International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Sentiment Analysis for Movie Recommendation

Ms.DEVADHARSHINI C
Department of Artificial Intelligence and Machine Learning
Sri Shakthi Institute of Engineering and Technology
Coimbatore, India

Ms. DHARSHANASHRI G
Department of Artificial Intelligence and Machine Learning
Sri Shakthi Institute of Engineering and Technology
Coimbatore, India

Mr.GOKUL K
Department of Artificial Intelligence and Machine Learning
Sri Shakthi Institute of Engineering and Technology
Coimbatore ,India

Mr. NISHANTH S
Department of Artificial Intelligence and Machine Learning
Sri Shakthi Institute of Engineering and Technology
Coimbatore, India

Mr.THIRUPPUGAL S
Department of Artificial Intelligence and Machine Learning
Sri Shakthi Institute of Engineering and Technology, Coimbatore, India

Ms.NIVEDHA S
Department of Artificial Intelligence and Machine Learning
Sri Shakthi Institute of Engineering and Technology, Coimbatore, India

***Abstract*— Our research proposes a movie recommendation system integrating sentiment analysis with LSTM neural networks. By extracting sentiments from user reviews, LSTM captures nuanced preferences, enhancing recommendation accuracy. Leveraging real-world movie datasets, our approach outperforms existing methods, offering personalized recommendations aligned with user sentiments. Through this fusion of NLP and deep learning, we strive to streamline movie selection, providing users with a more tailored and satisfying experience.**

Keywords—Sentiment Analysis, Natural Language Processing,, Long Short-Term Memory, Recurrent Neural Network.

I. INTRODUCTION

Sentiment analysis, a branch of natural language processing, plays a pivotal role in deciphering and understanding human emotions expressed in textual data. In this documentation,we delve into the realm of sentiment analysis using Long Short-Term Memory (LSTM) networks, a specialized type of recurrent neural network (RNN). Our focus lies on employing LSTM to discern sentiment in the IMDb dataset, a rich collection of movie reviews accompanied by sentiment labels.

Sentiment analysis, also known as opinion mining, enables us to gauge the subjective nature of language, unveiling sentiments such as positivity, negativity, or neutrality within textual content. Applied across various domains, sentiment analysis empowers businesses, policymakers, and researchers to glean valuable insights from massive volumes of textual data, ultimately informing decision-making processes.

II. LITERATURE REVIEW

Microblogging, particularly on platforms like Twitter, provides a significant and expanding source of opinions and discussions on various topics, encompassing diverse users ranging from politicians to celebrities. We collected a large corpus of positive, negative, and neutral tweets using the Twitter4j Java API, processed them by removing stop words, and applied machine learning algorithms for sentiment analysis. SVM with hybrid features demonstrated superior performance, achieving 84% accuracy, while Max Ent with bigram features outperformed Naïve Bayes. Future work includes developing an automatic sentiment classifier for multiple languages, starting with Hindi, to accommodate the growing trend of multilingual tweets on Twitter.

III. EXISTING SYSTEM

One of the existing problems in sentiment analysis on the IMDb movie dataset lies in the complexity and nuance of human language, particularly within the context of movie reviews. Traditional sentiment analysis methods often struggle to capture the subtleties, sarcasm, or colloquial expressions present in user-generated content. This leads to challenges in accurately categorizing sentiments as positive, negative, or neutral. Additionally, the IMDb dataset itself poses challenges such as varying review lengths, noisy data, and potential biases that may arise due to different user demographics or individual preferences. Conventional models may lack the capacity to effectively handle these issues, resulting in suboptimal performance and limited interpretability. Furthermore, the dynamic nature of the film industry introduces a need for sentiment analysis models that can adapt to evolving language trends and cultural shifts. Addressing these existing problems is crucial for advancing the accuracy and applicability of sentiment analysis in the domain of movie reviews and ensuring that insights derived from the analysis align more closely with the nuanced nature of user sentiments.

DRAWBACKS:

- Irrelevant data can be extracted.
- Time complexity can be high.
- Computational problems are occurred.
- Difficult to predict the similar users.

