|  |  |
| --- | --- |
| Internship Project Title | Automate sentiment analysis of textual comments and feedback |
| Project Title | Automate sentiment analysis of textual comments and feedback |
| Name of the Company | TCS ION |
| Name of the Industry Mentor | SWAPNA NIKALE |
| Name of the Institute | B.K. Birla College, Kalyan |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort (hrs.) | Project Environment | Tools used |
| 7/09/2023 | 20/10/2303 | 980hr | * Python as the primary programming language. * Windows operating system (as indicated by the file path in the code). * Commonly used Python libraries, including TextBlob, pandas, and matplotlib. | * Python: The primary programming language for the project. * TextBlob: A Python library for natural language processing and sentiment analysis. * pandas: A Python library for data manipulation and analysis. * matplotlib: A Python library for data visualization. * A dataset in CSV format from Kaggle is used as input data for sentiment analysis. |
| **Project Synopsis:**  The project appears to be developed in a local development environment using Python as the primary programming language.  The code snippet provided is running on a Windows operating system as indicated by the file path  (C:\Users\Viraj\Downloads\comments\sentiment-analysis.csv). It also uses various Python libraries and modules for text  analysis and data visualization. | | | | |
| **Solution Approach**:   * Load the textual comments and feedback data from a CSV file. * Analyze the sentiment of each comment using TextBlob and classify them as positive, negative, or neutral based on a polarity threshold. * Calculate the total count of each sentiment category and the average percentage distribution. * Save the results to a new CSV file. * Generate a sentiment analysis report and save it to a text file. * Create a bar chart to visualize the sentiment distribution. | | | | |
| **Assumptions:**   * The code assumes that the input dataset is in a specific CSV format with a column named "Text, Sentiment, Source, Date/Time, User ID, Location, Confidence Score." * It assumes that sentiment can be categorized as positive, negative, or neutral based on a polarity threshold of 0.2 and -0.2. * Non-string values in the dataset are assumed to be labeled as "N/A." * It assumes that the dataset is pre-cleaned and doesn't perform any additional data cleaning. | | | | |
| **Project Diagrams:**  No specific project diagrams are provided in the code or information 1 Graph is mentioned in code | | | | |
| **Algorithms:**  The project utilizes the TextBlob library for sentiment analysis, which internally uses a pre-trained model to analyze the sentiment of textual data.  The sentiment is determined based on the polarity score, which is compared to a threshold value to classify it as positive, negative, or neutral. | | | | |
| **Outcome:**  The project generates sentiment analysis results in the form of a CSV file ('results.csv'), which includes the sentiment labels for each comment.  It also produces a sentiment analysis report ('sentiment\_report.txt') with information about the average sentiment percentages, sentiment counts, and sentiment distribution.  A bar chart is created to visually represent the sentiment distribution. | | | | |
| **Exceptions considered:**  The code handles non-string values by labeling them as "N/A" to prevent errors during sentiment analysis.  The code does not explicitly handle other exceptions or errors, so it may not be robust to issues like missing or incorrect data in the input CSV. | | | | |
| **Enhancement Scope:**  Some potential enhancements to the project could include:  Handling data cleaning and preprocessing to improve the accuracy of sentiment analysis.  Providing more detailed insights, such as sentiment changes over time.  Developing a web-based or GUI application for easier interaction with the sentiment analysis tool.  Exploring different sentiment analysis libraries or models for potentially improved accuracy.  Adding additional visualization options or interactive charts for deeper analysis. | | | | |
| **Link to Code and executable file:**  I will provide the source code, dataset, as well as the output in DOCX format. | | | | |