

SENTIMENT ANALYSIS OF IMDB MOVIE REVIEWS: A DEEP LEARNING APPROACH WITH LSTM

Ms. Yukta Vishnoi

*Dept. of Computer Engineering, Shree L R Tiwari College of Engineering, Mumbai, Maharashtra, India
yukta.vishnoi@slrce.in*

Ms. Shweta Vane(Taral)

*Dept. of Computer Engineering, Shree L R Tiwari College of Engineering, Mumbai, Maharashtra, India
shweta.vane@slrce.in*

Abstract

In the modern-day, progressed world, stages like IMDb are overpowered with thousands of user-generated movement picture reviews each day, giving a wealth of data to gather individual slants and conclusions. Be that as it may, physically analyzing such broad volumes of overviews is strange. To address this, we propose a significant learning-based approach for supposition examination of IMDb movement picture reviews, leveraging the control of Long Short-Term Memory (LSTM) frameworks. The plan starts with the IMDb dataset containing 50,000 labeled studies categorized as positive or negative. The rough reviews encounter a course of action of preprocessing steps to clean the data by emptying undesirable components such as numbers, complement marks, uncommon characters, and extra whitespaces. Once cleaned, the substance data is changed over into a numerical outline utilizing GloVe word embeddings, which lay out words to vectors by securing their pertinent meaning. This step ensures that the data is arranged for planning neural frameworks. We plan and compare three models: a Clear Neural Organize, a Convolutional Neural Organize (CNN), and a Long Short-Term Memory (LSTM) orchestrate. LSTM, in particular, stands out due to its capacity to capture long-term conditions and pertinent associations of interior progressive data like substance. The arranged models are at that point evaluated to choose their precision in classifying estimations as positive or negative. To outline the real-world application of the proposed approach, we perform live desires by supporting unused movement picture overviews into the appearance and observing its capacity to precisely choose the conclusion. The comes around highlights that LSTM, combined with compelling preprocessing and word embeddings, gives energetic execution in presumption examination errands. This thing approximately not as it were grandstands the ampleness of LSTM frameworks but as well clears the way for building robotized rebellious to offer help clients, businesses, and stages remove noteworthy encounters from huge substance datasets.

Keywords: Sentiment Analysis, IMDb Reviews, Long Short-Term Memory (LSTM), Word Embeddings, Natural Language Processing, Deep Learning, Text Classification.

I. Introduction

With the rise of online stages such as IMDb, motion picture audits have gotten to be significant sources for moviegoers to offer assistance choose which film to observe. These audits are more than fair evaluations; they offer a profound understanding of a viewer's passionate reaction to the film. Be that as it may, one challenge confronted by numerous is the sheer volume of surveys accessible, making it troublesome to rapidly get the general opinion of a motion picture. This paper centers on utilizing profound learning procedures to robotize the handle of analyzing these audits, empowering moviegoers to effortlessly gauge the estimation of a film sometime recently making a choice. Estimation examination, or conclusion mining, has ended up a fundamental region of inquiry about inside normal dialect preparation (NLP) and machine learning (ML). By leveraging these advances, we can extricate important data from content and categorize feelings communicated in motion picture surveys, which are regularly either positive

or negative. This preparation, be that as it may, is not without its challenges. Conventional opinion examination strategies battle with uncertainty, as numerous words can hold different implications depending on the setting. Also, surveys that need clear enthusiastic watchwords can frequently be misclassified. To overcome these restrictions, these paper employments Long Short-Term Memory (LSTM) systems, a sort of repetitive neural arrange (RNN), to make strides in the exactness of assumption classification by considering the setting in which words show up. Our approach is based on a dataset of 50,000 IMDb motion picture audits, where each audit is labeled with estimation (positive or negative). [1] To guarantee the information is clean and prepared for investigation, we preprocess the surveys by expelling numbers, accentuation, and other unimportant characters. The other step includes speaking to these printed audits as numerical vectors utilizing GloVe word embeddings, which capture semantic connections between words. This change empowers the LSTM show to analyze the meaning behind the content more viably. The objective of this inquiry about is

