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): 224-231
DOI: 10.2478/bjlp-2022-004023

Efficient Analysis of Road Accident using Random Forest Comparison over Decision Tree Algorithm to Improve Accuracy to Predict the Severity of Accident

B.Charan Sai
Research Scholar, 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,

Ashwini S
Project Guide, Corresponding Author,
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,

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

Abstract

Aim: To analyze road accidents using Random Forest comparison over Decision Tree algorithms to improve the accuracy and to predict the severity of the accidents. **Materials and Methods:** Random Forest Algorithm and Decision Tree algorithm used to predict the accuracy percentage of Traffic accident severity analysis. The sample size was measured as 10. G-power is calculated for two different groups, alpha (0.05), power (80%). Here the sample size is 40 and the iteration of the algorithms is 30, to attain the significant value is 0.02. **Results:** The results achieved with p=0.02 shows that two groups are statistically significant. It was observed that the Random Forest algorithm obtains the accuracy as 85.80. It appears to have better accuracy than the Decision Tree algorithm 78.80%. **Conclusion:** This study concludes that the Random Forest Algorithm 85.80% performs significantly better than Decision Tree 78.80%.

Keywords:

Accident Severity, Tree Classifier, Decision Tree, Random Forest Algorithm, Machine learning, Novel tree specific Random Forest, Road accident.

INTRODUCTION

Every year car accidents cause hundreds and thousands of deaths world wide. According to a research conducted by the World Health Organization, There had been 1.35 million street visitors deaths globally in 2016 and with loads of heaps extra retaining excessive injuries and dwelling with prolonged time period fitness consequences. Globally, avenue visitors crashes are a main cause of death among younger people (Meena 2017), and the primary reasons of demise amongst the ones elderly 15-29 years. Road traffic injuries are currently estimated to be the eight leading cause of death across all age groups

globally (Darçın 2020), and are predicted to become the seventh leading cause of death by 2030 (AlMamlook et al. 2019). Here two algorithms Novel tree Specific Random Forest and Decision tree have been implemented to analyze the road accident Severity (Tiwari 2018). By comparing the existing System we can observe that there is a lack of prediction of the accident severity (Kwayu et al. 2021). In the existing system they have explained only about the accident that will happen or not (Ghandour, Hammoud, and Al-Hajj 2020).

There are 1350 papers published on Road accident Severity analysis in the research gate. 1620 papers published in google scholar for road accident analysis (Darçın 2020). This system will analyze road safety and the obtained accuracy is 85.80%. To achieve accuracy a novel tree classifier is used. This research study used novel tree specific Random Forest Algorithm to perform the operations to give the best exactness of reason for road accidents and to find the severity of accidents.Leveraging the tools and all the data information nowadays available, An extensive analysis to predict the traffic accidents and its severity would make a difference to death trolls (Alkheder, Taamneh, and Taamneh 2017). Analyzing a significant range of factors, including weather, light, type of road, time among others an accurate prediction of the severity of the accidents can be performed(Maulida and Mutijarsa 2021).Thus, the trends commonly lead to severe traffic accidents can help identify the highly severe accidents (Jiang and Deng 2020). This kind of information could be used by emergency services to send the exact required staff and equipment to the place of accidents (Bojino et al. 2022). This severe accident situation can be warned to nearby hospitals which can have all the equipment ready for a severe intention in advance (GharehGozlu 2021).consequently,road safety should be a prior interest for governments, local authorities and private companies investing in technologies that can help reduce accidents and improve overall driver safety (Panday et al. 2021). Data might contribute to determining the likeliness of a potential accident occurring might include information on previous accidents such as road conditions,weather conditions,exact time and place of the accident,type of vehicles involved in the accident(Alkhudhairy, Aldhalemi, and Saki 2022). This project aims to forecast the severity of accidents with previous information that could be given by a witness informing the emergency services(Schlögl and Stütz 2019).

