**CMPE 274 Hackathon Report  
Business Intelligence Technologies**

Rohit Dadlani

Bunpheng Chhay

Chirag Arora

**Enterprise Business Intelligence Platform RAG for Market Research & Food Security**

This document details the development and results of a data-driven application designed to address pressing issues related to food security and economic impacts on agriculture commodities, specifically phosphoric acid and phosphate fertilizers. The application leverages a combination of data visualization, machine learning, and natural language processing (NLP) to provide insights into commodity pricing trends and to facilitate an intelligent question-answering system about food security.

Introduction:

This project aims to utilize technological advancements in data analysis and artificial language understanding to provide clearer insights and accessible information regarding these topics. Specifically, the focus is on analyzing the Producer Price Index for phosphoric acid and phosphate fertilizers and examining global food security issues through a question-answering system developed with Haystack and Transformer technologies.

**Data Visualization of Commodity Prices**

The initial part of the project involves visualizing the Producer Price Index by Commodity for Chemicals and Allied Products, particularly phosphoric acid and phosphate fertilizers. This visualization is achieved through the use of Python's Plotly library, allowing for an interactive graphical representation of price trends over time. The data is sourced from a CSV file stored on Google Drive, showcasing the integration of cloud storage with data analytics.

**Key Features of the Visualization:**

* Time-range selectors to view data from different periods.
* Interactive line graphs displaying price trends.
* Dynamic loading of data from Google Drive to ensure real-time accuracy and accessibility.

This visualization serves as a foundational tool for analysts and stakeholders to observe market trends and make informed decisions regarding commodity investments and procurement.

Here are the Data Visualization Screenshot we had attached below:

Attach Output Screenshots of Visualization here:

A graph showing the growth of the company

Description automatically generated with medium confidence

**Development of the Question-Answering System**

The next phase involves the creation of an advanced question-answering system that utilizes a combination of Elasticsearch for document storage and retrieval, and the Haystack framework for processing and answering natural language queries.

Technical Implementation:

1. Document Storage and Retrieval: The system uses Elasticsearch to index and retrieve documents. These documents include detailed reports and PDFs related to food security and the agriculture sector.
2. PDF Processing: Documents are converted from PDF to text format, allowing for further processing and indexing.
3. Question Answering: Utilizes the deepset/roberta-base-squad2 model within Haystack to answer queries. This model is specifically fine-tuned for the question-answering tasks on the SQuAD dataset.

**Capabilities of the QA System:**

* Answer queries related to food security, such as the impact of economic conditions on food prices.
* Provide insights into the nutritional aspects in different scenarios, including conflict zones and economic downturns.

In the development phase we tried to solve this problem using different models and we also tried to integrate using various free APi’s. Here are the list of free open LLM API’s we tried to use:

* OpenAI free api
* GooseAi
* Gemini

We tried to integrate our small datasets first with all API’s. But some of the API’s provide very few features to perform free usage. For example some API offers only limited text entry or some does not offers the data to put on the cloud platforms.

For the visualization part we just use the basic machine learning library to predict the outcomes of the model. So, in that part we used the Machine learning libraries like numpy, pandas, etc. and for visualization diagrams we used matplotlib,seaborn,etc. to get the diagrams.

In the model building and API integration part we used the rasa model first to check the sustainability of the project. So, we integrated the openAI API to the rasa model that we have trained. But this model makes it slower to process our dataset. And after integrating the OpenAI API it nearly takes 2 hours to process the dataset and the accuracy of the model is very less that if we ask the question to the model that “How can I grow bananas?” so they provide the output in terms of apples. In that case we understood that the rasa model is not fit for this process.

After that we moved towards different API’s and used different models. We started working on the different models that suit more than we can go ahead and integrate the chatbot API on that model. So, we integrated the chatbot model and the whole code link I had attached below.

<https://drive.google.com/file/d/18LvRWlxSVv1mB4UZ10Q9-ZwnKQDRI9MT/view?usp=share_link>

This link shows the code for our jupyter notebook code that includes the final visualization part, the model output and the chatbot integration using open API. I had attached the screenshot outputs here below:

A screenshot of a computer

Description automatically generated

**Executive Report**

The question-answering system was tested with various queries related to global food security issues. Here are some highlighted inquiries and the insights provided by the system:

A table with text on it

Description automatically generated

**Conclusion**

The application developed during this hackathon effectively integrates data visualization and NLP-based querying to address and elucidate complex issues surrounding agriculture commodities and global food security. Through interactive graphs and an intelligent QA system, it provides valuable insights that can aid policymakers, analysts, and the public in making informed decisions.

The future scope of this project includes scaling the data sources to include more commodities and integrating more comprehensive reports and real-time data feeds to enhance the accuracy and relevance of the answers provided by the system.

**Colab NoteBook: https://drive.google.com/file/d/18LvRWlxSVv1mB4UZ10Q9-ZwnKQDRI9MT/view?usp=share\_link**