ISSN: 2581-3404 (Online) **IF: 5.445 (SJIF)**
International Journal of Innovative Research in Technology and Management, Volume-6, Issue-6, 2022.

Movie Reviews Prediction Using Machine Learning Techniques: A Survey

Ashwini Meshram[1], Chetan Agrawal[2], Prachi Tiwari[3]
Dept. of CSE, Radharaman Institute of Technology & Science, Bhopal, India[1, 2, 3]
ashu12.meshram@gmail.com[1], chetan.agrawal12@gmail.com[2], prachi.38@gmail.com[3]

Abstract: *In recent years, the predictive model has seen an increase in volume thanks to the application of machine learning. The film business is still quite significant, as seen by the hundreds of new films that are produced each year. The likelihood of a film's commercial success can be influenced by a wide range of elements, including film critics, actors, directors, actresses, and composers, amongst others. Several writers have proposed or implemented several approaches and algorithms for forecasting the success of movies, including KNN, SVM, Naive Bayes Classifier, Logistic regression, random forest, and others. These approaches and techniques include: In this study, we give a literature review on the topic of predicting the Reviews of a movie by employing a variety of machine learning algorithms. We also make some recommendations for further research.*

Keywords: Sentiment Analysis, Movie Review, Machine Learning, Deep Learning, SVM KNN.

Introduction

Machine Learning is a perception that consents the machine to acquire from examples and experience, and that moreover deprived of being overtly programmed. Thus, instead of you scripting the code, what you do is you feed data to the generic technique, and the technique/ machine builds the logic based on the given data. It permits the computer system or the machines to construct data-driven decisions rather than being explicitly programmed for carrying out a certain task. These programs or techniques are designed in a way that they learn and improve over time when are exposed to new data. Machine Learning Technique is trained using a training data set to create a model. When new input data is introduced to the ML technique, it predicts based on the model. The prediction is evaluated for accuracy and if the accuracy is acceptable, the Machine Learning technique is deployed. If the accuracy is not acceptable, the Machine Learning technique is trained again and again with an augmented training data set [1].

ISSN: 2581-3404 (Online) IF: 5.445 (SJIF)
International Journal of Innovative Research in Technology and Management, Volume-6, Issue-6, 2022.

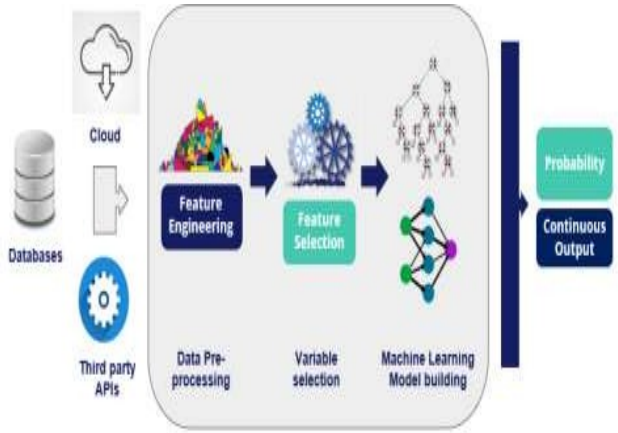

Fig.1: Working steps of Machine learning technique.

This is just a very high-level example as there are many factors and other steps involved shown in fig.1. While designing a machine (a software system), the programmer always has a specific purpose in mind. For instance, consider J. K. Rowling‟s Harry Potter Series and Robert Galbraith‟s Cormoran Strike Series. To confirm the claim that it was indeed Rowling who had written those books under the name Galbraith, two experts were engaged by The London Sunday Times, and using Forensic Machine Learning they were able to prove that the claim was true. They develop a machine learning algorithm and "trained" it with Rowling‟s as well as other writers writing examples to seek and learn the underlying patterns and then "test" the books by Galbraith. The algorithm concluded that Rowling‟s and Galbraith‟s writing matched the most in several aspects. So instead of designing an algorithm to address the problem directly, using Machine Learning, a researcher seek an approach through which the machine, i.e., the algorithm will come up with its solution based on the example or training data set provided to it initially [2].