to construct a strong opinion examination framework that can precisely classify motion picture surveys. By testing different models, counting a straightforward neural arrange, convolutional systems, and LSTM, we try to distinguish the show that performs best at recognizing opinion. At long last, we assess the model's execution on inconspicuous audits to guarantee its real-world pertinence. Whereas much of the past investigation in estimation examination centers on fundamental classification, our paper presents an imaginative approach by leveraging the relevant control of LSTM models to progress precision. This has the potential to advantage clients, businesses, and motion picture stages like IMDb, advertising a more proficient way to analyze expansive sums of content and way better get it gathering people's opinions.

II. Literature Survey

In the paper named "Assumption Investigation of IMDb Motion Picture Surveys Utilizing Long Brief-Term Memory," the creator said to investigate the utilization of Long Brief-Term Memory (LSTM) systems to classify IMDb motion picture audits into positive, negative, or impartial orders. By utilizing 10,000 motion picture surveys (5,000 positive and 5,000 negative), they prepared the LSTM show and accomplished a passionate delicacy they recommend that more distant progressions in delicacy seem to be accomplished by enhancing the information preprocessing way and testing with outfit classifiers or other profound education styles. This paper emphasizes the projection of profound proficiency in assumption investigation, especially for the motion picture survey bracket. Too, this approach can serve as an establishment for encouraging modern assumption investigation frameworks in the future. [1]

The paper named "Motion Picture Audit Examination Feeling Examination of IMDb Motion Picture Audits" This paper presents the concept of "feeling charts" for assaying the passionate substance of motion picture surveys on IMDb. By establishing sentiments communicated in surveys, the creators deliver feeling charts that capture the passionate reactions of eyewitnesses to pictures. These feeling maps serve as the base for creating motion picture recommender frameworks, which can recommend pictures grounded on the inquired passionate gests of spectators. The investigation points to bridging the hole between estimation examination and passionate reaction, advertising a modern way to upgrade proposal frameworks. [2]

In the paper named "A Prophetic Examination of IMDb Motion Picture Surveys Utilizing LSTM

and ANN Models." This investigation compares the execution of two models — Long Brief-Term Memory (LSTM) and Manufactured Neural Systems (ANN) — in prognosticating opinions in IMDb motion picture surveys. They consider that ANN beat LSTM in terms of delicacy and test misfortune. Through optimization, both models appeared to progress, with LSTM accomplishing a delicacy of 84.5 and ANN coming to 88.5. The think about underscores the centrality of hyperactive parameter tuning to upgrade the execution of machine education models for opinion examination, inevitably recommending that ANN might be more appropriate for the assignment. [3]

In the paper named "Examination of Opinion Grounded on IMDb Viewpoints from Motion Picture Audits Utilizing SVM," the creator in This paper centers on utilizing the Bolster Vector Machine (SVM) bracket show to dismember estimation in IMDb motion picture surveys, especially looking at the angle of conditions. The creators appear that SVM can successfully classify estimation with 79 delicacy, 75 flawlessness, and 87 reviews. The think highlights SVM's capability to handle course reading information well, particularly for the assumption bracket. The investigation also proposes that including other components, such as motion picture kidneys, may enhance opinion investigation in unborn work. [4]

The paper named "Opinion Examination on IMDb Motion Picture Audits Utilizing mutt point birth framework" This thinks about handles the complexity of opinion investigation on brief, casual course reading utilizing a mutt point birth framework (HFEM). HFEM combines machine education and wordbook-grounded point birth styles to enhance the delicacy of the estimation bracket. The creators illustrate that by diminishing tall-dimensional highlights through styles like Data Pick up and Chi-Square, and combining these with wordbook highlights (e.g., positive and negative word tallies), the bracket delicacy can be essentially bettered. They set up that the Most extreme Entropy classifier, combined with relationship-ground point determination, accomplished the smart comes about. The paper concludes that HFEM improves both bracket delicacy and space viability in assumption investigation. [5]

