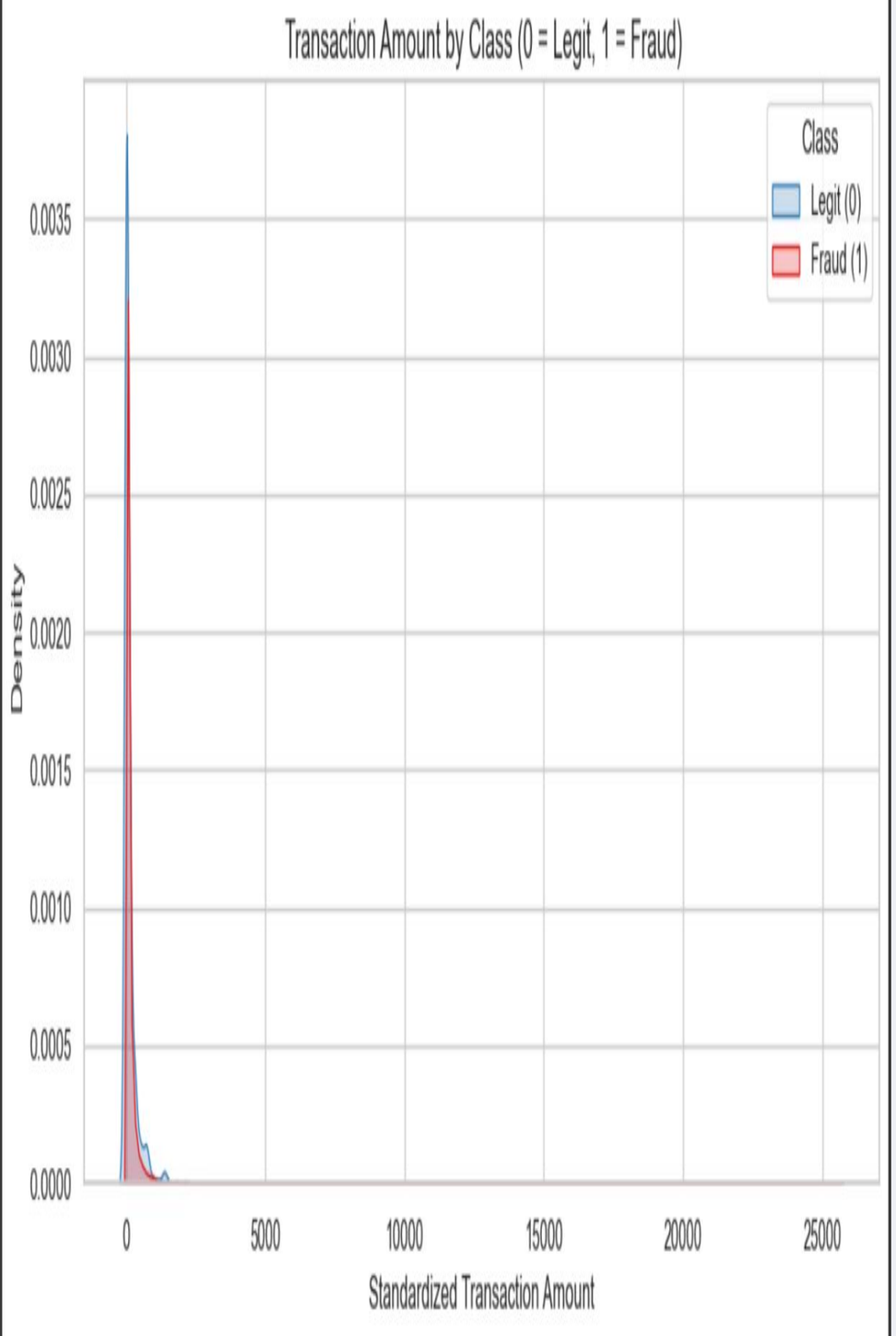
Credit Card Fraud Detection

Credit card fraud is a critical issue escalating with increased card usage. Global losses are projected to exceed $400 billion. Machine learning offers a modern defense strategy.

