UNIVERSIDADE ESTÁCIO DE SÁ

NITERÓI I

**Projeto de Análise de Dados de**

**Pacientes com Hepatite**

Alunos: Lucas Diniz de Araujo Pereira,

Nínive Fernandes de Farias Guimarães Natal

Professor: Marcelo Teixeira

2024

Niterói, Rio de Janeiro

**Introdução**

Este projeto tem como objetivo principal desenvolver uma árvore de decisão para a classificação de pacientes com hepatite, utilizando técnicas de aprendizado de máquina e linguagem Python.

1. **Definição do grupo de trabalho e parte interessada envolvida:**

Alunos Lucas Diniz de Araujo Pereira (Matrícula: 202203897411) e Nínive Fernandes de Farias Guimarães Natal (Matrícula: 202309146381)

1.2 **Identificar as partes interessadas: fazer contato com a comunidade e verificar o interesse na participação do projeto. Nesse contato pode ser necessária uma ou mais visitas ao local pretendido**

Não houve necessidade de realizar trabalhos de campo neste projeto.

1.3 **Realizar o diagnóstico das demandas necessárias e definir o escopo e a priorização**

Foi realizado o diagnóstico das demandas necessárias, priorizando a utilização de uma ferramenta que auxilie na classificação sobre o tratamento de pacientes com hepatite.

**1.4. Definição de cronograma de atuação/ visitas às partes interessadas**

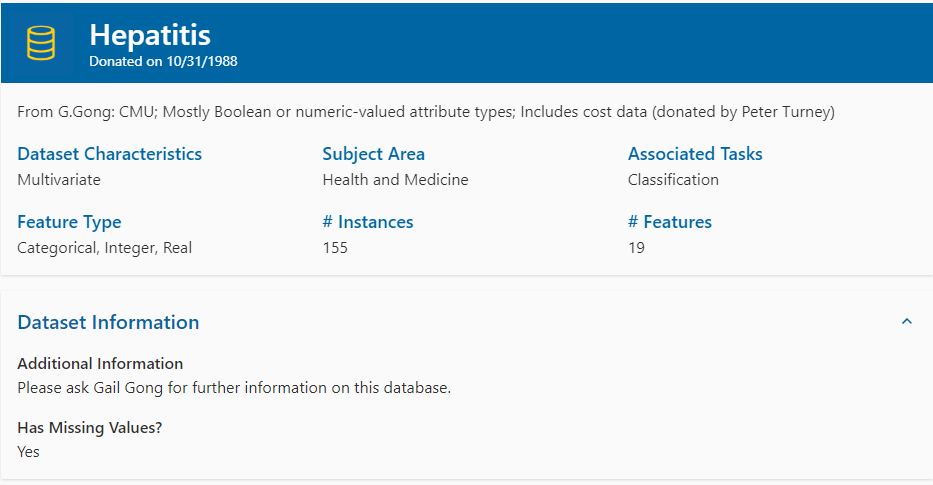
--

**2. Diagnóstico e teorização do projeto**

Este é um projeto feito em linguagem Python que faz a classificação de pacientes que possuem Hepatite, utilizando aprendizado de máquina. A base de dados dos pacientes é proveniente do website *UCI Machine Learning Repository.* O objetivo deste trabalho é realizar a análise de grandes dados e fazer a previsão de pacientes com a doença com antecedência a fim de viabilizar antecipação na tomada de decisão e realização, usando ferramentas e bibliotecas do Python, como Pandas, árvore de decisão e entre outros.

**3. Planejamento e desenvolvimento do projeto**

Como mencionado anteriormente, foi utilizado uma base de dados para classificação (database) do site [UCI Machine Learning Repository](https://archive.ics.uci.edu/dataset/46/hepatitis) contendo informações de pacientes com Hepatite.



Após baixar o documento, pegamos o arquivo “hepatitis.names”, selecionamos as colunas informadas nele, e colocamos em um novo arquivo chamado “hepatitis.csv”. Depois disso, abrimos o arquivo “hepatitis.data” que contém todos os dados dos pacientes que vamos precisar para o projeto; selecionamos eles e colocamos também no arquivo “hepatitis.csv”. Após esse processo, este arquivo está pronto para poder ser lido na linguagem Python.

