BALTIC JOURNAL OF LAW & POLITICS
A Journal of Vytautas Magnus University
VOLUME 15, NUMBER 4 (2022)
ISSN 2029-0454

Cite: *Baltic Journal of Law & Politics* 15:4 (2022): 357-367
DOI: 10.2478/bjlp-2022-0054038

A Novel Advertisement Recommendation System Using Random Forest over K++ Means Algorithms

A.Nanda Kishore
Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105.

L.Rama Parvathy
Project Guide,Corresponding Author, Department of Computer Science and Engineering,
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105.

Received: August 8, 2022; reviews: 2; accepted: November 29, 2022

Abstract

Aim: The aim of the work is to evaluate the accuracy in predicting movie recommendation using Novel Random Forest (RF) Algorithm and K-Mean algorithm. **Materials and Methods**: The classification algorithm is invoked on a movie dataset consisting of 5000 records. A framework for movie recommendation prediction has been proposed and developed that compares Novel Random Forest and K-Means Algorithm. Sample size was calculated as 8 in each group using G powers. Sample size was calculated using clinical analysis, with alpha and beta values of 0.05 and 0.5, 95% confidence, 80% pre-test power. **Results**: The Random Forest algorithm produces an accuracy of 89.98% when predicting movie recommendation on data sets following the root mean square error, while criteria based K-Mean give 86.98% and it is carried out by performing sample T-test with ± 2SD and significance value as $p=0.002$ ($p<0.005$, 2-tailed). **Conclusion:** The results show that the performance of the Novel random forest algorithm is better than that of the K-Means in terms of accuracy.

Keywords

Machine learning, Novel Random forest, K-Means, Movies, Recommendation, Classifier, Decision tree.

INTRODUCTION

Nowadays, finding the most relevant information is a hard and time-killing process, by using the machine learning recommendation system algorithms it will be easy to get personalized recommendations (Harrington 2012). In this paper the recommendation system which is novel random forest is used to recommend a list of movies to a user which is based on his previous watch history, similarity parameters and current opinions on the movies and this algorithm acts very dynamically considering the root mean square error for the both negative and positive feedbacks, this model works on the basis of similarity between contents of previously watched movie and the movies in the database(Liebowitz

2016). This is a world where everything is being automated and something is happening which is really cool. The trending technologies like data science,machine learning,artificial intelligence making everything so easy,The recommendation System too was a part of that (Mohan et al. 2021). Data is growing enormously and retrieving the most relevant thing is a hard and time-consuming process for the users. Some of the applications are Netflix,Amazon prime,Aha,Youtube,Hotstar (Shukla et al. 2019)

Many articles were published on recommendation system using random forest algorithm like ScienceDirect 1,854, articles, IEEE Xplore 180 journals, Google scholar 19,350 articles over the past 5 years.So to deal with this big data definitely there is a need of automated recommendation algorithm to solve queries and to give best results (Zhang et al. 2015), already saw that e-commerce industry has adopted recommendation system for its users and it got a huge response for that, and also recommendation system transformation was happened almost in all aspects like such as movie, music, news, web page and document recommendations(Pivarski et al. 2020). Many companies like google, amazon, ola, instagram, facebook adopted these systems and are going in the profit zone.(Tyagi and Jawdekar 2016). Table 1 describes different types of movies, will take each feature of the movie database which given by the user and try to calculate the importance of each feature and based on the importance and the decision will be taken as which feature should be a parent (Swamynathan 2019) and which feature should be a child as tree to project the list of movies which are closely related to the movie given by the user. (Paterek 2012)

Previously our team has a rich experience in working on various research projects across multiple disciplines (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).The research gap identified for many articles was not able to recommend personalized movie recommendations. So the proposed work will act like When user went to any online movie platform ,there will be a big list of movies which throwed on the users face but this won't be good it will be hard to the user to find the movie, so here developed an content based recommendation model to give movie recommendations to the user model will accepts a movie from the user and based on the content available on that movie, it will find the decision value of each feature of movie, and based on the feature importance the model will recommend similar movies to user, with the help of Euclidean Distance (Zhang et al. 2015) and Root Mean Square Error. The main goal of the research is to prove the accuracy of the random forest is better than the K-means and K-Means++ algorithm.