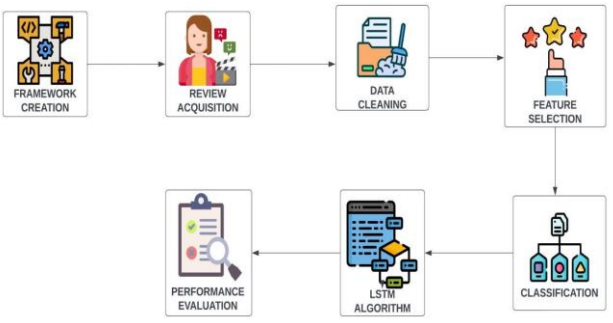
IV. PROPOSED SYSTEM

The proposed system for sentiment analysis on the IMDb movie dataset using LSTM algorithms aims to introduce an advanced and accurate model for understanding user sentiments expressed in movie reviews. The system will leverage the capabilities of LSTM, a type of recurrent neural network renowned for its ability to capture sequential dependencies in data. The architecture will involve preprocessing the IMDb dataset to handle noise, varying review lengths, and potential biases, ensuring a robust and scalable solution. The LSTM model will be trained to discern intricate patterns in the textual data, providing a nuanced analysis of sentiments over the course of each review. The system will not only categorize reviews into positive, negative, or neutral sentiments but also seek to understand the contextual evolution of sentiments, offering a more comprehensive understanding of user feedback. To enhance the system's applicability, it will be designed to generalize well across different movies and genres, making it adaptable to the diverse landscape of the IMDb dataset. The proposed system will contribute insights into the key factors influencing audience sentiments towards movies, facilitating a more granular analysis of the aspects of a film that resonate positively or negatively with viewers. By offering a sophisticated and scalable solution, the proposed system intends to provide valuable tools for filmmakers, studios, and researchers to gain deeper insights into audience opinions, preferences, and the intricate dynamics of sentiment expression in the context of movie reviews.

ADVANTAGES:

- Identify the similar users
- Reduce time and computational complexity
- Relevant data are extracted
- Robust in classification

V. BLOCK DIAGRAM



The framework creation phase for implementing a sentiment analysis movie system using the IMDb dataset involves acquiring, preprocessing, and organizing the relevant data for model training and evaluation. The IMDB dataset containing reviews and corresponding sentiment labels is obtained after the framework created. Data cleaning involves preprocessing to handle missing values, removing duplicates, and standardizing formats to ensure consistency. Feature selection involves identifying and choosing the most relevant and informative features from the dataset that is significant to the

sentiment classification task. The classification involves automatically assigning sentiment labels (e.g., positive, negative) to textual data based on the expressed opinions or emotions in the reviews. The LSTM model is compiled with a suitable loss function and optimizer. The trained LSTM model evaluated using performance metrics (accuracy, precision, recall, F1-score) predict sentiment on the test set of IMDb reviews.

HARDWARE REQUIREMENTS:

- CPU type : Intel Pentium 4
- Clock speed : 3.0 GHz
- Ram size : 4 GB
- Hard disk capacity : 100 GB
- Monitor type : 15 inch colour monitor
- Keyboard type : internet keyboard

SOFTWARE REQUIREMENTS:

- Operating system : Windows OS
- Front End : PYTHON
- Tool : JUPITER NOTEBOOK

VI. RESULT AND DISCUSION

READ THE DATA:

Out[5]:

	review	sentiment
0	One of the other reviewers has mentioned that ...	positive
1	A wonderful little production. The...	positive
2	I thought this was a wonderful way to spend ti...	positive
3	Basically there's a family where a little boy ...	negative
4	Petter Mattei's "Love in the Time of Money" is...	positive

The dataset is obtained and carefully organized to extract relevant features such as positive, negative reviews.

SENTIMENT LABELS:

```
Out[7]: 0        1
        1        1
        2        1
        3        0
        4        1
               ..
        49995    1
        49996    0
        49997    0
        49998    0
        49999    0
        Name: sentiment, Length: 50000, dtype: int64
```

The organized dataset is processed to label positive labelled sentiments as 1 and negative labelled sentiments as 0.

PREPROCSSING:

```
In [12]: len(tokenizer.word_index)

Out[12]: 112281
```

The dataset is preprocessed by divided into train and test dataset. Sentiment reviews are tokenized into separate words as per its index.