**4. Detalhamento técnico do projeto**

**4.1. Documentar um projeto de Tópicos de Big Data em Python que esteja alinhado aos objetivos e objetivos** **sociocomunitários deste plano de aprendizagem**

**4.1.1. Escopo do Projeto**

Desenvolver um sistema de classificação de pacientes com hepatite baseado em uma árvore de decisão.

**4.1.2. Requisitos Funcionais do sistema**

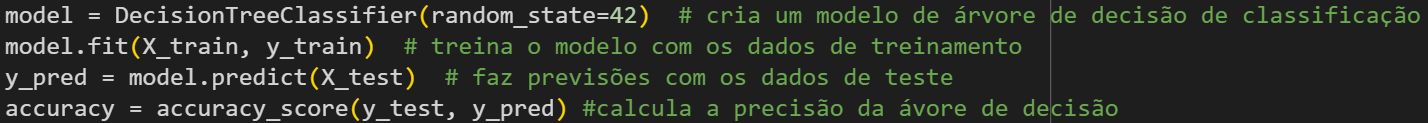
\* Importar e pré processar dados de um arquivo CSV.



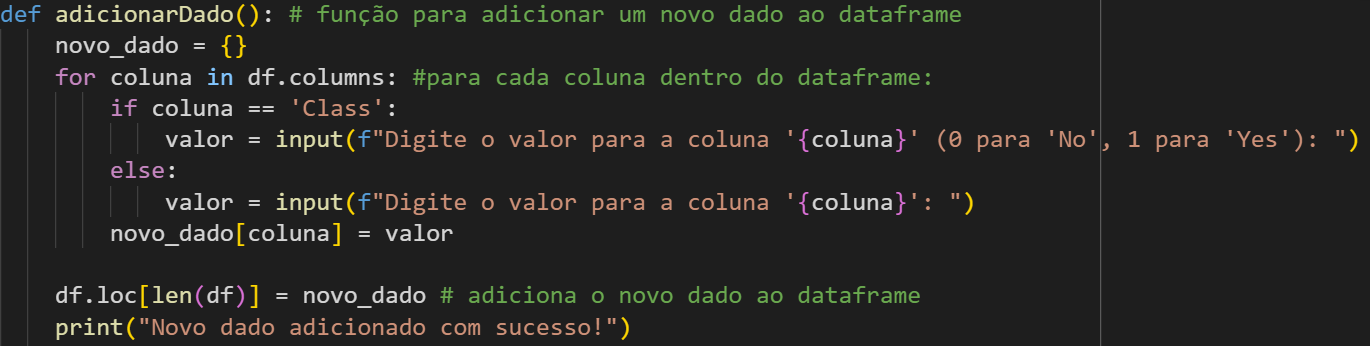
\* Treinar um modelo de árvore de decisão.



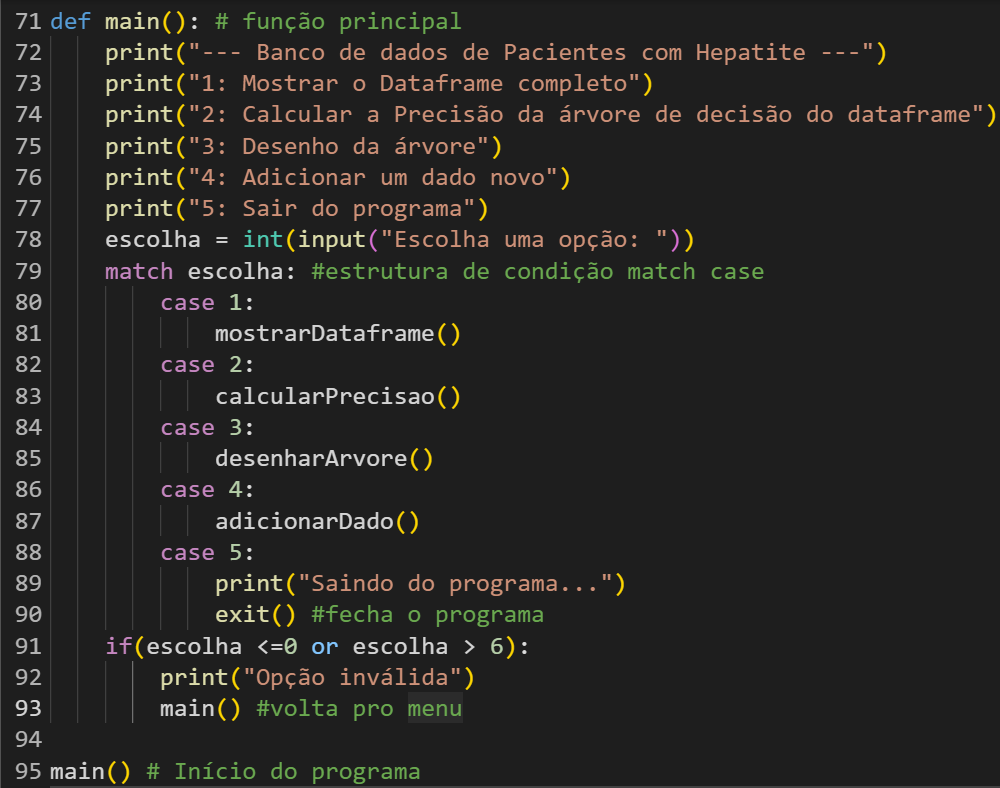
\* Avaliar a precisão do modelo.



