



Introduction

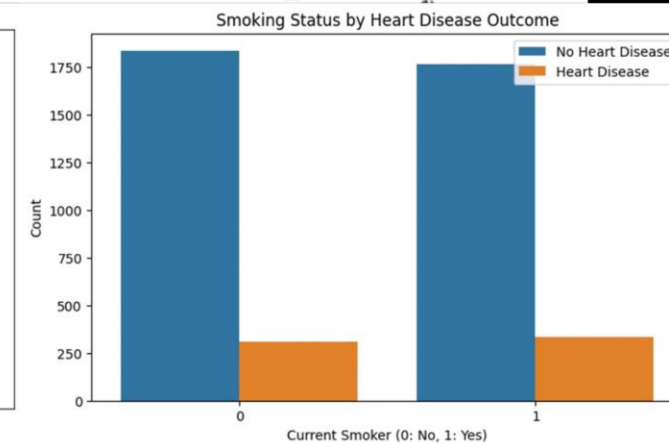
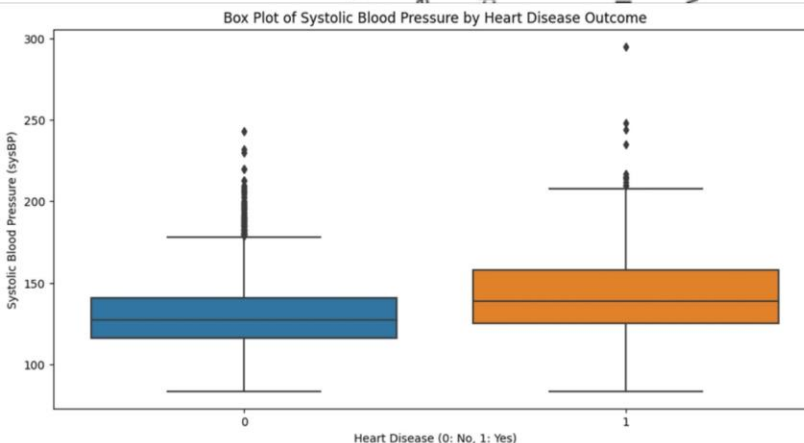
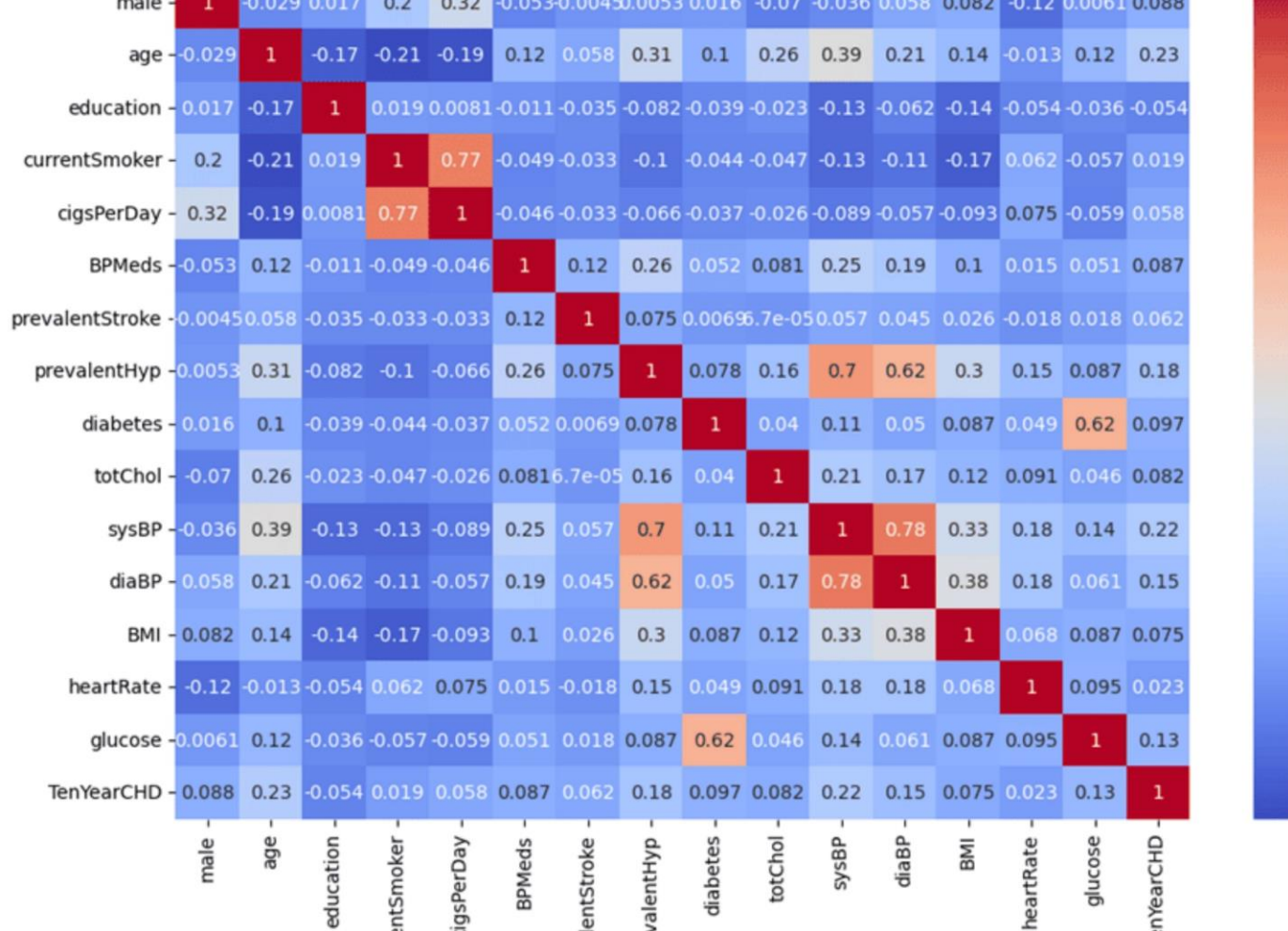
In this presentation, we compare **Logistic Regression** and **Gradient Boosting** for accurate *cardiovascular risk prediction*. We unravel the complexities of heart health to understand the best predictive model for patient care.

