COFORGE

SUMMER INTERNSHIP

# Introduction

* **Project Title** : Natural Language Processing Project for German Railway Client.
* **Name** : Md Athar
* **College** : Indian Institute of Technology (BHU) Varanasi.
* **Date of Joining** : 27th May, 2024
* **Last working Date**: 12th, July, 2024
* **Supervisor’s Name** : Shivani Gupta
* **Technology/ Tools Utilised** : Python, Spacy, NLTK, pandas, Numpy, scikit-learn, TF-IDF, streamlit, Matplotlib, Seaborn, Google colab, MS Excel etc.

# Objective

The objective of this project was to leverage Natural Language Processing (NLP) techniques to analyse customer reviews for the German railway system, develop a predictive model to determine the eligibility of customers for goodwill vouchers based on their feedback, and generate automated email responses addressing their issues. The ultimate goal was to enhance customer satisfaction and ensure a smoother journey experience by efficiently resolving customer complaints and rewarding loyal customers.

This objective is clear and comprehensive, highlighting the key aspects of my project:

* **Leveraging NLP**: Specifies the use of NLP techniques.
* **Analysing Customer Reviews**: States the initial task of analysing the customer feedback data.
* **Keywords Extraction and calculating Sentiment Score:** Keyword extraction using NLP (Natural Language Processing) involves identifying and extracting important words from customer reviews.
* **Predictive Model**: Mentions the development of a model to predict voucher eligibility.
* **Automated Email Generation**: Includes the task of generating emails based on customer issues.
* **Enhancing Customer Satisfaction**: Emphasises the overarching goal of improving the customer experience.



# Week Wise Work Done

## Week 1 : Understanding Company Processes and Initial Data Analysis

**Company Orientation:**

* Learned about the company processes, objectives, and goals.
* Understood the **scope of the project** and the specific requirements from the German railway client.

**Dataset Acquisition:**

* Received the dataset containing customer reviews and complaints from the **German railway system**.

**Initial Data Analysis:**

* Conducted an initial analysis of the dataset to understand its structure, content, and quality.
* Explored various techniques for **data cleaning** **and preprocessing.**
* Researched methods for effective data analysis, including **Natural** **Language Processing (NLP)** techniques.

## Week 2 : Data Cleaning and Preprocessing

**Data Cleaning:**

* Implemented data cleaning techniques to **handle missing values**, **remove duplicates**, and **correct errors** in the dataset.
* Standardised the format of the data to ensure consistency.
* Worked extensively with **MS Excel** to learn and apply various filtering techniques and features for data cleaning.

**Preprocessing:**

* Preprocessed the text data, including **tokenization**, **stemming**, and **lemmatization**.
* Removed **stop words** and performed **text normalisation** to prepare the data for analysis.

## Week 3 : Keyword Extraction and Sentiment Analysis

**Keyword Extraction and Text Preprocessing:**

* Used **spaCy** for keyword extraction, lemmatization, and removing stop words and punctuation.
* Applied **tokenization** using both **spaCy** and **NLTK** to segment text into individual tokens.

** Sentiment Analysis:**

* Utilised **TextBlob** and **VADER** to calculate sentiment scores for customer feedback, providing insights into the overall sentiment.

**Feature Extraction:**

* Applied **TF-IDF Vectorizer** and **One-Hot Encoding** to convert case descriptions into numerical data, making them suitable for machine learning models.
* **Features Extraction** from customer issue descriptions to calculate scores and predict outcomes.

## Week 4 : Model Development and Training

**Classification Model:**

* Trained and **built classification models** to categorise customer feedback based on predefined criteria.
* Evaluated different algorithms to determine the **best-performing model.**

**Model Validation:**

* Validated the model using cross-validation techniques to ensure **accuracy** and **reliability**.



## Week 5 : Automated Email Generation

**Email Generation System:**

* Designed and implemented a system to generate **automated emails** based on customer complaints.
* Used **Gemini** and **Llama2** models to create personalised email responses addressing customer issues.

**Testing and Refinement:**

* Tested the email generation system to ensure the quality and relevance of the generated emails.
* Refined the system based on feedback and performance metrics.

## Week 6 : Web Application Development (Streamlit)

**Web Application Creation:**

* Created a web application for customers to input their issues and receive guidance regarding their problems.
* Utilised **Streamlit** to develop the web application, ensuring a seamless and user-friendly interface.

**Model Integration:**

* **Integrated all trained NLP models**, including keyword extraction, sentiment analysis, and classification models, into the web application.
* Ensured that the application could effectively process customer input and provide appropriate directions and responses.

