FINAL PROJECT

NAME: ARUUN SRIRAM S.

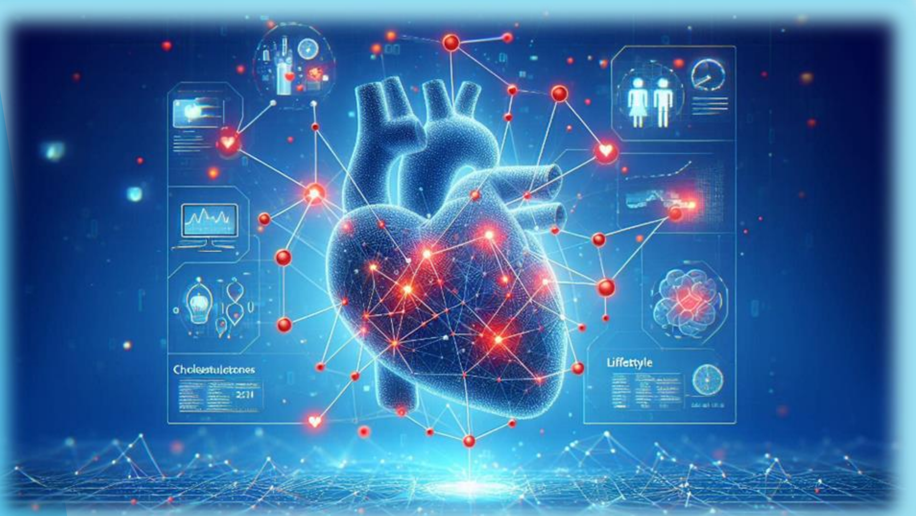
NAN MUDHALVAN ID: au821721243010

DEPT: B.TECH ARIFICIAL INTELLIGENCE AND DATA SCIENCE(3rd YEAR & 6th SEM).

COLLEGE: SIR ISSAC NEWTON COLLEGE OF ENGINEERING AND TECHNOLOGY.

MODEL NAME:

HEART ATTACK PREDICTION MODEL USING DIFFERENT CLASSIFICATION ALGORITHM



AGENDA

- ❖ PROBLEM STATEMENT.
- ❖ PROJECT OVERVIEW.
- ❖ WHO ARE THE END USERS?
- ❖ SOLUTION AND ITS VALUE PROPOSITION.
- ❖ THE WOW IN THE SOLUTION.
- ❖ DATA VISUALIZATION.
- ❖ CALCULATION OF RECALL VALUES.
- ❖ RESULT.

PROBLEM STATEMENT

- Despite advances in medical technology, heart disease remains a leading cause of mortality globally.
- Early detection of heart disease risk factors is crucial for effective prevention and intervention strategies.
- However, traditional methods of risk assessment may be limited in accuracy and efficiency.
- This project aims to address this challenge by leveraging machine learning algorithms to develop a predictive model for heart disease.
- By analyzing various patient data including demographic information, lifestyle factors, and clinical measurements, the goal is to create a tool that can reliably identify individuals at risk of developing heart disease, enabling healthcare professionals to intervene proactively and improve patient outcomes.

PROJECT OVERVIEW

1. **Data Collection:** Gather demographic information, lifestyle factors, medical history, and clinical measurements from patients.
2. **Data Preprocessing:** Clean the data, handle missing values, and encode variables for analysis.
3. **Model Selection:** Utilize a combination of machine learning algorithms such as Support Vector Machines (SVM), Naive Bayes, k-Nearest Neighbors (KNN), and Decision Tree Classifier to build the predictive model.
4. **Model Training:** Train the selected algorithms on the preprocessed data to learn patterns and relationships.
5. **Model Evaluation:** Assess the performance of each algorithm using metrics like accuracy, precision, and recall.
6. **Ensemble Learning:** Employ ensemble learning techniques to combine predictions from multiple algorithms for improved performance.
7. **Model Interpretability:** Ensure transparency and interpretability of the model's predictions to facilitate healthcare professionals' understanding and trust.

WHO ARE THE END USERS?

- **General Practitioners (GPs) and Primary Care Physicians:** They can utilize the predictive model during routine health check-ups to assess their patients' risk of developing heart disease and make informed decisions regarding preventive interventions or referrals to specialists.

- **Cardiologists:** Specialists in heart health can incorporate the predictive model into their diagnostic and treatment protocols to aid in early detection and management of heart disease among their patients.

- **Medical Researchers:** Professionals involved in cardiovascular research can use the predictive model to analyze trends in heart disease risk factors and outcomes, leading to advancements in preventive strategies and treatment modalities.