Cardiovascular Risk Factors

Understanding the impact of **age**, **gender**, *cholesterol levels*, and *blood pressure* on cardiovascular health is crucial. We explore how these factors influence predictive models.

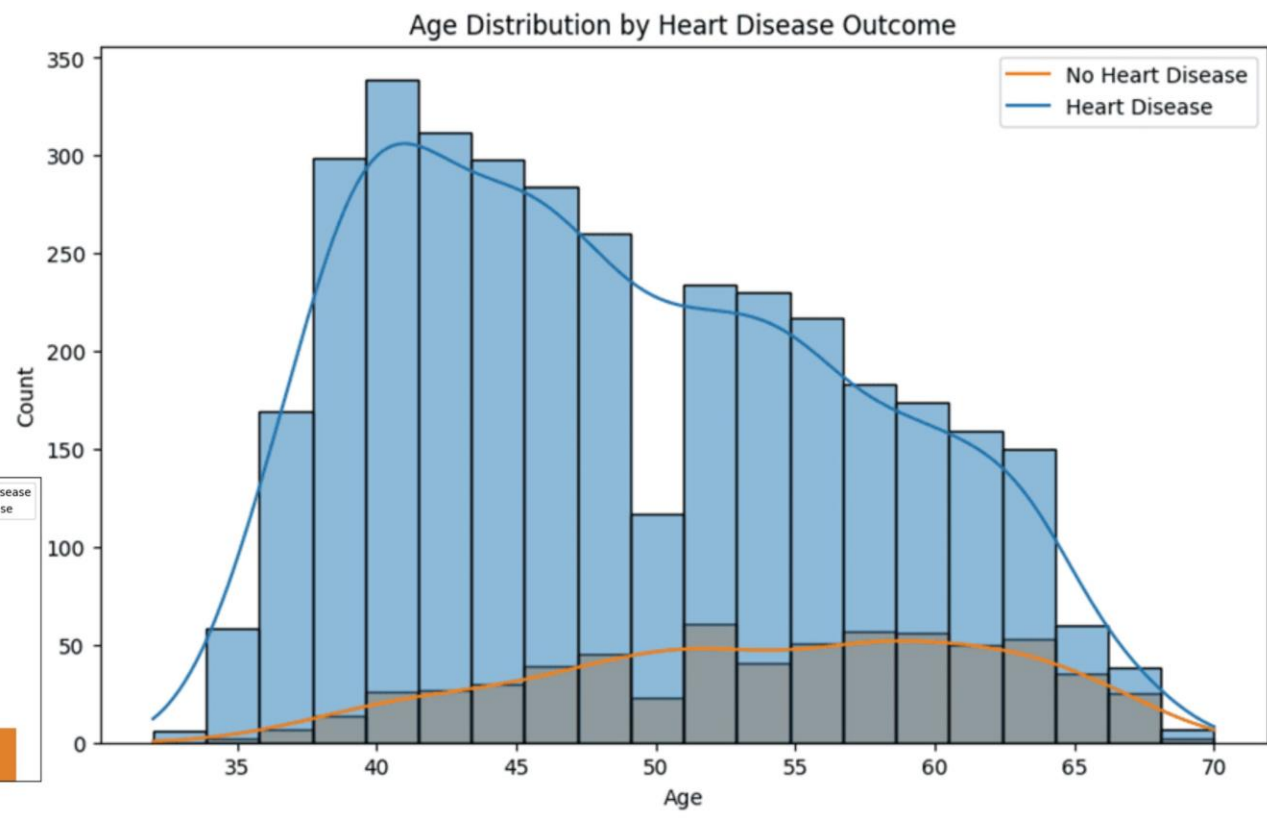
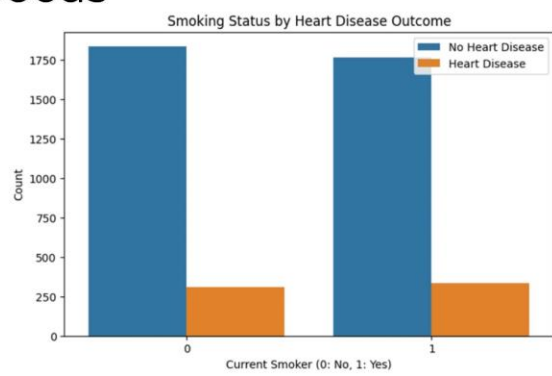
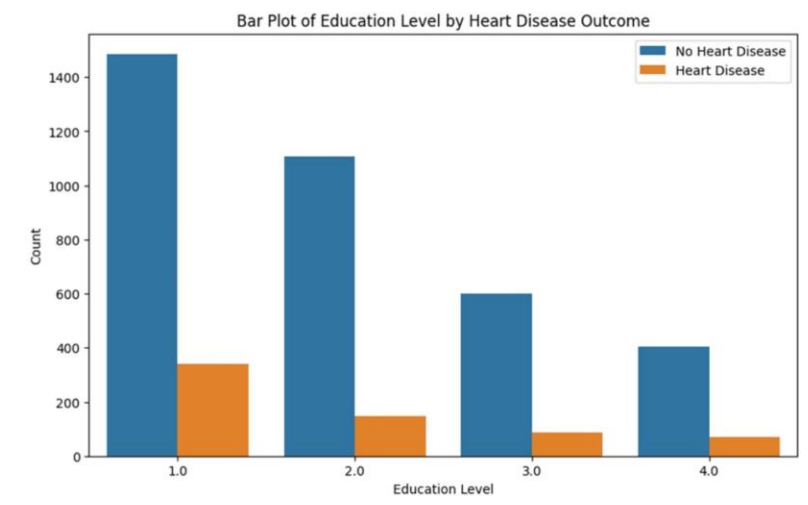
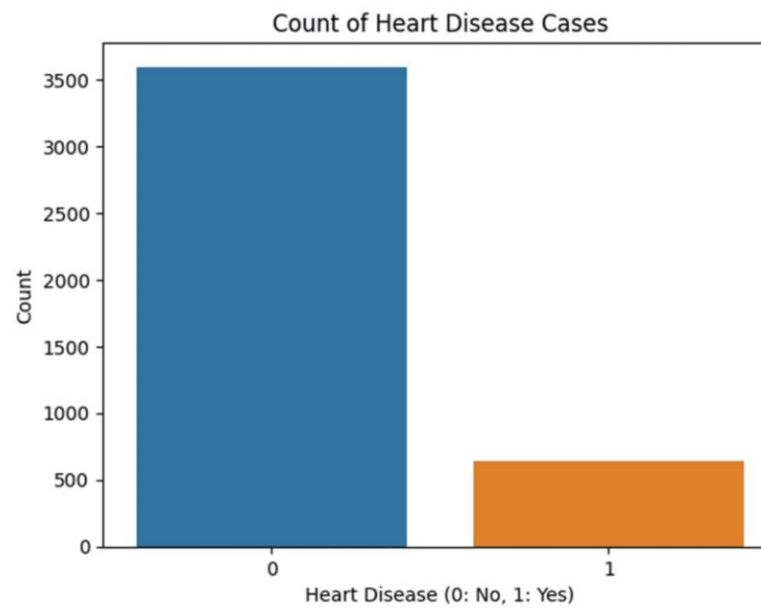
Exploratory Data ANALYSIS