MATERIALS AND METHODS

The research work is carried out in the Data Analytics Lab, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences. This research article has two groups and group 1 is K-Means++ algorithm(81.43) and group 2 is Random Forest (82.54). A group of 8 iterations are performed on each algorithm for calculation of better accuracy.some of the main attributes are genres, keywords, titles, overview, language. The sample size is calculated as 8 in each group using G power. The movie dataset has been used with a sample size of 5000 movies, 20 features and some missing values. Sample size was calculated using a dataset, with alpha and beta values 0.05 and 0.5, 95% confidence, pretest power 80% and enrolment ratio 1 (Ricci, Rokach, and Shapira 2015).

There are nearly 5000 records in the Movie lens dataset which is downloaded from www.kaggle.com. This dataset consists of genres like Thriller, Action, Drama, Narrative, Fantasy, Western,Fiction, Romantic, Comedy, Family action. From the movie lens dataset the content and rating extraction need to be done in order to give personalized recommendations.

ISSN 2029-0454
2022

The proposed work was performed in Google colab by importing datasets, Data cleaning, Content extraction, Splitting data into training and testing datasets and calculating the performance matrices. The Hardware and Software requirements for experimenting the work includes i3 processor,8th Generation, 500 GB HDD, 4 GB RAM, Windows OS, Python- Colab.

K-Means Algorithm

The Kmeans method is an iterative technique that attempts to split a dataset into K separate non-overlapping subgroups (clusters), each of which contains just one data point. It attempts to make intra-cluster data points as comparable as feasible while maintaining clusters as distinct (far) as possible. It distributes data points to clusters in such a way that the sum of the squared distances between them and the cluster's centroid (average of all the points in the cluster) is as little as possible. Within clusters, the less variance there is, the more homogenous (similar) the data points are. The following is how the k-means algorithm works: K is the number of clusters to specify. Initialize the centroids by shuffling the dataset and then picking K data points at random for the centroids without replacing them. Continue iterating until the centroids do not change. i.e. the clustering of data points does not change. Calculate the total of all data points' squared distances from all centroids .Assign each data point to the cluster that is closest to it (centroid). Calculate the cluster centroids by averaging all of the data points that correspond to each cluster.
Expectation-Maximization is the method used by kmeans to solve the problem. The data points are assigned to the nearest cluster in the E-step. The centroid of each cluster is computed in the M-step. A breakdown of how to solve it mathematically may be seen below as shown in fig. 1.

K-Means++ Algorithm

This approach guarantees that the centroids are correctly initialized and that the clustering quality is improved. The rest of the process is identical to the regular K-means algorithm, with the exception of initialization. That is, K-means++ is the conventional K-means algorithm with smarter centroids initialization.

Pseudocode for Novel K -means++ Recommendation Model
Input: Movie Dataset
Output: Recommendation accuracy

1) Read the training dataset as input
2) Preprocess the dataset and split into train and test
3) Load the data into variable
   SET x=movies[0:]
4) Select the Number k to decide number of clusters(genres)
   SET i=k
5) Select the random k centroids
   SET centroids=k
6) Assign each data point to the closest centroid
7) Calculate the variance and place a new centroid of each cluster.
8) Repeat step3 until all the data points ar been custard
9) Execute the following command
   SET kmeans = KMeans(n_clusters=i, init='k-means++', random_state= 42
10) Fit the training data to the k means model
   kmeans.fit(x)
11) Predict non-trained movie with the model
   SET Accuracy=Accuracy(Prediction[genre],movies[genre])
12) Print the Accuracy

Random Forest

Novel Random Forest is a well-known machine learning algorithm that uses the supervised learning method. In machine learning, it may be utilized for both classification and regression issues. It is based on ensemble learning, which is a method of integrating several classifiers to solve a complicated issue and increase the model's performance. Random Forest is a classifier that contains a number of decision trees on various subsets of a given dataset and takes the average to enhance the predicted accuracy of that dataset, according to the name. Rather than depending on a single decision tree, the novel random forest collects predictions from each tree and makes decisions based on the majority of votes as shown in fig. 2.

Because the random forest mixes numerous trees to forecast the dataset's class, some decision trees may correctly predict the output while others may not. However, when all of the trees are combined, the proper result is predicted. As a result, two assumptions for a better Random forest classifier are as follows:The dataset's feature variable should have some real values so that the classifier can predict correct outcomes rather than guesses.Each tree's predictions must have very low correlations.

