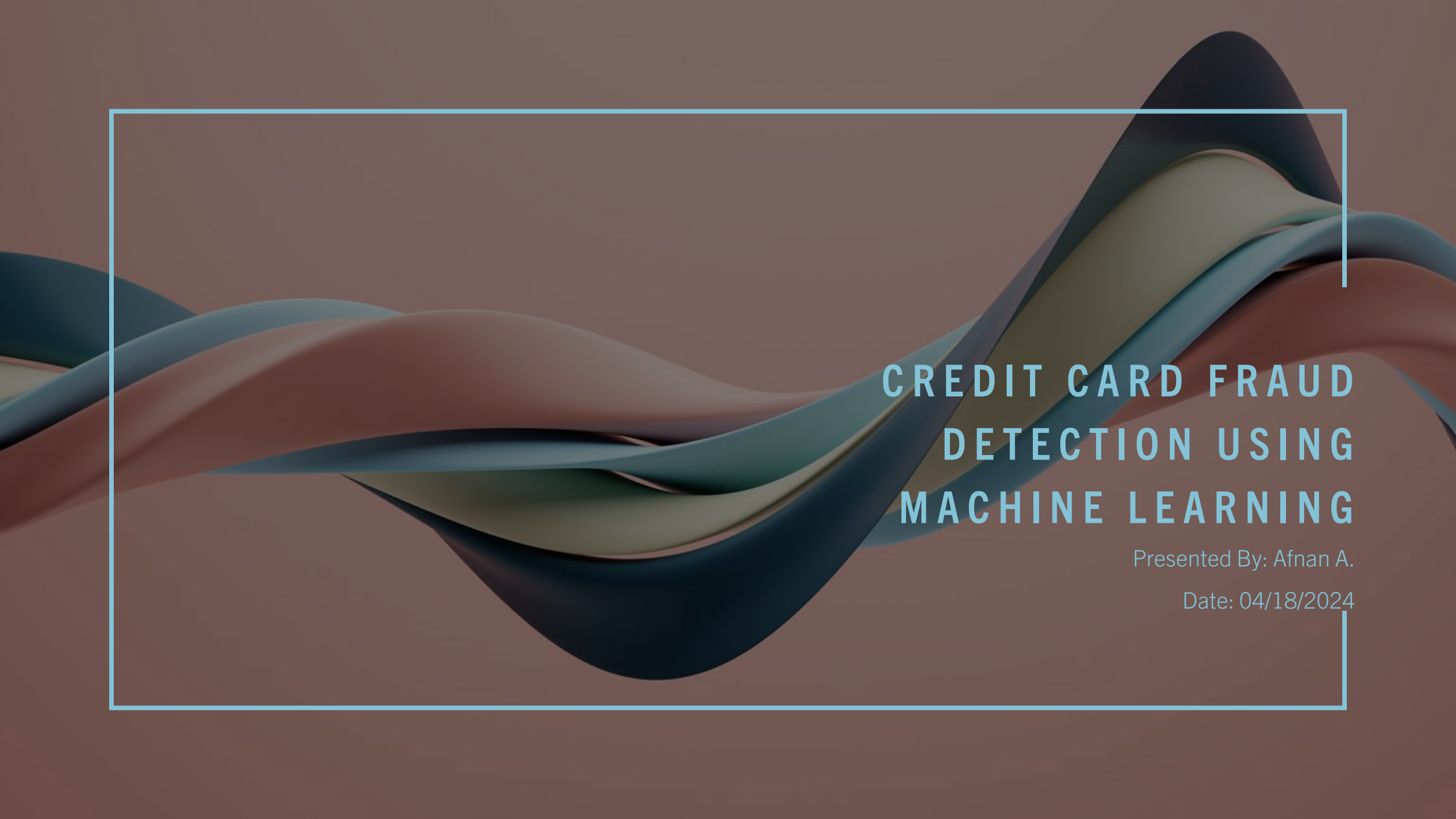


CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

Presented By: Afnan A.

Date: 04/18/2024

CREDIT CARD FRAUD DETECTION PROJECT

- Background: Credit card Fraud significantly impacts financial stability and consumer trust. With advancements in technology, the incidence of sophisticated fraudulent activities has risen, especially in the western United States.