Machine Learning Approach

Steps involved:

1. **Data Acquisition & Exploration**
2. **Data Preprocessing**
3. **Handling Class Imbalance**
4. **Dataset Splitting**
5. **Model Selection & Training**
6. **Model Evaluation**
7. **Results Summary**
8. **Conclusion**



🔍 Data Preprocessing

Load Dataset

Loaded `creditcard.csv`

Exploration:

`.head()`, `.info()`, `.describe()`

Problem Identified:

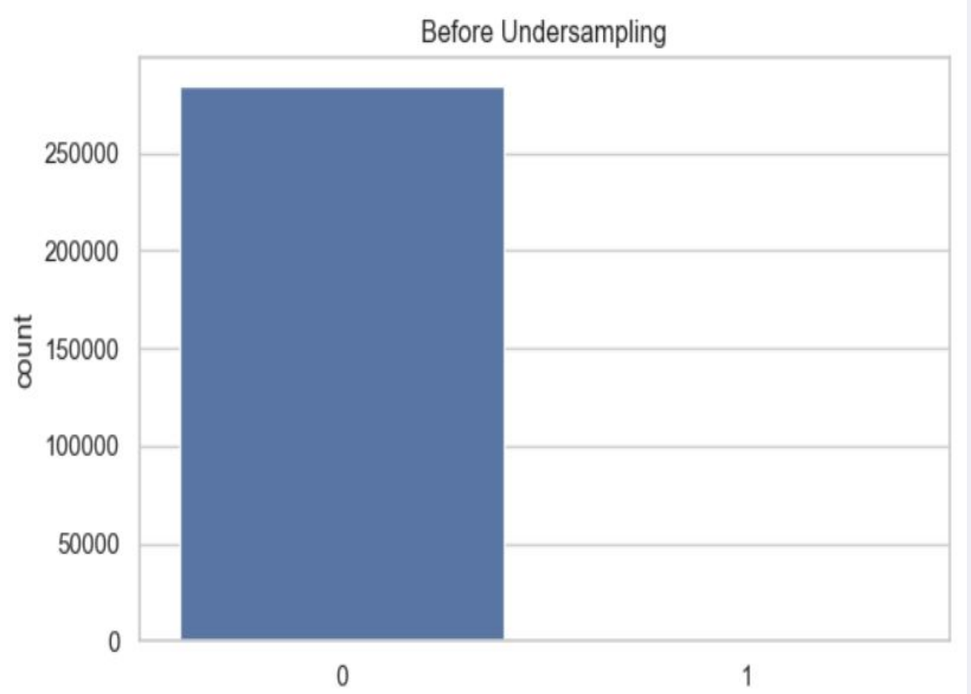
Highly imbalanced dataset (Fraud cases are rare)

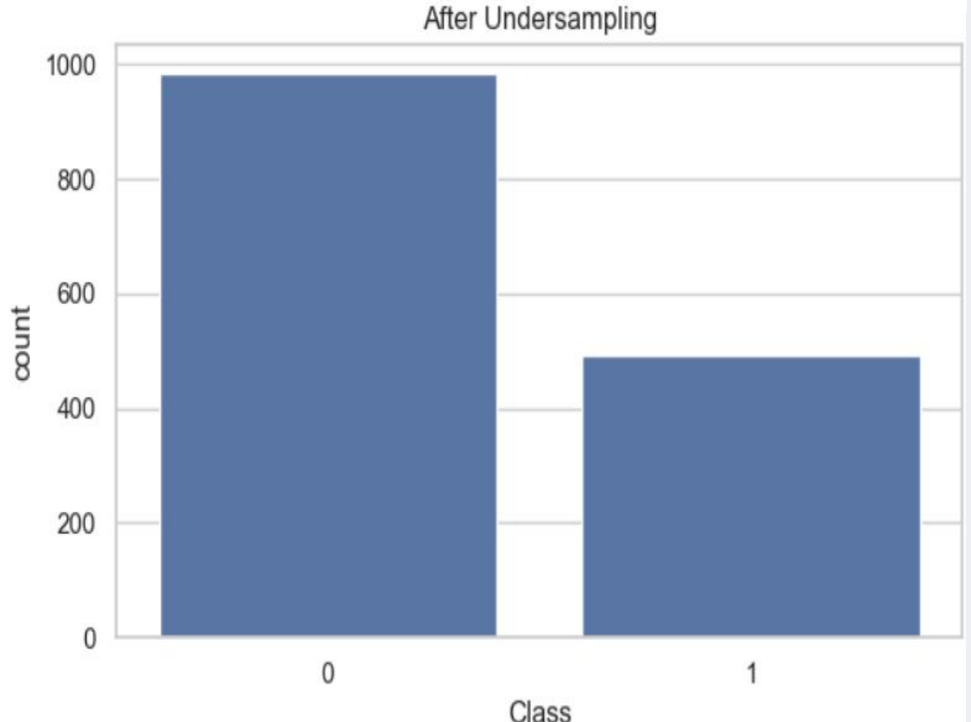
Visualizations:

