ENHANCING ACCURACY FOR BEST ROUTE ANALYSIS BY RANDOM FOREST ALGORITHM OVER BAGGING CLASSIFIER

G Santosh kumar[1] , Palanikumar S[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: To perform the best route analysis using a novel Random forest algorithm compared with Bagging tree classifier**.**
Material and Methods: The data set in this paper utilizes the publicly available Kaggle data set. The sample size of a best route prediction system with improved accuracy rate was sample 20 (Group 1=10 and Group 2=10) and calculation is performed utilizing G-power 0.8 with alpha and beta qualities are 0.05, 0.2 with a confidence interval at 95%. The best route analysis is performed by using Random forest (RF) with a number of samples (N=10) and Bagging tree (BT) Classifier with a number of samples (N=10) respectively.
Results: The novel Random forest algorithm has 95.00 percent higher accuracy rates when compared to the accuracy rate of BT is 88.33 percent. The study has a significance value of $p<0.05$ i.e. $p=0.021$.
Conclusion: The proposed random forest achieved significantly better classification than the bagging classifier for predicting the best route analysis**.**

Keywords: Route analysis, Novel Random forest, Bagging Classifier, Traffic, Prediction, Accuracy.

[1]Research Scholar, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Tamil Nadu,
[2*]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu. India. Pincode: 602105.

1. Introduction

The number of automobiles and resident journeys among urban residents has steadily increased as a result of the social economy's quick expansion and the rise in living conditions of people (Chen et al. 2017). As a result, traffic difficulties in major cities, both domestically and internationally, have gained attention (Cheng 2021). The prevalence of issues such a bad road system, inadequate amenities, heavy traffic, etc. has increased. One of the most crucial answers to this issue is traffic route direction. The precise prediction of users' travel route is becoming an increasingly essential subject for traffic advice. More researchers are interested in determining which component influences the prediction most or in proposing one effective route prediction method. This research proposed to develop a new approach to find the best route analysis using a novel random forest (RF) algorithm (Yu et al. 2018). The benefit of employing random forest is that it just needs a modest quantity of training data to identify the estimated parameters required for classification. The evaluation results demonstrate that the proposed random forest technique reduces the false positives significantly while maintaining the accuracy for the best path prediction (Sun, Zhuang, and Ma 2019).

Several approaches, such as machine learning and data mining have been used in travel route prediction over the past few years. IEEE Explore published 85 research papers, and Google Scholar found 78 articles. Karbassi and Barth (Karbassi and Barth 2003) addressed route prediction for smart vehicles for a car-sharing application. To determine the path a driver would follow between two specified starting and finishing drop-off points was their assignment. We don't depend on the passenger to enter their destination when doing our duties. Zhang et al. (C. Zhang et al. 2020) used machine learning and a particle extract to analyse GPS data to forecast people's destinations, routes, and even modes of transportation. To efficiently find route patterns, Cao et al. (Cao, Mamoulis, and Cheung 2005) developed a substring tree structure and enhanced level-wise mining technique. The bagging algorithm, developed by Giannotti et al. (Giannotti, Nanni, and Pedreschi 2006), is based on the tree method (Han et al. 2001) and can extract patterns from time annotated sequences. They also suggested an alternative version of the algorithm for mining trajectory patterns (Giannotti and Pedreschi 2008). To find lengthy, shareable route patterns, Gidófalvi and Pedersen (Gidófalvi and Pedersen 2009) presented a projection-based approach. A predicted location model was put forth by Karimi and Liu and route models were created by Simmons et al. (Simmons et al. 2006) using a Hidden Markov Model (HMM). To find route patterns, Froehlich and Krumm (Froehlich and Krumm 2008) suggested a Hausdorff distance-based clustering approach.

Our institution is keen on working on latest research trends and has extensive knowledge and research experience which resulted in quality publications (Rinesh et al. 2022; Sundararaman et al. 2022; Mohanavel et al. 2022; Ram et al. 2022; Dinesh Kumar et al. 2022; Vijayalakshmi et al. 2022; Sudhan et al. 2022; Kumar et al. 2022; Sathish et al. 2022; Mahesh et al. 2022; Yaashikaa et al. 2022). The main disadvantage of the bagging classifier approach is that it calculates all of the training examples for every test data during classification, leading to high temporal difficulty during the testing stage, further raising the computation complexity, and necessitating a significant amount of memory. This paper aims to develop a novel random forest (RF) algorithm to determine the most accurate route forecast in congested traffic. The algorithm has to find the shortest path between the source and destination location. The output results are compared with the bagging tree (BT) classifier. The experimental findings indicate that the proposed RF algorithm outperforms the bagging classifier in terms of accuracy and loss rate.

