BioGecko Vol 12 Issue 03 2023

ISSN NO: 2230-5807

Twitter Sentiment Analysis Based on Ordinal Regression

Mrs. S. Radhika[1], Dr.V.Anantha Krishna[2], M. Bhoomika[3], B. Nikitha[4], B. Priyanka[5]

[1]Assistant Professor, Sridevi Women's Engineering College, Hyderabad,India.
[2]Professor, Sridevi Women's Engineering College, Hyderabad,India.
[3,4,5]Computer Science and Engineering, Sridevi Women's Engineering College, B. Tech IV Year Hyderabad, India

Abstract
Finding out whether the writer encounters a positive or negative mindset requires a semantic analysis. It is possible to accumulate information and do analysis for categorization, forecasting, or knowledge extraction by expanding user interactions in online communities. Nous presents a Bi-LSTM and CNN algorithm for this categorization, using embedding texts as an emoticons-sense vector as inputs. The goal of NLP is to determine if a remark is favorable or unfavorable and to forecast a rating of five stars. Emojis enable a greater understanding of the value of a message. Emoji-sense vectors act as a foundation for the proposed CNN and LSTM techniques.

Keywords: Long Short-Term Memory, Semantic Classification, Social Network Analysis, Word Embedding, Emoji Sense Vectors, Natural Language Processing, Word2vec.

1. INTRODUCTION

Emojis are extracted from text throughout the pre-processing stage, and emoji-sense vectors are generated as an additional input. The process of transforming text into an individual, definitive shape is called text normalization. (For example, it corrects half spaces.) To generate root protagonists that are significant words for processing, lemmatization, stemming, and eliminating words with stops are used. The layer that integrates receives pre-processed data that is expressed as vectors of fixed length. Emoji sentiment data is used to create Emoji vectors. The mood is determined from 70,000 tweets that were collected and transcribed in 13 different European languages by 83 human annotators. Emoji-sense and a pair of text vectors were given to the model in separate stages. A bi-LSTM layer stack preserves word connections rather than concentrating on sequential order. One-dimensional CNNs are implemented for gathering local variables in a series.TheSoftMax algorithm layer forecasts a rating of five stars, whereas the sigmoid activation layer shows the likelihood of a factor. Both the CNN and LSTM scores are combined to get the final results. Greater generalization and less excessive fitting are the results of dropout layers.

2. RELATED WORK

A key component of deep neural networks involves successful word representation. A number of numerous antiquated techniques, including TFIDF, bags of words (which have a very large dimension), and preconditioned embedded words like Word2Vec. SACPC is a newly developed, shallow method for short text like comments that distinguishes senses using PLTSs and SVMs. Transformers are at the precise edge of the analysis of texts. An additional way of expressing words is by using transformers, which are then sent to the Bi-LSTM with care to gather the sense of emotion. Additional approaches to address the challenge include adversarial networks with generative algorithms, reinforcement training, sentiment-specific embedding of words models, cognition-based concentration models, intuitive thinking, and bidirectional encoder representations from transformers (BERT).Emoji embedding is an original method for creating emoji embedding for separately designing emoji for both good and negative emotive tweets.

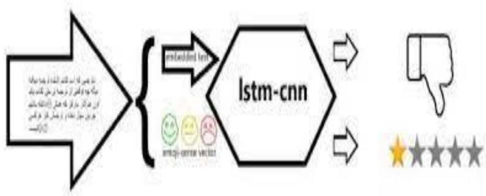

Figure: Model baseline

Existing System

In the field of NLP, sentiment analysis is a frequent job. Academics have employed both lexicon-based and machine learning techniques for sentiment analysis. The phrase "lexicon methods" relates to a precise numerical assessment of the emotion a word or phrase evokes, whether it be good or negative. These techniques determine the orientation of a document based on the semantic orientation of its words or phrases. Some ML strategies, like SVM, are used in machine learning techniques.Learning word vectors, mega-representing, automated approaches for sentiments similarity, and unstructured strategies for meaning similarity are further methods towards sentiment analysis. A crucial element of deep neural networks is effective word representation. There are several dated methods, including TFIDF, bag of words (which has a very large dimension), and preconditioned word embedding technologies like Word2Vec.

