D.J. Hemanth et al. (Eds.)

doi:10.3233/APC220039

A Novel Approach for Movie Suggestion System with Empirical Risk Minimization and Decision Tree Algorithms

D.Pavan Kumar [a] and A.ShriVindhya[b,1]
[a]*Research Scholar, Department of CSE, Saveetha School of Engineering,*
[b]*Project Guide, Department of CSE, Saveetha School of Engineering,*
[a,b]*Saveetha Institute of Medical and Technical Sciences,*
Saveetha University, Chennai, Tamilnadu, India

Abstract. The major objective of this study is to find accuracy in movie suggestion systems using machine learning algorithms namely Empirical Risk Minimisation Algorithm and Decision Tree Algorithm. Empirical Risk Minimisation algorithm and Decision Tree algorithm with sample size N = 10 was iterated 20 times for finding accuracy in movie suggestion systems.The Empirical riskminimisation algorithm produces a better accuracy of [92.02%] than the Decision tree algorithm [80.58%]. The significance of Empirical risk minimisation[$p<0.05$] for independent sample T-test is high.The comparison of results showed that the Empirical risk minimisation algorithm is significantly better for finding accuracy in movie suggestion systems when compared to the Decision tree algorithm

Keywords. Empirical Risk Minimisation Algorithm, Decision Tree Algorithm, Novel Movie Suggestion,Accuracy, Machine Learning.

1. Introduction

A suggestion system tries to forecast or filter preferences based on the client's preferences. Movies, research papers, news, books, music, search queries, items, and social tags are all examples of suggestion systems [1]. The primary goal of this study is to develop a revolutionary movie suggestion system using machine learningtechniques to suggest relevant items to a user based [2] on user previous information. These resultsdependon the user's search history. This study can be used in many applications like retrieving customer information [3], user choice and service [4], and to suggest a list of movies built on customer's interest [5].

Film suggestion frameworks give a mechanism to help clients in classifying users with comparative interests. This makes recommender frameworks basically a focal piece of sites and internet business applications. The essential goal is to propose a recommender framework through data clustering and computational intelligence [7]. These suggestion frameworks suggest movies depending on user information. It suggests the movies depend on the weight of the movies. Christakou and

[1]A.Shri Vindhya, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha University, Chennai, TamilNadu, India.Email id: shrivindhyaa.sse@saveetha.com

Stafylopatis[8] developed a system that provides precise movie recommendations. Boolean and fuzzy aggregation methods are used to integrate the results of movie filtering. The Movie Lens data was used to test this framework, which resulted in high accuracy computations. The accuracy value obtained on the movie dataset was 68.2% [SVM] and 76.9% [Apriori][9]. Client parallelism for film likes, assessed records of every film, and positive attitudes of customers were studied using various sorts of implicit feedback in movie suggestions [10], and these feedbacks were combined for collaborative filtering. [11]. Ituses a field-aware factorization model to predict film rating and to assist customers in choosing appropriate movies for different purposes. Our wide portfolio in research has translated into publications in numerous interdisciplinary projects. [12]. Now we are focusing on this topic. In the proposed work, an improvisation of a movie suggestion system is analyzed with higher classification accuracy using machine learning algorithms like empirical risk minimization algorithm and decision tree algorithm.

2. Materials and methods

The planned research review environment is finished inSaveetha School of Engineering. A total of two groups were employed in the study. Group 1 is the Empirical Risk Minimization algorithm and group 2 is Decision Tree algorithm. The total number of samples that were assessed on the proposed methodology was 10 in each group to identify various scales in the movie suggestion system [13]. The sample estimation is done using the GPower statistical software. The pre-power test is nearly 81%. Alpha error rate provides the difference between 2 algorithms. Movie suggestion system dataset namely movie metadata is collected from themovie lens website. It contains a total of 10,001 rows of movie data. The data processing includes removal of missing data and replacement of null values with mean or median values with standardization of data. The preprocessed data is given as an input for Empirical Risk Minimization and Decision Tree. 70 percent of the preprocessed data is used for training, while 30 percent is used for testing.