- **Public Health Officials:** The predictive model's insights can inform public health initiatives aimed at reducing the overall prevalence and impact of heart disease within communities through targeted interventions and health education programs.

SOLUTION AND ITS VALUE PROPOSITION

Solution: Develop a versatile predictive model for heart disease identification using a combination of machine learning algorithms including Support Vector Machines (SVM), Naive Bayes, k-Nearest Neighbors (KNN), and Decision Tree Classifier.

Proposition:

- **Comprehensive Algorithm Integration:** Develop a heart disease prediction model utilizing diverse machine learning algorithms including Support Vector Machine (SVM), Naive Bayes, k-Nearest Neighbors (kNN), Decision Tree Classifier, and Random Forest.
- **Robust Performance Evaluation:** Evaluate each algorithm's performance using metrics like accuracy, precision, recall, and F1-score to identify the most effective model for heart disease prediction.
- **Hyperparameter Optimization:** Employ GridSearchCV to fine-tune the hyperparameters of each algorithm, enhancing their predictive power and generalization performance.
- **Ensemble Learning:** Implement ensemble techniques such as model averaging or stacking to combine the predictions of multiple algorithms, potentially improving overall model accuracy and robustness.
- **Interpretability and Transparency:** Ensure the model's interpretability by providing insights into the feature importance and decision-making process, enabling healthcare professionals to understand and trust the predictions.
- **Scalable Deployment:** Design the model for easy integration into healthcare systems and seamless deployment in clinical settings, ensuring scalability and usability across different healthcare environments.

Overall, the heart disease prediction model leveraging SVM, Naive Bayes, KNN, and Decision Tree Classifier algorithms offers a robust and adaptable solution for early detection and prevention of heart disease, empowering healthcare professionals with actionable insights to improve patient outcomes and reduce the burden of cardiovascular diseases.

THE WOW IN YOUR SOLUTION

1.Reliable Tool for Early Detection: The predictive model is positioned as a reliable tool for early detection of heart disease, which is crucial for timely intervention and prevention strategies.

2.Improved Patient Outcomes: By accurately identifying individuals at risk of heart disease, the model contributes to improving patient outcomes, potentially leading to reduced morbidity and mortality rates associated with cardiovascular diseases.

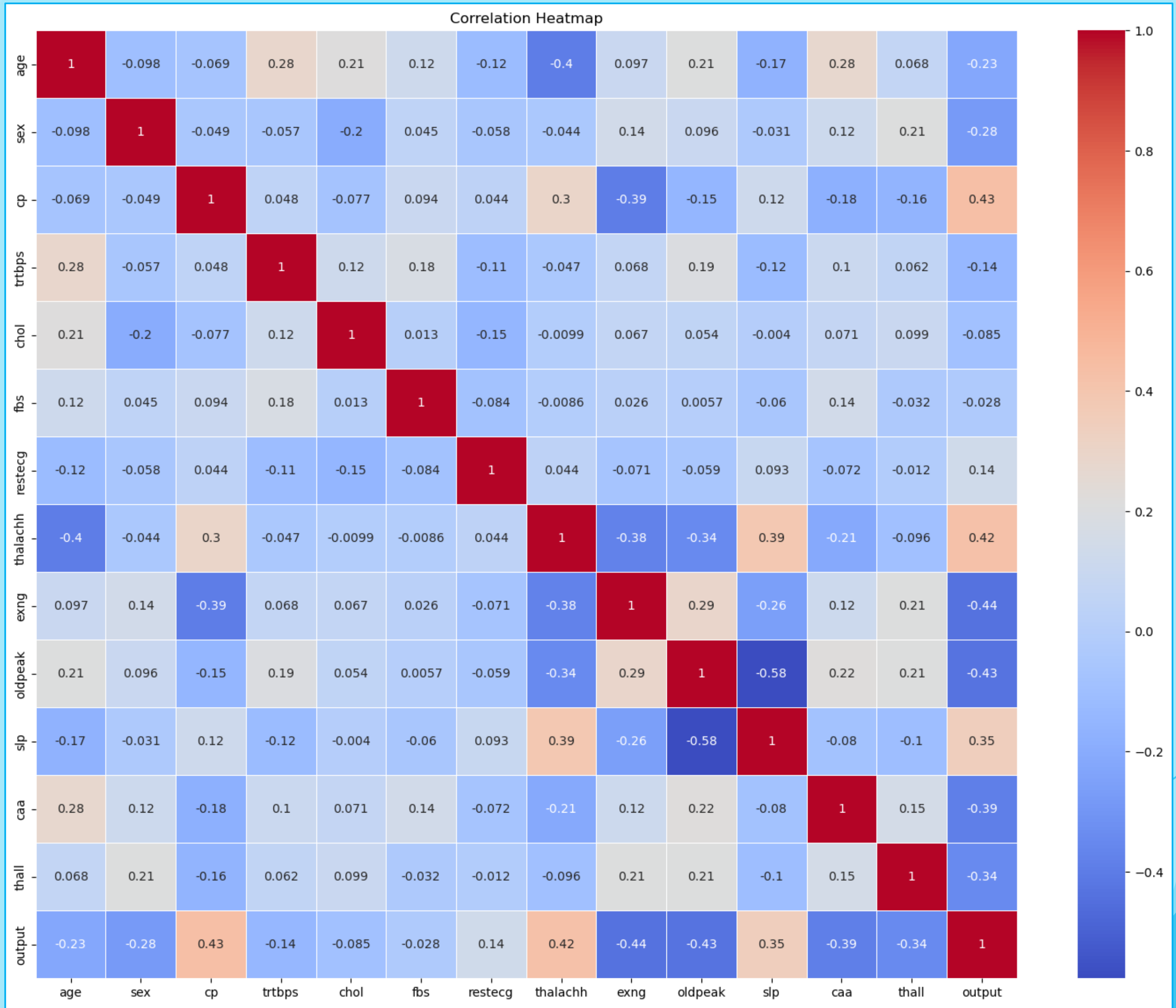
3.Healthcare Efficiency: The implementation of the predictive model enhances healthcare efficiency by enabling proactive management of heart disease, thereby optimizing resource allocation and reducing healthcare costs.

