IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

Movie Recommendation System with Sentiment Analysis

Saurabh Gautre[1], Sanskar Tannirwar[2], Pratik Gupta[3], Anup Gain[4], Prof. Devashri Kodgire[5]
Students, Department of Information Technology[1,2,3,4]
Professor, Department of Information Technology[5]
Rajiv Gandhi College of Engineering Research and Technology, Chandrapur, Maharashtra, India

Abstract: *A recommendation system is a system that, depending on certain data, makes suggestions to users for specific resources like books, movies, songs, etc. The characteristics of previously loved movies are typically used by movie recommendation systems to anticipate what movies a user would like. Such recommendation systems are advantageous for businesses that gather data from a lot of clients and want to successfully offer the finest recommendations. When creating a movie recommendation system, several variables may be taken into account, including the movie's genre, cast, and even director. The algorithms are capable of recommending movies based on a single attribute or a combination of two or more. The recommendation algorithm in this study is based on the kinds of genres that the user would want to watch. The method used to do this is content-based filtering with genre relevance. Movie Lens set of data is the one processed by the system. R is the data analysis programmed utilized.*

Keywords: Recommendation System

I. INTRODUCTION

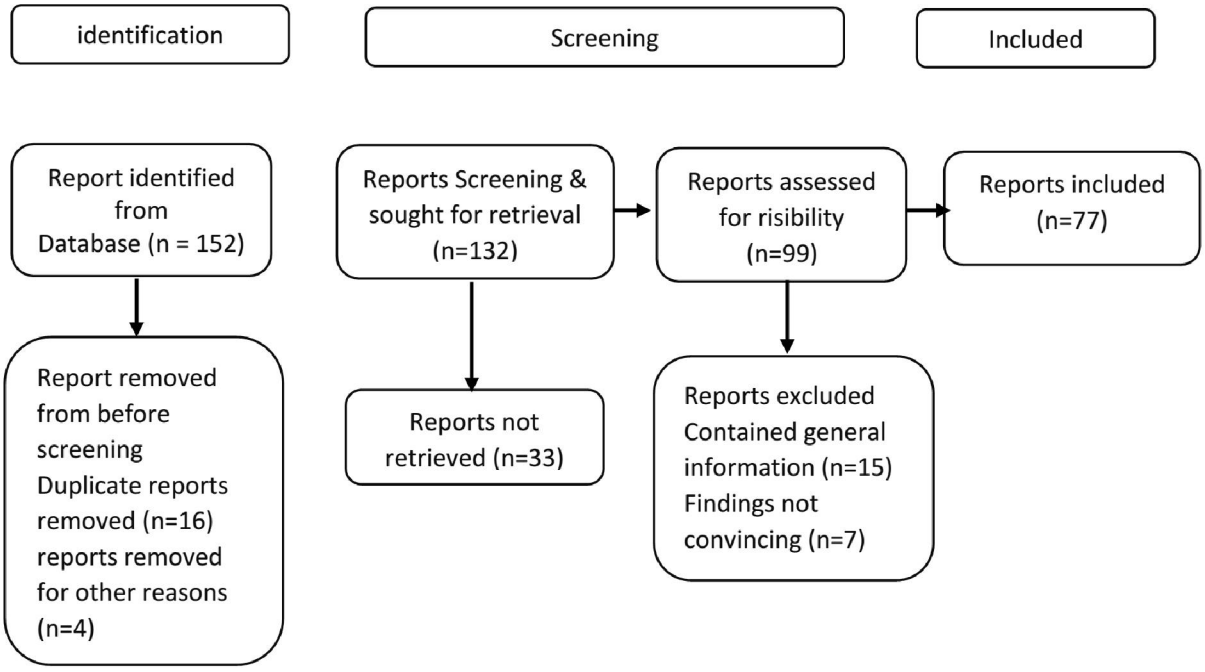
The amount of data transfers that take place every minute have drastically expanded in this era of the Internet. The vast volume of data has grown tremendously along with the number of Internet users. However, not all of the information on the Internet is useful or gives people what they want. Large amounts of data frequently turn out to be inconsistent, and without adequate processing, they become useless. In such circumstances, consumers must rerun their search several times before finding what they first sought after. Researchers have developed recommendation algorithms to address this issue. By considering their prior choices, a recommendation system gives users pertinent information. Data is filtered and uniquely modified to meet each user's needs. Recommendation systems have gained a lot of popularity as more and more data are made available online owing to their efficiency in delivering information quickly. Systems for making recommendations have been created in a variety of fields, including music, movies, news, and items in general. Nowadays, the vast majority of organizations use recommendation systems to meet client demands. Just a few examples are Netflix, Amazon, and LinkedIn. Among the millions of users who have subscribed to the portal, LinkedIn suggests relevant connections of individuals the user may know. The user can avoid conducting in-depth manual searches for persons in this way. The way Amazon's recommendation systems operate is to provide connected product recommendations for users to buy. If a certain client chooses to purchase books from the online store, Amazon offers recommendations for any recent additions to previously favored categories. In a manner quite similar to this, Netflix considers the kinds of shows a customer view and makes recommendations that are comparable to them. Recommendation systems may be roughly categorized into three groups depending on how they operate: content-based, collaborative, and hybrid approaches. A content-based recommendation system takes into account the user's previous activity and looks for patterns to suggest products that are similar to them. The user's past experiences and evaluations are analyzed using collaborative filtering, which compares them to those of other users. Recommendations are based on the ones that are most comparable. Collaborative and content-based filtering each have their own drawbacks. Researchers proposed a hybrid solution to solve this problem by combining the best aspects of both approaches. This study proposes a genre-based recommendation system that is content-based. The dataset utilized for this is a Movie Lens dataset with 9126 films that are categorized by categories. There are 11 different genres in all. 671 people provided their ratings for these movements, which have been compiled. Movies of similar genres are recommended to them based on the high- rated films that the audience gave them.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