\* Adicionar novos dados e visualizar a árvore de decisão.



**4.1.3 Modelar as classes do sistema**



*Classe principal do sistema com funções para mostrar o Dataframe completo, avaliar precisão, desenhar a árvore e adicionar novos dados.*

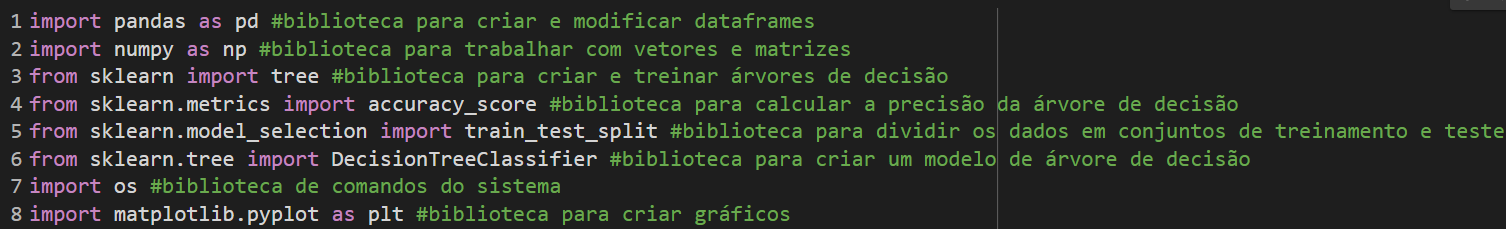
**4.1.4. Contato com a parte interessada para o processo de validação da proposta**

Não foi realizada pois não foi pretendemos procurar stakeholders.

**4.2. Desenvolver o sistema**

**4.2.1. Desenvolver as classes descritas**

Implementação das funções principais conforme descrito acima.



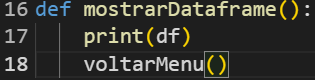
Primeiramente, é importado todas as bibliotecas e ferramentas necessárias para o projeto, como a principal a ser mencionada, a “sklearn” que é a responsável por criar e treinar árvores de decisão.

Logo após isso, é escrita a linha df=pd.read\_csv("hepatitis.csv", na\_values='?'), comando que lê o arquivo csv e preenche os valores ausentes com “?”.

