

# OTT v/s THEATRES: Ensemble Model of Opinion Mining on Social Media

Hiran Christa Bel B<sup>1</sup>, Raghavendra P<sup>1</sup>, Nithish Sai C<sup>1</sup>, Dr. J Sirisha Devi<sup>2</sup>

<sup>1</sup>Students, Department of Computer Science and Engineering

<sup>2</sup>Assistant Professor, Department of Computer Science and Engineering  
Institute of Aeronautical Engineering  
Dundigal, Hyderabad, Telangana-500043

**Abstract.** *Opinion Mining has gained interest lately due to the uptick in social media. People began expressing public and private outlooks on varied subjects and sharing their perspectives on online forums. In this report, we concentrate on the implementation of opinion classification with a long short-term memory (LSTM) network and with a convolutional neural network (CNN) deep learning model. Moreover, we present an ensemble model (LSTM-CNN) for opinion mining. This paper reviews the public mindset towards the OTT platforms and movie theatre with an LSTM-CNN ensemble model of Opinion Mining. Thus, Results categorize users' viewpoints through comments into "Ott biased" and "movie theatre biased" illustrated in a pie chart.*

**Keywords:** opinion mining, ensemble model, ott, theatres, lstm, cnn, sentiments.

## 1 Introduction

Sentiment Analysis or Opinion Mining is the text analysis technique that can analyze the text/tweets on social media. With this natural language processing (NLP) technique, we can determine if the data is positive, negative, or neutral in the state. One of the applications of these techniques involves the breakdown of social media messages automatically, based on the sentiments and feelings expressed. It plays a crucial role in commerce and research fields as it applies to any domain. So, this analysis is performed on the textual data to help monitor the businesses with their products and brands through customer feedback. Sentiment Analysis aids in improving decision-making, customer satisfaction, etc.

Sorting public statements or surveys manually is a perplexing task. A large amount of data is mined to obtain the necessary analytics. Hence, sentiment mining assists in analyzing unstructured data with greater efficiency without costing a lot of money. Based on your requirement, you can represent it according to your classifications to suit your conditions. Opinion Mining permits enterprises to comprehend their customer's intents by mining their feedback, replies, and social media discussions.

Many generic errors are triggered while operating on a single model while forecasting. To overcome these pitfall ensemble classifiers are introduced. This report seeks to convey advanced sentiment-primary techniques and tools to get analytics from the social forums. Corresponding to a single model, ensemble learning also permits better accuracy. General idea is to understand a collection of experts (classifiers) and is permitted to choose. Ensemble approaches hold more elevated predictive accuracy, corresponding to the models. Ensemble techniques are useful when there is unstructured knowledge in the dataset; Since various models converge to deal with this type of data. So, you will utilize ensemble learning ways once you need to boost the performance of machine learning models. as an instance to decrease the mean absolute error for regression models or extend the preciseness of classification models. The whole also results in a very stable model.

Online reviews are a generic way for users to share their thought on the services/products. With so much data, humans cannot read every review. With the changing lifestyles of people and the increasing use of smartphones with affordable internet services, OTT platforms are becoming increasingly widespread since the pandemic. The combat between cinema and OTT has both pros and cons. Streaming apps permit you to have entertainment in your comfort zone while watching a movie in a theatre has another level of joy to capture the cinematic experience. So, to find the people's perspectives regarding them, we need to analyze their comments with the LSTM-CNN ensemble model.

### **1.1 Problem Statement:**

The analysis of sentiments on data from social networks, namely Twitter or Facebook, is a research area with growing demands today. Even though significant work was accomplished, numerous challenges remain, including applying techniques developed for various data and certain data fields, decreasing time intervals, and enhancing model accuracy. In recent years, deep learning models have become widespread in the field of opinion mining, where their significant possibilities are showing. Several studies have focused solely on assembling a single model from a single (or a few) datasets in a relevant domain, for example, medical research, marketing tactics, and financial forecasts. When applied to specific fields, a single machine learning model is well-grounded. So, each deep learning technique has its pros and cons. The approach of combining two (or more) methods is introduced as a means of incorporating the advantages of both and thus fills some shortcomings of individual methods.

### **1.2 Objective:**

The purpose of this analysis is first, to study the ensemble model and second to examine the data according to the polarity of, "individuals who favor OTT" and "individuals

who favor movie theatre” and deliver the outcome utilizing an ensemble LSTM-CNN model of Opinion Mining.