II. METHODOLOGY

The procedure for gathering information for the literature review is described in this section. Information regarding movie recommender systems was gathered from sources that have undergone peer review. EBSCO Academic Search Premier, ScienceDirect, IEEE Library, ResearchGate, SpringerLink, and the ACM Portal were among the databases used. Additionally, Google Scholar was utilized to locate references for reviews of certain recommender system components. Search Descriptors: "Movie recommender systems," "movie personalization," "algorithms used in movie recommender systems," "filtering techniques in movie recommender systems," and "machine learning model metrics and measurement criteria" were a few of the keywords used to find information about movie recommender systems.

Inclusion Requirements: Papers containing information regarding recommender systems required to come from published, peer-reviewed sources in order to meet the inclusion criteria. To ensure the accuracy of the data they included for use in this investigation, the publication abstracts were examined. Papers that contained grey literature on recommendation systems were excluded from consideration. Figure 1 provide a summary of the research phases as well as the inclusion criteria for the publications.



2.1 Machine Learning Algorithm for Movie Recommendation System

These are the algorithms that are employed in data mining and information filtering to provide the desired results. Understanding how information filtering techniques operate is crucial to choosing the appropriate algorithm for a given job in recommender systems.

A. K-Means Clustering

This is one of the easiest methods for collaborative filtering that groups individuals according to their interests. Measurement of the Similarities: Finding similarities between the new user's features and those of the prior system users is the first stage. The algorithm always starts with the fundamental classifications so that the user may provide inputs and predictions can be created.

$$sim_{i,j} = \frac{\sum_{m\in(i\cap i)}(r_{i,m}-\overline{r_i})(r_{j,m}-\overline{r_j})}{\sqrt{\sum_{m\in(i\cap i)}(r_{i,m}-\overline{r_i})^2} \times \sqrt{\sum_{m\in(i\cap i)}(r_{j,m}-\overline{r_j})^2}}$$

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

Selection of the Neighbors: When creating the algorithm, many factors are constantly taken into account. The accuracy to be attained and the algorithm's execution time are the important metrics.