## Week 7

**System Testing:**

* Conducted thorough **testing of the entire system**, including the NLP models, email generation, and web application functionality.
* Identified and fixed bugs or issues that arose during testing.

**User Experience Enhancement:**

* Worked on improving the overall **user experience** based on feedback from initial testing and user trials.
* Focused on making the interface more intuitive and responsive.

**Interface Improvement:**

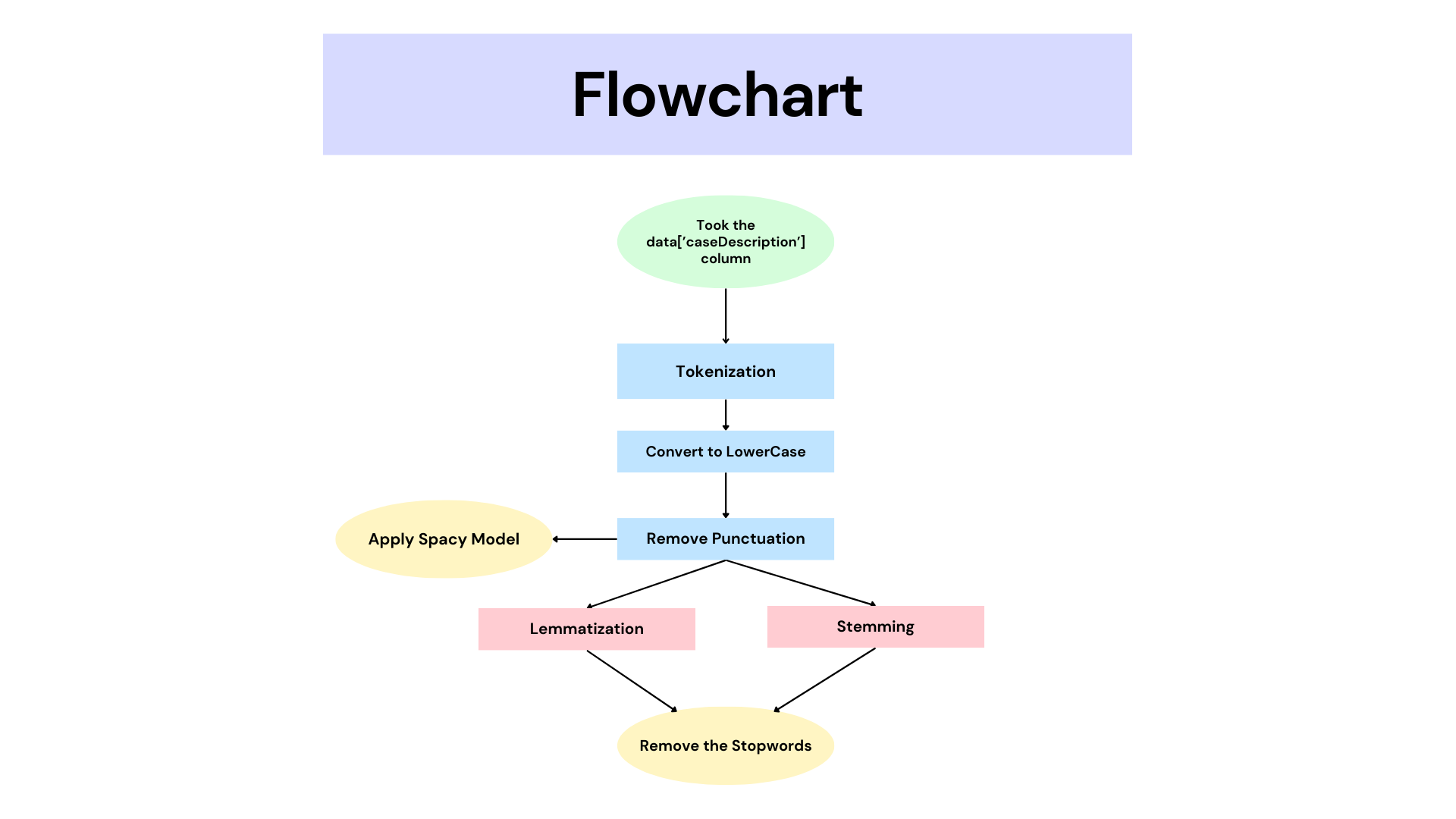
* **Enhanced the visual design** and usability of the web application interface to ensure a better user interaction.
* Incorporated additional features and optimizations to streamline user navigation and functionality.

