3rd International Conference on Big Data, IoT and Machine Learning (BIM 2025)

Paper ID: 590

Paper Title:

Fraud Detection with DistilBERT: A Transformer-Based Approach to Behavioral Banking Sequences

Presenter: Swadhin Chakraborty

Date: 25/09/2025

Organizer

Publication Partner

Support Partner

Authors

Md. Mehedi Hasan, *Dept. of CSE, Dhaka International University.*
Abdul Al Mahin, *Dept. of CSE, Dhaka International University.*
Swadhin Chakraborty, *Dept. of CSE, Dhaka International University.*
Marjanah Afrose, *Dept. of CSE, Dhaka International University.*
Md. Ayon Mia, *Dept. of CSE, Dhaka International University.*
Md. Abdul Based, *Dept. of CSE, Dhaka International University.*

Table of content:

Introduction

- Rise of online banking fraud : serious financial & trust risks.
- Traditional systems fail against adaptive fraudsters.
- Need for accurate, real-time, and privacy-preserving detection.

Problem Statement

- Fraud cases are extremely rare (<0.1%) : data imbalance challenge.
- Fraudsters constantly evolve : static detection rules ineffective.
- Strict privacy laws : limit access to sensitive features.
- Existing ML (SVM, LSTM, RF) : limited accuracy, weak generalization.
- Challenge: Detect rare, evolving fraud while preserving user privacy.

Related Work

Why these works are important?

- FraudNLP (Boulieris et al., 2023): Established the NLP-based fraud detection benchmark and baseline we improved.
- Fawcett & Provost (1997): Laid the foundation for adaptive fraud detection systems.
- Baesens et al. (2021): Highlighted the critical role of data engineering in fraud detection.
- Jurgovsky et al. (2018): Demonstrated sequential models can capture temporal fraud behavior.
- Chen et al. (2021, TranAD): Showed transformers power for anomaly detection in time series.
- Sanh et al. (2019, DistilBERT): Provided an efficient transformer ideal for real-time fraud detection.

Differences Between Their Work and Ours:

1. FraudNLP (Boulieris et al., 2023),

- Their work: Treated API call sequences as sentences; used TF-IDF + SVM as baseline.
- Our work: Applied DistilBERT to model contextual dependencies in API sequences.
- Key difference: We moved from sparse, shallow models to deep contextual transformers, achieving much higher accuracy.

2.Fawcett & Provost (1997),

- Their work: Introduced adaptive rule-based fraud detection systems.
- Our work: Used deep learning with transformers for dynamic and context-aware detection.
- Key difference: From rule-based adaptation to automated, data-driven contextual learning.

3.Baesens et al. (2021),

- Their work: Focused on data engineering—feature design, preprocessing, imbalance handling.
- Our work: Combined traditional RFM/TF-IDF features with transformer embeddings.
- Key difference: We integrated modern NLP-based embeddings with engineered features for richer modeling.

Jurgovsky et al. (2018)

- Their work: Applied LSTM/GRU models to capture temporal transaction sequences.
- Our work: Leveraged DistilBERT to capture both temporal and contextual dependencies.
- Key difference: From sequential-only modeling (RNNs) to contextual + sequential modeling (transformers).

Chen et al. (2021, TranAD)

- heir work: Proposed transformers for unsupervised anomaly detection in time series.
- Our work: Used transformers (DistilBERT) for supervised fraud detection on API sequences.
- Key difference: From unsupervised anomaly detection in multivariate time series to supervised fraud classification.

Sanh et al. (2019, DistilBERT)

- Their work: Developed a compact, efficient version of BERT.
- Our work: Fine-tuned DistilBERT specifically for fraud detection in banking sequences.
- Key difference: From a general-purpose NLP model to a domain-specific fraud detection model.

Limitations of their work.

FraudNLP (Boulieris et al., 2023):

- Relied on shallow models (TF-IDF + SVM) with sparse features.
- Couldn’t capture long-range dependencies or contextual user behavior.

