



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 5, May 2022

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.118

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)



# Sentimental Analysis using CNN

Aanchal Saxena<sup>1</sup>, Puneet Sharma<sup>2</sup>, Deepak Arora<sup>3</sup>

Department of Computer Science & Engineering, Amity University, Uttar Pradesh, Lucknow Campus, India

**ABSTRACT:** Sentiment analysis is a machine learning tool that analyses texts for positive or negative. Machines automatically learn how to detect sentiment without human input by training machine learning tools with examples of emotions in text. Sentiment analysis is performed by using techniques like Natural Language Pro-cessing (NLP), Machine Learning, Text Mining, and Semantic Approach. Support Vector Machine (SVM), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Random Forest, Nave Bayes, and Long Short-Term Memory (LSTM) are several machine learning ways employed in sentiment analysis. Sentiment analysis focuses on a text's polarity (positive, negative, or neutral), but it can also detect specific moods and emotions (angry, pleased, or sad), urgency (urgent, not urgent), and even intents (interested v. not interested).

Reason for sentimental analysis is every company is trying to figure out the sentiment of its customers if customers are satisfied, they will stay. This research paper could show a path to reduce customer churn. This is a report based on the concept of Computer Vision and Sentimental Analysis. In this report, authors have used computer vision to detect the emotion a person is exhibiting. The system should be able to determine the user's mood based on the emotion detected through the review.

With this report author has tried to segment IMDB reviews whether the movie gains positive vote or negative. This helps the directors and producers to get knowledge of what the audience wants from them, which type of movie is most supported. The genre which interests people etc. To do the sentimental analysis first search for a suitable dataset to do the research. Author searched for the dataset to train the system from Kaggle. After finding the dataset, author started to code. Before giving the review as an input to the model author performs following pre-processing steps: The length of each review should be made equal for the model to be working correctly, Author chooses the length of each review to be 500, If the review is longer than 500 words author is going to cut the extra part of the review, If the review is contains less than 500 words author pads the review with zeros to increase its length to 500.

Author concludes that the accuracy of the model is 85%. Author now picks up a random review and the result comes out to be negative. i.e. negative review. Author's model is now able to predict the sentiment of the review.

**KEYWORDS:** Machine Learning, Algorithms, CNN.

## I. INTRODUCTION

The method of analysing a user's emotion is known as sentiment analysis [1]. You can categorize their emotions as positive, negative, or neutral. It is a great research paper to understand how to perform sentiment analysis and its usage nowadays. Sentiment analysis is a machine learning tool that analyses texts for positive or negative. Machines automatically learn how to detect sentiment without human input by training machine learning tools with examples of emotions in text. [1]

Sentiment analysis is performed by using techniques like Natural Language Pro-cessing (NLP), Machine Learning, Text Mining, and Semantic Approach. Support Vector Machine (SVM), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Random Forest, Nave Bayes, and Long Short-Term Memory (LSTM) are a number of the machine learning ways employed in sentiment analysis [2].

For example, if we want to see if a product meets client needs or if there is a market for it, we can use this method. We can apply sentiment analysis to track the reviews for that product. Sentiment analysis is especially useful when dealing with vast amounts of unstructured material that we want to categorize by automatically labelling. Net Promoter Score (NPS) surveys are widely used to learn how customers feel about a product or service. Sentiment analysis has also grown in prominence as a result of its ability to analyze massive numbers of NPS data rapidly and consistently.

One of the biggest issues in natural language processing is sentiment analysis or opinion mining, which is the automatic extraction of the writer's attitude from the text. With over 7,000 articles written on the issue, it has become a major focus for the scientific community. Sentiment analysis algorithms are used across numerous social and review aggregation websites as a key aspect of the user interface. However, the scope of Sentiment Analysis applications extends much beyond that. It gives firms information by providing immediate feedback on items and measuring the effectiveness of their social marketing tactics. Similarly, it can be quite useful in political campaigns or any other platform that deals with public opinion. It can even be used in the stock market and algorithmic trading engines.



It is worth noting that proper sentiment analysis entails more than just determining the overall sentiment of a text or a particular paragraph. For example, in product reviews, the author normally does not confine his viewpoint to a particular feature of the product. The evaluations that cover various elements and provide a full list of pros and negatives are the most informative and beneficial. As a result, it is critical to be able to extract sentiments at a granular level and link each sentiment to the relevant element. On a higher level, the study can uncover complex attitude types rather than just positive or negative attitudes.

