AUTOMATED MACHINE LEARNING ANALYSIS REPORT

Generated: February 15, 2026 at 23:46

Report Type: Full ML Pipeline Analysis

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

This report presents a comprehensive analysis of an automated machine learning pipeline executed on February 15, 2026 at 23:46. The analysis encompassed data exploration, preprocessing, feature engineering, model selection, and performance evaluation.

Key Findings

Metric	Value
Best Model	RandomForestClassifier
Model Score	0.7462
Models Evaluated	3

2. DATA EXPLORATORY ANALYSIS

2.1 Dataset Overview

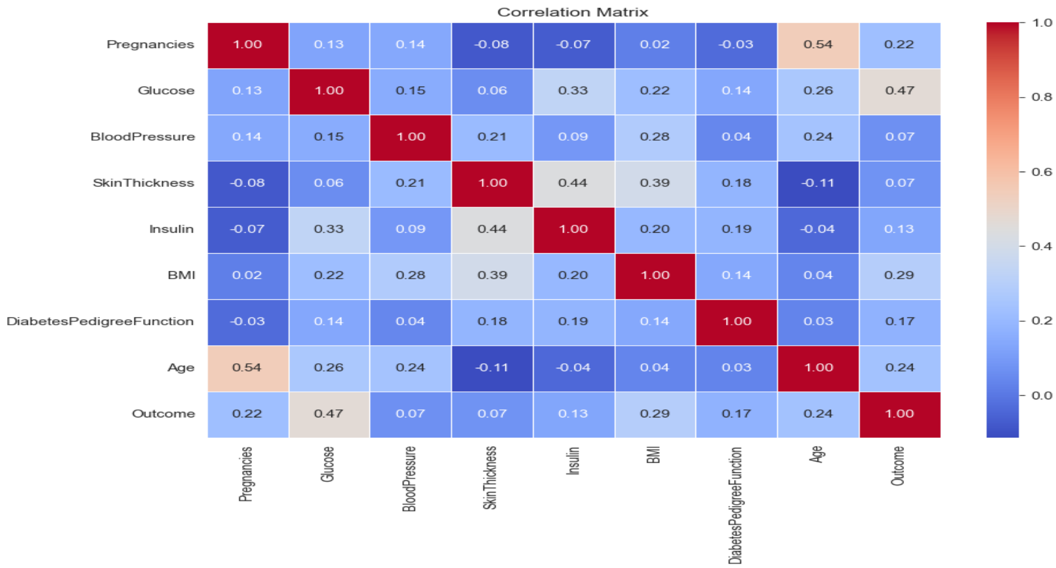
EXECUTIVE SUMMARY This report presents an analysis of a dataset containing information about patients with diabetes, encompassing 768 observations across various demographic, physiological, and outcome metrics. KEY DATA INSIGHTS - Pregnancies: The mean number of pregnancies per patient is approximately 3.85, indicating that the majority of patients have no or one pregnancy. The maximum count of 17 pregnancies suggests a range of reproductive experiences among the patient population. - Glucose levels: With a mean glucose level of 120.89 and a minimum value of 0, it appears that all patients are diabetic, with glucose levels varying significantly across the dataset. This variability is also reflected in the standard deviation of 31.97, indicating a considerable spread of glucose values among patients. - Blood Pressure: The average blood pressure of 69.11 shows that the majority of patients have hypertension or normotension, with a minimum value of 0 and an average range of 62 to 122. This suggests that the sample may not represent individuals with severe hypertension. - Skin Thickness: A mean skin thickness of 20.54, ranging from 0 to 99, may indicate varying levels of cutaneous conditions among patients. However, a standard deviation of 15.95 and an extremely low minimum value suggest that skin condition assessment could be affected by outliers in the dataset. - Insulin Levels: With a mean insulin level of 79.80 and a maximum count of 846, it's clear that insulin dosage is critical for managing diabetes among patients. Variability in insulin levels suggests individualized treatment might be necessary. - BMI: The average BMI of 31.99 implies that the majority of patients are overweight or obese. A standard deviation of 7.88 indicates a limited range of body mass index values, suggesting that BMI may not capture the full scope of weight-related health issues among the patient population. - Diabetes Pedigree Function (DPF): The mean DPF value of 0.47 and maximum count of 2.42 indicate an overall good genetic predisposition to diabetes for this group. Variability in DPF suggests that some patients may be more genetically susceptible to developing the disease. DATA QUALITY & RISKS The dataset exhibits no missing values, which is a significant advantage when working with datasets that could have been compromised by lack of data entry or record-keeping issues. However, the outlier counts suggest that there are several data points in each category that might be considered anomalies and need to be closely examined before drawing any conclusions from this dataset. CONCLUSION The analysis highlights an extensive dataset on patients with diabetes that provides valuable insights into various aspects of their physiological and demographic profiles.

