

PROJECT
OTT v/s THEATRES:
Ensembled model of opinion mining on social media.

Team (CSE036):
18951A0552 B. Hiran Christa Bel
18951A05B4 P. Raghavendra
18951A0599 C. Nithish Sai

Supervisor:
Dr. J Sirisha Devi
Associate Professor
CSE

HOD:
Dr . B.J.D Kalyani
Associate Professor
CSE-A,B

Abstract

Opinion Mining has gained interest lately due to the uptick in social media. People began expressing public and private beliefs on varied subjects and sharing their perspectives on online forums.

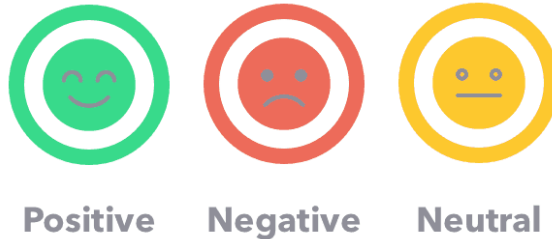
In this Presentation, we concentrate on

1. we present an ensemble model (LSTM-CNN) for opinion mining.
2. This presentation reviews the general public mindset towards the OTT platforms and Movie Theatre with an LSTM-CNN ensemble model of Opinion Mining.
3. Results categorize users' viewpoints through comments into “Ott biased” and “movie theatre biased” illustrated in a pie chart.

Introduction

- Sentiment analysis or Opinion Mining is the text analysis technique that can be used to analyze the text/tweets on social media
- With this natural language processing (NLP) technique we can determine if the data into
 1. positive
 2. negative or
 3. neutral in state.

Sentiment Analysis



Problem Statement & Objective

Problem Statement:

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.

Objective:

- The purpose of this analysis is first, to study the ensemble model
- second to examine the data according to the polarity of, “individuals who favour OTT” and “individuals who favour Movie Theatre” and deliver the outcome utilizing an ensemble LSTM-CNN model of Opinion Mining.

Literature Review

There are many techniques and complex algorithms can be used to perform sentiment analysis. One of the methods is applying sentiment analysis using customized dictionary on social media.

- This method filters the polarity of words from tweets based on their mean and standard deviation in
 - a. positive,
 - b. negative (by removing the neutral terms)
- A customized dictionary is designed to a business which identifies pattern of behavior in social media. It helps a business with their gains in losses.

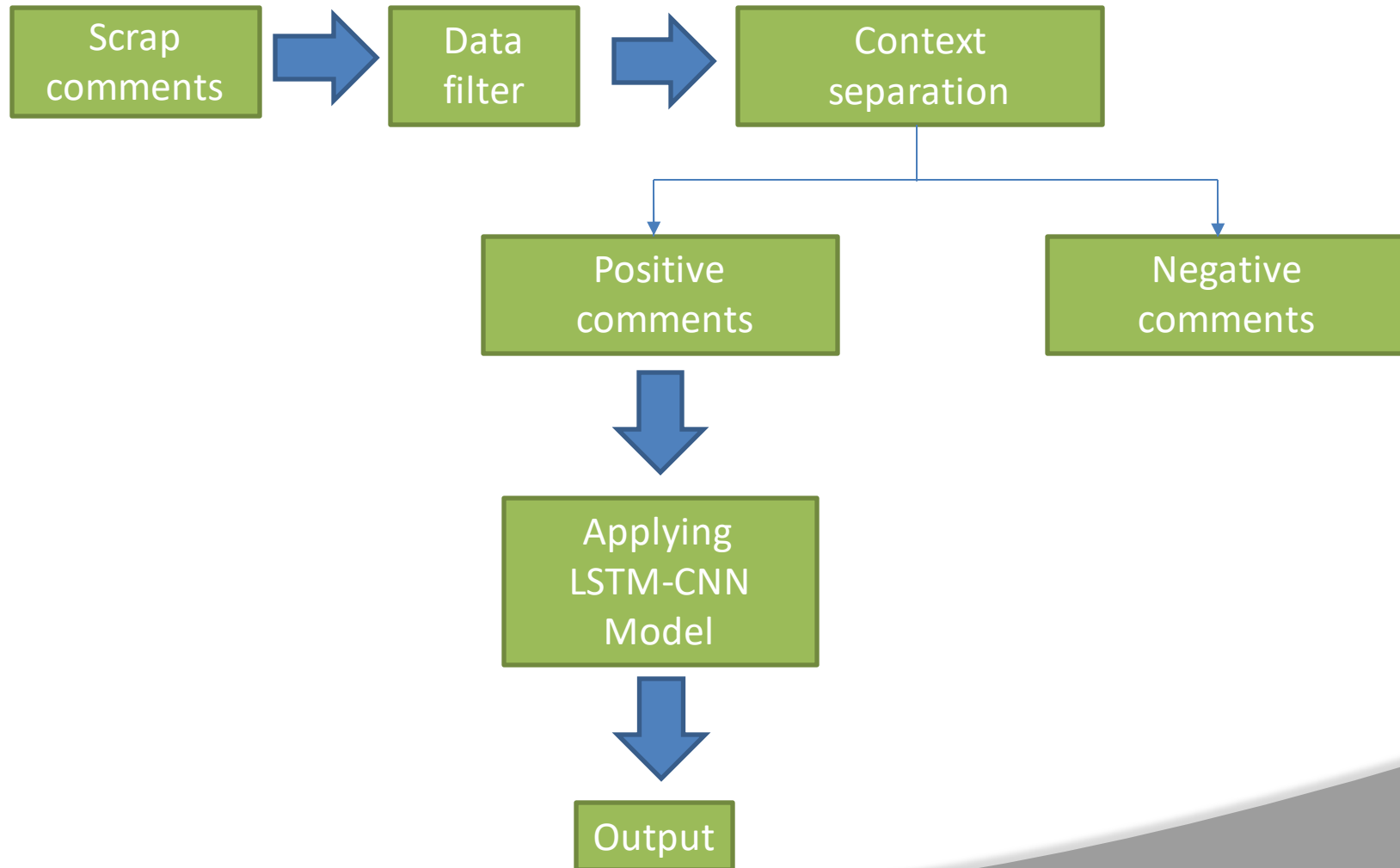
Proposed Solution

To develop an Ensemble model where multiple models are created and then combine them to produce improved results.

