EasyChair Preprint

№ 7678

Analysis and Performance Comparison of Various Machine Learning Based Algorithms

Vippathi Mani Mounica, Lanke Lahari, Sambhanu Satya Shivani and M.Siva Ganga Prasad

March 29, 2022

Analysis and Performance Comparison of Various Machine Learning Based Algorithms

Mani Mounica , Lanke Lahari, S. Satya Shivani, M. Siva Ganga Prasad Dept. of ECM
Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP India 522503

Abstract- **In this paper, we analyze and compare the performance of machine learning based algorithms like K-Nearest Neighbour, Random Forest, Logistic Regression and Decision Tree. These analyses use data from the dataset which contains the data of accidents severity which is collected from various sources and made into a single dataset. We explored the utilization of single and various algorithms for expectation and utilized four different machine learning approaches with both exactness and execution time execution utilized for the examinations. The outcomes showed most reliable outcomes and that the Random Forest approach was more exact over all blends of input data from the dataset.**

Keywords: **Accident Prediction, classification Techniques, Machine Learning**.

I. INTRODUCTION

The continuous event of car crashes straightforwardly compromises human existence and property wellbeing. The major study we can observe these days is Road Safety and accident prediction. The event of street auto collisions is fundamentally impacted by mathematical attributes of street, traffic stream, qualities of drivers and climate of street. Many examinations have been led to anticipate mishap frequencies and analyzes the characteristics of road accidents, including studies on the identification of dangerous areas analysis of the severity of accidents and analysis of the duration of accidents. Some studies have taken focus on the mechanism of accidents. No particular methodology accessible for the traffic police to predict which area is accident prone at a specific time. The car crash expectation is assuming a huge part in management of the traffic. The standard method of linear analysis cannot reflect the true scenario since the noise pollution and data amount are insufficient, resulting in unsatisfactory prediction results.

II. LITERATURE SURVEY

The profound learning as another AI technique was exceptionally worried by specialists and money managers. The profound learning hypothesis clarifies text, images and sounds which are typically used in the field of text, image and speech recognition, and the innovation of neural organization as an exceptionally productive deep learning method have been widely used in the anticipation of road accidents. Unlike the usual learning structure, deep learning could demonstrate complex non-direct particularities using the representation of conveyed and progressive components. Some of the ideas and works that got my attention for my idea are:

- For Lu Wenqi foreseeing the mishap seriousness by utilizing convolution neural Network.

- According to Fu Huilin, Zhou Yucai, traditional method of direct examinations cannot uncover the truly circumstance the consequence of expectation isn't agreeable. Contrasts customary BP organization and its proposed arrangement.

- By Rafael G. Mantovan, Ricardo Cerri,Joaquin Vanschoren, objective of delicate choice trees are improving hyper parameters. Four different optimization methods were studied.

There are different Machine learning algorithms and deep learning algorithms which can be achieve the idea into the practical way. Each algorithm has its significance, but a complete comparative study is required to conclude the best algorithm in the pool. Considering the factors and the type of dataset used can determine the conditions applied for the model testing. The primary testing issue is regarding how to learn strong and discriminative spatiotemporal component portrayals. Since few preparation tests of

auto collisions can be gathered, inadequate coding methods can be utilized for little information case. Nonetheless, most inadequate coding calculations which use standard regularization may not accomplish sufficient sparsity. The second difficult issue is about the example irregularity between car crashes and ordinary traffic to such an extent that identifier might want to lean toward typical traffic. The W-ELM can place more spotlight on car crash tests. Trial results in our gathered dataset have shown that this proposed auto collision identification calculation beats other best in class techniques as far as the component's sparsity and discovery execution.

II. THEORETICAL ANALYSIS

A. K-Nearest Neighbour

K-Nearest Neighbor is one of the top Machine Learning calculations fundamentally dependent on Supervised Learning strategy. The KNN calculation provides for the comparability between the new cases/measurements as soon as they are ready and the available cases and places the new case in the classification most comparable to the classes to be had. The KNN calculation stores all accessible information and characterizes another point of information dependent on comparability. This implies that when new information is displayed, it tends to be grouped into a well-suite class using the KNN calculation**.**

