

Heart Attack Prediction | A Data-Driven Approach

MILESTONE 2 – PROJECT 2

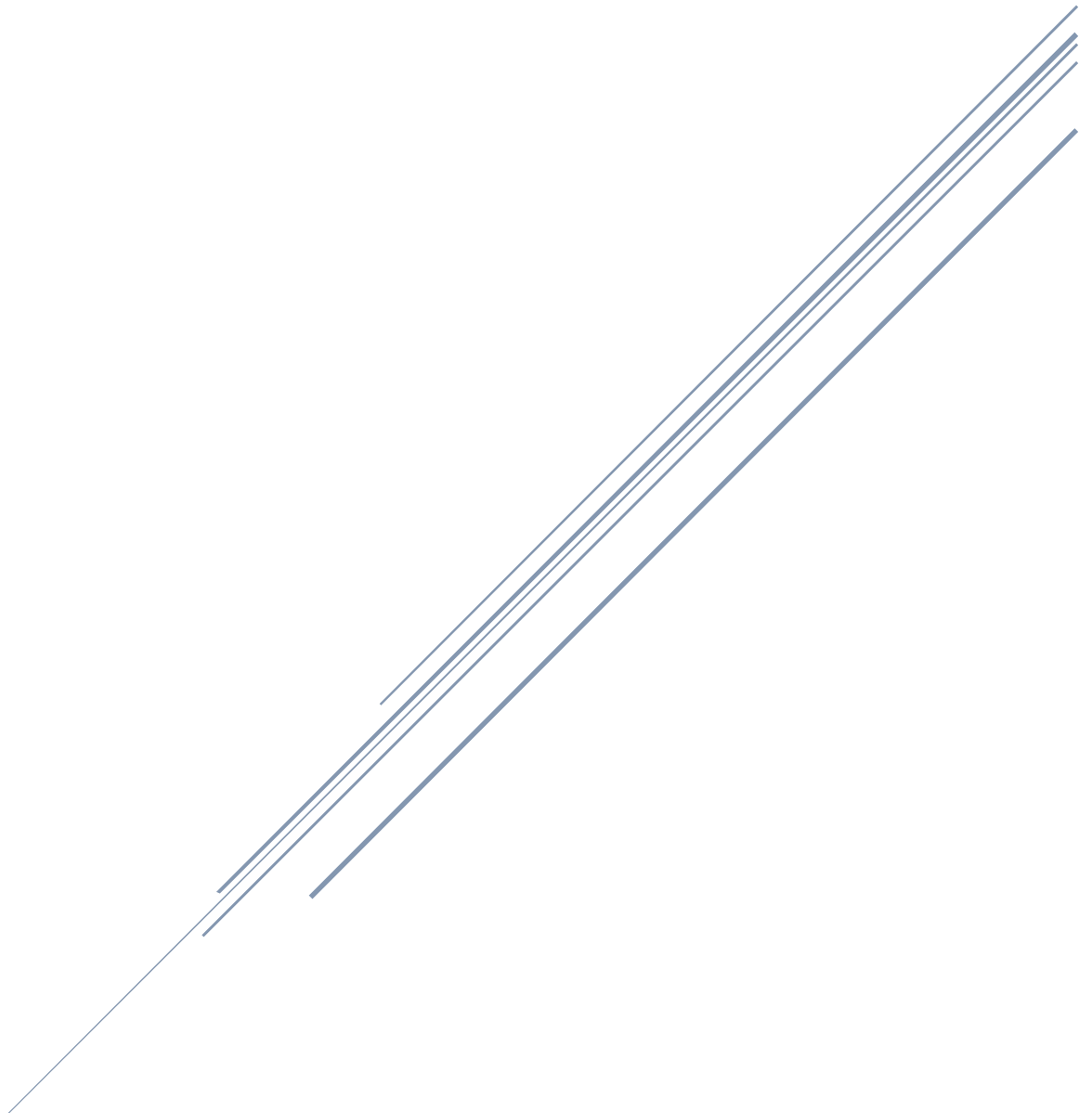


Table of Contents

Business Problem	2
Background/ History	3
Datasets.....	3
Data Preparation.....	4
Visualizations.....	5
Numeric Variables Distribution:	5
Percentage of Heart Attacks by Chest Pain Type	5
Distribution of Bloop Pressure by Gender	6
Distribution of Heart Attack Outcomes	7
Methods	7
Analysis	9
Random Forest Classifier	9
Logistic Regression	9
Support Vector Machine (SVM)	10
Naive Bayes	10
Conclusion	10
Assumptions.....	11
Limitations	11
Challenges.....	11
Future Uses/Additional Applications	12
Recommendations	12
Implementation Plan.....	12
Ethical Assessment.....	12
References.....	13
Questions	13

