**INTERNSHIP: PROJECT REPORT**

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

|  |  |
| --- | --- |
| Internship Project Title | Automate detection of different sentiments from textual comments and feedback |
| Project Title | Text Emotion Detector |
| Name of the Company | TCS ion |
| Name of the Industry Mentor | Debashis Roy |
| Name of the Institute | FEROZE GANDHI INSTITUTE OF ENGINEERING AND TECHNOLOGY |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort (hrs.) | Project Environment | Tools used |
| 16/06/2020 | 26/06/2020 | 45 | Spyder IDE, | Python IDE, Github |
| Project Synopsis:  To develop a deep learning algorithm to detect different types of sentiments contained in a collection of English Sentence or large paragraph. | |  |  |  |
| Solution Approach:  Step 1: Setting up the Development Environment  Step 2: Choosing Your Dataset  Step 3: Understanding What’s Inside Your Dataset  Step 4: Preprocessing the Data  Step 5: Feature Extraction  Step 6: Training Our Models  Step 7: Testing with different sentence and checking the accuracy and output. | |  |  |  |
| Assumptions:  We can get textual data from any website like a movie review website, or Amazon product reviews, and so on. Here, I’ll be using a labeled textual dataset consisting various tweets of different emotion. | |  |  |  |
| Project Diagrams:  C:\Users\RITIK SRIVASTAVA\Downloads\Screenshot_2020-06-23 Applied Machine Learning Part 3.png | |  |  |  |
| Algorithms:   1. Multinomial Naive Bayes Classifier 2. Linear SVM 3. logistic regression 4. Random Forest Classifier | |  |  |  |
| Outcome:  Different accuracy ratio of algorithm used are:   1. naive bayes count vectors accuracy 0.7764932562620424 2. lsvm using count vectors accuracy 0.7928709055876686 3. log reg count vectors accuracy 0.7851637764932563 4. random forest with count vectors accuracy 0.7524084778420038 | |  |  |  |
| Exceptions considered:   1. Extending further, one can think of replacing words with their most common synonyms. That could help in building better models. That is being skipped here. 2. The next biggest challenge with natural text is dealing with spelling mistakes, especially when it comes to tweets. Apart from that, what do we do about sarcasm or irony in the text? Due to the complexity of dealing with these issues we treat them as exception. | |  |  |  |
| Enhancement Scope:   1. Various popular algorithms can be used to get more scope of accuracy ratio which includes RNN (Recurrent Neural Network), LSTM (Long short-term memory), GRU (Gated Recurrent Units) etc. 2. Dataset of other field like Finance, Statistic, Education etc. can also be chosen to test versatile nature of the code. | |  |  |  |
| Link to Code and executable file:  https://github.com/rghf/TCS-AI-Internship/tree/master/final%20report | |  |  |  |

1