Prediction Computation: The nearest neighbors located in the system database are used to calculate the subsequent forecasts. The following formula yields the prediction:

$$prediction_{u,\,i} = \frac{\sum_{n \in Neighbors} (r_{n,i} - \overline{r_n})\; sim_{u,n}}{\sum_{n \in Neighbors} \mid sim_{u,n} \mid} + \overline{r_u}$$

Limitations of K-Means Clustering:

- Cold-Start Issue: When a new user enters the system, a prediction issue arises. There isn't much data available about the user, thus the algorithm can't really anticipate anything until the user starts providing data that can be associated and suggestions can be made based on the user or prior item attributes.
- Dataset sparsity: The recommendation algorithm evaluates a sizable quantity of information from the movie database. Users can't adequately evaluate the features since they can only access a small fraction of the database to search for a few specific items.
- Scalability: Balancing the processing time and the system's accuracy was one of the difficulties mentioned in the choosing of the neighbors. When there are few users or few movies in the database to recommend, the K-means filtering strategy is effective.

Principal Component Analysis K-Means

This content-based movie filtering approach outperforms the K-means clustering approach. Prior to giving clients suggestions, movies are categorized based on their main elements. Using the distance from the mean point, the K-means method determines how near a feature is to the centroid. However, the eigenvectors and eigenvalues are calculated using a covariance matrix created by the main component analysis. As a result, it increases the scalability of finding stronger parallels to propose movies. Assume a K-means algorithm computes the similarity of a single feature at a time to demonstrate this. This suggests that the accuracy and computation time have been compromised.

Principal Component Analysis Procedures

Data formulation is the initial phase in the process, during which the data is organized into tuples of size m + n. The potential tuples might be determined by either the user attributes, the movie feature characteristics, or a mix of both. The structure of the tuples is shown in Table below.

Structure of the tuples.

Tuple	Tuple Structure
A	ri ×ui
B	ri ×ii
C	ri ×ui×ii

the user attributes, the item characteristics, and the movie rating (ri, ui, and ii) (movie characteristics). While C will provide a 3D matrix of size mno, A and B will each provide a 2D matrix of dimension mn.

The covariance matrix is calculated for the dimension of the data that was formulated in the previous phase.

The eigenvalues and eigenvectors are calculated: A square matrix of the data's dimension will make up the estimated covariance matrix. It is employed to calculate the data's eigenvalues and eigenvectors. A future vector is created when the calculated eigenvectors are sorted in decreasing order based on the eigenvalues.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

Principal Component Analysis, Section 4.3 Self-Organizing Road Maps (PCA-SOM)

Self-organizing maps (SOMs) is a neural network-based technology that allows for unsupervised learning without the requirement for human supervision during the training process.

1. Obtain information without any categories or rankings;
2. Modelling data
3. Using unsupervised learning, SOM categorized the data to group together those with comparable qualities;
4. SOM's categorization is replaced with PCA, which examines the dataset's major components and generates additional classifications;
5. The choice to propose the idea

- Initialization: Following the collection of the data, random weights for the initial vectors are selected. The neurons in the data are represented by the vectors' weights, which are calculated as well
- Sampling: With a known probability, a known sample, x, is taken from the input space. The lattice is subjected to this activation sequence. In the new lattice, this pattern transfers the x dimension to be proportionate to the m-pattern.
- Similarity Matching: Using the smallest Euclidean distance between the neuron centroids, the optimal matching is identified at time step-n.

$$i(x) = arg(min||x - w_j|| \text{ where } j = 1, 2, \dots i$$

$$w_j\ (n+1) = w_j\ (n) + \eta(n) h_{ji(x)} (n)\big(x(n) - w_j(n)\big)$$

where hji(x)(n) is the neighborhood function of the i(x) winning neuron and (n) is the learning rate. For the best outcomes, these two works together dynamically.

Use PCA: The PCA procedure is employed further in the data processing to achieve a more accurate assessment after the synaptic weights are determined using the smallest Euclidian distance from the formula above.

Decision: Suggestions are made once similarities are matched.

SOM has excellent qualities that make it a useful tool in recommender systems, including:

Perspectives on the input area: The technique classifies the data using weight vectors and produces an output in the form of a feature map. The chilly beginning is much diminished the user can then enter data based on the aspects of the original output that are displayed.

Topological arrangement: In order for the feature map of SOM to function, a spatial position in the output grid must be mapped from the field of the input pattern.