1.1 Classification of Machine Learning

There are three important types of Machine Learning Techniques such as supervised learning, unsupervised learning, and reinforcement learning which we are discussing in detail:

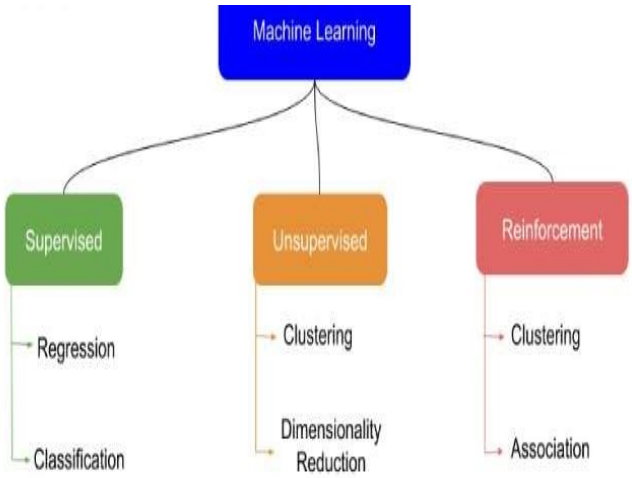

Fig.2: Classification of Machine Learning Techniques.

A. Supervised Learning

Supervised Learning is the most popular paradigm for performing machine learning operations. It is widely used for data where there is a precise mapping between input-output data. The dataset, in this case, is labeled, meaning that the algorithm identifies the features explicitly and carries out predictions or classification accordingly [2]. As the training period progresses, the algorithm can identify the relationships between the two variables such that we can predict a new outcome. The resulting Supervised learning algorithms are task-oriented. As we provide it with more and more examples, it can learn more properly so that it can undertake the task and yield the output more accurately. Some of the algorithms that come under supervised learning are as follows: Linear regression, random forest, support vector machine, artificial intelligence [3], etc.

There are two main types of supervised learning problems: they are classification which involves predicting a class label and regression which involves predicting a numerical value [3].

- Classification: Supervised learning problemthat involves predicting a class label.
- Regression: Supervised learning problemthat involves predicting a numerical label.

Both classification and regression problems may have one and more input variables and input variables may be any data type, such as numerical or categorical.

B. Unsupervised Learning

Unsupervised machine learning holds the advantage of being able to work with unlabeled data. This means that human labor is not required to make the dataset machine-readable, allowing much larger datasets to be worked on by the program. The model learns through observation and finds structures in the data. Once the model is given a dataset, it automatically finds patterns and relationships in the dataset by creating clusters in it [4].

In supervised learning, the labels allow the algorithm to find the exact nature of the relationship between any two data points. However, unsupervised learning does not have labels to work off of, resulting in the creation of hidden structures. Relationships between data points are perceived by the algorithm abstractly, with no input required from human beings. The creation of these hidden structures is what makes unsupervised learning algorithms versatile. Instead of a defined and set problem statement, unsupervised learning algorithms can adapt to the data by dynamically changing hidden structures.[4] This offers more post-deployment development than supervised learning algorithms. What it cannot do is add labels to the cluster, like it cannot say this is a group of apples or mangoes, but it will separate all the apples from mangoes. Suppose we presented images of apples, bananas, and mangoes to the model, so what it does, based on some patterns and relationships creates clusters and divides the dataset into those clusters. Now if new data is fed to the model, it adds it to one of the created clusters. Example of unsupervised learning is k-mean clustering, principle component analysis, SVD, FP-growth, etc.

There are many types of unsupervised learning, although there are two main problems that are often encountered by a practitioner: they are clustering which involves finding groups in the data, and density estimation which involves summarizing the distribution of data [4].

- Clustering: Unsupervised learning problem that involves finding groups in data.
- Density Estimation: Unsupervised learning problem that involves summarizing the distribution of data.

