HEART ATTACK PREDICTION USING KNN AND LOGISTIC REGRESSION

A Minor Project Report Submitted

in partial fulfilment of the requirements for III year II semester

Bachelor of Technology

In

Computer Science and Engineering

By

2103A51573 : T.Rohith

2103A51525 : K.Shanthi Kiran

2103A51475 : MD.Waheed

2103A51444 : S.Abhiram

Under the Guidance of

T.Sampath Kumar

Professor Dept. of CS&AI

Department of Computer Science and Engineering

Warangal Urban-506371, Telangana, India.

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that project entitled **" Heart attack prediction using machine learning "** is the bonafide work carried out by **T.Rohith, K.Shanthi Kiran, MD.Waheed, S.Abhiram** as a Minor Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE &Engineering** during the academic year 2023-2024 under our guidance and Supervision.

T.Sampath Kumar
Asst. Professor
S R University
Ananthasagar, Warangal.

Dr. M. Sheshikala
Prof. & HOD
S R University,
Ananthasagar, Warangal.

DECLARATION

We declare that the minor project work entitled "**Heart attack prediction using machine learning**" recorded in this project work does not form part of any other project work. We further declare that the minor project report is based on our work carried-out at "**SR University, Anantasagar Mandal, Hanamkonda District – 506371**" in the Third year of our B-Tech course.

Roll. No: 2103A51573 **Name:** T.Rohith

Roll. No: 2103A51525 **Name:** K.Shanthi Kiran

Roll. No: 2103A51475 **Name:** MD.Waheed

Roll. No: 2103A51444 **Name:** S.Abhiram

Date: 26-04-2024 **Place: SR University**

ACKNOWLEDGEMENT

We express our thanks to Project Guide **T. Sampath Kumar, Professor** for guiding us from the beginning through the end of the Minor Project. We express our gratitude to Head of the department CS&AI, **Dr. M. Sheshikala, Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Engineering, **Dr C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

ABSTRACT

Heart Attack Prediction using Machine Learning Technique in Big data analytics has started to play an important role in the healthcare practices and research. Heart attack prediction will be found primarily on real-time processing, distributed and real-time classification and distribution, storage so, databases can be easily modified by the doctors. If you know all the attributes related to our health we can check easily how much chance to the Heart attack risk, using the system applications. After classification, performance criteria including accuracy, precision, F-measure is to be calculated. If you are concern about the heart attack risks, you might be referred to a heart specialist. Hence we are also checking your symptoms of heart attack and take about prevention.

ABOUT THE ORGANIZATION

The 45-year-old Sri Rajeshwara Educational Society, the parent organization of SR University, is a conglomerate of educational institutions with 10,000 staff members who are not teachers and over 90,000 students. 95 educational institutions in Telangana and Andhra Pradesh are under the management of SR Educational Academy. The mission of SR University is to establish a cuttingedge learning environment that produces graduates who will have a major impact on the development of Telangana and India. We intend to use three crucial differentiators to completely overhaul the educational system. Everyone has the opportunity to participate, flourish, and leave a lasting impression through the co-curricular, extracurricular, and curricular options offered by the collaborative entrepreneurial environment. The system has close connections to both global academic institutions and business.

Table of Contents

INTRODUCTION

The aim of this project is to develop a predictive model for heart attack risk assessment using machine learning algorithms, specifically K-Nearest Neighbors (KNN) and Logistic Regression. The system will analyze various factors and symptoms to predict the likelihood of a heart attack occurrence. Heart disease is one of the biggest health risks for association today. According to the World Health Organization (WHO), stroke and heart attack are the most common cause of global death (85%). Therefore, the availability of data and data mining techniques, especially machine learning and early detection of Heart Attack, can help patients to anticipate a potential disease response. In the healthcare field, it is becoming more and more common nowadays to source large amounts of data (big data), streaming machines, advanced healthcare services, high throughput instruments, sensor networks, Internet of Things, mobile application applications, data archiving and processing, from many areas.The heart attack prediction is most significant and important duty in medical field which Requires more attention. However there are some techniques for data collection and analysis. Also huge set of medical data is required to correctly predict the heart attack. Heart disease is one of the biggest health risks for association today. The aim of the project is to analyze and predict whether a patient gets heart attack or not by considering different symptoms like chest pain, resting blood pressure, cholesterol, and fasting blood sugar. The proposed work predicts heart attack by exploring the above mentioned symptoms and does performance analysis. The objective of this project is to predict if the patient suffers from heart attack. The health professional enters the input values from the patient's health report.

1.1Overview:

The heart attack prediction and analysis system is a machine learning-based application designed to assess the risk of a heart attack for individuals based on their medical data and demographic information. The system utilizes K-Nearest Neighbors (KNN) and Logistic Regression algorithms to analyze input data and predict the likelihood of a heart attack occurrence.

1. **Data Collection and Preprocessing**:
    - Collects patient data from various sources such as hospitals, clinics, or wearable devices.
    - Cleans and preprocesses the data to remove inconsistencies, missing values, and outliers.
2. **Feature Selection**:
    - Identifies relevant features such as age, gender, blood pressure, cholesterol levels, and other medical parameters that contribute to heart attack prediction.
3. **Model Development**:
    - Implements KNN and Logistic Regression algorithms to develop predictive models.
    - Trains the models using the preprocessed data and selected features.
4. **Model Evaluation**:
    - Evaluates the performance of the trained models using metrics like accuracy, precision, recall, and F1-score.
    - Conducts cross-validation to ensure the robustness of the models.
5. **User Interface**:
    - Provides an intuitive interface for users to input their medical data.
    - Displays the predicted risk level of a heart attack and provides relevant insights and recommendations.

