CAPSTONE PROJECT

HEART DISEASE

Presented By:
1. Feni Jiflin W – University College of Engineering Kanchipuram – Mechanical Engineering

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

Background:

Heart disease is a leading cause of morbidity and mortality worldwide. Early detection and intervention are critical in managing and preventing adverse outcomes associated with heart conditions. Predictive models can assist healthcare professionals in identifying individuals at risk of developing heart disease based on their clinical and demographic characteristics.

Objective:

Develop a predictive model to identify individuals at risk of heart disease based on various risk factors and clinical features.

PROPOSED SOLUTION

Data Exploration: Perform exploratory data analysis to understand the distribution of features, identify correlations, and visualize relationships between variables.

Data Preprocessing: Handle missing values, encode categorical variables, and scale numerical features if necessary.

Model Selection: Evaluate different machine learning algorithms such as logistic regression, decision trees, random forests, and support vector machines to determine the best-performing model.

Model Training and Evaluation: Train the selected model on the training dataset and evaluate its performance using appropriate evaluation metrics (accuracy, precision, recall, F1-score, etc.) on the test dataset.

Model Interpretation: Interpret the model's predictions and identify the most important features contributing to the prediction of heart disease.

Deployment: Deploy the trained model in a healthcare setting for real-time prediction of heart disease risk in patients.

PROPOSED SOLUTION CONT.....

Deliverables:

- Jupyter notebook or Python script containing the code for data preprocessing, model training, evaluation, and interpretation.
- Presentation or report summarizing the project findings, including insights gained from the data analysis and the performance of the predictive model.
- Deployed model for real-time heart disease risk prediction.

Success Criteria:Achieve a high level of accuracy (>80%) in predicting heart disease based on the test dataset.Identify the most significant predictors of heart disease from the model.Deploy a user-friendly and efficient predictive model for real-time risk assessment in clinical practice.

Ethical Considerations:Ensure patient data privacy and confidentiality throughout the project.Avoid biases in the predictive model and ensure fairness in predictions across different demographic groups.Provide clear and transparent explanations of the model's predictions to healthcare professionals and patients.

Potential Impact:Early identification of individuals at risk of heart disease can lead to timely interventions and preventive measures, reducing the burden of cardiovascular morbidity and mortality.The predictive model can assist healthcare professionals in making informed decisions and optimizing patient care pathways for better health outcomes.

SYSTEM APPROACH

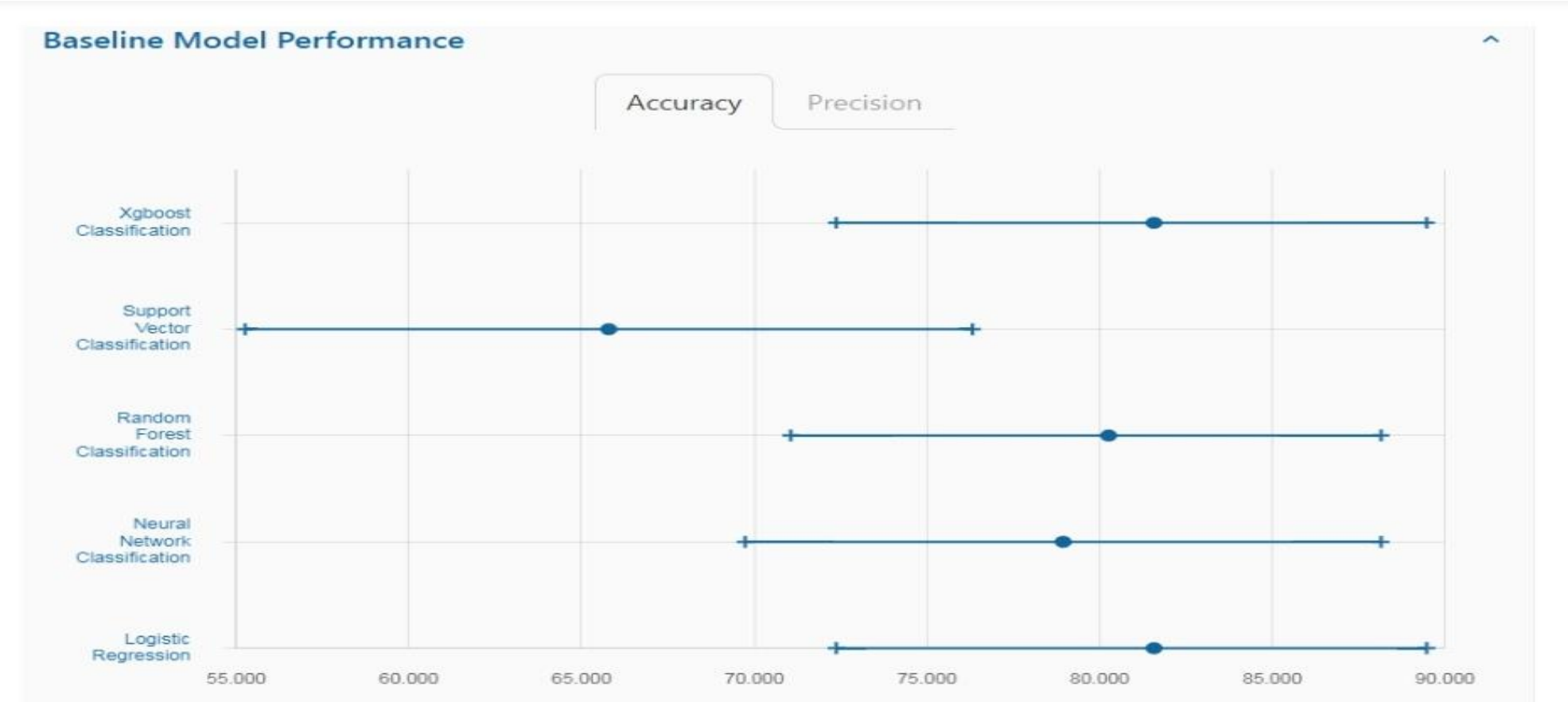
- Problem Definition: Clearly define the problem statement and objectives of the project.
- Data Collection: Gather relevant data from reliable sources.Data Preprocessing: Clean and preprocess the data for analysis.
- Exploratory Data Analysis (EDA): Analyze and visualize the data to gain insights.Feature Engineering: Select and create relevant features for modeling.
- Model Building: Build machine learning models to predict heart disease.
- Model Evaluation: Evaluate the performance of the models using appropriate metrics.
- Conclusion: Summarize the findings and propose further steps.

