



AIN429 Data Mining Lab Project Proposal

**Customer Support Chatbot using Knowledge Graphs and
Llama Model**

**Kemal Şahin, 2200765021
Burak Kurt, 2200765010**

Course Name: AIN429 Data Mining Laboratory

Date: November 9, 2024

Contents

1 Problem	2
2 Data	2
3 Methodology	2
4 Plan	3
5 References	3



1 Problem

In today's digital world, customer support demands have increased as companies manage a high volume of customer inquiries on social media. Traditional customer service solutions can struggle with managing diverse queries efficiently and providing personalized responses. This project aims to build a customer support chatbot that can:

- Automatically detect key entities and intents in customer inquiries,
- Retrieve relevant information from a knowledge graph and JSON-based dataset,
- Generate concise and contextually accurate responses using the Llama 3.1 instruct model.

The chatbot will streamline customer support on social media platforms, enhancing response accuracy and efficiency for customer support teams.

2 Data

The primary dataset for this project is the *Customer Support on Twitter* dataset from Kaggle, which consists of consumer conversations with customer support agents on Twitter. Key fields include:

- `tweet_id`, `author_id`, `inbound`, `created_at`, and `text` fields for each tweet,
- `response_tweet_id` and `in_response_to_tweet_id` for conversation flow.

This dataset provides an extensive collection of real-world customer queries and support interactions suitable for building conversational AI models.

3 Methodology

The chatbot will be developed through the following key steps:

1. **Data Preparation:** Download and clean the dataset to handle missing values and anonymize sensitive information.
 2. **Entity and Intent Detection Using Llama:** Use the Llama 3.1 instruct model to identify key entities (e.g., product, issue type) and intents (e.g., complaint, request for assistance) in tweets via structured prompts.
 3. **Knowledge Graph and Ontology Creation:** Develop a taxonomy and ontology of common issues, organizing them into a knowledge graph using tools like Neo4j.
 4. **JSON Lookup Dataset:** Structure a JSON dataset with frequently asked questions, common issues, and responses for efficient query-based lookup.
 5. **Response Generation with Llama:** Integrate Llama to generate responses, using data from the JSON lookup or knowledge graph. If no data is available, the model will respond with a fallback response.
-

4 Plan

The project will be completed over five weeks with the following schedule:

1. Week 1: Data Preparation and Exploration

- Download and explore the dataset, analyzing its structure and content.
- Clean the data by handling missing values and anonymizing sensitive information.
- Segment conversations to capture context, organizing data by conversation threads.

2. Week 2: Entity and Intent Extraction with Llama

- Set up and configure the Llama 3.1 instruct model.
- Use Llama to identify and extract key entities (e.g., products, issue types) and classify intents (e.g., complaints, inquiries) from the dataset.
- Store extracted entities and intents as structured data for further use.

3. Week 3: Taxonomy and Ontology Development

- Design a taxonomy to organize the most common entities and intents into structured categories.
- Develop an ontology by defining relationships between these categories using an ontology tool (e.g., Protégé).
- Begin creating a basic knowledge graph structure to store relationships and hierarchies.

4. Week 4: Knowledge Graph and JSON Dataset Creation

- Populate the knowledge graph with extracted entities, intents, and relationships.
- Design and create a JSON lookup dataset containing frequently asked questions, common issues, and responses.
- Ensure that entries in the JSON dataset align with the structure of the knowledge graph.

5. Week 5: Llama Integration and Response Generation

- Integrate Llama with the knowledge graph and JSON lookup dataset for response generation.
- Test and refine the chatbot's response accuracy, making adjustments to prompts and data retrieval.
- Conduct final tests on example queries to evaluate chatbot performance.

5 References

1. Thoughtvector. (n.d.). *Customer Support on Twitter Dataset*. Retrieved from <https://www.kaggle.com/datasets/thoughtvector/customer-support-on-twitter>.
 2. OpenAI. (2023). *Llama 3.1 Instruct Model Documentation*. OpenAI.
-