## 2 Literature Review

Milene Dias Almeida and Vinicius Mothé Maia and Roberto Tommasetti studied developing a customized dictionary based on perceptual mapping, assembling the sentiment indicator according to a business enterprise or an organization. (Maia, 2021)

Most studies are conducted on the Twitter datasets as Twitter is the best online forum to collect users' sentiments. Opinion mining is a helpful tool for analyzing the user's tweets to get the required insights. (Sarlan, 2014) Özgür Ağralı and Ömer Aydın examined the Twitter sentiments about the user's opinion on the Metaverse when Mark Zuckerberg revealed to revise its name to Meta. An open-source, Vader (Valence Aware Dictionary and Sentiment Reasoner) was used to analyze the twitter sentiments. (Ağralı, 2021)

Alsayat researched the hybrid ensemble model using long-term short-term memory (LSTM) network over the Twitter datasets. This paper introduces custom ensemble learning with deep learning, and data polarization is done on the Crowd Flower platform. The experiment is performed on the three distinct types of datasets through various hidden layers and neurons as model parameters. (Alsayat, 2022)

Ghosh and Sanyal used feature selection methods namely Gini Index, IG, and Chi-Square are evaluated on three datasets (Electronic reviews, kitchen reviews, and IMDb movie reviews) individually as well as an integrated approach. At last trained the data with LR (logical regression), SMO, RF, and MNB classifiers with this feature vector model. (Ghosh, 2018)

Matthew Whitehead and Larry Yaeger worked on the problem of classifying the written text of public opinions. This paper reported the accuracy of ensemble models that gives higher efficiencies. (Larry, 2018) Vohra and teraiya worked on the sentiments of customers' reviews about the services and products. (Vohra, 2013) Tweetfeel to examine the real-time tweets. It is an exceptional Application. (Das, 2018)

Xia R, Zong C, and Li S proposed an ensemble framework combining different classifiers and subsets of algorithms. (Xia, 2011) They worked on the three different classifiers namely, SVM, NB, and ME. White et al. researched on the SVM. In the paper, they proposed four different ensemble techniques such as bagging, random subspace, and boosting. Through random subspace best execution was accomplished. (Whitehead, 2022)

### 3 Research Method

An opinion is a belief of a person regarding a specific thing that helps in decision-making. Due to advancements in technology and people are more actively providing their viewpoints on topics on online forums, which has made it an advantage for the business enterprises to analyse the users' reviews to get insights about their products/services. This research area is challenging because it is new and many problems must be solved. Analysing the sentiments is quite tough as many contradictory words have an opposite polarity to the result that was predicted.

Our report aims to know polarity towards the OTT and theatres by analysing the data. Due to the pandemic, people preferred OTT platforms, but now movies are being screened in theatres. So, currently what people are biased can be determined in this paper.

There are various techniques to perform sentiment analysis, mostly individual models are used to analyse the sentiments. As the technology keeps advancing many new features are developed to get accurate predictions. Ensemble Learning, Hybrid Models are the new techniques which are used to fulfil drawbacks of the individual model. In this paper we choose an ensemble (LSTM-CNN) model to analyse the users' sentiments with the help of GloVe from Tensor Flow.

The proposed classification methods outlined in the steps are precise as follows:

**1. Collection of Data:** We have collected the datasets from the YouTube comments by using the YouTube API. The name, comment, reply author, reply, these four columns are collected into spreadsheet using SpreadsheetApp. Then the reply author and reply columns are manually sorted into the name and comments columns.

**2. Pre-processing:** The raw data is an unsupervised data with noise and inconsistency. *Batchgen* is used to clean the datasets. With the help of regular expressions all the unwanted characters and emoji are removed from the text. The datasets are divided into "goodfile" and "badfile".

**3. Vocabulary:** By using *Glove* from Tensor Flow sentiments of the comments are polarized and vocabulary is built. Then the vocabulary is processed.

**4. Model Selection:** Selecting a model is a crucial step while working in a machine learning field. According to the requirements of the project the wise choice must be

made to make better predictions. We chose LSTM-CNN ensemble model. So, the datasets are trained on LSTM-CNN ensemble model. Datasets are split into batches and each batch is trained individually.

We selected LSTM and CNN to design an ensemble model because of their unique features that are mentioned below:

*Convolutional Neural Networks (CNNs):*

Primarily CNNs are designed for image classification that can distinguish pictures of dogs vs. Pictures of cat. But here, we use these networks to catch negative phrases. For example, “don’t like” / “don’t want”.

