Credit Card Fraud Detection

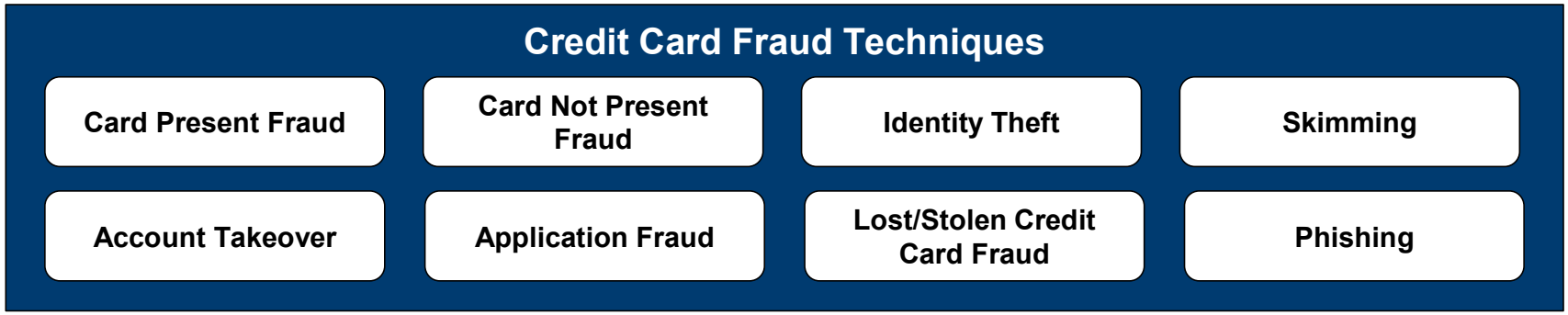
Leveraging Machine Learning to Protect Digital Payments

Arup Chakraborty Fernando Calderon Jabali Shah

GitHub: https://github.com/teslanando/AAI510_ML_Final_Project

The Growing Threat of Credit Card Fraud

- Digital payments are evolving, but so are cyber criminals.
- 114,348 reported cases of credit card fraud in 2023(Federal Trade Commission).
- Detecting fraud is challenging given the high volume of daily transactions.



Impact on Business

Financial Loss

- Globally credit card fraud causes over ~$30B in annual losses
- In the US, it results in ~$12B in losses each year
- Businesses often pass on the costs of fraud prevention and losses to consumers through higher prices
- Victims of fraud may face financial loss, emotional distress, and the inconvenience of resolving fraudulent charges.

Damage to Reputation

- Target Data Breach (2013): The breach significantly damaged Target's reputation, resulting in lost sales and a dip in stock prices
- Equifax Data Breach (2017): The breach led to widespread criticism, loss of consumer trust, and significant financial losses due to legal fines and settlements.
- Reputational damage results in loss of customer trust, decreased loyalty and competitive disadvantage

Legal & Operational Consequences

- Handling fraud cases can be time-consuming, requiring resources to investigate and resolve disputes
- Rapid advancements in fraud tactics and technologies require continuous innovation and investment in sophisticated fraud detection systems and algorithms.
- Balancing the need for robust security with user convenience in authentication methods is an ongoing challenge for businesses and financial institutions