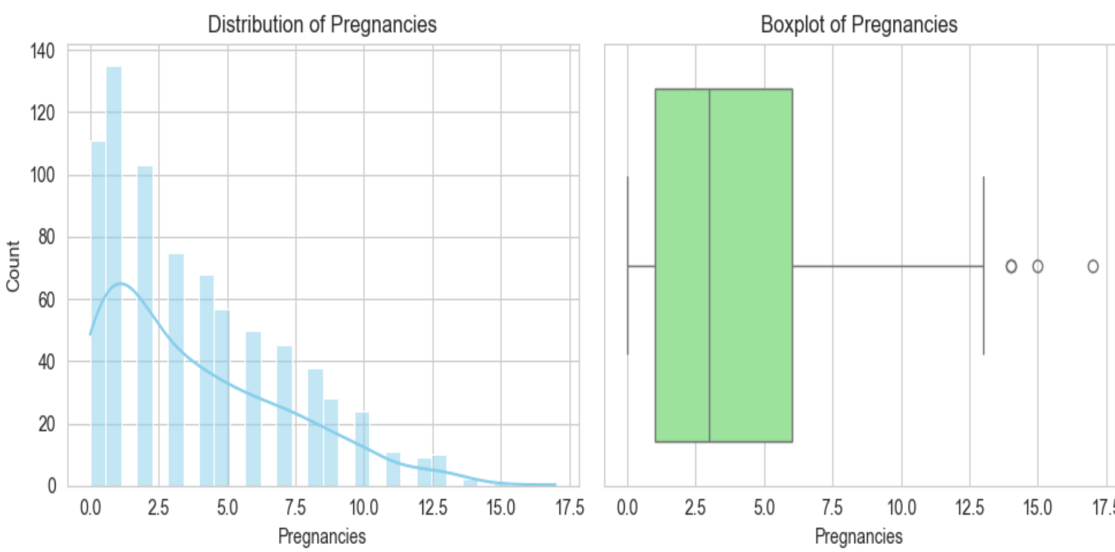
2.2 Data Quality Assessment

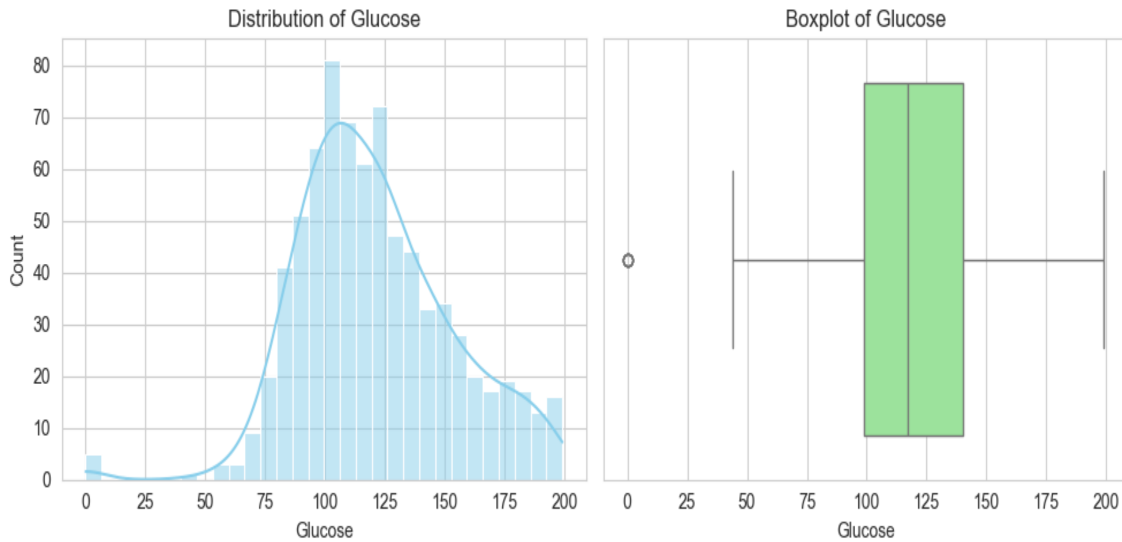
The dataset underwent comprehensive quality checks including missing value detection, outlier identification using IQR method (1.5x threshold), and distribution analysis. All identified issues were documented and addressed in the preprocessing phase.