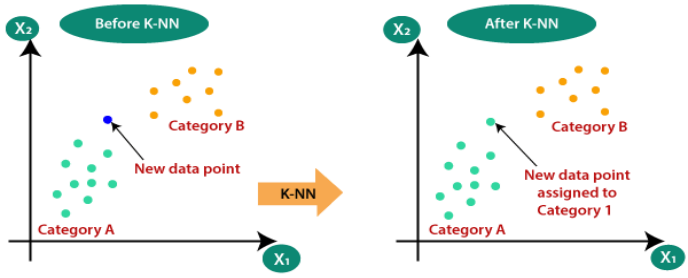


Fig 1: Before KNN vs After KNN

B. Random Forest

A random forest is a technique this is applied to take care of relapse and arrangement issues. It fabricates decision trees on various examples and takes their larger part vote in favor of arrangement and normal in the event of relapse. A random forest calculation comprises of numerous decision trees.[7]

Random Forest enjoys numerous upper hands over various classifiers. This calculation is truly steady. Regardless of whether a new information point is presented in the dataset the by and large calculation isn't impacted much since new information may sway one tree, however, it is exceedingly unlikely that it will impact all the trees. When there are both category and numerical characteristics, the Random Forest computation performs wonderfully.

C. Logistic Regression

Logistic regression, like linear regression, uses a condition as the representation. To predict an output value, input esteems (x) are linked directly using loads or coefficient esteems (y).

Logistic regression is basically a regulated grouping calculation. In a characterization issue, the goal variable (or result), y, can take simply disqualifies for a given association of capabilities X.[8].

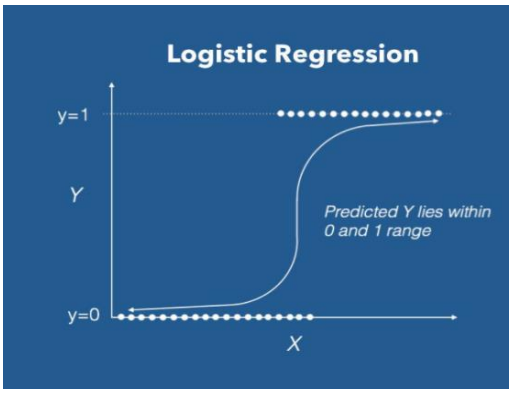


Fig 2: Logistic Regression

D. Decision Tree

Decision Trees are a type of Supervised Machine Learning. The goal of this calculation is to make a model that predicts the really well worth of an goal variable where internal nodes address the elements of a dataset, the decision rules are addressed by the branches, and the result is addressed by each leaf hub.[9]

In a Decision tree, the tree can be clarified by two substances, in particular decision hubs and leaves. Furthermore, the decision hubs are the place where the information is parted. The most recent improvements in the sign handling have protracted the traditional

strategies for the wavelet change to the areas that are with anomalies, that is the diagram.[12]

IV. PROCESS OF EXECUTION

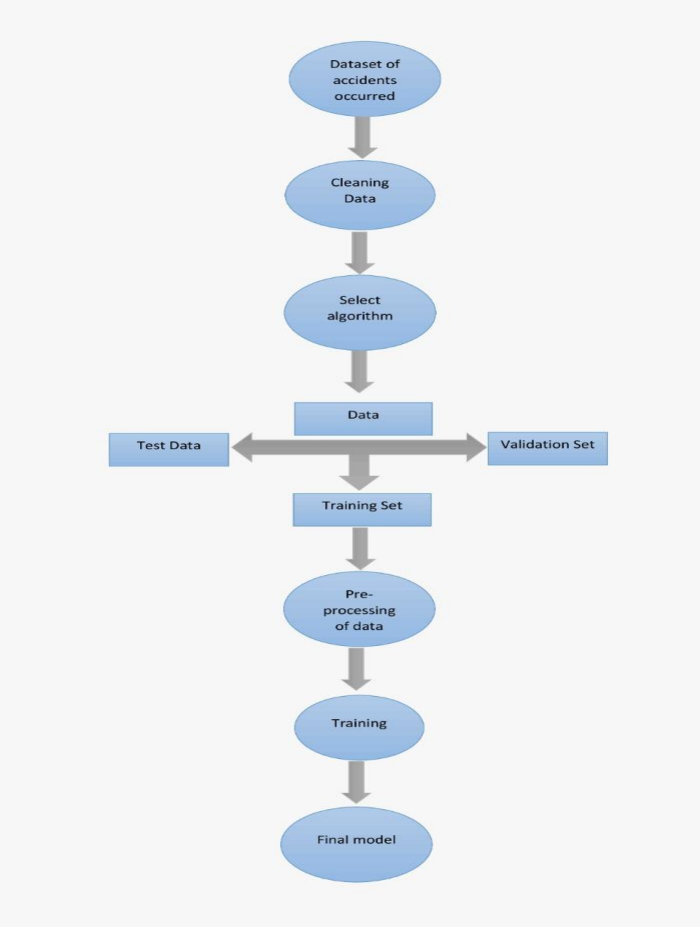
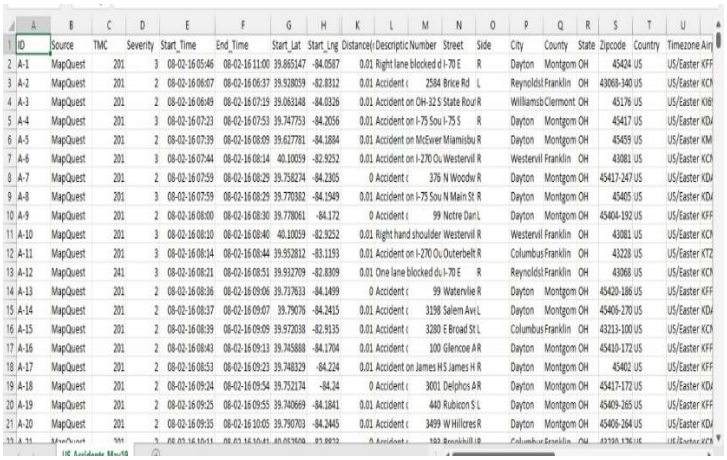