Deep learning algorithms have recently demonstrated outstanding performance in natural language processing applications, such as sentiment analysis, across numerous datasets. These models do not require pre-defined features chosen by an engineer; instead, they can learn sophisticated features from the dataset on their own. Although each individual unit in these neural networks is quite simple, these models are capable of learning more intricate decision limits by stacking layers of non-linear units on top of each other. The feature extraction is left to the neural network, which represents words in a high-dimensional vector space. As a result, these models may map words with comparable semantic and syntactic qualities to neighboring areas in their coordinate system, similar to how people comprehend words. Recursive Neural Networks, for example, are capable of grasping the structure of phrases efficiently. Deep learning models are a natural fit for a task like sentiment analysis because of these properties.

## II. RELATED WORK

Sentiment analysis serves as a key aspect of that deal with customers on online portals and websites for the companies. They always do that to classify the comment as a query, complaint, suggestion, opinion, or just love for the product. So, they can easily go through the comments or questions and give priority to what they need to manage first and even order them in a manner that looks better. Companies can even sometimes try to delete content that has a negative sentiment connected to it [2].

It is a simple way for understanding and analysing public reception and perception of various ideas and concepts, such as a newly launched product, an event, or a government policy.[3] Sentiment Analysis is also used for spam filtering and eliminating unwanted content.

Sentiment analysis can be classified depending on the scope into document-level sentiment analysis, sentence level sentiment analysis, and sub sentence level or phrase level sentiment analysis. Also, a common classification is based on what should be done with the data or a reason for the sentiment analysis. Examples of which are: Simple classification of text into positive, negative, or neutral. It might also advance into fine-grained responses like positive or somewhat positive.

- **Convolutions**

Convolutions are window functions applied to a matrix that achieve specific results (e.g., image blurring, edge detection). The window here is called the kernel, filter, or feature detector. The visualization shows six 3x3 kernels multiplying their elementwise values through the first matrix and then adding them. CNNs are just multiple layers of convolutions with activation functions like relu that allow nonlinear relationships to model. By applying this set of dot products, we will retrieve appropriate information from images, from edges at more superficial levels to identifying whole objects at deeper neural network levels. Unlike traditional neural networks, which simply smooth the input, CNNs can extract spatial relationships that are particularly useful for image data[3].

- **Customer Segmentation**

As per customization of people in our daily lives, customer segmentation enables a company to customize its relationships with the customers[4]. In marketing, a corporation might segment shoppers or buyers according to the standards of Segmentation associated & a broad array of causes. Demographic Segmentation includes gender, age, occupation, marital status, income. Geographic Segmentation includes country, state, city of residence and Specific towns or counties. Technographic Segmentation includes technologies, software, and mobile devices. Psychographic Segmentation includes personal attitudes, values, interests, and personality traits. Behavioural Segmentation includes actions or inactions, spending/consumption habits, feature use, session frequency, browsing history, and average order value [7].

## III. IMPLEMENTATION

To do the sentimental analysis first search for a suitable dataset to do the research[4]. Author searched for the dataset to train the system from Kaggle. After finding the dataset, author started to code. Consider Figure.1, Import the basic python libraries, e.g.- NumPy, pandas, from sklearn import suitable libraries.

```
In [ ]: import tensorflow as tf

In [ ]: import tensorflow.keras as keras
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

**Figure 1. Importing the dataset**

After importing the necessary libraries now author will read the data files, we have two data files here. first data file contains the imdb reviews and their corresponding sentiments which can be either positive or negative, author is going to use this file as our training data. the second file is also similar to the first file, but author is going to use it as the test data. The string data cannot pass to author's model directly, so author need to convert string data to the integer format. For this, map each distinct word as a distinct integer for e.g.{'this':14 , 'the':1}. Author already has a file that includes mapping from words to integers, so author loads that file.

The word index file contains mapping from words to integers. Next author converts the word index data frame into a python dictionary so that we can use it for converting our reviews from string to integer format. Now author define a function review encoder that encodes the reviews into integer format according to the mapping specified by word index file. Author splits the reviews from their corresponding sentiments so that author preprocess the reviews and sentiments separately and then later pass it to the model.

Before transforming the reviews as integers author tokenizes or splits the review on the basis of whitespaces for e.g., the string "The movie was wonderful" becomes ["The" , "movie" , "was" , "wonderful" ]. Since author tokenized the reviews now; apply the review encoder function to each review and transform the reviews into integer format. Figure 2 depicts that if sentiment is positive assign 1, if negative then 0. Encode train labels and test labels. Then to balance the length of each review we use Keras. Now author does word embedding.