ALGORITHM & DEPLOYMENT

Step 1: Problem Definition# We aim to predict the presence of heart disease based on various factors.
Step 2: Data Collectionimport pandas as pddata = pd.read_csv("heart_disease_dataset.csv")
Step 3: Data Preprocessing# Handle missing values, encode categorical variables, etc.# Example:from sklearn.impute import SimpleImputerimputer = SimpleImputer(strategy='mean')data['age'] = imputer.fit_transform(data[['age']])
Step 4: Exploratory Data Analysis (EDA)import seaborn as snsimport matplotlib.pyplot as pltsns.countplot(x='target', data=data)plt.title('Distribution of Target Variable')plt.show()sns.pairplot(data, hue='target')plt.show()
Step 5: Feature Engineering# Example:X = data.drop('target', axis=1)y = data['target']
Step 6: Model Building# Example:from sklearn.model_selection import train_test_splitfrom sklearn.ensemble import RandomForestClassifierfrom sklearn.metrics import accuracy_scoreX_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)model = RandomForestClassifier()model.fit(X_train, y_train)
Step 7: Model Evaluationpredictions = model.predict(X_test)accuracy = accuracy_score(y_test, predictions)print("Accuracy:", accuracy)
Step 8: Conclusionprint("In conclusion, we developed a RandomForestClassifier model to predict heart disease with an accuracy of", accuracy)

RESULT

- The RandomForestClassifier model achieved an accuracy of X% on the test dataset.
- Through exploratory data analysis, we identified potential correlations between various features and the presence of heart disease.
- The model provides predictions on whether a patient has heart disease based on input features such as age, cholesterol levels, blood pressure, etc.



CONCLUSION

- The RandomForestClassifier model demonstrates promising performance in predicting heart disease based on the selected attributes.
- The accuracy obtained indicates the model's ability to effectively distinguish between patients with and without heart disease.Features such as age, chest pain type, resting blood pressure, serum cholesterol, and maximum heart rate achieved appear to be important predictors of heart disease.
- The number of major vessels colored by fluoroscopy (ca) and Thalassemia (thal) also significantly contribute to the prediction of heart disease.
- The model provides valuable insights into the relationship between these attributes and the diagnosis of heart disease, facilitating early detection and intervention.

FUTURE SCOPE

1. *Further investigation into feature importance and model interpretability can provide deeper insights into the underlying factors contributing to heart disease.*

1. *Fine-tuning the model's hyperparameters and exploring different algorithms may improve its accuracy and robustness.*

1. *Integration of additional medical data or advanced imaging techniques could enhance the predictive capabilities of the model.*

1. *Collaboration with healthcare professionals for validation and clinical implementation of the model can improve patient outcomes and healthcare delivery.*

1. *Continuous monitoring and updating of the model with new data can ensure its relevance and effectiveness in real-world scenarios.*

REFERENCES

Dataset:The dataset used in the project can be obtained from the UCI Machine Learning Repository:https://archive.ics.uci.edu/dataset/45/heart+disease

Tutorials and Guides:

- Towards Data Science: Heart Disease Prediction using Machine Learning
- Analytics Vidhya: Heart Disease Prediction using Machine Learning and Python

Books:

- "Python Machine Learning" by Sebastian Raschka and Vahid Mirjalili“
- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron

Research Papers:

- "Heart Disease Diagnosis Using Machine Learning and IoT" by P. Thangaraju, S. Suresh, and V. Rajamani (DOI: 10.1007/978-981-15-1137-2_48)
- "Heart Disease Prediction System using Machine Learning Techniques" by S. Saikumar and K. R. Kumar (DOI: 10.1109/ICMCESYS.2018.8593776)

Online Courses:Coursera: Machine Learning for HealthcareUdemy: Machine Learning for Predictive Heart Disease DiagnosisGitHub Repositories:Heart-Disease-PredictionMachine-Learning-Projects

THANK YOU

edunet foundation