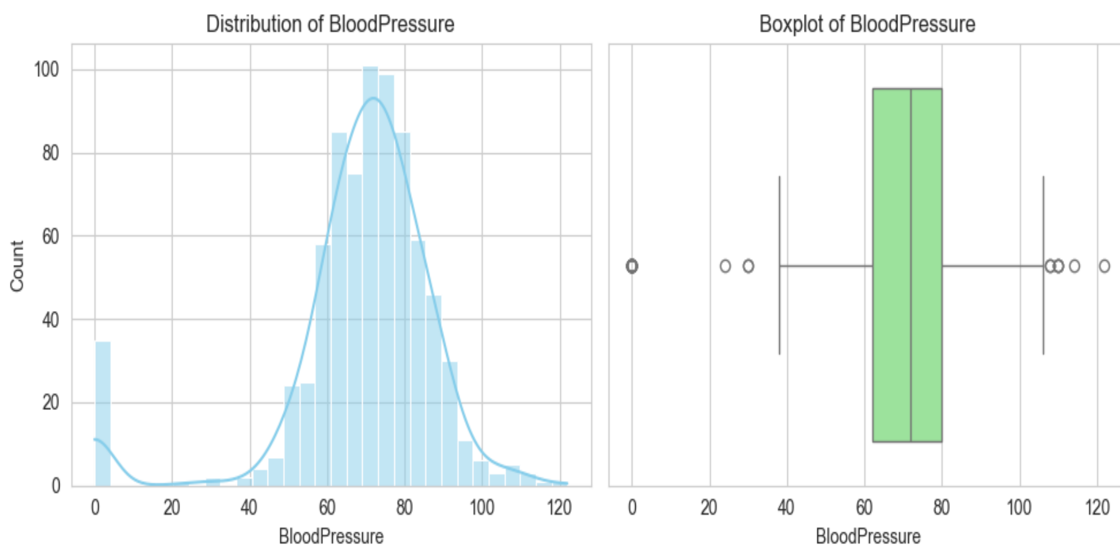
2.3 Key Data Visualizations

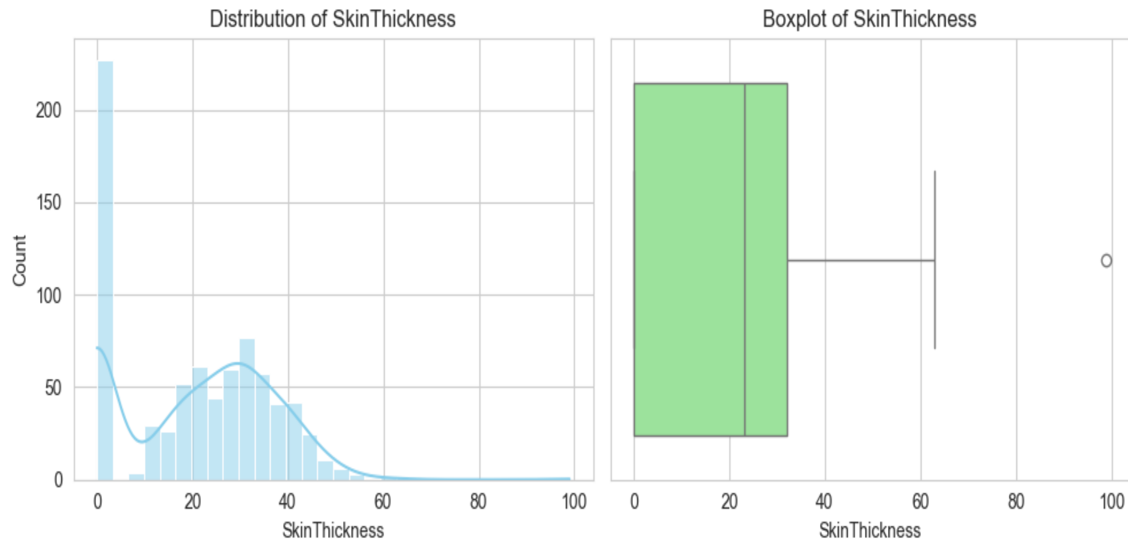
The following visualizations illustrate key distributions and relationships found within