2. Materials and Methods

This work was carried out in the Networking laboratory, Department of Information Technology, Saveetha School of Engineering. In this study, the dataset was collected from the kaggle repository. The database is divided by the amount of 75% training and 25% testing. Two sets are taken and 10 data samples for each set, total number of samples considered are 20. Group 1 was a bagging classifier algorithm and Group 2 was a novel random forest (RF) algorithm. The output is obtained by using Matlab software for the prediction of the best route in traffic. The calculation is performed utilizing G-power 0.8 with alpha and beta qualities 0.05, 0.2 with a confidence interval at 95% (Liu et al. 2018).

Bagging Classifier

Bagging is a technique for creating numerous classifier iterations in order to obtain an aggregated one. Bootstrap repeats from the training dataset are used to create the various classifier versions. Bootstrap aggregating gives the term "bagging" its origin. The notion is that each classifier serves as a member for each unit, much like in a parliament. The largest number of votes cast by the B classifiers determines the final classification. The newly introduced classifier will compile all the data gathered from the bootstrap duplicates.

Pseudo code for Bagging classifier
Input – To find the short distance path in traffic _Input Features

Output - Accuracy
Step 1: Generate the bootstrap sample from the dataset.
Step 2: Train a base learner $h_k$from the bootstrap sample.
Step 3: Extract the bth bootstrap sample on sth slave from the data matrix.
Step 4: Fit the classifier Csb(xi) to the bth bootstrapped training sample
Step 5: Create the bagging tree node with the best attribute.
Step 6: Make predictions and verify their accuracy.

Random Forest
Random forest is a supervised algorithm for guided learning. The "forest" it creates is made up of a group of decision trees that are often trained using the "bagging" method. The main concept of the bagging method is that the final result is improved by combining learning models. A random forest's hyperparameters are pretty comparable to a D-or tree's a bagging algorithm's. Fortunately, employing a random forest classifier-class eliminates the need to integrate a decision tree and a bagging classifier. You can also solve regression issues with random forest by using the algorithm's regressor. As the trees grow, the random forest increases the model's randomness. It chooses the highest value from a selection of attributes at random, not the most significant, when dividing a node.

Pseudo code for Random Forest
Step 1: Input: Best route prediction in traffic _Input Features
Step 2: Output: Accuracy
Step 3: Random Forest function (input features IF = 1... n)
Step 4: Perform while (conditioning)
Step 5: Pick 'k' features at random from the input characteristics IF = 1...n.
Step 6: Choose the Training and Validation dataset for the Input features IF = 1...n.
Step 7: Using the best divide point out of "k" characteristics, calculate the root node to build the tree.
Step 8: Use the best divide point to split the node into the current node.
Step 9: Repetition of steps 1 through 5 is required to grasp all nodes.
Step 10: Repeat steps 4 through 6 to create a random forest tree for Trees T.
Step 11: End while
Step 12: Return Classification outcomes

Statistical Analysis
Matlab software is used to generate the results (Knight 2019). A monitor with a resolution of 1024x768 pixels was required to train these datasets (10th gen, i5, 12GB RAM, 500 GB HDD). The software program IBM SPSS is employed in this study for statistical analysis (Yockey 2017). The independent sample t test was used to determine the mean, standard deviation, and standard error mean statistical significance between the groups, and then the two groups were compared using SPSS software to obtain accurate values for the two different groups, which were then used with the graph to calculate the significant value with maximum accuracy (95.00 percent), mean value (95 percent), and standard deviation value (0.12323). Accuracy is a dependent variable, while random forest and Bagging classifiers are independent variables.

3. Results

The accuracy rate of the RF classifier compared to the BT classifier is shown in Figure 1. The RF classifier has a higher accuracy rate of 95.00 when compared to the BT classifier, which has 88.33. The RF classifier is significantly different from the BT classifier ($p<0.05$ independent sample test). RF, BT accuracy rates are plotted on the X-axis. Y-axis: Mean accuracy rate for keyword identification, ± 1 SD with 95 percent confidence interval.

Table 1 presents the evaluation metrics of the comparison of the RF classifier with the BT classifier. The RF classifier has a 95.00 accuracy rate, whereas the BT classifier has 88.33, respectively. In all parameters, the RF classifier outperforms the BT in the prediction of the best route analysis, with a higher accuracy rate.

Table 2 displays the statistical computations for RF, BT classifier, such as mean, standard deviation, and standard error mean. In the t-test, the accuracy rate parameter is used. The RF classifier has a mean accuracy rate of 95.00, while the BT classifier has 88.33, respectively. The RF classifier has a standard deviation of 0.12323, while BT has a standard deviation of 1.72843 respectively. The RF classifier has a Standard Error Mean of 0.12894, while BT has a Standard Error Mean of 0.82931 respectively.

Table 3 shows the statistical computations for independent samples of RF compared to the BT classifier. The accuracy rate has a significance level of 0.021. The RF classifier is compared to BT using an Independent samples T-test with a confidence interval of 95 percent and a threshold of significance of 0.82738. The significance level is 0.001, the significance level is two-tailed, the mean difference, the standard error difference, and the lower and upper interval difference are all included in this independent sample test.

4. Discussion

In this research, a novel random forest algorithm with proposed architecture and bagging tree algorithm is implemented for best route prediction in road traffic. The performance of the proposed algorithms is studied and implemented in Matlab software. The proposed novel RF gives high accuracy and is more efficient than the BT algorithm. Random forest algorithm has significantly higher accuracy, about 95.00 percent compared to the BT method of 88.33 percent. The novel Random grid algorithm seems to give more consistent results with a minimal standard deviation.

Some similar discoveries are Simmons et al. (Simmons et al. 2006) anticipate destinations and routes, based on their understanding of the road network. Although the following road section is the only one related to the present one in 94% of the cases, they claim a prediction accuracy of up to 98.7%. In order to produce a model of drivers' route choosing behaviour under the effect of traveller information, Lee et al. (Lee et al. 2005) integrated SVM and Genetic method and created a GNN model. The accuracy for the multinomial Logit model varied from 70.5% to 75.2%, the accuracy for NN varied from 94.2% to 95.3%, and the accuracy for the proposed model varied from 93.3% to 96.3% based on the expressed preference data gathered from Korean drivers. To describe how travellers choose their modes of transportation, Zhang and Xie (Y. Zhang and Xie 2008) developed an SVM model and evaluated its effectiveness against Logistic regression and Support Vector Machine models. The outcomes demonstrate that the Logistic regression performed worse than both LR and SVM. When given information, Gupta et al. (Gupta et al. 2018) built a standard random forest to mimic drivers' route selection between sideways and freeways. The findings indicate that with adequate adjustment, the RF can replicate participants' actual choices 90% to 93% of the time.

The proposed method contains limitations, such as an excessive usage of data. Furthermore, it is difficult to update a random forest with new samples. Adding upgrades and treatments to the nodes to be trimmed selection can also improve output accuracy. In the future, the findings of this methodology will be compared to those of other categorization approaches.

5. Conclusion

In this study, we developed a machine learning based technique for predicting the best route analysis**.** The suggested model incorporates a novel random forest (RF) classifier and a Bagging tree (BT) model, with the RF classifier achieving the greatest accuracy values. The RF classifier's accuracy rate is 95.00 percent higher than the BT model's accuracy rate of 88.33 percent for predicting the best route analysis**.**

Declarations

Conflict of interests

No irreconcilable situation in this original copy

Authors Contribution

Author G.Santosh kumar was involved in data collection, data analysis, manuscript writing. Author Palanikumar.S was involved in the Action process, Data verification and validation, and Critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this research work successfully.

Funding: We thank the following organization for providing financial support that enabled us to complete the study.

1.Azacus Techno Solutions, Chennai.
2.Saveetha University
3.Saveetha Institute of Medical and Technical Sciences
4.Saveetha School of Engineering

6. References

Cao, Huiping, N. Mamoulis, and D. W. Cheung. 2005. “Mining Frequent Spatio-Temporal Sequential Patterns.” In Fifth IEEE International Conference on Data Mining (ICDM’05), 8 pp. – .

Chen, Dawei, Dongwoo Kim, Lexing Xie, Minjeong Shin, Aditya Krishna Menon, Cheng Soon Ong, Iman Avazpour, and John Grundy. 2017. “PathRec: Visual Analysis of Travel Route Recommendations.” In Proceedings of the Eleventh ACM Conference on Recommender Systems, 364–

65. RecSys ’17. New York, NY, USA: Association for Computing Machinery.

Cheng, X. 2021. “A Travel Route Recommendation Algorithm Based on Interest Theme and Distance Matching.” EURASIP Journal on Advances in Signal Processing. https://link.springer.com/article/10.1186/s13634-021-00759-x.

Dinesh Kumar, M., V. Godvin Sharmila, Gopalakrishnan Kumar, Jeong-Hoon Park, Siham Yousuf Al-Qaradawi, and J. Rajesh Banu. 2022. “Surfactant Induced Microwave Disintegration for Enhanced Biohydrogen Production from Macroalgae Biomass: Thermodynamics and Energetics.” Bioresource Technology 350 (April): 126904.

Froehlich, Jon, and John Krumm. 2008. “Route Prediction from Trip Observations.” In Society of Automotive Engineers (SAE) 2008 World Congress, April 2008. https://www.microsoft.com/en-us/research/publication/route-prediction-trip-observations/.

Giannotti, Fosca, Mirco Nanni, and Dino Pedreschi. 2006. “Efficient Mining of Temporally Annotated Sequences.” In Proceedings of the 2006 SIAM International Conference on Data Mining (SDM), 348–59. Proceedings. Society for Industrial and Applied Mathematics.

Giannotti, Fosca, and Dino Pedreschi. 2008. Mobility, Data Mining and Privacy: Geographic Knowledge Discovery. Springer.

Gidófalvi, Győző, and Torben Bach Pedersen. 2009. “Mining Long, Sharable Patterns in Trajectories of Moving Objects.” GeoInformatica 13 (1): 27–55.

Gupta, Bharat, Shivam Awasthi, Rudraksha Gupta, Likhama Ram, Pramod Kumar, Bakshi Rohit Prasad, and Sonali Agarwal. 2018. “Taxi Travel Time Prediction Using Ensemble-Based Random Forest and Gradient Boosting Model.” In Advances in Big Data and Cloud Computing, 63–78. Springer.

Han, Jiawei, Jian Pei, Behzad Mortazavi-Asl, Helen Pinto, Qiming Chen, Umeshwar Dayal, and Meichun Hsu. 2001. “Prefixspan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth.” In Proceedings of the 17th International Conference on Data Engineering, 215–24. Citeseer.

Karbassi, A., and M. Barth. 2003. “Vehicle Route Prediction and Time of Arrival Estimation Techniques for Improved Transportation System Management.” In IEEE IV2003 Intelligent Vehicles Symposium. Proceedings (Cat. No.03TH8683), 511–16.

Knight, Andrew. 2019. Basics of MatLab® and beyond. Chapman and Hall/CRC.

Kumar, J. Aravind, J. Aravind Kumar, S. Sathish, T. Krithiga, T. R. Praveenkumar, S. Lokesh, D. Prabu, A. Annam Renita, P. Prakash, and M. Rajasimman. 2022. “A Comprehensive Review on Bio-Hydrogen Production from Brewery Industrial Wastewater and Its Treatment Methodologies.” Fuel. https://doi.org/10.1016/j.fuel.2022.123594.

Lee, S. B., Y. C. Kim, M. Namgung, and J. W. Kim. 2005. “Development of Route Choice Behavior Model Using Linkage of Neural Network and Genetic Algorithm with Trip Information.” KSCE Journal of Civil. https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/BF02829045&casa_token=8_QKcrkU_m8AAAAA:JKid3fq2wUUoqAMk5QYSk7lGHj3WKnMq1p2Ty6k_FKQD8XYx66yKs8T_86CNRJ1XnLvg8zWP8gXcTAXF.

Liu, Y., M. Hansen, D. J. Lovell, and M. O. Ball. 2018. “Predicting Aircraft Trajectory Choice–a Nominal Route Approach.” Proc. of the International. https://blog.umd.edu/nextor/files/2019/02/ICRAT2018_paper-1l5foji.pdf.