Business Problem

Heart attacks are a global leading cause of sudden and fatal deaths, often requiring immediate medical intervention. The problem with heart attacks, also known as myocardial infarctions, is multifaceted and has significant implications for public health, individuals, and healthcare systems. The following is an overview of the key issues associated with heart attacks:

- **Lifestyle-Related Risk Factors:** Unhealthy behaviours like poor diet, physical inactivity, smoking, and excessive alcohol consumption significantly heighten the risk of heart attacks.
- **Aging Population:** As the global population ages, heart attack incidence is anticipated to increase, given that age is an unmodifiable risk factor.
- **Healthcare Costs:** Treating heart attacks places substantial financial burdens on healthcare systems, insurers, and individuals, covering hospitalizations, surgeries, medications, and long-term care.
- **Treatment Gaps:** Timely access to effective treatment is crucial during a heart attack, but inadequate healthcare facilities in rural or underserved areas result in delays and poorer outcomes.

Addressing this problem necessitates a holistic approach, encompassing prevention, early detection, access to quality healthcare, and improving cardiovascular health in individuals and populations.

Background/ History

Heart disease has a deep historical footprint, with traces back to ancient Egypt long before modern medical interventions existed. Despite significant advancements in heart care, it remains the leading cause of death in the United States. The Centers for Disease Control and Prevention (CDC) reports that approximately 659,000 people in the U.S. succumb to heart disease annually, accounting for 1 in every 4 deaths.

The origins of heart disease can be dated back to ancient civilizations. The earliest documented case of coronary atherosclerosis, characterized by artery plaque buildup that can lead to heart attacks, was found in an Egyptian princess who lived between 1580 and 1550 B.C. This discovery challenges earlier beliefs that heart disease was less prevalent in ancient times. Over the centuries, mankind has tirelessly made efforts to enhance heart disease care, despite the limited options available for a considerable part of this journey. ([Baystate Health | Healthcare in Western MA](#)).

Datasets

The dataset used in this project is sourced from Kaggle([Fahad Mehfooz](#)). This dataset has around 12 features with around 900 rows. The following are the features available in this dataset.

- i. Age
- ii. Sex
- iii. ChestPainType
- iv. RestingBP
- v. Cholesterol

- vi. FastingBS
- vii. RestingECG
- viii. MaxHR
- ix. ExerciseAngina
- x. Oldpeak
- xi. ST_Slope
- xii. HeartDisease