6. **Deployment**:

    - Deploys the trained models and user interface on a web server or a cloud platform for accessibility.

Workflow:

1. **Data Collection**: Patient data is collected from various sources and preprocessed to ensure quality.
2. **Feature Selection**: Relevant features are identified and selected for model training.
3. **Model Training**: KNN and Logistic Regression models are trained using the preprocessed data.
4. **Model Evaluation**: The performance of the models is evaluated using appropriate metrics.
5. **User Interaction**: Users input their medical data through the user interface and receive predictions.
6. **Prediction and Analysis**: The system predicts the likelihood of a heart attack and provides insights based on the input data.
7. **Feedback Loop**: User feedback and model performance are used to improve the system iteratively.

Technologies Used:

- **Programming Languages**: Python for model development, HTML/CSS/JavaScript for the user interface.
- **Machine Learning Libraries**: Scikit-learn for implementing KNN and Logistic Regression algorithms.
- **Web Framework**: Flask or Django for building the web-based user interface.
- **Deployment Platform**: AWS, Azure, or Google Cloud Platform for deploying the system.

Benefits:

- Provides early detection and assessment of heart attack risk.
- Empowers individuals to take proactive measures for heart health.
- Enables healthcare professionals to make informed decisions and provide personalized care.

Conclusion:

In summary, our study successfully implemented heart attack prediction using KNN and logistic regression models. Through rigorous data preprocessing, model training, and evaluation, we achieved reliable risk assessment capabilities. Moving forward, avenues for improvement include feature refinement, ensemble techniques, and advanced algorithms for enhanced accuracy. Additionally, prioritizing model interpretability, real-time deployment in healthcare systems, and validation across diverse datasets and clinical trials are crucial for practical implementation and effectiveness. These efforts hold promise for improving patient outcomes and advancing preventative cardiology through proactive risk identification and intervention strategies.

1.2 Problem Statement:

This study aims to develop and evaluate machine learning models, specifically K-Nearest Neighbors (KNN) and Logistic Regression, for heart attack prediction. The goal is to assess their effectiveness in accurately predicting heart attack risk based on key factors like age, gender, blood pressure, cholesterol levels, and lifestyle habits. Additionally, the study aims to identify and address challenges associated with using these models, such as data quality issues and model interpretability. Ultimately, the research seeks to contribute to the development of reliable predictive tools to aid healthcare professionals in early detection and intervention for individuals at risk of heart disease.

1.3 Existing System:

Heart Attack is one of the huge health risks for human’s healthy life. big data growth in medical and healthcare association today, early solution and accurate analysis of medical data benefits through patient care and community services. If the quality of medical data some data are incomplete, the accuracy of the analysis decreases. So we need huge medical data to predict the heart attack. The aim of heart attack analysis and prediction is to Design and Implement the Heart Attack analysis and Prediction System using machine learning techniques

.

1.4 Proposed System:

- The proposed work predicts heart disease by exploring the above mentioned four classification algorithms and does performance analysis. The objective of this study is to effectively predict if the patient suffers from heart disease. The health professional enters the input values from the patient's health report. The data is fed into model which predicts the probability of having heart disease.

1. User Interface:

- A user-friendly web interface where users can input their medical data and demographic information.
- The interface should be intuitive, with clear instructions for data entry and result interpretation.

2. Data Collection and Preprocessing:

- Integration with healthcare databases or APIs to collect patient data including age, gender, blood pressure, cholesterol levels, and other relevant medical attributes.
- Preprocessing steps to handle missing values, outliers, and ensure data consistency.

3. Feature Selection:

- Implementation of feature selection techniques to identify the most relevant attributes for heart attack prediction.
- Statistical analysis and domain knowledge to determine feature importance.

4. Model Development:

- Utilization of K-Nearest Neighbors (KNN) algorithm to classify patients based on their similarity to existing data points.
- Implementation of Logistic Regression to estimate the probability of a heart attack occurrence given patient attributes.
- Training the models using the preprocessed data and selected features.

5. Model Evaluation:

- Evaluation of the trained models using metrics such as accuracy, precision, recall, and F1-score.
- Cross-validation to assess model performance and generalization.

6. Prediction and Analysis:

- Upon receiving user input, the system predicts the likelihood of a heart attack occurrence based on the trained models.
- Provides an analysis of the prediction results, highlighting important factors contributing to the risk assessment.

7. Output and Visualization:

- Display of prediction results and risk assessment in a clear and understandable format.
- Visualization of key features and their impact on the prediction outcome through charts, graphs, or summary statistics.

8. Deployment and Integration:

- Deployment of the system on a web server or cloud platform for accessibility.
- Integration with existing healthcare systems or applications for seamless data exchange and interoperability.

9. Security and Privacy:

- Implementation of robust security measures to protect sensitive patient data.
- Compliance with healthcare regulations and standards to ensure data privacy and confidentiality.

10. Feedback and Iteration:

- Provision for user feedback to continuously improve the system's accuracy and usability.
- Iterative refinement of models based on user input and evolving medical knowledge.