The novel random forest is formed in two phases: the first is to combine N decision trees to build the random forest, and the second is to make predictions for each tree created in the first phase. The following stages and graphic can be used to demonstrate the working process:

Pseudocode for Random Forest Recommendation Model
Input: Movie Dataset
Output:Recommendation accuracy

1) Read the training dataset as input
2) Preprocess the dataset and split into train and test
3) Select K data points from the data set
   SET data=movies[0:k]
   SET X_train,X_test,y_train,y_test TO train_test_split(X,Y,
   SET test_size TO 0.20,
4) Build the decision trees for all the data points available in the data data frame
   SET clf=DecisionTreeClassifier()
5) Fit the data into decision tree classifier object
   SET clf=clf.fit(X_train,y_train)
6) Choose no of DT should build
   SET N=no
7) Repeat 4 and 5 steps for N times.
8) Create Random Forest object
   classifier=RandomForest(n_estimators=10,criterion="entropy")
9) Fitting the data into RD model
   classifier.fit(X_train,y_train)
10) Predict non-trained movie with the model
   SET Accuracy=Accuracy(Prediction[genre],movies[genre])
11) Print the Accuracy

The dataset utilized is movie_db_kaagle, which is a subset of the Movie Lens dataset. A total of 5000 people have rated 4804 films/included in the data collection The ratings are on a scale of one to ten.a scale of 0 to 5, with 0 representing unrated films. The table below shows a sample dataset as shown in table 2. The following format is used to express the data: User id, Movie id, Rating, Timestamp id as shown in Table. 2. The rows of the data sets were passed into the random forest machine learning model and it will be iterable for given numbers In each iteration the decision of the model will be changed due to the change in the root values of the model By name itself, by doing different random iterations it will give a final decision by clubbing all the minor or sub decisions.

Statistical Analysis

In addition to the experimental analysis, the work was statistically evaluated using the Statistical Package for the Social Sciences (SPSS). Analysis was performed to obtain the mean, standard deviation and mean of standard errors. To compare the parameters of the two groups, an independent variable T test was performed. In the analysis, the independent variables are Movie name,budget,tagline. The dependent variables that affect the output are accuracy and precision.finally determines the accuracy graph between k-means, k-means++ and random forest algorithms

RESULTS

Table 3 shows the different iterations of the two groups. Accuracy and Precision were calculated for novel random forest and K-Means. Table 4(Average) Two-group analysis shows that RF has higher accuracy (89.98%) and Precision (85.73%) than K-Means. Table 5 shows the statistical analysis of RF and K-Means with different sets of test data. Table 6 shows the Independent Samples T-test - RF seems to be significantly better than K-Means ($p=0.002$). The average accuracy of the RF model appears to be higher than that of the K-Means model. In addition, the precision of RF is much higher than the K-Means. The performance of the novel random forest algorithm is superior to that of the K-Means algorithm. There is no significant difference between the two groups. Therefore, RF is better than K-Means by considering the graph which is shown in Fig. 3.

DISCUSSION

Recommender systems are a type of information filtering system that is different from the rest. The transmission of items chosen from a huge collection is the subject of information filtering.things the user is likely to find interesting or useful, and which can be categorized as a task. A user model is induced based on training data, allowing the filtering system to function.to categorize unseen elements into a positive class c (relevant to the user) or a negative class n (not relevant to the user). The items that the user discovered make up the training set interesting. These elements combine to generate training instances, each of which has a unique attribute.

There are numerous ways to portray terms in order to use them as a foundation for the learning component. The vector space model is a commonly used representation approach. (Deisenroth, Aldo Faisal, and Ong 2020) A document D is represented as an m-dimensional vector in the vector space model, where each dimension corresponds to a separate term and m is the total number of terms used in the collection of documents. The document vector is represented as, where wi denotes the importance of the phrase ti and ti denotes the weight of the term. Weight wi is 0 if document D does not contain the phrase ti. The tf-idf scheme can be used to calculate term weights. The terms are given an a and the frequency with which they occur over the full d in this method. In (Ropero et al. 2009; Liu, Sung, and Yao 2015) shows up at least once Tf-assumptions idf's are based on two characteristics. For starters, the more times a term appears in a document, the more likely it is to be misunderstood. In (Alshuraiqi 2020) its relevance to the document's subject. Second, the more times a term appears in all, the better. The more documents in the collection, the worse it is at distinguishing between them.