**Project Documentation:**

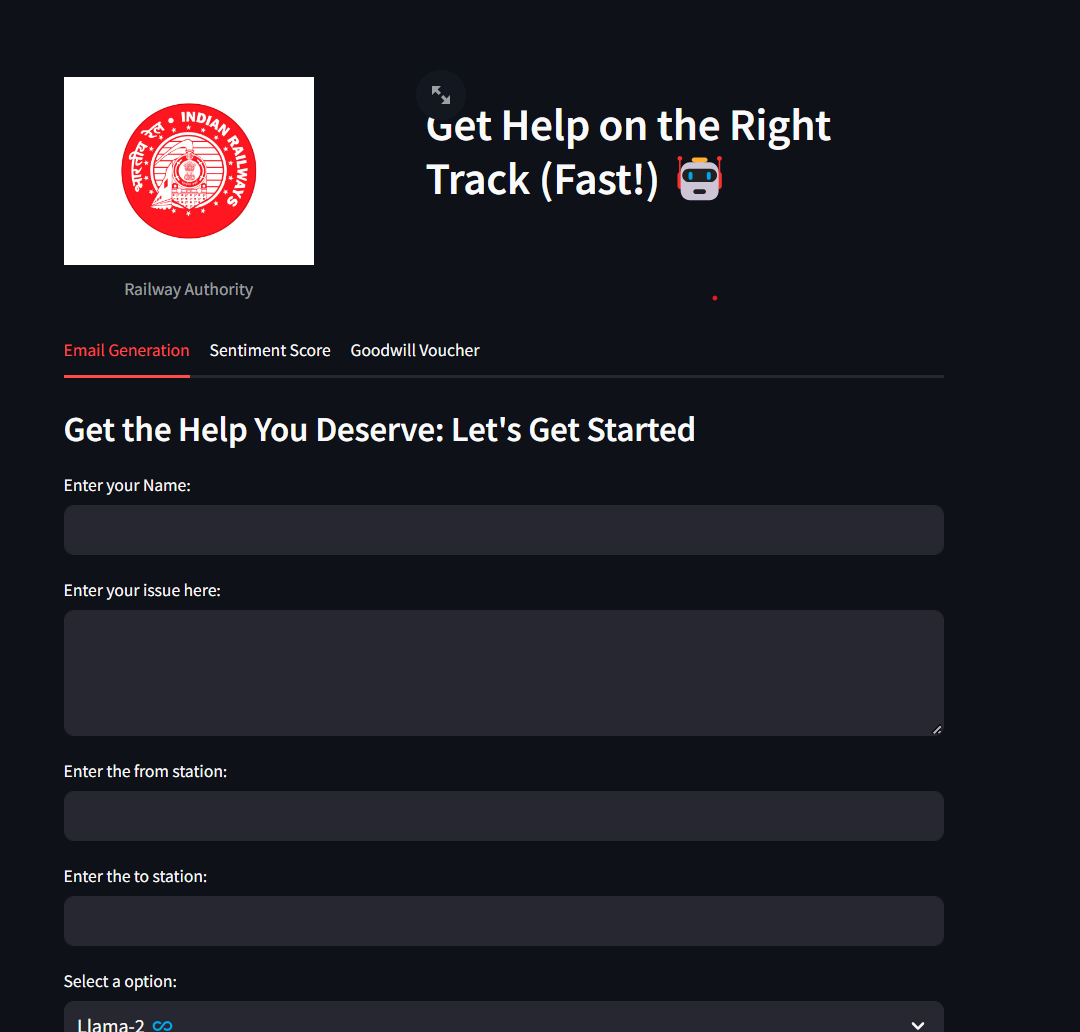
* Created detailed documentation outlining the project objectives, methodology, results, and conclusions.
* Included technical details, user guides, and troubleshooting information.