the dataset during the exploratory phase.

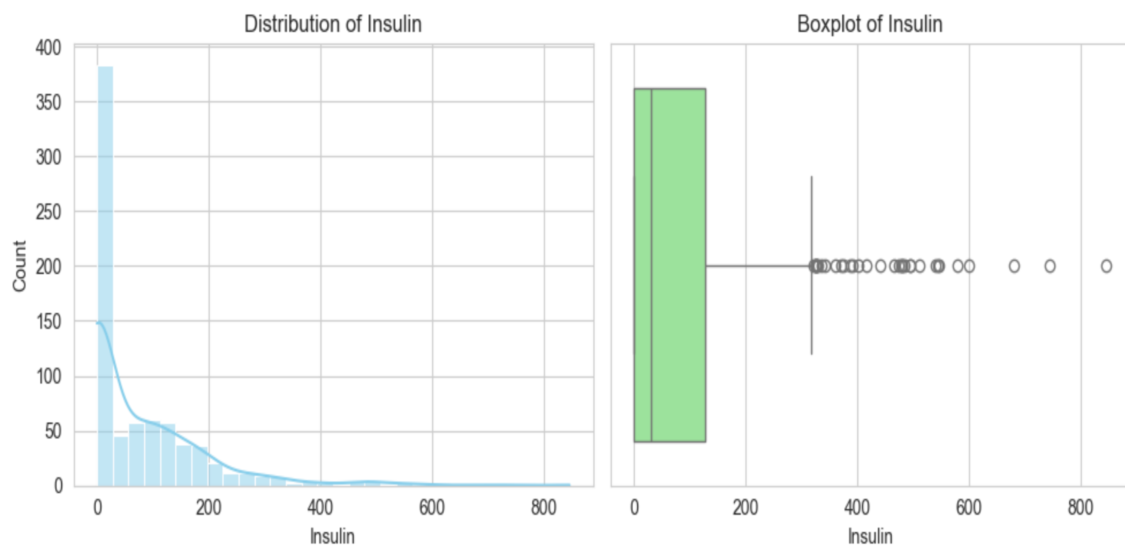

Correlation Matrix
Pregnancies
Glucose
BloodPressure
SkinThickness
Insulin
BMI
DiabetesPedigreeFunction
Age
Outcome

