

# **CAPSTONE PROJECT**

## **FINAL REPORT :PREDICTIVE ANALYSIS FOR HEART DISEASE RISK ASSESSMENT**



SUBMITTED BY  
-PRERANA DHAKAL

# OVERVIEW



**Business overview** : Highlights both immediate and longterm advantages such as enhanced, reduced readmission rate and integration of predictive models in diagnosis



**Scorecard Highlights**: Demonstrate model accuracy and reliability in identifying key risk predictors like age and high cholestrols level



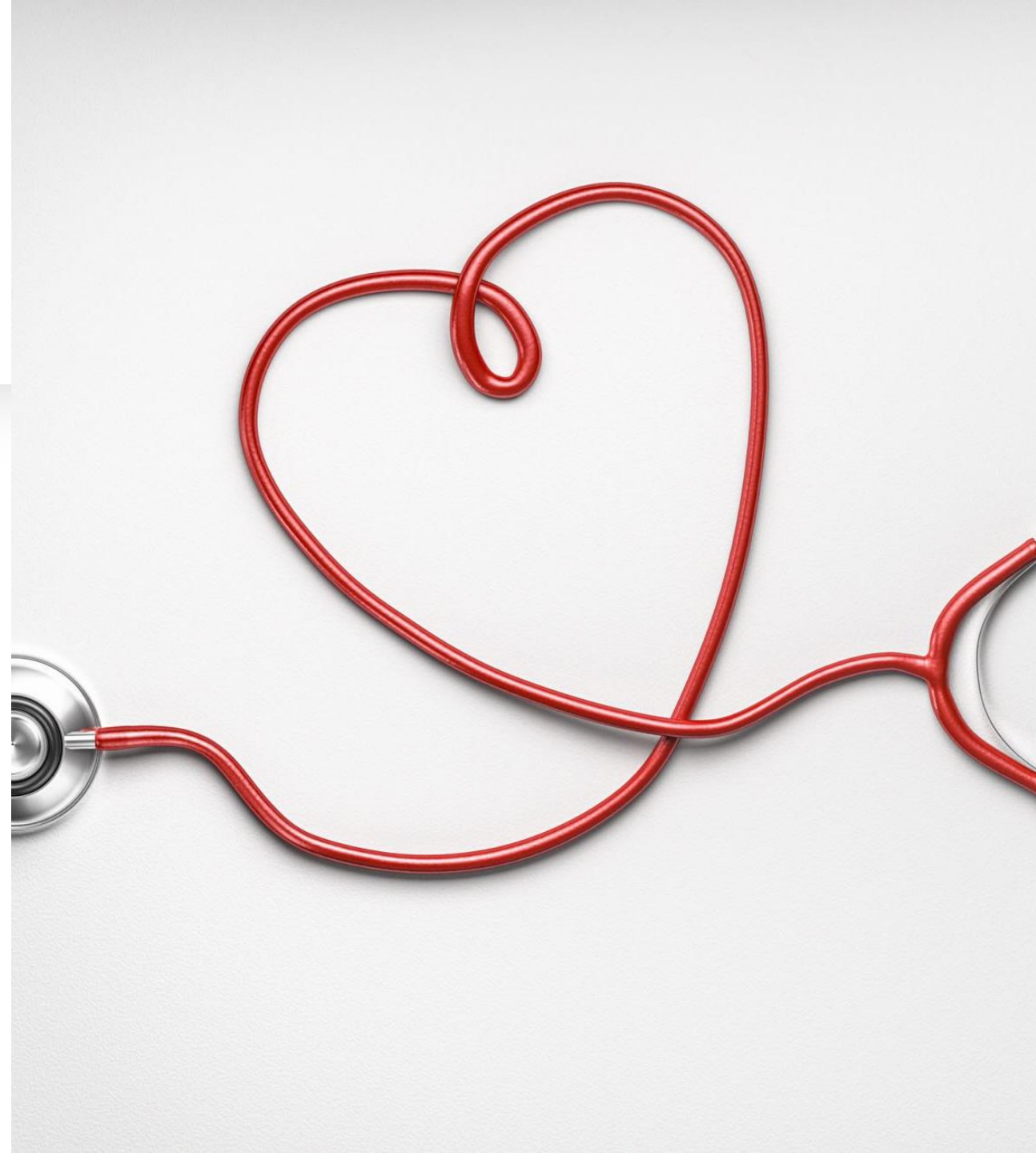
**Recommendation**: Adopt predictive analysis in Routine practice



**Conclusion**: Identification key predictors of heart disease, demonstrating the efficacy of predictive analytics in healthcare.

## **KEY QUESTIONS**

- What are the primary risk factors contributing to heart disease and how can health professionals address them effectively?
- How accurate is the model in identifying individuals at high risk for heart disease ?
- What are the methods used to clean and determine the analysis of the dataset?