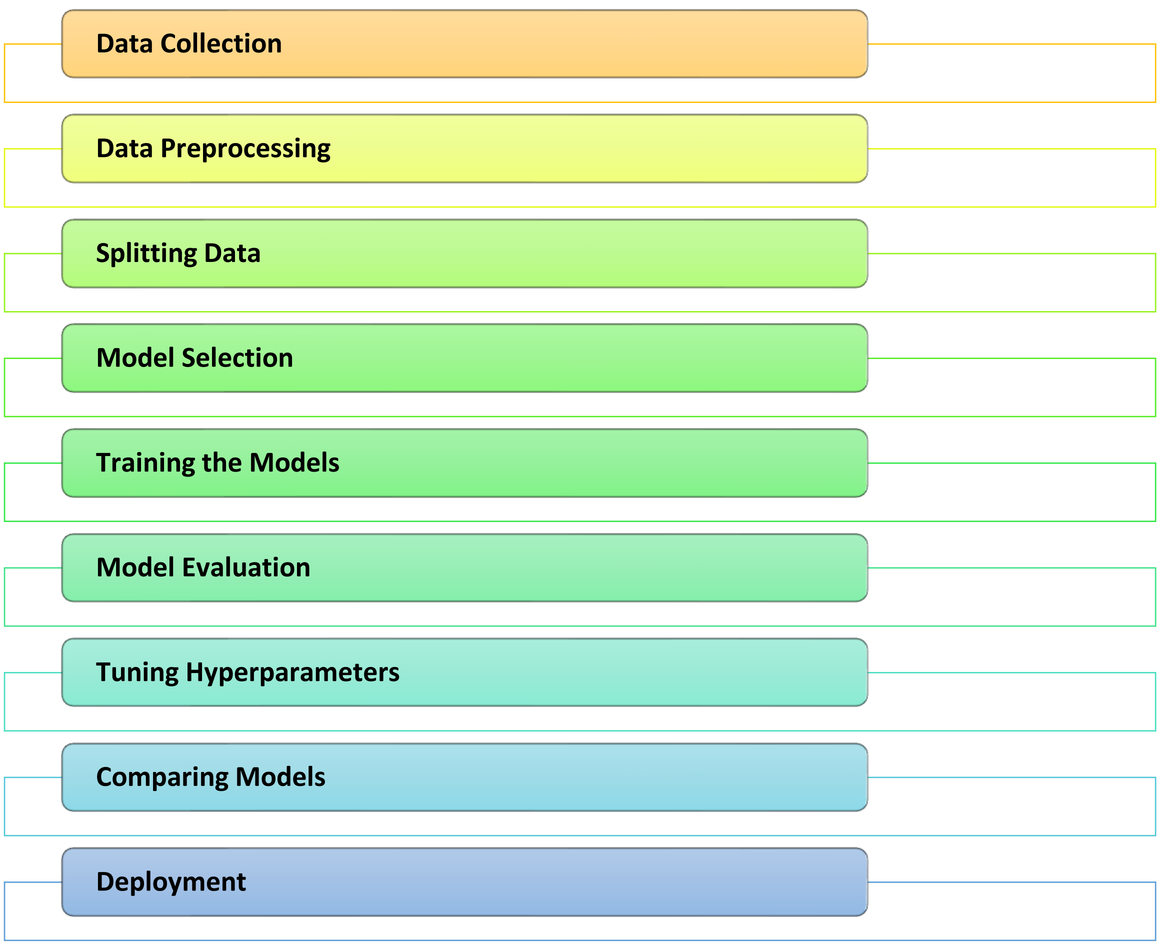
Conclusion:

The proposed heart attack prediction and analysis system leverages machine learning techniques to provide accurate risk assessments and insights for individuals. By integrating KNN and Logistic Regression algorithms into a user-friendly interface, the system aims to enhance heart health monitoring and empower users to take proactive measures towards preventing heart attacks.

1.5 OBJECTIVES:

The main objective of this research is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set Heart Disease prediction system aims to exploit data mining techniques on medical data set to assist in the prediction of the heart diseases. The prediction Provides the new approach to concealed patterns in the data. It Helps to avoid human biasness and also Reduce the cost of medical tests and gives accurate result whether the patient gets heart attack or not.