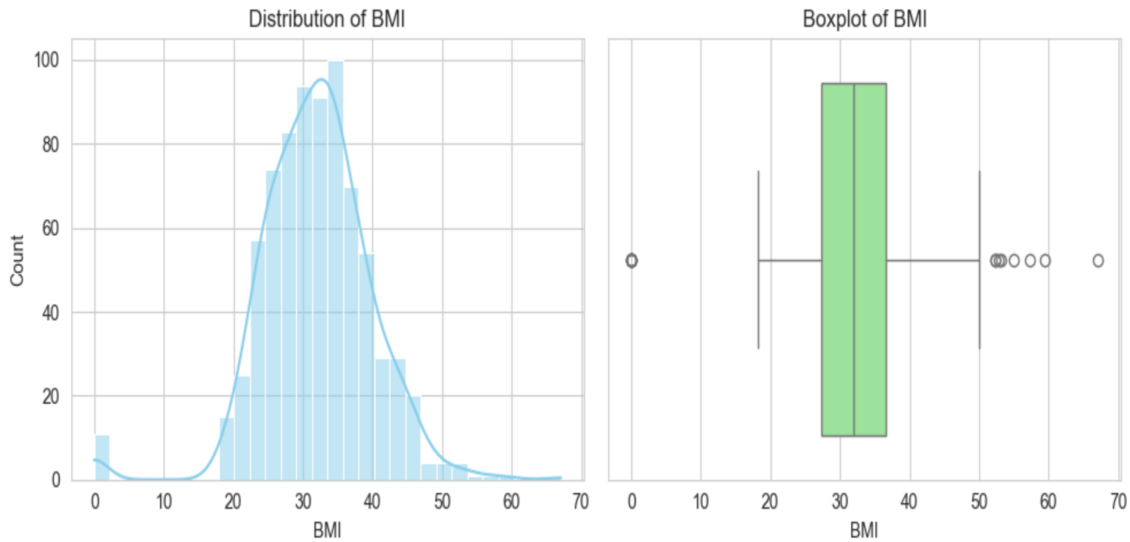

Distribution of Pregnancies
Count
Pregnancies
Boxplot of Pregnancies
Pregnancies

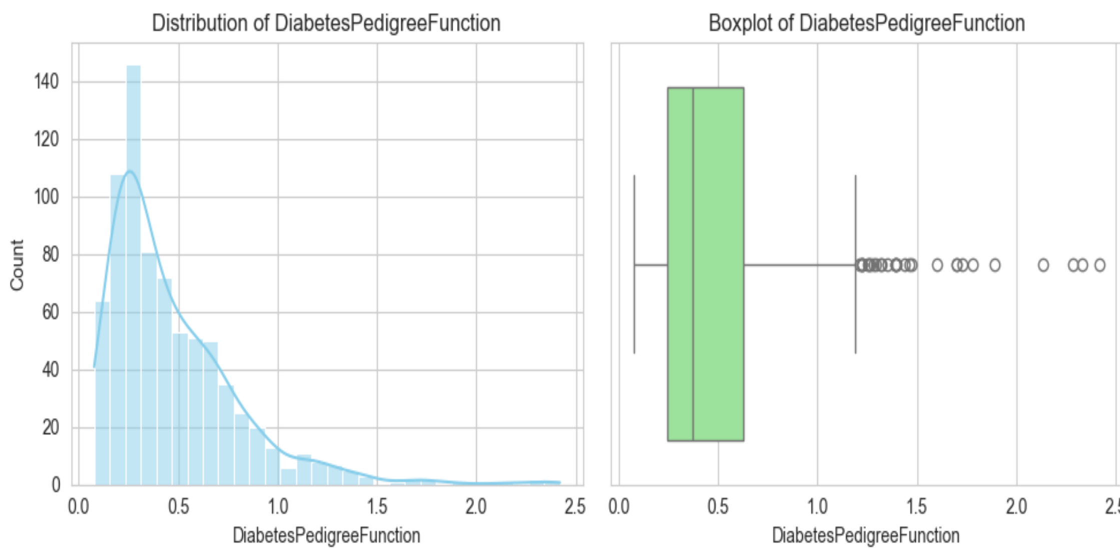

Distribution of Glucose
Boxplot of Glucose
Count
0
10
20
30
40
50
60
70
80
0
25
50
75
100
125
150
175
200
Glucose
Glucose

