# Developing a Large Tabular Model for Predicting Credit Approvals and Analyzing Customer Behavior in Financial Institutions

TD Bank

Columbia University Data Science Institute

## Background / Objective

### Background:

This project explores the use of Tabular Foundational Models (TabFM), specifically Llama2-GTL (Generative Tabular Learning), for predicting credit approval. While traditional machine models have demonstrated effectiveness in predicting customer behavior, they often require significant retraining and feature engineering, which can be resource-intensive and time-consuming. GTL models, leveraging zero-shot and in-context learning, offers the potential for a more efficient approach by reducing the need for extensive retraining and manual preprocessing.

### Objective:

- Establish baseline models using traditional machine learning algorithms, including logistic regression, decision trees, random forest, and XGBoost.
- Implement Large Tabular Foundation Model based off LlaMA-2 GTL without fine-tuning and Compare the performance of traditional machine learning models with that of TabFM based on customer behavior patterns to support data-driven strategies in predicting if a customer will get approved for a credit product.

## Methodology and Key Observations

### Methodology:

For traditional machine learning models, Various preprocessing techniques, such as feature engineering and SMOTE for handling data imbalance, are employed. When it comes to LlaMa-2 GTL different models, including the 7B (quantized and unquantized) and 13B (quantized) versions were tested using different prompt types (T-table vs. T-anony). The results show that the T-table prompt technique, which provides more information to the model, consistently outperforms the T-anony prompt.

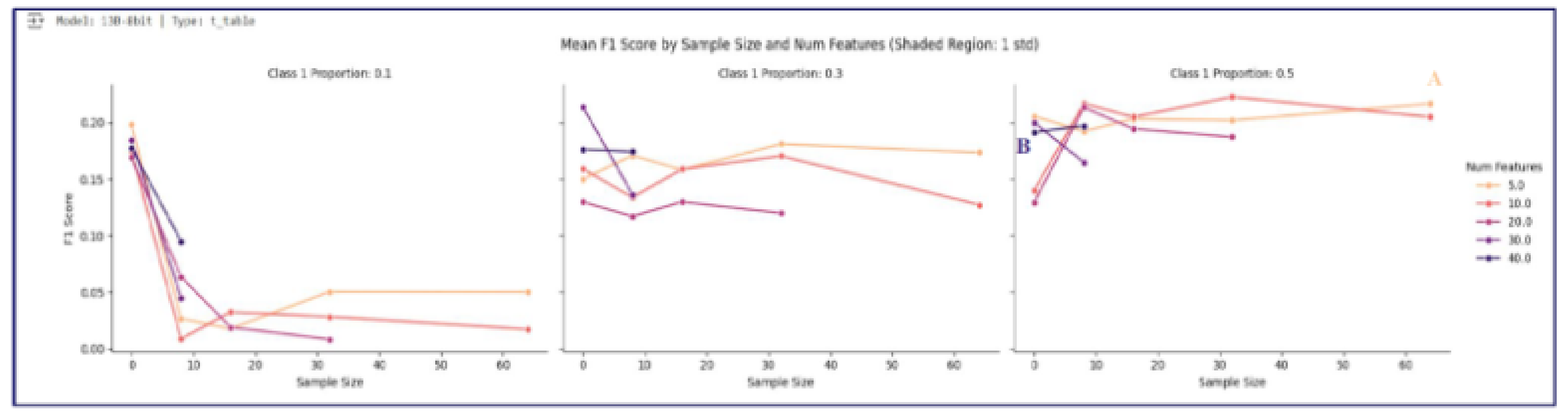
Additionally, the larger 13B model outperforms the smaller 7B model, underscoring the importance of model size in achieving better performance. The project further reveals that balanced in-context examples (equal number of majority and minority class examples) when prompting improves F1 scores. A larger sample size with fewer features also yields better results compared to a smaller sample size with more features, adhering to the context length constraints of large language models like Llama2.

Finally, comparing Llama2-GTL against the traditional models, the 13B-GTL-8bit with the T-table prompt type performed on average slightly worse than logistic regression but outperformed decision trees, random forests, and XGBoost. However, the stable performance across varying sample sizes positions Llama2-GTL as a promising tool for future credit approval prediction tasks, offering significant computational and efficiency advantages over traditional methods.

### Key Observations:

- T-Table beats T-anony
- 13B-GTL-8 bit beats 7B-GTL-8 bit
- Higher the Class 1 proportion, better the F1 score
- A large sample size with a small number of features is preferred over a small sample size with a large number of features.

In conclusion, Llama2-GTL shows potential for improving financial predictions with reduced retraining, marking a step forward in utilizing generative learning in financial applications.




**Team Members:** Atharva Kulkani, Minh Dang, Nicolás Corgono, Ryu Sonoda, Varun Agarwal, Yu Zheng Lim