Ensemble methods usually produces more accurate solutions than a single model would.

Our presentation 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.

FLOW CHART



1. DATA WRANGLING

Step1: Use google Api to access specific youtube comments.

Step2: Select a specific youtube video to create a list of data(with plenty of comments)

Step3: Use python to scrape comments from the video

Step4:You can filter the comments if you want like "top level comments" or "first comment"

Code Snippet:

```
import pandas as pd
def create_df_author_comments():
    authorname = []
    comments = []
    for i in range(len(response["items"])):
        authorname.append(response["items"][i]["snippet"]["topLevelComment"]["snippet"]["authorDisplayName"])
        comments.append(response["items"][i]["snippet"]["topLevelComment"]["snippet"]["textOriginal"])
    df_1 = pd.DataFrame(comments, index = authorname, columns=["Comments"])
    return df_1
df_1 = create_df_author_comments()
df_1
```



...	...
bot	India should make its own otd platform.
aditya singh	Two different things \nActually if you want to...
Paras	There are many films which can enjoyed on big ...
Atul Verma	With introduction of Ott platform definitely r...
Ruchita singh	Ott cant replace cinema experience but it will...

100 rows × 1 columns

2. PRE-PROCESSING:

The raw data is an unsupervised data with noise.

Batchgen is used to clean the datasets. With the help of regular expressions all the unwanted characters and emojis are removed from the text.

The datasets are divided into ``goodfile" and ``badfile".

3. VOCABULARY:

By using Glove from TensorFlow sentiments of the comments are polarized and vocabulary is built. Then the vocabulary is processed.

4. MODEL SELECTION - LSTM-CNN Ensembled model

We are building an LSTM-CNN Ensembled Model because of their unique features.

Convolutional Neural Networks(CNNs):

Convolutional Neural Networks (CNNs) are networks initially created for image-related tasks that can learn to capture specific features regardless of locality.

In our particular case, it could capture a negative phrase such as "don't like" regardless of where it happens in the tweet.

- I **don't like** watching on OTT platforms.
- That's the one thing I really **don't like**.
- I saw the movie, and I **don't like** how it ended.

Long-Term Short-Term Memory (LSTMs):

LSTMs are the networks that remembers 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 who 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.

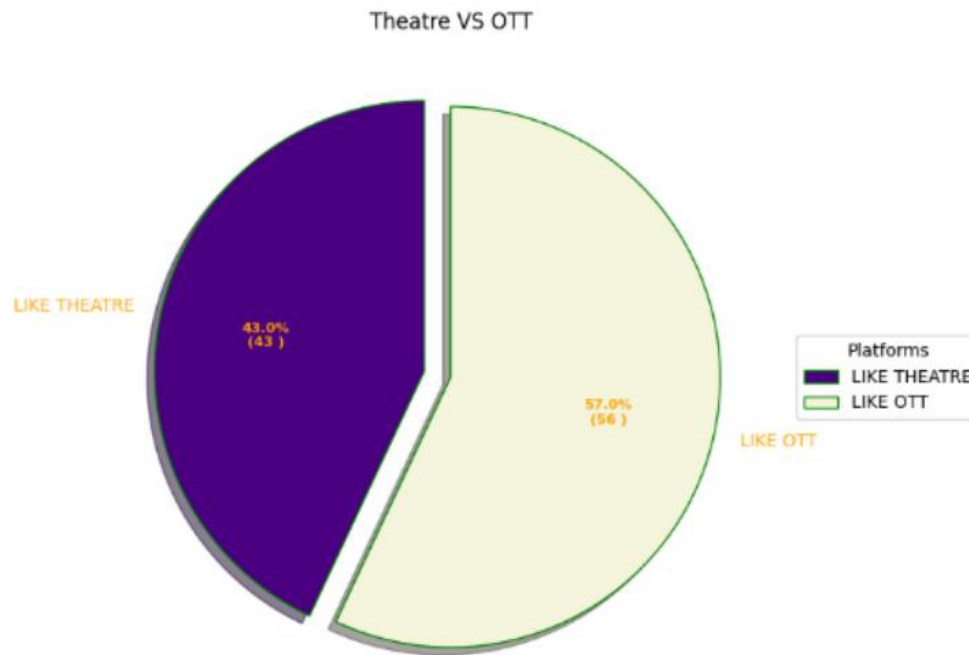
LSTM-CNN Model:

1. Firstly, an embedded layer is created
2. 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”.
3. In CNN networking, a convolution layer and maxpooling layer is applied for each filter.
4. After applying those layers, the pooled features are combined and final predictions are made.

Results

Plotting results on the Chart:

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



Conclusion

By combining CNNs and LSTMs we are able to harness both the CNN's ability in recognizing local patterns, and the LSTM's ability to harness the text's ordering. However, the ordering of the layers in our models will play a crucial role on how well they perform.

As mentioned, LSTM-CNN ensemble model yields effective outcomes. The methodology that we presented has some disadvantages as follows:

- The datasets are collected from a single source. These datasets 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.

THANK YOU

TEAM: 18951A0552 B. Hiran Christa Bel

18951A05B4 P. Raghavendra

18951A0599 C. Nithish Sai