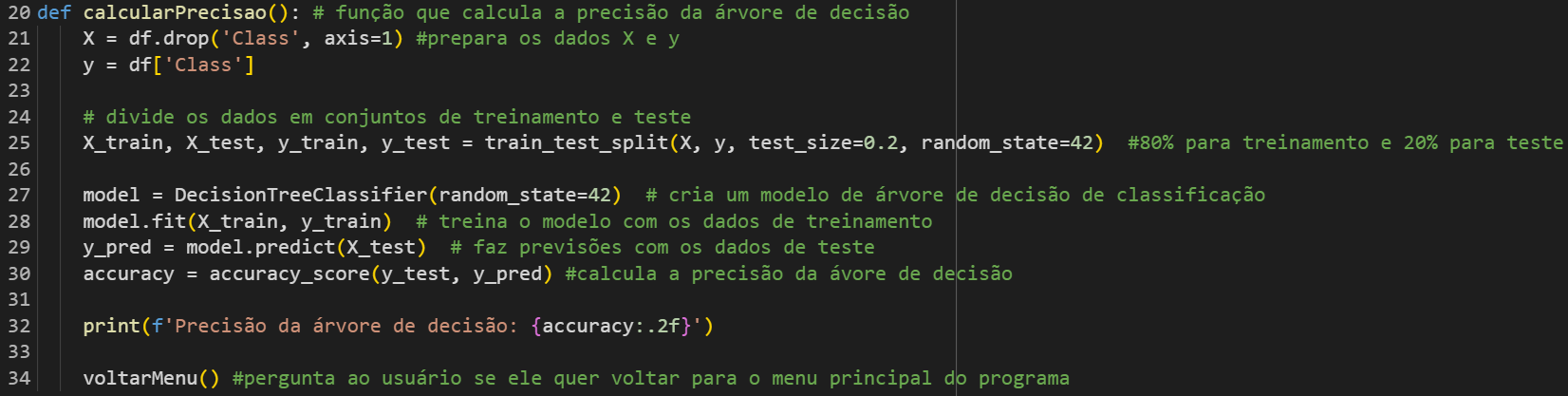


Depois de importar o arquivo, são feitos dois comandos de dataframe. O primeiro “df.fillna” Preenche os valores ausentes com a média das colunas numéricas, a fim de que não haja valores nulos na tabela.

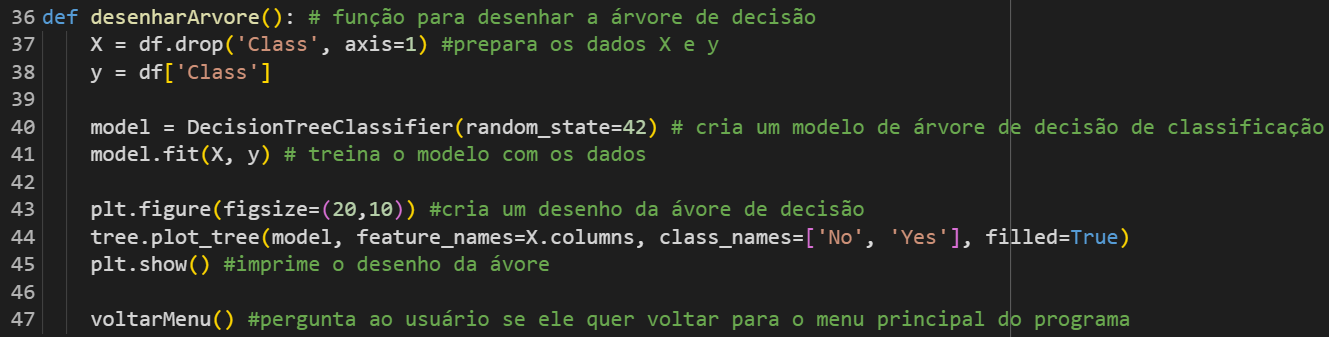
O segundo “df.fillna” preenche valores ausentes com a moda das colunas categóricas.