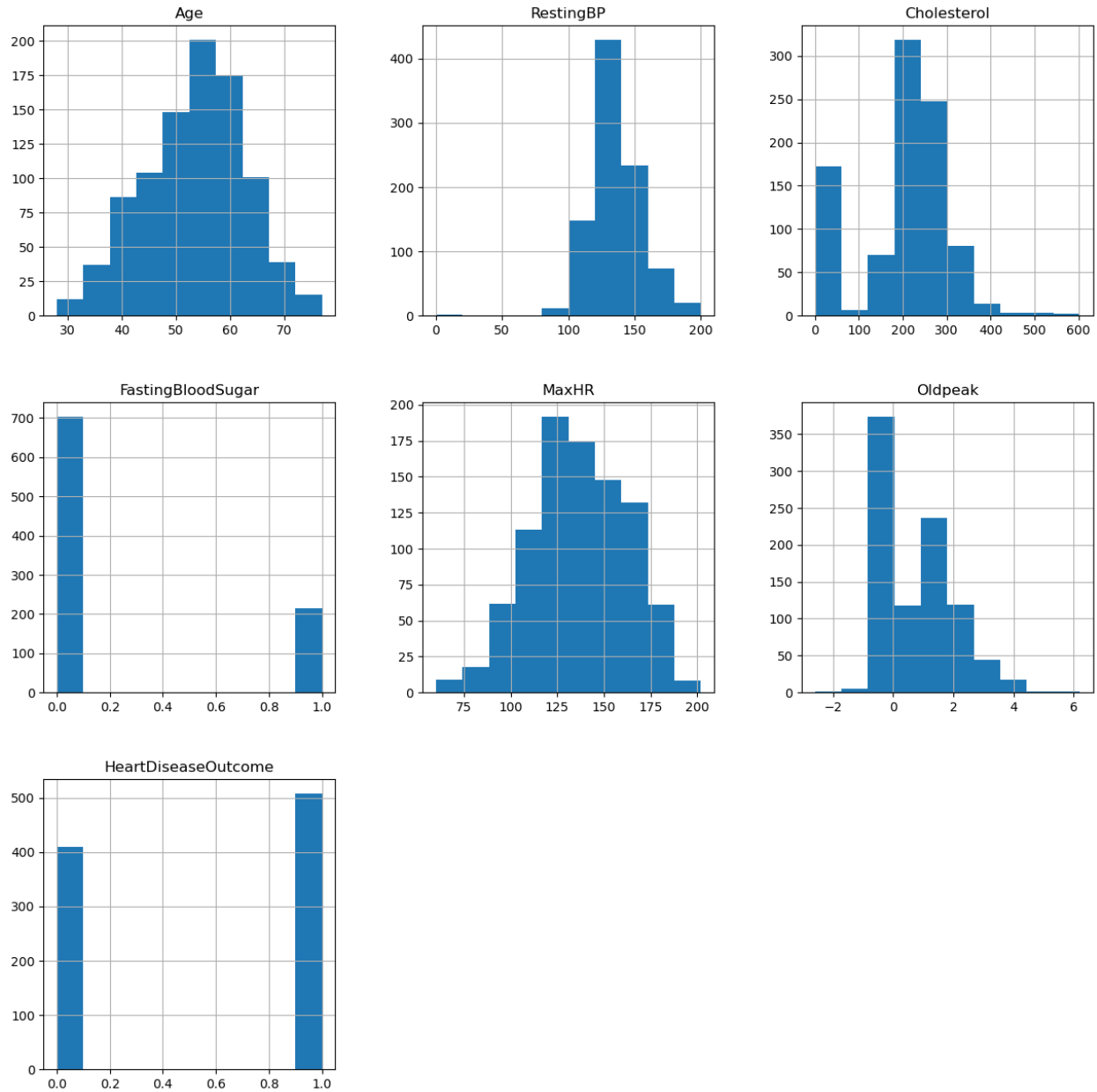
Data Preparation

The following steps were performed to prepare the data for modeling.

- i. Checked for null rows/columns in the data.
- ii. Performed check for duplicates.
- iii. Renamed columns.

Visualizations

Numeric Variables Distribution:

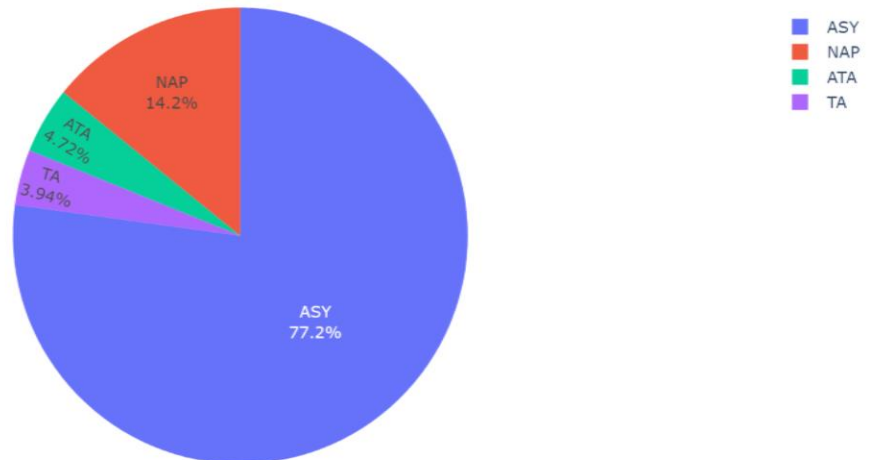


Percentage of Heart Attacks by Chest Pain Type

Chest Pain types in the dataset are as follows:

- “TA”: typical angina
- “ATA”: atypical angina
- “NAP”: non-anginal pain
- “ASY”: asymptomatic