- KDE Plot for transaction amount
- PCA for dimensionality reduction and fraud separation

Cleaning:

- Dropped missing `Class` values
- Dropped `Time` column
- Standardized `Amount` feature





⚖️ Handling Imbalance

Extreme Class Imbalance:

- Very few fraudulent transactions

Solution:

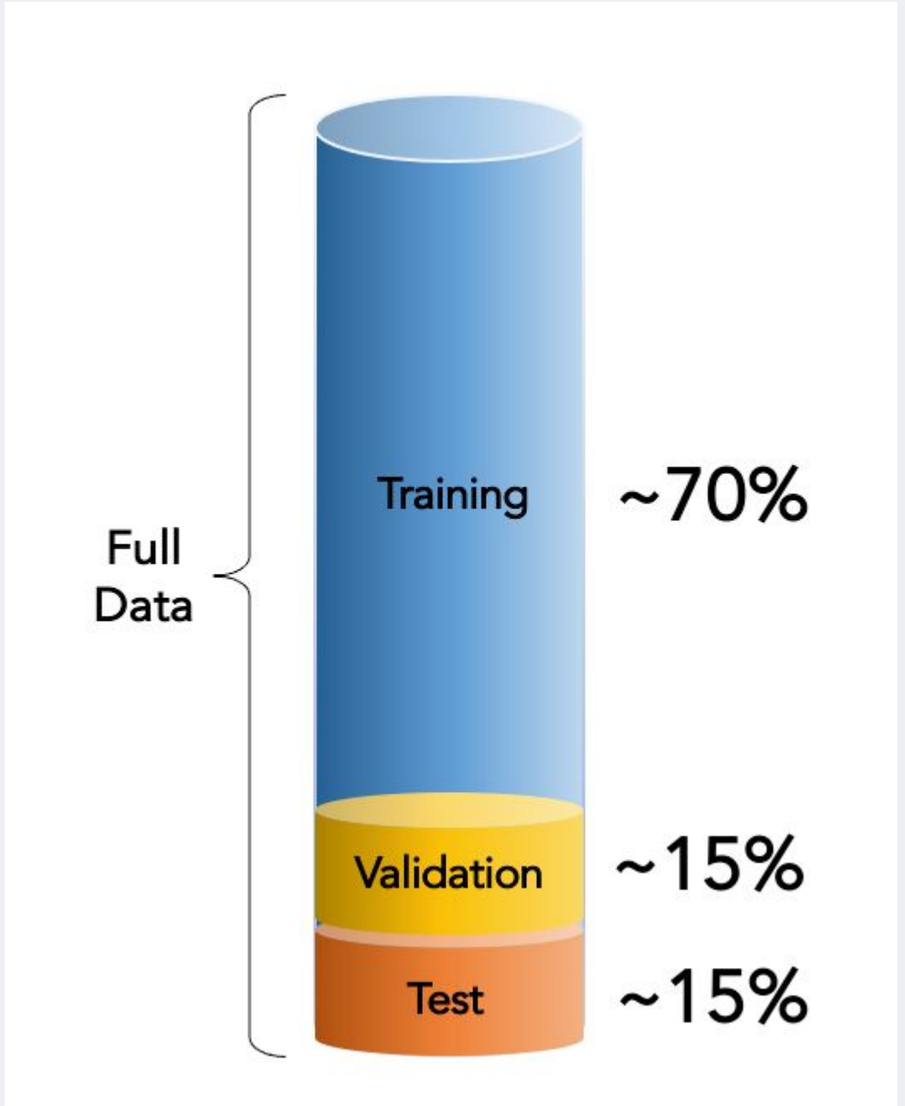
- Used **RandomUnderSampler** to balance dataset (50:50 ratio)

Visual Proof:

- Bar plots before and after resampling

Train-Test Split

- **80% Training**, **20% Testing**
- **Stratified Split** (maintains class ratio)



🤖 Model Selection

1. 🌲 **Random Forest Classifier**

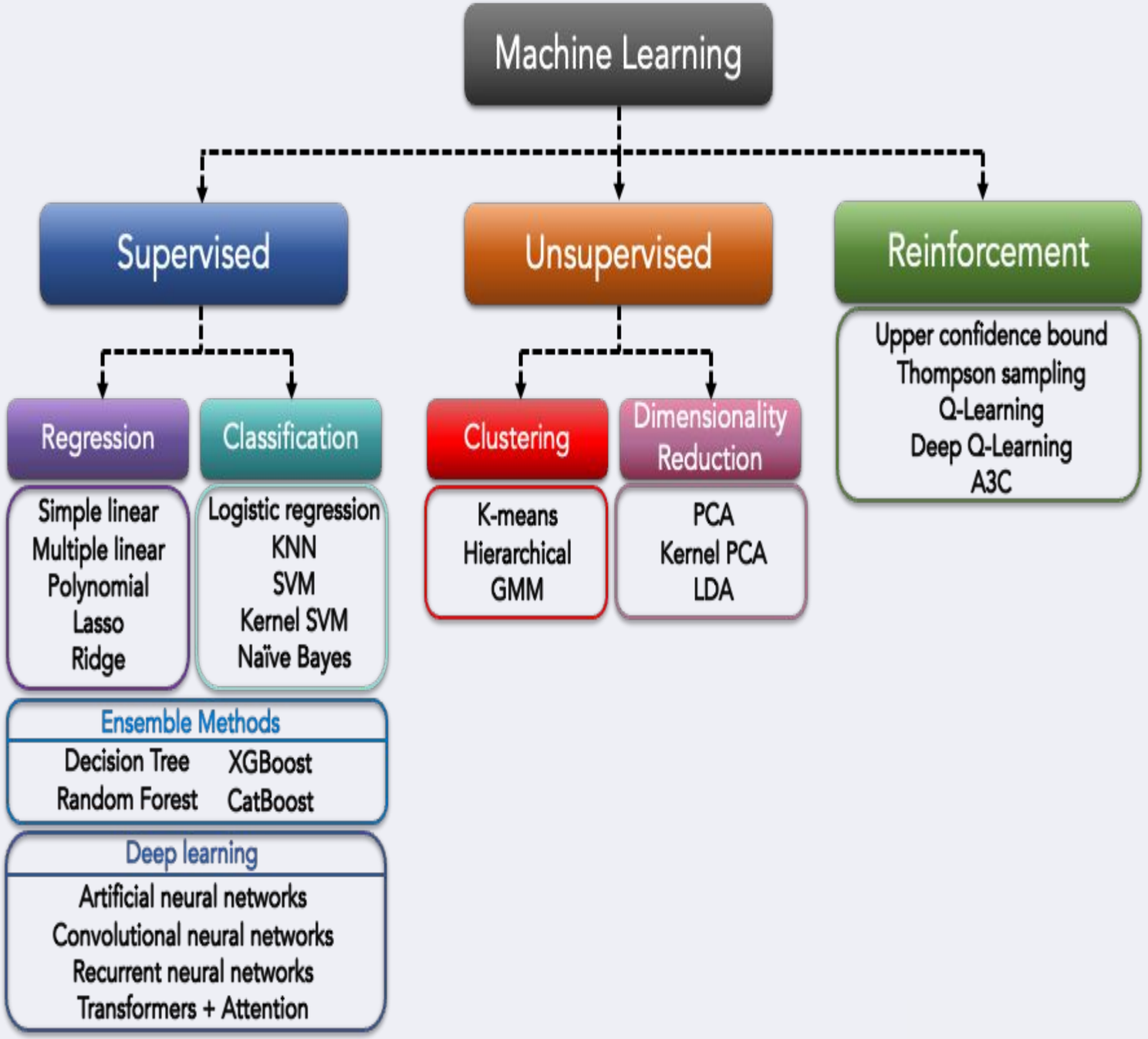
2. ⚡ **XGBoost Classifier**

1. 🌲 Random Forest Classifier

- `n_estimators=200`
- `random_state=42`

2. ⚡ XGBoost Classifier

- `n_estimators=300`
- `max_depth=6`
- `learning_rate=0.05`
- `eval_metric='logloss'`
- `use_label_encoder=False`