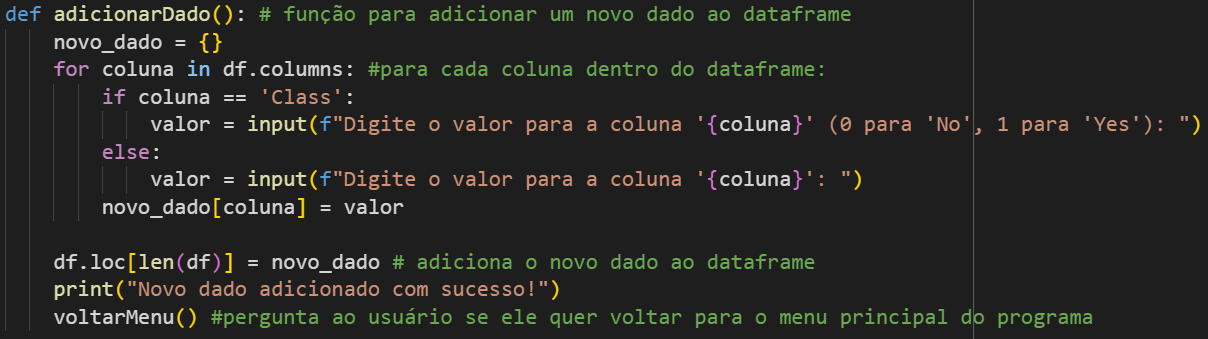
A primeira classe do programa mostrarDataframe() realiza uma tarefa simples de imprimir a tabela do dataframe para o usuário, no console do programa. Na terceira linha chama a função voltarMenu() que será explicada mais abaixo.



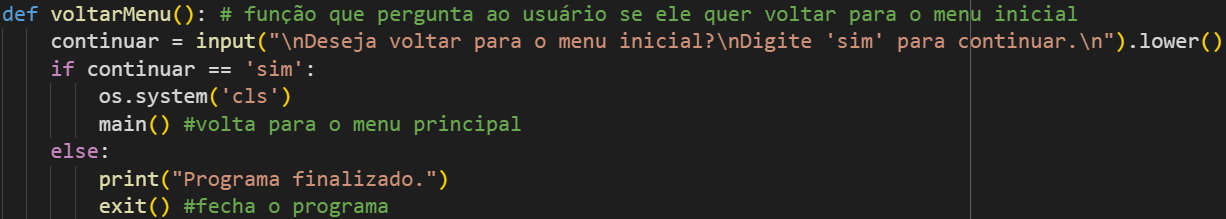
A segunda classe do programa se chama calcularPrecisao(). É neste bloco em que a árvore de decisão será treinada. No final da execução, será mostrado um print para o usuário mostrando a precisão que a árvore obteve.



Na classe desenharArvore() é o bloco onde vai ser criado a representação gráfica da árvore de decisão. Na linha 45 é executado o comando em que imprime a imagem na tela.



Na classe adicionarDado() foi construída para caso se o usuário quiser adicionar um dado novo no dataframe. O laço ‘for’ junto com o ‘if’ vai percorrer por cada uma das colunas presentes no dataframe e ir perguntando ao usuário para digitar ‘0’ para ‘não’ e ‘1’ para ‘Sim’. Depois executar todo o laço, o programa vai mostrar na tela do usuário que o novo dado foi adicionado com sucesso.



Por fim, a última classe voltarMenu(), que é executada sempre no final de todas as outras classes mostradas acima. Um bloco simples que pergunta ao usuário se ele deseja voltar para o menu principal do programa. Caso digite ‘sim’, ele chama a função main() onde está localizado o bloco principal do programa. Caso digite qualquer outra palavra além de ‘sim’, programa é finalizado.

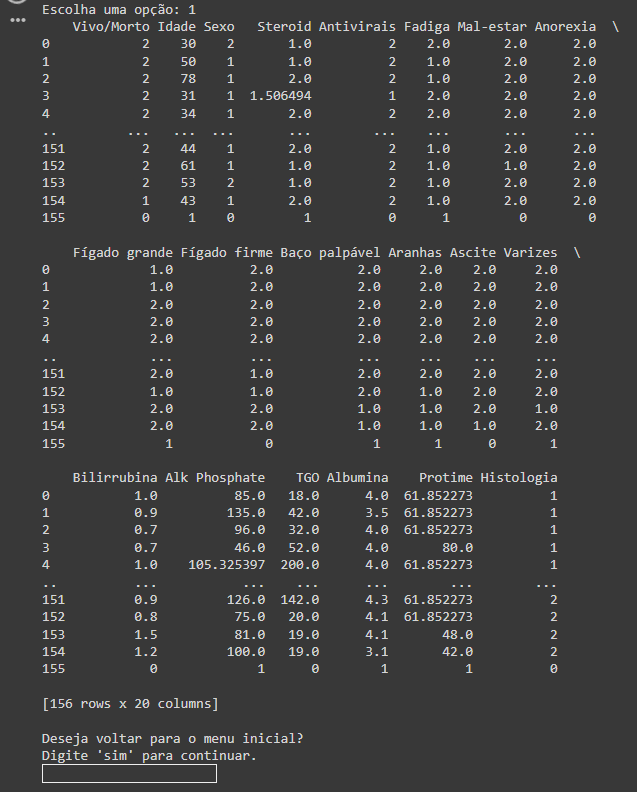
**4.2.2. Criar as interfaces do Sistema**

Neste sistema não foi necessário o uso de interfaces.

