IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, April 2024

Sentiment Analysis using Deep Learning

Prof. R. N. Muneshwar[1], Mr. Yuvraj Prakash Hase[2], Mr. Pradip Balu Karwar[3], Mr. Shubham Nitin Hingmire[4], Mr. Bhushan Vilas Gopale[5]
Department of Information Technology[1,2,3,4,5]
Amrutvahini College of Engineering, Sangamner India

Abstract: *The use of online platforms and the internet is increasing daily. Businesses and political organizations might benefit from knowing public opinion when making strategic decisions. Given this, sentiment analysis is crucial for determining the polarity of the public's opinions. These days, this vast amount of data can be put to good use. Sentiment analysis of text posts can yield knowledge and information useful for social contexts, business intelligence, Internet of Things (IOT) mood-triggered devices, and citizen opinion polling.*

The sentiment analysis based on Emotional Recognition (ER) is the primary focus. The primary goal is to deploy the algorithm for sentiment prediction across all online platforms using emoji, length words, and generic terms. Additionally, we will contrast the conventional understanding of sentiment analysis. We can attempt to regulate some illicit actions posted on social media and movie review websites by using sentiment analysis.

The process is broken down into six parts in this model: feature engineering is the third phase, data preprocessing is the second, and data overview is the first. Model selection comes in fourth, model evaluation comes in fifth, and model deployment comes in last.

We can strive to get results in two classes: good and bad sentiment, or positive and negative. Two columns are attempted to be created: one for the data sample and one for the outcome. Thus, the forementioned procedure concludes that using the length words and emoji from the data will boost the sentiment analysis's accuracy.

Keywords: Sentiment, LSTM, NLP

I. INTRODUCTION

The online platform's continuously growing data can be used for a variety of analysis processes in order to meet targets or goals for a range of fields. The decision that will be made in the future can be predicted using the data. The information can also be used to analyze user sentiment. These days, users of online platforms can share information on anything with any emotion.so that a study of their mood or emotions is possible.

Examination of internet platforms (social networking, movie reviews, etc.) Sentiment text posts can provide knowledge and information useful in social contexts, corporate intelligence, and citizen opinion polling. Sometimes a person's feelings, whether positive or negative, can be used to anticipate their mental state or behavior. You can utilize online reviews of movies to determine if a film is good or awful. evaluations can be utilized by others to evaluate the film; certain evaluations may have a greater influence on other users or customers.

Reviews and analysis can also be used to forecast the success of other sectors. Certain death situations may be connected to an individual's activities on online platforms, wherever they may have delivered a particular remark. Sentiment analysis can be used to track certain activities or the globe at large, as well as to enhance the quality of the content.

A. Problem Statement

These days, a variety of reviews can be posted on websites, such as the sections dedicated to movie reviews. It is possible to forecast if a film will be good or terrible using the analysis of the review. In this situation, the feedback can also be analyzed using the different reviews. Sentiment analysis can be used to stop some criminal acts or forecast user

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, April 2024

reactions for businesses. Thus, we can attempt to create a natural language processing-based system for sentiment prediction. Thus, make an effort to employ length words in your analysis to improve system accuracy.

B. Objective

- Fundamental emotions/Emotions: Give the sentiment analysis system the opportunity to acquire the fundamental sentiments connected to particular domains so that it may comprehend and use them.
- Use in Business: Make the sentiment analysis system available for use in the business sector to improve the caliber of the product. For example, movie reviews are used to determine if a film is good or poor.
- Domain Expansion: To guarantee the sentiment analysis system's adaptability to a wide range of user needs, allow it to continuously expand its capabilities by learning data such as reviews in new domains, such as movie reviews, social media, or feedback forms
- Hybrid Procedural Knowledge: To expand the sentiment analysis system into other domains or to halt some criminal actions that can be detected through online platform reviews.