Mahesh, Narayanan, Srinivasan Balakumar, Uthaman Danya, Shanmugasundaram Shyamalagowri, Palanisamy Suresh Babu, Jeyaseelan Aravind, Murugesan Kamaraj, and Muthusamy Govarthanan. 2022. “A Review on Mitigation of Emerging Contaminants in an Aqueous Environment Using Microbial Bio-Machines as Sustainable Tools: Progress and Limitations.” Journal of Water Process Engineering. https://doi.org/10.1016/j.jwpe.2022.102712.

Mohanavel, Vinayagam, K. Ravi Kumar, T. Sathish, Palanivel Velmurugan, Alagar Karthick, M. Ravichandran, Saleh Alfarraj, Hesham S. Almoallim, Shanmugam Sureshkumar, and J. Isaac JoshuaRamesh Lalvani. 2022. “Investigation on Inorganic Salts K2TiF6 and KBF4 to Develop Nanoparticles Based TiB2 Reinforcement Aluminium Composites.” Bioinorganic Chemistry and Applications 2022 (January): 8559402.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. “Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications.” Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102102.

Rinesh, S., K. Maheswari, B. Arthi, P. Sherubha,

A. Vijay, S. Sridhar, T. Rajendran, and Yosef Asrat Waji. 2022. "Investigations on Brain Tumor Classification Using Hybrid Machine Learning Algorithms." Journal of Healthcare Engineering 2022 (February): 2761847.

Sathish, T., V. Mohanavel, M. Arunkumar, K. Rajan, Manzoore Elahi M. Soudagar, M. A. Mujtaba, Saleh H. Salmen, Sami Al Obaid, H. Fayaz, and S. Sivakumar. 2022. "Utilization of Azadirachta Indica Biodiesel, Ethanol and Diesel Blends for Diesel Engine Applications with Engine Emission Profile." Fuel. https://doi.org/10.1016/j.fuel.2022.123798.

Simmons, R., B. Browning, Yilu Zhang, and V. Sadekar. 2006. "Learning to Predict Driver Route and Destination Intent." In 2006 IEEE Intelligent Transportation Systems Conference, 127–32.

Sudhan, M. B., M. Sinthuja, S. Pravinth Raja, J. Amutharaj, G. Charlyn Pushpa Latha, S. Sheeba Rachel, T. Anitha, T. Rajendran, and Yosef Asrat Waji. 2022. "Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model." Journal of Healthcare Engineering 2022 (February): 1601354.

Sundararaman, Sathish, J. Aravind Kumar, Prabu Deivasigamani, and Yuvarajan Devarajan. 2022. "Emerging Pharma Residue Contaminants: Occurrence, Monitoring, Risk and Fate Assessment – A Challenge to Water Resource Management." Science of The Total Environment. https://doi.org/10.1016/j.scitotenv.2022.153897.

Sun, Junjie, Chenyi Zhuang, and Qiang Ma. 2019. "User Transition Pattern Analysis for Travel Route Recommendation." IEICE Transactions on Information and Systems E102.D (12): 2472–84.

Vijayalakshmi, V. J., Prakash Arumugam, A. Ananthi Christy, and R. Brindha. 2022. "Simultaneous Allocation of EV Charging Stations and Renewable Energy Sources: An Elite RERNN-m2MPA Approach." International Journal of Energy Research. https://doi.org/10.1002/er.7780.

Yaashikaa, P. R., P. Senthil Kumar, S. Jeevanantham, and R. Saravanan. 2022. "A Review on Bioremediation Approach for Heavy Metal Detoxification and Accumulation in Plants." Environmental Pollution 301 (May): 119035.

Yockey, Ronald D. 2017. SPSS® Demystified: A Simple Guide and Reference. Routledge.

Yu, Bin, Huaizhu Wang, Wenxuan Shan, and Baozhen Yao. 2018. "Prediction of Bus Travel Time Using Random Forests Based on Near Neighbors." Computer-Aided Civil and Infrastructure Engineering 33 (4): 333–50.

Zhang, Chengkai, Junchi Bin, Wells Wang, Xiang Peng, Rui Wang, Richard Halldearn, and Zheng Liu. 2020. "AIS Data Driven General Vessel Destination Prediction: A Random Forest Based Approach." Transportation Research Part C: Emerging Technologies 118 (September): 102729.

Zhang, Yunlong, and Yuanchang Xie. 2008. "Travel Mode Choice Modeling with Support Vector Machines." Transportation Research Record 2076 (1): 141–50.

Dey, N., Kumar, G., Vickram, A. S., Mohan, M., Singhania, R. R., Patel, A. K., ... & Ponnusamy, V. K. (2022). Nanotechnology-assisted production of value-added biopotent energy-yielding products from lignocellulosic biomass refinery–a review. Bioresource Technology, 344, 126171.

Tables and Figures

Table 1. The evaluation metrics of the comparison of the RF classifier with the BT classifier has been presented. The RF classifier has a 95.00 accuracy rate, whereas the BT classifier has 88.33, respectively. In all parameters, the RF classifier outperforms the BT in the prediction of the best route analysis, with a higher accuracy rate.

SI.No.	Test Size	ACCURACY RATE	
		RF	BT
1	Test1	93.69	85.51

2	Test2	93.95	86.92
3	Test3	93.97	86.25
4	Test4	93.78	86.81
5	Test5	94.76	87.73
6	Test6	94.84	87.92
7	Test7	94.52	87.57
8	Test8	94.33	88.45
9	Test9	95.13	88.22
10	Test10	95.43	88.35
Average Test Results		95.00	88.33

Table 2. The statistical computations for RF, BT classifier, such as mean, standard deviation, and standard error mean. In the t-test, the accuracy rate parameter is used. The RF classifier has a mean accuracy rate of 95.00, while the BT classifier has 88.33, respectively. The RF classifier has a standard deviation of 0.12323, while BT has a standard deviation of 1.72843. The RF classifier has a standard error mean of 0.12894, while BT has a standard error mean of 0.82931.

Group		**N**	**Mean**	**Standard Deviation**	**Standard Error Mean**
Accuracy Rate	BT	10	88.33	1.72843	0.82931
	RF	10	95.00	0.12323	0.12894

Table 3. The statistical computations for independent samples of RF compared to the BT classifier. The accuracy rate has a significance level of 0.021. The RF classifier is compared to BT using an Independent samples T-test with a confidence interval of 95 percent and a threshold of significance of 0.82738. The significance level is 0.001, the significance level is two-tailed, the mean difference, the standard error difference, and the lower and upper interval difference are all included in this independent sample test.

Group		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 (Lower)	95% Confidence Interval (Upper)
Accuracy	Equal variances assumed	8.342	0.021	16.234	18	.001	12.78322	0.82738	11.62738	13.62738
	Equal variances not assumed			16.234	12.345	.001	11.12823	0.21273	10.11273	12.2342

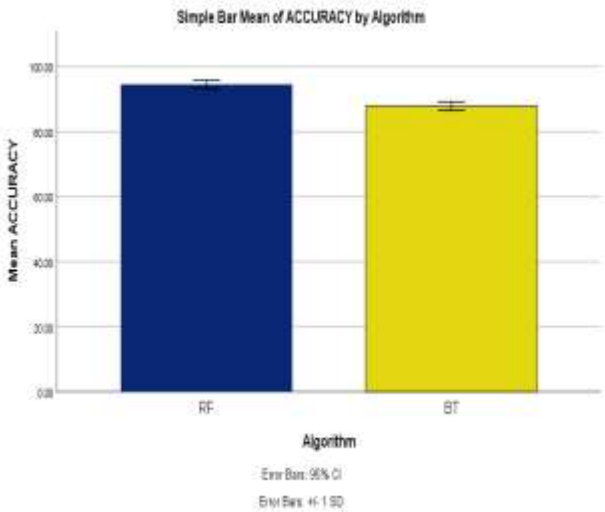

Fig. 1. The accuracy rate of the RF classifier compared to the BT classifier is shown in the bar graph. The RF classifier has a higher accuracy rate of 95.00 when compared to the BT classifier, which has 88.33. There is a significant difference between RF classifier and BT model ($p<0.05$ Independent sample test. RF, BT accuracy rates are plotted on the X-axis. Y-axis: Mean accuracy rate for keyword identification, ± 1 SD with 95 percent confidence interval.