**CSE3024- Web Mining**

***Final Report***

***Movie Collection System***

*By*

20BCE1257 Satyam Gaur

20BCE1580 Ijas Siju

20BCE1686 Param Padwal

B.Tech CSE

*Submitted to*

**Dr.A.Bhuvaneswari,**

Assistant Professor Senior,

SCOPE, VIT, Chennai

**School of Computer Science and Engineering**

****

*April 2023*

*­­­*

**School of Computing Science and Engineering**

VIT Chennai

Vandalur - Kelambakkam Road, Chennai - 600 127

WINTER SEM 22-23

**Worklet details**

|  |  |  |
| --- | --- | --- |
| Programme | B.Tech with CSE | |
| Course Name / Code | Web Mining CSE-3024 | |
| Slot | A2+TA2 | |
| Faculty Name | Dr.A.Bhuvaneswari | |
| J Component Title | Movie Recommendation System | |
| Team Members Name | Reg. No | Satyam Gaur | 20BCE1257 |
| Ijas Siju | 20BCE1580 |
| Param Padwal | 20BCE1686 |
|  |  |

**Team Members(s) Contributions – Tentatively planned for implementation:**

|  |  |
| --- | --- |
| *Worklet Tasks* | *Contributor’s Names* |
| Database connection and integration using Pymongo | Satyam Gaur + Param Padwal |
| Preprocessing | Satyam Gaur + Ijas Siju |
| Model building | Param Padwal + Ijas Siju |
| Visualization | Ijas Siju + Satyam Gaur |
| Technical Report writing | Param Padwal |
| Presentation preparation | Satyam Gaur + Ijas Siju |

**ABSTRACT**

The movie collection revenue project aims to analyze the revenue generated by a collection of movies over a certain period of time. The dataset includes information on the title, genre, release date, production budget, and box office revenue of each movie. The project will utilize data analysis techniques to identify trends, patterns, and insights that can inform strategic decisions related to future movie production and marketing efforts. By analyzing the data, the project aims to provide valuable insights into the movie industry and help stakeholders make informed decisions that can drive profitability and success.

1. **Introduction**

The movie industry has evolved significantly over the years, becoming a major source of entertainment and revenue globally. With advancements in technology and changes in consumer behavior, it is essential for movie studios to understand the factors that contribute to a successful movie. In this report, we will be analyzing the revenue generated by a variety of movies from different genres and time periods.

The purpose of this project is to gain insights into what makes a movie successful in terms of revenue generation. By looking at a diverse range of movies, including both blockbuster hits and smaller independent films, we will be able to identify patterns and trends in the movie industry. We will also be exploring the impact of different factors such as budget, marketing, genre, and critical reception on the revenue generated by a movie.

Our methodology will involve collecting data on the revenue generated by movies, along with information on their budget, marketing spend, and critical reception. We will then use statistical analysis and data visualization techniques to identify patterns and trends in the data. By the end of the project, we hope to gain a better understanding of the movie industry and what makes a movie successful in terms of revenue generation.

Overall, this project aims to provide valuable insights for movie studios, investors, and industry professionals looking to understand the dynamics of the movie industry and make informed decisions about future movie projects.