- Objective: Apply machine learning techniques to detect fraudulent activities within credit card transactions with a target detection rate of at least 95% accuracy.

PROJECT OVERVIEW AND DATA INSIGHTS

- Scope:
  1. Data Cleaning: Handling missing values, outliers, and standardizing features.
  2. Data Exploration: Analyzing transaction patterns and features influencing fraud detection.
  3. Model Training: Utilized advanced machine learning techniques to develop the fraud detection model.
  4. Evaluation: Assessed the model's performance to ensure it meets our accuracy targets.
- Results:
  - Successfully developed a fraud detection system that meets our goal of 95% accuracy.

2

DATA WRANGLING & EDA

SUMMARY AND INTEGRITY CHECK

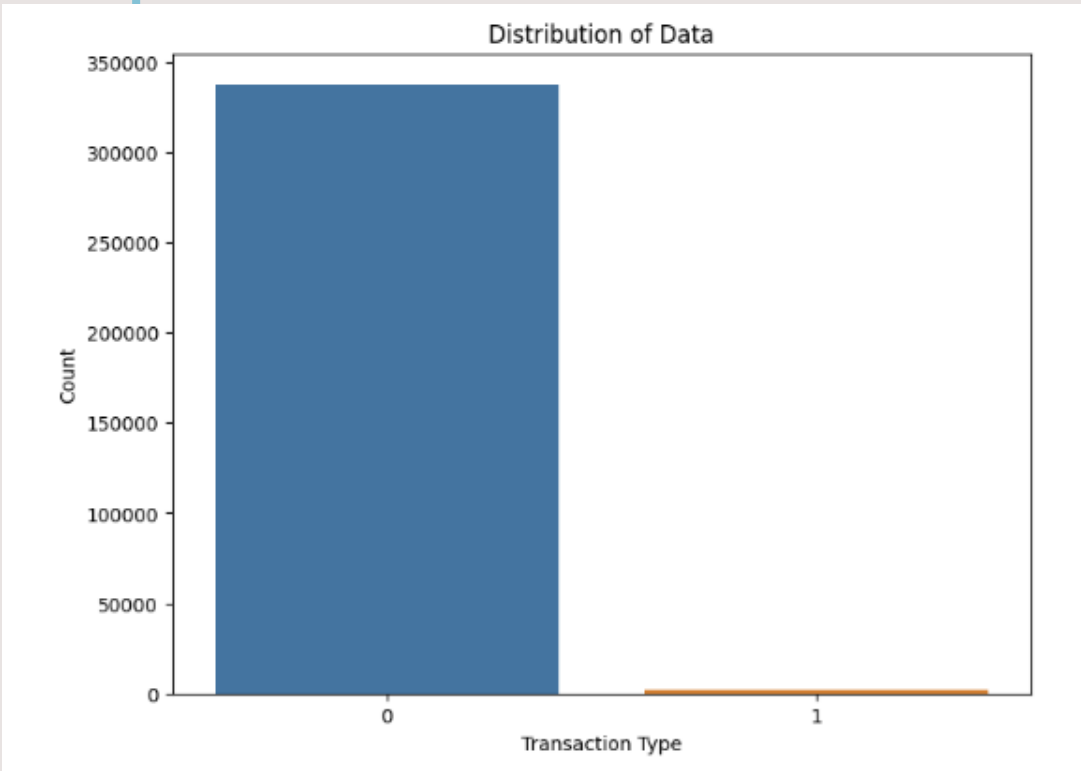
- No missing values: Dataset with 339,607 entries and 15 columns verified for completeness.
- Descriptive Statistics:
  - Numerical data ranges significantly, e.g., transaction amounts vary widely.
  - Categorical data shows diversity in merchants, jobs, and categories.

```
Out[72]:
trans_date_trans_time    0
merchant                 0
category                 0
amt                      0
city                     0
state                    0
lat                      0
long                     0
city_pop                 0
job                      0
dob                      0
trans_num                0
merch_lat                0
merch_long               0
is_fraud                 0
dtype: int64
```