Cardiovascular diseases (CVD) are the leading cause of global mortality, necessitating accurate risk assessment for proactive intervention. The objective of this project is to leverage advanced machine learning techniques to predict the 10-year risk of cardiovascular disease, thereby enabling early preventive measures. Dataset Overview: A comprehensive dataset comprising a myriad of health metrics, including age, cholesterol levels, blood pressure, and lifestyle factors, facilitating a nuanced exploration of factors influencing cardiovascular health.



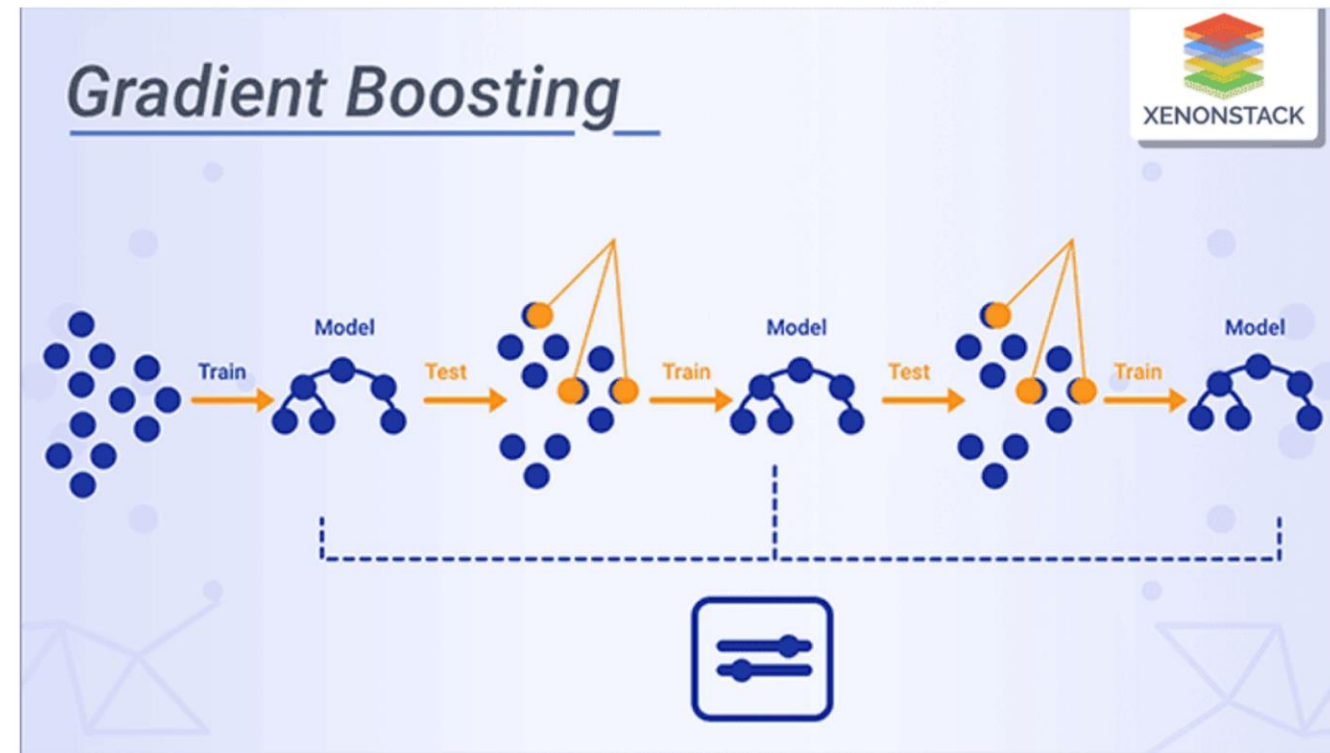
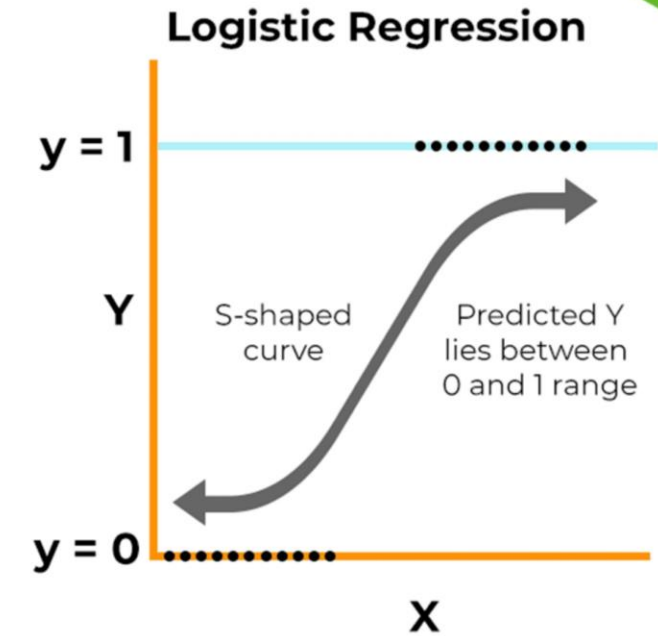
Feature Importance