4.Utilization of Advanced Algorithms: Leveraging a combination of advanced machine learning algorithms, including SVM, Naive Bayes, KNN, and Decision Tree Classifier, underscores the sophistication and effectiveness of the developed model.

5.Continuous Monitoring and Adaptation: The establishment of mechanisms for continuous monitoring and improvement ensures that the model remains effective and relevant over time, adapting to evolving healthcare data and practices.

DATA VISUALIZATION

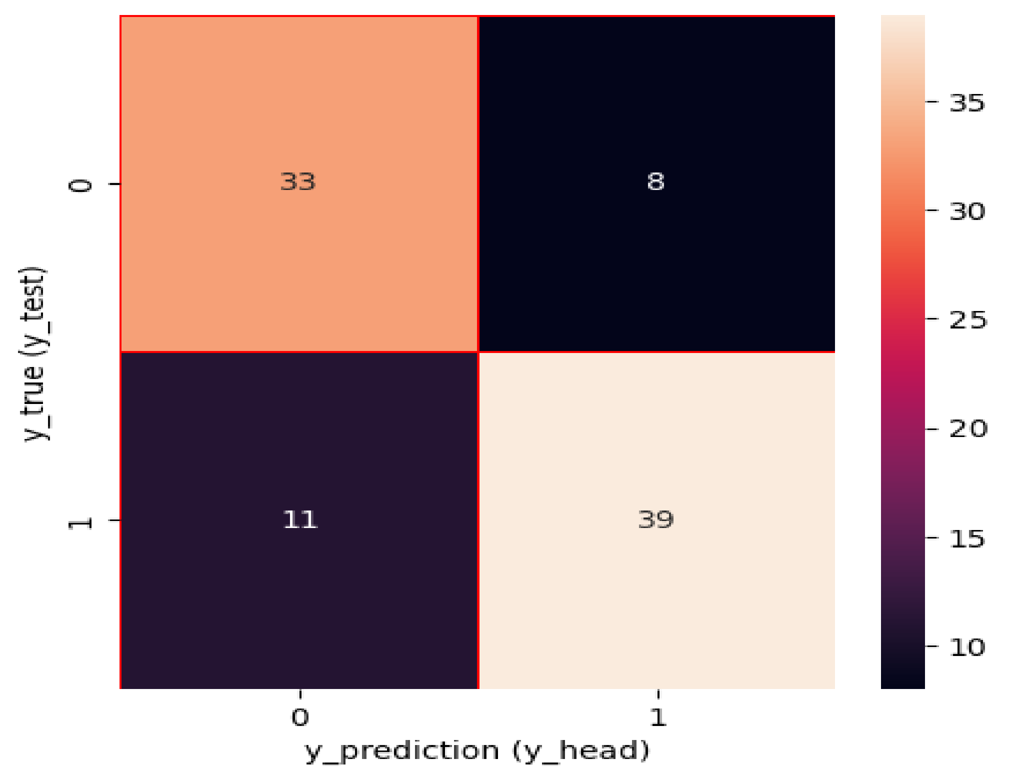
➢ CORRELATION HEATMAP



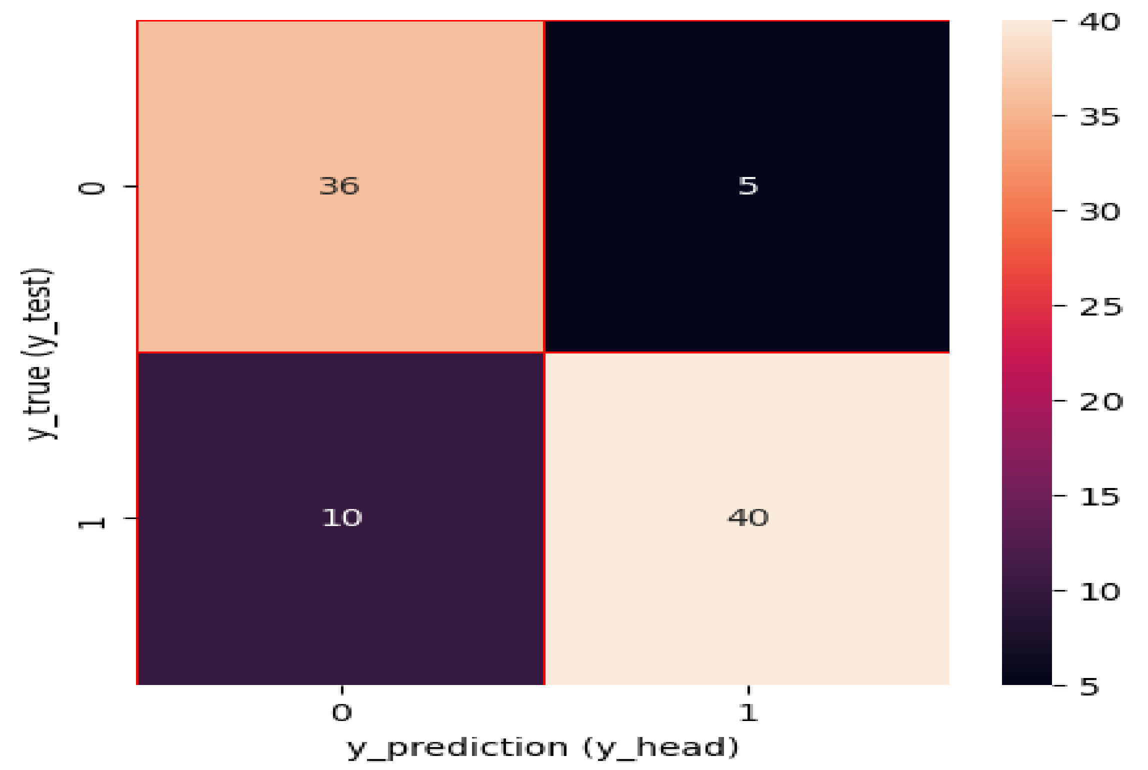
DATA VISUALIZATION

HEATMAP FOR CONFUSION MATRIX:

HEATMAP
CONFUSION MATRIX 1
(LOGISTIC REGRESSION):



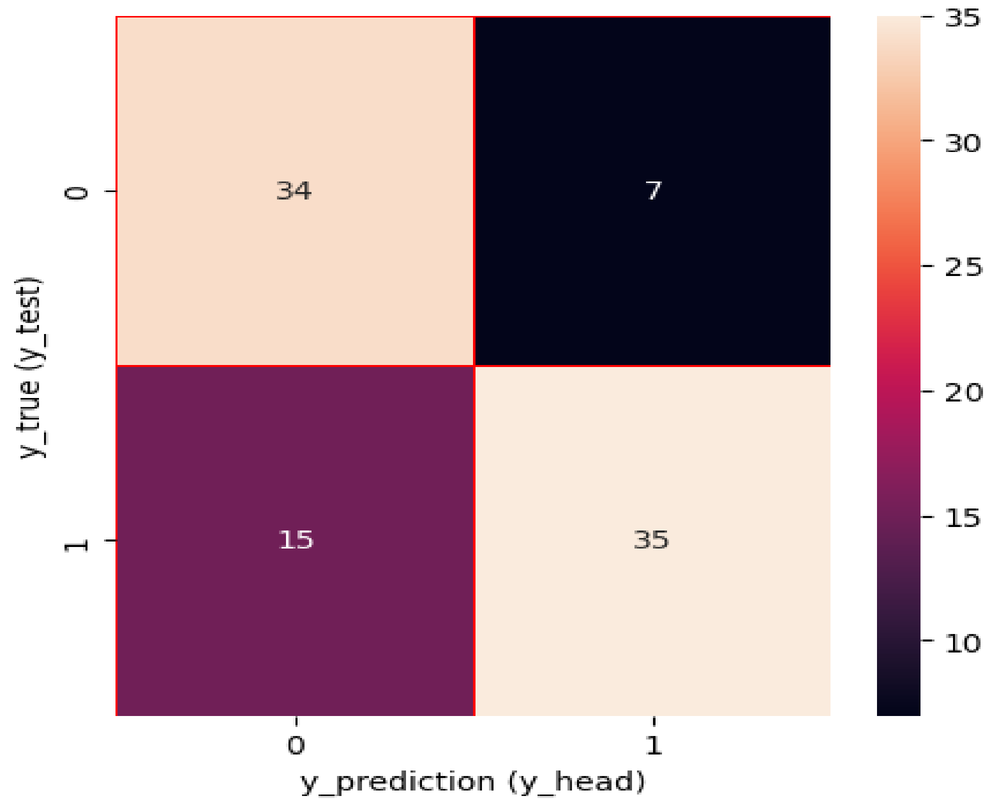
HEATMAP
CONFUSION MATRIX 2
(GAUSSIAN NB):



