Capstone Project

Cover Page

- **AI classification model for financial industry complaints**
- **Group 5 - Jean Miranda, Richard Contreras, Ricardo Ibanez**
- **Date - 7 Dec 2024**
- **Boston University/ MBA+MSDT Program**

Executive Summary

The AI-driven classification model for financial industry complaints uses OpenAI's GPT-4o-mini to automate complaint categorization and routing via a Streamlit chatbot and Jira integration. This scalable solution improves efficiency, reduces handling times, and enhances customer satisfaction. Achieving over 80% accuracy at the product level, it highlights areas for refinement in issue-level classification. Cost-effective with a high ROI, the system addresses key operational challenges and positions financial institutions for streamlined complaint management.

1. Introduction (~0.5 Page)

1.1 Problem Statement

The financial industry handles a multitude of customer complaints daily, ranging from billing issues to allegations of fraud. The efficiency of addressing these complaints is affected by existing complaint management systems, which require significant manual input and are constrained by rigid, rule-based classification systems. These limitations lead to increased handling times, customer dissatisfaction, improper initial contact with pertinent parties, and potential regulatory risks. This project proposes an AI-driven solution that utilizes a pre-trained Large Language Model (LLM) to automate the classification and routing of complaints, thus improving response times and accuracy.

Key Decisions:

API Model: We opted to use the OpenAI API instead of a pretrained model to minimize initial capital expenditures and to leverage the advanced capabilities of cutting-edge models like GPT- 4o-mini.

Prompt Engineering: We have opted for prompt engineering over fine-tuning due to the additional effort and costs associated with managing multiple categories in fine-tuning using API. Instead, we are focusing on refining prompts through instructions, format, length, and persona, while also evaluating the inclusion of context.

User Interface: We have narrowed the prototype scope to a chat interface that allows users to enter their issues and receive an acknowledgment indicating the product team to which the issue has been assigned. Subsequently, we are generating a task in Jira with the description of the issue and a priority for those cases related with fraud.

2. Research and Literature Review (~0.5 Page)

2.1 Related Work

Consumer complaints of consumer financial protection bureau via two-stage residual one-dimensional convolutional neural network (TSR1DCNN) (1): The study introduces the Two-Stage Residual One-Dimensional Convolutional Neural Network (TSR1DCNN) as an advanced model for processing consumer complaints, particularly those managed by the Consumer Financial Protection Bureau (CFPB). The CFPB receives a high volume of complaints, making efficient processing essential to ensuring timely resolutions.

TSR1DCNN uses deep learning to classify and prioritize these complaints, outperforming traditional models such as one-dimensional CNNs, LSTMs, and Bidirectional LSTMs. With a dataset of 555,957 complaints, TSR1DCNN achieved 78.07% accuracy on training data and 76.53% on test data, demonstrating its robustness in handling textual complaint data.