Density Matching: After the input has been entered, any changes to the input distribution are equally reflected in the output grid, ensuring that the greatest densities with the most matches and the lowest densities with the fewest matches are both accurately represented

Feature Selection: To successfully match the similarities to the grid, the SOM algorithm chooses the optimal features for the non-linear distribution in the input data.

How cosine similarity works

Cosine similarity is a statistic that is used to assess how similar two papers are, regardless of the size of the documents. The cosine of the angle made by two vectors projected onto a multidimensional space is computed. The cosine similarity is useful since it increases the likelihood that the two comparable documents will be orientated closer together, even if they are separated by a large Euclidean distance because of the size of the documents. The cosine similarity increases with decreasing angle.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

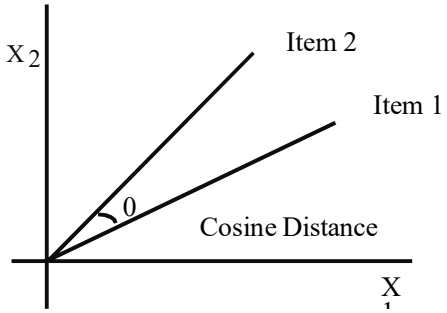


Fig 2: Work of cosine similarity

III. BLOCK DIAGRAM

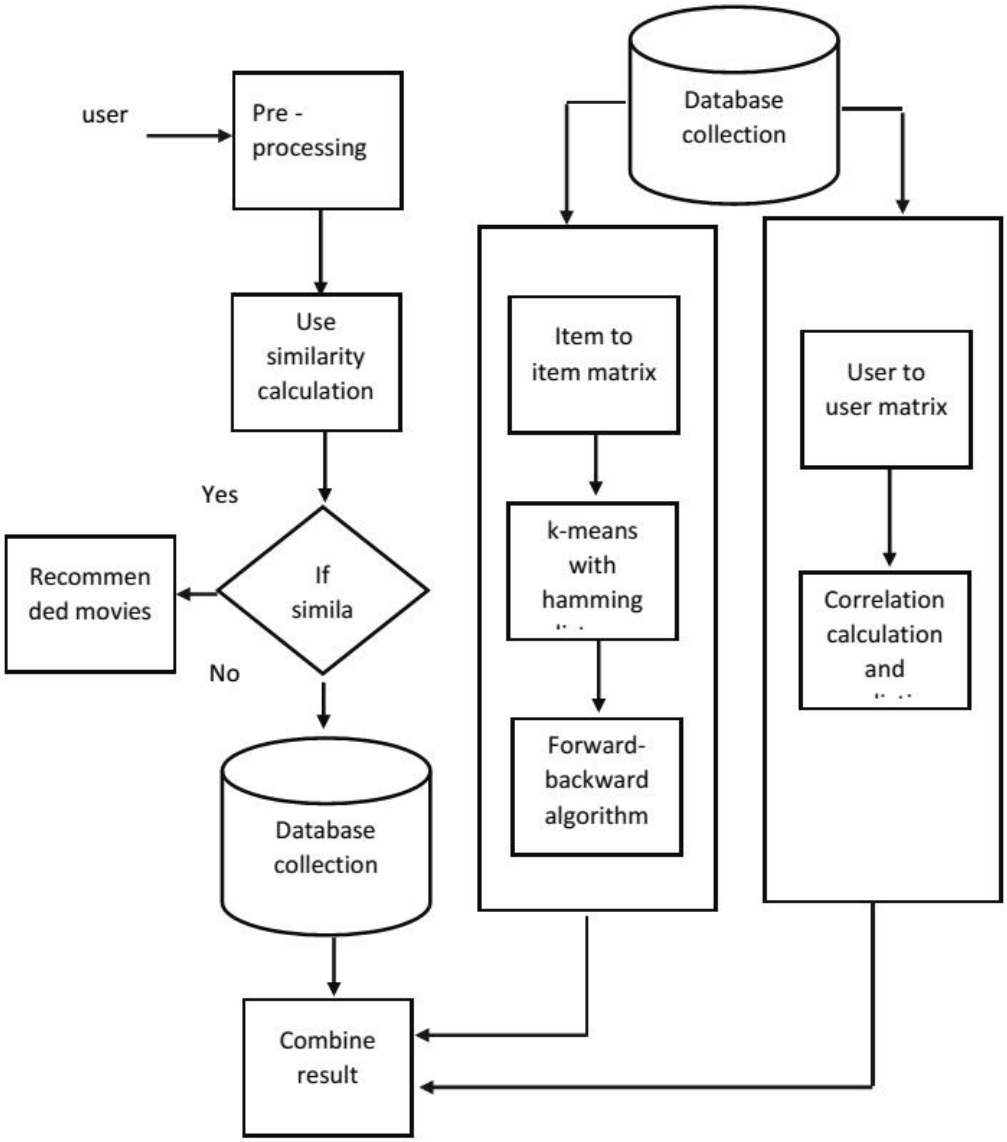


Fig 3: Block Diagram of Movie Recommendation

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

IV. ARCHITECTURE

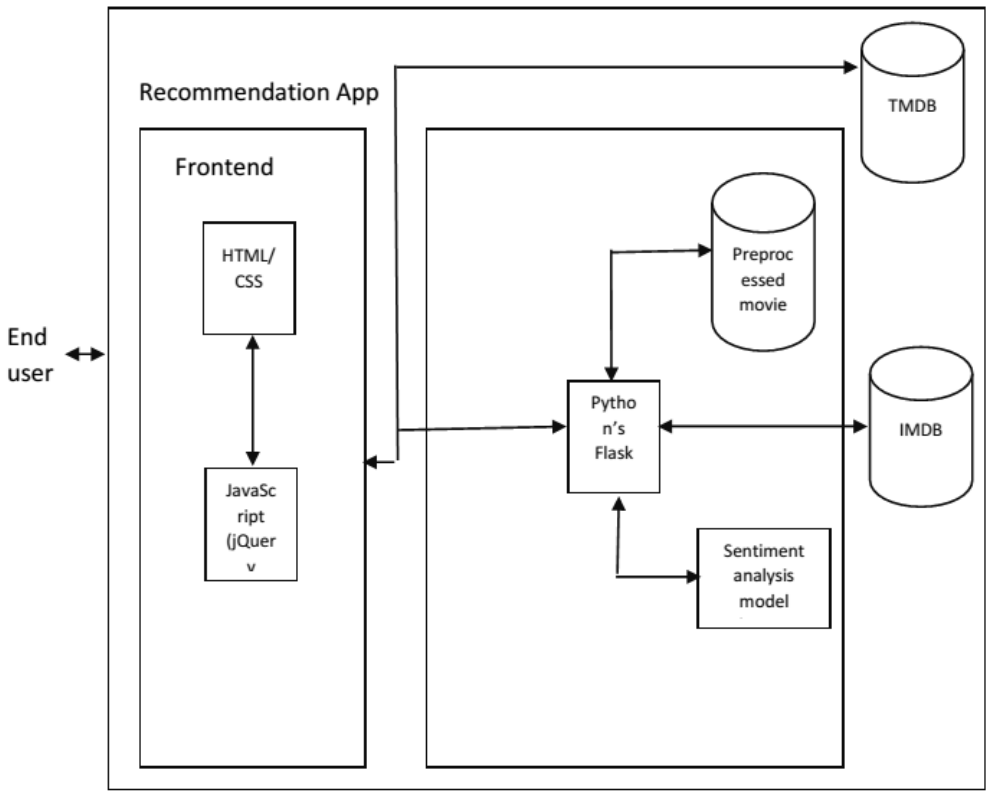


Fig: 4 Architecture of movie recommendation with Sentiment analysis

V. FUTURE SCOPE

When we don't have enough user ratings for a movie or when user ratings for a particular movie are very high or low, the cosine similarity computation does not perform properly. Other approaches of computing similarity, including modified cosine similarity, can be utilized to improve this project.

The user vectors Ux and Uy are normalized before the adjusted cosine similarity, which is comparable to cosine similarity, is calculated by computing the cosine of the angle between them. However, unlike cosine similarity, adjusted cosine similarity substitutes the deviation between each user's raw item rating and their average item rating (denoted Ru) when computing the dot product of the two user vectors.

$$sim(u_x, u_y) = \frac{\sum_{i \in P_{u_x,u_y}} (r_{u_x,i} - \bar{r}_{u_x})(r_{u_y,i} - \bar{r}_{u_y})}{\sqrt{\sum_{i \in P_{u_x,u_y}} (r_{u_x,i} - \bar{r}_{u_x})^2} \sqrt{\sum_{i \in P_{u_x,u_y}} (r_{u_y,i} - \bar{r}_{u_y})^2}}$$

The computation of adjusted cosine similarity is represented by the following equation: where the subset of items I I that both users have rated is represented by Pux,uy, User Ux's rating on item I is represented by Rux, and User Uy's rating is represented by Ruy, The primary benefit of this strategy is that, in item-based collaborative filtering, the item vectors are made up of ratings from many users, many of whom have distinct rating scales.

VI. CONCLUSION

This essay is mainly split into two halves. One of them focuses on the sentiment analysis, while the other one is a movie recommendation system. The paper thoroughly examines both systems and draws some significant findings. The Cosine Similarity algorithm has been utilized for the Movie Recommendation System to suggest the best films that are relevant to the movie the user submitted based on several parameters such as the genre of the movie, overview, the cast, and the ratings provided to the movie. Even after multiple testing, Cosine Similarity has produced respectable findings and has

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

been pretty reliable in terms of suggesting the films.

In this study, sentiment analysis is also crucial. In essence, it seeks to categories the evaluations as favorable or unfavorable. For the same purpose, two algorithms have been utilized. The first is NB, while the second is SVC. Since there is a great deal of variation in reviews, it is crucial to select the optimal method for classification. This is the major motivation behind utilizing two algorithms to categories reviews. Finally, the experimental findings indicate that SVM has a very slight accuracy advantage over N.

Here are a few outcomes of this study that have been mentioned:

1 Improving Sentiment Analysis's accuracy to better classify ironic or sarcastic evaluations.

2 Analysis of the reviews in other languages outside English's sentiment.

3 Personalized movie recommendations based on user preferences (cast, genre, year of release, etc.).

The technology has certain limitations even if it is quite precise. One of them is that the system won't suggest movies if the user-entered movie isn't included in the dataset or if the user doesn't input the name of the movie similarly to how it is in the dataset. The language barrier when performing the emotional analysis is still another drawback. Only reviews that have been written in English so far may be evaluated. If reviews are sardonic or caustic, the Sentimental Analysis incorrectly classifies them as well.

REFERENCES

[1]. Y. Rahhal, A. Jafar, and N. Nassar, Deep multi-criteria collaborative filtering model for recommendations that is new

[2]. Towards cognitive recommender systems: A. Beheshti, S. Yakhchi, S. Mousaeirad, S.M. Ghafari, S.R. Goluguri, and M.A. Edrisi

[3]. V. Rana, S. Sharma, and M. Malhotra, Based on a hybrid filtering technique, an automatic recommendation system 1-16 in Educ. Inf. Technol., 27 (2021),

[4]. S. Chauhan, N. Rajput, Various sentiment analysis methods are examined pp. 75–79 in Int. J. Computer Science and Mobile Computing, 8 (2) (2019).

[5]. A.A. Zulfiqar, M. Azeem, T. Mahmood, C. Xiao, Z. Shaukat, and Sentiment analysis on IMDB using neural networks and a lexicon, SN Appl. Sci., 2 (2) (2020), pp. 1–10.

[6]. K. De, S. Kumar, and P.P. Roy, IEEE Trans. Comput. Soc. Syst., 7 "Movie recommendation system employing sentiment analysis from microblogging data" (4)

[7]. A. Khan, M.E. Hossain, S. Uddin, and M.A. Moni, Contrasting various illness prediction supervised machine learning methods

[8]. B. Adeel, K. Dashtipour, M. Gogate, H. Larijani, and A. Hussain, Deep learning is used to analyse the sentiment of Persian movie reviews.

[9]. Soubraylu, S., and Rajalakshmi, R., Sentiment analysis using a hybrid convolutional bidirectional recurrent neural network for movie reviews