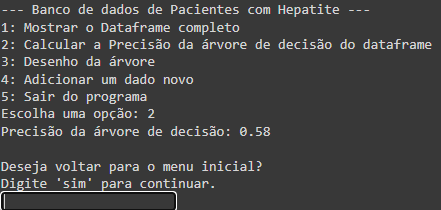
**4.3. Testar e implantar o sistema**

**4.3.1. Testes de unidade**

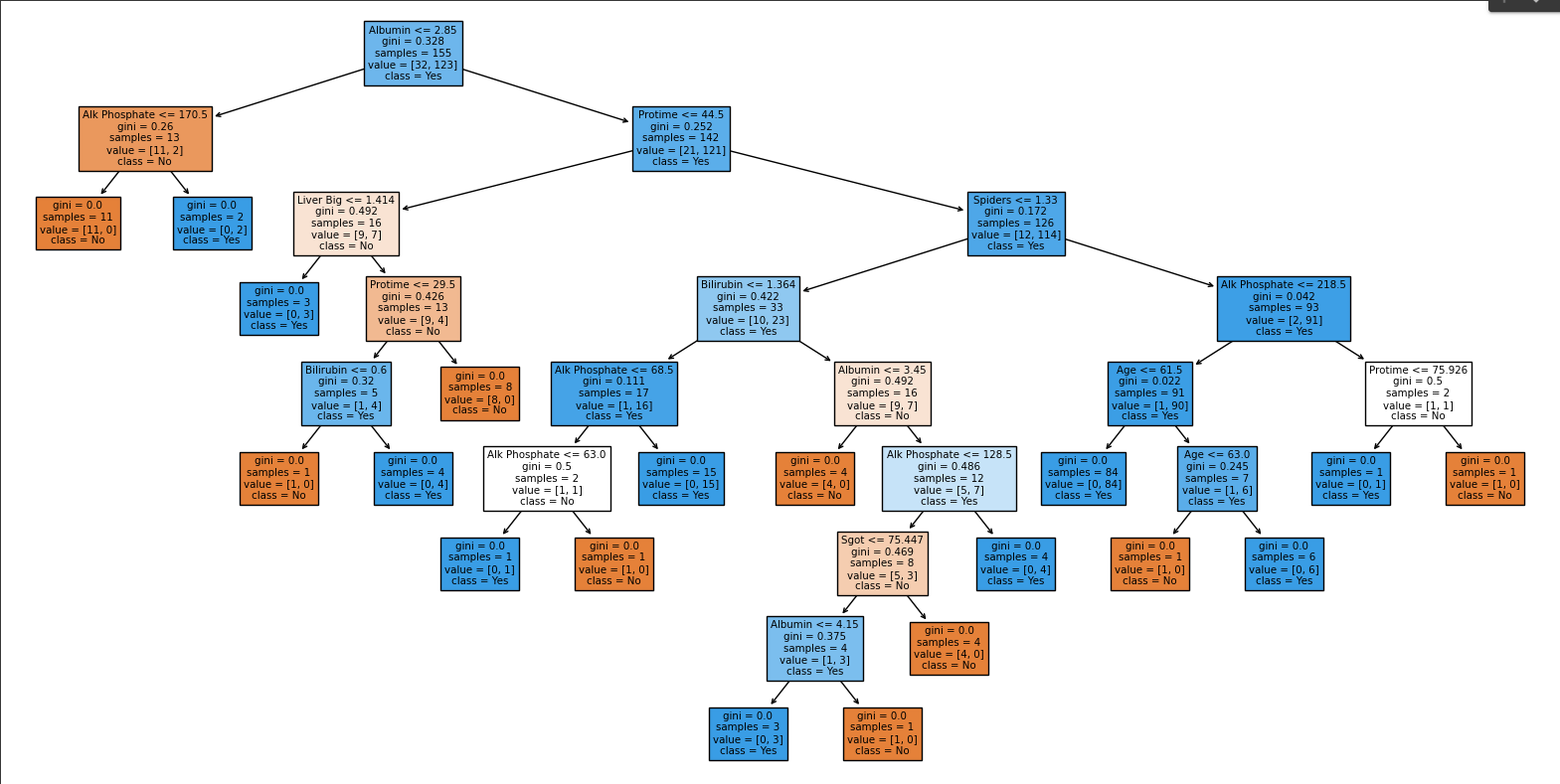
Teste de funções individuais do programa. (Executados através da plataforma Google Colab).