Fig 3: Flowchart of proposed system

Four distinct machine learning models were used in this work for categorization logic. The models were created using supervised learning approaches. Collisions are categorised. Depending on the severity of the injury, several degrees are possible. The accident data has been gathered and analysed. The data has been analysed, aggregated, and classified. Using the most appropriate algorithm depending on several factors. This may be used to analyse and locate issues, as well as the underlying causes of such difficulties. For the blunders, as well as tips on how to prevent them in the future. In our review, crash harm statistics has been gotten from 49 territories of US and 27 provinces of India.

DATA SET:



ID	Source	TMC	Severity	Start_Time	End_Time	Start_Lat	Start_Lng	Distance(i	Descriptic	Number	Street	Side	City	County	State	Zipcode	Country	Timezone Air
A-1	MapQuest	201	3	08-02-16 05:46	08-02-16 11:00	39.865147	-84.0587	0.01	Right lane blocked d	I-70 E		R	Dayton	Montgom	OH	45424	US	US/Easter KFF
A-2	MapQuest	201	2	08-02-16 06:07	08-02-16 06:37	39.928059	-82.8312	0.01	Accident (	2584	Brice Rd	L	Reynoldsl	Franklin	OH	43068-340	US	US/Easter KCH
A-3	MapQuest	201	2	08-02-16 06:49	08-02-16 07:19	39.063148	-84.0326	0.01	Accident on OH-32 S	State Rou		R	Williamsb	Clermont	OH	45176	US	US/Easter KYH
A-4	MapQuest	201	3	08-02-16 07:23	08-02-16 07:53	39.747753	-84.2056	0.01	Accident on I-75 Sou	I-75 S		R	Dayton	Montgom	OH	45417	US	US/Easter KDA
A-5	MapQuest	201	2	08-02-16 07:39	08-02-16 08:09	39.627781	-84.1884	0.01	Accident on McEwer	Miamisbu		R	Dayton	Montgom	OH	45459	US	US/Easter KMI
A-6	MapQuest	201	3	08-02-16 07:44	08-02-16 08:14	40.10059	-82.9252	0.01	Accident on I-270 Ou	Westervil		R	Westervil	Franklin	OH	43081	US	US/Easter KCH
A-7	MapQuest	201	2	08-02-16 07:59	08-02-16 08:29	39.758274	-84.2305	0	Accident (	376	N Woodw	R	Dayton	Montgom	OH	45417-247	US	US/Easter KDA
A-8	MapQuest	201	3	08-02-16 07:59	08-02-16 08:29	39.770382	-84.1949	0.01	Accident on I-75 Sou	N Main St		R	Dayton	Montgom	OH	45405	US	US/Easter KDA
A-9	MapQuest	201	2	08-02-16 08:00	08-02-16 08:30	39.778061	-84.172	0	Accident (	99	Notre Dan	L	Dayton	Montgom	OH	45404-192	US	US/Easter KFF
A-10	MapQuest	201	3	08-02-16 08:10	08-02-16 08:40	40.10059	-82.9252	0.01	Right hand shoulder	Westervil		R	Westervil	Franklin	OH	43081	US	US/Easter KCH
A-11	MapQuest	201	3	08-02-16 08:14	08-02-16 08:44	39.952812	-83.1193	0.01	Accident on I-270 Ou	Outerbelt		R	Columbus	Franklin	OH	43228	US	US/Easter KTZ
A-12	MapQuest	241	3	08-02-16 08:21	08-02-16 08:51	39.932709	-82.8309	0.01	One lane blocked du	I-70 E		R	Reynoldsl	Franklin	OH	43068	US	US/Easter KCH
A-13	MapQuest	201	2	08-02-16 08:36	08-02-16 09:06	39.737633	-84.1499	0	Accident (	99	Watervlie	R	Dayton	Montgom	OH	45420-186	US	US/Easter KFF
A-14	MapQuest	201	2	08-02-16 08:37	08-02-16 09:07	39.79076	-84.2415	0.01	Accident (	3198	Salem Av	L	Dayton	Montgom	OH	45406-270	US	US/Easter KDA
A-15	MapQuest	201	2	08-02-16 08:39	08-02-16 09:09	39.972038	-82.9135	0.01	Accident (	3280	E Broad St	L	Columbus	Franklin	OH	43213-100	US	US/Easter KCH
A-16	MapQuest	201	2	08-02-16 08:43	08-02-16 09:13	39.745888	-84.1704	0.01	Accident (	100	Glencoe A	R	Dayton	Montgom	OH	45410-172	US	US/Easter KFF
A-17	MapQuest	201	2	08-02-16 08:53	08-02-16 09:23	39.748329	-84.224	0.01	Accident on James H	S James H		R	Dayton	Montgom	OH	45402	US	US/Easter KFF
A-18	MapQuest	201	2	08-02-16 09:24	08-02-16 09:54	39.752174	-84.24	0	Accident (	3001	Delphos A	R	Dayton	Montgom	OH	45417-172	US	US/Easter KDA
A-19	MapQuest	201	2	08-02-16 09:25	08-02-16 09:55	39.740669	-84.1841	0.01	Accident (	440	Rubicon S	L	Dayton	Montgom	OH	45409-265	US	US/Easter KFF
A-20	MapQuest	201	2	08-02-16 09:35	08-02-16 10:05	39.790703	-84.2445	0.01	Accident (	3499	W Hillcres	R	Dayton	Montgom	OH	45406-264	US	US/Easter KDA