3. Solution Design and Methodology (~2 Pages)

3.1 Solution Overview

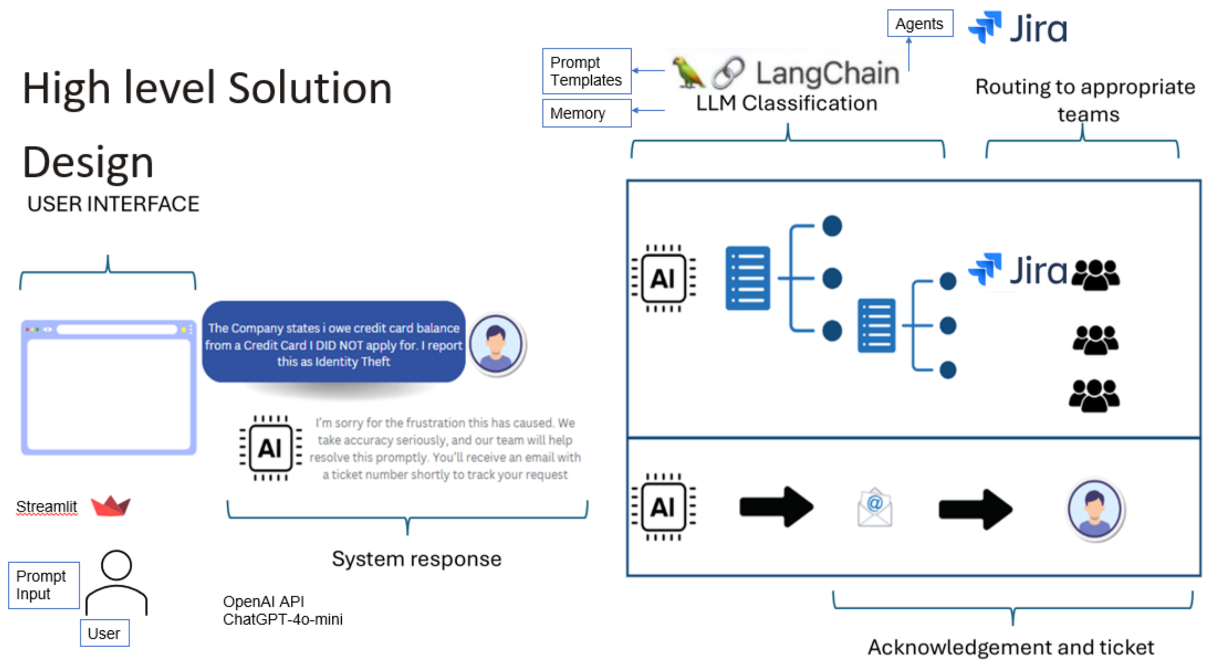
Our solution architecture is designed to integrate a pre-trained large language model (LLM) via the OpenAI API with the existing infrastructure of financial institutions to classify customer complaints across various categories of Products (e.g. debt collection, credit card, checking accounts) and Issues (e.g. incorrect information on your report, improper use of your report, attempts to collect debt not owed, and account management). Reflecting on the user experience (UX) of the solution, we have defined a system that utilizes a chat interface, allowing users to input their issues. In addition to this, these issues will then be categorized and assigned a task in Jira for further processing by the resolution team. This integration will automate routing of complaints to the appropriate responsible via Jira, significantly reducing processing time.

- Chatbot Interface: Developed with Streamlit, allowing users to input complaints and receive categorized responses.
- Classification System: Uses sequential instructions to classify complaints hierarchically by Product, Sub-product, and Issue.
- Jira Integration: Automatically creates and prioritizes tasks for resolution teams based on complaint classification.

3.2 System Architecture

Model. We selected GPT-4o-mini for its proven capability in understanding and generating human-like text. GPT-4's extensive training on diverse datasets makes it appropriate for understanding the language often found in customer complaints, which is crucial for accurate classification. We chose GPT-4o-mini over GPT-4o because it demonstrated higher accuracy in our specific use case, particularly in distinguishing complaint categories.

Front-End: The user interface is built using Streamlit, which enables a web-based version of the chatbot for seamless user interaction. All necessary API keys are securely configured within the platform, facilitating access to both GPT-4o-Mini for natural language processing and Jira for task management. The chat interface is designed to engage users in a friendly and intuitive manner through the LLM, collecting detailed information about their issues while guiding them through the process. Once the required details are gathered, the system confirms the reported issue to the user and assures them that it is being processed by the support team, creating a streamlined and transparent user experience.



Back-End (Tools): JIRA is seamlessly integrated into the chat ecosystem to manage customer complaints by converting them into actionable tasks. Once a complaint is classified, a corresponding task is automatically created in JIRA, assigned to the appropriate team or individual based on predefined rules. JIRA facilitates efficient issue tracking by enabling priority management and monitoring the assignment timeline. Additionally, it enhances communication by sending automated email notifications for each assigned task, ensuring immediate review and resolution. This integration streamlines complaint handling while providing robust tracking and accountability throughout the resolution process.

GitHub: https://github.com/rjcontrerasr/IS883-LLM_Project-Financial_complaints

Prototype link: https://is883-group-project-g5.streamlit.app/

3.3 Use Cases

Use Case 1 - General Complaint Classification and Task Creation: A customer submits a complaint through the Streamlit interface, describing an issue with unexpected charges on their credit card. The system processes the input using the GPT-4o-mini model, which classifies the complaint into the categories "Credit Card" (product), "General-purpose credit card or charge card" (sub-product), and "Problem with a purchase shown on your statement" (issue). Once the classification is complete, the system automatically creates a task in Jira, detailing the classification and customer-provided information. The task is assigned to the relevant team for resolution with a High Priority to ensure prompt action.