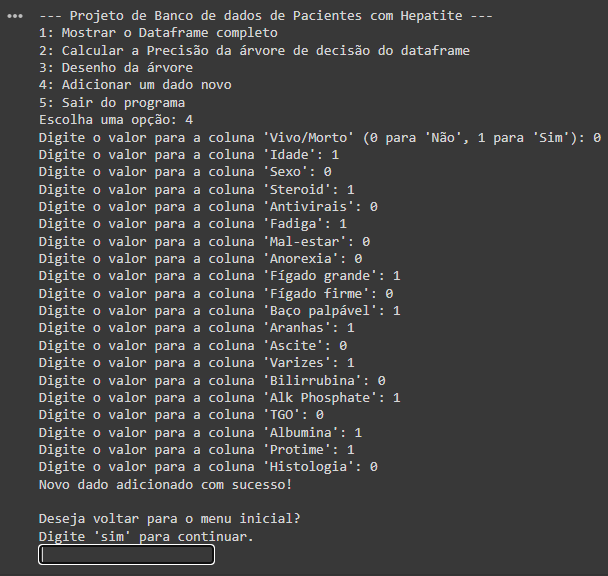
Resultado quando o usuário escolher “mostrar o dataframe”:



Resultado da precisão da árvore de decisão:



Desenho da árvore de decisão.



Resultado após adicionar um novo dado no dataframe.

**4.3.2. Realizar o teste de integração com todas as partes do sistema**

Verificação do funcionamento completo do sistema, incluindo o fluxo de dados do início ao fim.

**4.3.3. Implantar o sistema na instituição**

Nesta parte ocorreria a utilização desse sistema para os profissionais de saúde interessados, porém este projeto foi feito apenas para teste e estudos, não sendo feito para esse fim.

**4.3.4. Treinar o usuário responsável**

Sessões de treinamento para garantir que os usuários saibam como operar o sistema da forma correta.

**5. Conclusão**

Este projeto exemplifica a aplicação prática de técnicas de ciência de dados para resolver problemas reais na área da saúde, especificamente na classificação de pacientes com hepatite. Através do desenvolvimento de um sistema de árvore de decisão, foi possível aplicar conceitos teóricos em um contexto prático, proporcionando uma ferramenta útil para os profissionais de saúde envolvidos.

**5.1 Relato Individual**

Desenvolver esse trabalho de classificação de pacientes se mostrou ser eficiente e bem-sucedida em vários aspectos. O código desenvolvido é claro e de fácil entendimento para possíveis melhorias e futuras extensões, além da precisão da árvore ter conseguido atingir uma boa precisão.

O projeto foi bastante útil. Ele não apenas permitiu a aplicação dos conhecimentos adquiridos ao longo do curso, mas também destacou a importância da análise de dados e do machine learning na área da saúde.

Agradeço ao professor Marcelo Teixeira pelo apoio e orientação durante todo o desenvolvimento do projeto, fornecendo informações valiosas e feedback contínuo.

**6. Apresentação dos Projetos: nesta etapa, cada grupo deve apresentar o projeto e os resultados obtidos em cada trabalho. Para garantir a participação de todos os alunos no grupo, o docente pode definir um plano de ação, combinado anteriormente com a turma. O docente deve avaliar a entrega coletiva, a entrega individual e a avaliação de reação da parte interessada, presentes no Roteiro de Extensão, quanto à eficiência e eficácia do atingimento dos objetivos do projeto.**