**SENTIMENT ANALYSIS ON CUSTOMER FEEDBACK**

Bhagya Sree Bokka-1002145347

Kireeti Mattaparti-1002099884

Poorna Pallavi Tekkali-1002160637

**INTRODUCTION:**

* Sentiment analysis is used to analyze raw text to drive objective quantitative results using natural language, processing machine learning, and other data analytics techniques. This involves instructing computers to recognize the sentiment conveyed in the text.
* The goal of studying Natural Language Processing (NLP) is to enable computers to process spoken and written language in a manner that is similar to that of humans. This field combines machine learning, linguistics, and computer science.
* NLP's sentiment analysis heavily relies on data mining methods to categorize thoughts and segment text into primarily binary classifications, such as positive and negative.
* Sentiment analysis can read beyond simple sentences and detect sarcasm, read common chat acronyms (LOL, ROFL, etc.), and correct common mistakes like misused and misspelled words.
* Our project mainly uses LSTM and GRU techniques to identify different patterns like whether the given text is positive or negative or neutral.

**DATA PRE-PROCESSING:**

* It is important to perform data cleaning and processing before carrying out data analysis.
* Cleaning involves not only enhancing and standardizing the dataset but also eliminating extraneous elements like noise and additional tags. Conversely, data transformation into a vector format, achieved through various preprocessing procedures, is a critical step.
* In the field of Natural Language Processing (NLP), pre-processing involves crucial techniques such as Stemming and Lemmatization, with the order of these tasks being significant. The dataset is processed in a structured sequence of steps.

Below are the steps we followed in Data pre-processing.

* Lower Casting
* Contrast Expand
* Noise Removal
* Correcting Typo
* Removing Stop words
* Lemmatization

**Model Implementation:**

* Long-Term-Short Memory (LSTM) and Gated Recurrent Unit(GRU) are types of recurrent neural networks that are particularly effective in processing sequential data, such as text, are the types of recurrent neural networks. In the context of sentiment analysis, for example in analyzing customer feedback, they are frequently applied.
* An analysis of customer feedback sentiment can be carried out on both the LSTM and GRU networks. A series of words representing the feedback text would normally be used as an input to the network.
* Using techniques such as word embeddings, all words in the sequence are usually represented by vectors.

**REFRENCES USED:**

•Pankaj, P., Pandey, P., Muskan, M., Soni, N.: Sentiment analysis on customer feedback data: Amazon product reviews. In: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 320–322 (2019)​  
<https://ieeexplore.ieee.org/abstract/document/8862258>​

•Summarization of customer reviews for a product on a website using natural language processing​  
<https://ieeexplore.ieee.org/abstract/document/7732392?casa_token=TBdd1zuycNkAAAAA:EHpZ-PfE3clFW-pEszLaSd9k2IA33OJvf1v8zP-HzyyNa99VYFZ8VcjlQn2LcJ9TSquPFHIICsE>​

•Mining of customer review feedback using sentiment analysis for smart phone product​  
<https://link.springer.com/chapter/10.1007/978-3-030-86165-0_21>​

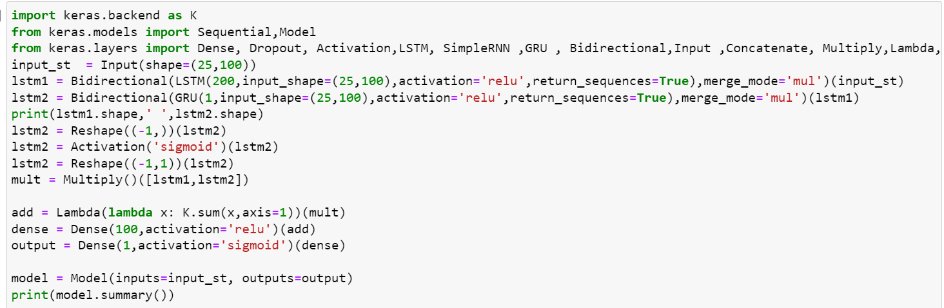
**DIFFERENCES**

* For sentiment classification, the papers are likely to use traditional machine learning algorithms such as Support Vector Machines SVM, Naive Bayes or Decision Trees. ​
* Deep neural network techniques, such as Convolutional Neural Networks and CNNs or Recurrent Neural Networks and RNNs to model sequences and sentiment analysis can also be applied.​
* Whereas our project completely uses LSTM and GRU techniques which are completely different from the techniques used in references.​
* LSTM and GRU models can capture temporal dependencies and contextual information in sequential data, making them well-suited for modeling natural language.​
* They are capable of learning from both short-term and long-term dependencies in text data, enabling them to capture complex linguistic patterns and semantics.​