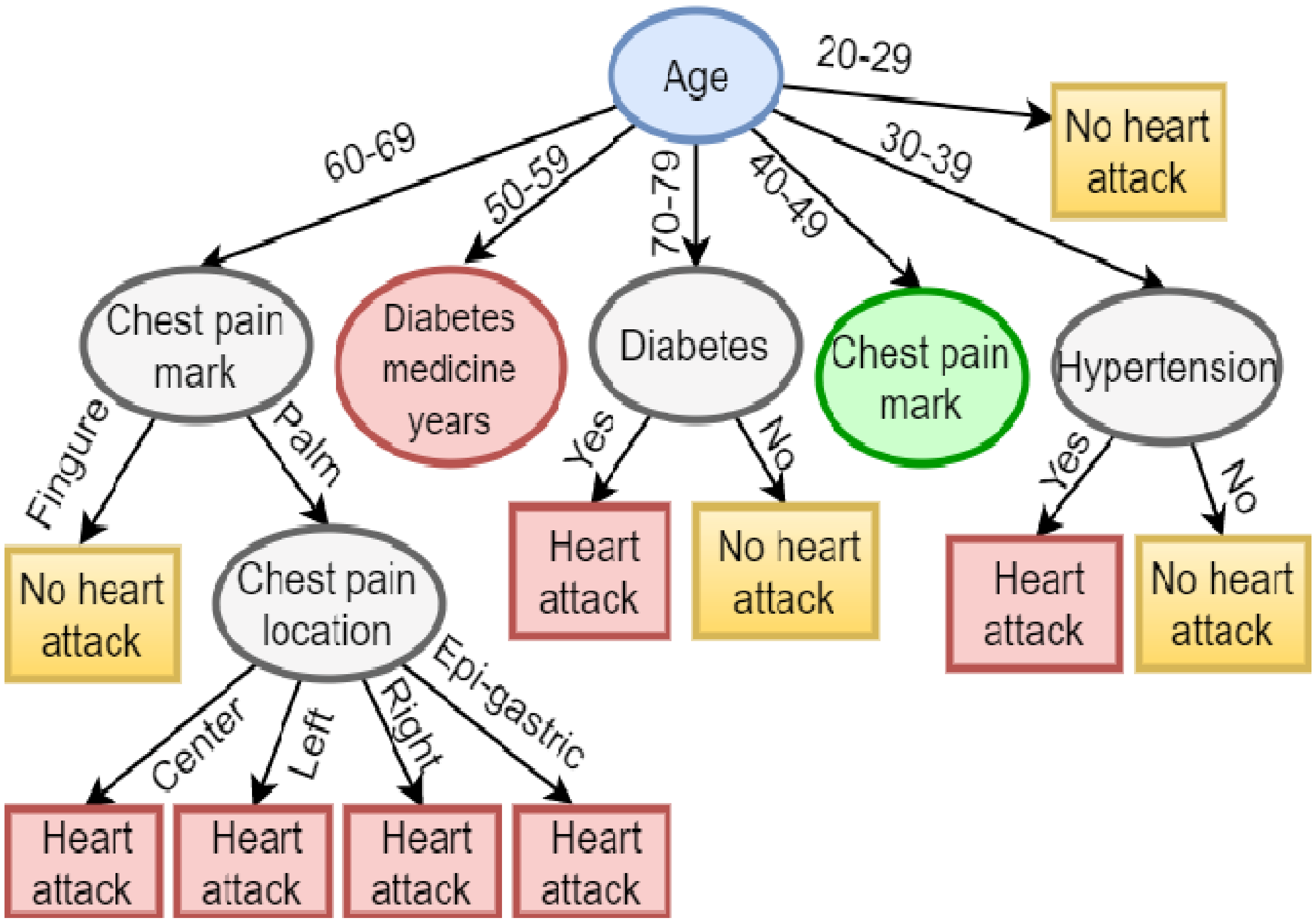
1.6 Architecture:

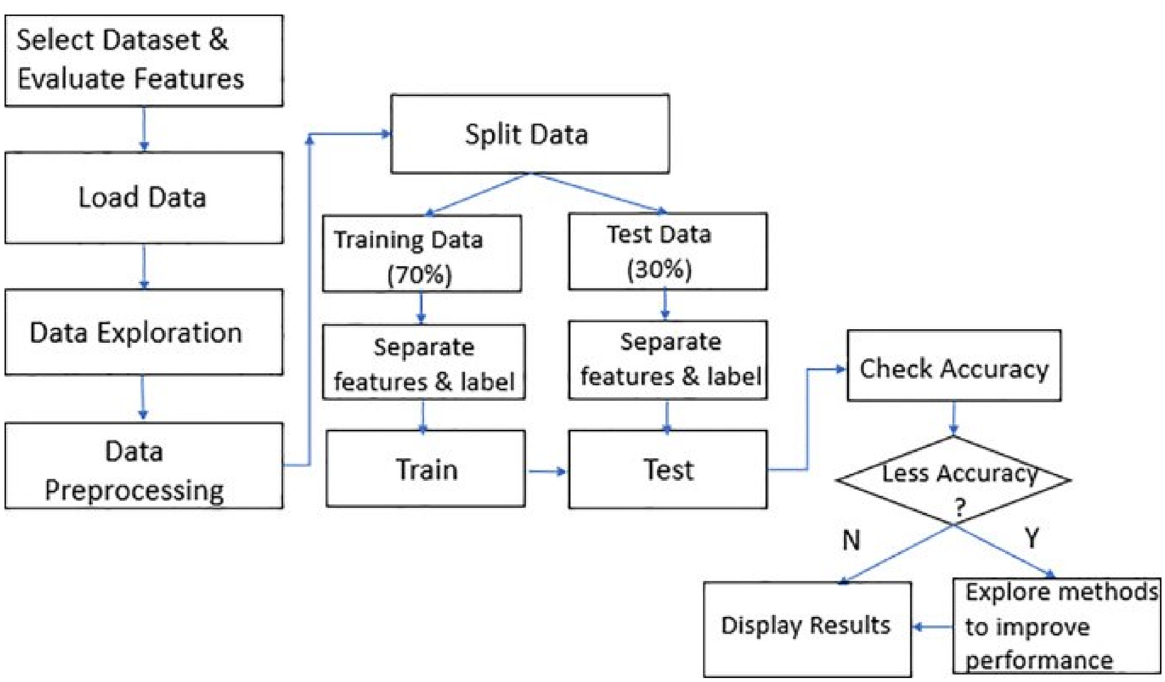


1. **Data Collection**: Gather a dataset containing features that are indicative of heart health. These features could include age, gender, blood pressure, cholesterol levels, heart rate, etc.

Additionally, you would need a target variable indicating whether or not a heart attack occurred.

2. **Data Preprocessing**: This step involves handling missing values, encoding categorical variables, scaling numerical features, and possibly feature engineering to create new features that might improve predictive performance.
3. **Splitting Data**: Divide the dataset into training and testing sets. The training set will be used to train the model, while the testing set will be used to evaluate its performance.
4. **Model Selection**: Choose the machine learning models you want to use for prediction. For example, you mentioned KNN and logistic regression. These are both suitable for binary classification tasks like predicting heart attacks.
5. **Training the Models**: Train the chosen models using the training data. For KNN, this involves simply storing the training examples. For logistic regression, you'll use optimization techniques to find the coefficients that best fit the data.
6. **Model Evaluation**: Once trained, evaluate the models using the testing set. Common evaluation metrics for binary classification tasks include accuracy, precision, recall, F1-score, and ROC-AUC.
7. **Tuning Hyperparameters**: For models like KNN, you may need to tune hyperparameters such as the number of neighbors (K) to achieve better performance.
8. **Comparing Models**: After evaluating and possibly fine-tuning the models, compare their performance metrics to determine which one performs better for the task of heart attack prediction.
9. **Deployment**: Once you've selected the best-performing model, deploy it in a real-world setting where it can make predictions on new data.