Percentage of Heart Attacks by Chest Pain Type

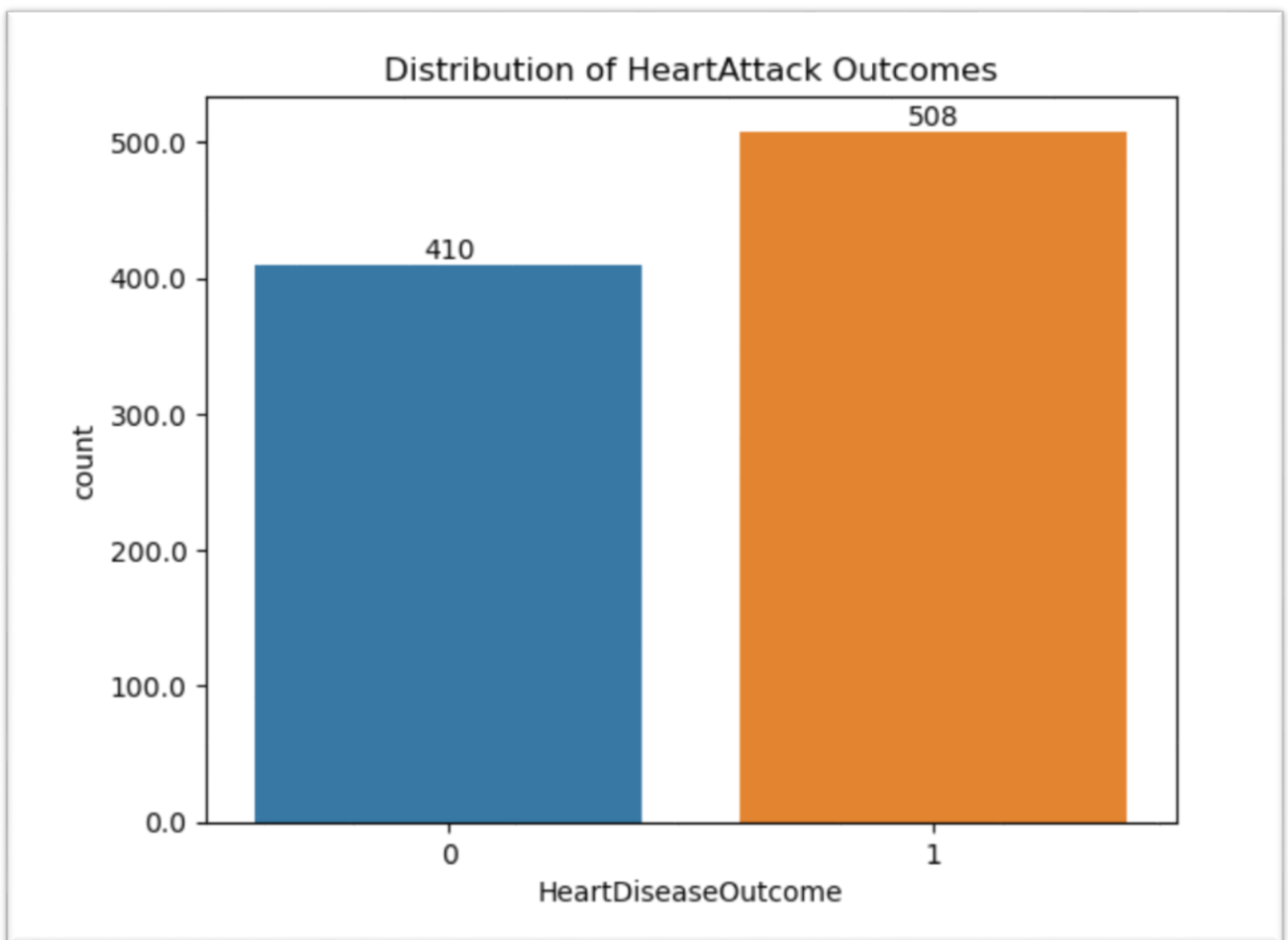


Distribution of Blood Pressure by Gender

Distribution of Resting Blood Pressure by Gender



Distribution of Heart Attack Outcomes



To construct an effective model, it was essential to check the dataset's balance. The current dataset seems to have a reasonable balance.

Methods

The subsequent step was to partition the data into training and testing datasets.

The following models were developed, with their respective outcomes recorded.

- i. Logistic Regression: Logistic Regression is a well-suited model for predicting heart attacks due to its interpretability, computational efficiency, and effectiveness in

- binary classification. Its scalability, regularization capabilities, and well-defined probability estimates make it a widely accepted and trusted choice in the medical field.
- ii. **Random Forest:** A Random Forest model is a strong choice for predicting heart attacks due to its ability to handle complex relationships in the data, manage feature importance, and reduce overfitting. Its ensemble of decision trees provides high accuracy, and it's robust to noisy data. Furthermore, it can accommodate both numerical and categorical features, making it versatile for medical datasets.
 - iii. **Support Vector Machine (SVM):** A Support Vector Machine (SVM) is a strong choice for predicting heart attacks due to its ability to handle complex and non-linear relationships in the data. SVMs excel at separating data into different classes, making them effective for binary classification tasks like heart attack prediction. They work well with both numerical and categorical features and can be fine-tuned for optimal performance.
 - iv. **Naive Bayes:** Naive Bayes is a viable choice for predicting heart attacks due to its simplicity, efficiency, and effectiveness in handling categorical and numerical features commonly found in medical data. It is particularly well-suited when feature independence assumptions hold reasonably true, making it a quick and efficient choice. Naive Bayes can provide valuable insights into feature importance and conditional probabilities, aiding in understanding risk factors.