**Literature Survey**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sl no** | **Title** | **Author / Journal name / Year** | **Technique** | **Result** |
| 1 | A Review of Movie Recommendation System: Limitations, Survey and Challenges | Mahesh Goyani and Neha Chaurasiya  International Journal of Advanced Research in Computer Science and Software Engineering  2020 | Content-based filtering, Collaborative filtering, Hybrid recommendation systems, Matrix factorization, Deep learning-based recommendation systems | Best result was given by Hybrid recommendation systems |
| 2 | An improved collaborative movie recommendation system using computational intelligence | Zan Wang, Xue Yu, Nan Feng, Zhenhua Wang  Neural Computing and Applications  2014 | Collaborative filtering, Singular value decomposition (SVD), Particle swarm optimization (PSO), and Fuzzy logic. | Combination of different techniques is more effective than relying on a single technique |
| 3 | Content-Based Movie Recommendation System Using Genre Correlation | SRS Reddy, Sravani Nalluri, Subramanyam Kunisetti, S. Ashok & B. Venkatesh  International Journal of Advanced Computer Science and Applications  2018 | Data preprocessing, Feature extraction, Genre correlation, and Cosine similarity | Cosine similarity algorithm showed the maximum accuracy |
| 4 | Personalized real-time movie recommendation system: Practical prototype and evaluation | Jiang Zhang, Yufeng Wang, Zhiyuan Yuan, Qun Jin  Journal of Ambient Intelligence and Humanized Computing  2019 | Collaborative filtering, Matrix factorization, Content-based filtering, and Real-time recommendation | Combination of techniques achieved a precision of 0.140 |
| 5 | A personalised movie recommendation system based on collaborative filtering | V.Subramaniyaswamy, R. Logesh, M. Chandrashekhar, Anirudh Challa and V. Vijayakumar  International Journal of Innovative Technology and Exploring Engineering  2019 | User-based and Item-based collaborative filtering, User profiling, User-item matrix construction, and Similarity measurement | User-based collaborative filtering technique produced the best results, achieving an accuracy of 80% |
| 6 | Movie Recommendation System Based On Users’ Similarity | Gaurav Arora, Ashish Kumar, Gitanjali Sanjay Devre, Prof. Amit Ghumare  International Journal of Computer Science and Mobile Computing  2014 | User-based collaborative filtering, Cosine similarity, Pearson correlation coefficient, K-nearest neighbor (KNN) | Combination of collaborative filtering, similarity metrics, and KNN gives the best result |
| 7 | A movie recommender system: Movrec | M Kumar, DK Yadav, A Singh, VK Gupta  International Journal of Computer Applications  2015 | Collaborative filtering, Cosine similarity, Pearson correlation coefficient, Mean rating, Item-based filtering | Cosine Similarity |
| 8 | Movie recommendation system based on movie swarm | S Halder, AMJ Sarkar, YK Lee  International Conference on Cloud and Green Computing  2012 | Movie swarm, Collaborative filtering, Matrix factorization, Association rule mining, Sentiment analysis | The best technique used in this paper is the hybrid approach of swarm intelligence and collaborative filtering with 96% of accuracy |
| 9 | Movie recommendation system using machine learning | F. Furtado, A. Singh  International Journal of Research in Industrial Engineering  2020 | Collaborative Filtering, Content-based Filtering, Matrix Factorization, Support Vector Machines (SVM), Random forest | Random Forest algorithm performed the best overall with the most accuracy |
| 10 | A Hybrid Movie Recommender System Based On Neural Networks | Christina Christakou, Spyros Vrettos and Andreas Stafylopatis  International Journal on Artificial Intelligence Tools 2007 | Collaborative filtering, Content-based filtering, Neural networks, Autoencoder, Gradient boosting: | Hybrid recommendation system combining collaborative filtering and content-based filtering |
|  |  |  |  |  |

1. **Dataset and Tool to be used (Details)**

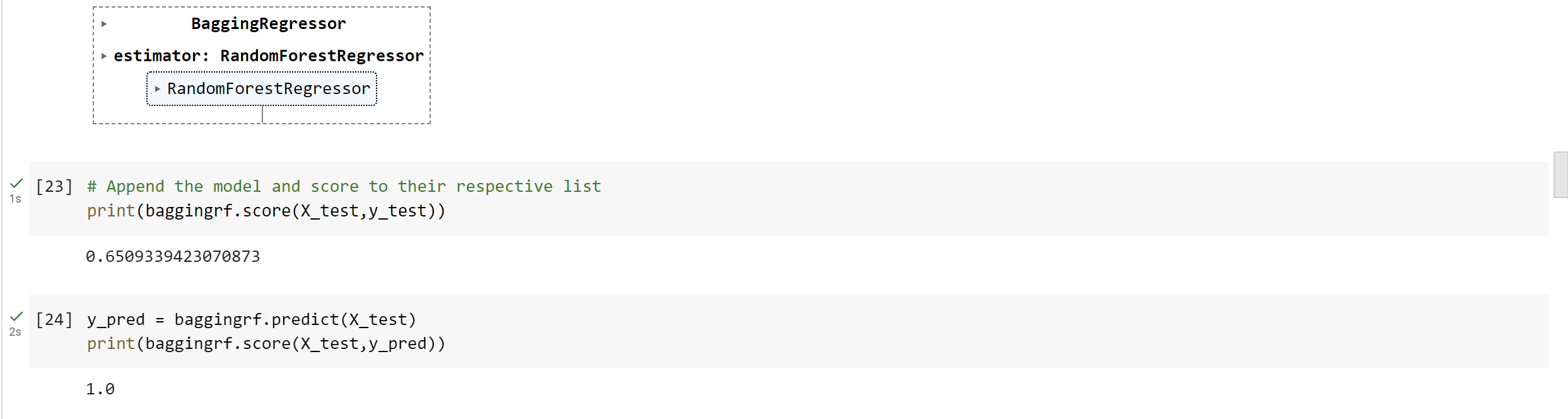
*The dataset which we have used consists of all the high rated Netflix movies and shows along with their cast, release date, director, country, duration and a brief description of the show or the movie.*