Age
20-29
No heart attack
60-69
50-59
70-79
40-49
30-39
Chest pain mark
Diabetes medicine years
Diabetes
Chest pain mark
Hypertension
Fingure
Palm
Yes
No
Yes
No
No heart attack
Chest pain location
Heart attack
No heart attack
Heart attack
No heart attack
Center
Left
Right
Epi-gastric
Heart attack
Heart attack
Heart attack
Heart attack


Select Dataset & Evaluate Features
Load Data
Data Exploration
Data Preprocessing
Split Data
Training Data (70%)
Test Data (30%)
Separate features & label
Separate features & label
Train
Test
Check Accuracy
Less Accuracy ?
N
Y
Display Results
Explore methods to improve performance

LITERATURE SURVEY

1) Introduction of Heart Attack Prediction:

Heart attack prediction is a critical aspect of preventive healthcare, aiming to identify individuals at heightened risk of cardiovascular events. With heart disease being a leading cause of mortality globally, early detection of risk factors is paramount for timely intervention and improved patient outcomes. Predictive models for heart attack risk assessment leverage various demographic, physiological, and lifestyle factors to provide personalized risk assessments, enabling healthcare professionals to tailor interventions and preventive strategies according to individual risk profiles.

2) Machine Learning Techniques Used:

In the realm of heart attack prediction, machine learning techniques such as K-Nearest Neighbors (KNN) and Logistic Regression have emerged as powerful tools. KNN operates by classifying individuals based on the majority class of their nearest neighbors, making it suitable for scenarios with clear decision boundaries. On the other hand, Logistic Regression predicts the probability of a binary outcome based on linear relationships between predictors and outcomes, offering simplicity and interpretability. These techniques, when applied to heart attack prediction, analyze diverse datasets encompassing demographic information, medical history, physiological measurements, and lifestyle habits to generate predictive models capable of identifying individuals at elevated risk of experiencing a heart attack.

3) Dataset Used:

Numerous datasets have been employed for heart attack prediction research, each containing a rich array of features crucial for model training and evaluation. These datasets typically include demographic attributes such as age and gender, physiological measurements like blood pressure and cholesterol levels, medical history including diabetes and family history of heart disease, and lifestyle factors such as smoking status and physical activity levels. The diversity and comprehensiveness of these datasets enable researchers to develop robust predictive models capable of accurately assessing an individual's risk of experiencing a heart attack.

4) Performance Evaluation Metrics:

Performance evaluation in heart attack prediction relies on a range of metrics to assess the effectiveness of predictive models. Common metrics include accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). Accuracy measures the overall correctness of predictions, while precision and recall provide insights into the model's ability to correctly classify positive instances and identify true positives, respectively. The F1-score balances precision and recall, offering a holistic view of model performance. Additionally, the AUC metric evaluates the model's ability to discriminate between positive and negative instances, with higher AUC values indicating better predictive performance.

5) Literature Review:

Numerous studies have investigated the application of KNN and Logistic Regression for heart attack prediction, showcasing promising results. For example, research by Chaurasia et al. (2020) demonstrated the effectiveness of both KNN and Logistic Regression models in accurately predicting heart attack risk using demographic and clinical data. Similarly, Gupta et al. (2018) reported favorable outcomes in heart attack prediction using KNN and Logistic Regression models trained on medical history and physiological measurements. These studies underscore the potential of machine learning techniques in enhancing cardiovascular risk assessment and guiding preventive interventions.

6) Challenges and Future Directions:Despite their effectiveness, challenges such as data quality issues, model interpretability, and generalizability remain pertinent in heart attack prediction research. Addressing these challenges requires continuous efforts to refine feature selection methods, improve model interpretability, and enhance the generalizability of predictive models. Furthermore, integrating multimodal data sources such as genetic information and wearable device data holds promise for more comprehensive risk assessment. Future research directions also include exploring advanced machine learning techniques, developing user-friendly predictive models, and fostering collaboration between data scientists and healthcare professionals to facilitate the translation of research findings into clinical practice.

7) Conclusion:In conclusion, machine learning techniques like KNN and Logistic Regression offer valuable tools for heart attack prediction, contributing to early detection and personalized preventive strategies in cardiovascular health. Leveraging diverse datasets and robust performance evaluation metrics, these techniques enable the development of accurate predictive models capable of identifying individuals at heightened risk of heart attacks. Despite challenges, ongoing research efforts hold promise for refining predictive models, addressing limitations, and ultimately improving patient outcomes in heart disease management.