Credit Card Fraud Detection techniques

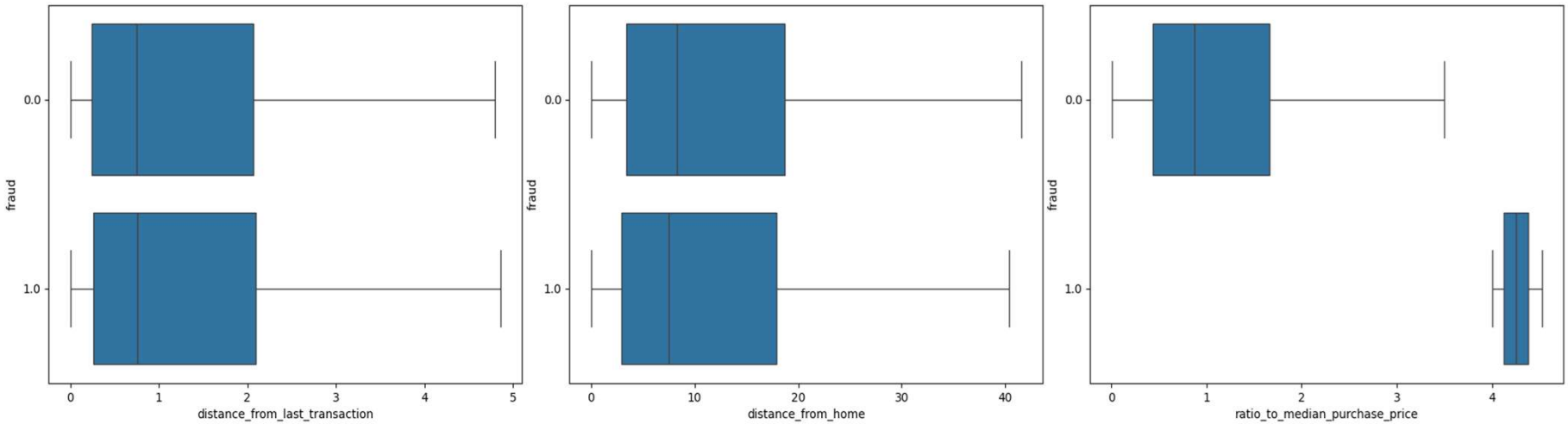
Traditional Fraud Detection

- Traditional credit card fraud detection relies on predefined rules and patterns but often lacks to keep pace with evolving tactics, there are primarily two approaches that stand out as follows:
- Rule-Based Systems:
  - Relies on predefined rules and thresholds set by experts.
  - Simple and easy to implement.
  - High false positive rate and limited adaptability.
  - Requires constant updating to keep up with new fraud techniques.
- Statistical Methods:
  - Uses statistical techniques to identify outliers and anomalies.
  - Limited adaptability to new fraud patterns.
  - Requires significant statistical expertise.

Machine Learning based detection

- Fraud detection with machine learning leverages the power of advanced algorithms for identifying patterns of suspicious activity efficiently
- Supervised Learning:
  - Uses labeled datasets to train predictive models.
  - High accuracy and adaptability to new fraud techniques.
  - Reduces false positives significantly.
  - Requires large labeled datasets and potential for overfitting.
- Unsupervised Learning:
  - Detects anomalies without labeled data.
  - Useful when labeled data is unavailable.
  - Can uncover new fraud patterns.
  - Lower accuracy and higher false positive rate.
- Semi-Supervised Learning:
  - Complex to implement but provides benefits of both supervised and unsupervised learning

Characteristics of Fraudulent Transactions

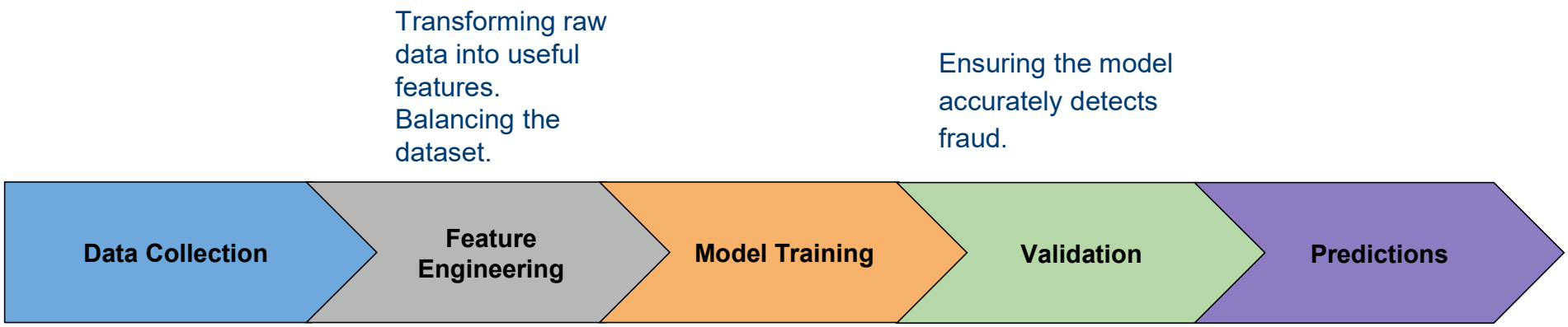


- Fraudulent transactions often seem to occur slightly further from the last transaction compared to normal ones.
- Fraudulent transactions tend to happen further from the cardholder's home as compared to actual transactions.
- These transactions are generally more costly than typical ones.