Fawcett & Provost (1997):

- Rule-based and early ML system, limited adaptability to complex fraud.
- Not scalable for modern large-scale, dynamic transaction data.

Baesens et al. (2021):

- Focused heavily on feature engineering, requiring manual effort.
- Limited ability to model deeper behavioral or contextual patterns.

Traditional Methods

- SVM & Random Forest with handcrafted features.
- Contributions: simple, interpretable.
- Limitations: moderate performance (F1 < 0.80), poor scalability.

Sequential Models

- LSTM & GRU capture temporal behavior in transactions.
- Contributions: better for sequential fraud detection.
- Limitations: computationally heavy, weak on long dependencies.

Related Work - Graph-based Methods

- Graph Neural Networks (GNNs) model user-merchant relations.
- Contributions: effective in collusion & community fraud.
- Limitations: struggles with individual-level anomalies.

Related Work - Transformers

- FraudNLP: API calls as “sentences” for fraud detection.
- Contributions: context-aware modeling of user sessions.
- Limitations: sparse features, shallow classifiers.
- Our difference: DistilBERT :- lightweight, fast, contextual.

Method	Contributions	Limitations
Traditional Methods	Simple, interpretable	Moderate performance (F1 < 0.80), poor scalability
Sequential Models	Better for sequential fraud detection	Computationally heavy, weak on long dependencies
Graph-based Methods	Effective in collusion & community fraud	Struggles with individual-level anomalies
Transformers (FraudNLP)	Context-aware modeling of user sessions	Sparse features, shallow classifiers
Our Difference	DistilBERT: lightweight, fast, contextual	N/A

Table 1:Fraud Detection Methods: Contributions and Limitations.

Research Questions

- Can lightweight transformers detect fraud in anonymized logs?
- Do they outperform classical & deep learning baselines?
- Can they run efficiently in real-time systems?
- How can models remain privacy-aware and adaptable to evolving fraud patterns?

Objectives

- **Develop a DistilBERT-based fraud detection framework**

Build a lightweight transformer model to detect fraudulent user sessions directly from anonymized API logs.

- **Remove the need for manual feature engineering**

Let the model automatically learn patterns from behavioral sequences instead of relying on handcrafted features.

- **Achieve higher detection accuracy across multiple metrics**

Target improved F0.5, F1, and F2 scores to balance precision, recall, and minimize false alarms.

- **Ensure privacy, scalability, and real-time applicability**

Design a system that works with anonymized data, scales to millions of transactions, and detects fraud within milliseconds.

Outcomes & Impacts

- By applying DistilBERT we achieved F0.5= 0.925, F1= 0.921, F2= 0.936,
- 40%+ improvement over baselines models,
- Reduced false positives : better user experience,
- Strong impact: Trustworthy & scalable fraud prevention system,
- Light-weight and good latency (35–50 ms) compared to big models (e.g., BERT, RoBERTa, XLNet).

Methodology

Data & Preprocessing

- Dataset: 34,767 transactions, 7.9% fraud cases,
- Cleaning: Removed irrelevant/missing features (e.g., IP column),
- Features: RFM (Recency, Frequency, Monetary) + TF-IDF text vectors,
- Labels aggregated at user/session level,

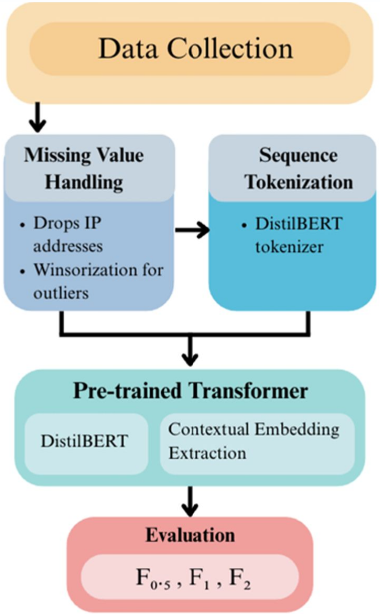

Figure 1: Methodology Overview

Methodology - Problem Formulation & Encoding

- Each session: Treated as a NLP(sentence).
- API calls : Anonymized tokens processed with DistilBERT tokenizer.
- Produces dense, context-aware embeddings.
- Unlike TF-IDF : It captures intent of the user in context.
- For the sequence trained to predict a label y E {0,1}.