C. Reinforcement Learning

Reinforcement learning directly takes inspiration from how human beings learn from data in their lives. It features an algorithm that improves upon itself and learns from new situations using a trial-and-error method. Favorable outputs are encouraged or „reinforced", and non-favorable outputs are discouraged or „punished". Based on the psychological concept of conditioning, reinforcement learning works by putting the algorithm in a work environment with an interpreter and a reward system.
In every iteration of the algorithm, the output result is given to the interpreter, which decides whether the outcome is favorable or not [5].
In the case of the program finding the correct solution, the interpreter reinforces the solution by providing a reward to the algorithm. If the outcome is not favorable, the algorithm is forced to reiterate until it finds a better result. In most cases, the reward system is directly tied to the effectiveness of the result [5].
In typical reinforcement learning use cases, such as finding the shortest route between two points on a map, the solution is not an absolute value. Instead, it takes on a score of effectiveness, expressed in a percentage value. The higher this percentage value is, the more reward is given to the algorithm. Thus, the program is trained to give the best possible solution for the best possible reward [5]. This simple feedback reward is known as a reinforcement signal.

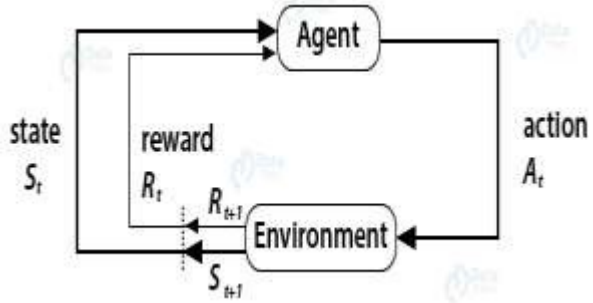

Fig. 3: Example of reinforcement learning.

The agent in the environment is required to take actions that are based on the current state. This type of learning is different from Supervised Learning in the sense that the training data in the former has output mapping provided such that the model is capable of learning the correct answer. Whereas, in the case of reinforcement learning, there is no answer key provided to the agent when they have to perform a particular task. When there is no training dataset, it learns from its own experience.

D. Semi-Supervised Learning

In this type of learning, the given data are a mixture of classified and unclassified data. This combination of labeled and unlabeled data is used to generate an appropriate model for the classification of data. In most situations, labeled data is scarce and unlabeled data is in abundance (as discussed previously in the unsupervised learning description) [6]. The target of semi-supervised classification is to learn a model that will predict classes of future test data better than that from the model generated by using the labeled data alone.
The way we learn is similar to the process of semi-supervised learning. A child is supplied with:

- Unlabeled data provided by the environment. The surroundings of a child are full of unlabeled data in the beginning.
- Labeled data from the supervisor. For example, a father teaches his children about the names (labels) of objects by pointing toward them and uttering their names.

II. Related Work

Though various researchers have contributed to ML and numerous algorithms and techniques have been introduced as mentioned earlier, if it is closely studied most of the practical ML approach includes three main supervised algorithms or their variants. These three are namely, Naive Bayes, Support Vector Machine, and Decision Tree. The majority of researchers have utilized the concept of these three, be it directly or with a boosting algorithm to enhance the efficiency further. The literature related to movie success prediction is discussed below:

A. Prediction using SVM Classifiers

Hemant Kumar and Santosh Kumar [11] proposed a narrative approach to constructing and using a semantic relation between actor, director, genre, budget, ratings, producer, and other certain essential attributes, thus alleviating the hypothesis into forecasting. Distinctively we will comply with the formulation and composition of linear regression, Support Vector Machine, and decision tree as an augmented amalgamated algorithm to assemble features for classification and predictions. Subsequently, the proposed investigational consequences will illustrate the precise and efficient results and prediction thereof. The proposed framework predicts the achievement of a motion picture in light of its gainfulness by utilizing chronicled information from different sources. Utilizing informal community examination and content mining methods, the framework naturally separates a few gatherings of highlights, including "who" is on the best composition (actor and director) what a film is about, "when" a motion picture will be released, and in addition "semi variety" highlights that match "who" with "what", and "when" with "what". The examination comes about with motion pictures amid years‟ time frame demonstrated that the framework beats benchmark techniques by a substantial edge in anticipating motion picture productivity. Novel highlights we proposed likewise made extraordinary commitments to the expectation. Moreover, to plan a choice emotionally supportive network with reasonable utilities, our investigation of key factors for motion picture productivity may likewise have suggestions for hypothetical research on group execution and the achievement of imaginative work. Komal Gothwal et al. [12] developed a mathematical model for predicting the success class such as flop, hit, and neutral of the movies. For doing this we have to develop a methodology in which the historical data of each component such as actor, actress, director, and music that influences the success or failure of a movie is given due to weightage and then based on multiple thresholds calculated based on descriptive statistics of the dataset of each component it is given class flop, hit, neutral label. Based on the weightage of historical data of each film crew the movie will be labeled as neutral, hit, or flop. This system helps to find out whether the movie is a super hit, hit, or flop based on historical data of the actor, actress, music director, writer, director, marketing budget, and release date of the new movie. If the movie releases on weekend, the new movie will get higher weightage, or if the movie releases on weekdays new movie will get low weightage. The factors such as actor, actress, director, writer, music director, and marketing budget historical data of each component are calculated and movie success is predicted. This application helps to find out the review of the new movie.

V. Subramaniyaswamy et al. [13] analyze the efficiency of using multiple linear regression and Support Vector Machine Classification to predict the box-office success of movies while analyzing the influence of variables like trailer views, Wikipedia page views, critic ratings, and time of release. Nahid Quader et al. [14] proposed a decision support system for the movie investment sector using machine learning techniques. This research helps investors associated with this business for avoiding investment risks. The system predicts an approximate success rate of a movie based on its profitability by analyzing historical data from different sources like IMDb, Rotten Tomatoes, Box Office Mojo, and Metacritic. Using Support Vector Machine (SVM), Neural Network,

and Natural Language Processing the system predicts a movie box office profit based on some pre-released features and post-released features. This paper shows Neural Network gives an accuracy of 84.1% for pre-released features and 89.27% for all features while SVM has 83.44% and 88.87% accuracy for pre-released features and all features respectively when one away prediction is considered. Moreover, we figure out that budget, IMDb votes, and no. of screens are the most important features which play a vital role while predicting a movie‟s box- office success.

B. Prediction using KNN Classifier

Dipak Gaikar et al. [15] developed a mathematical model for predicting the rating and success classes such as hit, flop, and neutral of the movies. To do this, we have used machine learning and a data mining algorithm. The algorithm used for classification is k-NN. Popularity factor of various movie parameters like an actor, actress, director, writer, budget, etc. is collected which helps in the movie success prediction. This project helps the director or producer of the movie to decide the parameters such as actor, actress, etc. of the movie. This project also helps the user to decide whether to book a ticket in advance or not based on upcoming movie predictions.

George H. Chen and Devavrat [16] Shah focus on nonasymptotic statistical guarantees, which we state in the form of how many training data and what algorithm parameters ensure that the Nearest Neighbor prediction method achieves a user-specified error tolerance. We begin with the most general of such results for nearest neighbor and related kernel regression and classification in general metric spaces. In such settings in which we assume very little structure, what enables successful prediction is smoothness in the function being estimated for regression, and a low probability of landing near the decision boundary for classification. In practice, these conditions could be difficult to verify empirically for a real dataset. We then cover recent theoretical guarantees on nearest neighbor prediction in the three case studies of time series forecasting, recommending products to people over time, and delineating human organs in medical images by looking at image patches. In these case studies, clustering structure, which is easier to verify in data and more readily interpretable by practitioners, enables successful prediction.

Ladislav Peska and Peter Vojtas [17] described details of our approach to the RecSys Challenge 2014: User Engagement as Evaluation. The challenge was based on a dataset, which contains tweets that are generated when users rate movies on IMDb (using the iOS app on a smartphone). The challenge for participants is to rank such tweets by expected user interaction, which is expressed in terms of retweets and favorite counts. During experiments, we have tested several current off-the-shelf prediction techniques and proposed a variant of item biased k- NN algorithm, which better reflects user engagement and the nature of the movie domain content-based attributes. Our final solution (placed in the third quartile of the challenge leader board) is an aggregation of several runs of this algorithm and some off-the-shelf predictors.

