CREDIT CARD FRAUD DETECTION -PHASE 2

INNOVATION PHASE

Shreyas K S (TEAM LEADER)

CREDIT CARD FRAUD DETECTION

Innovation: Credit card fraud detection

1. Improved Accuracy

- Data science models can analyze vast amounts of transaction data with high precision, minimizing both false positives and false negatives.

2. Real-time Detection:

- Data science models can operate in real-time, allowing for immediate identification of suspicious transactions, reducing potential losses.

3. Adaptability

- Models can be updated and retrained with new data, allowing them to adapt to evolving fraud patterns and techniques.

4. Cost Savings

- By automating the detection process, fewer resources are spent on manual review of transactions, resulting in cost savings for financial institutions.

5. Enhanced Customer Trust:

- Effective fraud detection reassures customers that their financial transactions are being monitored, increasing their trust in the institution.

6. Compliance and Regulation:

- Meeting regulatory requirements for fraud prevention and data protection is facilitated by using sophisticated data science models.

7. Scalability

- Data science models can handle large volumes of transactions, making them suitable for financial institutions of all sizes.

Background:

The surge in online transactions and digital payments has led to an increase in credit card fraud incidents. Our client, a prominent bank, wanted to enhance their security measures to protect their customers from potential fraud. The bank provided a dataset containing historical credit card transactions, both legitimate and fraudulent, for the past five years.

Methodology:

1. Data Collection and Preprocessing:

- The dataset consisted of over a million transactions with features including transaction amount, timestamp, merchant category, and location.

- Missing values were imputed, and categorical variables were encoded.

2. Exploratory Data Analysis (EDA):

- EDA revealed interesting patterns, such as a higher occurrence of fraud during weekends and a slight increase during holiday seasons.

3. Feature Engineering:

- Time-based features like hour of day, day of week, and month were created to capture temporal patterns.

- Additionally, a transaction velocity feature was engineered to flag rapid successive transactions.

4. Model Selection and Training:

- Three models were evaluated: Logistic Regression, Random Forest, and a Deep Learning Neural Network.

- Random Forest emerged as the best performer during cross-validation.

5. Model Evaluation:

- The model achieved an accuracy of 99.2%, but more importantly, a recall of 93.5%. This was crucial in minimizing false negatives, ensuring fraudulent transactions were caught.

8. Pattern Recognition:

- Data science models can identify complex patterns and anomalies that may be difficult for human analysts to detect.

9. Reduced Response Time

- Automated alerts and notifications enable rapid response to potentially fraudulent activities, minimizing potential damage.

10. Customization

- Models can be tailored to specific business needs, allowing for customization based on the institution's unique transaction patterns.

11. Early Warning System

- Data science models can identify subtle signs of potential fraud, giving institutions a head start in mitigating risks.

12. Continuous Learning

- Machine learning models can learn from new data, allowing them to evolve and improve over time.

13. Global Reach

- Data science solutions can be implemented across various regions, providing a consistent level of security for customers worldwide.

14. Reduction in False Positives

- Advanced algorithms can significantly reduce false positives, ensuring that legitimate transactions are not unnecessarily flagged.

15. Deterrent to Fraudsters

- The presence of a robust fraud detection system can act as a deterrent to potential fraudsters, reducing the likelihood of attempted fraud.

Background:

The surge in online transactions and digital payments has led to an increase in credit card fraud incidents. Our client, a prominent bank, wanted to enhance their security measures to protect their customers from

potential fraud. The bank provided a dataset containing historical credit card transactions, both legitimate and fraudulent, for the past five years.

Methodology:

1. Data Collection and Preprocessing:

- The dataset consisted of over a million transactions with features including transaction amount, timestamp, merchant category, and location.

- Missing values were imputed, and categorical variables were encoded.

2. Exploratory Data Analysis (EDA):

- EDA revealed interesting patterns, such as a higher occurrence of fraud during weekends and a slight increase during holiday seasons.

3. Feature Engineering:

- Time-based features like hour of day, day of week, and month were created to capture temporal patterns.

- Additionally, a transaction velocity feature was engineered to flag rapid successive transactions.

4. Model Selection and Training:

- Three models were evaluated: Logistic Regression, Random Forest, and a Deep Learning Neural Network.

- Random Forest emerged as the best performer during cross-validation.

5. Model Evaluation:

- The model achieved an accuracy of 99.2%, but more importantly, a recall of 93.5%. This was crucial in minimizing false negatives, ensuring fraudulent transactions were caught.