Credit Card Fraud Detection Report

Executive Summary

As a new player in the credit card industry, [Company Name] has positioned itself as one of the safest cards to use in the western United States. In line with this vision, we have developed a machine learning model to identify fraudulent credit card transactions. This model aims to prevent fraud by flagging potentially fraudulent transactions while minimising the risk of missing any real instances of fraud. The model has been built with the company's directive to err on the side of caution, ensuring we prioritise fraud detection even if it means some legitimate transactions are flagged as fraudulent.

Problem Overview

Credit card fraud is a growing concern for financial institutions. As a new company in the market, [Company Name] needs to demonstrate robust fraud detection capabilities to build customer trust and ensure the safety of cardholders. The goal of this project is to accurately predict fraudulent transactions from a dataset of credit card transactions, considering features like transaction amount, merchant information, and customer demographics.

The key business challenge is to balance fraud detection precision with minimising the number of false negatives—fraudulent transactions that go undetected. The company prioritises catching fraud at the cost of occasionally flagging legitimate transactions, as it's more critical to prevent losses than to inconvenience customers.

Key Insights from the Model

1. **Fraud Detection Accuracy:**
    - **Overall Accuracy:** The model achieved an impressive 99.9% accuracy, but the critical metrics for fraud detection—precision and recall—tell a more detailed story.
    - **Precision:** 83% – This means that 83% of the transactions flagged as fraudulent were truly fraudulent. This indicates the model is good at minimising false positives (flagging legitimate transactions as fraud).
    - **Recall:** 64% – This means that the model successfully identified 64% of all actual fraudulent transactions. While this is a solid result, there's room to further increase the recall and catch more fraud.
2. **Model Performance in the Context of Business Goals:**
    - The model aligns well with the company's goal of being cautious and prioritising fraud detection. Although we are flagging some legitimate transactions, the focus on maximising recall ensures that fewer fraudulent transactions go undetected.
3. **Handling Imbalanced Data:**

- Fraudulent transactions are rare, so we used SMOTE (Synthetic Minority Over-sampling Technique) to address the imbalance in the dataset. This helped improve the model's ability to detect fraud without being biased towards the larger number of non-fraudulent transactions.

Business Impact

- **Fraud Prevention:** By detecting fraudulent transactions early, the company can significantly reduce potential losses. While some legitimate transactions may be flagged, this approach ensures that the company is protecting cardholders from significant financial harm.

- **Customer Trust and Retention:** Customers are more likely to trust a credit card company that offers reliable fraud protection. By catching fraudulent transactions, the company strengthens its reputation as a safe and secure financial institution.

- **Operational Efficiency:** Automating fraud detection reduces the need for manual review of transactions, lowering operational costs and allowing the company to focus resources on other business-critical areas.

Recommendations

1. **Fine-Tuning the Model:**

    - **Increase Recall:** While the model performs well in terms of precision, the recall could be improved. One potential way to do this is by adjusting the classification threshold to flag more transactions as fraudulent. This would increase the number of frauds detected but might also lead to more false positives.

    - **Experiment with Different Algorithms:** While the Random Forest model is performing well, testing other machine learning algorithms (e.g., Gradient Boosting, XGBoost) could further improve the model's ability to detect fraud, especially when balancing precision and recall.

2. **Real-Time Deployment:**

    - **Deployment in Production:** The next step is to deploy the model in real-time systems to flag suspicious transactions instantly. This would allow the company to prevent fraud before it occurs and act quickly when a fraudulent transaction is flagged.

    - **Integration with Existing Systems:** The model should be seamlessly integrated into the company's transaction processing system, ensuring that flagged transactions are either blocked or flagged for review in real-time.

3. **Regular Model Updates:**

    - **Retraining with New Data:** Since fraudulent behavior evolves over time, it's important to retrain the model regularly with new transaction data. This ensures

the model adapts to changing patterns in fraud.

- **Monitor Model Performance:** Ongoing monitoring will be necessary to evaluate the model's performance in the real world. Adjustments can be made to ensure it continues to meet the company's fraud detection goals.

4. **Customer Communication Strategy:**

   - **Handling False Positives:** It's crucial to have a clear customer communication strategy in place for cases where legitimate transactions are flagged. Transparency about why transactions were flagged and how customers can resolve these issues will help maintain customer trust.

   - **Improved Customer Experience:** As fraud is prevented, the company should consider offering customers extra tools or support to manage their accounts in the event of a flag, ensuring a positive customer experience even when issues arise.

2. **Continuous Monitoring:**
   - Even after implementation, continuous monitoring of the model's performance is crucial. Performance metrics should be regularly reviewed to ensure that the system continues to meet fraud prevention goals. This will help identify areas for improvement, such as fine-tuning thresholds or exploring new features that could enhance fraud detection.

Conclusion

The fraud detection model developed for [Company Name] performs well in detecting fraudulent transactions, with a high degree of accuracy and a focus on minimizing false negatives. By prioritizing recall, the company is better positioned to protect its customers and reduce the risk of financial loss. Moving forward, the model should be fine-tuned, deployed for real-time detection, and regularly updated to keep pace with evolving fraud patterns.

The next steps involve real-time integration of the model into the company's systems, improving recall for better fraud detection, and ensuring clear communication with customers in case of false positives.

Model and Data Access

The model, along with its code and raw data, is available in the following repository for further review:

https://github.com/luluhsu727/data-science-portfolio/tree/main/Fraud%20Detection

---