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).Government should be highly interested in accurate predictions of the severity of an accident, in order to reduce the time of arrival and make a more efficient use of the resources, and thus save a significant amount of people each year (Taamneh and Taamneh 2021).Others interested could be private companies investing in technologies aiming to improve road safeness (Jiang and Deng 2020)Based on the observation of the above literature survey, The implementation of the road accident severity analysis is made with the machine learning algorithms. The Novel tree specific Random Forest set of rules seems to carry out appreciably higher than the selection tree set of rules. The proposed work is used for better accuracy and precision.

MATERIALS AND METHODS

The research study was done in the Data Analytics lab, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (SIMATS). Two groups were identified for the study setting where the group one Novel tree specific Random forest group two is the decision tree. Using The G power sample sizes are totally 20 sample sizes are totally 20 sample sizes out for study 95% confidence and pretest power 80%.

Dataset for this algorithm is traffic crashes from the data.world website. This data set contains 12720 rows and 61 columns; the attribute information is

object_id,crash_seve, etc (Maulida and Mutijarsa 2021). To train the Novel tree specific Random Forest the class is used from the sk learn from the svc library. In this case, use the optimal values as parameters. It will be used for the fitting of the algorithm to the dataset and will show the prediction of accident severity. To train the Decision Tree use the decision novel tree classifier from the sklearn.tree library. Decision novel tree classifier also takes criterion as parameter.

Decision Tree Algorithm

Decision tree set of rules is a supervised mastering that may be used for each type and regression. It is a graphical illustration for buying all of the Possible answers to a hassle primarily based totally at the given condition.

Uses cart set of rules which stands for class and regression tree set of rules, there are various algorithms in device getting to know, so deciding on the exceptional set of rules for the given records set and hassle is the principle factor to bear in mind even as growing the device getting to know the model.

Step 1: Import and read the dataset.
Step 2: Extract the data.
Step 3: Split the data set.
Step 4: Fit the Decision tree.
Step 5: Predict the test set.
Step 6: Calculate the accuracy.

Random Forest Algorithm

Random Forest set of rules is a broadly used gadget gaining knowledge of a set of rules that belongs to the supervised gaining knowledge of technique, it could be used for each type and regression evaluation in gadget gaining knowledge. It is a machine learning model, It is primarily based totally at the standards of gadget gaining knowledge of machine learning.

Random wooded area is a type that is composed of some of choice timber on diverse subsets of the given statistics set and takes the common to enhance the predictive accuracy of the statistics set. The more quantity of better accuracy and stops the trouble of overfitting. Random wooded vicinity works in ways: first is to create the random wooded vicinity thru such as N desire tree, and second is to make predictions for each tree created withinside the primary phase regression.

It is able to cope with each class and regression tasks. It is able to cope with huge datasets with excessive dimensionality. It complements the accuracy of the version and stops the overfitting issue.it takes much less schooling instances in comparison to different algorithms. It predicts output with higher accuracy, even for the maximum vital dataset it runs efficiently. It also can preserve accuracy while a large percentage of facts is missing

Step 1: Import libraries.
Step 2: Read and load the data set.
Step 3: Generate the RF classifier criterion as a parameter.
Step 4: Gini was used as a parameter value.
Step 5: Predict the result for every sample.
Step 6: Calculate the accuracy.

Statistical Analysis

For statistical implementation, the software program for use right here is IBM SPSS V26.0. Statistical package deal for social sciences is used to calculate the statistical calculations inclusive of mean, preferred deviation, and additionally to devise the graphs etc(Wang et al. 2015). The impartial variables are Url, Label and the based variable is

'accuracy'. In SPSS, the dataset is ready through the use of 10 as pattern length for every organization and accuracy is given because of the trying out variable.