# METHODS AND FINDINGS

## Methods for analysis:

- Statistical Software: Python, Pandas, Scikit-learn, Matplotlib
- Data Requirements: Age, cholesterol levels, blood pressure
- Exploratory Analysis: Patterns, correlations, anomalies
- Data Cleaning: Handling missing values, outliers
- Training, Validation, and Testing: Split data for model development

## High level data

- Define the Problem Statement
- Data Collection and Preparation
- Exploratory Data Analysis (EDA)
- Addressing Outliers
- Model Training and Validation

# **FINDINGS**



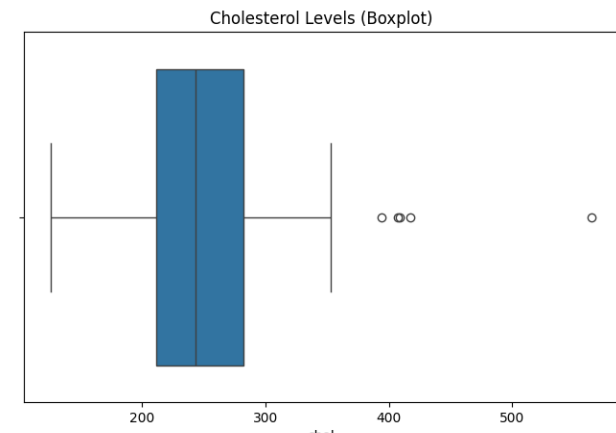
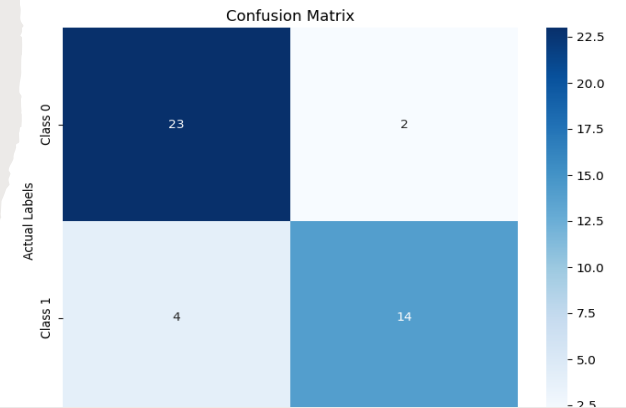
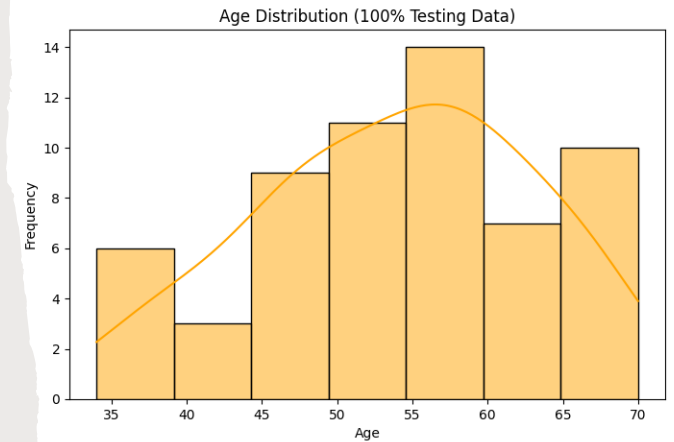
Age and cholesterol levels were found to be the most significant predictors of heart disease.



The model achieved a validation accuracy of approximately 80% when predicting heart disease using 20% of the validation dataset.



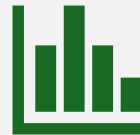
Addressing outliers improved model accuracy and reduced prediction error.



## **RECOMMENDATION**



Adopt Predictive Analytics in Routine Practice



Enhance Data Collection Systems



Focus on Public Awareness Campaigns

# **FURTHER QUESTIONS**



HOW CAN WE IMPROVE THE  
ACCURACY OF THE PREDICTIVE  
MODEL FURTHER?



WHAT ADDITIONAL DATA SOURCES  
CAN BE INTEGRATED TO ENHANCE  
THE MODEL?



WHAT RESOURCES ARE NEEDED TO  
IMPLEMENT THESE  
RECOMMENDATIONS EFFECTIVELY?

# **CONCLUSION**

- **Highlights**

- Early identification of heart disease risk is crucial
- Predictive models offer significant benefits
- Implementation and continuous refinement are key

- **Urgency**

- Immediate action is required to integrate predictive models
- Continuous improvement ensures long-term effectiveness





# **REFERENCE**

- UCL Machine Learning Repository
- Python Documentation
- Pandas Documentation
- Scikit-learn Documentation
- Matplotlib Documentation





THANK  
You! 😊