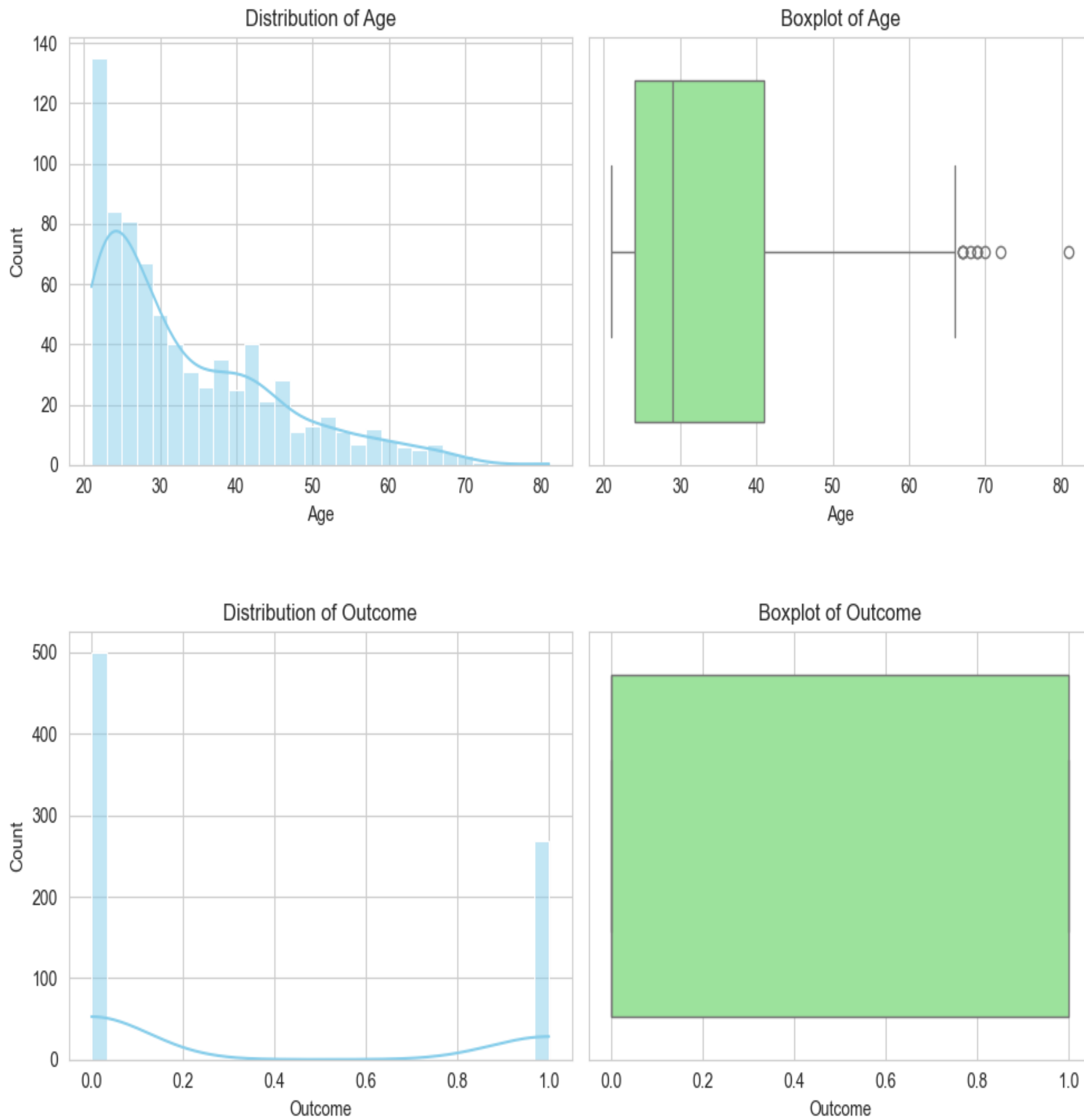

Distribution of BloodPressure
Boxplot of BloodPressure
Count
0
20
40
60
80
100
0
20
40
60
80
100
120
BloodPressure
BloodPressure