The limitation of this research is that it cannot give appropriate results for smaller data. In this model it is not able to consider all given feature variable parameters for training. The future scope of proposed work will be movie prediction using Novel Random Based Filtering. Our future work will focus on faster and more accurate movie prediction through the use of advanced image processing techniques.

CONCLUSION

Novel Random forest filtering technique that uses the mean to improve accuracy. The work shows that the prediction of accuracy for movie prediction using Novel Random Forest (RF) is better than the K-Means. RF works significantly better than K-Means at accurately predicting the movies, but the average error is slightly greater than K-Means. Therefore, it is concluded that the novel Random Forest provides acceptable accuracy compared to K-Means.

DECLARATION

Conflict of interest

No conflict of interest in this manuscript.

Author Contribution

Author ANK was involved in data collection, data analysis, algorithm framing, implementation and manuscript writing. Author RP was involved in designing the work flow, guidance and review of manuscript.

Acknowledgements

We thank Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (formerly Saveetha University) for providing facilities and continued assistance to complete this study.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. QBEC INFOSOL Pvt. Ltd.
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

REFERENCES

Alshuraiqi, Hamza Sulimansallam. 2020. “Improved Term Frequency Inverse Document Frequency (TF-IDF) Method for Arabic Text Classification.” *International Journal of Advanced Trends in Computer Science and Engineering*. https://doi.org/10.30534/ijatcse/2020/11952020.

Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. “A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres.” *Materials Today: Proceedings* 33 (January): 798–805.

Deisenroth, Marc Peter, A. Aldo Faisal, and Cheng Soon Ong. 2020. *Mathematics for Machine Learning*. Cambridge University Press.

Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. “Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition.” *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.

Harrington, Peter. 2012. *Machine Learning in Action*. Simon and Schuster.

Liebowitz, Jay. 2016. *Big Data and Business Analytics*. CRC Press.

Liu, Hsiang-Chuan, Wen-Pei Sung, and Wenli Yao. 2015. *Computing, Control, Information and Education Engineering: Proceedings of the 2015 Second International Conference on Computer, Intelligent and Education Technology (CICET 2015), April 11-12, 2015, Guilin, P.R. China*. CRC Press.

Mohan, Neeraj, Ruchi Singla, Priyanka Kaushal, and Seifedine Kadry. 2021. “Artificial Intelligence, Machine Learning, and Data Science Technologies.” https://doi.org/10.1201/9781003153405.

Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. “Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia.” *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.

Paterek, Arkadiusz. 2012. *Predicting Movie Ratings and Recommender Systems: A 195-Page Monograph on Machine Learning, Recommender Systems, and the Netflix Prize*. Arkadiusz Paterek.

Pivarski, Jim, Ianna Osborne, Pratyush Das, Anish Biswas, and Peter Elmer. 2020. “Awkward Array: JSON-like Data, NumPy-like Idioms.” *Proceedings of the 19th Python in Science Conference*. https://doi.org/10.25080/majora-342d178e-00b.

Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. “Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City.” *Oral Health & Preventive Dentistry* 18 (1): 379–86.

Ricci, Francesco, Lior Rokach, and Bracha Shapira. 2015. *Recommender Systems Handbook*. Springer.

Ropero, Jorge, Ariel Gómez, Carlos León, and Alejandro Carrasco. 2009. “TERM WEIGHTING: NOVEL FUZZY LOGIC BASED METHOD VS. CLASSICAL TF-IDF METHOD FOR WEB INFORMATION EXTRACTION.” *Proceedings of the 11th International Conference on Enterprise Information*. https://doi.org/10.5220/0001982901300137.

Sathish, T., and S. Karthick. 2020. “Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite.” *Materials Today: Proceedings* 33 (January): 2555–58.

Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. “Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid.” In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.

Shukla, Rajesh Kumar, Jitendra Agrawal, Sanjeev Sharma, and Geetam Singh Tomer. 2019. *Data, Engineering and Applications: Volume 1*. Springer.

Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. “Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies.” *Materials Today Communications* 25 (December): 101603.

Swamynathan, Manohar. 2019. *Mastering Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics Using Python*. Apress.

Tyagi, Ramkumari, and Anand Jawdekar. 2016. “An Advanced Recommendation System for E-Commerce Users.” *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*. https://doi.org/10.1109/cdan.2016.7570876.

