

Supervised Machine Learning for Detecting Online Transaction Fraud: A Data-Driven Approach



-LAKKISETTY
SAIVISHNU VARDHAN
2412185

Machine Learning and Fraud Detection

RESEARCH BACKGROUND

- Detecting fraud in financial transactions is a crucial responsibility for regulatory agencies, corporations, and banks. Machine learning (ML) models have been widely used for fraud detection, improving detection accuracy over traditional rule-based methods.
- However, issues including high false positive rates, unbalanced data, and a lack of real-time decision-making plague many of the ML-based fraud detection systems now in use.
- In order to improve fraud detection efficiency and accuracy, this study investigates supervised machine learning algorithms.

RESEARCH PROBLEM

- Conventional methods of detecting fraud rely on rule-based algorithms, which have trouble identifying emerging fraud trends.
- Although ML models increase accuracy, they are difficult to interpret, which undermines financial organizations' credibility.
- Standard ML models are less successful when there is a class imbalance (fraud occurrences are few relative to genuine transactions).
- While accuracy is a major focus of current research, real-time fraud detection and model transparency are still issues.

RESEARCH AIM & OBJECTIVES

Aim:

- Develop a Supervised ML-based fraud detection model that balances accuracy, interpretability, and real-time applicability.

Objectives:

- To Improve fraud detection accuracy using ML models.
- To handle imbalanced datasets using SMOTE & cost-sensitive learning.
- To Implement real-time fraud alerts for financial institutions.

LITERATURE REVIEW

Existing Research on Fraud Detection Using Machine Learning:

- Patel et al. (2021) [1]: Random Forest and logistic regression models increase the precision of financial transaction fraud detection.
- According to Sharma et al. (2023) [2], changing fraud trends provide difficulties for real-time fraud detection.
- The Jadhav group (2022) [3] – Cost-sensitive and SMOTE learning address class imbalance in fraud detection.
- According to Verma et al. (2023) [4], class-weight tweaking enhances fraud detection in infrequent fraud situations.

•Key Research Findings:

Fraud detection works well with supervised machine learning models. Model performance is impacted by class imbalance, necessitating specific handling methods. The growth of fraud patterns and delay make real-time fraud detection difficult.

LITERATURE REVIEW (CONT.)

Identified Research Gaps:

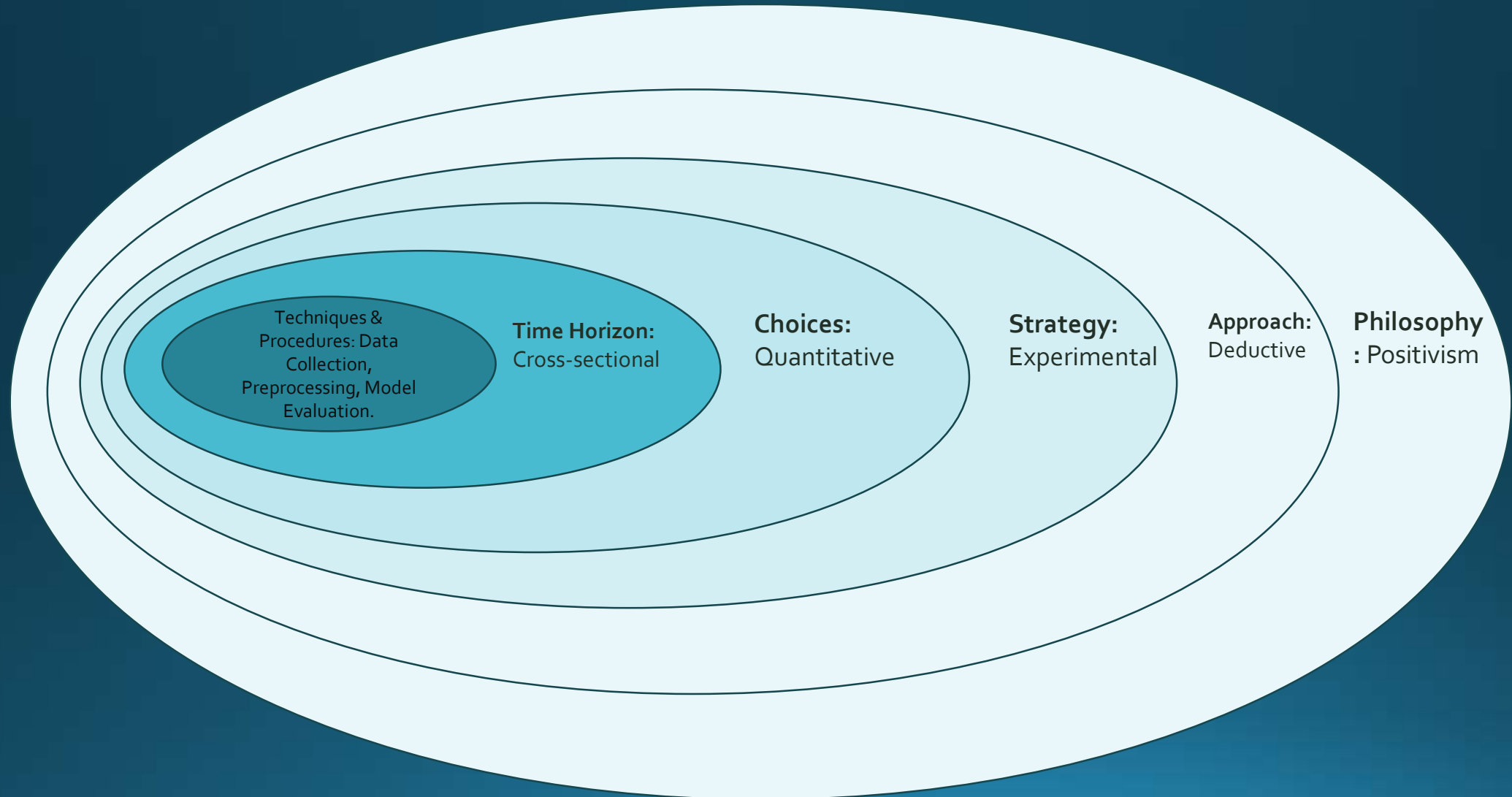
Lack of interpretability in ML models – Financial organizations have a hard time making decisions because of the black-box nature.

Need for real-time fraud detection -Instead of making predictions in real time, current machine learning algorithms rely on batch processing.

Handling fraud pattern evolution- Because fraudsters are always changing, it is essential to create adaptive machine learning models.

RESEARCH METHODOLOGY

RESEARCH ONION FRAMEWORK



RESEARCH METHODOLOGY (CONT.)

Philosophy (Positivism):

- Focuses on **objective data-driven research** using measurable outcomes.

Approach (Deductive):

- Uses **existing fraud detection theories** to develop and test hypotheses.

Strategy (Experimental):

- Involves **testing multiple ML models** on fraud datasets to find the best-performing one.

Choices (Quantitative):

- Uses **structured numerical data** (transaction amounts, timestamps, fraud labels) for analysis.

Time Horizon (Cross-sectional):