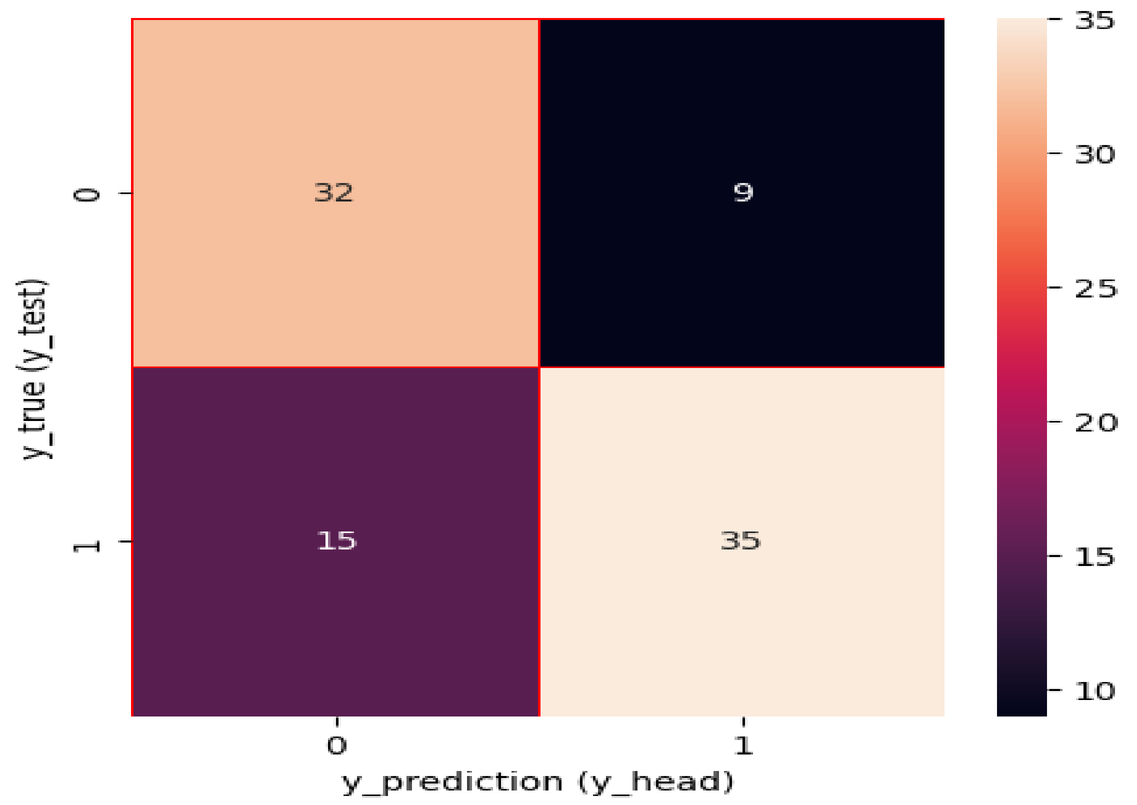
DATA VISUALIZATION

HEATMAP FOR CONFUSION MATRIX(CONT.):

HEATMAP
CONFUSION MATRIX 3
(KNEIGHBOR CLASSIFIER):