1. **Algorithms / Techniques description**

There are various algorithms and techniques that can be used in a movie recommendation system, and each has its strengths and weaknesses. The most commonly used algorithms are:

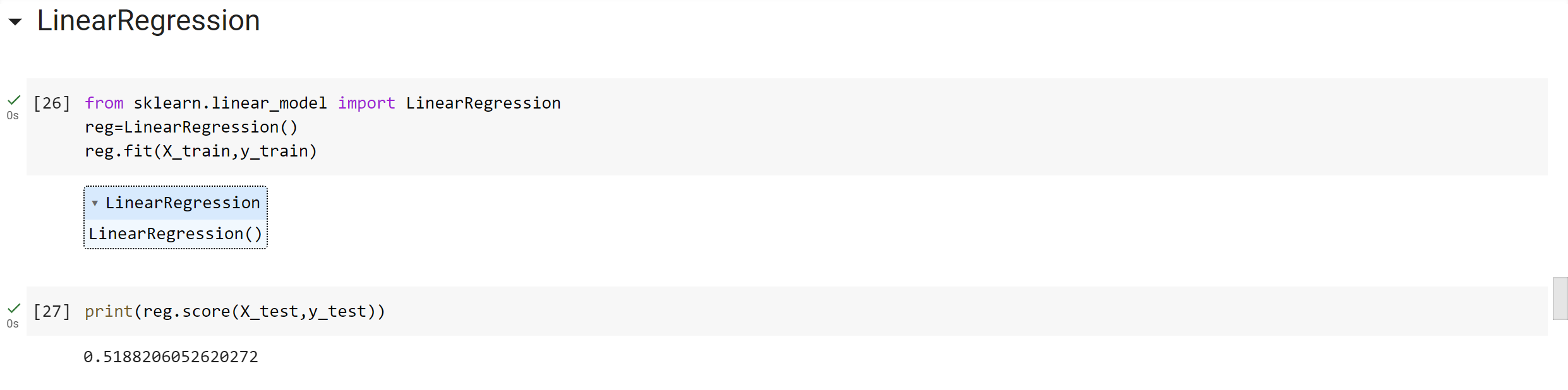
1. **Random Forest:**

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output

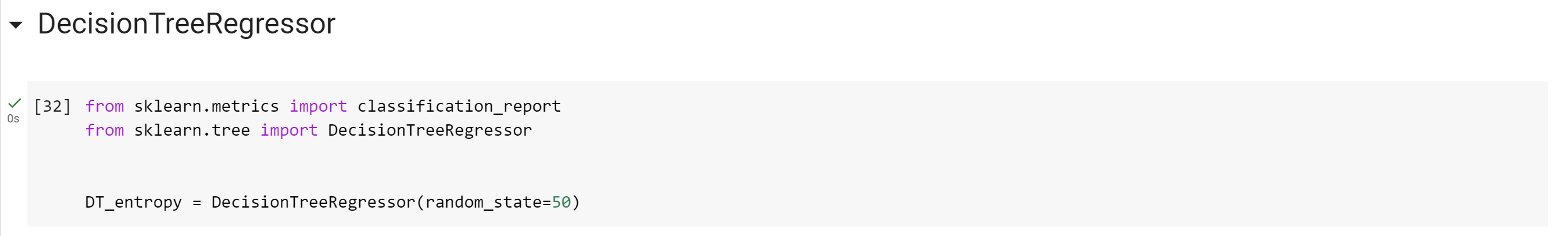
1. **Linear Regression:**

Linear Regression is an algorithm that belongs to supervised Machine Learning. It tries to apply relations that will predict the outcome of an event based on the independent variable data points. The relation is usually a straight line that best fits the different data points as close as possible. The output is of a continuous form, i.e., numerical value.