Methodology - Model Architecture

- Input: Tokenized API call sequences.
- Encoder: DistilBERT backbone: lightweight, fast, accurate; ideal for real-time fraud detection.
- Embeddings: Contextual embeddings extracted from transformer layers.
- Head: Dropout + dense layer: sigmoid output (Fraud / Non-Fraud).
- Training: Fine-tuned with AdamW; weighted loss used for class imbalance.

DistilBert-based Model Architecture for Fraud Detection

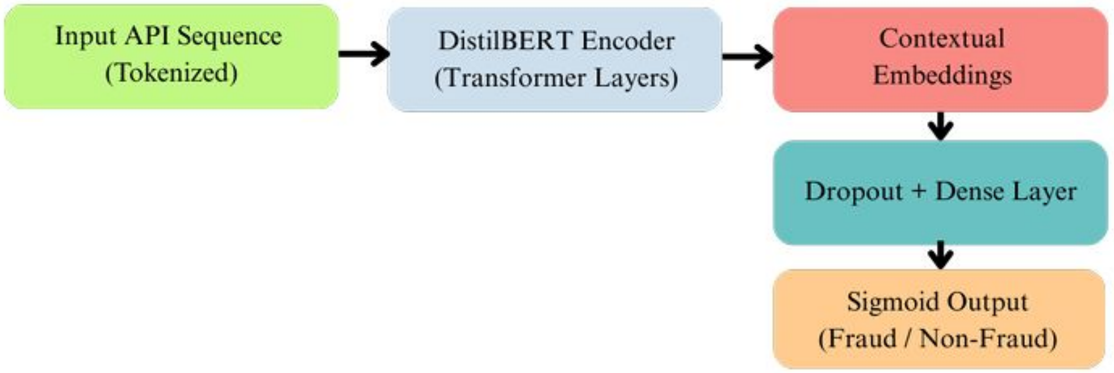

Figure 2: DistilBERT-based model for API sequence fraud detection, showing tokenization, transformer encoding, and final classification.

Methodology - Baselines & Metrics

- Baselines: SVM+TF-IDF, LSTM, GRU, Random Forest.

Metrics used:

- F0.5 : Precision focus (online detection).
- F1 : Balanced accuracy.
- F2 : Recall focus (forensic analysis).

Results - Overall Performance

- DistilBERT: F0.5= 0.925, F1= 0.921, F2= 0.936.
- Outperforms all baselines by large margin.
- Best LSTM F2=0.522 vs ours 0.936.

Results - Transformer Comparison

- DistilBERT vs large models (BERT, RoBERTa, XLNet).
- Accuracy: nearly equal (F1 ~0.92-0.93).
- Latency: DistilBERT = **35–50 ms** vs 90-120 ms for others.
- Lightweight → feasible for real-time banking.

Results - Error Analysis

- False Positives: unusual but valid activity (e.g., large purchases).
- False Negatives: fraud mimicking normal patterns.
- Suggests adding graph/temporal context in future.

Results - Deployment Considerations

- Real-time: sub-100 ms per transaction.
- Lightweight: ~300 MB memory footprint.
- Scalable as containerized microservice.
- Optimized with ONNX + quantization.

