Natural Language Processing Course End Project

Start Assignment

- Due Oct 29 by 8pm
- Points 100
- Submitting a file upload

COURSE 5

Career Simulation Projects

Streamlining the Customer Grievance Process

Context:

As a data scientist working for the front office of a major American multinational bank, you are responsible for enhancing customer service and ensuring compliance with financial regulations. Your current assignment involves analysing the customer complaints the bank has received over the past year.

Problems:

The current time-consuming manual process for daily triaging and reviewing of customer complaints
The complaints data is currently underutilized in enhancing the quality of products and services.

Objective:

The goal is to use NLP techniques, such as text classification and sentiment analysis, to efficiently gain insights into the underlying causes of customer grievances. By leveraging these methods, we aim to better understand and address customer grievances, ultimately improving our grievance redressal process.

Steps to be done:

1. Prepare text data using appropriate NLP techniques.
2. Efficiently identify the primary factors behind customer grievances using sentiment analysis.
3. Convert insights from sentiment analysis into actionable business strategies for Retail

Banking.

Data understanding:

1. Read data in python environment.
2. Check if the variables have correct datatypes. Make changes wherever necessary.
3. Find the date range
4. Define a function named preprocessing that executes the following series of pre-processing steps in order:
   - Convert text to lowercase
   - Remove numbers
   - Remove stopwords
   - Remove punctuation
   - Apply lemmatization
5. Clean the text under 'Complaint Description' using the above function
6. Convert the pre-processed text into a matrix of TF-IDF features for downstream modelling.
7. In order to effectively manage the process, it is critically important to categorise the complaint and pass on to the concerned product department. Consider the department as a target variable and build a classification model.
8. Use SentimentIntensityAnalyzer to predict sentiments from the complaints. The SentimentIntensityAnalyzer is a class from the vaderSentiment library designed for sentiment analysis. It evaluates text to determine the sentiment scores across four categories: positive, negative, neutral, and compound. The compound score is a normalized value between -1 (most extreme negative) and +1 (most extreme positive), providing an overall sentiment rating. This analyzer is particularly effective for social media and other informal texts, as it can interpret emoticons, acronyms, and slang. It is widely used for tasks like sentiment classification, opinion mining, and customer feedback analysis. Its ease of use and accuracy make it a valuable tool in NLP.
9. How can the score be used by the bank? Share your insights.

Input dataset: Dataset (https://docs.google.com/spreadsheets/d/1oY7aXafWfNCrBZwk8Lb6_m2rdDpFotaD/edit?usp=drive_link&ouid=100315178870191646624&rtpof=true&sd=true)

NLP Project Rubric

Criteria	Ratings			Pts
1. Data Understanding & Preprocessing	**20 pts** **Full Points** Data types corrected, date range identified, and preprocessing function implemented with all required steps.	**10 pts** **Incomplete** Some preprocessing steps missing or not clearly implemented. Minor issues with data handling.	**0 pts** **No Points** No preprocessing or incorrect handling of data types and missing date range.	20 pts
2. Text Feature Engineering	**20 pts** **Full Points** TF-IDF matrix correctly created and used for downstream tasks.	**10 pts** **Incomplete** TF-IDF matrix created but contains issues or is not clearly used for modeling.	**0 pts** **No Marks** No TF-IDF features created or completely misapplied.	20 pts
3. Complaint Classification	**20 pts** **Full Points** Classification model built with proper workflow and department as target variable. Results evaluated clearly.	**10 pts** **Incomplete** Model built but missing evaluation or department classification not clearly justified.	**0 pts** **No Marks** No model built or irrelevant target used.	20 pts
4. Transformer-based Modeling & Sentiment Analysis	**20 pts** **Full Points** Transformer model used effectively, and SentimentIntensityAnalyzer applied with results interpreted.	**10 pts** **Incomplete** One of the two methods implemented or results not interpreted thoroughly.	**0 pts** **No Marks** No transformer model or sentiment analysis used.	20 pts
5. Business Insights & Strategy Recommendations	**20 pts** **Full Points** Clear, actionable insights derived from analysis. Shows deep	**10 pts** **Incomplete** Insights mentioned but lacking depth or clear linkage to business	**0 pts** **No Marks** No meaningful insights or strategy recommendations	20 pts

Criteria	Ratings			Pts
	understanding of how the bank can use sentiment scores.	strategies.	provided.	
				Total Points: 100