Use Case 2 - Fraudulent Complaint with Highest Priority: A customer reports fraudulent activity on their checking account through the complaint submission interface. The model categorizes the complaint under "Checking or saving account" (product), "Checking Account" (sub-product), and "Problem with fraud alerts or security freezes" (issue). Recognizing the critical nature of the complaint, the system creates a task in Jira, assigns it to the dedicated fraud resolution team, and sets the priority to Highest Priority. This ensures immediate action is taken to mitigate potential risks.

Use Case 3 - Handling Incomplete Complaint Information: A customer submits a complaint stating only "There's a mistake with my payment" without specifying the product or issue. The system engages the user through a chat feature, prompting them to provide additional details, such as the product and specific issue. For instance, the chat asks, "Could you clarify which product this issue is related to?" Once the customer specifies "Personal Loan" the model classifies the complaint as "Payday loan, title loan, personal loan, or advance loan" (product), "Installment loan" (sub-product), and "Problem when making payments" (issue). The system then creates a task in Jira, assigns it to the appropriate team, and sets the priority to High to ensure timely resolution, following the same process as for fully detailed submissions.

4. Implementation and Development (~0.5 Page)

4.1 Technical Approach

Our solution is an AI Financial classification system designed to improve how financial institutions handle customer complaints. By using the OpenAI API with a pre-trained GPT-4o-mini model, this system automates the classification and routing of complaints to enhance response time, accuracy, and customer satisfaction, and creates a task in Jira that is assigned to the technical team to provide support.

The high-level system architecture is illustrated in the diagram below. The user interface, hosted on Streamlit, provides access to the financial support system through an intuitive chatbot interface. When a user initiates an interaction, the system uses a large language model (LLM) to guide the conversation.

Through carefully designed prompt engineering and sequential chains, the chatbot asks targeted questions to gather detailed information about the user's issue. The process is designed to give a balance between maintaining a formal tone and being user-friendly and easy to understand.

Once the user has fully described their issue, the LLM automatically classifies the complaint by product, subproduct, and issue type. The system acknowledges receipt of the complaint and provides a clear confirmation to the user.

Finally, the system integrates with Jira to create a task. This task includes a descriptive title based on the reported issue, product and subproduct, and is assigned to an appropriate team in Jira, streamlining the resolution process. This end-to-end flow ensures a seamless and efficient user experience while automating key steps in complaint management.

5. Financial Modeling and Cost Analysis (~1 Page)

5.1 Financial Model

For the financial modeling and ROI, we have assumed that the model is replacing 2 full time workers with a minimum wage of $15.00 per hour, working 40 hours a week (**see Exhibit 1**).

Category	**Cost/Benefit**	**Monthly Expense/Revenue ($)**	**Assumptions**
API Usage	Cost	14	10,000 complaints/month at $0.0014/token
Cloud Storage	Cost	20.48	1TB monthly storage at $0.02/GB **(4)**
Development Costs	Cost	500	One-time cost spread over 12 months
Training Costs	Cost	300	Training and prompt engineering costs
Savings from Automation	Benefit	-4800	Replacing 2 full time workers with a wage of $15hr

For the sensitivity analysis we have considered an increase in the complaints, we initially estimated 10K complaints a month, if that were to grow considerably, the ROI would still be positive.

	10K complaints	**15K Complaints**	**20K Complaints**
Category	**Annual Cost ($)**	**Annual Cost ($)**	**Annual Cost ($)**
API Usage	168	252	336
Cloud Storage	246	246	246
Development Costs	6,000	6,000	6,000
Training Costs	3,600	3,600	3,600
Savings from Automation	(57,600)	(57,600)	(57,600)
ROI	**475.2%**	**470.4%**	**465.7%**

6. Evaluation Plan (~2 Pages)

6.1 Testing and Validation

The testing plan focused on classification accuracy at Product, Sub-product, and Issue levels, using real public data from financial institutions.

6.2 Success Metrics

Quantitative:

- **Classification Accuracy:** Exceeded 80% for Product and Sub-product levels but remained low at the Issue level (See exhibit 2 for evaluation).
- **Umbrella Categorization Accuracy:** Improved to around 60% the issue level, but still insufficient.

Qualitative:

- **User Interaction Quality:** Assessed through chatbot tests, focusing on friendly responses and information-seeking capabilities.

6.3 Strengths and Limitations

Strengths:

- High accuracy at broader categorization levels (Product and Sun-product).
- Umbrella categorization demonstrated potential for simplifying complex issues.