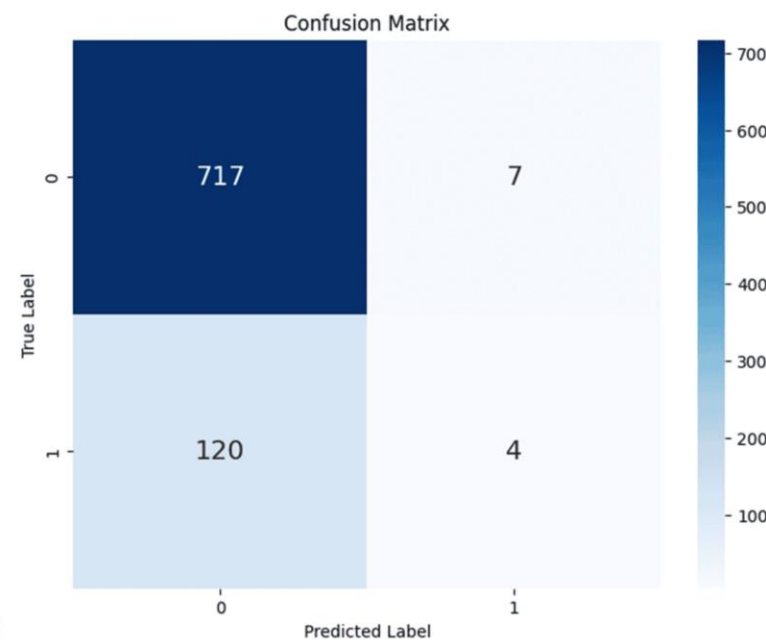
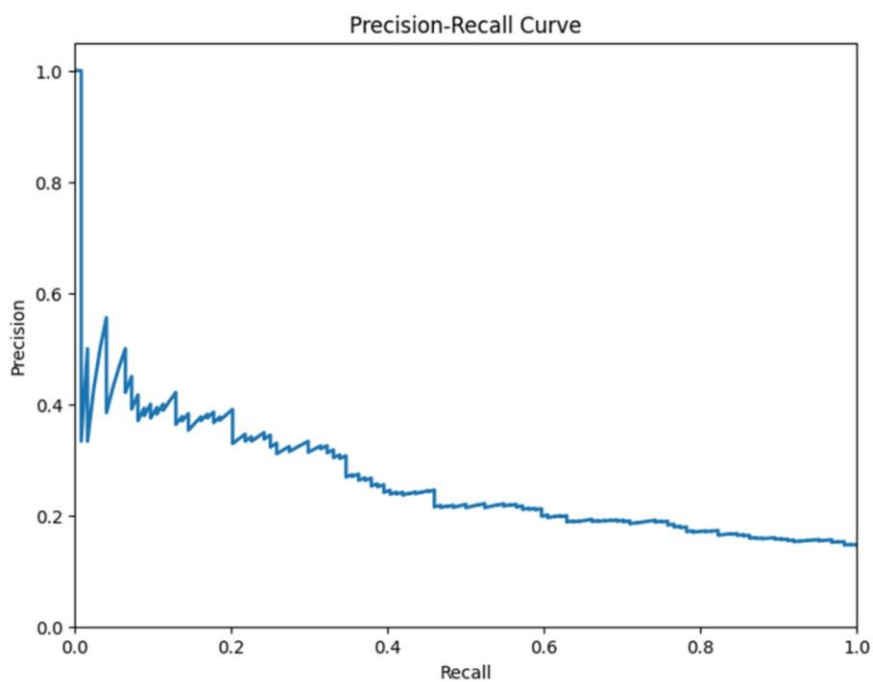
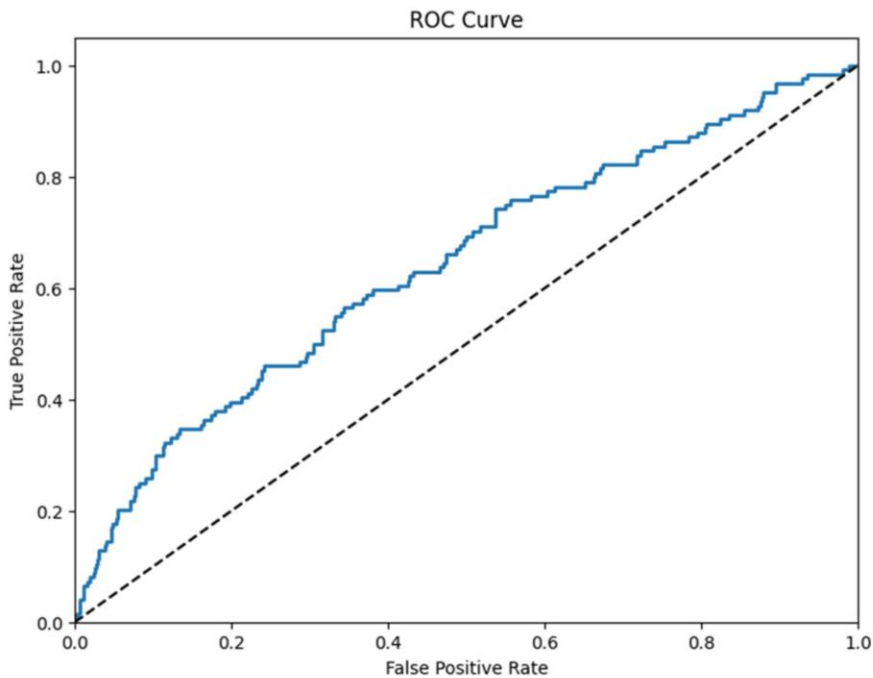
Utilized feature importance analysis, including techniques like permutation importance, to unveil the critical predictors influencing cardiovascular risk. Top predictors include age, cholesterol levels, and blood pressure, emphasizing their significance in determining disease risk. Visual Representation: Graphical representation of feature importance, showcasing the impact of each predictor on the model's predictive power and highlighting areas of focus for preventive measures



