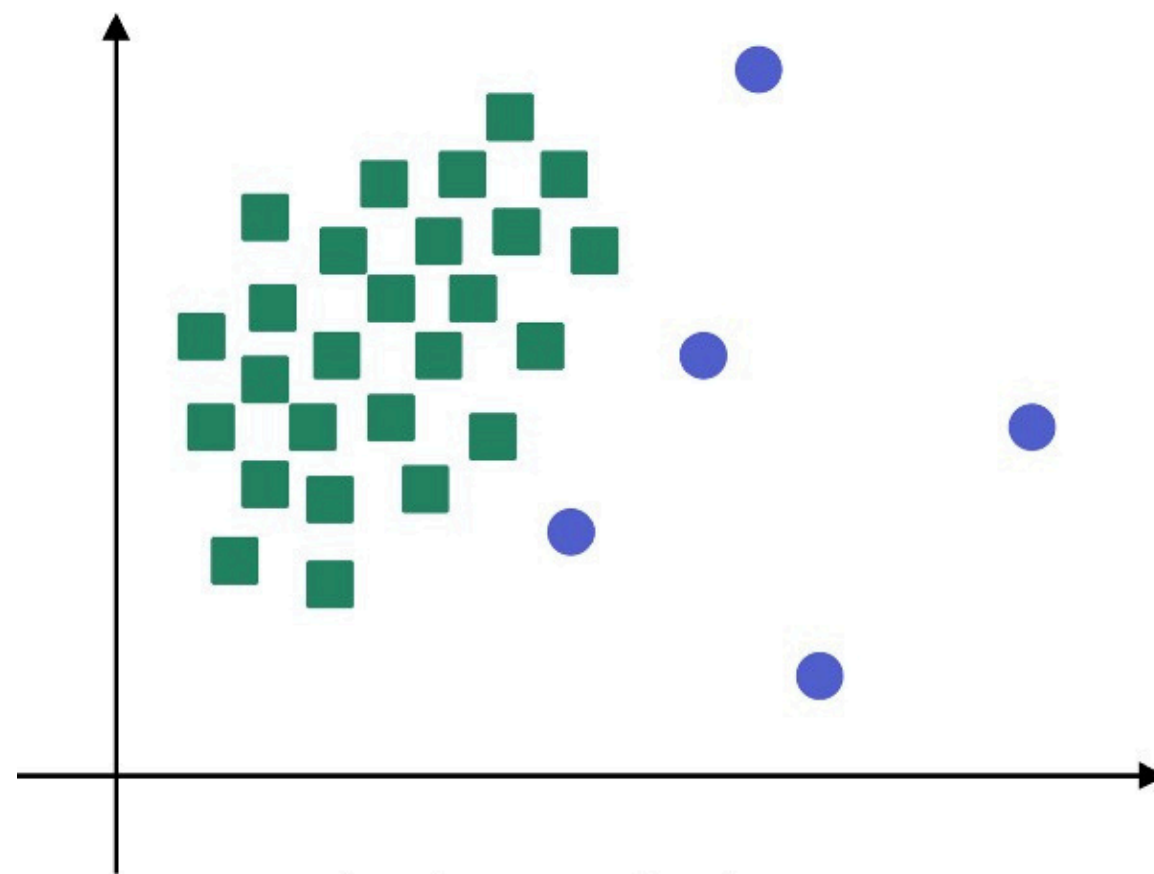


Multi-class classification:

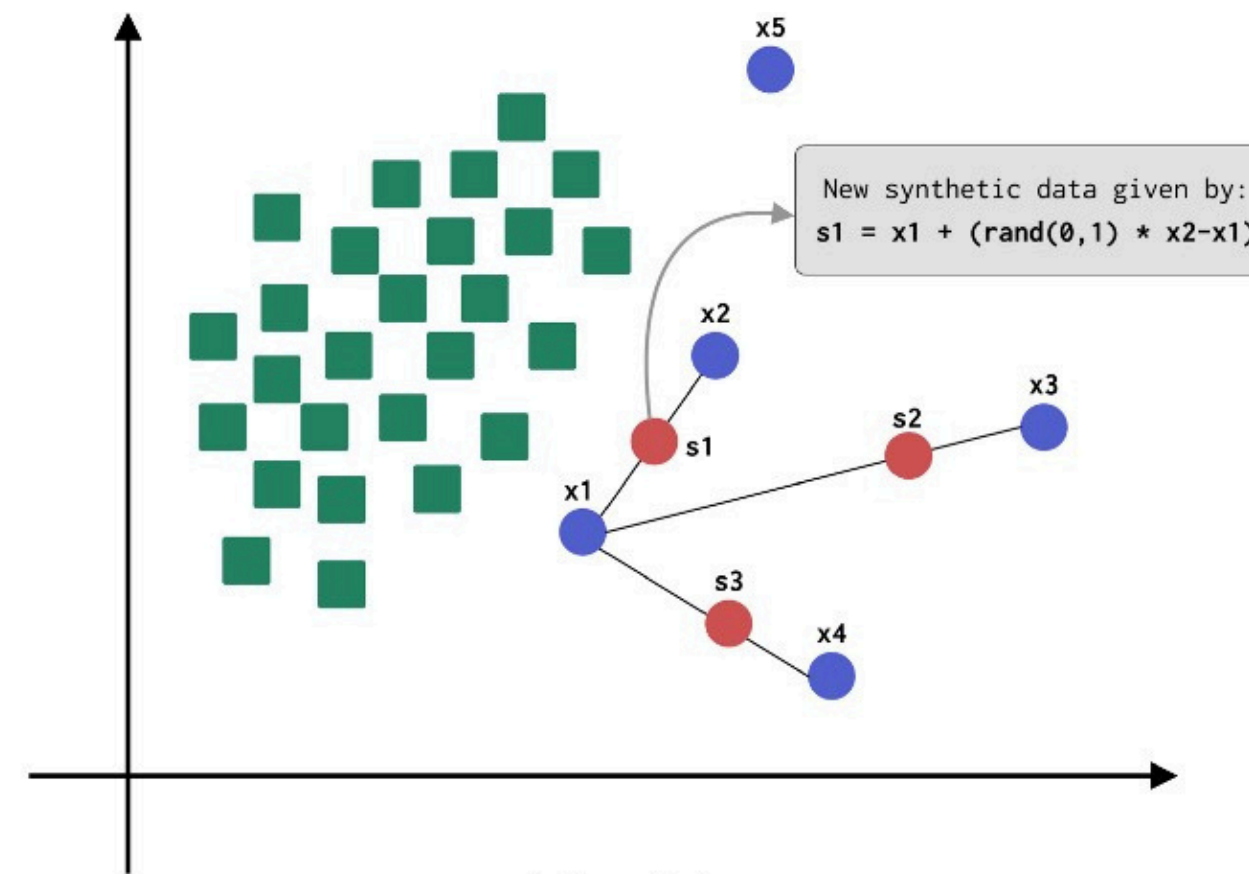


SMOTE

Essa abordagem gera exemplos sintéticos da classe minoritária, permitindo que o modelo aprenda de forma mais equilibrada e evite viés em relação à classe majoritária.



a) Class Imbalance



b) SMOTE

Resultados

```
[ ] from sklearn.metrics import classification_report

# 📄 Classification Report
report = classification_report(all_targets, all_preds, target_names=[f"Classe {i}" for i in range(num_classes)])
print("\n📄 Classification Report:")
print(report)
```



```
📄 Classification Report:
              precision    recall  f1-score   support

Classe 0      0.62      0.58      0.60       500
Classe 1      0.68      0.74      0.71       500
Classe 2      0.81      0.90      0.85       500
Classe 3      0.61      0.46      0.52       500
Classe 4      0.93      0.97      0.95       500
Classe 5      0.74      0.77      0.75       500
Classe 6      0.64      0.66      0.65       500

 accuracy      0.72      0.72      0.72    3500
 macro avg      0.72      0.72      0.72    3500
 weighted avg   0.72      0.72      0.72    3500
```

Classes

Mood disorder	266
Addictive disorder	186
Trauma and stress related disorder	128
Schizophrenia	117
Anxiety disorder	107
Healthy control	95
Obsessive compulsive disorder	46

Comentários e Melhorias

- **Base 01**
 - Bom resultado e bem balanceado, mas pouco suporte
- **Base 02**
 - Resultados satisfatórios, poderia ser testados mais hiperparametros ou modelos mais complexos
- **Base 03**
 - Problema mais complexo
 - Como estratégia de melhoria, pode-se considerar remover ou agrupar classes com baixo desempenho, especialmente se elas forem pouco representativas ou pouco relevantes para a tarefa final.