Limitations:

- Challenges with FAISS vectors implementation to improve accuracy. Limited by the absence of formal definitions for 87 distinct issues, making context embedding infeasible. Prototype definitions generated with ChatGPT lacked clear category distinctions.
- Data cleaning imperative. Real-world financial data requires significant cleaning to address inconsistencies and ambiguous descriptions, which hinder the model's ability to differentiate subtle variations in issue categories.

Recommendations:

- Standardized Definitions: Develop and validate definitions for all 87 issues with input from domain experts, providing a strong foundation for future classifications.
- Refinement of Categorization: Enhance the umbrella categorization approach with expert feedback to ensure real-world applicability and scalability.
- Focus on Dataset Cleaning: Prioritize resolving inconsistencies and ambiguities to enhance issue-level accuracy.

GitHub (performance): https://github.com/rjcontrerasr/IS883-LLM_Project-Financial_complaints/blob/main/Model_performance.ipynb

7. Project Management and Timeline (~1 Page)

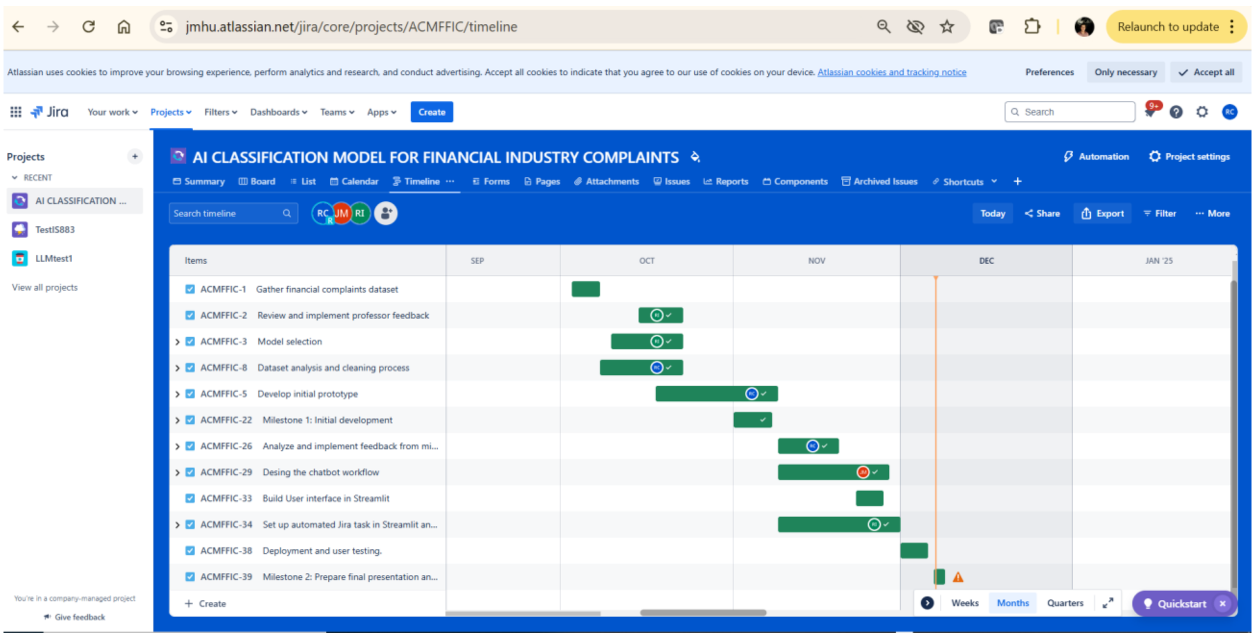
7.1 Timeline and Milestones

The project was organized into key phases, focusing on coding tasks to improve classification accuracy and refine the model's performance. Each of us took on an equal share of responsibilities, working on implementing and testing different solutions, including embedding issue definitions using FAISS vectors and developing an umbrella categorization approach. A significant part of our work involved coding the logic for these enhancements, iteratively testing their impact on classification accuracy, and debugging to address inconsistencies.

7.2 Team Dashboard and Work Distribution

In this second part of the project, the chatbot workflow, including its structure, classification logic, and notification flows, was designed and implemented. Additionally, the user interface was developed using Streamlit, and Jira task assignment automation was seamlessly integrated into the core application.