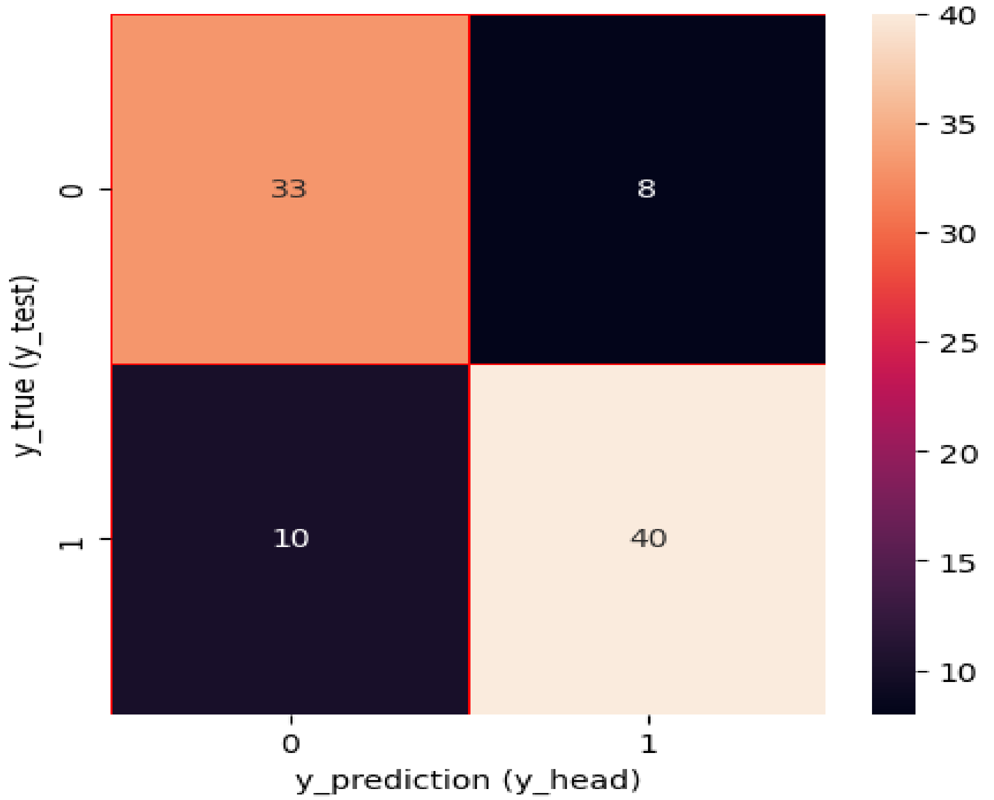
HEATMAP
CONFUSION MATRIX 4
(DECISIONTREE CLASSIFIER):



