**1.2 SEGUNDO DESAFIO: TREINANDO UM MODELO DE MACHINE LEARNIG DE REVIEW DE CLIENTES**

**Etapa : Importação de Bibliotecas,** **Carregamento dos Dados e Vetorização de Texto (TF-IDF):**

Nesta etapa, começamos carregando os dados do arquivo 'review\_clientes\_melhores\_compras.csv'. Para preparar os dados para análise, realizamos as seguintes ações:

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Limpeza de Texto: Removemos caracteres especiais e pontuações das avaliações dos clientes.

Conversão de Texto para Minúsculas: Padronizamos o texto convertendo-o para minúsculas, evitando discrepâncias devido à formatação.

Remoção de Stopwords: Eliminamos palavras comuns que não contribuem significativamente para a análise.

Finalmente, convertemos o texto em representações numéricas usando a técnica TF-IDF e dividimos os dados em conjuntos de treinamento e teste.

Tela de celular com aplicativo aberto

Descrição gerada automaticamente

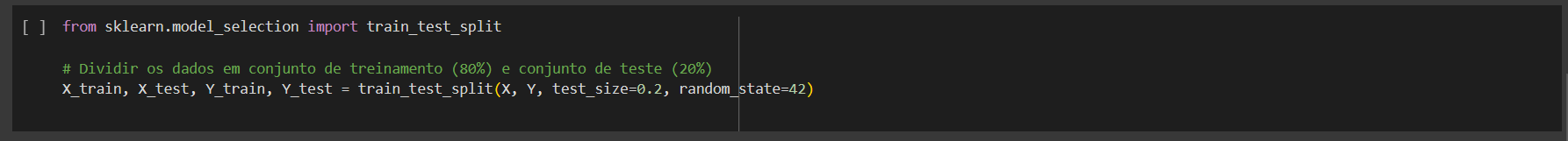
Um vetorizador TF-IDF é criado usando TfidfVectorizer. Ele é configurado para considerar no máximo 5000 recursos.

O vetorizador TF-IDF é aplicado aos textos limpos na coluna 'review\_cliente', resultando em uma representação numérica das avaliações dos clientes.

A variável alvo 'sentimento' é atribuída a 'Y'.

**Etapa : Divisão dos Dados:**

Os dados são divididos em conjuntos de treinamento (80%) e teste (20%) usando train\_test\_split. O parâmetro random\_state é configurado para garantir a reprodutibilidade.



**Etapa: Avaliação do Modelo**: escolhemos o modelo Random Forest para prever avaliações positivas e negativas. Treinamos o modelo com base nos dados de treinamento, permitindo que ele aprendesse padrões nas avaliações dos clientes. O modelo treinado é usado para fazer previsões nos dados de teste.

As seguintes métricas de desempenho são calculadas:

Acurácia usando accuracy\_score.

Matriz de Confusão usando confusion\_matrix.

Relatório de Classificação usando classification\_report.

Curva ROC usando roc\_curve e auc.

Texto

Descrição gerada automaticamente

**Etapa: Visualização da Curva ROC: e Relatório**

A Curva ROC é plotada usando matplotlib para visualizar o desempenho do modelo.

Exibição de Métricas:

A acurácia, a matriz de confusão e o relatório de classificação são exibidos na saída para avaliação do modelo.

O código, em resumo, carrega os dados, pré-processa o texto, treina um modelo Random Forest para prever as avaliações dos clientes e avalia o desempenho do modelo com várias métricas. O resultado inclui a acurácia do modelo e informações detalhadas sobre como ele se saiu na classificação de avaliações positivas e negativas.

Interface gráfica do usuário

Descrição gerada automaticamente

Avaliamos o desempenho do modelo usando várias métricas, incluindo acurácia, matriz de confusão, precisão, recall, F1-score e Curva ROC. Objetivo era selecionar um modelo com acurácia superior a 80%. Com base nas métricas, escolhemos o modelo Random Forest.