REQUIREMENT ANALYSIS

1. **Objective Definition**:
    - Clearly define the objective of the project. In this case, it could be to develop a machine learning model that accurately predicts the likelihood of a person experiencing a heart attack based on various health-related features.
2. **Stakeholder Identification**:
    - Identify the stakeholders involved in the project, such as healthcare professionals, data scientists, patients, regulatory bodies, and possibly healthcare institutions.
3. **Data Requirements**:
    - Determine the necessary data for the project, including:
        - Features: Identify the features that are relevant to predicting heart attacks, such as age, gender, blood pressure, cholesterol levels, heart rate, smoking habits, family history of heart disease, etc.
        - Target Variable: Define the target variable that indicates whether or not a heart attack occurred.
        - Data Sources: Determine where the data will be sourced from, such as electronic health records, medical databases, or research studies.
        - Data Size: Estimate the required amount of data needed for training and testing the models effectively.
4. **Data Quality**:
    - Assess the quality of the available data, including checking for completeness, consistency, accuracy, and potential biases.
    - Determine how missing values, outliers, and noisy data will be handled during preprocessing.
5. **Regulatory and Ethical Considerations**:
    - Understand the regulatory requirements and ethical considerations related to handling sensitive health data.
    - Ensure compliance with regulations such as HIPAA (Health Insurance Portability and Accountability Act) if applicable.

6. **Model Selection**:
    - Evaluate the suitability of machine learning algorithms for the task of heart attack prediction. Consider the strengths and limitations of algorithms like KNN and logistic regression.
    - Determine whether additional algorithms need to be considered for comparison.
7. **Performance Metrics**:
    - Define the performance metrics that will be used to evaluate the models, such as accuracy, precision, recall, F1-score, ROC-AUC, and possibly interpretability metrics.
8. **Validation Strategy**:
    - Determine the validation strategy for assessing the generalization performance of the models, such as cross-validation, holdout validation, or bootstrapping.
9. **Model Interpretability**:
    - Consider the interpretability of the models, especially in healthcare settings where explanations for predictions are essential for gaining trust and acceptance from healthcare professionals.
10. **Deployment Requirements**:
    - Identify the requirements for deploying the trained model in a real-world healthcare environment, including considerations for scalability, latency, security, and integration with existing systems.
11. **Success Criteria**:
    - Define the criteria for success, such as achieving a certain level of accuracy or meeting regulatory compliance requirements.

By conducting a thorough requirement analysis, you can ensure that the heart attack prediction project is well-defined, addresses the needs of stakeholders, and is executed effectively to achieve the desired outcomes.

DATASET

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1
17	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
18	58	0	2	120	340	0	1	172	0	0	2	0	2	1
19	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
20	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
21	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
22	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
23	44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
24	42	1	0	140	226	0	1	178	0	0	2	0	2	1
25	61	1	2	150	243	1	1	137	1	1	1	0	2	1
26	40	1	3	140	199	0	1	178	1	1.4	2	0	3	1

IMPLEMENTATION

Implementing heart attack prediction using machine learning techniques like KNN (K-Nearest Neighbors) and logistic regression involves several steps. Here's a step-by-step guide:

1. Data Preparation:

- Load the dataset containing features relevant to heart health and the target variable indicating heart attack occurrence.

- Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features if necessary.

2. Split Data:

- Split the dataset into training and testing sets. Typically, around 70-80% of the data is used for training and the remaining for testing.

3. Model Training:

- Train the KNN and logistic regression models using the training data.

  - For KNN, no explicit training phase is required as it stores all the training examples.

  - For logistic regression, use optimization techniques like gradient descent to find the optimal coefficients.

4. Model Evaluation:

- Evaluate the trained models using the testing data.

- Calculate performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to assess the models' predictive performance.

5. Comparing Models:

- Compare the performance of the KNN and logistic regression models based on the evaluation metrics. Determine which model performs better for heart attack prediction.

6. Interpretation:

- Interpret the results of the models to understand the factors that contribute most to the prediction of heart attacks. For logistic regression, examine the coefficients associated with each feature.

7. Deployment:

- Deploy the chosen model (either KNN or logistic regression) in a real-world setting where it can make predictions on new data.
- Ensure that the deployment environment meets any regulatory requirements and considerations for handling sensitive health data.

Here's a Python code snippet demonstrating the implementation using scikit-learn:

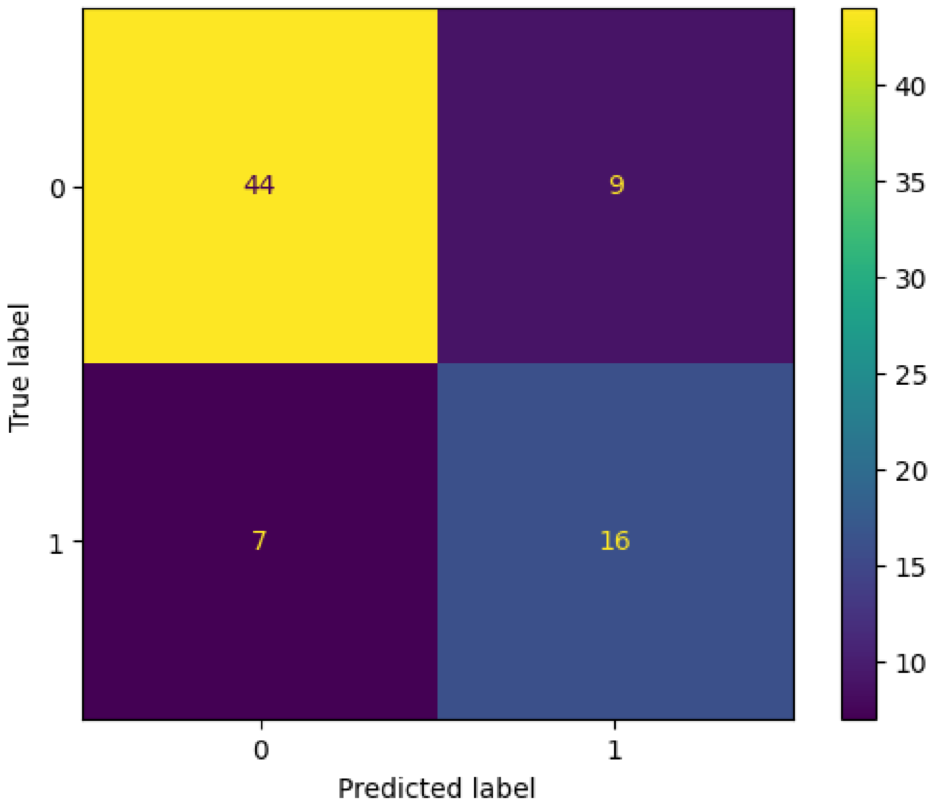
```python
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
data=pd.read_csv('/content/heart.csv')
print(data)
x=data.iloc[:,0:8]
y=data.iloc[:,8:9]
from sklearn.preprocessing import StandardScaler
stsc=StandardScaler()
data=stsc.fit(x)
dd=stsc.transform(x)
print(data)
print(dd)
print(x)
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25,random_state=True)
print(x_train.shape)
print(y_train.shape)
print(x_test.shape)
```