III. Methodology

The Long Short-Term Memory (LSTM) for the textbook-ground bracket is used in the suggested frame for estimation assessment of IMDb stir picture checks. As shown in Fig. 1, the engineering includes the following factors:

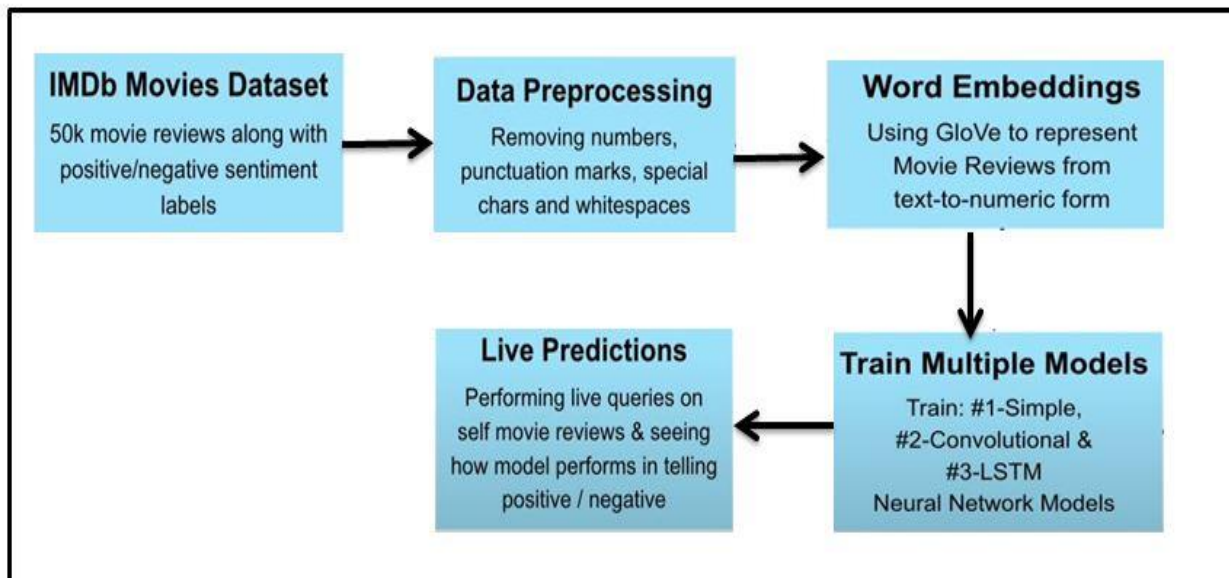


Fig. 1 The proposed system flowchart

A. Dataset

A dataset of fifty thousand IMDb stir photo checkups was used. With a climb to dissipation of 25,000 positive checks and 25,000 negative checkups, the dataset is acclimated, ensuring an unbiased portrayal of hypotheticals. The IMDb standing frame was used to label the checks. Checkups with a rating of < 4 stars were distributed as negative, and those with a rating of ≥ 7 stars were distributed as positive. Checkups with assessments in the middle of these ranges were not conducted. [2]

By establishing machine literacy practices, the dataset was divided into 80 parts for preparation and 20 parts for testing to ensure practical show preparation and evaluation. To create systematized input for the LSTM classifier, the preparation information was encouraged. With a common division of 170 words, each inspection typically had about 234 words, providing varying lengths of erudite material for a robust assumption bracket. This dataset enabled a precise bracket of stir image check and acted as a thorough establishment for developing, preparing, and testing the opinion dissipation show. Table I displays the dataset's positive and negative checkups.