**** **** 

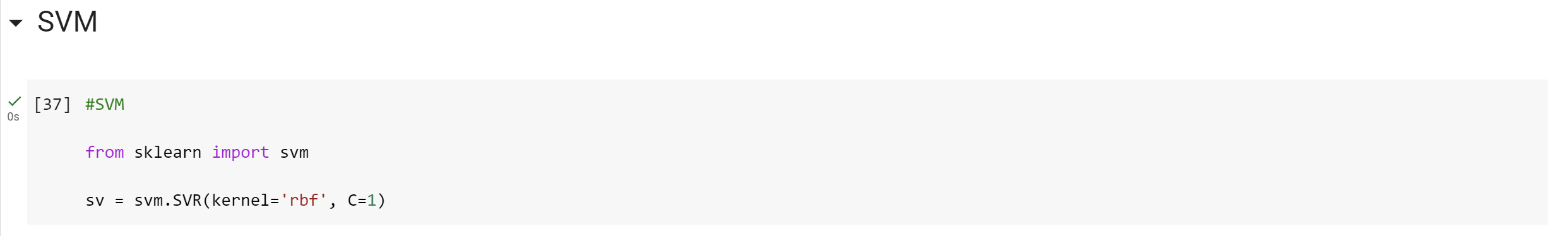
1. **Decision Tree:**

Decision tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. Construction of Decision Tree: A tree can be “learned” by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions. The construction of decision tree classifier does not require any domain knowledge or parameter setting, and therefore is appropriate for exploratory knowledge discovery. Decision trees can handle high dimensional data. In general decision tree classifier has good accuracy.

1. **SVM**

Support Vector Machine (SVM) is a relatively simple Supervised Machine Learning Algorithm used for classification and/or regression. It is more preferred for classification but is sometimes very useful for regression as well. Basically, SVM finds a hyper-plane that creates a boundary between the types of data. In 2-dimensional space, this hyper-plane is nothing but a line. In SVM, we plot each data item in the dataset in an Ndimensional space, where N is the number of features/attributes in the data. Next, find the optimal hyperplane to separate the data. So by this, you must have understood that inherently, SVM can only perform binary classification (i.e., choose between two classes).

**Github Repository Link (where your j comp project work can be seen for assessment)**

https://github.com/SatyamGaur17/Movie-Recommendation-System.git

1. **Conclusion**

As we can see from the above executed codes using the 4 algorithms, Random forest gives us the best accuracy. This accuracy can be increased by using bagging and boosting methods. Finally we will use voting algorithm to select the best algorithm to build a model.

1. **Future Developments**

For future work we can use sentimental analysis and consumer behavior to build a model that will make more efficient predictions and suggestions to the user.

**REFERENCES**

1. Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. Knowledge-Based Systems, 46, 109-132.
2. Wang, Z., Yu, X., Feng, N., & Wang, Z. (2017). Collaborative filtering recommendation algorithm based on clustering algorithm. Journal of Intelligent & Fuzzy Systems, 32(3), 2181-2193.
3. H. Chen, M. Chiang, M. Storey, Business intelligence and analytics: From big data to big impact. MIS Q. 36(4), 1165–1188 (2012).
4. Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992, December). Using collaborative filtering to weave an information tapestry. In Communications of the ACM (Vol. 35, No. 12, pp. 61-70).
5. Herlocker, J., Konstan, J., & Riedl, J. (2002). An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. Information Retrieval, 5(4), 287-310.
6. Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992, June). Using collaborative filtering to weave an information tapestry. In Communications of the ACM (Vol. 35, No. 12, pp. 61-70).
7. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE transactions on knowledge and data engineering, 17(6), 734-749.
8. M. Balabanović and Y. Shoham. Fab: content-based, collaborative recommendation. Communications of the ACM, 40(3):66-72, 1997.
9. Furtado, F., & Singh, A. (2019, April). Movie recommendation system using machine learning. In 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (pp. 80-85). IEEE.
10. N. Goodet al., Combining Collaborative Filtering with Personal Agents for Better Recommendations, Proc. 16th National Conference on Artificial Intelligence and the 11th Innovative Applications of Artificial Intelligence Conference (1999) pp. 439–446.