📊 Evaluation Metrics Used

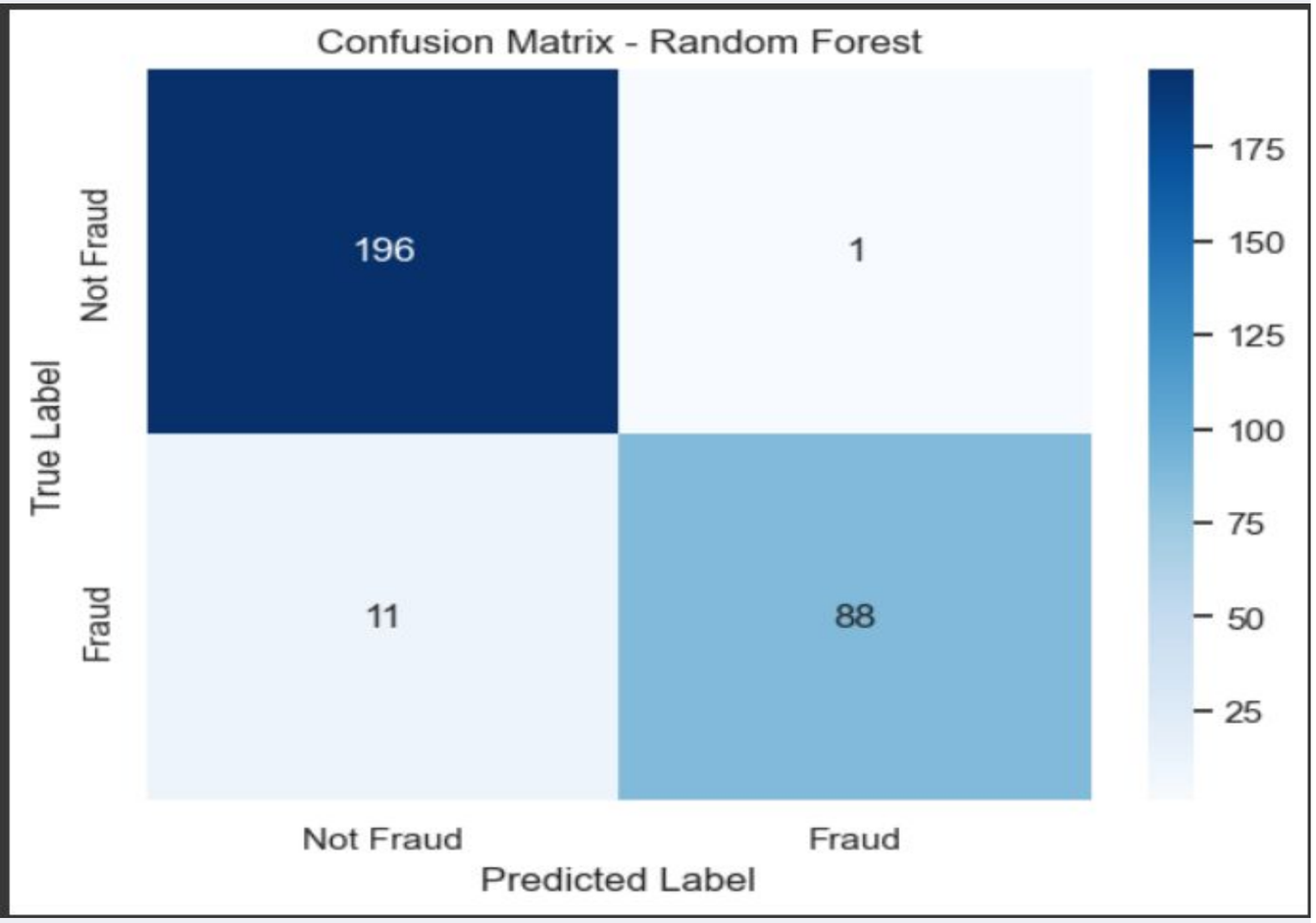
- **Accuracy**: Correct predictions over total predictions
- **Precision**: Correctly predicted frauds among predicted frauds
- **Recall**: Correctly predicted frauds among actual frauds
- **F1 Score**: Harmonic mean of Precision and Recall
- **Confusion Matrix**
- **ROC Curve & AUC Score**

