Financial Analysis Driven by Machine Learning

Computer Science and Engineering
University of Puerto Rico Mayagüez Campus

Alejandro Figueroa Ramirez, Alexander Rivera De Jesús, Diego Paris Figueroa, Diego Rosario Vega

Advisor: Dr. Wilson Rivera Gallego

Department of Computer Science and Engineering

1 Problem Statement

Efficiently analyzing large volumes of financial transactions remains a significant challenge for both individuals and businesses. Manual methods are slow and prone to errors, straining to keep pace with digital transactions. Therefore, an automated, machine learning-based system would provide a quick and accurate financial analysis.

2 Problem Background

The manual process of categorizing bank transactions poses a significant challenge for our client company that manages hedge funds, impacting the efficiency and accuracy in their financial analysis. With each client's statements containing thousands of transactions, the current method is labor-intensive, time-consuming, and susceptible to bias.

This inefficiency hinders the company's ability to scale and serve its growing clientele. Thus, encouraging the integration of AI technology to streamline the categorization process, freeing up valuable time for employees and ensuring accurate financial analysis.

3 Objectives

Considering these challenges, we established the following SMART objectives:

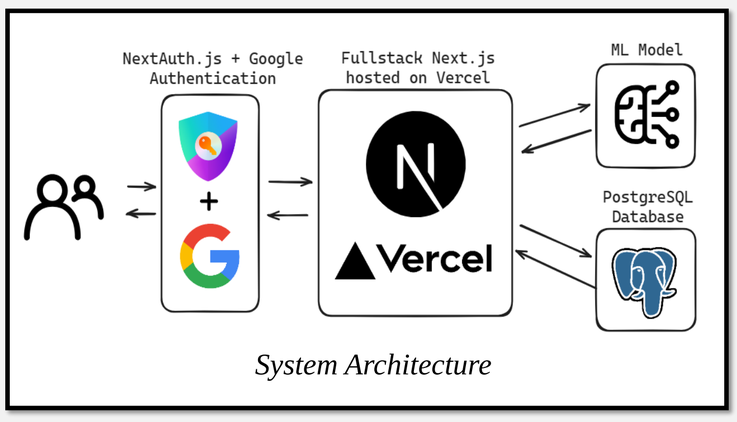
1. **Ensure Accuracy:** Make a ML model to classify bank statement transactions into categories with at least **95%** accuracy.
2. **Increase Efficiency:** Reduce the completion time for classifying bank statement transactions, aiming for at least a time reduction of **90%**.

These objectives are designed to guide the project towards creating a robust and efficient automated system, addressing the client's need for fast and accurate financial analysis while providing a scalable foundation for future expansions.

4 Technical Approach

System Architecture Overview
We utilized Next.js for front-end and back-end functionality, facilitating full-stack development. For user authentication, NextAuth.js was integrated with Google to ensure secure sign-ins.

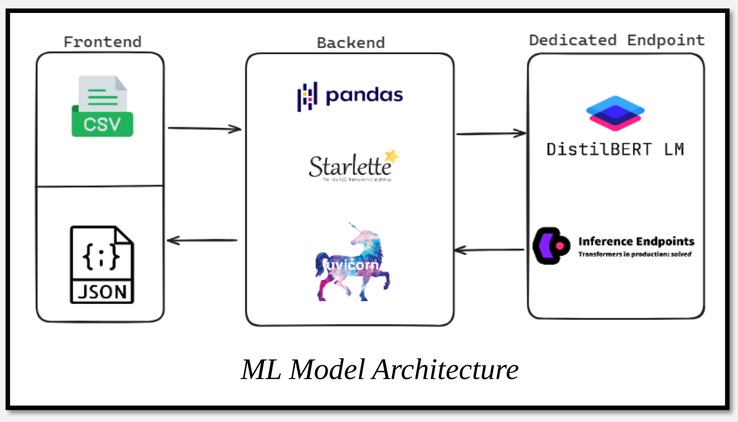
The platform and PostgreSQL database were hosted on Vercel to benefit from optimal performance and zero-config deployment, reducing technical debt and accelerating development cycles. Complimenting this, the ML model provides enhanced data processing capabilities. This architecture ensures robust scalability and performance.



System Architecture

ML Model Architecture Overview
The backend receives user CSV files through our API, made using Starlette and Uvicorn. Data is then cleaned and formatted using the Pandas library. Afterward, the data is passed into our machine learning model implemented with the DistilBert Language Model.

This model is hosted in a dedicated endpoint, through HuggingFace, which receives the formatted data and responds with the categorizations. The backend receives the response, joins them with the original file and sends it into our platform.



ML Model Architecture

5 Results

The integration of a ML model to streamline the categorization process of bank statements has produced outstanding results.

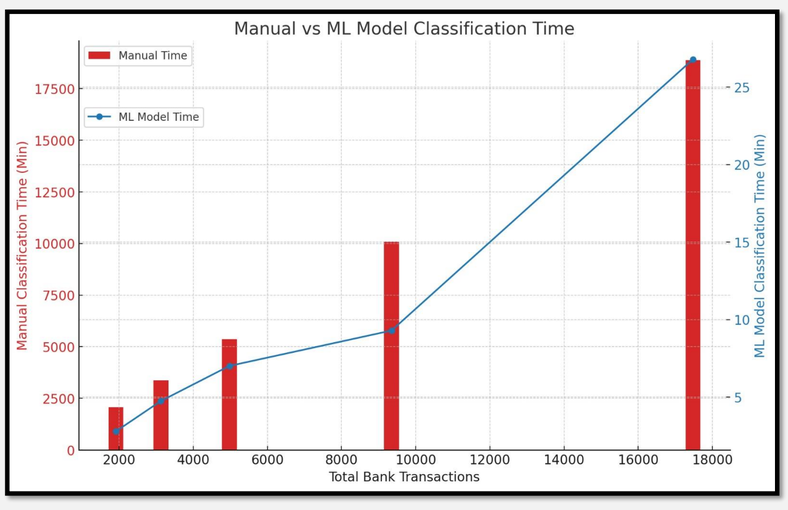
ML Model Accuracy
A total of 1,330 new transactions were sent to the ML model, and it classified them correctly with an accuracy of **99.98%**. Successfully achieving an accuracy rate of at least **95%**.

Efficiency Increase
The manual classification process averages **1.08 minutes per transaction**, while the ML model significantly reduces this time to just **0.001 minutes**, achieving an average increase of **99.87%.**

Disclaimer: The following measurements were calculated using a domestic computer running Windows 11. Performance may increase using dedicated hardware.

TABLE 1: Manual vs ML Model Classification Efficiency

Total Bank Transactions	Manual Classification Time (Min.)	ML Model Classification Time (Min.)	Percentage Increase
1,914	2,067.12	2.80	99.86 %
3,129	3,379.32	4.76	99.86 %
4,977	5,375.16	7.02	99.87 %
9,348	10,095.84	9.30	99.91 %
17,475	18,873	26.78	99.86 %



6 Conclusion

The implementation of our ML-based system provides a significant advancement in the field of financial analysis. We have enabled the client company to scale effectively with their growing clientele while maintaining high accuracy and performance.

R References

1. V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “Distilbert, a distilled version of Bert: Smaller, faster, cheaper and lighter,” arXiv.org, https://arxiv.org/abs/1910.01108 (accessed Feb. 16, 2024).
2. “Huggin Face Transformers,” huggingface.co. https://huggingface.co/docs/transformers/index (accessed Apr. 01, 2024).