```
In [17]: def encode_sentiments(x):
if x=='positive':
return 1
else:
return 0

train_labels=train_labels.apply(encode_sentiments)
test_labels=test_labels.apply(encode_sentiments)
```

**Figure 2. Encoding the sentiments**

Before giving the review as an input to the model author performs following pre-processing steps: The length of each review should be made equal for the model to be working correctly, Author choses the length of each review to be 500, If the review is longer than 500 words author is going to cut the extra part of the review, If the review is containing less than 500 words author pads the review with zeros to increase its length to 500.

Word embedding layer is of length 16 which convert each word into vector of length 16 and author converts each review into length of 500. Hence a model is created[9]. Now the author creates a model since it is neural network model for this author divides it into layers. Author uses relu for activation layer, and Keras to build the model. First layer is word embedding layer, second one is global average pooling layer, third layer is a dense layer with extreme hidden units which helps the activation function, and final is output layer which uses sigmoid as activation function[6].



```
In [21]: #training the model
history=model.fit(train_data,train_labels,epochs=30,batch_size=512,validation_data=(test_data,test_labels))

Epoch 1/30
49/49 [=====] - 3s 68ms/step - loss: 0.6918 - accuracy: 0.5525 - val_loss: 0.6891 - val_accuracy: 0.68
82
Epoch 2/30
49/49 [=====] - 3s 62ms/step - loss: 0.6819 - accuracy: 0.6914 - val_loss: 0.6726 - val_accuracy: 0.72
99
Epoch 3/30
49/49 [=====] - 6s 114ms/step - loss: 0.6532 - accuracy: 0.7538 - val_loss: 0.6357 - val_accuracy: 0.7
691
Epoch 4/30
49/49 [=====] - 3s 70ms/step - loss: 0.6032 - accuracy: 0.7767 - val_loss: 0.5833 - val_accuracy: 0.77
66
Epoch 5/30
49/49 [=====] - 3s 64ms/step - loss: 0.5390 - accuracy: 0.8165 - val_loss: 0.5221 - val_accuracy: 0.81
80
Epoch 6/30
49/49 [=====] - 3s 68ms/step - loss: 0.4748 - accuracy: 0.8450 - val_loss: 0.4676 - val_accuracy: 0.83
62
Epoch 7/30
49/49 [=====] - 3s 63ms/step - loss: 0.4187 - accuracy: 0.8630 - val_loss: 0.4240 - val_accuracy: 0.84
42
Epoch 8/30
49/49 [=====] - 4s 72ms/step - loss: 0.3742 - accuracy: 0.8752 - val_loss: 0.3893 - val_accuracy: 0.85
80
```

Figure 3. Training the dataset

After creating model, author now compiles the model and use adam optimizer and binary cross entropy[8] loss as a loss function for the model and we use accuracy as a metric to evaluate our model. Author has designed the model and compiled also. Now author can train the model using the training dataset, as shown in the figure 3. Store each epoch in the history variable to forward training data and test data to execute it.

#### IV. EXPERIMENTAL RESULTS

In this research work the authors has tried to segment IMDB reviews whether the movie gains positive vote or negative vote. This helps the directors and producers to get knowledge of what the audience wants from them , which type of movie is most supported. It is believed that sentiment analysis is becoming increasingly relevant as the amount of large-scale online user-generated content grows. The use of social media data for prediction and forecasting is in high demand. Data sentiment analysis can uncover a wealth of useful information, and it has a wide range of applications, including public opinion analysis, product evaluations analysis, and so on. Authors textures that the outcomes of our sentiment analysis can provide additional motivation to other scholars.

After the completion of training the model, it has accuracy of 95% on training dataset and 88.58% on test dataset. From figure 4 , author concludes that the accuracy is 85%.

