**INTERNSHIP: PROJECT REPORT**

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

|  |  |
| --- | --- |
| Internship Project Title | Rio-45 |
| Project Title | Automate Detection of different Sentiments from Textual Comments and Feedback. |
| Name of the Company |  |
| Name of the Industry Mentor | Debashis Roy |
| Name of the Institute | Sri venkateswara engineering college |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort (hrs.) | Project Environment | Tools used |
| 18-7-2022 | 31-7-2022 | 45 | Google collab | Google colab,numpy,scikit |

**Project Synopsis: -**

**Project Description:**

Project Objective & Brief: To develop a deep learning algorithm to detect different types of sentiment contained in a collection of English Sentences or a large paragraph.

1. **Approaches used For Text Classification:**

* 1.Rule Based Systems
* 2. Text Classification Algorithms :
* Some of the most popular machine learning algorithms used for creating text classification models include the naive bayes family of algorithms, support vector machines, Regressions, and deep learning algorithms with CNN and RNN.
* 3. Machine Learning based Systems:

**2)Performance metrics included:**

* **Accuracy**
* **Precision**
* **Recall**
* **F1 Score**

**Solution Approach:**

**Rule Based Systems:**

Rule-based approaches classify text into organized groups by using a set of linguistic rules.

Each rule comprises of a pattern based on semantics and its predicted category.

Machine Learning based Systems:

Text classification based on past observations.

By using training data, the algorithm can learn the different associations between pieces of text and that a particular output (i.e. tags) is expected for a particular input (i.e. text).

Feature extraction: Transforms each text into a numerical representation in the form of a vector. E.g. bag of words [a vector represents the frequency in a predefined dictionary of words ]

The algorithm is fed with training data consisting of feature sets.

Once trained with enough training samples, the machine learning model can begin to make accurate predictions on unseen text with similar feature sets.

**Text Classification Algorithms**

Some of the most popular machine learning algorithms for creating text classification models include the naive bayes family of algorithms, support vector machines, Regressions, and deep learning algorithms with CNN and RNN. Metrics and Evaluation Cross-validation is a common method to evaluate the performance of a text classifier.

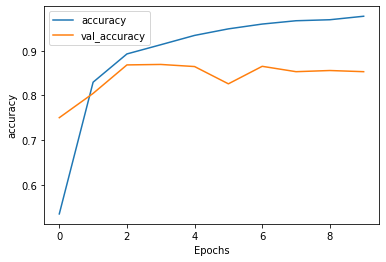
It consists in splitting the training dataset randomly into equal-length sets.

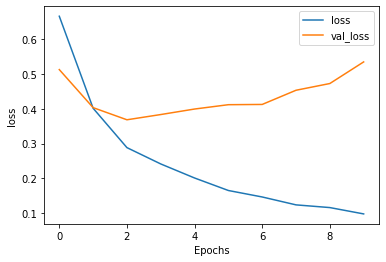
For each set, a text classifier is trained with the remaining samples (e.g. 75% of the samples).

The classifiers make predictions on their respective sets and the results are compared against the human-annotated tags.

With these results, a performance metrics is built, that are useful for a quick assessment on how well a classifier works.

**Project Diagrams:**

****



**Project Outcome & Link to Code and executable file** **:**

🡪[**ProjectLink**](https://colab.research.google.com/drive/1CRuKjsbvMzprlpfaC1nI2bCQZdi9s19Y#scrollTo=bxaLdNyFKRjZ) **(click here)**

**Data Sources to solve problem:**

Data Sets from any open source data sources can be used. Model building and solution engineering is NOT restricted to data.

Example:

* 1)https://ai.stanford.edu/~amaas/data/sentiment/
* 2)https://data.world/crowdflower/sentiment-analysis-in-text
* 3)https://www.kaggle.com/crowdflower/twitter-airline-sentiment
* 4)Test Accuracy achieved with IMDB dataset using following algorithms:

**Conclusion :**

Developed a **deep learning algorithm** to **detect different types of sentiment** contained in a collection of English Sentences or a large paragraph with an accuracy of **85%**.