C. Using Logistic Regression

Aashya Khanduja [18] carried out to correctly estimate the probability with which an unreleased movie will be successful in the given market. With further analysis and modification, this study can be used as early as in predicting if a movie‟s script will provide a favorable outcome for the production house. Using predictive modeling, we identify patterns and anomalies in a data set comprising historical information. This obtained order aids in foreseeing a value associated with new information and provides a probability attached to it. Its applications have proven to be invaluable in the field of social science, but we wish to extend it into further arenas. This paper presents the steps involved in developing a Logistic Regression model based on various parameters affecting movie ticket sales. Moghaddam et al. [19] proposed a novel method for predicting the popularity of movies. The method is based on hybrid visually-driven features, representative of the movie

content, which can be used to effectively predict not only the movie's popularity but also the average rating of the movie. Our extensive experiments on a large dataset of more than 13‟000 movie trailers show that the proposed hybrid approach achieves promising results by exploiting the visual Attractiveness features of movies in comparison to the other baseline features.

D. Using Random Forest

D. ABİDİN [20] Predicting movie success with machine learning algorithms has become a very popular research area. Many algorithms can be applied to a data set to make movie success predictions if the data set is prepared and represented properly. In this study, we explained how IMDB movie data was used for movie rating prediction. The data set extracted from IMDB was formatted and prepared for data mining algorithms.
These algorithms were executed on the WEKA application environment and the performances in movie ratings and confusion matrices were obtained. The seven machine learning algorithms used have performed well on the data set with varying performance ratings of 73.5% to 92.7%. The Random Forest algorithm had the best performance of 92.7%. This is the highest score obtained among similar studies. Suchita V et al. [21] classify movie reviews into positive or negative polarity by using machine learning algorithms such as Naïvebayes, support vector machine, and random forest. This approach makes the use of a machine learning classifier‟s set of configuration parameters known as hyper parameters of Random Forest and SVM, which are required to be tuned before a model gets trained. In this approach, if the classifier makes use of this hyper parameter then the random forest and SVM model results in high accuracy. The results obtained by this approach provide an accuracy of 84.29% for NaiveBayes, 96% for SVM, and 95% for the random forest. After making use of the hyper parameter the accuracy increased to 97.42% for SVM, and 96% for the random forest.

III. Applications of Machine Learning

Machine Learning (ML) is a buzzword in the technology world right now and for good reason, it represents a major step forward in how computers can learn. The need for Machine Learning Engineers is high in demand and this surge is due to evolving technology and the generation of huge amounts of data aka Big Data. The ML is used in various sectors nowadays in which some of the real-world applications are discussing below:

A. Traffic Alerts (Maps)

Now, Google Maps is probably the app we use whenever we go out and require assistance in directions and traffic. The other day I was traveling to another city and took the expressway and Maps suggested: “Despite the Heavy Traffic, you are on the fastest route”. But, how does it know that?
It‟s a combination of People currently using the service, Historic Data of that route collected over time, and a few tricks acquired from other companies. Everyone using maps is providing their location, average speed, and the route in which they are traveling which in turn helps Google collect massive Data about the traffic, which makes them predict the upcoming traffic and adjust your route according to it.

Social Media (Facebook)

One of the most common applications of Machine Learning is Automatic Friend Tagging Suggestions on Facebook or any other social media platform. Facebook uses face detection and Image recognition to automatically find the face of the person which matches its Database and hence suggests we tag that person based on DeepFace. Facebook‟s Deep Learning project DeepFace is responsible for the recognition of faces and identifying which person is in the picture. It also provides Alt Tags (Alternative Tags) to images already uploaded on Facebook.

B. Oil and Gas

This is perhaps the industry that needs the application of machine learning the most. Right from analyzing underground minerals and finding new energy sources to streaming oil distribution, ML applications for this industry are vast and are still expanding.