TRAIN THE MODELS:

```
X train Shape : (40000,)
y train Shape : (40000,)

X test Shape : (10000,)
y test Shape : (10000,)
```

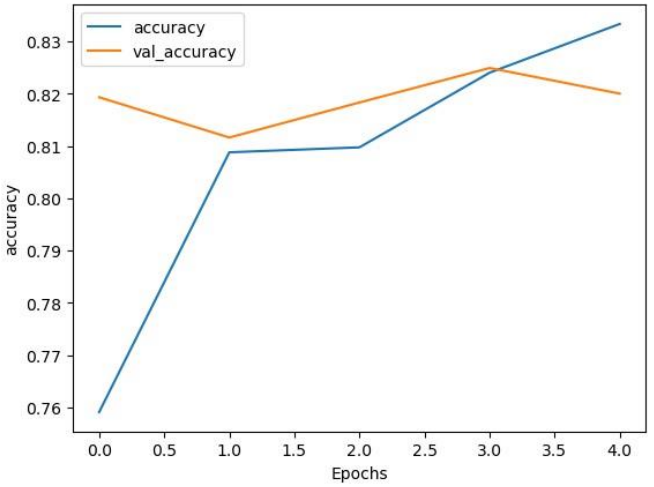
Train and text data are divided to train the model from the train dataset and to test the test dataset to evaluate the model.

LSTM MODEL SUMMARY:

```
In [20]: model.summary()
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 120, 16)           160000

 bidirectional (Bidirectiona (None, 120, 128)          41472
 l)

 bidirectional_1 (Bidirectio (None, 64)                41216
 nal)

 dense (Dense)               (None, 6)                 390

 dense_1 (Dense)             (None, 1)                 7

=================================================================
Total params: 99,085
Trainable params: 99,085
Non-trainable params: 0
_________________________________________________________________
```

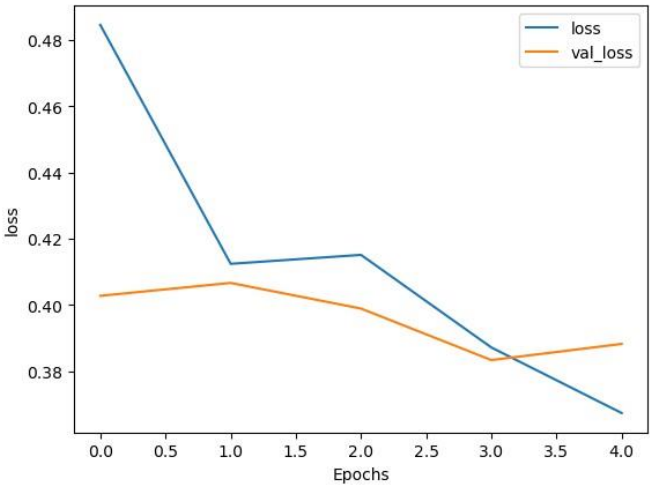
Model architecture created and model sizes are fixed to create the LSTM model.

TRAINING ACCURACY:



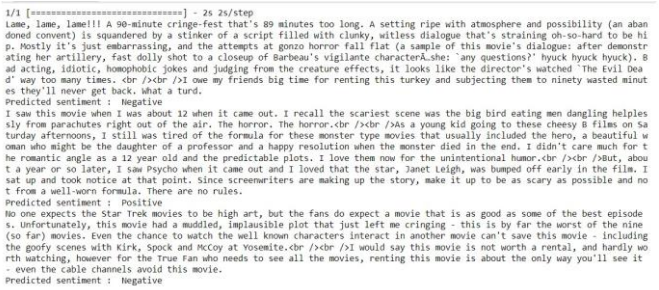
Accuracy obtained during training, the accuracy and value accuracy obtained from each training epoch is depicted in a graph.

