**INTERNSHIP: PROJECT REPORT**

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

|  |  |
| --- | --- |
| Internship Project Title | TCS iON: Rio 45- Automate detection of different sentiments from textual comments and feedback |
| Project Title | TCS iON: Rio 45-Automate detection of different sentiments from textual comments and feedback |
| Name of the Company | TCS iON |
| Name of the Industry Mentor | Debashis Roy |
| Name of the Institute | IIT Kanpur |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End date | Total effort | Project environment | Tools used |
| 17-07-2022 | 25-07-2022 | 45 | Chrome, Window 10, | Python, Google colaboratory, Ms word |

|  |
| --- |
| **Project Synopsis:**  Objective of this project is to develop a deep learning algorithm to detect different types of sentiment contained in a collection of English Sentences or a large paragraph. The main aim is to understand the deep learning process how the model will automatically detect the review is positive or negative. The process of developing the model is divide into several process like text cleaning, preprocessing, Model fitting, and then testing the model and used different libraries for this project. I have used the IMDB review dataset downloaded from Kaggle containing the review and the sentiments labelled as 1, 0 as positive and negative respectively. |
| **Solution Approach:**   * Taking any dataset for sentiment analysis like imdb movie review dataset or tweet dataset from kaggle or any other webpage. * Then analysis the data having two column one is review column having feedback on movies and the sentiment column having categorical text (positive and negative) then checked that is there any null value in the data or not otherwise fill it or remove it if it is less and then convert this sentiment column in “1” for positive and “0” for negative. * I did the text cleaning and preprocessing of the data. In text cleaning , i used the stopwords in english from nltk module to remove the english word and punctuation from text then i used the porterstemmer for stemming of the words. Then i tokenized the word by using tokenizer. Then padding a tokenized words. * Initialising a sequential model. Spliting the data into testing and training data into ratio of 2:8 using scikit library. Import the modules of sequential and different layer like embedding, LSTM, Dense and Dropout from tensorflow. * Then fit the model on training data with epoch=25 and batch size 200 and predict the Y value for the test data. And then compute the accuracy (y\_test and y\_pred) by using sklearn.metrics is 83%. |
| **Solution Workflow**  The workflow diagram is given below: |
| **Link to code and dataset**  Data set link:https://drive.google.com/drive/folders/1lYareqBuxNHtobOTnURLosz3bfiqHwCS?usp=sharing  Code link:  https://drive.google.com/drive/folders/1lYareqBuxNHtobOTnURLosz3bfiqHwCS?usp=sharing |
| **Results:**   * I was able to successfully complete the sentiment analysis on the movie review with implementing the sequential modelling. * I was able to attain the accuracy of 84.6% which is pretty good |
| **Enhancement Scope :**  Sentiment analysis has been an important tool for brands seeking to gain a deeper understanding of their customers' thoughts and emotions. It is a relatively straightforward form of analytics that assists brands in identifying key areas of weakness (negative sentiments) and strengths (positive sentiments). Sentiment analysis is gaining traction within other organizations. During Brexit and the 2016 United States presidential election, these data tools were utilized to measure emotions and predict the outcomes of these events. This has caused non-brand organizations to implement sentiment analysis for their own purposes. In addition, the insights gained from these tools are becoming much more indepth due to the emergence of social media platforms and functions. The future of sentiment analysis will continue to delve deeper, far beyond the number of likes, comments, and shares, to reach and fully comprehend the significance of social media interactions and what they reveal about consumers behind the screens. This forecast also predicts that sentiment analysis will be used by individuals in the public eye, governments, nonprofits, educational institutions, and numerous other organizations. The other machine learning algorithms can also be applied such as naïve bayes etc to increase the accuracy of the model. |

1