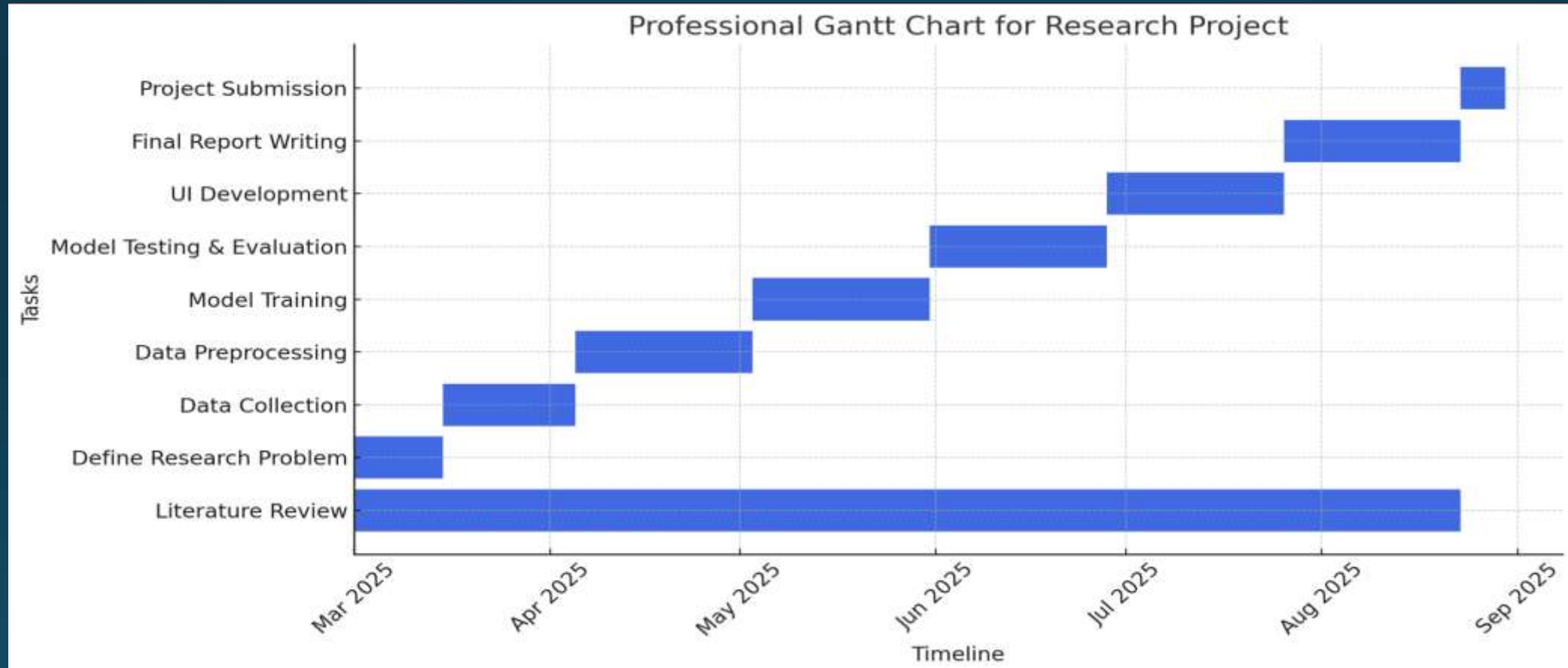
- Analyzes **historical fraud data** from a specific period to train models.

Techniques & Procedures:

- Includes **data collection, preprocessing, model evaluation, and fraud detection performance testing.**

PROJECT PLANNING & RESOURCES

- Gantt chart



- This is the proposed Gantt Chart for this project.

RESOURCES

Hardware: HPC, Cloud (AWS/GCP), Local Storage.

Software & Tools: Python (ML libraries), PostgreSQL, Tableau/Power BI.

Datasets: Kaggle, PaySim, Banking Transactions.

PROJECT IMPACT & LIMITATIONS

Impact:

Reduces financial losses from banking and online transactions [1].

Improves fraud detection accuracy using machine learning models [2].

Enables real-time fraud detection for secure transactions [3].

Limitations:

Data Availability Issues – Fraud cases are rare, making training difficult [4].

Computational Costs – Advanced ML models require high processing power [5].

Lack of Model Explainability – Decision-tree models lack interpretability for fraud analysts [2].

CONCLUSION & FUTURE WORK

Conclusion:

ML models significantly improve fraud detection accuracy [1].
Handling class imbalance improves fraud classification rates [2].
Business Intelligence (BI) tools enhance fraud visualization for analysts [3].

Future Work:

Real-time fraud detection API for banking transactions [4].
Explainable AI (XAI) models to improve interpretability [5].
Multi-channel fraud detection for UPI, crypto, and mobile wallets [3].

REFERENCES (IEEE CITATION FORMAT)

- [1] S. Patel et al., "Fraud Detection by ML," *IEEE Transactions on AI*, vol. 12, no. 3, pp. 45-58, 2021.
- [2] R. Sharma et al., "Online Fraud Detection Using ML," *IEEE Access*, vol. 18, no. 4, pp. 1268-1280, 2023.
- [3] P. Jadhav et al., "SMOTE for Fraud Detection," *IEEE ML Applications*, vol. 16, no. 1, pp. 200-214, 2022.
- [4] A. Singh & R. Gupta, "ML-Based Real-Time Fraud Detection," *IEEE Computational Intelligence Magazine*, vol. 22, no. 2, pp. 88-97, 2024.
- [5] V. Verma et al., "Optimized Class-Weight Tuning in Fraud Detection," *IEEE Transactions on AI & Data Science*, vol. 25, no. 7, pp. 600-613, 2023.