Venu, Harish, and Prabhu Appavu. 2021. “Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine.” *International Journal of Ambient Energy* 42 (14): 1588–94.

Zhang, Mimi, Jianping Chai, Yan Wang, and Xinran Wang. 2015. “The Internet Movie Customer Satisfaction Model Based on User Generated Content.” *2015 8th International Symposium on Computational Intelligence and Design (ISCID)*. https://doi.org/10.1109/iscid.2015.282.

TABLES AND FIGURES

Table 1: Type of Movies

Film Genres	Description
Action	Action film is a film genre in which the protagonist is thrust into a series of events that typically involve violence and physical feats.
Adventure	Adventure movies are a genre of movies. They contain many of the same features of action movies, but are usually set in exotic locations.

ISSN 2029-0454
2022

Animated	Animated Films are ones in which individual drawings, paintings, or illustrations are photographed frame by frame (stop-frame cinematography).
Comedy	A comedy film is a category of film which emphasizes humor. These films are designed to make the audience laugh through amusement. Films in this style traditionally have a happy ending.
Drama	Drama Films are serious presentations or stories with settings or life situations that portray realistic characters in conflict with either themselves, others, or forces of nature.
Fantasy	Fantasy films are films that belong to the fantasy genre with fantastic themes, usually magic, supernatural events, mythology, folklore, or exotic fantasy worlds.
Historical	A historical film is a fiction film showing past events or set within a historical period. This extensive genre shares territory with the biopic, costume drama, heritage film, and epic film.
Horror	Horror is a film genre that seeks to elicit fear or disgust in its audience for entertainment purposes. Horror films often explore dark subject matter and may deal with transgressive topics or themes.
Musical	Musical film is a film genre in which songs by the characters are interwoven into the narrative, sometimes accompanied by singing and dancing.
Romance	Romance movies are romantic love stories recorded in visual media for broadcast in theaters and on television that focus on passion, emotion.
Science Fiction	Science fiction (or sci-fi) is a film genre that uses speculative, fictional science-based depictions of phenomena that are not fully accepted by mainstream science.
Thriller	Thriller is a genre of fiction, having numerous, often overlapping subgenres. Thrillers are characterized and defined by the moods they elicit, giving viewers heightened feelings of suspense, excitement, surprise, anticipation and anxiety.
Western	Westerns often stress the harshness and isolation of the wilderness, and frequently set the action in an arid, desolate landscape.

Table 2. Schema of ratings given by users to movies

User_id	**Movie_id**	**Rating**	**Timestamp**
540	3856	5	245124542
890	4186	4	548451515
382	1028	2	548845454

Table 3: Accuracy and precision for the Random Forest(RF) and K-Means Algorithm

	Accuracy		**Presion**	
Iteration No	**RF**	**K-Means**	**RF**	**K-Means**
1	81.41	86.98	81.23	80.10
2	83.98	83.23	84.24	82.84
3	79.90	79.24	79.56	80.13
4	89.98	85.56	83.64	79.85
5	85.54	81.22	85.73	83.98
6	87.56	83.64	82.43	79.99
7	89.98	82.99	84.48	82.12
8	86.05	81.43	80.53	73.85

Table 4. Experimental analysis in Google colab for Accuracy, precision for RF and K-Means,. RF provides better Accuracy (89.98%) and Precision (85.73%) than K-Means

MODEL	**ACCURACY**	**PRECISION**
RF	89.98	85.73
K-Means	86.98	83.98

Table 5. Group Statistics Results-RF has an mean accuracy (85.55%), std.deviation (3.68), whereas forCBF has mean accuracy (83.03%), std.deviation (2.46)

	RF,K-Means	**N**	**MEAN**	**STD.DEVIATION**	**STD.ERROR MEAN**
Accuracy	**RF**	8	85.5500	3.68524	1.30293
	K-MEANS	8	83.0363	2.46742	0.87236

Table 6. Independent Samples T-test - RF seems to be significantly better than APR (p=0.002)

Accuracy	**Independent Samples Test**							
	Levene's Test for Equality of Variances					**T-test for Equality of Means**		
	F	**Sig**	**t**	**df**	**Sig(2-tailed)**	**Mean Difference**	**Std.Error Difference**	**95% Confidence Interval of the Difference**