Disadvantages Of Existing System

1) The effectiveness of machine learning has decreased.
2) It's essential for complex neural networks to effectively depict text.
3) Everyone is unable to forecast or recognize international characteristics using the current available methods.

Problem Definition

In the field of NLP, sentiment analysis is a frequent job. Machine learning and lexicon-based methods for sentiment analysis have both been used by academics.The phrase "lexicon methods" refers to a precise numerical assessment of the emotion that a word or phrase elicits, whether it be either favorable or adverse.These techniques determine the orientation of a document based on the semantic orientation of its words or phrases. Some ML strategies, like SVM, are used in machine learning techniques.

Proposed System

With integrated text and emoji-sense vectors as inputs, we recommend a Bi-LSTMand CNNmodel to determine whether a remark is favorable or negative and predict a 5-star ranking. Emojis facilitate a deeper comprehension of the meaning behind a communication. Emoji-sense vectors serve as the foundation for the suggested CNN and LSTM methods.

Advantages of Proposed System

1) Emojis and embedded text are input into the CNN and LSTM output, and the model makes predictions after a few dense and average levels.
2)Emojis facilitate a deeper comprehension of the meaning behind a communication.
3) The CNN and Bi-LSTM ensemble enhance the resilience and precision of the model.

Algorithms

Convolutional Neural Network (CNN):

Nous will develop a 6-layer deep neural network that can distinguish between a single picture and another in order to show how to construct a convolutional neural network-based image classification. We will construct an exceedingly tiny network that can also be operated by a central processing unit (CPU).The traditional neural networks that are excellent at classifying images possess numerous additional attributes as well as requiring plenty of time for training on a standard processor. Nevertheless, our goal is to demonstrate how to use TensorFlow to create a convolutional neural network that can be utilized in the real world.

Activation processes come in a variety of forms. Sigmoid is a well-liked triggering function. The term "sigmoid neuron" refers to a neuron that activates using the sigmoid function. Neurons are classified into different types, such as RELU and TanH, based on their firing mechanisms. A layer, which is the next building component of neural networks, is created by stacking neurons in a continuous line. The picture below features layers.

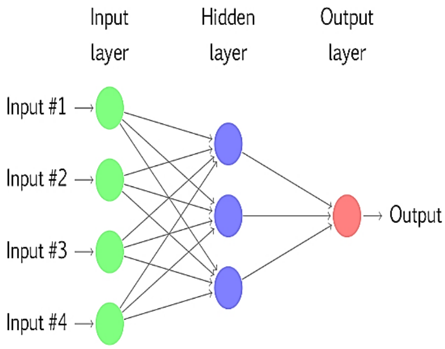

Figure: CNN layer

Long Short-term Memory (LSTM):

The term "long short-term memory" (LSTM) refers to an artificial RNN design used in deep learning. The LSTM has feedback links, in contrast to conventional feedforward neural networks. It can handle complete data sequences as well as individual data elements (such as images) and formats (such as speech or video). As an illustration, LSTM can be used for jobs like linked, unsegmented handwriting recognition, voice recognition, and anomaly detection in network data, also known as IDSs (intrusion detection systems). An input gate, an output gate, an ignore gate, and a cell make up a typical LSTM device. The three gates control how information enters and leaves the cell, and the cell can retain numbers for any length of time.

LSTM networks are particularly suited to classifying, processing, and creating forecasts based on time series data since there may be ambiguous lags between major events in a time series. In order to solve the disappearing gradient problem that might occur when training standard RNNs, LSTMs were developed. Due to their relative insensitivity to gap length, LSTM frequently outperform RNNs, hidden Markov models, and other sequence learning methods.