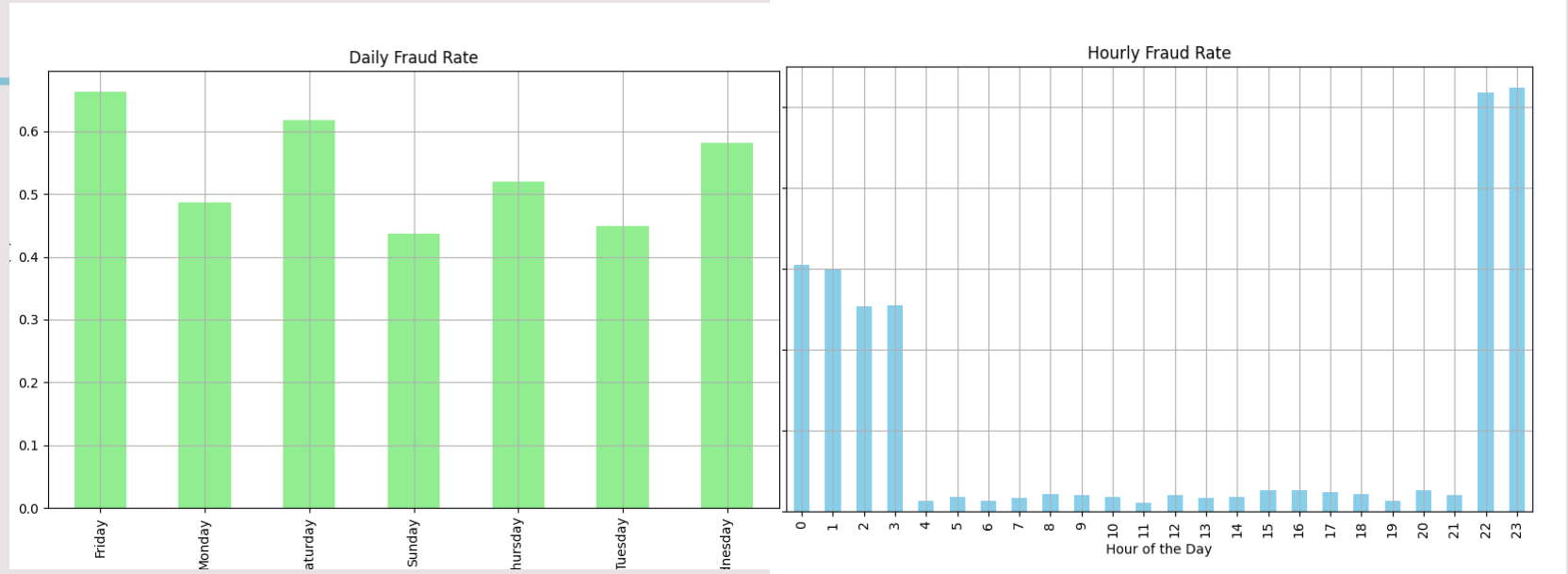
No missing data

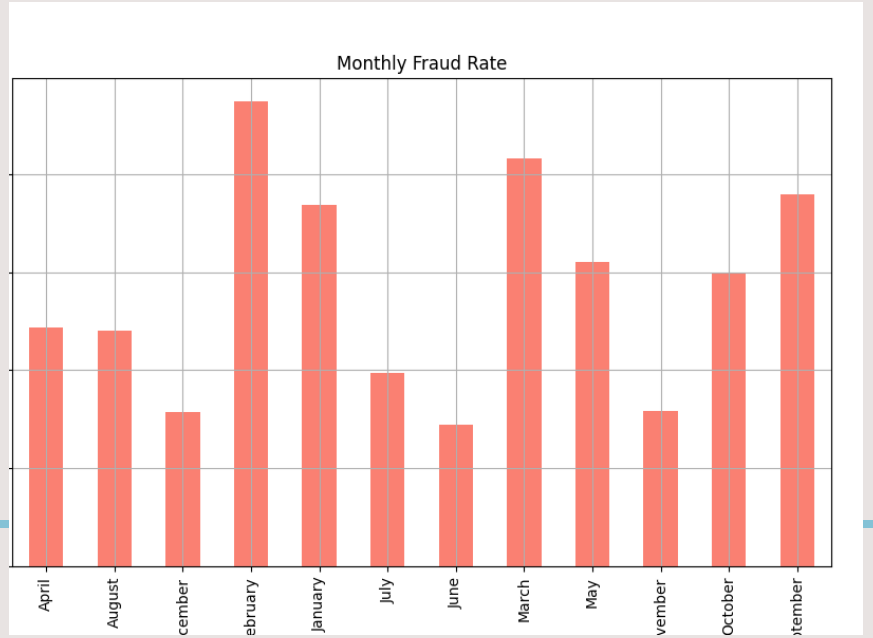
TRANSACTION DISTRIBUTION

- Class Imbalance: Legitimate transactions greatly outnumber fraudulent cases, indicating a need for class imbalance strategies in model training.