# Results and Discussion

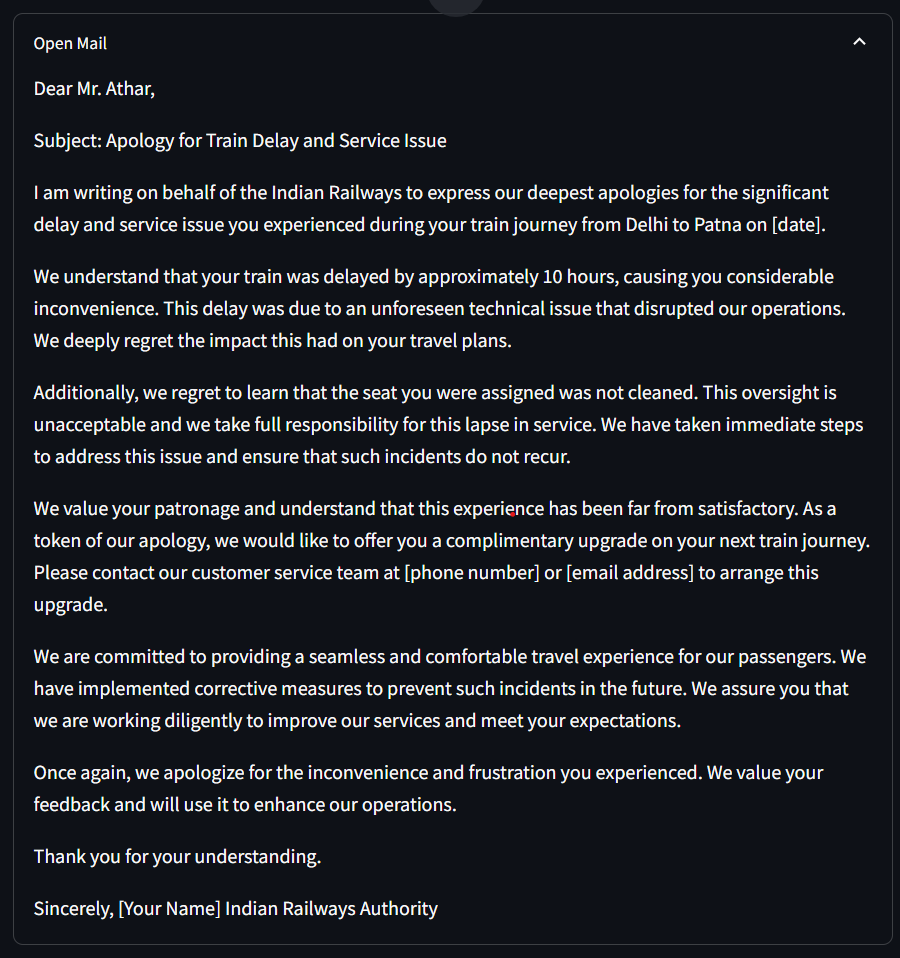
## Keyword extraction Flowchart



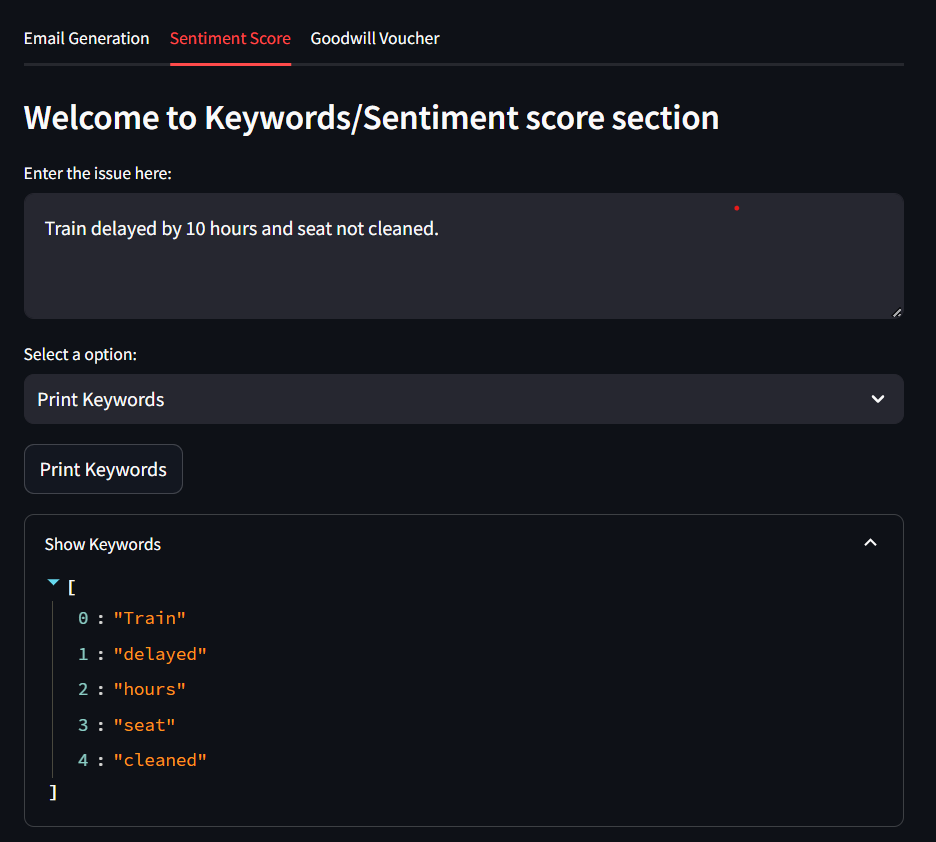
## Streamlit Web Application:



## Automated Email Generation:



## Keyword Extraction:



## Good Will Voucher Prediction: