Proceedings of the International Conference on Edge Computing and Applications (ICECAA 2022)
IEEE Xplore Part Number: CFP22BV8-ART; ISBN: 978-1-6654-8232-5

2022 International Conference on Edge Computing and Applications (ICECAA) | 978-1-6654-8232-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICECAA55415.2022.9936227

Comparative Analysis using K - Nearest Neighbour with Artificial Neural Network to Improve Accuracy for Predicting Road Accidents

T. Deva Prakash
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.

thalladevaprakash17@saveetha.com

Nagaraju V
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.

nagarajuv.sse@saveetha.com

Abstract - The purpose of this study is to use machine learning approaches to improve the accuracy of modern road accident prediction systems like the K-Nearest Neighbour Algorithm and Artificial Neural Networks Algorithm. Materials and techniques used include the K-Nearest Neighbour technique and the Artificial Neural Networks algorithm with sample size N=10, iterated 20 times in parallel to test the accuracy of forecasting road accidents. p0.05 indicates the significance of the K-Nearest Neighbour method. When comparing the results of the two algorithms, it is discovered that the K-Nearest Neighbour approach (81.22%) outperforms the Artificial Neural Networks algorithm (69.22%) in terms of accuracy in forecasting road accidents.

Keywords - K-Nearest neighbour algorithm, Artificial neural network algorithm, Novel Prediction, Road accidents, Machine Learning.

I. INTRODUCTION

Road traffic accidents are shaking the world, killing thousands of people and destroying property on a daily basis without discrimination, but they haven't focused on reducing the severity. In any case, it is one of the dangerous episodes on the planet because of death and property harm. Recognizing the essential street car crash elements will assist with giving a suitable answer to limit the unfavourable impact of seriousness on human and property loss. [1]. Because of the fast advancement of finance and innovation, the versatility of city occupants has expanded step by step in metropolitan regions, which likewise brought amazing difficulties like gridlock and caused extra long stretches of sitting in gridlocks. At the point when traffic requests surpass the street limit, blockage happens. It has been demonstrated to be testing for thruway offices to react with the current circumstance of clog [2]. A number of circumstances can benefit from this research, including improving road infrastructure for safety management [3], road authorities [4], road designers, and road safety professionals [4].

Road accident prediction has been the subject of 43 papers in Science Direct and 132 articles in Google Scholar. Learning begins with perception or information, and AI is an application that offers frameworks the capacity to take in and improve a fact without being communicated. Machine learning based strategies utilized in rush hour gridlock expectation by catching the examples or connections between information series without demonstrating the actual traffic cycle, for example, Artificial Neural Networks (ANN) and K-Nearest neighbor (KNN). Among different variables, climate conditions enormously affect [5] traffic stream attributes and driving practices. Transcendent variable classes are influenced by nasty climate for example, traffic interest, traffic wellbeing etc. KNN approach has been applied for determining traffic. A few analysts have applied KNN to figure traffic stream rates [6]. In addition, the investigations were led under various settings, for example, traffic boundaries considered, the nature of datasets and their sources, collection intervals, forecast step durations, production horizons, [7] and performance measure considered. The traffic estimating techniques have been step by step moving from conventional factual models to computational knowledge approaches. [8] Computational knowledge approaches are regularly used to portray high dimensional and non-direct connections. As a result of the collaboration with a variety of writers from across

universities, the project was completed quickly and accurately [9 -13].

Existing systems' tactics and procedures have a low accuracy rate, low reliability, and aren't particularly convincing when it comes to predicting road accidents. The ANN's unique prediction stage is slow with huge data and sensitive to data scale and irrelevant features. In the proposed work, machine learning technologies such as the KNN algorithm and ANN are used to improve the accuracy of forecasting road accidents.

II. MATERIALS AND METHODS

Saveetha School of Engineering, and Saveetha Institute of Medical and Technical Sciences, are responsible for the survey in the proposed research. The quantity of gatherings utilized for the exploration is the latter. The former gathering is KNN calculation and latter gathering is ANN calculation. The absolute number of tests that are surveyed on the proposed strategy is 10 in each gathering to distinguish different scales in street mishaps in informal organizations. The example assessment is completed with 80% GPower. The alpha blunder rate is a value of 0.05 that shows the difference between two calculations. The enlistment rate is roughly 1%.

Road accidents dataset namely dft RoadSafety_Accidents is collected from UCI repository website. It contains a total of 900 rows of accident data. Prediction of Road accidents dataset contains 9 attributes like place, time, date, year, no. of vehicles, weather, way of road, country, and information to police. The dataset needs to be processed before applying it to the novel prediction machine learning model. In the data collection procedure, the various data are collected based on the different locations and stored in the CSV file for the further analysis and clustering of the data through the data mining process. The data processing includes removal of missing data and replacement of null values with mean or median values with standardization of data. The preprocessed data is given as an input for KNN and ANN. From the preprocessed data 70% is given as training data and 30% is given as testing data.

A. K-Nearest Neighbors Algorithm

The KNN algorithm is a form of supervised machine learning method that is used to find the closest neighbors. KNN is incredibly simple to implement in its most fundamental structure, but it performs exceedingly difficult classification jobs once it has been trained. It is referred to as a "lazy novel learning algorithm" because it does not have a specific training phase. In contrast, when classifying a new data point or instance, it uses the entire set of data for training purposes. It's worth noting that KNN is a non-parametric learning approach, which means it doesn't make any assumptions about the data. This is an important feature because a large portion of the information in today's world does not adhere to any theoretical assumptions, such as linear separability, uniform distribution, and so on.

Pseudocode for K-Nearest Neighbors Algorithm

Input: Dataset trained
Output: Acuracy
Import the required packages and dataset
Define x and y variables as movie and rating.
x_train, x_test, y_train, y_test equal to train_test_split (x,y, test_size = 0.1,random_state = 56)
from sklearn. neighbors import K-NeighborsClassifier
classifier = K-NeighborsClassifier ()
classifier.fit (x_train, y_train)
classifier. Score (x_test, y_test)
y_pred = classifier. predict (x_test)
score=accuracy_score (y_test, y_pred)
print score

B. Artificial Neural Networks Algorithm

Computer models based on human brains are known as ANNs. Vocal recognition, picture identification, and robotics using ANN are just a few of the recent developments in Artificial Intelligence that have been made.

Pseudocode for Artificial Neural Networks Algorithm

Input: Dataset trained
Output: Accuracy
Import the required packages and dataset
Define x and y variables as movie and rating.
x_train, x_test, y_train, y_test equal to train_test_split (x,y, test_ size=0.1,random_state = 56)
from sklearn. neighbors import Artificial Neural Network Classifier
classifier = A Neural Network Classifier()
classifier.fit (x_train, y_train)
classifier.score (x_test, y_test)
y_pred=classifier.predict (x_test)
score=accuracy_score (y_test, y_pred)
print score

Jupyter scratch pad is used to evaluate the KNN calculation and the ANN calculation. The framework is a 64-bit operating system with an x64 processor. Intel i5 processor with 4GB RAM is the required equipment. The operating system Windows is used, as well as the instrument Jupyter note pad with Python writing computer programmes is used to assess KNN calculation and ANN calculation. The framework type is a 64-bit working system, and x64 based processor. The equipment necessities are intel i5 processor with 4GB RAM.

The street accident dataset is dftRoadSafety_ Accidents gathered from UCI archive site. Cleaning the data, such as removing extraneous characteristics and connecting and reorganising them, is performed. The substance included in the dataset is revealed through information analysis. Data pre-processing is done after the dataset is cleansed by deleting invalid qualities or invalid qualities. Pre-processing of information is done. After information pre-processing, the dataset is secluded into two areas such as, testing set and training sets. 70% of the dataset is taken as a training test and 30% of the dataset is taken as a testing test. In the wake of separating the dataset, the calculation is fitted. By surveying the calculation with train and test sets, the necessary limit exactness rate is expected.

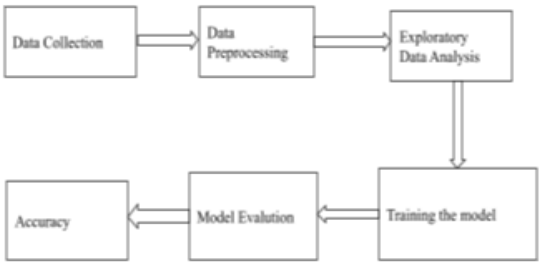

Fig. 1. System architecture

The steps included are pre-processing of data, splitting of training data and testing data, training data processed using machine classifiers, and finally predicting the outcome.

C. Statistical Analysis

IBM SPSS version 21 is used for the statistical analysis. After 20 iterations, the accuracy of the KNN algorithm and the ANN are evaluated using the SPSS programme. The mean, standard deviation, and standard deviation error for the two groups are included in the data. Each cycle's precision study yields the predicted precision. Using an independent sample t test, the value obtained by iteration is tested. The input features for the road accidents dataset, such as accident, location, latitude, and longitude, are the independent variables in this study. The output variables in this study are information to the police, the number of cars, the day, time, year, and the weather.

III. RESULTS

The road accidents assist users to make decisions in complex information areas where the data accessible is large. The accuracy rate of the road accident is predicted with the help of machine learning techniques. In this study, KNN and ANN can improve the enhancement of road accidents. Results show that KNN Algorithm predicts with greater accuracy (81.22%) than ANN accuracy (69.22%), and KNN Algorithm has got better significance than ANN. Table 2 represents the significant values of the accuracies that vary for different sample size.

TABLE 1. ACCURACIES OF THE KNN AND ANN

Sample	KNN Accuracy (%)	ANN Accuracy (%)
1	72.82	61.32
2	73.21	62.58
3	75.65	64.75
4	76.96	65.98
5	78.95	66.26
6	82.31	71.00
7	84.25	73.22
8	85.98	74.54
9	89.56	75.32
10	92.54	77.25

The mean and standard deviation for KNN are 81.2231 and 6.82929 respectively, and ANN algorithm has the mean of 69.2218 and standard deviation of 5.72435. The KNN approach has a standard error of 2.15961, whereas the ANN algorithm has a standard error of 1.81020. A p-value of 0.04 or lower indicates a significant difference.

TABLE 2. GROUP STATISTICS RESULTS

Algorithm	N	Mean	Std. Deviation	Std. Error Mean
K-Nearest Neighbors	10	81.2231	6.82929	2.15961
Artificial Neural Networks	10	69.2218	5.72435	1.81020

Table 3. has the comparison of accuracies of the novel road accidents prediction system using KNN algorithm and ANN algorithm with significance p<0.05. KNN proved with mean difference =12.00130, standard error difference =2.81793 that it has obtained better significance than ANN with value of p0.05 in independent sample test.

TABLE III. COMPARISON OF ACCURACIES

Accuracy	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	
								Lower	Upper
Equal Variances assumed	.292	.046	-4.01	18	.001	12.00130	2.81793	6.0810	17.9215
Equal variances not assumed			-4.01	9.213	.001	12.00130	2.81793	6.0680	17.9345

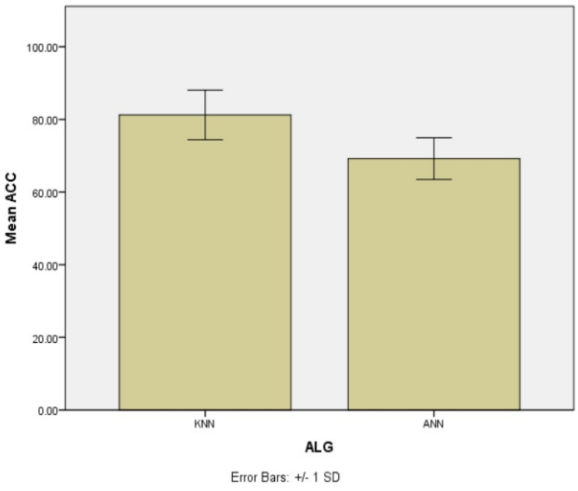

Fig. 2. Comparison graph (KNN and ANN) in terms of mean accuracy

Figure 2 depicts the visual diagram in relation to the mean accuracy calculated using KNN and ANN calculations.

The accuracy of ANN is 69.2218 and KNN is 81.2231. KNN computation has a 2.81793 error rate, while ANN has a 2.82486 error rate. The illustrations of the accuracy calculations are analysed using the t-test, and a truly critical contrast is seen at P 0.05. The KNN model has obtained 81.22% precision. When contrasted, the calculations of the proposed KNN has accomplished preferable execution over the ANN calculation. The reference chart addresses the correlation of mean accuracy of KNN' calculation with ANN calculation, wherein the KNN calculation has obtained more huge outcomes than the ANN calculation.

IV. DISCUSSION

The study was explored based on the use of data from the road accident by using SVM, K-Means Clustering, and ranking SVM which produced accuracies of 75%, 66% and 56% respectively. In the accuracy of the Indian Road Conditions and Accident Risk Predictions utilising a deep Learning technique domain developed in real-time was

90.27%. In Lasso and Linear Regression algorithms made predictions in real time accidents and obtained accuracies of 89.36% and 75.91% respectively. proposed a comparison between SVM, ANN (88%), logistic (63%) and Random Forest (76%). proposed Road Traffic Prediction using KNN and Optimized Multilayer Perceptron like logistic, ANN and CNN. There are no opposition findings in the study of predicting road accidents. It is shown from survey that the KNN algorithm has better accuracy when compared with ANN algorithm.

The limits of the proposed model is that the accuracy of this novel forecast system relies upon the nature of the information. It is sensitive to the size of the information and unimportant highlights. Moreover, it requires high memory to store all the preparation information. The equipment reliance, unexplained working of the network, confirmation of legitimate network structures, the trouble of showing the issue to the network, the length of the network, are obscure. Later on, the work can be upgraded by applying the ideal calculation strategies which diminishes memory capacity, with the most least mean error and with great precision.

V. CONCLUSION

The proposed novel prediction framework of road accidents in social networks has attained the accuracy of 81.22% by KNN algorithm which is higher when compared to that of the ANN technique that has an accuracy of 69.22%. The proposed framework proves that the KNN algorithm has better significant accuracy than the ANN algorithm.

REFERENCES

[1] S. S. Yassin and Pooja, "Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach," *SN Applied Sciences*, vol. 2, no. 9, pp. 1–13, Aug. 2020.

[2] "Predicting Road Accidents: A Rare-events Modeling Approach," *Transportation Research Procedia*, vol. 14, pp. 3399–3405, Jan. 2016.

[3] M. Antonio, "Live Prediction of Traffic Accident Risks Using Machine Learning and Google Maps," *Towards Data Science*, Oct. 14, 2019. https://towardsdatascience.com/live-prediction-of-traffic-accident-risks-using-machine-learning-and-google-maps-d2eeffb9389e (accessed Jul. 30, 2021).

[4] Hossein Anisi, "Evaluation of the Cyber Security Risk Models (CSRM) in Cloud Computing", Journal of Machines and Computing, vol.2, no.3, pp. 124-133, July 2022. doi: 10.53759/7669/jmc202202017.

[5] "Accident prediction models for urban roads," *Accid. Anal. Prev.*, vol. 35, no. 2, pp. 273–285, Mar. 2003.

[6] R. Jagannathan, S. Petrovic, G. Powell, and M. Roberts, "Predicting Road Accidents Based on Current and Historical Spatio-temporal Traffic Flow Data," in *Computational Logistics*, Sep. 2013, pp. 83–97.

[7] "Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt," *Accid. Anal. Prev.*, vol. 36, no. 2, pp. 149–163, Mar. 2004.

[8] "Predicting wet weather accidents," *Accid. Anal. Prev.*, vol. 13, no. 2, pp. 83–99, Jun. 1981.

[9] S. Gheena and D. Ezhilarasan, "Syringic acid triggers reactive oxygen species-mediated cytotoxicity in HepG2 cells," *Hum. Exp. Toxicol.*, vol. 38, no. 6, pp. 694–702, Jun. 2019.

[10] P. Neelakantan, D. Grotra, and S. Sharma, "Retreatability of 2 mineral trioxide aggregate-based root canal sealers: a cone-beam computed tomography analysis," *J. Endod.*, vol. 39, no. 7, pp. 893–896, Jul. 2013.

[11] M. C. Putchala, P. Ramani, H. J. Sherlin, P. Premkumar, and A. Natesan, "Ascorbic acid and its pro-oxidant activity as a therapy for tumours of oral cavity -- a systematic review," *Arch. Oral Biol.*, vol. 58, no. 6, pp. 563–574, Jun. 2013.

[12] P. U. A. Wahab, M. Madhulaxmi, P. Senthilnathan, M. R. Muthusekhar, Y. Vohra, and R. P. Abhinav, "Scalpel Versus Diathermy in Wound Healing After Mucosal Incisions: A Split-Mouth Study," *J. Oral Maxillofac. Surg.*, vol. 76, no. 6, pp. 1160–1164, Jun. 2018.

[13] S. I. DeSouza, M. R. Rashmi, A. P. Vasanthi, S. M. Joseph, and R. Rodrigues, "Mobile phones: the next step towards healthcare delivery in rural India?," *PLoS One*, vol. 9, no. 8, p. e104895, Aug. 2014.