C. Transportation and Commuting (Uber)

If you have used an app to book a cab, you are already using Machine Learning to an extent. It provides a personalized application that is unique to you. Automatically detects your location and provides options to either go home or office or any other frequent place based on your History and Patterns. It uses a Machine Learning algorithm layered on top of Historic Trip Data to make a more accurate ETA prediction. With the implementation of Machine Learning, they saw a 26% accuracy in Delivery and Pickup [9].

D. Products Recommendations

Suppose you check an item on Amazon, but you do not buy it then and there. But the next day, you‟re watching videos on YouTube and suddenly you see an ad for the same item. You switch to Facebook, and there also you see the same ad. So how does this happen? Well, this happens because Google tracks your search history, and recommends ads based on your search history. This is one of the coolest applications of Machine Learning. 35% of Amazon‟s revenue is generated by Product Recommendations [9].

E. Healthcare

With the advent of wearable sensors and devices that use data to access the health of a patient in real-time, ML is becoming a fast-growing trend in healthcare. Sensors in wearable provide real-time patient information, such as overall health condition, heartbeat, blood pressure, and other vital parameters. Doctors and medical experts can use this information to analyze the health condition of an individual, draw a pattern from the patient history, and predict the occurrence of any ailments in the future. The technology also empowers medical experts to analyze data to identify trends that facilitate better diagnoses and treatment.[8]

F. Speech Recognition

All current speech recognition systems available in the market use machine learning approaches to train the system for better accuracy. In practice, most such systems implement learning in two distinct phases: pre-shipping speaker independent training and post-shipping speaker-dependent training. [7]

G. Computer Vision

The majority of recent vision systems, e.g., facial recognition software‟s, systems capable of automatic classification of microscopic images of cells, employ machine learning approaches for better accuracy. For example, the US Post Office uses a computer vision system with a handwriting analyzer thus trained to sort letters with handwritten addresses automatically with an accuracy level as high as 85%.

H. Government

Government agencies like utilities and public safety have a specific need for ML, as they have multiple data sources, which can be mined for identifying useful patterns and insights. For example, sensor data can be analyzed to identify ways to minimize costs and increase efficiency. Furthermore, ML can also be used to minimize identity theft and detect fraud.[10]

I. Bio-Surveillance

Several government initiatives to track probable outbreaks of diseases use ML algorithms. Consider the RODS project in western Pennsylvania. This project collects admissions reports to emergency rooms in the hospitals there, and an ML software system is trained using the profiles of admitted patients to detect aberrant symptoms, their patterns, and areal distribution. Research is ongoing to incorporate some additional data in the system, like over-the-counter medicines‟ purchase history to provide more training data. The complexity of this kind of complex and dynamic data set can be handled efficiently using automated learning methods only. [7]

J. Robot or Automation Control

ML methods are largely used in robot and automated systems. For example, consider the use of ML to obtain control tactics for stable flight and aerobatics of helicopters. The self-driving cars developed by Google use ML to train from collected terrain data. [7]

K. Empirical Science Experiments

A large group of data-intensive science disciplines uses ML methods in several of its research. For example, ML is being implemented in genetics, to identify unusual celestial objects in astronomy, and Neuroscience and psychological analysis. The other small scale yet important application of ML involves spam filtering, fraud detection, topic identification, and predictive analytics (e.g., weather forecast, stock market prediction, market survey, etc.).

L. Financial Services

Companies in the financial sector can identify key insights in financial data as well as prevent any occurrences of financial fraud, with the help of machine learning technology. The technology is also used to identify opportunities for investments and trade. Usage of cyber surveillance helps in identifying those individuals or institutions which are prone to financial risk and taking necessary actions in time to prevent fraud.

IV. Conclusion

The statistical analysis of data is effectively performed by machine learning techniques. It also gains information from historical data. The machine learning algorithm is classified into three categories namely supervised, unsupervised, and reinforcement learning and each classification technique has its merits and demerits. In this paper, we present the comparative analysis of the most used algorithm such as decision tree, support vector machine, and naïvebayes classifier and it is found that NaiveBayes gives more accurate results than the other two. This paper presented the review of literature for movie prediction and also presented the applications of machine learning in different sectors but after studying the literature it is found that these techniques have advantages and disadvantages. So for the effective and accurate prediction of movie success in the future, need to develop an ensemble classifier that can predict the data more accurately and effectively.