ISSN 2029-0454
2022

								Lower	Upper
Equal variances assumed	1.264	.280	1.603	14	0.002	2.51375	1.56801	0.84929	5.87679
Equal variances not assumed			1.603	12.226	0.002	2.51375	1.56801	-0.89566	5.92316

Input: Training data(DS)
Output: Initial centroids of K-means
Initialize: p←Uniformly select a data point from DS randomly

while |d|<k
do
Sample s ϵ DS with probability $\frac{d^2\{s,p\}}{\phi s(p)}$
$p \leftarrow p \cup \{s\}$
end

Fig. 1. Probability for random data points in data sets

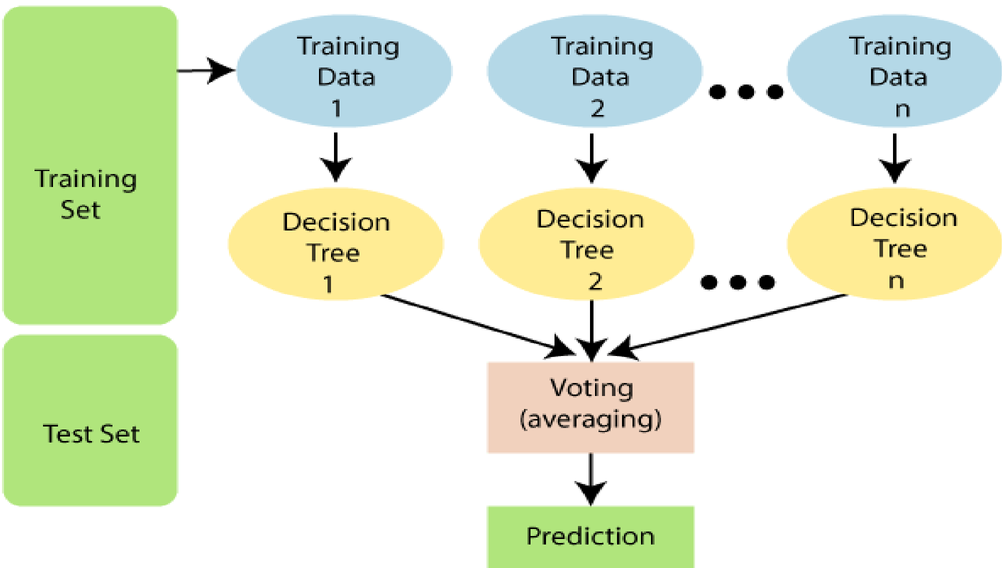

Fig. 2. Random Forest Algorithm architecture

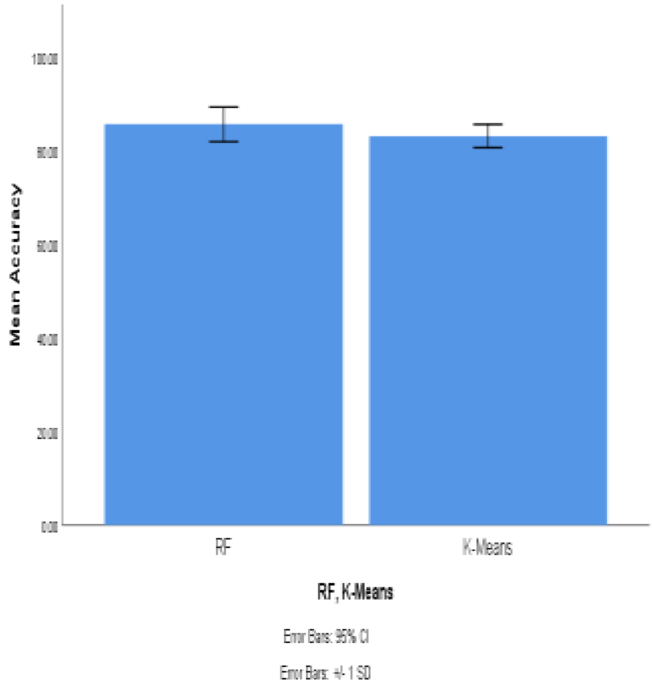

Fig. 3. Comparison of mean accuracy of RF and K-Means algorithms. RF appears to produce more consistent results with higher accuracy. X-axis: RF vs K-Means. Y-axis: Mean Accuracy. The error bars are 83% for both algorithms. The Standard Deviation Error Bars are +/- 2 SE.