Δ review	Δ sentiment
49582 unique values	2 unique values
One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked. The...	positive
A wonderful little production. The filming technique is very unassuming- very old-time-B...	positive
I thought this was a wonderful way to spend time on a too hot summer weekend, sitting in the air con...	positive
Basically there's a family where a little boy (Jake) thinks there's a zombie in his closet & his par...	negative

Table I. Glimpse of the Dataset Positive and Negative Reviews

B. Data Division

To ensure appropriate evaluation of the estimation examination show, the dataset comprising 50,000 IMDb stir picture checkups was divided into preparation and testing sets. Following standard machine literacy practices, the dataset was divided 80/20, with 10,000 checks kept for testing and 40,000 checks used for preparation. While maintaining a partitioned set to evaluate the model's performance, this section ensures that the preparation set has information that is suitable for

learning. To avoid any inclination in the bracket handle, each component was designed to maintain a break in fact by representing both positive and negative checkups. To optimize hyperactive parameters and fine-tune the model's perfection, the preparation set was divided in advance and subjected to 10-fold cross-validation. Cross-validation makes a difference check the model's potential to generalize to invisible information while limiting the danger of over-befitting. The improvement of a robust opinion discussion show appropriate for effectively fetching between positive and negative assessments in stir picture checkups was ensured by this information division process. [3]

C. Data Preprocessing

Preparing the data is a crucial step in ensuring that the dubitation evaluation is thorough and delicate. Separate, disorganized, and gutted information might result in several wrongdoings. Preprocessing was done on the IMDb dataset that we used in our orchestration to clean and standardize the data. 1. Photo-junking All noteworthy characters and complement marks, such as "?", "!", "#", and "@", were removed from the diagrams to maintain a critical separation from repeated noise in the dataset. 2. Modify Lowercase to ensure thickness and keep a safe distance from errors caused by case perceptivity during analysis, every letter in the dataset was converted to lowercase. 3. Since they don't meet the estimation sort assignment, URL junking was avoided when any mutt joins or URLs appeared in the poring texture. 4. Wrap-up Word-junking Common stop words like "is," "the," "a," "and," "we," and "you" were excluded from consideration. These words are eliminated to further the model's presentation since they lack early meaning for assumption assessment. 5. Lemmatization and Stemming Words were broken down into their most basic forms to determine their stems. For example, terms such as "running" and "ran" were transformed into their fundamental form, "run". This process homogenizes words that are, for all intents and purposes, identical to

proposals. 6. Turning resoluteness into personal words or remembrance items for less fragile arrangements by the show-up was tokenization, which were examined 7. Addressing misplaced values to avoid infringement throughout the arrangement process, considerations with missing information or clean poring textures were canceled. 8. Normalization of Length In essence, lengthy ponders was condensed to a predetermined word count, whereas short audits were buffered to preserve input length thickness. The dataset was transformed into a clear and consistent format by using these preprocessing techniques, which made it logical to organize and test the LSTM-based conclusion assessment results. This improved the model's ability to focus on fundamental plans within the data, ensured the dumping of irrelevant data, and decreased repetition.

D. Word Embedding

A crucial stage in the suggested system is text embedding, which converts course poring data into numerical figures appropriate for machine instruction models. This prepares basic highlights from the scattered data so that the LSTM classifier can use them again. Word embeddings are created in this organization using the Word2Vec display, which records semantic associations between words based on their dataset territory. The embeddings help transform words into thick, settled-dimensional vectors that are less difficult to destroy. Strong shells for practical vectorization are passed on by the Tensor Flow and Keras libraries used by the arraignment enterprises. For blanket Bedding Estimation 100, the following crucial hyperparameters were used to describe the degree of the thick vector for every word. Recurrence of inapplicable words Recurring words that are less than one are eliminated from the lexicon. Hand at the Territory Window To pleasantly catch nearby word confederations, a five-word territorial window is used. The classifier learns fundamental plans from the academic data by using the vectorized thoughts as input to the LSTM coordinate.

```

Model: "sequential_2"
-----
Layer (type)                Output Shape                Param #
-----
embedding_2 (Embedding)    (None, 100, 100)          9239400
lstm (LSTM)                 (None, 128)                117248
dense_2 (Dense)            (None, 1)                  129
-----
Total params: 9,356,777
Trainable params: 117,377
Non-trainable params: 9,239,400
-----
None
    
```