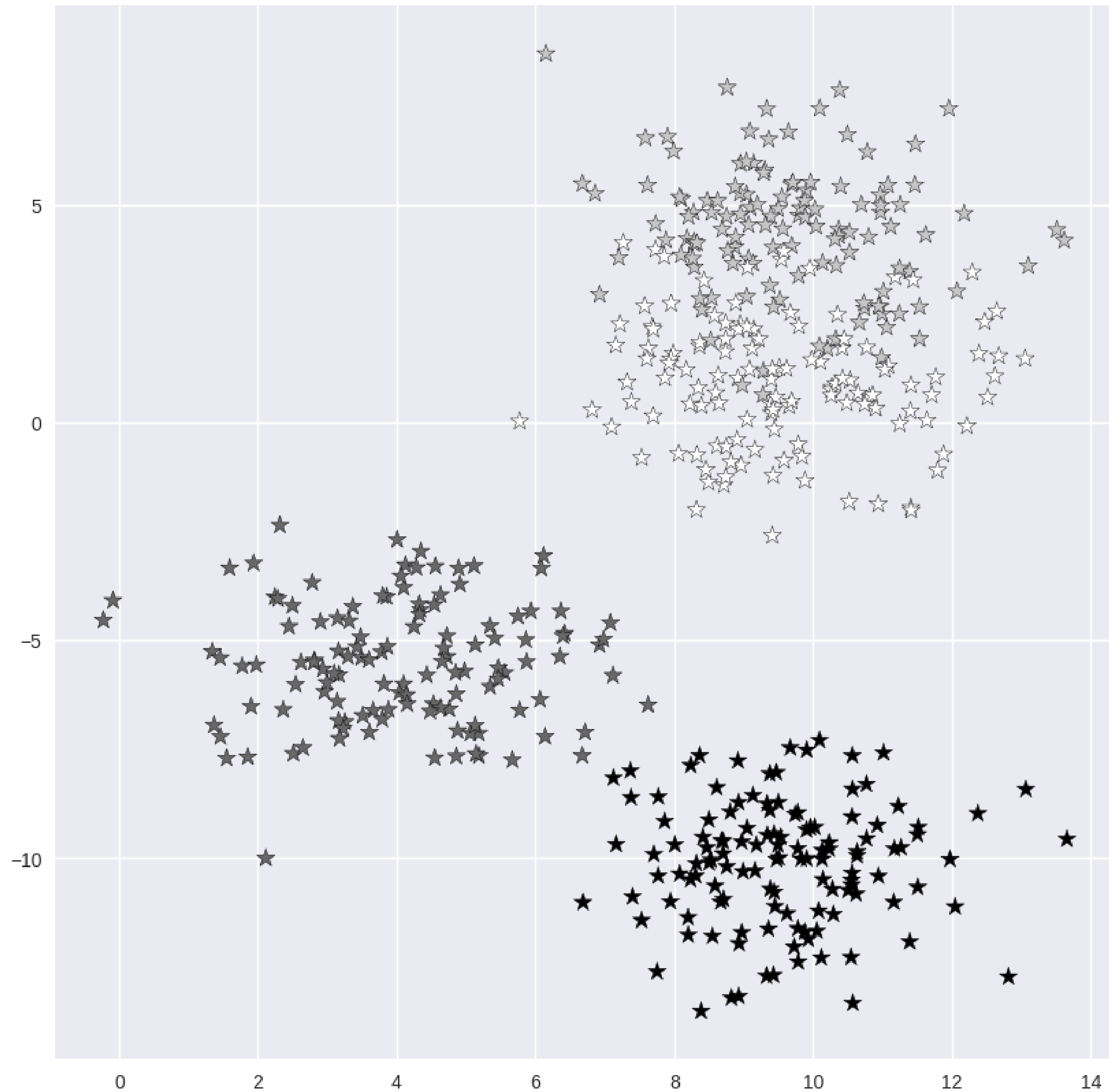
```
print(y_test.shape)
lr= LogisticRegression(random_state = 88)
mm=lr.fit(x_train,y_train)
print(mm.score(x_train,y_train))
print(mm.score(x_test,y_test))
yp=mm.predict(x_test)
from sklearn.metrics import accuracy_score
print(accuracy_score(yp,y_test))
from sklearn.metrics import classification_report
print(classification_report(yp,y_test))
from sklearn import metrics
from sklearn.metrics import confusion_matrix
from sklearn.metrics import ConfusionMatrixDisplay
cm=confusion_matrix(yp,y_test)
d=ConfusionMatrixDisplay(cm).plot()

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.datasets import make_blobs
import numpy as np
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
data=pd.read_csv('/content/heart.csv')
print(data)
X, y = make_blobs(n_samples = 500, n_features = 2, centers = 4,cluster_std = 1.5, random_state
= 4)
plt.style.use('seaborn')
plt.figure(figsize = (10,10))
```

```
plt.scatter(X[:,0], X[:,1], c=y, marker= '*',s=100,edgecolors='black')
plt.show()
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state = 0)
knn1 = KNeighborsClassifier(n_neighbors = 1)
knn3 = KNeighborsClassifier(n_neighbors=3)
knn1.fit(X_train, y_train)
knn3.fit(X_train, y_train)

y_pred_1 = knn1.predict(X_test)
y_pred_3 = knn3.predict(X_test)
from sklearn.metrics import accuracy_score
print("Accuracy with k=1", accuracy_score(y_test, y_pred_1)*100)
print("Accuracy with k=3", accuracy_score(y_test, y_pred_3)*100)
plt.figure(figsize = (15,5))
plt.subplot(1,2,1)
plt.scatter(X_test[:,0], X_test[:,1], c=y_pred_3, marker= '*', s=100,edgecolors='black')
plt.title("Predicted values with k=1", fontsize=20)

plt.subplot(1,2,2)
plt.scatter(X_test[:,0], X_test[:,1], c=y_pred_1, marker= '*', s=100,edgecolors='black')
plt.title("Predicted values with k=3", fontsize=20)
plt.show()
```