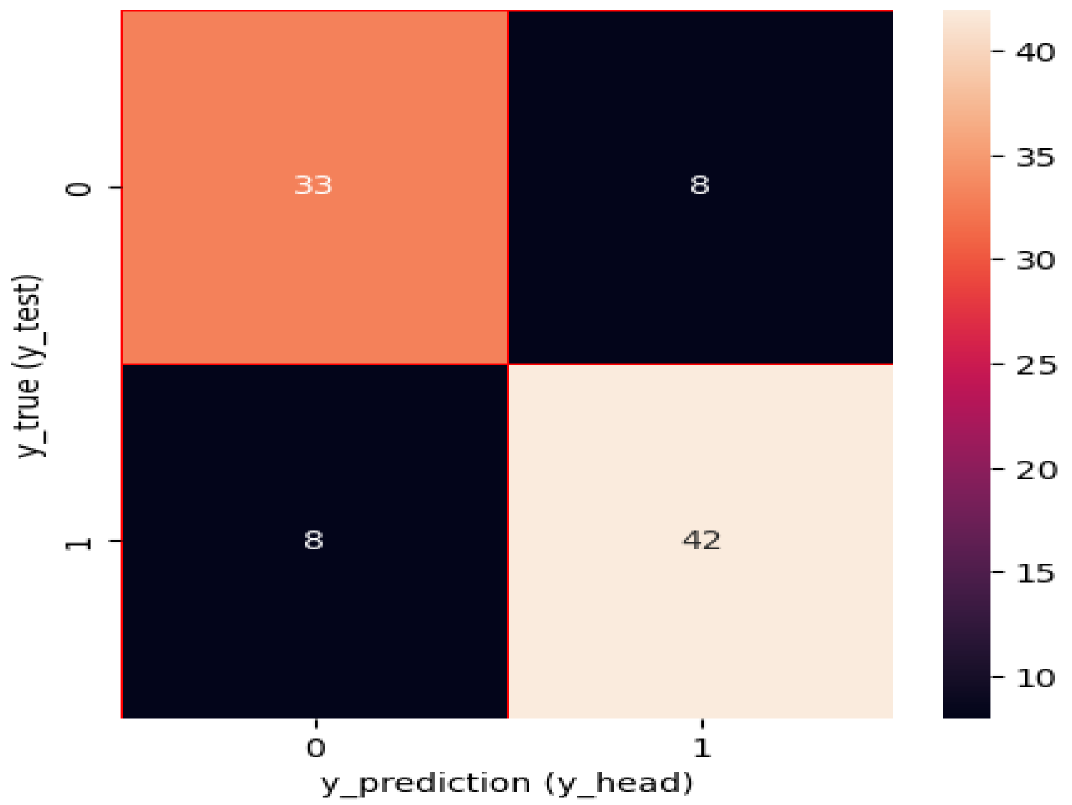
DATA VISUALIZATION

HEATMAP FOR CONFUSION MATRIX(CONT.):

HEATMAP CONFUSION MATRIX 5 (SUPPORT VECTOR MACHINE):



HEATMAP CONFUSION MATRIX 6 (RANDOM FOREST):



CALCULATION OF RECALL VALUES:

Let’s calculate the recall for each of the models based on the accuracy attained by each model:

1.Logistic Regression:

•TP: 0.7912087912087912 (from test accuracy)
•FN: 1 - 0.7912087912087912 = 0.2087912087912088
•Recall = (\frac{0.7912087912087912}{0.7912087912087912 + 0.2087912087912088} = 0.7912087912087912)

2.K-Nearest Neighbors (KNN):

•TP: 0.7582417582417582 (from test accuracy)
•FN: 1 - 0.7582417582417582 = 0.2417582417582418
•Recall = (\frac{0.7582417582417582}{0.7582417582417582 + 0.2417582417582418} = 0.7582417582417582)

3.Support Vector Machine (SVM):

•TP: 0.8021978021978022 (from test accuracy)
•FN: 1 - 0.8021978021978022 = 0.1978021978021978
•Recall = (\frac{0.8021978021978022}{0.8021978021978022 + 0.1978021978021978} = 0.8021978021978022)

4.Naive Bayes:

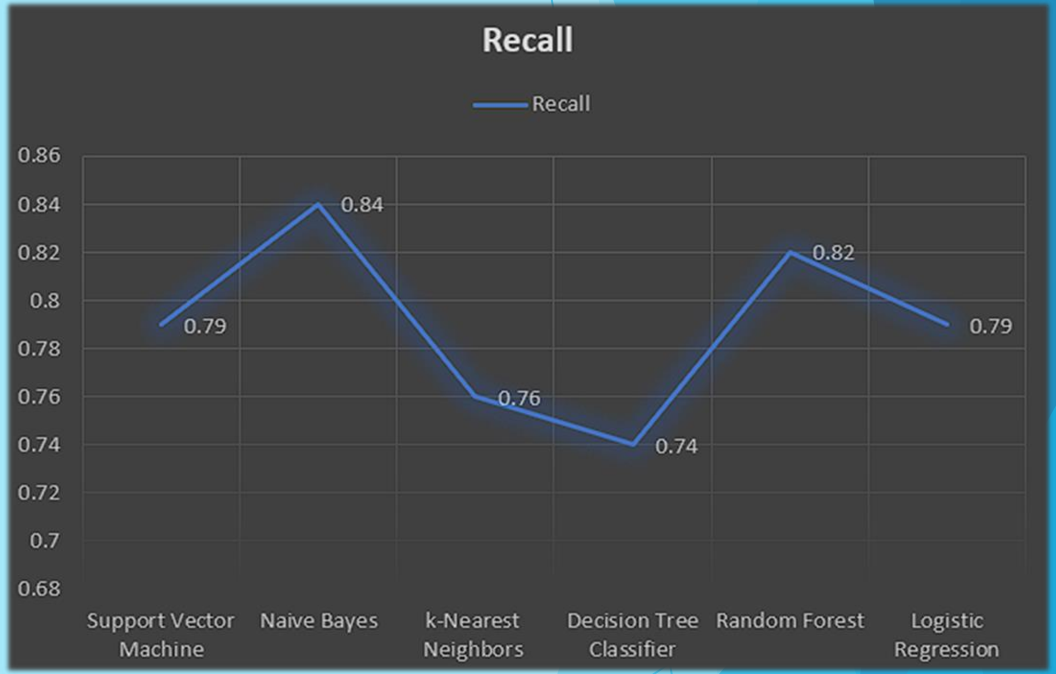
•TP: 0.8351648351648352 (from test accuracy)
•FN: 1 - 0.8351648351648352 = 0.1648351648351648
•Recall = (\frac{0.8351648351648352}{0.8351648351648352 + 0.1648351648351648} = 0.8351648351648352)