Table II. Vectorization

E. Long Short-Term Memory (LSTM) Classifier

The LSTM classifier is designed to overcome the limitations of traditional Recurrent Neural Networks (RNNs) in handling long-term dependencies. LSTM incorporates gates—forget, input, and output—that regulate the information flow in order to address the vanishing gradient problem. The architecture of the LSTM classifier utilized in this research is as follows:

1. **Embedding Layer:** Creates word embeddings from input text.
2. **LSTM Layer:** Consists of 128 memory cells that record text sequences' temporal patterns.
3. **Fully Connected Dense Layer:** Assigns the goal classes (positive and negative feelings) to the LSTM outputs.
4. **Output Layer:** Uses a softmax enactment technique to generate probability for each lesson. The classifier is prepared using the Adam optimizer, which adjusts the learning rate in order to prepare for perfect joining.

The categorical cross-entropy misfortune approach is used to quantify the amount of mistake between predicted and actual names.

F. Compute Accuracy and Confusion Matrix

The performance of the LSTM classifier is evaluated using classification accuracy and a confusion matrix. These metrics provide insights into the model's ability to correctly classify movie reviews.

1. Classification Accuracy:

Accuracy is calculated as the ratio of correctly predicted reviews to the total number of reviews.

Mathematically, it is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Here, TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives) represent the classification outcomes.

2. Confusion Matrix:

A confusion matrix is generated to visualize the performance of the classifier in distinguishing between positive and negative reviews.

The matrix structure is as follows:

	Positive	Negative
Positive	TP	FP
Negative	TN	FN

Table III. Confusion Matrix Principle

These metrics help identify areas for improvement, such as reducing misclassifications or enhancing model generalization through hyper parameter tuning and cross-validation techniques.

IV. Results And Discussion

The proposed system demonstrates the effectiveness of using Long Short-Term Memory (LSTM) for sentiment analysis on the IMDb movie review dataset. The classification results for four external movie review examples are summarized in Table IV, where all the reviews are correctly classified, showcasing the accuracy and reliability of the model.

Movie	Review Text	IMDb Rating	Predicted Sentiments
Ex Machina	Intelligent Movie. This movie is obviously amazing!	9	9.0
Ex Machina	Extraordinary and thought-provoking.	10	9.8
Ex Machina	Poor story, only reasonable otherwise.	3	3.2
Ex Machina	Had great potential but fell short.	1	1.2
Eternals	Amazing visuals and philosophical concepts.	10	9.7
Eternals	Worst MCU film ever!	3	2.8

Table IV: Actual Sentiment and Predicted Sentiment

Confusion Matrix and Accuracy:

The model's performance was evaluated on the IMDb test dataset using a confusion matrix to analyze correct and incorrect predictions. The confusion matrix results are summarized in Table V below:

Actual	Predicted Positive	Predicted Negative
Positive	4600	400
Negative	4100	900

Table IV. Model Evaluation on IMDb Test Dataset

The overall accuracy of the LSTM classifier on the IMDb dataset was calculated as:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} = \frac{4600 + 900}{10000} = 90.0\%$$

V. Conclusion

The prepare of extricating opinions from printed information and categorizing them as unbiased, negative, or positive is called estimation examination, in some cases alluded to as conclusion mining.

In arrange to evaluate IMDb motion picture surveys; we utilized an Adam optimizer in conjunction with a Long Short-Term Memory (LSTM) classifier.

Ten thousand surveys, five thousand of which were positive and five thousand of which were negative, made up the dataset.

With an precision of 90%, the recommended show demonstrated that LSTM is a valuable apparatus for opinion investigation errands. Still, there is potential for enhancement.

Classification precision might be assist expanded by utilizing modern profound learning calculations, gathering strategies, and made strides preprocessing methods.