II. RELATED WORK

Certainly! In the context of sentiment analysis utilizing deep learning, let's compare Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks:

1. CNN:

Although their primary application is in image processing, CNNs can also be modified to perform text classification tasks such as sentiment analysis. CNNs use filters to convolve over the input text in text categorization in order to capture patterns and characteristics at different levels of abstraction. Word order is less important for jobs like sentiment analysis, where CNNs are good at identifying local patterns in the data. Long-range dependencies and comprehending the context of a sentence or document as a whole may be difficult for CNNs to capture. For training, they might need a lot of labeled data, which could be a problem in some situations. Nevertheless, long-range dependencies and sequential information included in text data may be difficult for CNNs to capture.

2. RNN:

RNNs are made to handle sequential data by storing information about past inputs in a hidden state. RNNs are able to collect word dependencies and contextual information over the course of the full sequence in sentiment analysis. But the vanishing gradient issue plagues conventional RNNs, making it difficult for them to identify long-term relationships in text data. Susceptible to issues with vanishing and inflating gradients, which may compromise the stability of training. Computationally more costly than CNNs, particularly when dealing with lengthy sequences.

3. LSTM:

LSTMs are a kind of RNN that uses gating methods to solve the vanishing gradient issue. Long-term dependencies in sequential data can be effectively captured by LSTMs because they preserve a cell state that allows them to selectively retain or forget information over time. LSTMs are useful for modeling complicated dependencies and capturing subtle sentiment information found in text data in sentiment analysis. By adding gating mechanisms, conventional RNNs can overcome their vanishing gradient issue. They work well for recording long-term dependence because they preserve a memory cell that allows them to selectively recall or forget information over time. And can adjust the cell state in an adaptable manner, making them suitable for handling sequences of different durations.

CNNs work effectively in sentiment analysis applications when feature extraction and local context are important considerations. They can be useful for tasks like sentiment categorization at the sentence level and are particularly good at identifying patterns in brief text sequences.

In sentiment analysis, RNNs including LSTMs are useful when comprehending long-range dependencies and the context of the entire text. Their ability to extract subtle sentiment information and contextual clues spanning several words or phrases makes them appropriate for applications such as sentiment analysis in conversation threads or document-level sentiment analysis.

DOI: 10.48175/IJARSCT-17478

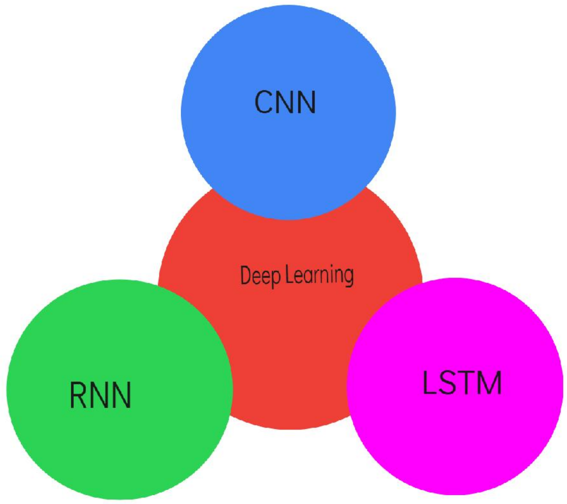

Fig. Models in Deep Learning.

The benefit of using LSTMs is that they can mitigate the vanishing gradient problem while capturing long-term interdependence. For sentiment analysis jobs involving lengthy text sequences or documents with intricate architecture, this makes them very helpful.

To summarise, the selection of CNNs, RNNs, or LSTMs for sentiment analysis is contingent upon the particular demands of the task, including the duration of text sequences, the relative significance of local versus global context, and the computational resources at hand. Every architecture has pros and downsides, and choosing the best one requires weighing trade-offs between computational effectiveness, model complexity, and performance for the particular sentiment analysis task at hand.

Materials and Methods

- The methodology of the study is covered in this part. The programming environments will be covered in the first section.
- We will look at the data collection process and location in the second section, along with the data preparation technique.
- The final portion will include specifics about the sentiment analysis algorithm that takes the longer word into account.

Programming Environment

One of the most widely used programming languages for data research, machine learning, and natural language processing (NLP) is Python. Numerous NLP and ML approaches are available in Python, which can be utilized to address a wide range of issues. Python was selected as the investigation's programming language because of its large library and practicality. One of the many Python tools that facilitate working with data in human languages is NLTK.It supports a number of necessary modules, including seaborn, pandas, matplotlib, NumPy, and beautiful soup. There are more feature extraction techniques included as well.