ANALYZING FRAUD TRENDS OVER TIME





3

MODELLING

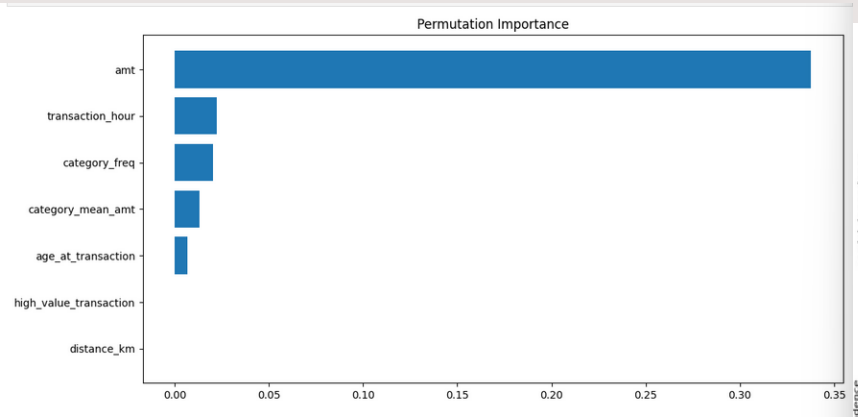
MODEL PERFORMANCE AND BUSINESS IMPACT

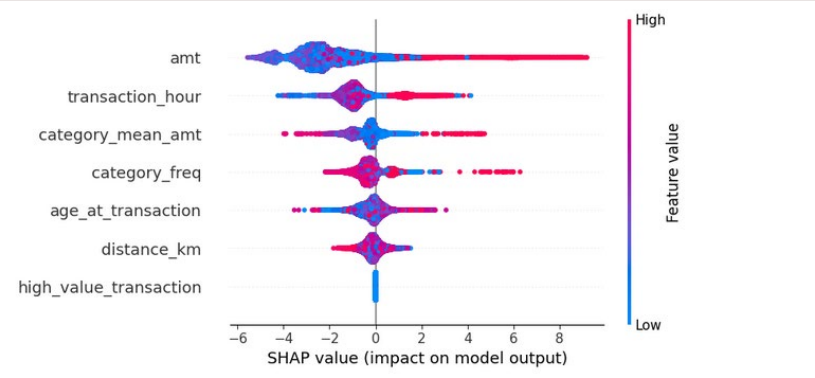
- Model Choice
  - Selected XGBoost due to its excellent balance of speed and accuracy, proving to be highly effective in handling unbalanced datasets typical of fraud detection scenarios.
- Business Impact:
  - Enhanced Security: The improved detection rate minimizes potential financial losses from fraud, protecting both the company's assets and customer transactions.
  - Increased Trust: High accuracy in fraud detection boosts customer confidence, reinforcing our reputation as a safe and trustworthy service provider.
  - Operational Efficiency: Streamlining fraud detection processes reduces the need for manual review, allowing resources to be allocated to other critical areas of operation.

