TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 2414 - 2420

Movie Recommendation System Using Machine Learning Algorithms

[1]Anupriya Koneru, [2]D.Siddhabhi, [3]T. Shyam Kumar, [4]K.B Vamsi
[1] Associate Professor, Department of Information Technology, Lakireddy Bali Reddy College of Engineering (Autonomous), Mylavaram, Krishna DT, Andhra Pradesh, India. smartykoneru@gmail.com
[2,3,4] Student, B.Tech, Department of Information Technology, Lakireddy Bali Reddy College of Engineering (Autonomous), Mylavaram, Krishna DT, Andhra Pradesh, India.

Article Info
Volume 83
Page Number: 2414 - 2420
Publication Issue:
May - June 2020

Article History
Article Received*: 11August 2019*
Revised*: 18November 2019*
Accepted*: 23January 2020*
Publication*:10May2020*

Abstract:

Due to extravagant advantages of the big data, the recommendation systems are commonly used in different areas and technologies, including social networking, e-commerce and a vast range of web-based services. The film recommendation feature is very important in our lives because of its ability to provide enhanced entertainment for the user. Like this type of recommendation system, a selection of movies can be recommended to users based on their interest, or movie popularities. In today's world, there is having many more personalized movie recommendation systems that are making use of movie databases which are freely accessible (e.g. Netflix, MovieLens and ErosNow), and enhanced performance and metrics. However, there is a fundamental issue which is still being ignored by recommendation system. Collaborative filtering is one of the main effective strategies for improvising the recommendation system but lacks with time complexity when working on huge data. So hereby in order to overcome the issue used a KNN (K Nearest Neighbor), Decision Tree and Logistic Regression algorithms which are mainly responsible for improvised performance and reduced time complexity of the Movie Recommendation System.

Keywords: *Recommender system, movies, Machine learning, KNN, logistic regression, decision tree.*

I. INTRODUCTION

By the involvement and usage of big data and Cloud Computing [1]-[5] among multiple applications, storing data is not an issue. But huge availability of data confuses the user to take decision. To overcome this, a number of recommendation systems were developed to help people. This paper is regarding recommendation system for movies based on the user's interest and let them make a decision which movies have to watch in huge database. The film industry is a prodigious video producer. It has already been stated by the year 2000 that more than 4500 films are released around the world each year, covering around 9000 hours of footage. With that massive amount of knowledge, there is a lot of need of technology that must be involved to help viewers enable to access movies based on their interest by conveniently and thereby providing movie distribution. Each and every user will have different opinion likes and dislikes. Commonly, a single customer's opinion may depend on a large range of criteria, such as season, mood, or type of work being done by the user. The type of music you would like to hear during the workout, for example, can differ objectively as much as you prepare dinner. So they must find new areas and new paths to understand more about the customer, whilst still understanding almost everything about what is known about the customer info. Many researchers worked in this area to find sentiment analysis [6], [7] on the reviews. But this work is extended by considering the user's

May – June 2020
ISSN: 0193-4120 Page No. 2414 - 2420

profile and information as one of the parameter. There are two main important methods that are mostly used for the recommender systems. Among them one is content based filtering, in which we make an attempt to structure the user's preferences by using the data retrieved, and give suggestions to that profile based on its information. The other one is collaborative filtering, in which we try to cluster users having alike preferences together and hence use data about the group to make recommendations to the customer. But it lacks with time complexity when working on huge data. So hereby in order to overcome the issue, Machine learning algorithms [8]-[10] KNN (K Nearest Neighbor), Logistic regression algorithm and Decision tree are used which are mainly responsible for improvised performance and reduced time complexity.

II. Literature Survey

A. Survey on KNN Algorithm

KNN algorithm [11] is also called as K Nearest Neighbor classification algorithm. It is based upon the technique of supervised learning. Also, KNN is a simple algorithm that stores all existing cases and classifies new existing cases by a similarity measure. KNN had already used this type of non-parametric technique for classification and regression in the statistical estimation approach and pattern recognition method in the early 1970s. But mostly it is used for classification problems. This algorithm is often known as the Lazy Learner Algorithm, since it does not immediately try to learn from the training dataset, but then stores the dataset and classification time of the dataset, and instead performs the operation on that particular dataset. It is shown in figure 1.

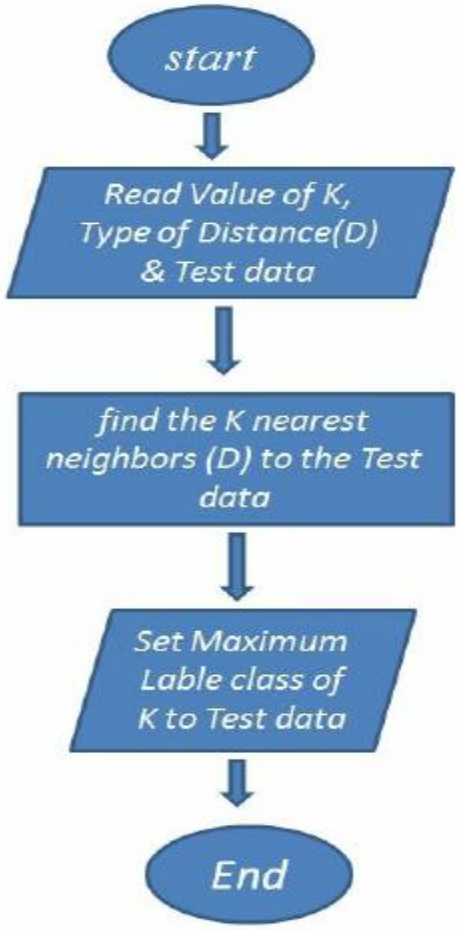

Fig. 1. KNN algorithm

B. Survey on Logistic Regression Algorithm

Logistic regression is a quantifiable model that uses a definite capacity in its fundamental structure to interpret a corresponding variable ward, while a lot of construction is dynamically complex. It is shown in figure 2. In a calculated regression [12] (or logistic regression), regression studies evaluate the parameters of a strategy model (a type of binary regression). A parallel model has science that has a linking variable with two possible characteristics, such as the passing / bomb that describes a pointing variable with the two attributes called '0' and '1'.

C. Survey on Decision Tree Algorithm

Decision tree [13] is also one of the predictive modeling methodology used in various fields such as statistics, data mining and machine learning as well. Decision trees are designed using an algorithmic

May – June 2020
ISSN: 0193-4120 Page No. 2414 - 2420

approach which identifies different ways of splitting a collection of data based on different conditions. It is also one of the most commonly used methods of supervised learning and a practical one. Decision trees are also known as the non parametric based supervised learning method. Example is shown in figure 3. This method is also been used for both the tasks like the classification and regression tasks. This is called the decision tree algorithm.

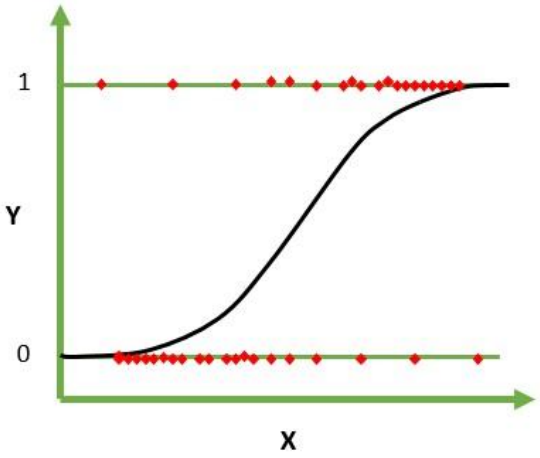

Fig. 2. Logistic Regression

D. Survey on related methods

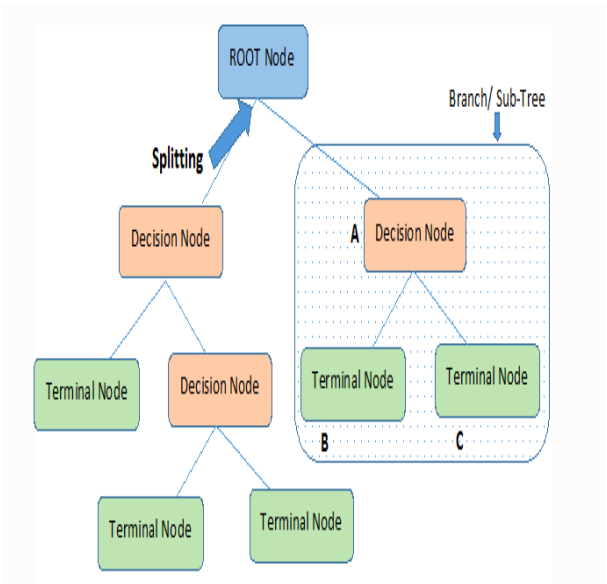

Fig. 3. Decision tree example

MOVREC [14] is a D.K. Yadav et al implemented movie recommendation program focused on a collective filtering approach. It uses user-based data. The data is evaluated and the users who are first agreed with the movie with the uppermost score are recommended for that movie. The program also provides the user with the option to pick the attributes they want to recommend.

Luis M Capos et al. [15] looked at common recommender schemes, like Content-based Filtering and Collaborative Filtering. Because all of these have their own drawbacks, Author has suggested a novel method which incorporates Bayesian Network with Collaborative Filtering. The method being implemented is designed for the specified difficulty and includes distributions of probabilities to draw effective conclusions.

Problem and provides distributions of probabilities for valid inferences. A hybrid system has been implemented by Harpreet Kaur et al. [16]. The program uses both a combination of information and collaborative filtering algorithm. Often, the context of the movies is remembered during recommendation. In the suggestion, the user-user relationship and user-item relationship plays a part.

Balabanovic et al. [17] presented a recommendation system framework which is applied on various applications. It uses diverse distinctiveness, like author, genre, and other words. To draw this sort of data, a TF-IDF is applied.

III. Proposed System

The proposed system uses three algorithms for recommending movies to the user based on the rating, genres and genome_score. Figure 4 shows the work flow of the recommendation system.

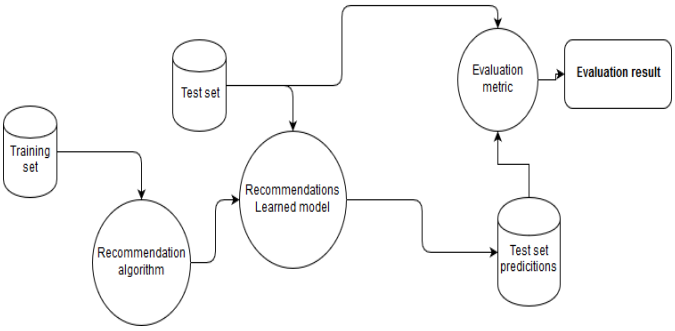

Fig. 4. Work flow for Movie Recommendation System

May – June 2020
ISSN: 0193-4120 Page No. 2414 - 2420

Considered sample datasets such as movies, ratings, genome_score and consolidate into one dataset. The graph is plotted between ratings and titles which is shown in figure 5. Based on the ratings and genome_score we can say that weather the movie is good, average or poor. It is shown in figure 6.

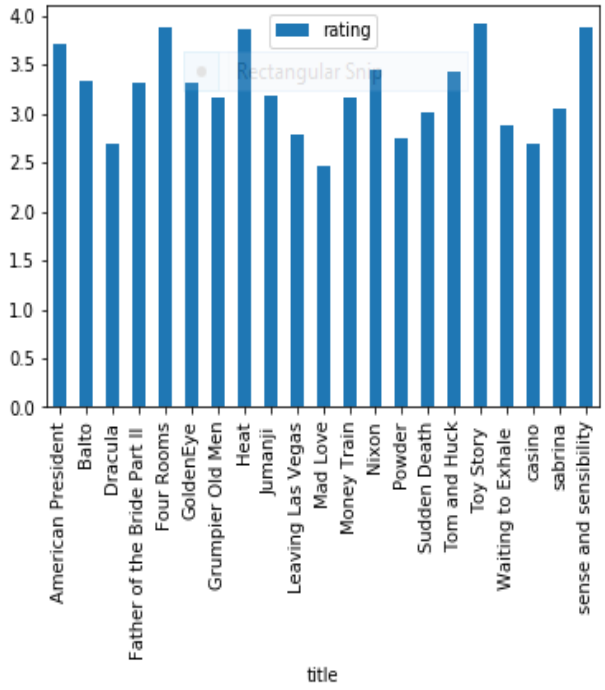

Fig. 5. Rating of the sample data

```
df['outcome'].unique()
```

```
array(['Above average', 'poor', 'average', 'good'], dtype=object)
```

Fig. 6. Categorization of ratings

The algorithm takes the input from the user as a movie name and provides related movie Ids' based on genres, ratings and genome_score. This is shown in figure 7.

```
inp=fun3('Powder')
mcode=fun4('Powder')
inp,mcode
```

```
(0, 13)
```

```
df9=df[df['genres']==inp]
df9['movieId'].unique()
```

```
array([ 6,  7,  9, 10, 19, 20])
```

Fig. 7. Interactive recommendation system

Here Powder is the movie which a user has entered; based on the movie's rating, genres and genome_score we can recommend movies. By using logistic regression, the algorithm takes a movie name as input and the probabilities as output. It is shown in figure 8.

```
tlist=[]
def fun33(value):

    for x in dict11:
        if dict11[x]==value:
            tlist.append(x)

for k in titles:
    fun33(k)
print("Recommended movies...")
print(set(tlist))
```

```
Recommended movies...
{'Powder', 'Heat', 'sabrina', 'GoldenEye', 'Sudden Death', 'Leaving Las Vegas'}
```

Fig. 8. Recommended movies for the given input.

IV. Result

Apply the algorithms on the training data collected to build a recommendation model. To determine the algorithm's accuracy, use this model on each test data instance. Draw a confusion matrix for each of these algorithms and find the accuracy of these algorithms. Compare the accuracy and find the best model for the dataset to provide better results. Here we used algorithms named logistic regression, KNN and decision tree.

May – June 2020
ISSN: 0193-4120 Page No. 2414 - 2420

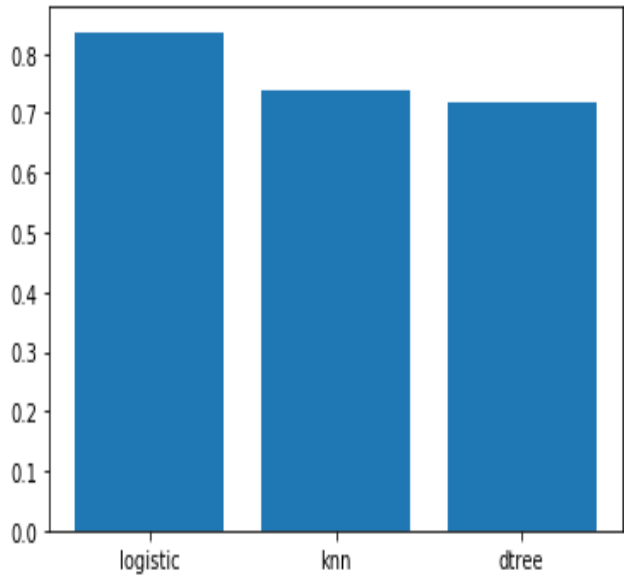

Fig. 9. Accuracy of algorithms when the input is 'Powder'

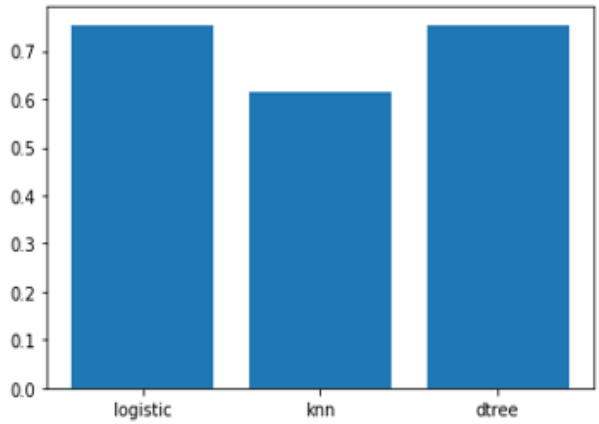

Fig. 10. Accuracy of algorithms when the input is 'Sense and Sensibility'

The accuracy values of the proposed recommendation system for three algorithms are shown in figure 9. It shows that the Logistic regression is showing high accuracy when the movie title is "Powder".

The accuracy values of the proposed recommendation system for three algorithms are shown in figure 10 when the given input is "Sense and Sensibility". It shows that the Logistic regression is showing high accuracy.

The accuracy values of the proposed recommendation system for three algorithms are shown in figure 11 when the given input is "Nixon". It shows that the Logistic regression is showing high accuracy.

The accuracy values of the proposed recommendation system for three algorithms are shown in figure 12 when the given input is "Heat". It shows that the three algorithms are showing same accuracy values.

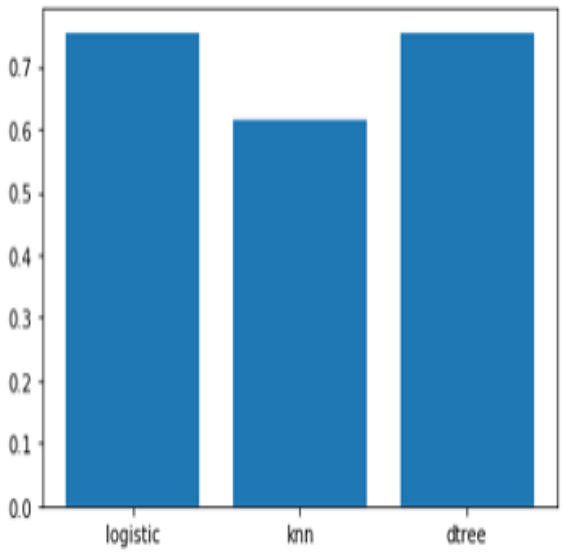

Fig. 11. Accuracy of algorithms when the input is 'Nixon'

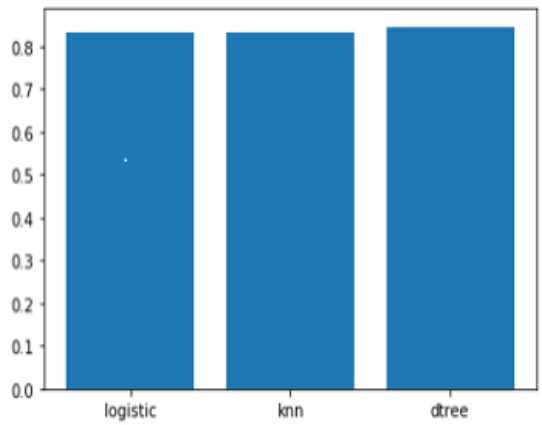

Fig. 12. Accuracy of algorithms when the input is 'Heat'

After observing the accuracy of these algorithms, the Logistic regression is showing high accuracy in the process of suggesting movies to the users. So, Movie Recommendation system using Logistic regression is accurate to suggest the best movies to the end users.

V. Conclusion

In this paper, developed fast and scalable movie recommendations by using the algorithms like

May – June 2020
ISSN: 0193-4120 Page No. 2414 - 2420

logistic regression, K Nearest Neighbor and decision tree. After comparing the accuracy of these algorithms the Logistic Regression has given high accuracy. This work could be extended by considering the Neural Networks in future.

References:

1. A. Koneru and M. S. Latha, "Three tier architecture: To select CSP through CloudServiceBroker in multicloud environment," *2016 International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, 2016,pp.1-6. doi:10.1109/INVENTIVE.2016.7823236 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7823236&isnumber=7823173
2. *AnupriyaKoneru and Sreelatha M, "A Blind Key Signature Mechanism for Cloud Brokerage System," journal of Advanced Research in Dynamic and Control Systems, Issue: 13-Special issue, 2018, pp. 770-776*
3. AnupriyaKoneru and Sreelatha M, "Broker decision verification system using MR cloud tree," International Journal of Engineering & Technology, 7 (4) (2018) 3306-3311.
4. Koneru, Anupriya & Sreelatha, M.. (2017). Cloud service broker: Selection of providers using DTRFV evaluation. Journal of Theoretical and Applied Information Technology. 95. 3551-3559.
5. *A. Koneru and M. Sreelatha, "A comprehensive study on cloud service brokering architecture," 2017 International Conference on Computing Methodologies and Communication (ICCMC), Erode, 2017, pp. 47-53. doi:10.1109/ICCMC.2017.8282517 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8282517&isnumber=8282515*
6. A.Koneru,N.B. Naga Sai Rajani Bhavani, K. Purushottama Rao, G. Sai Prakash, I. Pavan Kumar and V. Venkat Kumar, "Sentiment Analysis on Top Five Cloud Service Providers in the Market," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, 2018, pp. 293-297. doi: 10.1109/ICOEI.2018.8553970
7. Purushottama Rao K., Koneru A., Naga Raju D. (2019) OEFC Algorithm—Sentiment Analysis on Goods and Service Tax System in India. In: Mallick P., Balas V., Bhoi A., Zobaa A. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, vol 768. Springer, Singapore
8. K. Lavanya, L. S. S. Reddy and B. Eswara Reddy ,"A Study of High-Dimensional Data Imputation Using Additive LASSO Regression Model", Computational Intelligence in Data Mining,2019 Advances in Intelligent Systems and Computing 711.
9. Rama Devi Burri, Ram Burri, Ramesh Reddy Bojja, Srinivasarao Buraga,"Insurance claim Analysis using Machine learning Algorithms, "International journal of innovative technology and Exploring Engineering (IJITEE), Volume-8, Issue-6S4, April-2019, ISSN: 22278-3075.
10. SambasivaraoChindam, Rama Devi Burri" Snoezelen Bubble Tube – ATherapy for the Mentally Challenged People". International Journal ofRecent Technology and Engineering (IJRTE) ,Volume-7, Issue-6S5, April 2019,ISSN: 2277-3878.
11. Altman, Naomi S. *(1992)*. "An introduction to kernel and nearest-neighbor nonparametric regression" (PDF). *The American Statistician. 46 (3):175-185.* doi:10.1080/00031305. 1992.10475879. hdl:1813/31637
12. Tolles, Juliana; Meurer, William J (2016). "Logistic Regression Relating Patient Characteristics to Outcomes". JAMA. 316 (5):533–4. doi:10.1001 /jama.2016.7653. ISSN 0098-7484. OCLC 6823603312. PMID 27483067
13. Rokach, Lior; Maimon, O. (2008). Data mining with decision trees: theory and applications. World Scientific Pub Co Inc. ISBN 978-9812771711.
14. Ankur Singh, Manoj Kumar, D.KYadav, Vijay Kr- Based on "A movie recommender system: MOVREC's" International Journal of

May – June 2020
ISSN: 0193-4120 Page No. 2414 - 2420

Computer Applications (0975 – 8887) Volume 124 – No.3, August 2015.

15. Luis M. De Campos, Juan M. Fernández-Luna*, Juan F. Huete, Miguel A. Rueda-Morales; "Combination of content-based and collective recommendations: a hybrid approach based on Bayesian networks", Intemational journal of approximate reasoning, updated 2010.
16. Harpreet Kaur Virk, Er. Maninder Singh based on - " Analysis and Design of Hybrid Online Movie Recommender System ", International Journal of Innovations in Engineering and Technology Volume 5 Issue 2,April 2015.
17. M. Balabanovic and Y. Shoham. Fab: content-based, col- ´ laborative recommendation. Communications of the ACM, 40(3):66–72, 1997.