```
In [25]: loss,accuracy=model.evaluate(test_data,test_labels)

782/782 [=====] - ETA: 0s - loss: 0.2985 - accuracy: 0.88 - 2s 2ms/step - loss: 0.2974 - accuracy: 0.8
851
```

Figure 4. Display Accuracy

Pick a random review from dataset and test it for the model. As an output we can test if sentiment is positive or negative consider Figure 5. Use the model. Predict function to predict the output 0 -1. If the value is less than 0.5 then the model can be assumed to be negative, if more than 0.5 then it can be positive[1].

```
In [28]: index=np.random.randint(1,1000)
user_review=test_reviews.loc[index]
print(user_review)

Reviews      <START i just read an extra long review on the...
Sentiment                                negative
Name: 51, dtype: object
```

Figure 5. Sentimental analysis of a random review

```
In [29]: user_review=test_data[index]
user_review=np.array([user_review])
if (model.predict(user_review)>0.5).astype("int32"):
    print("positive sentiment")
else:
    print("negative sentiment")

negative sentiment
```

**Figure 6. Integer format of above review to check prediction of model**

Figure 6 tells that the sentiment of random review comes out to be negative. The proposed model is now able to predict the sentiment of the review.

## V. CONCLUSION

Today, machine-learning methods and algorithms are highly advanced and successful for solving many real-world problems. It is the most recent topic in the IT industry today. Machine learning makes a software system learn independently and makes it able of predict outcomes. It is being used extensively for data interpretation and for the storage, extracting, manipulating, and retrieving data from data warehouses[9]. In this paper, author proposed an approach to understand situations in the real world with a sentiment analysis on deep learning techniques. With the proposed method, it is likely to predict user satisfaction of a product, happiness with some environment or destroy situation after disasters. Recently, deep learning can solve problems in computer vision or voice recognition, and convolutional neural network (CNN) works good for image as well as text analysis and classification. With this research paper author has tried to segment IMDB reviews whether the movie gains positive vote, negative or neutral vote. This helps the directors and producers to get knowledge of what the audience wants from them, which type of movie is most supported. The genre which interests people; etc.

Despite the fact that this work is very new and fundamental right now and restricted too. The authors have some plans for the future. In future, Author will try to apply this sentimental analysis using CNN in real-time also. Productivity of the algorithms will then be assessed using various evaluation metrics. The authors have been working on mixing algorithms that are used in this work and are aiming to make a hybrid algorithm model, which will demonstrate more accuracy. In the field of machine learning, work on future involves the utilization of all the machine learning techniques in each possible way[11].

## REFERENCES

- [1] X. Ouyang, P. Zhou, C. H. Li, and L. Liu, "Sentiment Analysis Using Convolutional Neural Network," 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, 2015, pp. 2359-2364, Doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.349.
- [2] Rani, S.; Kumar, P.: A sentiment analysis system to improve teaching and learning. Computer **50**(5), 36–43 (2017)
- [3] A. Zouzou and I. E. Azami, "Text sentiment analysis with CNN & GRU model using GloVe," 2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS), 2021, pp. 1-5, doi: 10.1109/ICDS53782.2021.9626715.
- [4] Shirani-Mehr, H.: Applications of deep learning to sentiment analysis of movie reviews. Technical report, Stanford University (2014)
- [5] Sharma, Y.; Mangat, V.; Kaur, M.: A practical approach to sentiment analysis of Hindi tweets. In: Proceedings of 1st International Conference on Next Generation Computing Technologies, pp. 677–680 (2015)
- [6] Du zhengLei. Sentiment analysis for short text of micro-blog [D]. BeiJing: Beijing Information Science and Technology University, 2013.
- [7] B Pang and L Lee, "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval, vol. 2.1–2, pp. 1-135, 2008.
- [8] A Krizhevsky, I Sutskever and G E Hinton, "Imagenet classification with deep convolutional neural networks", Neural Information Processing Systems, pp. 1097-1105, 2012.



- [9] A.U. Rehman, A.K. Malik, B. Raza and W. Ali, "A Hybrid CNN-LSTM Model for", Improving Accuracy of Movie Reviews Sentiment Analysis. *Multimed. Tools Appl.*, vol. 78, pp. 26597-26613, 2019.
- [10] Ain, QuraTul et al. "Sentiment Analysis Using Deep Learning Techniques: A Review." *International Journal of Advanced Computer Science and Applications* 8 (2017): n. pag.
- [11] Rehman, Anwar Ur et al. "A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis." *Multimedia Tools and Applications* (2019): 1-17.



**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.118**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details