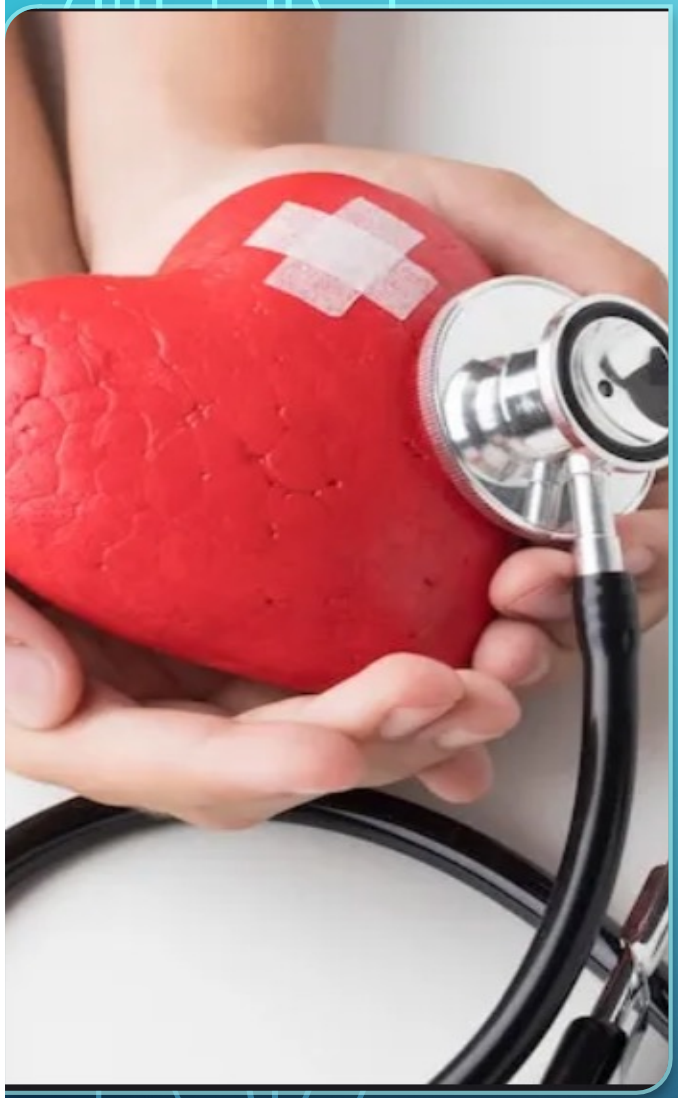


PREDICTION OF HEART DISEASES USING RANDOM FOREST

AAKASH SINGH (2100290109001)

LALIT KISHOR (2200290109008)

SUHEL KHAN (2200290109016)

SUPERVISION :

PROF. GAURAV PARASHAR

INTRODUCTION

- Heart disease is a leading global cause of mortality.
- Early detection is vital for effective treatment.
- Machine Learning (ML) helps identify patterns for better diagnosis.

PROBLEM STATEMENT

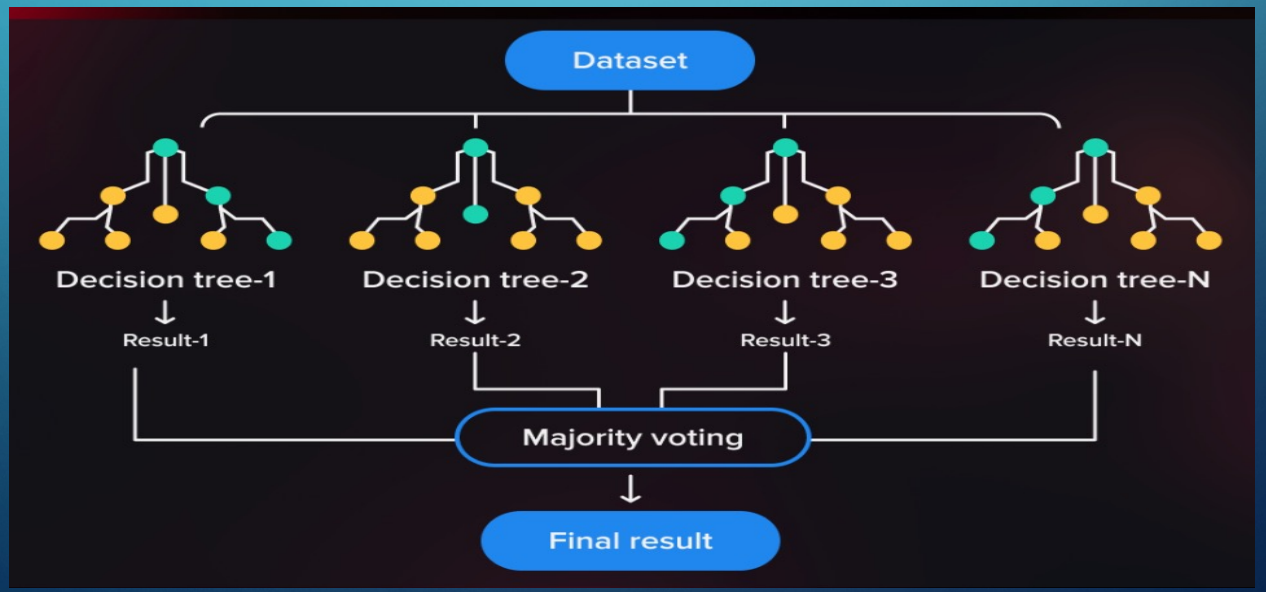
- Traditional diagnostic methods may miss patterns in data.
- Need for a reliable, automated model for heart disease prediction.
- Aim: Build an accurate ML model for early detection.

DATASET OVERVIEW

- Used dataset with 303 records and 14 clinical features.
- Includes: age, sex, chest pain type, cholesterol, etc.
- Sourced from trusted medical databases (e.g., Kaggle).

RANDOM FOREST ALGORITHM

- Ensemble learning model combining decision trees.
- Reduces overfitting and increases prediction accuracy.
- Performs well with both categorical and numerical data.



DATA PREPROCESSING

- Handled missing values using imputation.
- Normalized data for consistency.
- Used feature selection methods like PCA.

PERFORMANCE METRICS

- Accuracy: 89.92%
- Sensitivity: 91.58%, Specificity: 87.67%
- AUC: 94.16% - strong discrimination ability.

COMPARATIVE ANALYSIS

- Random Forest outperformed Logistic Regression, SVM, Naive Bayes.
- Best overall balance of accuracy and interpretability.
- SVM showed highest accuracy (94.41%) but less interpretability.

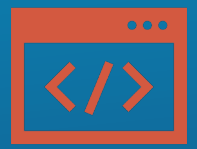
IMPLEMENTATION

- Frontend: HTML, CSS, JavaScript.
- Backend: Python and Jupyter Notebook.
- Integrated model to website for real-time prediction.

CONCLUSION

Random Forest model effectively predicts heart disease.

High accuracy and robust against noise.



User-friendly web interface enhances accessibility.

FUTURE SCOPE

1

Integrate real-time health data from wearables.

2

Use larger, diverse datasets for broader generalization.

3

Expand model for other chronic disease predictions.

THANK
YOU