RESULTS

Table 1 indicates accuracy samples For Decision Tree and Random Forest Algorithm and in comparison each of the algorithms With accuracy.
Table 2 indicates the results achieved with p=0.02 shows that two groups are statistically significant.
Table 3 and Table 4 represents Independent pattern check for importance and preferred blunders determination. For each proposed and present algorithm 10 iterations have been taken for every new release the expected accuracy became referred to for analyzing accuracy. The outcomes of statistical applications of social sciences (IBM-SPSS v21) used for statistics analysis.
Table 5 represents the results achieved with p=0.02 (p>0.05) shows that two groups are statistically significant. Confidence c program language period of the distinction as decrease and top values range.
Figure 1 indicates the bar graph(fig. 1) is plotted with the aid of deciding on suggested accuracy on Y-axis and the Group on X-axis. From the graph, it's miles clean that Random Forest has drastically better accuracy 85% than Decision Tree 78%. The blunders bars are proven withinside the graph and the mistake charge is much less for Random Forest Compared to Decision Tree. From Fig. 1 it was represented by a simple bar Mean of Accuracy Decision Tree error range (0.89 - 0.82) and Random Forest error range (0.89 - 0.82).

DISCUSSION

In this research work, Random Forest has higher Accuracy than Decision Tree algorithm used Random Forest algorithm and Decision Tree methods to validate the feature selection method. From the experimental results performed in Jupyter, the accuracy of Random Forest is 85.80%, while Decision Tree provides the accuracy of 78.80%. This shows that Novel tree specific Random Forest is better than the Decision Tree. According to the SPSS plot, the proposed Random Forest classifier performs better in terms of accuracy(85.80%) than the Decision Tree algorithm.

To support this research work the studies are better in both experimental and statistical analysis(AlMamlook et al. 2019). Accuracy assessment cannot provide a better result on larger data sets. In addition, in Random Forest, the average error seems to be higher than Decision Tree. To oppose this work it would be preferable if the average error could be considerably reduced(Bull, United Nations. Economic Commission for Latin America and the Caribbean, and Deutsche Gesellschaft für Technische Zusammenarbeit 2003). However, the work can be improved by applying optimization algorithm techniques, to achieve lower mean error(Alkheder, Taamneh, and Taamneh 2017). Feature selection algorithms can be used prior to classification to improve the classification accuracy of classifiers. Therefore, thanks to the data mining algorithms, the computation time can be reduced and the accuracy of the classification of classifiers can be improved.

The limitation of the Random Forest is high-dimensional space and has a low convergence rate in the iterative process(Gómez-García et al. 2022). In future, the work can be enhanced by applying the optimal networks technique with the least mean error with good accuracy. Hence, the selection algorithms used to reduce the computation time and improve the better accuracy.

CONCLUSION

In this research work to find better accuracy we use two algorithms in predicting road accident severity, Random Forest and Decision Tree are used. Random Forest algorithm performed better accuracy than Decision Tree.

DECLARATIONS

Conflicts of Interests
The authors do not have any conflict of interest associated with this manuscript.

Authors Contributions
Author B.C concerned in statistics collection, statistics analysis, manuscript, writing. Author A.S concerned in conceptualization, statistics validation, crucial overview of manuscript

Acknowledgement
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 work successfully.

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

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

REFERENCES

Alkheder, Sharaf, Madhar Taamneh, and Salah Taamneh. 2017. "Severity Prediction of Traffic Accident Using an Artificial Neural Network." *Journal of Forecasting*. https://doi.org/10.1002/for.2425.

Alkhudhairy, Miaad K., Ali A. Aldhalemi, and Morteza Saki. 2022. "Analysis of Deaths and Injuries Related to Road Traffic Crashes in Al-Najaf Governorate of Iraq from 2017 to 2019." *The Nigerian Postgraduate Medical Journal* 29 (1): 56–62.

AlMamlook, Rabia Emhamed, Keneth Morgan Kwayu, Maha Reda Alkasisbeh, and Abdulbaset Ali Frefer. 2019. "Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity." *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*. https://doi.org/10.1109/jeeit.2019.8717393.

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.