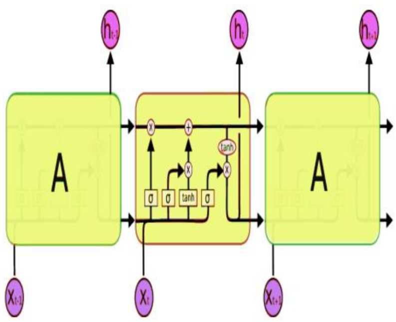

Figure: LSTM layer

BioGecko
Vol 12 Issue 03 2023
ISSN NO: 2230-5807

Modules
1) Data exploration
2) Processing
3) Splitting data into train & test
4) User signup & login
5) User input
6) Prediction

System Architecture

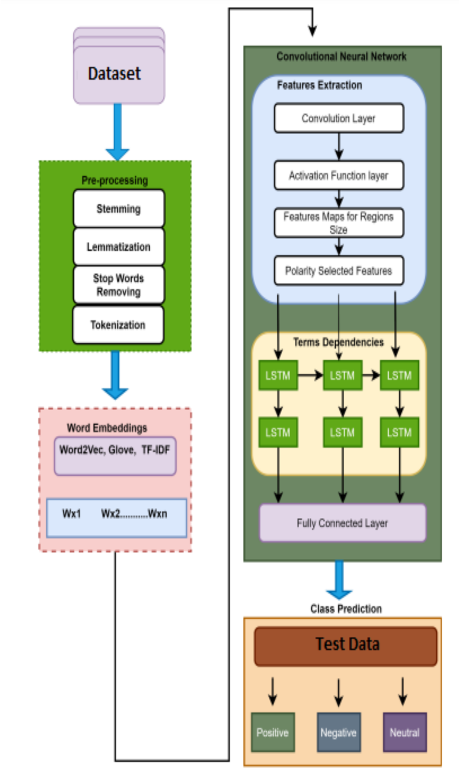

Figure: Architectural view of performing sentiment analysis.

Implementation
1) Install Python 3.7 IDLE before starting the project.
2) Use the dataset to execute the command below after download.
3) Launch the server with python app.py from the command line.
4) Start your browser and type the URL in the command window.

ISSN NO: 2230-5807

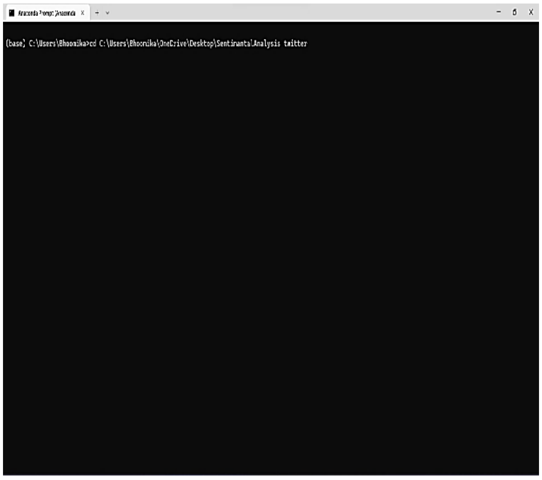

Figure: Command prompt used for running algorithms with dataset.

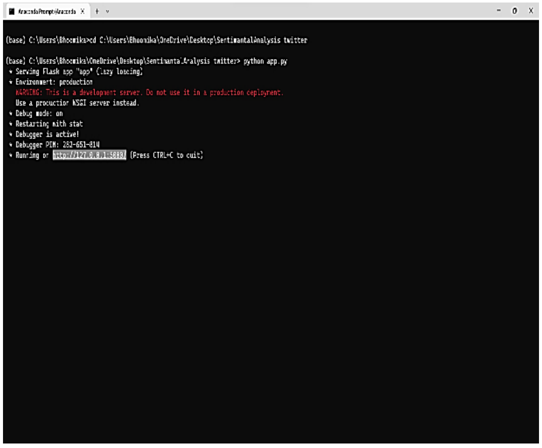

Figure: Copying the URL from command prompt.

Evaluation of Strategies

In the method of testing, test data is stored and utilized to evaluate each module separately before validation in the field is applied. The system testing that follows ensures that every aspect of the system operates as a cohesive whole. The test evidence should be selected so that it can withstand any scenario. Evaluation is actually the stage of installation designed to make sure the system operates correctly and effectively before the real operation starts. The techniques for evaluation used during the evaluation time are described in the paragraphs that follow.