Fig 5: Collected accident record for US

Long time crash dataset turned into accrued, a sum of 1155332 and 28533 injuries had been utilized to study the review. Various ascribes has been gathered for every dataset like climate, avenue condition, region type, No. Of vehicles involved walkers, signal boards and so on.

V. RESULTS AND DISCUSSION

Expecting Mishap Seriousness with various AI Calculations. In this survey, four plan artificial intelligence computations were evaluated. These are Logistic Regression, K Nearest Neighbors (KNN), Random Forest and Decision Trees. As you are able to find in Figure 15, Random Forest computation is reliably the victor and KNN is the continue onward on the overview in states of accuracy. The accuracy is extremely high for Montgomery Area in PA.

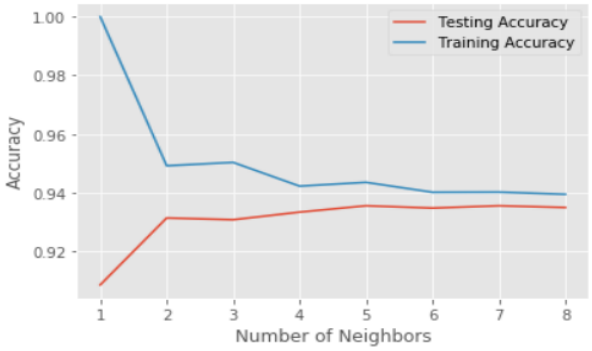


Fig 4: KNN varying number of neighbors

Plotting the graph between accuracy score versus algorithm
Comparing the accuracy scores, we get:

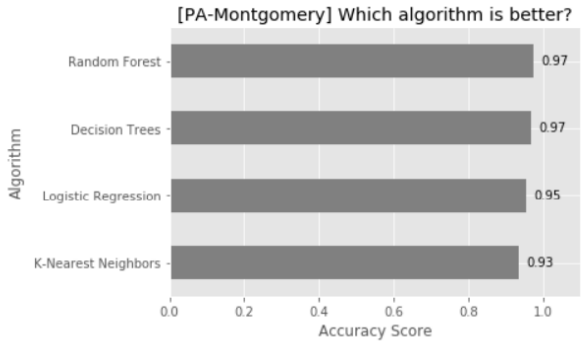


Fig 5: Performance comparison of RF, LR, DT, KNN

VI. CONCLUSION

This paper introduced an audit of calculations and models used to dissect, describe, and estimate street mishaps. The calculations and models involve information mining and AI strategies, introduced in the writing from the year2015, for including arising research draws near. Inside the strategies and calculations surveyed, neural organizations can be featured for their exactness to order information, while profound learning techniques are a generally original way to deal with anticipate streetcar crashes with high accuracy and fusing the utilization of the combination of a few information sources, for example, auto collision reports, land use, street foundation, climate conditions and segment data. It very well may be expressed that the best outcomes are acquired when at least two logical procedures are consolidated, so that investigation of the got results is reinforced. For example, the utilization of a neural organization joined with a standard framework (Castro and Kim, 2016) or hereditary calculation matched with a standard framework and choice tree shows improvement in the forecast dependability and accuracy of the outcomes, when contrasting and the outcomes acquired by the execution of a solitary procedure. Among the future difficulties in street traffic anticipating lies the upgrade of the extent of the proposed models and expectations by the consolidation of heterogeneous information sources, that including spatial information, data from traffic volume, traffic insights, video, sound and message and opinion from web-based media, that many creators agree that can work on the accuracy and precision of the examination and forecasts. Hardships in road setbacks are agonizing, to the public similarly as a non-modern country like us. Thus, it has transformed into an essential need to control and coordinate traffic with a significant level system to reduce the amount of road disasters in our country. By playing it safe, considering forecast or admonitions of a refined framework might forestall car crashes. In addition, it's an essential requirement for our nation now, to handle what is going on where consistently such endless people were killed in a fender bender and bit by bit this rate is getting extended. The execution of man-made intelligence is a utilitarian and a remarkable method for managing take an exact decision with the experience to manage the current situation and the disclosures of the assessment part can be proposed to traffic specialists for diminishing the number of disasters. We can use proposed ways of managing complete computer-based intelligence here because of their exhibited and higher precision to predict car crash earnestness. In addition, to make it more practical, we will attempt to make a recommender framework by utilizing these methodologies that can give an expectation to the car crash and can caution the street client. Later, it will be our attempt to make a portable application by executing this strategy to give a precise expectation to the client and make it extremely helpful and gainful too.

REFERENCES

[1] Lu Wenqi, Luo Dongyu & Yan Menghua, “A Model of Traffic Accident Prediction” INSPEC Accession Number: 17239218 DOI: 10.1109/ICITE.2017.8056908

[2] Thineswaran Gunasegaran Yu-N Cheah, “Evolutionary Cross validation” INSPEC Accession Number: 17285520 DOI: 10.1109/ICITECH.2017.8079960

[3] Simon Bernard, Laurent Heutte and Sebastien Adam, “On the Selection of Decision Trees in Random Forests” INSPEC Accession Number: 10802866 DOI: 10.1109/IJCNN.2009.5178693

[4] Rafael G.Mantovan,, Ricardo Cerri, Joaquin Vanschoren, “Hyper-parameter Tuning of a Decision Tree Induction Algorithm” INSPEC Accession Number: 16651860 DOI: 10.1109/bracis.2016.018

[5] Fu Huilin, Zhou Yucai, “The Traffic Accident Prediction Based on Neural Network”, 2011

[6] Lin, L., Wang, Q., Sadek, A.W., 2014. Data mining and complex networks algorithms for traffic accident analysis. In: Transportation Research Board 93rd Annual Meeting (No. 14- 4172).

[7] Gunasegaran, T., & Cheah, Y.-N. (2017). Evolutionary cross validation. 2017 8th International Conference on Information Technology (ICIT). doi:10.1109/icitech.2017.8079960

[8] Bernard, S., Heutte, L., & Adam, S. (2009). On the selection of decision trees in Random Forests. 2009 International Joint Conference on Neural Networks. doi:10.1109/ijcnn.2009.5178693

[9] Matthew Bihis ; Sohini Roychowdhury, “A generalized flow for multi-class and binary classification tasks: An Azure ML approach”

[10] Abdel-Aty, M., N. Uddin, and A. Pande. Split Models for Predicting Multivehicle Collisions during High-Speed and Low-Speed Operating Conditions on Freeways. In Transportation Research Record: Journal of the Transportation Research Board, No. 1908,