Choosing the Right Model

Rigorous model selection process involved considering Logistic Regression for interpretability and Gradient Boosting for capturing intricate relationships. Decision grounded in aligning model choices with the dataset's characteristics to ensure robust and nuanced risk predictions. Models selected based on the trade-off between model complexity and interpretability, aiming for a balance that aligns with the project's goals.



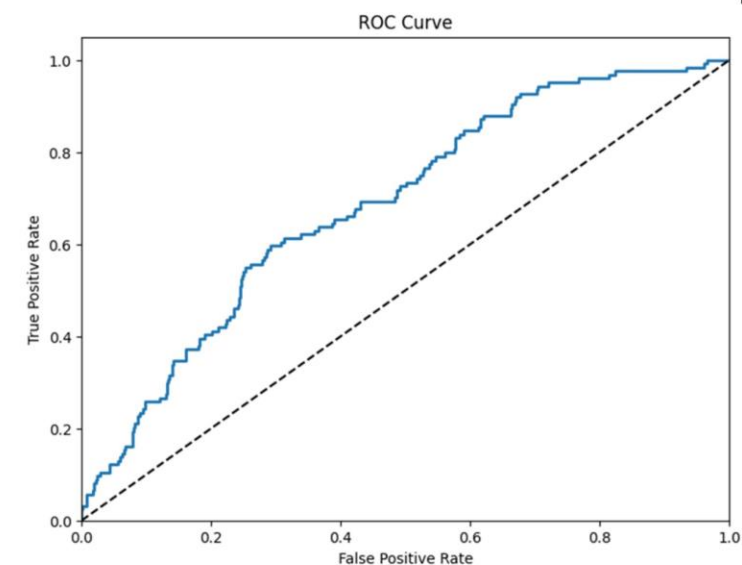
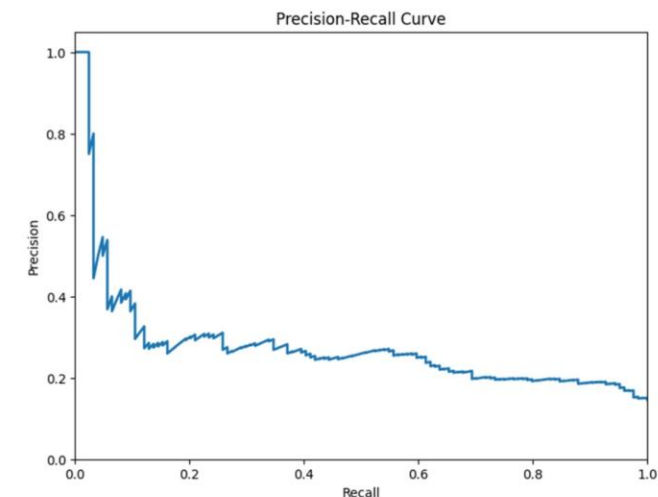
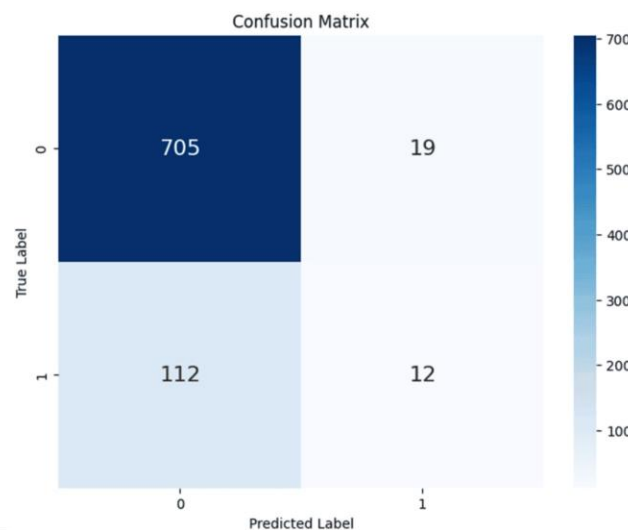


Performance Evolution

Evaluation Metrics: Confusion Matrix: $\begin{bmatrix} 719 & 5 \\ 115 & 9 \end{bmatrix}$ Classification Report: Accuracy: 85.35% Precision: 64% Recall: 7% F1-Score: 13% ROC-AUC: 0.7059 Interpretation: While achieving commendable overall accuracy, Logistic Regression faces challenges in identifying positive cases, as indicated by the low recall. The ROC-AUC score indicates a fair discriminatory power of the model.

