BUSINESS ENTERPRISE DATA ANALYTICS

Logistic Regression Model Report – Heart Attack Detection

Background

This project focuses on building a logistic regression model to identify patients likely to suffer a heart attack based on medical and demographic indicators. The dataset contains 1,319 observations and includes variables such as Age, Gender, Blood Pressure, Heart Rate, Blood Sugar, and two cardiac biomarkers (CK-MB and Troponin). The target variable is 'Result', indicating whether the patient had a heart attack (positive) or not (negative).

Objective

1. Build a logistic regression model to predict heart attack likelihood.
2. Evaluate variable significance and interpret coefficients.
3. Check multicollinearity.
4. Validate model using cross-validation.
5. Save the model for later use in a Streamlit app.

Exploratory Data Analysis (EDA)

Initial analysis showed that Age, CK-MB, and Troponin were the strongest predictors. Males had a higher incidence rate of heart attacks. Most patients were clustered between 60–70 years of age. Correlation analysis highlighted a strong relationship between Systolic and Diastolic blood pressure, leading to removal of Diastolic BP to avoid multicollinearity.

Multicollinearity Check

Variance Inflation Factor (VIF) was used to assess multicollinearity. All remaining features had VIF values below 5, confirming independence. The final features included: Age, Gender, Heart Rate, Systolic Blood Pressure, Blood Sugar, CK-MB, and Troponin.

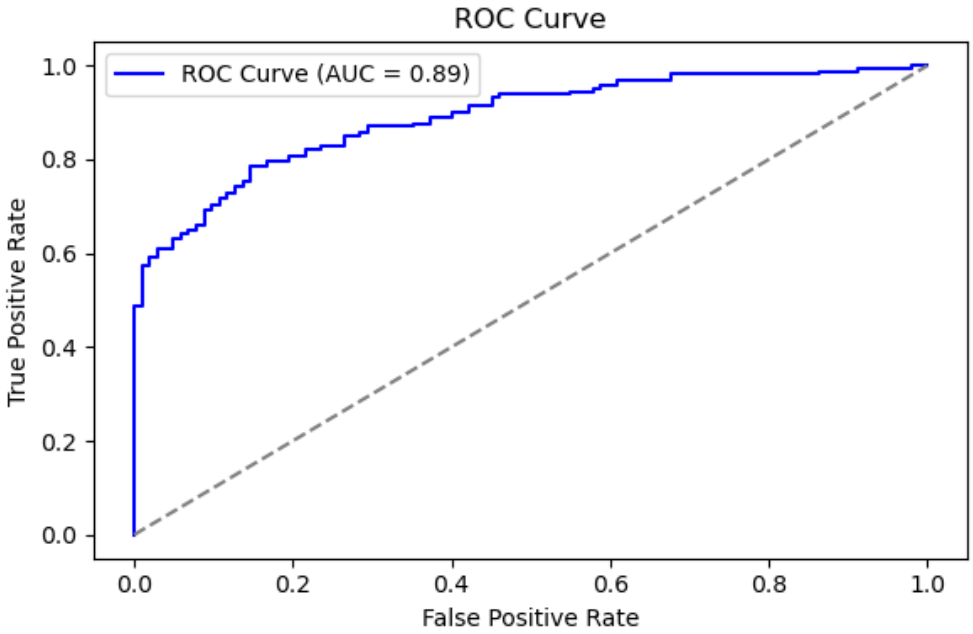
Model Summary and Interpretation

The logistic regression model was trained using scikit-learn and statsmodels. Key coefficients indicated that Troponin and CK-MB were the strongest positive predictors. Age also significantly increased the odds of a heart attack. The statsmodels summary confirmed the statistical significance ($p < 0.05$) of the main predictors.

Model Evaluation

Confusion matrix, classification report, and ROC curve were used to assess performance:
- Accuracy: 80.7%
- Precision (Heart Attack): 83.6%
- Recall (Heart Attack): 85.2%
- AUC Score: 0.89

Cross-validation (5-fold) showed a mean accuracy of 99.9%, indicating strong generalization.



Model Deployment

The final logistic regression model and scaler were saved using joblib and will be used to create an interactive Streamlit app. Files were saved as:
- models/logreg_heart_attack_model.pkl
- models/standard_scaler.pkl