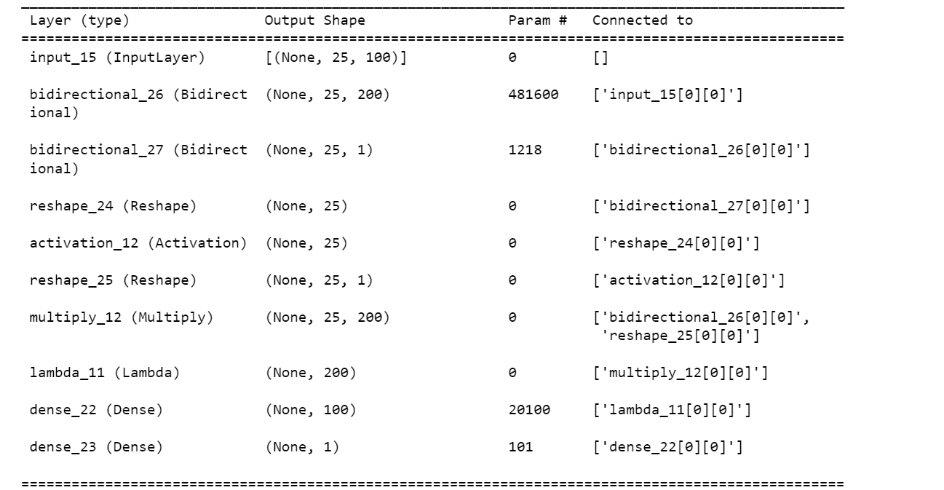
By incorporating word embeddings and learning from large amounts of text data, LSTM and GRU models can achieve state-of-the-art performance on various NLP tasks.

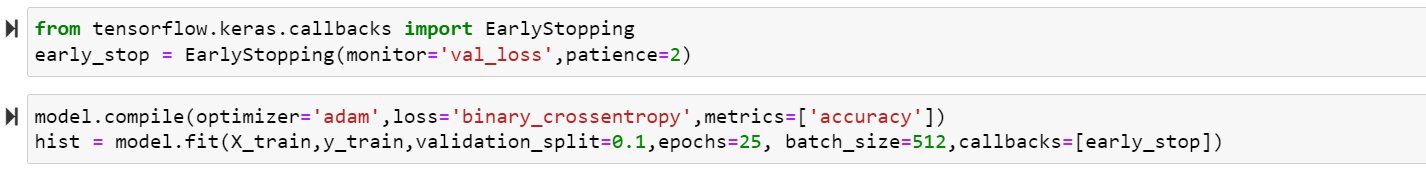
**ENHANCEMENT:**

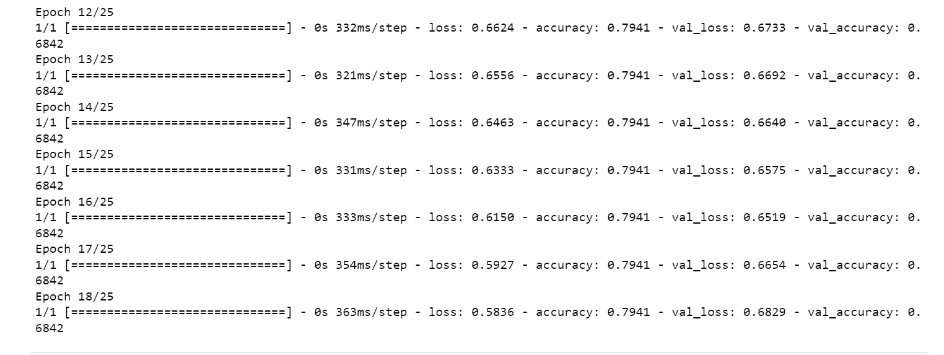
In NLP research papers, researchers often start with simpler models like Naive Bayes classifiers as baselines to establish a performance benchmark. They then explore more complex and powerful models like LSTM and GRU to improve accuracy and address the limitations of simpler models. LSTM and GRU models excel in capturing sequential dependencies and contextual information in text data, which can lead to enhanced accuracy compared to Naive Bayes classifiers, especially in tasks where context and semantics play a crucial role.​



We are implementing LSTM AND GRU directly from the Keras module. We are experimenting with different layers of bidirectional LSTM and bidirectional GRU. We are combining those layers to train our model to obtain enhanced accuracy.​

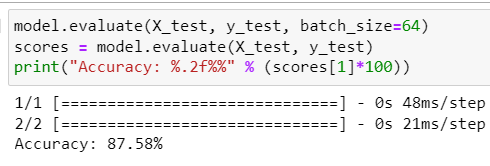






**ACCURACY**:

Initially we got 85% accuracy, however, when we changed the model with 5 epochs, and by setting the batch size as 64, we got 87.58% as the accuracy.



**CONCLUSION:**

* Initially, we were able to improve the accuracy to almost 87.58% of the LSTM model compared to the reference paper which had an accuracy of around 80%
* LSTM is very effective in capturing the long-term dependencies in sequential data accounting for the higher performance over all other models.
* Bi-directional LSTMs and GRUs are powerful tools for sequence-based tasks, when your data benefits from context in both directions and when you need a more efficient model with fewer parameters.