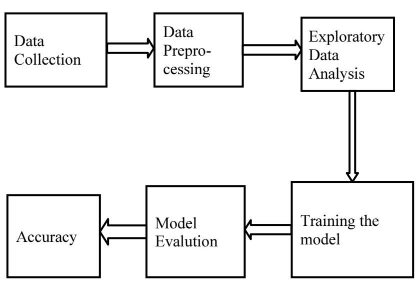

Figure 1. Architecture Diagram for Machine Learning classifiers

2.1 Empirical Risk Minimisation Algorithm

In statistical knowledge theory, empirical risk minimization (ERM) is a criterion that is used to put hypothetical limitations on their presentation. The key premise is that while it is impossible to predict how excellent an algorithm will perform in practice since the real data distribution on which the model will operate is unknown, it may be judged by its presentation based on a given training data set. The performance of the prediction system is analysed by the empirical risk and hypothesis. They are given in the equation [1].

$$Remp(h) = 1/n(L(h(xi), yi)(1)$$

Because the circulation P(x,y) is unknown to the learning algorithm, the risk R(h) cannot be estimated from equation (1); this referred to as uncertain learning situation. However, in the training set, by averaging the loss function, you may get an approximation of the risk, known as empirical risk. According to the empirical risk reduction principle, the learning algorithm must select a hypothesis h ˆ that reduces the empirical risk.

Pseudocode : Empirical Risk Minimisation algorithm
Input : Training set
Output : Classifier trained accuracy
Import required packages and data set
Define x and y variables
x_train,x_test, y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=42)
From sklearn import erm
model= erm.Empirical Risk MinimisationClassifier()
model.fit(x_train,y_train)
predictions_df=rf.predict1(x_test)
score1=accuracy_score1(y_test,y_pred)
Print score

2.2 Decision Tree Algorithm

Decision Tree algorithm is one of the supervised learning methods. Unlike other controlled learning calculations, the choice tree calculation may also be used to deal with relapse and characterization problems. By integrating upfront choice standards obtained from past data, a decision tree was used to construct a training model which was used to anticipate the value or class of an objective variable.We start at the bottom of the tree in Decision Tree algorithm to avoid a class mark for a value. Then compare the values of the root attribute with the attribute of the record. Follow the branch that corresponds to that value and go to the next node based on the comparison.
Pseudocode : Decision Tree Algorithm
Input : Training set
Output **:** Classifier Trained Accuracy
Import required packages and data set
Define x and y variables

```
x_train,x_test, y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=42)
From sklearn import tree
df= tree.DecisionTreeClassifier()
df.fit(x_train,y_train)
predictions_df=rf.predict(x_test)
score=accuracy_score(y_test,y_pred)
Print score.
```

The platform used to evaluate Empirical Risk Minimisation algorithm and Decision Tree algorithm was jupyter notebook. A 64-bit operating system with an x64-based CPU is the system type. The hardware requirements are intel i5 processor with 8GB RAM. The operating system which is used for implementation is Windows and the tools that were used are jupyter notebook with python programming.

The implementation of statistical analysis was done by utilizing IBM SPSS version 2.1 for data. For the Empirical Risk Minimisation method and the Decision tree algorithm, 10 iterations were completed, and the expected accuracy was recorded for each iteration to be analysed in the SPSS tool. The statistics consist of mean, standard deviation and standard deviation error of 2 groups. An independent sample t-test is performed on value acquired from iteration. The independent variables in this research are input attributes for movie_metadata dataset like language genres, user-Id. The dependent variables are output variables - id, title, rating, title year.

3. Results

Owing to the rapid rise of information on the World Wide Web, finding valuable data on the web has become one of the critical and significant issues in research areas. The recommendation system assists users to make decisions in complex information areas where the data accessible is large.

Table 1. has the accuracy of Empirical Risk Minimization algorithm with (92.02%) and Decision Tree algorithm with (80.58%).

Sample	Empirical Risk Minimisation Accuracy(%)	Decision Tree Accuracy(%)
1	74.45	61.20
2	76.66	63.87
3	78.63	64.02
4	79.23	65.78
5	81.71	66.32
6	83.98	70.12
7	85.12	72.54
8	87.47	74.01
9	89.21	78.58
10	92.02	80.58

The accuracy rate of the movie suggestion is predicted with the help of machine learning algorithms. In this study, Empirical Risk Minimization and DecisionTree can enhance movie suggestions, when the ratings of the movie are increased.The results of the independent samples t-test demonstrated that Empirical Risk Minimization Algorithm proved to be more accurate(92.02%) than Decision tree accuracy (80.58%). Empirical Risk Minimization Algorithm got better significance than Decision tree. Table. 2 represents significance values.