Leveraging Machine Learning for Fraud Detection

Machine Learning and Its Capabilities	Patterns and Anomalies Detection in Transaction	Real-Time Fraud Detection to Minimize Losses
Machine learning (ML) enables computers to learn from data and make predictions. ML can recognize patterns, adapt to new data, and improve over time.	Analyzes vast amounts of transaction data to identify normal and fraudulent behaviors. Flags anomalies for further investigation.	Processes transactions instantly to detect and prevent fraud. Reduces financial losses and limits customer impact.

Developing the Fraud Detection Model



Model Performance

1. **Random Forest**:
   - **ROC-AUC**: 0.99999996
   - Extremely high performance, indicating it can effectively distinguish between fraudulent and non-fraudulent transactions.
2. **XGBoost**:
   - **ROC-AUC**: 0.99990074
   - Nearly matches the Random Forest in performance, suggesting it is also a very robust choice for fraud detection.
3. **Logistic Regression**:
   - **ROC-AUC**: 0.97958758
   - Performs well but not as strong as the other models, making it a simpler, more interpretable option with slightly less efficacy.
4. **SVM**:
   - **ROC-AUC**: 0.9990844632206718
   - Another strong contender, demonstrating high accuracy and robustness.

Considerations for Model Selection:

- **Interpretability**: Logistic Regression offers more transparency and is easier to interpret, which can be valuable in understanding the model's decisions and gaining regulatory approval.
- **Performance**: Random Forest, XGBoost, and SVM provide excellent detection capabilities with high ROC-AUC scores, making them suitable for maximizing detection rates.
- **Complexity and Resources**: More complex models like Random Forest and XGBoost require more computational resources for training and might be harder to deploy and maintain compared to simpler models like Logistic Regression.
- **Overfitting**: The extremely high scores suggest potential overfitting. It's crucial to validate these models in real-world settings and possibly regularize or tune them further to ensure they generalize well to new data.

Business Benefits

Significant Reduction in Financial Losses

- **Drastically reduces fraud-related financial losses.**
- **Machine learning models detect fraudulent transactions in real-time, saving millions annually.**

Improved Customer Trust and Loyalty

- **Customers feel safer knowing their transactions are protected by advanced fraud detection systems.**
- **Increased customer satisfaction, retention, and positive word-of-mouth, driving long-term loyalty.**

Compliance with Regulatory Requirements

- **An effective ML-based fraud detection system helps meet GDPR, CCPA, PCI-DSS compliance requirements by securing customer data and preventing breaches.**
- **This helps avoid penalties and maintaining reputation**

Managing Risks in ML-Based Fraud Detection

Business Risks

- **False Positives:** Mitigate with continuous model improvement and customer feedback integration.
- **False Negatives:** Enhance model sensitivity and update regularly.
- **Operational Disruptions:** Gradual integration with robust testing and IT support.

Ethical Risks

- **Bias and Fairness:** Regular audits, diverse training data, fairness constraints.
- **Transparency:** Implement explainable AI techniques.
- **Customer Trust:** Clear communication of benefits and privacy measures.

Regulatory Risks

- **Data Privacy:** Ensure compliance with laws (GDPR, CCPA, PCI DSS), regular audits.
- **Data Security:** Advanced encryption, multi-factor authentication, continuous monitoring.
- **Regulatory Changes:** Stay informed, maintain flexibility, involve legal experts.

Recommendations

1. **Deployment**: Given the high performance of Random Forest and XGBoost, they are strong candidates for deployment. SVM can also be considered if the performance difference is negligible in practical applications.
2. **Monitoring**: Implement a robust monitoring system to track model performance and update models as new transaction data becomes available.
3. **Validation**: Continuously validate the models against new, real-world data to ensure they maintain high performance and generalize well beyond the training data.
4. **Combination Approach**: Consider using an ensemble of these models to leverage the strengths of each and further improve detection accuracy.

By carefully evaluating the trade-offs between performance, interpretability, and resource requirements, the best model for this specific needs can be selected and deployed effectively to mitigate financial fraud.

Acknowledgements

Professor: Mirsardar Esmaeili

Program: Applied Artificial Intelligence