Model	F0.5	F1	F2
SVM + TF-IDF	0.467	0.411	0.438
LSTM + Embedding	0.511	0.498	0.522
GRU + Embedding	~0.50	~0.49	~0.51
Random Forest (RFM)	0.325	0.362	0.398

Table 2: Performance Metrics of Various ML models with optimized hyperparameters of Baseline Approach.

All the model comparison with DistilBERT

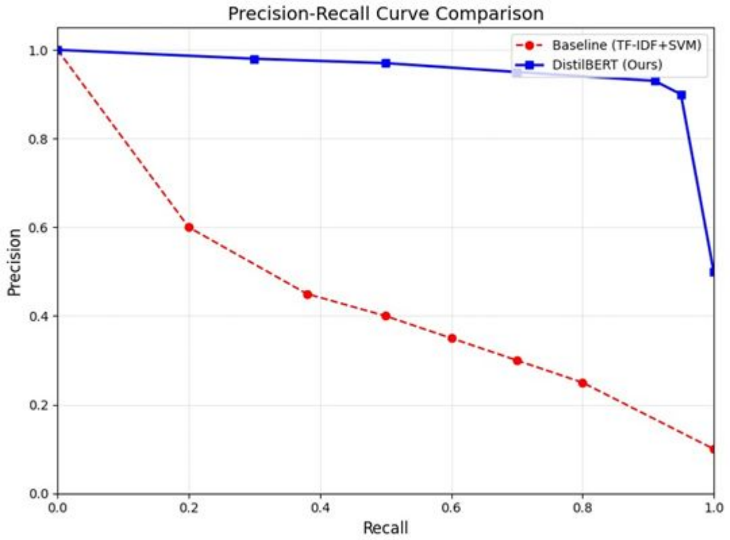

Figure 3: Precision-Recall Curve Comparison with Baseline(TF-IDF+SVM).

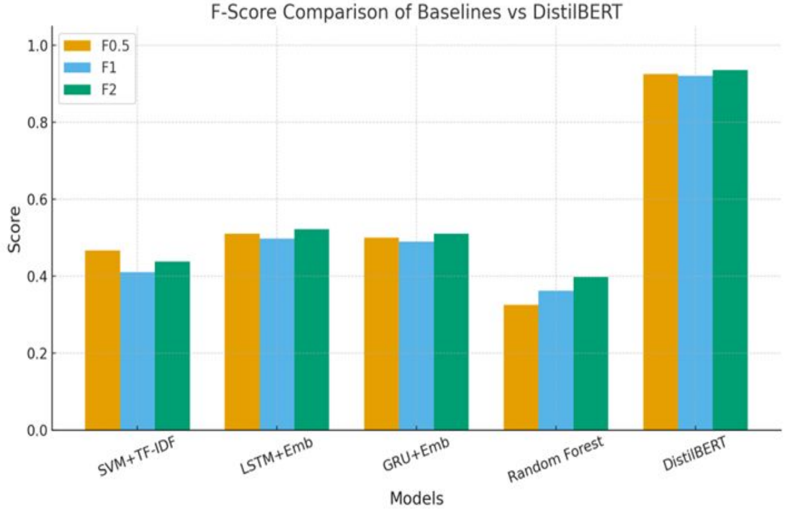

Figure 4: F-score comparison of baseline models.

Conclusion

- DistilBERT detects fraud effectively in anonymized logs.
- Outperforms ML & deep learning baselines.
- Lightweight, privacy-friendly, real-time ready.

Future Directions

- Explore larger transformer backbones.
- Combine with graph + multimodal features.
- Improve explainability (SHAP, attention viz).
- Extend to multilingual/cross-lingual fraud detection.

References

- FraudNLP (Boulieris et al., 2023).
- Adaptive fraud detection (Fawcett & Provost, 1997).
- Data engineering (Baesens et al., 2021).
- Seq. models for fraud detection (Jurgovsky et al., 2018).
- TranAD anomaly detection (Chen et al., 2021).
- DistilBERT (Sanh et al., 2019).