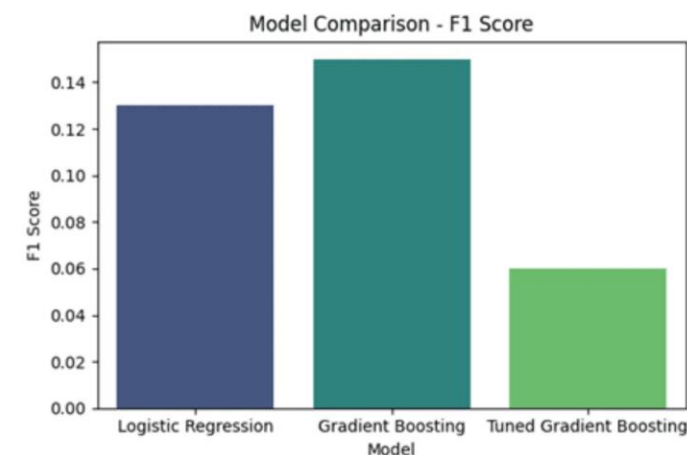
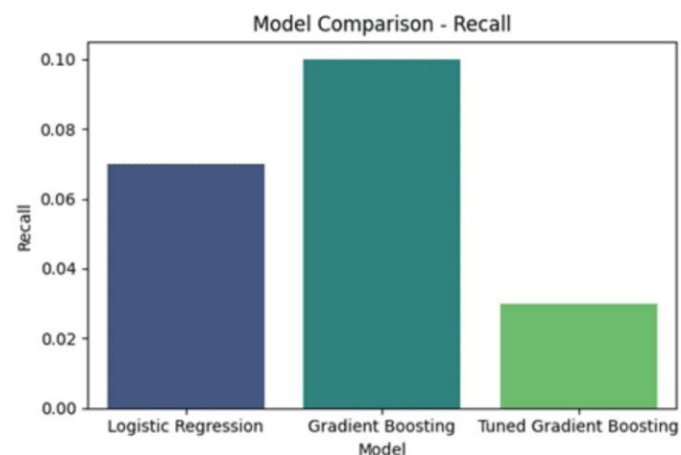
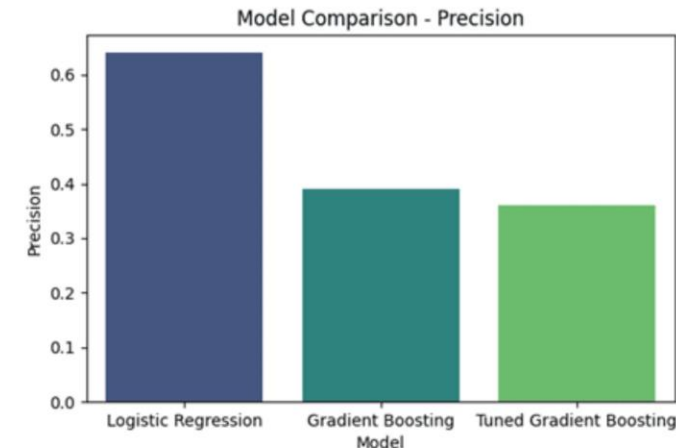
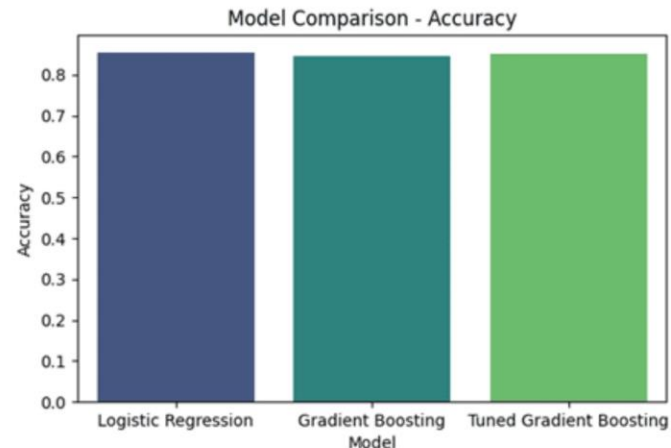
Model Performance - Gradient Boosting

Evaluation Metrics: Confusion Matrix: $\begin{bmatrix} 705 & 19 \\ 112 & 12 \end{bmatrix}$ Classification Report: Accuracy: 84.55% Precision: 39% Recall: 10% F1-Score: 15% ROC-AUC: 0.6865 Interpretation: Gradient Boosting demonstrates higher accuracy but shows a trade-off between precision and recall, reflecting the model's inherent complexity. The ROC-AUC score indicates a moderate discriminatory power of the model.



Model Comparison

Thorough comparison using metrics: Accuracy, Precision, Recall, F1-Score, and AUC to provide a holistic view of model performance. Insights: Logistic Regression excels in overall accuracy, while Gradient Boosting showcases superior precision. A nuanced understanding of trade-offs based on specific needs is imperative for model selection. Visual Comparison: A side-by-side visual representation of performance metrics, offering a comprehensive overview of the strengths and weaknesses of each model.



Conclusion

Summary: Logistic Regression, with its simplicity and commendable accuracy, coexists with the complexity and trade-offs of Gradient Boosting. Next Steps: Suggested avenues for improvement include exploring advanced ensemble methods, hyperparameter fine-tuning, and feature engineering for continuous model refinement and improved risk predictions. Implications: The findings hold significant implications for healthcare professionals, policymakers, and individuals in enhancing cardiovascular risk assessment and preventive strategies.