References

[1] https://www.edureka.co/blog/what-is-machine- learning/
[2] Christopher M. Bishop, “Pattern Recognition and Machine Learning (Information Science and Statistics)”, 2006. Page-3
[3] Russel, “Artificial Intelligence: A Modern Approach”, January 1, 2015
[4] Hastie et al. “The Elements of Statistical Learning: Data Mining,Inference, and Prediction”, Second Edition (Springer Series in Statistics), 2016, pp. 28. [online]. Available: https://www.amazon.com/Elements-Statistical- Learning-Prediction-Statistics.

[5] Richard S. Sutton et al., "Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning",2018, second edition, pp.-2. [Online]. Available: https://www.amazon.com/Reinforcement- Learning-Introduction-Adaptive-Computation.
[6] Mohammed et al., "Machine Learning Algorithms and Applications", 2017, CRC Press Taylor & Francis, London, ISBN-13:978-1- 4987-0538-7.
[7] Das and Behera, "A Survey on Machine Learning: Concept, Algorithms, and Applications", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 2, February 2017, pp. 1301-1309.
[8] [Online]: Available. https://www.outsource2india.com/software/articl es/machine-learning-applications-how-it-works- who-uses-it.asp
[9] [Online]: Available. https://www.edureka.co/blog/machine-learning- applications/
[10] Vaseem Naiyer, Jitendra Sheetlani, Harsh Pratap Singh, "Software Quality Prediction Using Machine Learning Application", Smart Intelligent Computing and Applications, Springer 2020. Pp. 319-327.
[11] Hemant Kumar and Santosh Kumar, "Predicting Movie Success or Failure using Linear Regression & SVM over Map-Reduce in Hadoop", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 6, Issue 6, June 2018, pp 6426-6433.
[12] Komal Gothwal et al., "Movie Success Prediction", IOSR Journal of Engineering (IOSRJEN), Volume 6, PP 66-69, ISSN (e): 2250-3021, ISSN (p): 2278-8719.
[13] V. Subramaniyaswamy et al., "Predicting movie box office success using multiple regression and SVM" International Conference on Intelligent Sustainable Systems (ICISS), 2017. In proceeding of IEEExplore.
[14] Nahid Quader et al., "A Machine Learning Approach to Predict Movie Box-Office Success", 20th International Conference of Computer and Information Technology (ICCIT), 22-24 December 2017. In proceeding of IEEExplore.
[15] Dipak Gaikar et al., "Movie Success Prediction Using Popularity Factor from Social Media", International Research Journal of Engineering and Technology (IRJET), Volume: 06, Issue: 04, Apr 2019.
[16] George H. Chen, Devavrat Shah, "Explaining the Success of Nearest Neighbor Methods in Prediction", Vol. XX, No. XX, pp 1–250. DOI: XXX.
[17] Ladislav Peska and Peter Vojtas "Biased k-NN Similarity Content Based Prediction of Movie Tweets Popularity", Dateso 2015, pp. 101–110, CEUR-WS.org/Vol-1343.
[18] Aashya Khanduja, "Logistic Regression for Predicting Movie‟s Success", Openreview.net, 2018.
[19] Farshad B. Moghaddam et al., "Predicting Movie Popularity and Ratings with Visual Features", 14th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP),2019. In proceeding of IEEExplore, 2019
[20] D. ABİDİN, C. BOSTANCI, A. SİTE, "Movie Rating Prediction with Machine Learning Algorithms on IMDB Data Set", International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES‟18), May 11-13, 2018 Safranbolu, Turkey.
[21] Suchita V. Wawre, Sachin N. Deshmukh, "Sentimental Analysis of Movie Review using Machine Learning Algorithm with Tuned Hypeparameter", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 6, June 2016.