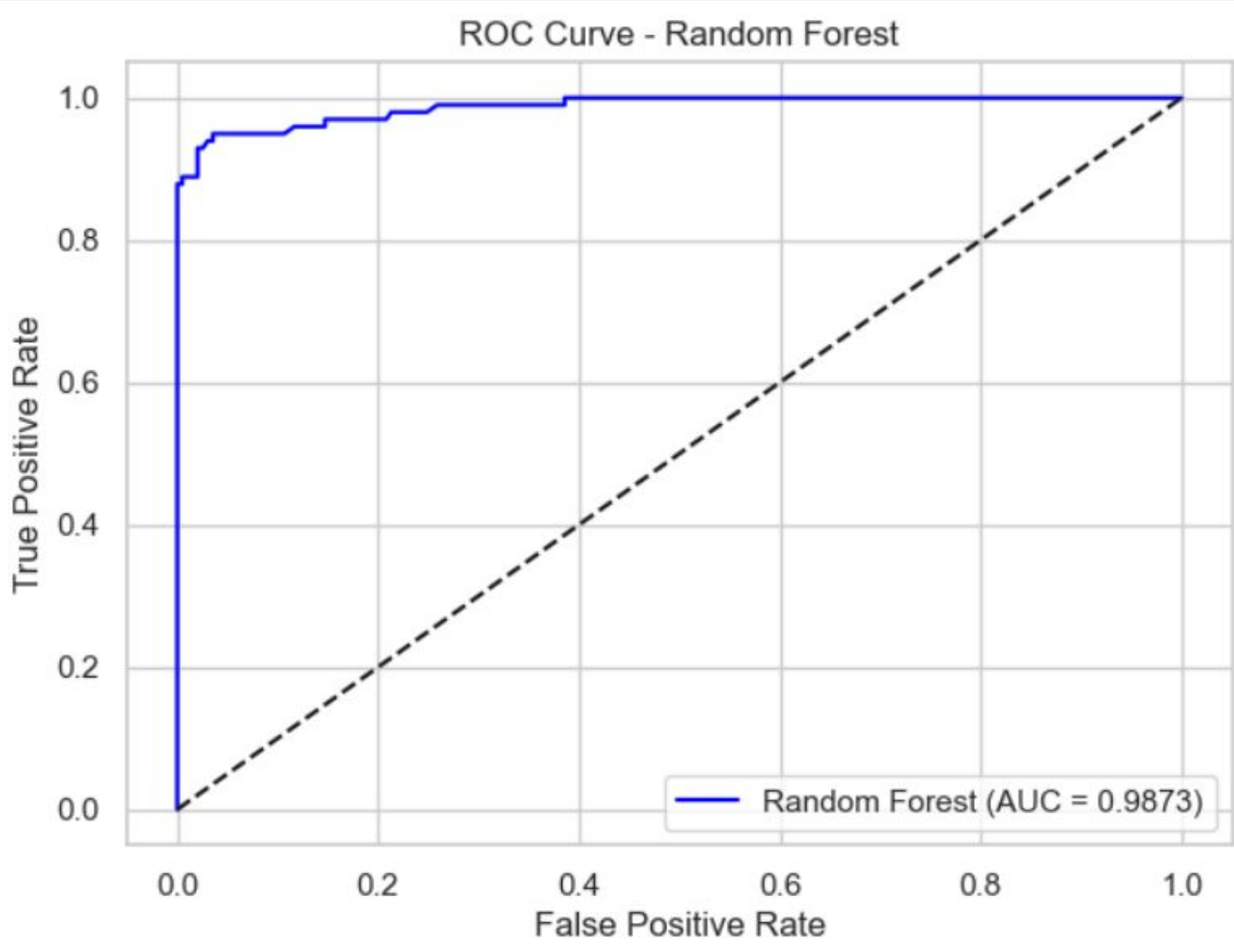
📈 Random Forest Results

📈 XGBoost Results

🏆 Conclusion

- **Random Forest**:
  - Excellent precision.
  - Slightly lower recall.
  - Good overall performance.





XGBoost:

- Higher recall and slightly better F1 score.

- Best choice for fraud detection where **catching frauds** is critical.