|  |  |
| --- | --- |
| Internship Project Title | TCS iON RIO-125 : 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 | Atulya Kaushik |
| Name of the Institute | PrepInsta Technologies Pvt. LTd. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort (hrs.) | Project Environment | Tools used |
| 05-05-2023 | 03-08-2023 | 125 | Chrome, Windows 11,  Android 11 | MS Office |
| **Project Synopsis:**  This project aims to develop a deep learning algorithm that automates the detection and segmentation of different sentiments (positive, negative, and neutral) within a collection of English sentences or a large paragraph. The algorithm will leverage advanced techniques such as recurrent neural networks (RNN) or transformer-based models like BERT to effectively learn the underlying patterns and nuances in the text. The project will begin with the collection of a diverse and representative dataset of labeled English sentences or paragraphs covering a wide range of sentiments. This dataset will serve as the foundation for training and evaluating the deep learning algorithm. Next, the algorithm will be developed and fine-tuned using the labeled data. By learning from the relationships between the input text and the corresponding sentiment labels, the algorithm will gain the ability to accurately identify and segment sentiments within the text.  The training process will involve optimizing the model's parameters, selecting appropriate network architectures, and fine-tuning hyperparameters to achieve the best possible performance. Rigorous evaluation using appropriate metrics such as accuracy, precision, recall, and F1-score will ensure the reliability and effectiveness of the developed algorithm.The expected outcome of this project is a robust deep learning algorithm capable of automatically detecting and segmenting different sentiments within a given text. The algorithm will be accompanied by a detailed evaluation report, showcasing its reasonable accuracy and performance on unseen data.By automating sentiment analysis, businesses can efficiently process and analyze large volumes of textual comments and feedback. This, in turn, can provide valuable insights for decision-making, customer satisfaction improvement, and overall enhancement of business strategies and operations. | | | | |
| **Solution Approach:**  Rule based system:   * 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 * 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. | | | | |
| **Assumptions:**  1. Sufficient labeled data for training: Abundant labeled data available for training the algorithm on diverse sentiments.  2. Accurate sentiment labels: Reliable sentiment labels accurately represent expressed sentiments in training data.  3. Effectiveness of deep learning models: Deep learning models capture patterns and nuances in textual sentiment analysis effectively.  4. Adequate computational resources: Adequate computing resources available for efficient training and evaluation of the deep learning algorithm.  5. Generalization to unseen data: Algorithm generalizes effectively, accurately detecting and segmenting sentiments in diverse English text.  6. Reasonable accuracy expectation: Algorithm attains reasonable sentiment detection accuracy, accounting for variations in data quality and model complexity. | | | | |
| **Project Diagrams:** | | | | |
| **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.  1.It consists in splitting the training dataset randomly into equal-length sets.  2.For each set, a text classifier is trained with the remaining samples (e.g. 75% of the samples).  3.The classifiers make predictions on their respective sets and the results are compared against the human-annotated tags.  4.With these results, a performance metrics is built, that are useful for a quick assessment on how well a classifier works. | | | | |
| Outcome:  1.Algorithms to detect different types of sentiment from a paragraph.  2.Detailed presentation with proof of reasonable accuracy. | | | | |
| Exceptions considered:  Exceptions considered in automating sentiment analysis of textual comments and feedback include sarcasm/irony, ambiguity, negation/contrast, domain-specific language, misspelled/informal text, out-of-vocabulary words, and biased training data. | | | | |
| Enhancement Scope:  1.Real-time sentiment monitoring enables businesses to receive instant alerts when negative sentiments are detected, facilitating prompt actions to address customer concerns.  2. Social media integration enables real-time analysis and tracking of sentiments expressed across platforms, aiding businesses in monitoring brand reputation, tracking sentiment trends, and engaging with customers for improved satisfaction and brand perception | | | | |
| Link to Code and executable file:   1. <https://colab.research.google.com/drive/1qS0osCdJKgfL1gdtzjuKl6g2nijwC6Qi?usp=sharing> 2. https://colab.research.google.com/drive/1MrqQt4Qmz7UrQu5cbT-yqeOL8s3ynYoA?usp=sharing | | | | |