5.Decision Tree:

•TP: 0.7362637362637363 (from test accuracy)
•FN: 1 - 0.7362637362637363 = 0.2637362637362637
•Recall = (\frac{0.7362637362637363}{0.7362637362637363 + 0.2637362637362637} = 0.7362637362637363)

6.Random Forest:

•TP: 0.8241758241758241 (from test accuracy)
•FN: 1 - 0.8241758241758241 = 0.1758241758241759
•Recall = (\frac{0.8241758241758241}{0.8241758241758241 + 0.1758241758241759} = 0.8241758241758241)



These recall values indicate how well each model captures positive instances.
Higher recall values are desirable, as they indicate better performance in correctly identifying actual positive cases.

RESULTS:

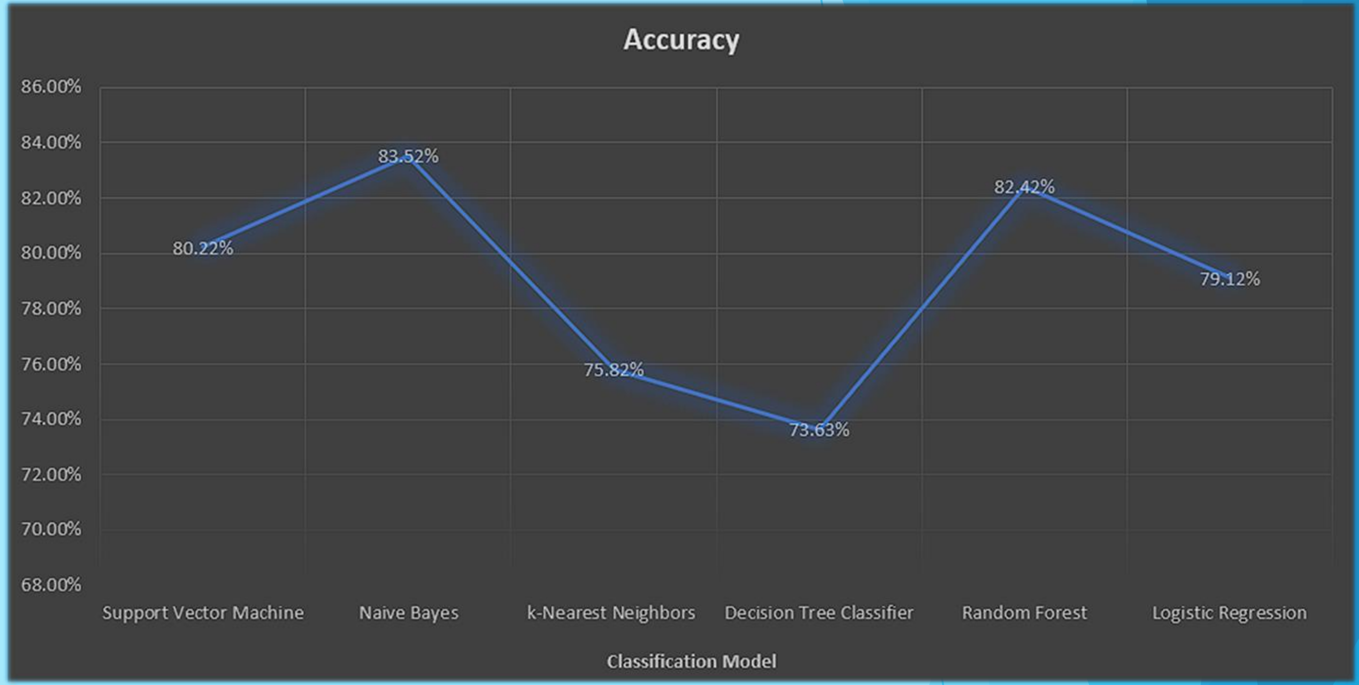
1. Naive Bayes achieved the highest test accuracy of approximately 83.5%, making it a strong contender.

2. Support Vector Machine (SVM) performed well with an accuracy of around 80.2%.

3. Random Forest also delivered a commendable performance, achieving an accuracy of about 82.4%.

4. Logistic Regression and K-Nearest Neighbors (KNN) had accuracies of approximately 79.1% and 75.8%, respectively.

5. However, the Decision Tree model lagged behind, achieving an accuracy of only 73.6%.



- In summary, Naive Bayes stands out as the top performer, closely followed by Random Forest and SVM.