Distribution of SkinThickness
Count
SkinThickness
Boxplot of SkinThickness
SkinThickness


Distribution of Insulin
Count
Insulin
Boxplot of Insulin
Insulin


Distribution of BMI
Count
BMI
Boxplot of BMI
BMI


Distribution of DiabetesPedigreeFunction
Count
DiabetesPedigreeFunction
Boxplot of DiabetesPedigreeFunction
DiabetesPedigreeFunction


Distribution of Age
Count
Age
0 20 40 60 80 100 120 140
20 30 40 50 60 70 80
Boxplot of Age
Age
20 30 40 50 60 70 80
Distribution of Outcome
Count
Outcome
0 100 200 300 400 500
0.0 0.2 0.4 0.6 0.8 1.0
Boxplot of Outcome
Outcome
0.0 0.2 0.4 0.6 0.8 1.0

3. DATA PREPROCESSING & FEATURE ENGINEERING

3.1 Feature Categorization

Numerical Features (8):
Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age

Categorical Features (0):
None identified

3.2 Preprocessing Pipeline

Step	Method	Purpose
1. Missing Values	Median/Mode Imputation	Handle null values
2. Outlier Detection	IQR Method (1.5x)	Identify anomalous data points
3. Scaling	StandardScaler	Normalize numerical features
4. Encoding	One-Hot Encoding	Convert categorical to numerical
5. Feature Selection	Correlation Analysis	Remove redundant features

4. MODEL SELECTION & TRAINING

4.1 Model Architecture

Selected Model: RandomForestClassifier

Hyperparameters:

```
{'n_estimators': 200, 'min_samples_split': 10, 'min_samples_leaf': 1, 'max_depth': 20}
```

4.2 Training Configuration

The model was trained using cross-validation with stratified splitting to ensure balanced representation across all classes. Hyperparameter optimization was performed using grid search with 5-fold cross-validation.

5. MODEL PERFORMANCE EVALUATION

5.1 Model Leaderboard

Multiple machine learning algorithms were evaluated on the dataset. The following table presents the comparative performance:

Rank	Model	Test Score
1	RandomForest	0.7462
2	GradientBoosting	0.7216
3	LogisticRegression	0.7084