Analysis

The StandardScaler preprocessing technique was used to standardize or normalize the numerical features in the dataset.

The models were then built, and the outcomes were recorded as follows.

Random Forest Classifier

Heart Attack Outcome	Precision	Recall	F1-Score
0 – No Heart Attack	0.85	0.88	0.87
1 – Heart Attack	0.91	0.89	0.90

Accuracy	ROC-AUC Score
88.6%	0.885

Logistic Regression

Heart Attack Outcome	Precision	Recall	F1-Score
0 – No Heart Attack	0.77	0.88	0.82
1 – Heart Attack	0.91	0.81	0.86

Accuracy	ROC-AUC Score
84.23%	0.848

Support Vector Machine (SVM)

Heart Attack Outcome	Precision	Recall	F1-Score
0 – No Heart Attack	0.82	0.86	0.84
1 – Heart Attack	0.89	0.87	0.88

Accuracy	ROC-AUC Score
86.4%	0.86

Naive Bayes

Heart Attack Outcome	Precision	Recall	F1-Score
0 – No Heart Attack	0.73	0.87	0.79
1 – Heart Attack	0.89	0.77	0.82

Accuracy	ROC-AUC Score
80.9%	0.82

Conclusion

In summary, this project aimed to predict the likelihood of heart attacks using different machine learning models. We explored models such as, logistic regression, random forests, Naïve Bayes, and Support Vector Machine.

Model	Random Forest Classifier	Logistic Regression	Support Vector Machine	Naïve Bayes
Accuracy	88.6%	84.23%	86.4%	80.9%

Each model showed its own strengths and trade-offs. The Random Forest and Logistic Regression models achieved high accuracy and a good balance of precision and recall, making them well-suited for this prediction task. The SVM model also performed well, with a balanced precision-recall trade-off. The Naive Bayes model, while not as accurate as the others, provided reasonable results.

Assumptions

- It is assumed that the input data is accurate, complete, and representative of the target population.
- The assumption that the selected features have clinical validity in predicting heart attacks. Medical expertise and domain knowledge are crucial for selecting and validating these features.

Limitations

Some limitations of heart attack prediction with machine learning include data quality issues, the challenge of capturing complex interactions among risk factors, potential model overfitting, and the need for extensive datasets for robust predictions.

Challenges

Complex and dynamic interactions among risk factors and the changing nature of heart disease present challenges in building accurate predictive models.

Future Uses/Additional Applications

- i. These models can help identify individuals at high risk of heart attacks, allowing for early intervention and preventive care strategies.
- ii. Educational programs and apps can use these models to provide personalized health advice and risk factor management.

Recommendations

This project can be expanded to collaborate with healthcare professionals and institutions to clinically validate the predictive models in real-world clinical settings to ensure their accuracy and effectiveness.

Implementation Plan

The plan involves collecting and preparing patient data, developing predictive models, and validating them with healthcare professionals. Prioritizing ethical data handling, integrating the models into clinical workflows, and providing training for clinicians and patient education are some key steps to be implemented.

Ethical Assessment

Following are a few of the many ethical considerations vital for this project.

- i. Ensuring patient data privacy and obtaining informed consent.
- ii. Adhering to healthcare regulations, such as HIPAA, and obtaining necessary approvals.
- iii. Respecting patients' autonomy and providing them with control over their data.
- iv. Engaging with healthcare professionals, patients, and institutions to incorporate their ethical perspectives.

References

The history of heart disease dates back to Egyptian pharaohs. Baystate Health | Healthcare in Western MA. (n.d.) - <https://www.baystatehealth.org/news/2022/02/history-of-heart-disease#:~:text=The%20American%20College%20of%20Cardiology,common%20in%20ancient%20times%20that>

Fahad Mehfooz. HeartAttack prediction with 91.8 % Accuracy, Kaggle - <https://www.kaggle.com/code/fahadmehfooz/heartattack-prediction-with-91-8-accuracy/input?select=heart.csv>

Questions

1. What machine learning algorithms are employed?
2. What is the impact of these models on patient care?
3. Are regulatory approvals required?
4. What are the potential risks associated with using predictive models?
5. What happens if a model provides a false positive or false negative result?
6. What role do AI and machine learning play in early diagnosis compared to traditional methods?
7. How do you ensure that the models remain up-to-date with the latest medical research?
8. What are the key risk factors the models consider for heart attack prediction?
9. What metrics are used to evaluate model performance?
10. What is the impact of these models on patient care?