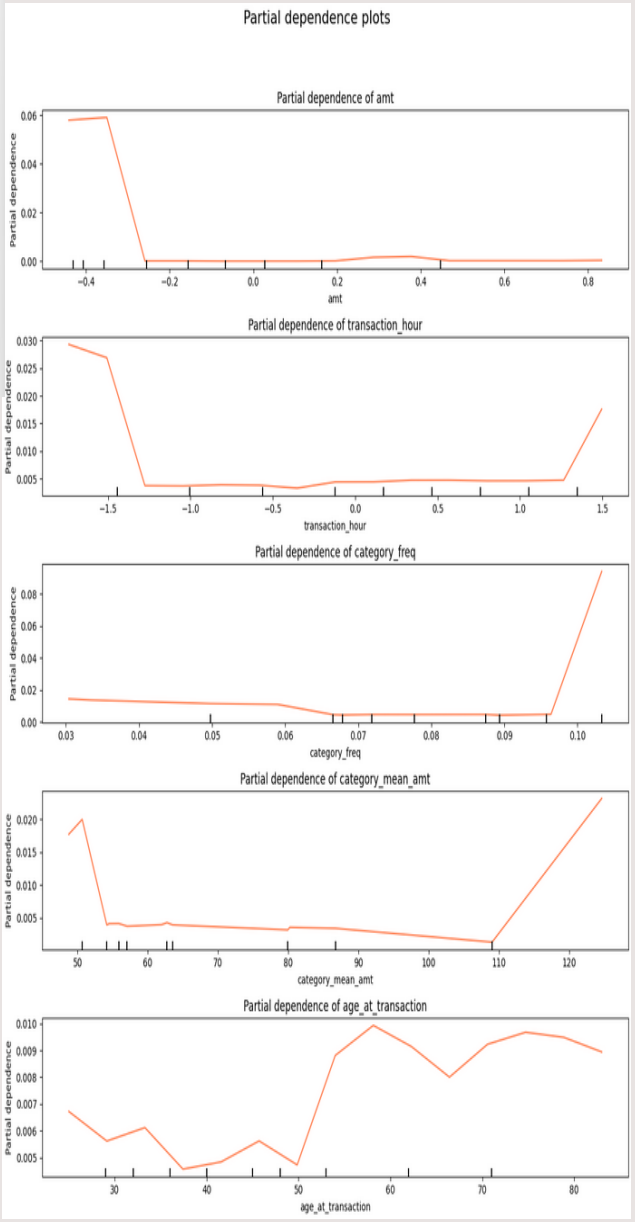
4

Decoding the Predictive Patterns

MODEL INSIGHTS AND INTERPRETATION







- Permutation Importance: Highlights 'amt' and 'transaction_hour' as critical predictors in the XGBoost model, confirming the pivotal role of transaction amount and timing in fraud detection.
- Partial Dependence Plots: Demonstrates how features like 'amt' exert a non-linear effect on fraud predictions, indicating complex relationships between transaction values and fraud likelihood.
- SHAP Values: Offers detailed insights into the impact of individual features, with 'amt' having a substantial and variable influence, thus validating its importance in the model's decision-making process.

5
CONCLUSION

CONSIDERATIONS

- Synthetic Data Limitations: The current model is based on simulated data, which might not capture all real-world complexities and behavioral patterns.
- Regional Specificity: The data and model are localized to the Western United States, which could limit applicability to different geographic areas with varying fraud dynamics.
- Data Privacy Compliance: Continuous monitoring and updating of data privacy measures to ensure compliance while utilizing data effectively for fraud detection.

FUTURE DIRECTIONS

- Real-World Application: Plan to deploy the model in a controlled real-world environment to validate its practical effectiveness and refine it based on actual transaction data.
- Broader Data Integration: To enhance generalization, consider incorporating data from various regions and countries, taking into account different types of fraud that may not have been represented in the synthetic dataset.
- Continuous Model Evolution: Implement mechanisms for the model to evolve with new fraud tactics using techniques like online learning or reinforcement learning to adapt to emerging fraud patterns.

CONCLUSION

Achievements:

- Objective Achieved: The project successfully met its goal of detecting fraudulent transactions with at least 95% accuracy using the XGBoost model.
- Impactful Data Handling: Comprehensive data cleaning and feature engineering laid a solid foundation for effective modeling, demonstrating the critical role of detailed transaction data in fraud detection.
- Strategic Feature Selection: A structured approach to feature selection ensured the model's high accuracy, proving the importance of choosing the right features.

Future Outlook:

- Commitment to Excellence: We remain dedicated to enhancing our fraud detection capabilities, continuing to invest in technology and processes that protect our customers and business.
- Adaptive and Responsive: The project will continue to evolve in response to new fraud tactics and changing market conditions, ensuring it remains at the forefront of fraud detection technology.

- Any Questions?