1. I **do not like** chocolate.
2. I **do not want** to eat.

*Long-Term Short-Term Memory (LSTMs):*

LSTMs are the networks that remember the data. It has a memory from which the decisions are made. LSTM networks can identify the changes in the text “*I love watching movies in theatres but I ended up hating it.*” By looking at this illustration, it has two meanings, firstly he used to love watching movies, normally other models capture this meaning which conflicts the sentiments but with LSTM networking can capture the true sentiments of the whole sentence.

### 3.1 LSTM-CNN Model:

Firstly, an embedded layer is created and the data is passed to LSTM layer. So, at this level of LSTM layer the whole sentence is captured without losing its original sentiments as it can remember the statements. For example, “I prefer Ott but now I want to watch in theatres.” Illustration states that “they said that they prefer Ott” initially but at the end of the sentence their viewpoints have changed. Now they want to watch a movie in theatre. Hence, LSTM layer captures this positive sentiment, “they like Movie theatre”. In CNN networking, a convolution layer and maxpooling layer is applied for each filter. After applying those layers, the pooled features are combined and final predictions are made

### 3.2 Flow Chart:

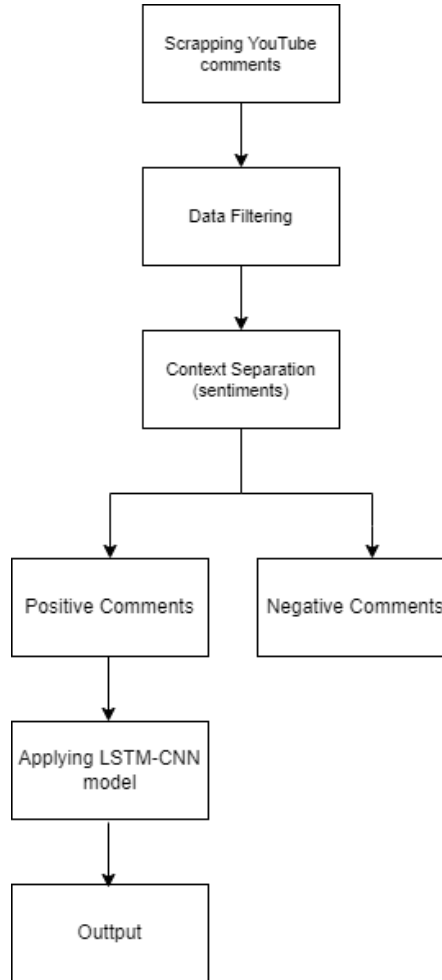


Fig (1) Flow Chart of the entire process

## 4 Results and Discussions

### 4.1 Plotting on the Chart

The percentage of people's outlook on OTT and theatres can be represented on a chart by importing pyplot from Matplotlib after applying the LSTM-CNN ensemble model. The pie chart displays the percentage of ``LIKE OTT" and ``LIKE THEATRE". In the figure (2), 43 percent of people are theatre biased and 57 percent of people are Ott biased.

From the chart statistics, it is evident that most of the people are in favor of OTT platforms than Theatres.

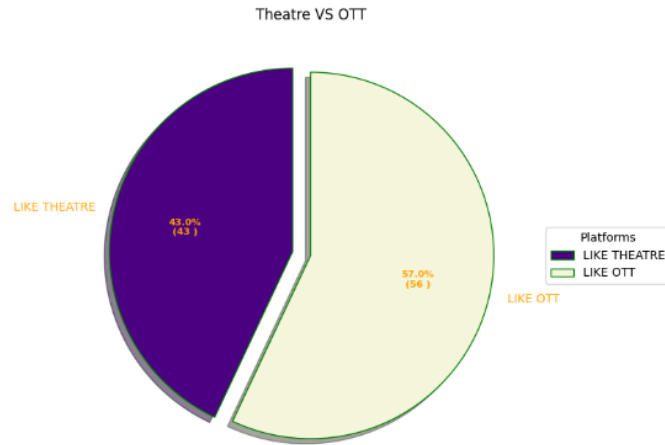


Figure (2) – Pie Chart representing the statistics of OTT vs THEATRE

## 5 Conclusion

The pursuit of this paper is to study the integrated model using LSTM and CNN models to design an ensemble model to enhance the performance of two individual machine learning models namely convolutional neural network (CNN) and Long-Term Short-Term Memory (LSTM). As mentioned, LSTM-CNN ensemble model yields effective outcomes. The methodology that we presented has some disadvantages are as follows:

- The datasets are collected from a lone source. These data are the comments scrapped from YouTube.
- analysis is done only on the positive insights. For example, this paper only talks about the people, who are biased towards either OTT or theatres.

