**Model Performance Report**

**Introduction**

Liver disease is a major health concern worldwide, and early detection is crucial for effective treatment. This project aims to develop a machine learning-based liverdisease prediction model using patient medical records. The dataset includes key biomarkers like bilirubin levels, enzyme counts, and protein ratios to help classify patients into diseased or non-diseased categories.

Various machine learning models, including Logistic Regression, Decision Trees**,** Random Forest, Gradient Boosting, and XGBoost, were evaluated to determine the most effective approach. The report presents data preprocessing techniques and model performance comparison.

**Basis of Model Design**

* The dataset contains medical records with features like **bilirubin levels, enzyme levels, and proteins**, which are critical for diagnosing liver disease.
* **Feature Engineering**: Categorical values (Gender) were converted into numerical values.
* **Data Cleaning**: Duplicates and outliers were removed to improve model accuracy.
* **Standardization**: The dataset was standardized to ensure better performance across different models.
* **Train-Test Split**: 80%-20% split was used to train and evaluate models.

**Model Comparison**

The following models were trained and evaluated:

|  |  |
| --- | --- |
| **Model** | **Accuracy (%)** |
| Logistic Regression | 71.4 |
| K-Nearest Neighbors (KNN) | 68.3 |
| Support Vector Classifier (SVC) | 72.1 |
| Decision Tree Classifier | 74.6 |
| Random Forest Classifier | 75.9 |
| Gradient Boosting Classifier | 76.2 |
| XGBoost Classifier | **77.4** |

**Best Model:** XGBoost performed the best with an accuracy of **77.4%**, making it the most suitable model for production.

**Recommendation**

**XGBoost** is the best model for production due to its highest accuracy and ability to handle complex relationships.

**Conclusion**

This project successfully developed a machine learning model to predict liver disease based on medical attributes. Several models were evaluated, with **XGBoost achieving the highest accuracy (77.4%)**, making it the best choice for deployment. Challenges like **data imbalance, missing values, and outliers** were addressed using preprocessing techniques.