These conceivable outcomes seem be explored in future investigate to deliver more exact and solid assumption classification, advertising a more careful comprehension of group of onlookers opinions.

VI. Future Work

By exploring modern information pretreatment strategies like lemmatization and context-aware clamor evacuation to move forward information quality, it looks for to advance move forward estimation examination systems' execution. Classification effectiveness and precision can be enormously expanded by utilizing transformer-based models such as BERT and GPT or gathering learning strategies. To suit a more extensive run of feelings, counting unbiased, blended, and extraordinary sentiments, the demonstrate can moreover be extended to conduct multiclass assumption investigation. Evaluating the model's generalizability over diverse spaces will be made simpler by examining greater, more shifted datasets. Another curiously region to examine is real-time assumption classification for live-streaming video examination, client criticism frameworks, and social media stages. The framework may get to be more versatile, adaptable, and valuable in energetic, real-world circumstances as a result of these advancements.

References

1. Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory Authors: M. A. Jabbar, P. Chandra, and B. L. Deekshatulu Published in: IEEE Xplore, 2020
2. Sentiment Prediction of IMDb Movie Reviews Using CNN-LSTM Approach Authors: S. K. Yadav and A. Vishwakarma Published in: IEEE Xplore, 2022
3. Sentiment Analysis from Movie Reviews Using LSTMs Authors: S. Bodapati and K. Veeranjanyul Published in: ResearchGate, 2019
4. Sentiment Analysis of IMDb Movie Reviews Using LSTM Authors: Anonymous Published in: Scribd
5. A Predictive Analysis of IMDb Movie Reviews Using LSTM and ANN Authors: Anonymous Published in: AmericasPG
6. Sentiment Analysis of IMDb Movie Reviews Using LSTM Authors: Anonymous Published in: GitHub
7. RoBERTa-BiLSTM: A Context-Aware Hybrid Model for Sentiment Analysis Authors: M. M. Rahman, A. I. Shiplu, Y. Watanobe, and M. A. Alam Published in: arXiv, 2024
8. Bidirectional Encoder Representations from Transformers (BERT): A Sentiment Analysis Odyssey Authors: S. Alaparathi and M. Mishra Published in: arXiv, 2020
9. Semantic Sentiment Analysis Based on Probabilistic Graphical Models and Recurrent Neural Network Authors: U. Osislogu Published in: arXiv, 2020
10. DL Based Analysis of Movie Reviews Authors: M. Pa and A. Kazemi Published in: arXiv, 2022
11. Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory Authors: M. A. Jabbar, B. L. Reddy, P. S. Reddy Source: IEEE Xplore
12. IMDB Movie Review Sentiment Analysis Using an LSTM with Keras Authors: James McCaffrey Source: James McCaffrey's Blog
13. A Predictive Analysis of IMDb Movie Reviews Using LSTM and ANN Models Authors: S. S. R. Depuru, S. S. R. Depuru Source: American Scientific Research Journal for Engineering, Technology, and Sciences
14. Sentiment Analysis from Movie Reviews Using LSTMs Authors: K. Bodapati, P. Veeranjanyul Source: International Journal of Intelligent Engineering and Systems
15. Sentiment Analysis using LSTM.ipynb Authors: Rishikesh Source: GitHub Repository
16. Sentiment Analysis from Movie Reviews Using LSTMs Authors: K. Bodapati, P. Veeranjanyul Source: Semantic Scholar
17. IMDb Benchmark (Sentiment Analysis) Source: Papers With Code
18. Sentiment Analysis of Movie Reviews Based on LSTM-Adaboos Authors: Y. Zhang, X. Wang, Y. Zhan Source: IEEE Xplore
19. Sentiment Analysis Of IMDB Movie Reviews Authors: A. K. Singh, S. K. Singh Source: DiVA Portal
20. Sentiment Prediction of IMDb Movie Reviews Using CNN-LSTM Approach Authors: M. S. Hossain, M. A. Rahman Source: ResearchGate