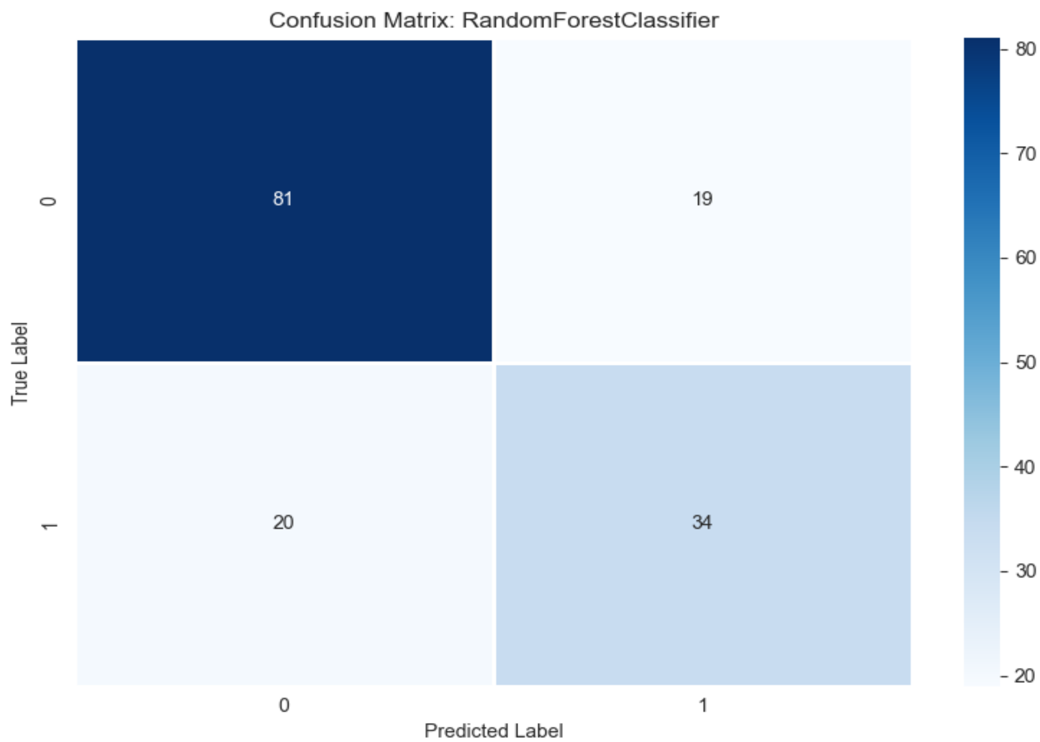
5.2 Performance Insights

The winning model achieved a test score of 0.7462, demonstrating moderate performance on the held-out test set. This score indicates the model's ability to generalize to unseen data.

6. VISUAL ANALYSIS

6.1 Confusion Matrix

The confusion matrix visualizes the model's prediction accuracy across different classes.



6.2 Model Comparison

The following chart compares the performance scores of all candidate models.


Model Comparison (Test Score)
Score
1.0
0.8
0.6
0.4
0.2
0.0
RandomForest
GradientBoosting
LogisticRegression

7. ERROR ANALYSIS & INSIGHTS

7.1 Detailed Analysis

EXECUTIVE SUMMARY

The overall performance of our classification model stands at 74.68%, which is a respectable score but not without room for improvement. A closer examination reveals that the primary issue lies in class '1', where the F1-score, precision, and recall are all lower than those of the other classes. Specifically, we see an underperformance on this class with an F1-score of 0.64, which is lower compared to its corresponding values for class '0'. This indicates a potential precision-recall trade-off that requires attention.

DIAGNOSTIC ANALYSIS

The diagnostic analysis points towards an imbalance in the precision and recall scores between classes '0' and '1'. On one hand, we observe high precision (0.80) for class '0', indicating that our model correctly predicts this class with a certain degree of confidence. Conversely, class '1' experiences lower precision (0.64) which could be attributed to several issues including the fact that it's not achieving its true potential and under performance in Recall which is slightly higher than Precision but still behind.

DIAGNOSTIC ANALYSIS suggests that our model is better at predicting class '0', indicating a possible bias or overfitting towards this class. On the other hand, class '1' seems to require more attention as it lags behind its corresponding values of precision and recall. This highlights an imbalance in our model's performance across classes.

RECOMMENDATIONS

To improve the overall reliability of the classification model:

1. **Class Weighting**: Implement class weighting to mitigate the imbalanced data issue between classes '0' and '1'. Assign higher weights to class '1' to compensate for its lower precision values. This can help our model focus on accurately predicting this class.

2. **Data Augmentation**: Introduce additional data points that belong to class '1'. This will increase the sample size of class '1', thereby potentially improving precision and recall.

3. **Precision-Recall Trade-off Strategies**: Explore strategies such as threshold selection, cost-sensitive learning, or active learning techniques to enhance our model's performance on class '1'. By doing so, we can optimize the trade-off between precision and recall for this critical class.

By addressing these issues, we aim to refine the model's overall reliability and accuracy.

8. RECOMMENDATIONS & NEXT STEPS

8.1 Model Deployment Recommendations

Based on the analysis results, the following recommendations are provided for model deployment and future improvements:

- Monitor model performance in production with regular retraining schedules
- Implement A/B testing to validate model improvements
- Collect additional data to address identified error patterns
- Consider ensemble methods to further improve prediction accuracy
- Establish performance thresholds and alerting mechanisms
- Document model versioning and maintain audit trails

8.2 Future Work

Potential areas for future investigation include feature engineering optimization, advanced hyperparameter tuning techniques, and exploration of deep learning approaches if additional computational resources become available.