Dataset

The casual conversations that took place between various friend groups on Facebook, Twitter, and chat, as well as a movie review dataset, were the datasets used in this. It has been discovered that the majority of young people use informal text that is, brief messages with longer words to communicate their emotions. Since Python can handle these kinds of files more readily, the data files are prepared in the Comma Separated Values (CSV) format.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, April 2024

III. PROPOSED METHODOLOGY

Phases 1 through 5 are outlined in the methodology of the proposed system. The unstructured raw data from the various sources is used as the input for the tokenization stage.

Tokenization

During this stage, several delimiters are used to separate the data into tokens. Delimiters such as Space, Comma, Hash (#), and @ are typically utilized.

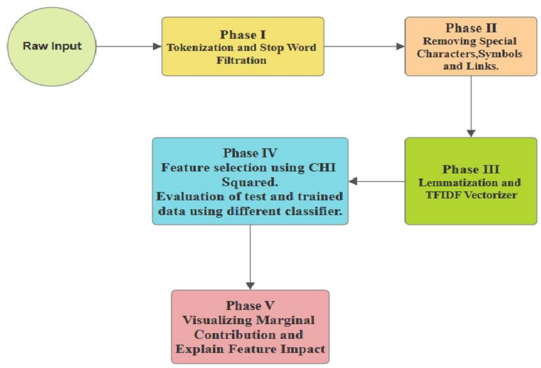

Fig: Architecture

Stop Word Removal

Stop word elimination is done in this phase using the tokens created during tokenization. A stop word is a term that is frequently used but has no inherent meaning, like "the." The sentiment is not determined in any way by these words. Thus, they are eliminated. Additional emojis are eliminated from the text as well.

Normalization

The remaining tokens are routed to this stage, which is where token normalization happens. The same tokens will be forwarded to phase IV concurrently. In the Senti WordNet, each content word has a Senti-score. In the Senti-Score generation process, the Senti-score of the normalized word will be taken from the Senti-WordNet.

LSTM

We started by cleaning our data before building the LSTM network. The model was constructed layer by layer in a sequential fashion using the Sequential API provided by Keras.

Our LSTM network's architecture is as follows:

- Layer of embedding: Our integer sequences, which stood for words, were changed into dense vectors of a set size by this layer. When managing categorical data especially textual data the Embedding layer plays a critical role.
- Layers of LSTM: Two LSTM layers, each with 128 units, were then added. Because LSTM layers can record the temporal connections between sequence items, they are useful for processing sequences, including text. Dropout was also incorporated throughout these layers to prevent overfitting.
- Dense layers: Two dense (completely connected) layers were inserted after the LSTM layers. The last one, which was the output layer, contained three units (matching to our three sentiment classes), while the first had 64 units. We obtained the probability for each class by using the'softmax' activation function in our output layer.

IV. RESULTS

The quality and quantity of the dataset, the LSTM model's architecture, the training process, and the assessment metrics employed are some of the variables that could affect the outcomes of sentiment analysis using LSTM. On common

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, April 2024

datasets such as IMDb movie reviews, LSTM models for sentiment analysis often attain accuracies between 80% and 90%. These figures, however, may differ based on the particular task and dataset.

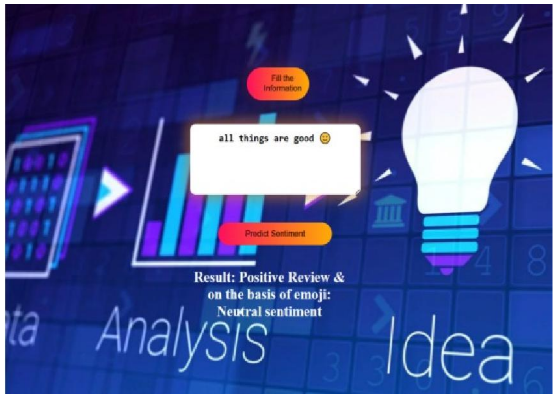

Fig. Result predicted

In the above image you can see that the simple text and a emoji which is "all things are good" is given as a input and the result is predicted positive with neutral sentiment based on text and emoji.