The following testing techniques were used in this project:

1) **System Testing:**Testing an entire, fully functioning software application is referred to as "system examination."
2) **Module Testing:**Module testing is a form of software evaluation that examines specific categories, processes, subdivisions, or subroutines within a program.
3) **Integration Testing:**This stage of software evaluation involves combining and evaluating different software elements individually.

3. RESULTS

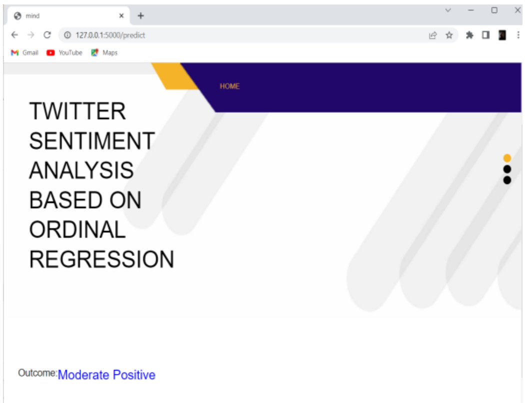

Figure: Result-1

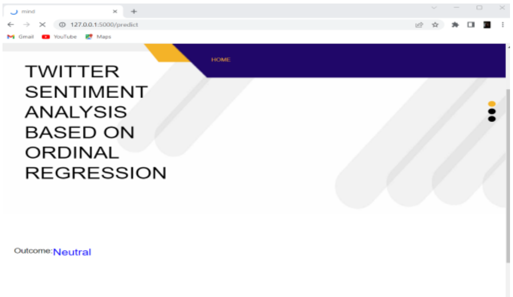

Figure: Result-2

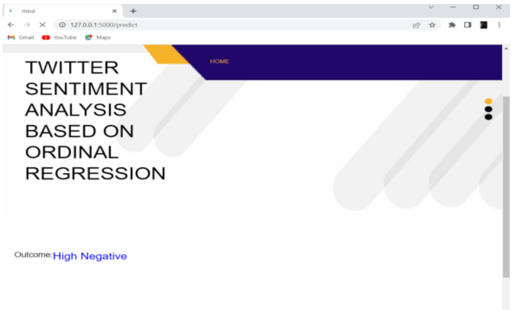

Figure: Result-3

4. CONCLUSION

Because CNNs learn local qualities first, the more CNN layers that exist in a collection of CNNs, the greater the worldwide features are collected. Two distinct points of view are put forth for these two models when processing remarks. Studies show that using ensemble models enhances model precision and resilience. Emojis play a big part in text comprehension in this study. Emojis are combined with a vector, which is then fed to thick layers with processed and inserted text. Results demonstrate the direct impact of the emoticon vector on output. Results indicate that efficiency can be successfully improved by using word embedding. The results may change depending on the social media. Emojis are used more frequently and are simpler to categorize with greater precision in remarks on Twitter or Instagram.

Future Aims

Machines that have been trained can also contend with the most advanced systems for an identical job. The effectiveness of the framework is hypothetically improved by using a preconditioned model. I'm confident that the effort to forecast and delve further into the meanings of emoji will become its own part of mood assessment as they grow into an even more crucial component of interpersonal interaction. Other organizations in the industry can build on this initiative if it turns out to be helpful enough.

5. REFERENCES

1. P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," Association for Computational Linguistics, pp. 417-424, 2002.
2. S. Baccianella, A. Esuli and F. Sebastiani, SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, Valletta, Malta: European Language Resources Association, 2010.
3. H. Kanayama and T. Nasukawa, "Fully Automatic Lexicon Expansion for Domain-oriented Sentiment Analysis," Association for Computational Linguistics, p. 355–363, 2006.
4. M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," Computational Linguistics, vol. 37, no. 2, pp. 267-307, 2011.
5. T. Mullen and N. Collier, "Sentiment Analysis using Support Vector Machines with Diverse Information Sources," Association for Computational Linguistics, pp. 412-418, 2004.