TRAINING LOSS:



Loss obtained during training, the loss and value loss obtained from each training epoch is depicted in a graph

OUTPUT:


1/1 [==============================] - 2s 2s/step
Lame, lame, lame!!! A 90-minute cringe-fest that's 89 minutes too long. A setting ripe with atmosphere and possibility (an aban doned convent) is squandered by a stinker of a script filled with clunky, witless dialogue that's straining oh-so-hard to be hi p. Mostly it's just embarrassing, and the attempts at gonzo horror fall flat (a sample of this movie's dialogue: after demonstr ating her artillery, fast dolly shot to a closeup of Barbeau's vigilante characterÂ…she: `any questions?' hyuck hyuck hyuck). B ad acting, idiotic, homophobic jokes and judging from the creature effects, it looks like the director's watched `The Evil Dea d' way too many times.

I owe my friends big time for renting this turkey and subjecting them to ninety wasted minut es they'll never get back. What a turd.
Predicted sentiment : Negative
I saw this movie when I was about 12 when it came out. I recall the scariest scene was the big bird eating men dangling helples sly from parachutes right out of the air. The horror. The horror.

As a young kid going to these cheesy B films on Sa turday afternoons, I still was tired of the formula for these monster type movies that usually included the hero, a beautiful w oman who might be the daughter of a professor and a happy resolution when the monster died in the end. I didn't care much for t he romantic angle as a 12 year old and the predictable plots. I love them now for the unintentional humor.

But, abou t a year or so later, I saw Psycho when it came out and I loved that the star, Janet Leigh, was bumped off early in the film. I sat up and took notice at that point. Since screenwriters are making up the story, make it up to be as scary as possible and no t from a well-worn formula. There are no rules.
Predicted sentiment : Positive
No one expects the Star Trek movies to be high art, but the fans do expect a movie that is as good as some of the best episode s. Unfortunately, this movie had a muddled, implausible plot that just left me cringing - this is by far the worst of the nine (so far) movies. Even the chance to watch the well known characters interact in another movie can't save this movie - including the goofy scenes with Kirk, Spock and McCoy at Yosemite.

I would say this movie is not worth a rental, and hardly wo rth watching, however for the True Fan who needs to see all the movies, renting this movie is about the only way you'll see it - even the cable channels avoid this movie.
Predicted sentiment : Negative

Model is trained and predicts the relevant features into positive, negative reviews.

VII. CONCLUSION

The sentiment analysis program applied to the IMDb movie dataset using LSTM algorithms presents a valuable tool for extracting nuanced insights from user reviews. The ability to discern and categorize sentiments within the complex fabric of movie critiques offers a deeper understanding of audience reactions, providing filmmakers, studios, and audiences with valuable feedback. As a future enhancement, the program could be refined to handle additional layers of complexity, such as sentiment intensity or emotion analysis, to provide even more detailed insights into user opinions.

VIII. REFERENCES

- Joshi, Rohit, and Rajkumar Tekchandani. "Comparative analysis of Twitter data using supervised classifiers." Inventive Computation Technologies (ICICT), International Conference on. Vol. 3. IEEE, 2016.
- Indriani, Fatma, and Dodon T. Nugrahadi. "Comparison of Naive Bayes smoothing methods for Twitter sentiment analysis." Advanced Computer Science and Information Systems (ICACSIS), 2016 International Conference on. IEEE, 2016.
- Hussein, Doaa Mohey El-Din Mohamed. "A survey on sentiment analysis challenges." Journal of King Saud University-Engineering Sciences (2016).
- Neri, Federico, et al. "Sentiment analysis on social media." Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on. IEEE, 2012.
- Vinodhini, G., and R. M. Chandrasekaran. "Sentiment analysis and opinion mining: a survey." International Journal 2.6 (2012): 282-292.
- Mouthami, K., K. Nirmala Devi, and V. Murali Bhaskaran. "Sentiment analysis and classification based on textual reviews." Information communication and embedded systems (ICICES), 2013 international conference on. IEEE, 2013.
- Cho, Sang-Hyun, and Hang-Bong Kang. "Text sentiment classification for SNS-based marketing using domain sentiment dictionary." Consumer Electronics (ICCE), 2012 IEEE International Conference on. IEEE, 2012
- Khan, Aurangzeb, and Baharum Baharudin. "Sentiment classification using sentencelevel semantic orientation of opinion terms from blogs." National Postgraduate Conference (NPC), 2011. IEEE, 2011.