We all collaborated on testing different models initially, but we divided tasks to work more efficiently once a prototype was in place. One team member focused on testing prompts and model versions to improve accuracy, another refined the categorization process and experimented with chains, and the third worked on integrating the code with tools and agents and transitioning it from Colab to the Streamlit application. Despite these focused responsibilities, we supported one another throughout the process, addressing questions and ensuring the project advanced cohesively.



https://jmhu.atlassian.net/jira/core/projects/ACMFFIC/timeline[1]

8. Conclusion (~0.5 Page)

8.1 Summary of Outcomes

Our AI Classification Model for Financial Industry Complaints successfully automated key aspects of complaint management. By using GPT-4o-mini, our model achieved over 80% accuracy in classifying complaints by product and subproduct while integrating with Jira to automatically create and assign tasks to relevant teams. This reduced complaint handling time, improved routing precision, and streamlined resolution processes. In addition, the user experience was enhanced with a chatbot interface developed on Streamlit, allowing customers to submit complaints easily while receiving instant acknowledgments.

8.2 Future Work and Recommendations

To enhance our model, we recommend cleaning and standardizing the dataset to eliminate ambiguities and inconsistencies. Additionally, establishing standardized, expert-reviewed definitions for complaint issues would significantly enhance classification accuracy by providing the model with clearer context and more precise criteria for the classification process.

---

[1] Please review your email (elhamod@bu.edu) to accept the invitation as a team member of the project.

Appendix (Optional)

Exhibit 1

On average, each complaint contains approximately 200 tokens, and the prompt instructions another 200 tokens. With our current model, we assume handling over 10,000 complaints per month, averaging 4M tokens, for a total cost of $10 per 4M input tokens **(5)**. Similarly, for output tokens, we average about 40 tokens for product and issue categorization. This would result in a total output cost of $4 for those 10,000 complaints, assuming we can deduct it from the $10 charge per 1 million output tokens.

Model	Pricing	Pricing with Batch API*
gpt-4o	$2.50 / 1M input tokens	$1.25 / 1M input tokens
	$1.25 / 1M cached** input tokens	
	$10.00 / 1M output tokens	$5.00 / 1M output tokens

ref: https://openai.com/api/pricing/

Exhibit 2

Performance across different modeling approaches

Model Attempt	Description	Product Accuracy	Sub-product Accuracy	Issue Accuracy	Umbrella Issue category Category
Sub-product Classification	Classifies complaints into their respective sub-product categories	N/A	0.8	N/A	N/A
Issue Classification	Classifies customer complaints directly by issue categories	N/A	N/A	0.12	N/A
Sub-product and Issue Classification	Sequential classification into sub-product and then issue categories	N/A	0.79	0.3	N/A
Product, Sub-product, and Issue Classification	Hierarchical classification into product, sub-product, and issue categories	0.85	0.79	0.34	N/A
Filtered Hierarchical Classification	Hierarchical classification with filtered categories and prompt improvements	0.85	0.8	0.43	N/A
Group Issues under broader umbrella categorie	Pre-defined mapping to classify 87 issues into 21 broader categories.	0.9	0.8	N/A	0.2
Group Issues under broader umbrella categorie - improvement.	Pre-defined mapping to classify 87 issues into 21 broader categories. Add clasification issue step	0.92	0.84	0.42	0.62

References:

(1) Oyewola, D., Omotehinwa, T. O., & Dada, E. G. (2023, August). Consumer complaints of consumer financial protection bureau via two-stage residual one-dimensional convolutional neural network (TSR1DCNN). Retrieved from https://www.researchgate.net/publication/372822603

(2) Consumer Financial Protection Bureau. (2024). *Consumer complaints search (from April 5, 2024, through October 5, 2024)*. Retrieved from https://www.consumerfinance.gov/data-research/consumercomplaints/search/?date_received_max=2024-10-05&date_received_min=2024-04-05&page=1&searchField=all&size=25&sort=created_date_desc&tab=List

(3) Consumer Financial Protection Bureau. (2023). *Consumer complaints search (from October 5, 2022, through April 4, 2023)*. Retrieved from https://www.consumerfinance.gov/data-research/consumercomplaints/search/?date_received_max=2023-04-04&date_received_min=2022-10-05&page=1&searchField=all&size=25&sort=created_date_desc&tab=List

(4) Google Cloud. (2023). North America storage pricing. Retrieved December 7, 2023, from https://cloud.google.com/storage/pricing#north-america

(5) OpenAI**.** (n.d.). *API pricing*. Retrieved from https://openai.com/api/pricing/