# Title: **Predictive-Transaction-Intelligence-using-for-BFSI**

**1. Introduction**

The Banking, Financial Services, and Insurance (BFSI) sector generates massive volumes of transaction data every day. With the rise in digital payments, mobile banking, and online financial services, the risk of fraudulent transactions has also increased significantly. This project, **Predictive Transaction Intelligence**, aims to leverage Artificial Intelligence (AI) and **Large Language Models (LLMs)** to analyze customer transaction patterns and predict future transactions while simultaneously detecting fraud risks in real-time.

**2. Background**

Financial institutions rely heavily on risk assessment and fraud detection mechanisms to protect customers and prevent monetary losses. Traditional fraud detection models are often rule-based, static, and unable to adapt to evolving fraud strategies. With advancements in AI, particularly **machine learning and LLMs**, it is possible to analyze historical behavioral data and identify anomalies with greater accuracy and speed.

By combining **transaction history analysis, behavioral modeling, and predictive AI**, BFSI firms can not only detect fraud but also forecast customer transaction needs, thereby improving customer experience and operational efficiency.

**3. Problem Statement**

Despite the availability of digital tools, the BFSI sector faces challenges such as:

* Increasing **fraudulent transaction activities** (identity theft, phishing, money laundering, etc.)
* **Static rule-based systems** that fail to adapt to new fraud techniques
* Lack of real-time **transaction prediction and risk assessment**
* Inefficient customer service in terms of **personalized financial recommendations**

Hence, there is a need for a robust AI-driven system that can:

1. Predict future transactions based on customer history.
2. Detect fraudulent activities in real-time with high accuracy.

**4. Objectives**

The main objectives of this project are:

* To design a **predictive intelligence system** for BFSI transactions.
* To integrate **Large Language Models (LLMs)** for analyzing transaction patterns and customer behavior.
* To detect and flag **fraudulent transactions in real time**.
* To provide BFSI institutions with actionable insights for **risk management and customer service**.

**5. Proposed Solution**

The proposed solution leverages **AI + LLMs + predictive analytics**:

* Use **transaction history and behavior analysis** to model customer spending patterns.
* Apply **supervised learning algorithms** to detect fraud using historical labeled data.
* Implement **real-time anomaly detection** for transactions that deviate from normal patterns.
* Integrate **LLMs for contextual understanding**, allowing the system to explain transaction risks and predict customer needs.
* Provide BFSI institutions with a **dashboard and alerting system** to monitor transactions.

**6. Dataset Description**

**Dataset 1: PaySim Synthetic Financial Datasets**

📌 [Link](https://www.kaggle.com/datasets/ealaxi/paysim1)

* Simulated financial transactions based on real-world data.
* Fields include: transaction type (CASH-IN, CASH-OUT, TRANSFER), amount, old and new balances, fraud flag.
* Highly useful for fraud detection research.

**Dataset 2: Fraud Detection Dataset (Kartik2112)**

📌 [Link](https://www.kaggle.com/datasets/kartik2112/fraud-detection?select=fraudTrain.csv)

* Transaction data with customer demographic details.
* Features: transaction date, category, amount, merchant details, fraud label.
* Helps in modeling both fraud detection and customer behavioral patterns

13-10-25

Present dataset is not suitable.  
LLM models for financial industry:

* FinGPT
* Instruct-FinGPT
* FinMA
* [BloombergGPT](https://www.bloomberg.com/company/press/bloomberggpt-50-billion-parameter-llm-tuned-finance/)

1. FinMA -7B
2. FinGPT – 7b -lora
3. InternLM-7B
4. Falcon – 7B
5. Mixtral – 7B
6. CFGPT-sft-7B-Full
7. Bhaichuan-7B

The two benchmarks are:

FLARE(**F**inancial **L**anguage understanding and P**R**ediction **E**valuation) benchmark : consisting of 5 tasks and 9 datasets  
- Sentiment analysis: dataset [1, 2]  
- News headline classification: dataset [3]  
- Named entity recognition (NER): dataset [4]  
- Question answering (Q&A): dataset [9, 10]  
- Stock price movement: dataset [7, 8, 9]  
Datasets can be found both in [[github](https://github.com/The-FinAI/PIXIU" \t "_blank)] and [[huggingface](https://huggingface.co/TheFinAI" \t "_blank)]

FinGPT benchmark: consisting of 4 tasks and 7 datasets  
- Sentiment analysis: dataset [1, 2, 11, 12]  
- Named entity analysis (NER): dataset [4]  
- News headline classification: dataset [3]  
- Relation Extraction: dataset [13]  
Datasets can be found both in [[github](https://github.com/AI4Finance-Foundation/FinGPT" \t "_blank)] and [[huggingface](https://huggingface.co/FinGPT" \t "_blank)]

References: <https://meisinlee.medium.com/llm-related-resources-for-the-finance-domain-210413bbae0a#5958>

## 14-10-25:

I used "all-MiniLM-L6-v2" – LLM Model

Deployment:

User login, risk prediction, transaction adding, transaction prediction,

Transaction categorization, suggestions using llm.

Intro

Problem statement

Goal

Dataset

Data preprocess

Model

Deployment

Functions

Result and insights

Challenges and insights  
Future scope.   
  
XGBoost works by building a series of decision trees sequentially, with each new tree correcting the errors of the previous ones. It starts with an initial prediction (like the average of the target values), calculates the errors (residuals) from this prediction, and then builds a new, simple "weak learner" tree to predict those residuals. The predictions from this new tree are added to the old ones to improve the overall model, and the process repeats for a set number of iterations to minimize error