Table 2 shows the standard deviation and mean. For Empirical Risk Minimization, they hold the values of 5.74259 and 82.8280 respectively. On the other hand, the values for Decision Tree are 6.57722 and 69.7020. By performing statistical analysis of 10 samples, Empirical Risk Minimization algorithm obtained 5.74259 standard deviation, while for Decision Tree it is 6.57722 (Table. 2). The significant value is (0.04) smaller than 0.05, showing that our hypothesis is good.

	Algorithm	N	Mean	Std.Deviation	Std.Error.Mean
Accuracy	ERM	10	82.8280	5.74259	0.79901
Accuracy	Decision Tree	10	69.7020	6.57722	0.91597

The independent samples T-test, the evaluation of accuracy for novel based movie suggestion systems using Empirical Risk Minimization algorithm and Decision Tree algorithm with standard error difference 2.76111 and the significance error is fewer than $p<0.05$.

Table 1 has the accuracy of Empirical Risk Minimization algorithm with (92.02%) and Decision Tree algorithm with (80.58%).Table 2 has standard deviation and mean for the algorithm Empirical Risk Minimization is 5.74259 and 82.8280. The Decision Tree algorithm is 6.57722 and 69.7020. By performing statistical analysis of 10 samples, Empirical Risk Minimization algorithm obtained 5.74259 standard deviation, while Decision Tree is 6.57722 (Table. 2).

The significant value is (0.04)less than 0.05, showing that our hypothesis is good. Figure. 2 has the bar graph representation in comparison mean of accuracy with Empirical Risk Minimization algorithm and Decision Tree. Empirical Risk Minimization is 82.8280 and Decision Tree is 69.7020. The accuracy of two algorithms was compared using the T-test, and there was a significant difference ($P<0.05$). The Empirical Risk Minimization method had an accuracy of 82.82 percent.

The bar graph illustrates the comparison of mean accuracy of the Empirical Risk Minimization algorithm with Decision Tree. The Empirical Risk Minimization algorithm got more significant results than the Decision Tree.

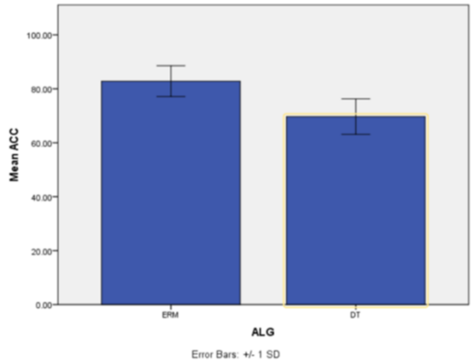

Figure 2. shows that suggested Empirical Risk Minimization method performed better than the Decision Tree technique.

4. Discussion

Empirical Risk Minimization algorithm has better accuracy of 92.02% than Decision Tree which has an accuracy of 80.58%. The standard deviation and mean accuracy for the Empirical Risk Minimization algorithm is 5.74259 and 82.8280 and for the Decision Tree is 6.57722 and 69.7020. This study [1] used Apache Mahout to implement a movie recommender system employing two collaborative filtering techniques. [14] The primary aim of the proposed model is to deliver a recommending system based on data diffusion and fame in social media streams using k-means clustering algorithm and linear regression with accuracy of 71% and 53%. [15] discovers the dissimilar characteristics and potentials of two different forecast techniques which include in recommendation systems utilizing logistic regression, SVM, and ERM, Collaborative Filtering and Content-based Filtering had accuracy of 52 percent, 61 percent, and 74 percent, respectively. [16] movie suggestion system approach based on machine learning algorithms like KNN (75%), ERM (82%), KNN (57%).

5. Conclusion

The proposed framework provides a novel approach to find accurate ratings in movie suggestion systems.It was found the Empirical Risk Minimization algorithm has an accuracy of 92.02% compared to Decision Tree techniques that has the accuracy of 80.58%. The proposed framework proves that the Empirical Risk Minimizationalgorithm is more accurate than the Decision Tree algorithm.