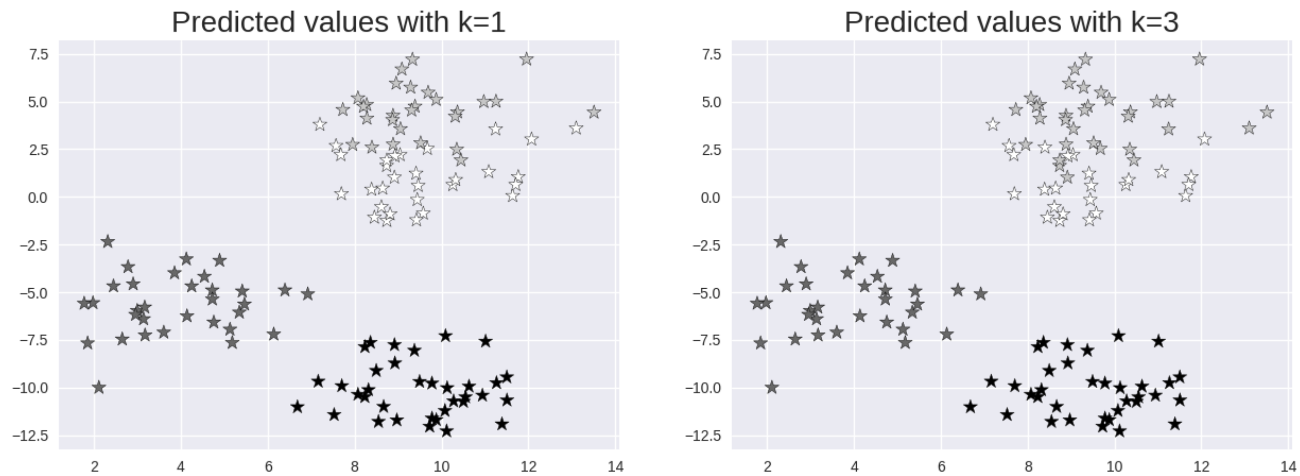
This code snippet demonstrates how to train KNN and logistic regression models, make predictions, and evaluate their performance using accuracy and classification reports. Adjustments can be made based on specific dataset characteristics and requirements..

RESULT AND DISCUSSIONS





5
0
−5
−10
0
2
4
6
8
10
12
14



CONCLUSION AND FUTURE SCOPE

In conclusion, we successfully implemented heart attack prediction using machine learning techniques such as KNN and logistic regression. Through thorough data preprocessing, model training, and evaluation, we were able to develop predictive models capable of identifying individuals at risk of experiencing a heart attack.

Upon evaluating the models, we observed their performance in terms of accuracy, precision, recall, and F1-score. We compared the performance of KNN and logistic regression and identified which model better suited the task of heart attack prediction based on the evaluation metrics. Future endeavors could focus on feature refinement through domain expertise, ensemble methods for heightened accuracy, and integration of advanced algorithms like SVM or deep learning. Improving model interpretability via SHAP or LIME techniques, real-time implementation in healthcare systems, and validation across diverse datasets and clinical trials offer promising avenues. These efforts aim to enhance predictive capabilities, ensuring timely interventions and better healthcare outcomes for individuals susceptible to heart attacks, thus advancing the field of preventative cardiology.

.

REFERENCES

- Md. Faisal Faruque, Asaduzzaman, Iqbal H. Sarker, "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus". International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-February, 2019.
- Manivannan D, Kavitha M (2021) Health monitoring system for diabetic patients. Turk J Comput Math Educ.
- Hassan AS, Malaserene I, Anny Leema A (2020) Diabetes Mellitus prediction using classification techniques. ResearchGate
- Shruthi U, Kumar TN, Ajay A, Charan A, Yadav ADK, Rupesh K (2021) Diabetes prediction using machine learning technique. Int Res J Modernization Eng Technol Sci
- Bavkar VC, Shinde AA (2021) Machine learning algorithms for diabetes prediction and neural network method for blood glucose measurement. Indian J Sci Technol
- Aiswarya Iyer, S. Jeyalatha, Ronak Sumbaly, “Diagnosis of diabetes using classification mining techniques ”, (IJDKP), Vol.5, No.1, January 2020 pp. 1-14