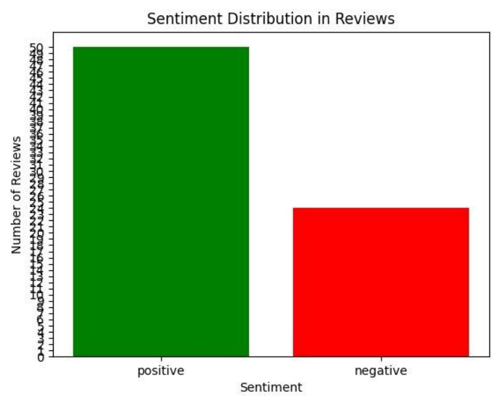

Fig. Positve and Negative Sentiments Predicted.

This image shows that how many positive and negative sentiments are predicted by the model. The positive sentiments are nearly 49 and the negative sentiments are nearly 26.

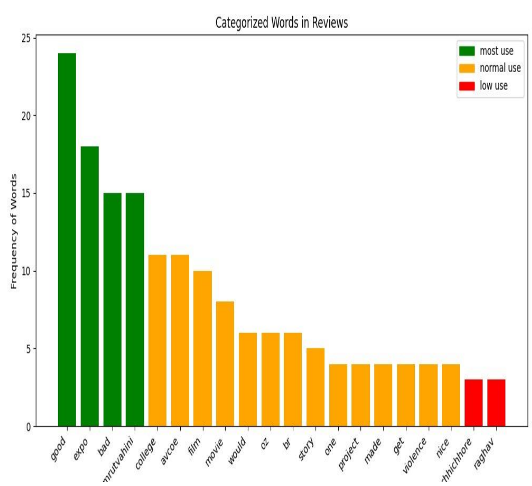

Fig. Word Frequency and Word Categorization.

In the above image we can see the frequency of words used and their categorization. The categorization is done in three types most used, normal used and low used. In this "good" word is mostly used with frequency near 24, "college" word is normally used with count near 11 then "raghav" word is low used with count 3.

V. CONCLUSION AND FUTURE SCOPE

To sum up, this project aims to develop a sentiment analysis system that is more user-centric and adaptable. It seeks to create a sentiment analysis system that can effortlessly grasp and execute a wide range of tasks while learning and adapting to new difficulties by focusing on functional requirements like context awareness, adaptability, and natural language understanding. Tight non-functional requirements pertaining to security, dependability, and privacy will guarantee the system's credibility and safeguard data at the same time. The aim of this project is to close the gap between present sentiment analysis system capabilities and user and business expectations. This will enable the creation of a system that can more accurately anticipate sentiment based on user reviews. Both social and business domains can use it.

We obtained remarkable results in predicting the sentiment of imdb movie reviews with an LSTM-based neural network. This approach provides a strong tool for figuring out consumer sentiment and directing corporate choices. Still, this is but one step in the vast field of NLP. Future research might look into alternative neural network topologies, make use of pre-trained embeddings, or use this model for additional natural language processing tasks

REFERENCES

[1] Improving Sentiment Analysis in Social Media By Handling Lengthened Words. Ashima Kukkar, Rajni Mohana, Aman Sharma, Anand Nayyar and Mohd. Asif Shah,2023

[2] M. Birjali, M. Kasri, and A. Beni-Hssane, ''A comprehensive survey on sentiment analysis: Approaches, challenges and trends,'' Knowl.-Based Syst., vol. 226, Aug. 2021.

[3] K. Jindal and R. Aron, ''WITHDRAWN: A systematic study of sentiment analysis for social media data,'' Mater. Today, Proc., vol. 2021

[4] A Conneau, H. Schwenk, L. Barrault, and Y. Lecun, ''Very deep convolutional networks for text classification,'2016.

[5] D. Tang, B. Qin, and T. Liu, ''Document modeling with gated recurrent neural network for sentiment classification,"2015.

[6] The Effects of Emoji in Sentiment Analysis Mohammed O. Shiha, Serkan Ayvaz.2017.

[7] Using Word Lengthening to Detect Sentiment in Microblogs, Samuel Brody, Nicholas Diakopoulos,2011