**Problem Statement**

Fraud detection is a classic example of a highly **imbalanced classification problem**, where fraudulent transactions form a very small proportion of the dataset. The challenge lies in correctly identifying these rare fraud cases (minority class) without being overwhelmed by false positives. Traditional evaluation metrics like accuracy are misleading in such scenarios because a model can achieve high accuracy simply by predicting the majority class (non-fraud) correctly.

The dataset was evaluated using several machine learning models, including:

* **RandomForestClassifier**
* **DecisionTreeClassifier**
* **XGBoostClassifier**
* **GradientBoostingClassifier**
* **LogisticRegression**

with different resampling techniques to handle class imbalance:

* **SMOTE** (Synthetic Minority Oversampling Technique)
* **Tomek Links** (undersampling of overlapping majority-class points)
* **SMOTE + Tomek** (combined approach)

Key metrics used for evaluation include **precision**, **recall**, **F1-score**, **F2-score**, **PR AUC**, and **ROC AUC**, with a focus on **recall and F2-score** because detecting fraudulent transactions (true positives) is far more critical than overall accuracy.

**Observations from model performance:**

| **Model** | **Resampling** | **Recall** | **Precision** | **F1-score** | **Notes** |
| --- | --- | --- | --- | --- | --- |
| RandomForestClassifier | SMOTE / SMOTE+Tomek | 0.444 | 0.085 | 0.143 | Balanced detection of frauds with moderate recall and acceptable false positives. |
| DecisionTreeClassifier | Tomek | 0.667 | 0.073 | 0.132 | Slightly higher recall but lower overall balance (F1). |
| RandomForestClassifier | Tomek | 0.889 | 0.052 | 0.099 | Very high recall but extremely low precision, predicting almost all cases as fraud → impractical. |
| XGBoost / GradientBoosting / LogisticRegression | Various | 0.0 | 0.0 | 0.0 | Failed to detect any fraud cases. |

**Insights:**

* Many models (XGBoost, GradientBoosting, LogisticRegression) failed to capture fraud instances entirely, indicating poor minority class detection.
* Models with very high recall but extremely low precision (RandomForest with Tomek) are not practical due to excessive false positives.
* A **trade-off between recall and precision** is required for operational feasibility, emphasizing the need for a **balanced model** with the highest F2-score.

**Conclusion**

Based on the evaluation:

✅ **Selected Model:** **RandomForestClassifier with SMOTE or SMOTE+Tomek**

**Reasons for selection:**

1. **Effective fraud detection:** It predicts actual fraud cases, unlike models that fail to detect any minority class instances.
2. **Balanced performance:** Achieves the highest **F2-score (~0.282)** among models that detect fraud, balancing recall and precision with emphasis on recall.
3. **Moderate recall (0.444):** Detects a reasonable proportion of frauds without generating an unmanageable number of false positives.
4. **Resampling advantage:** Using **SMOTE or SMOTE+Tomek** effectively addresses class imbalance, improving minority class representation and model learning.

**Recommendations for further improvement:**

* **Threshold tuning:** Adjusting the decision threshold can increase recall if business priority demands catching more frauds.
* **Ensemble methods:** Combining RandomForest with XGBoost may further improve detection performance and F2-score.
* **Cost-sensitive evaluation:** Penalize false negatives more heavily to prioritize fraud detection.
* **F2-score driven selection:** Continue using F2-score instead of accuracy for final model comparison to ensure minority class performance remains the primary focus.

**Summary:** RandomForestClassifier with SMOTE or SMOTE+Tomek provides the **best balance between detecting frauds and controlling false positives**, making it the optimal choice for this imbalanced fraud detection problem.