3. Liver Cirrhosis Stage Prediction (livercirrhosis.ipynb)

- **Problem Statement**: The project aims to predict the stage of liver cirrhosis based on clinical and demographic data. This appears to be a multi-class classification problem (predicting 'Stage').
- **Dataset Used**
  :
  - The dataset was loaded from liver_cirrhosis.csv.
  - The target variable is 'Stage'.
  - Features include 'N_Days', 'Status', 'Drug', 'Age', 'Sex', 'Ascites', 'Hepatomegaly', 'Spiders', 'Edema', 'Bilirubin', 'Cholesterol', 'Albumin', 'Copper', 'Alk_Phos', 'SGOT', 'Tryglicerides', 'Platelets', 'Prothrombin'.

- **Methodology and Approach**
  :
  - **Data Preprocessing**
    :
    - The 'Status' column was dropped.
    - Categorical features ('Sex', 'Drug', 'Ascites', 'Hepatomegaly', 'Spiders', 'Edema') were manually encoded into numerical representations using map.
    - The 'Stage' column was converted to an integer type.
    - Rows with missing values were dropped using dropna().
    - Numerical features were scaled using StandardScaler.

  - **Model Training**
    :
    - The data was split into training (80%) and testing (20%) sets.
    - Three classification models were trained using
      GridSearchCV
      for hyperparameter tuning:
      - Logistic Regression (LogisticRegression with multi_class='ovr')
      - Decision Tree Classifier (DecisionTreeClassifier)
      - Random Forest Classifier (RandomForestClassifier)

- **Model Evaluation**
  :
  - Models were evaluated based on accuracy, Mean Squared Error (MSE), classification reports, and confusion matrices.

- **Results and Conclusion**
  :
  - **Logistic Regression**: Best Params: {'C': 1}. Accuracy: 55.36%.
  - **Decision Tree**: Best Params: {'max_depth': 3, 'min_samples_split': 2}. Accuracy: 58.93%.
  - **Random Forest**: Best Params: {'max_depth': 7, 'n_estimators': 100}. Accuracy: 60.71%.
  - The Random Forest classifier achieved the highest accuracy among the models tested.
  - The best model (Random Forest, in this case) was saved to best_liver_stage_model.pkl.