**INTERNSHIP: PROJECT REPORT**

------------------------------------------------------------------------------------------------------------------------------------------

|  |  |
| --- | --- |
| Internship Project Title | RIO-45: Automate detection of different sentiments from textual comments and feedback |
| Project Title | Sentiment Analysis |
| Name of the Company | TCS iON |
| Name of the Industry Mentor | Mr. Debashis Roy / Rushikesh |
| Name of the Institute | University of Waterloo |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort (hrs.) | Project Environment | Tools used |
| 14.06.2021 | 23.06.2021 | 50 | Python | NLTK, Re, Sklearn |
| Project Synopsis: | | Sentiment Analysis model, using NLTK, for cleaning the input, and Multinomial Naïve Bayes algorithm, for sentiment score generation, which returns ‘Positive’, ‘Negative’ or ‘Neutral’, depending upon the inputted paragraph. It has been tested to have an accuracy of 86%. | | |
| Solution Approach: | | Cleaned the input paragraph text using NLTK module, so as to remove the unnecessary phrases and vectorized it to make it computer-understandable. Research was done on various suitable algorithms suitable for sentiment detection. Support Vector Machine (SVM) and Naïve Bayes were found to be the best. The cleaned text was passed into an instance of Multinomial Naïve Bayes, to get the sentiment score. Faced a few issues executing the plan, initially, but I did overcome them, eventually. | | |
| Assumptions: | | Paragraphs containing both Positive and Negative sentiments are not inputted. If it does happen, the model will return an output (Positive or Negative) favoring the sentiment which occurs in more number of words, instead of returning Neutral. | | |
| Project Diagrams: | |  | | |
| Algorithms: | | (in order)   * Cleaning the inputted paragraph, by removing stopwords (from NLTK module); * Vectorization; * Multinomial Naïve Bayes for Sentiment detection | | |
| Outcome: | | By applying the above algorithm, the generated model yields 86% accurate output . | | |
| Exceptions considered: | | * Empty input will not throw an erroneous output. | | |
| Enhancement Scope: | | * The accuracy can be improved further. This can be attained by adding more elements into the ‘train.pkl’ file. * Model can be enhanced to be able to yield the correct output for inputs with mixed sentiments | | |
| Link to Code and executable file: | | <https://github.com/madhuv-sharma/sentiment-analysis/> | | |