References

[1] Wu C-SM, Garg D, Bhandary U. Movie Recommendation System Using Collaborative Filtering [Internet]. 2018 IEEE 9th International Conference on Software Engineering and Service Science [ICSESS]. 2018. Available from: http://dx.doi.org/10.1109/icsess.2018.8663822

[2] An In-Depth Guide to How Recommender Systems Work [Internet]. [cited 2021 May 19]. Available from: https://builtin.com/data-science/recommender-systems
[3] Aljunid MF, Manjaiah DH. Movie Recommender System Based on Collaborative Filtering Using Apache Spark. In: Data Management, Analytics and Innovation. Springer, Singapore; 2019. p. 283–95.
[4] Website [Internet]. [cited 2021 May 19]. Available from: https://www.academia.edu/download/37143750/06_Paper_31071422_IJCSIS_Camera_Ready_pp._40-45.pdf
[5] Sharma P, Yadav L. Movie Recommendation System Using Item Based Collaborative Filtering [Internet]. Vol. 8, International Journal of Innovative Research in Computer Science & Technology. 2020. Available from: http://dx.doi.org/10.21276/ijircst.2020.8.4.2
[6] Dhanalakshmi R, Sinha BB. An effective movie recommender system enhanced with time series analysis of user rating behavior [Internet]. Vol. 1, International Journal of Mathematics in Operational Research. 2020. p. 1. Available from: http://dx.doi.org/10.1504/ijmor.2020.10032451
[7] Himel MT, Uddin MN, Hossain MA, Jang YM. Weight based movie recommendation system using K-means algorithm [Internet]. 2017 International Conference on Information and Communication Technology Convergence [ICTC]. 2017. Available from: http://dx.doi.org/10.1109/ictc.2017.8190928
[8] Christakou C, Stafylopatis A. A hybrid movie recommender system based on neural networks [Internet]. 5th International Conference on Intelligent Systems Design and Applications [ISDA'05]. 2005. Available from: http://dx.doi.org/10.1109/isda.2005.9
[9] Singh T, Nayyar A, Solanki A. Multilingual Opinion Mining Movie Recommendation System Using RNN. In: Proceedings of First International Conference on Computing, Communications, and Cyber-Security [IC4S 2019]. Springer, Singapore; 2020. p. 589–605.
[10] Zhou M, Hu T, Zhang W, Wang Q, Kong L, Zhou M, et al. COVID-19 pandemic: Impacts on air quality and economy before, during and after lockdown in China in 2020. Environ Technol. 2022 Mar 4;1–50.
[11] Lang F, Liang L, Huang K, Chen T, Zhu S. Movie Recommendation System for Educational Purposes Based on Field-Aware Factorization Machine [Internet]. Mobile Networks and Applications. 2021. Available from: http://dx.doi.org/10.1007/s11036-021-01775-9
[12] Venu H, Subramani L, Raju VD. Emission reduction in a DI diesel engine using exhaust gas recirculation [EGR] of palm biodiesel blended with TiO2 nano additives. Renewable Energy. 2019 Sep;140:245–63.
[13] Priscilla S, Naveena C. Social Balance Theory Based Hybrid Movie Recommendation System. 2020 [cited 2021 May 19]; Available from: http://dx.doi.org/10.1166/jctn.2020.9012
[14] Yazdi KM, Yazdi AM, Khodayi S, Hou J, Zhou W, Saedy S, et al. Improving Recommender Systems Accuracy in Social Networks Using Popularity [Internet]. 2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies [PDCAT]. 2019. Available from: http://dx.doi.org/10.1109/pdcat46702.2019.00062
[15] Jena KC, Mishra S, Sahoo S, Mishra BK. Principles, techniques and evaluation of recommendation systems [Internet]. 2017 International Conference on Inventive Systems and Control [ICISC]. 2017. Available from: http://dx.doi.org/10.1109/icisc.2017.8068649
[16] Chavare SR, Awati CJ, Shirgave SK. Smart Recommender System using Deep Learning [Internet]. 2021 6th International Conference on Inventive Computation Technologies [ICICT]. 2021. Available from: http://dx.doi.org/10.1109/icict50816.2021.9358580