Bojino, Alessandro, Fabio Roccia, Kirsten Carlaw, Peter Aquilina, Euan Rae, Sean Laverick, Irene Romeo, et al. 2022. "A Multicentric Prospective Analysis of Maxillofacial Trauma in the Elderly Population." *Dental Traumatology: Official Publication of International Association for Dental Traumatology*, February. https://doi.org/10.1111/edt.12736.

Bull, Alberto, United Nations. Economic Commission for Latin America and the Caribbean, and Deutsche Gesellschaft für Technische Zusammenarbeit. 2003. *Traffic Congestion: The Problem and How to Deal with It*. Santiago, Chile : United Nations, Economic Commission for Latin America and the Caribbean.

Darçın, Murat. 2020. *Accident Analysis and Prevention*. BoD – Books on Demand.

Ghandour, Ali J., Huda Hammoud, and Samar Al-Hajj. 2020. "Analyzing Factors Associated

with Fatal Road Crashes: A Machine Learning Approach." *International Journal of Environmental Research and Public Health* 17 (11). https://doi.org/10.3390/ijerph17114111.

GharehGozlu, Hanieh. 2021. "Analysis of Road Accident Data and Determining Affecting Factors by Using Regression Models and Decision Tree." *International Journal of Business Intelligence and Data Mining*. https://doi.org/10.1504/ijbidm.2021.10024451.

Gómez-García, Lourdes, Elisa Hidalgo-Solórzano, Ricardo Pérez-Núñez, Vanessa F. Jacobo-Zepeda, Ricardo G. Ascencio-Tene, Jeffrey C. Lunnen, and Amber Mehmood. 2022. "Factors Associated with the Severity of Road Traffic Injuries from Emergency Department Based Surveillance System in Two Mexican Cities." *BMC Emergency Medicine* 22 (1): 20.

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.

Jiang, Hui, and Hongxing Deng. 2020. "Traffic Incident Detection Method Based on Factor Analysis and Weighted Random Forest." *IEEE Access*. https://doi.org/10.1109/access.2020.3023961.

Kwayu, Keneth Morgan, Valerian Kwigizile, Kevin Lee, and Jun-Seok Oh. 2021. "Discovering Latent Themes in Traffic Fatal Crash Narratives Using Text Mining Analytics and Network Topology." *Accident; Analysis and Prevention* 150 (February): 105899.

Maulida, Nabila Rahmi, and Kusprasapta Mutijarsa. 2021. "Traffic Density Classification Using Multilayer Perceptron and Random Forest Method." *2021 International Seminar on Intelligent Technology and Its Applications (ISITIA)*. https://doi.org/10.1109/isitia52817.2021.9502269.

Meena, Krishan Kumar. 2017. "Road Traffic Accidents and Road Safety Behavior in India." *Indian Journal of Public Health Research & Development*. https://doi.org/10.5958/0976-5506.2017.00028.6.

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.

Panday, Pushpender, Aman Vikram, Anoop Chawla, and Sudipto Mukherjee. 2021. "Prediction of Lower Extremity Injuries in Car-Pedestrian Crashes - Real-World Accident Study." *Traffic Injury Prevention* 22 (2): 173–76.

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.

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.

Schlögl, Matthias, and Rainer Stütz. 2019. "Methodological Considerations with Data Uncertainty in Road Safety Analysis." *Accident Analysis & Prevention*. https://doi.org/10.1016/j.aap.2017.02.001.

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.

Taamneh, Salah, and Madhar M. Taamneh. 2021. "A Machine Learning Approach for Building an Adaptive, Real-Time Decision Support System for Emergency Response

to Road Traffic Injuries." *International Journal of Injury Control and Safety Promotion* 28 (2): 222–32.
Tiwari, Prayag. 2018. *Accident Analysis by Using Data Mining Techniques*. GRIN Verlag.
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.
Wang, Jianqiang, Yang Zheng, Xiaofei Li, Chenfei Yu, Kenji Kodaka, and Keqiang Li. 2015. "Driving Risk Assessment Using near-Crash Database through Data Mining of Tree-Based Model." *Accident Analysis & Prevention*. https://doi.org/10.1016/j.aap.2015.07.007.

TABLES AND FIGURES

Table 1. Comparison of Prediction Accuracy between Random forest and Decision Tree Algorithm. The Random Forest attained accuracy 85% compared to Decision tree accuracy 79 %.

Execution	Random forest Algorithm	Decision Tree Algorithm
1	89.23	82.12
2	88.56	81.37
3	84.80	79.89
4	83.96	78.72
5	85.24	79.45
6	86.64.	76.67
7	87.76	79.68
8	85.12	81.32
9	86.67	78.21
10	85.94	79.63

Table 2. Group statistics results(mean of Random Forest is 85 is more compared to Decision Tree and Standard error mean for Random forest is .374 and Decision tree is .270)

ALGORITHM	N	MEAN	Std.Deviation	Std.Error Mean
Random Forest	10	85.80	.837	.270
Decision Tree	10	78.80	.837	.374

Table 3. The results achieved with p=0.02 (p>0.05) shows that two groups are statistically significant.

	Levene's Test for Equality of	T-test for Equality of Means

		Variance								
		f	Sig	t	df	Sig.(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence of the Differences	
									Lower	Upper
ACCURACY	Equal variances assumed	.000	0.02	13.229	8	.002	7.000	.529	5.780	8.220
	Equal variances not assumed			13.229	8.000	.002	7.000	.529	5.780	8.220

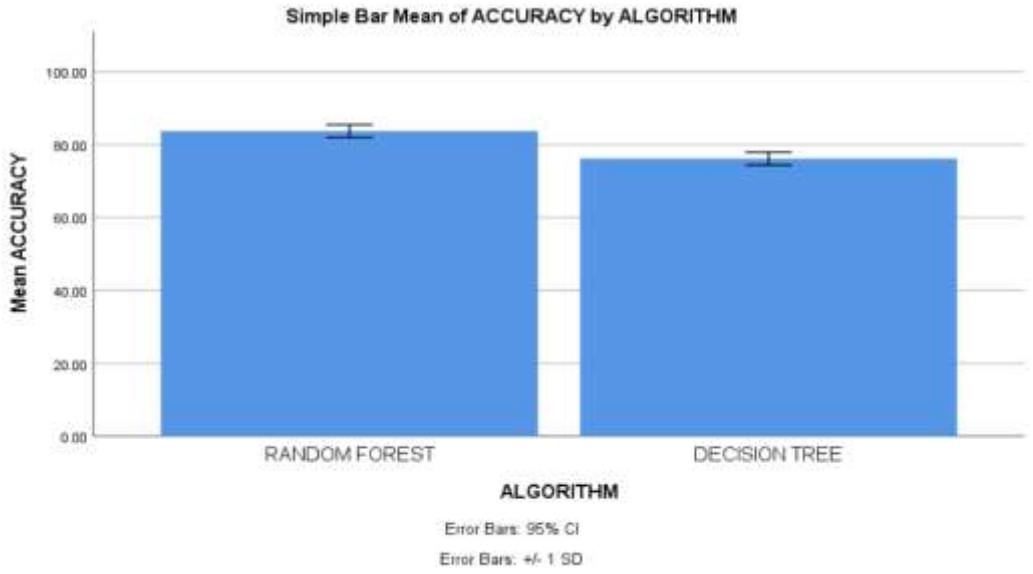

Fig. 1. Bar chart representing the comparison of mean accuracy af Random Forest Algorithm And Decision Tree Algorithm . Mean accuracy of Random Forest Algorithm is 85.80% appears to be better than Decision Tree which is 78.80%. The x-axis represents Random Forest and Decision tree and Y-axis Represents the mean Accuracy ± 1SD.