## 6 References

1. Ağralı, Ö., & Aydın, Ö. (2021). Tweet Classification and Sentiment Analysis on Metaverse Related Messages. *Journal of Metaverse*, 1(1),25-30
2. Alsayat, A. Improving Sentiment Analysis for Social Media Applications Using an Ensemble Deep Learning Language Model. ("Dr. Ahmed Alsayat - Google Scholar") *Arab J Sci Eng* 47, 2499–2511 (2022). Doi: doi.org/10.1007/s13369-021-06227-w
3. Ben Chamblee- Data Scientist, Unsupervised Sentiment Analysis of Yelp Reviews Using Natural Language Processing, Springboard School of Data,15th June 2021.
4. Das, S.; Behera, R.K.; Kumar, M.; Rath, S.K.: Real-time sentiment analysis of twitter streaming data. *Proc. Comp. Sci.* 132, 956–964 (2018). Doi: <https://doi.org/10.1016/j.procs.2018.05.111>.
5. Ghosh M., Sanyal, G. An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning. ("Leveraging explainable machine learning to identify gait biomechanical ...") *J Big Data* 5, 44 (2018), Doi: <https://doi.org/10.1186/s40537-018-0152-5>
6. Milene Dias Almeida and Vinicius Mothé Maia and Roberto Tommasetti and Rodrigo de Oliveira Leite, Sentiment analysis based on a social media customized dictionary, *MethodsX*, 8,2215-0161 (2021). Doi: <https://doi.org/10.1016/j.mex.2021.101449>
7. Mining Using Ensemble Classification Models, Whitehead and Larry Yaeger, Indiana University, School of Informatics, 901 E. 10th St., Bloomington, IN 47408, mewhiteh, larryy@indiana.edu ("Sentiment Mining Using Ensemble Classification Models")
8. Pedro M. Sosa, Twitter Sentiment Analysis using lstm-cnn models, University of California, Santa Barbara, Graduate Student June 7th 2017, From: <https://konukoi.com/blog/2018/02/19/twitter-sentiment-analysis-using-combined-lstm-cnn-models/>
9. Sarlan, A., Nadam, C., & Basri, S. (2014, November). Twitter sentiment analysis. "In Proceedings of the 6th International conference on Information Technology and Multimedia (pp. 212-216)." ("Journal of Metaverse » Submission » Tweet Classification and Sentiment ...") IEEE. 10.1109/ICIMU.2014.7066632
10. Scrape Comments with replies from learning Orbis on September 21st 2020, [github:MAN1986/LearningOrbis/blob/master/scrapeCommentsWithReplies.gs](https://github.com/MAN1986/LearningOrbis/blob/master/scrapeCommentsWithReplies.gs)
11. Shangyan Li, Kevin Liu, HW1: Classification, Artidoro Pagnoni, February 1st, 2018, [apagnoni@college.harvard.edu](mailto:apagnoni@college.harvard.edu), GitHub: [artidoro/sentiment-analysis/blob/master/report.pdf](https://github.com/artidoro/sentiment-analysis/blob/master/report.pdf)
12. Twitter Sentiment Analysis Training Corpus datasets 2022 From: [thinknook.com, twitter-sentiment-analysis-training-corpus-dataset-2012-09-22](https://thinknook.com/twitter-sentiment-analysis-training-corpus-dataset-2012-09-22)
13. Vohra, M.S.; Teraiya, J.: Applications and challenges for sentiment analysis: a survey. *Int. J. Eng. Res. Technol.* 2(2), 1–6 (2013)



14. Whitehead M, Yaeger L. Sentiment mining using ensemble classification models. In: International conference on systems, computing sciences and software engineering (SCSS 08). ("Ensemble of feature sets and classification algorithms for sentiment ...") Berlin: Springer; 2008.
15. Xia R, Zong C, Li S. Ensemble of feature sets and classification algorithms for sentiment classification. *Inf Sci.* 2011; 181:1138–52.