**Revolutionizing Liver Care: Predicting Liver Cirrhosis Using Advanced Machine Learning Techniques**

**1. Introduction**

**1.1. Project Overview**

This project aims to build a machine learning model to predict liver cirrhosis based on various patient features. By analysing these features, the model will classify patients into risk categories, aiding in early diagnosis and treatment.

**1.2. Objectives**

* Collect and prepare a dataset of liver health characteristics.
* Perform exploratory data analysis (EDA) and visualize the data.
* Build and evaluate multiple machine learning models.
* Optimize the best-performing model using hyperparameter tuning.
* Deploy the final model for practical use.

**2. Project Initialization and Planning Phase**

**2.1. Define Problem Statement**

The goal is to classify patients' risk levels for liver cirrhosis based on their medical data. Accurate prediction will support better management and early intervention for liver health.

**2.2. Project Proposal (Proposed Solution)**

The solution involves developing several machine learning models to predict liver cirrhosis. We will select and optimize the best model based on performance metrics to achieve the highest accuracy.

**2.3. Initial Project Planning**

Initial planning included setting up the project environment, defining objectives, and outlining the workflow for data collection, preprocessing, model development, and evaluation.

**3. Data Collection and Preprocessing Phase**

**3.1. Data Collection Plan and Raw Data Sources Identified**

The dataset for this project was sourced from Kaggle, containing patient data relevant to liver cirrhosis prediction

(Dataset link: https://www.kaggle.com/datasets/bhavanipriya222/liver-cirrhosis-prediction).

**3.2. Data Quality Report**

* **Data Shape:** The dataset initially comprised [number of rows, number of columns] rows and columns.
* **Missing Values:** Handled by dropping rows with missing values.

**3.3. Data Exploration and Preprocessing**

* **Univariate Analysis:** Histograms were plotted for numerical features.
* **Bivariate Analysis:** Scatter plots and pair plots explored relationships between features.
* **Outlier Handling:** Outliers were detected and managed using the IQR method.

**4. Model Development Phase**

**4.1. Feature Selection Report**

Features relevant to liver cirrhosis prediction were selected, and data scaling was applied to standardize the input.

**4.2. Model Selection Report**

* **Models Tested:** Naive Bayes, Random Forest, Logistic Regression, Ridge Classifier, Support Vector Classifier, KNN, XG Boost.
* **Evaluation Metrics:** Accuracy, Confusion Matrix, Classification Report.

**4.3. Initial Model Training Code, Model Validation and Evaluation Report**

* **Code:** Model training and evaluation steps were implemented for each algorithm.
* **Validation:** Models were validated using a test set, with performance metrics recorded. The KNN model achieved the highest accuracy of 86.32%.

**5. Model Optimization and Tuning Phase**

**5.1. Hyperparameter Tuning Documentation**

* **KNN:** Optimized by tuning the number of neighbors and distance metrics.
* **XG Boost:** Hyperparameters tuned for learning rate, max depth, and n\_estimators.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **5.2. Performance Metrics Comparison Report Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Naive Bayes | 35.79% | 0.00 | 0.00 | 0.00 |
| Random Forest | 35.79 | 0.00 | 0.00 | 0.00 |
| Logistic Regression CV | 81.58% | 91.80 | 79.43% | 86.49 |
| Ridge Classifier | 84.21% | 93.44 | 83.82 | 88.37 |
| Support Vector Classifier | 35.79% | 0.00 | 0.00 | 0.00 |