2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS)

An Effective Prediction Modelling of Traffic Accidents based on Categorical Boosting with Grid Search

1st Peruri Sudheshna
School of Computer Science and Engineering
Malla Reddy Engineering College for Women (UGC -Autonomous)
Maisammaguda, Hyderabad, India
sudheshnamrecw@gmail.com

2nd Haassan M.M. AlJawahry
Department of Computers Techniques Engineering
College of Technical Engineering
The Islamic University
Najaf, Iraq
eng.iu.hassanaljawahry@gmail.com

3rd Rekha Phadke
Department of Electronics and Communication Engineering
Nitte Meenakshi Institute of Technology
Bengaluru, India
rekha.phadke@nmit.ac.in

4th S Sivasubramanian
Department of Information Technology
New Prince Shri Bhavani College of Engineering and Technology
Chennai, India
sivasubramaniyan.it@newprinceshribhavani.com

5th Tatiraju.V.Rajani Kanth
TVR Consulting Services Private Limited
Hyderabad, India
tvrajani55@gmail.com

2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS) | 979-8-3315-0496-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICIICS63763.2024.10859490

***Abstract*—Recent days, the alarming rise in road traffic accidents has devastating consequences globally, causing a significant surge in injuries and fatalities. This trend is impaired by projected doubling of vehicles worldwide by 2040 which is placed by unprecedented strain on transportation infrastructure. Traditional approaches for modelling traffic accidents had significant challenges which include difficulty in integrating diverse data sources, complexity and difficulty in understanding model decisions. Therefore, this research proposes Categorical Boosting (CatBoost) integrated with Grid Search for predictive modelling of traffic accidents. Initially, the data is taken from Road Traffic Accident (RTA) dataset which is taken from department of police, Addis Ababa Sub city. Then, this data is preprocessed by using Random Projection Paradigm (RPP) and Improved Linear Discriminant Analysis (ILDA) which effectively reduced dataset dimensions, simplified complex traffic data. After that, the model is selected by using CatBoost which effectively processes the categorical data and ensured robust modeling, the hyperparameters are tuned by using grid search optimization. Finally, the model is evaluated by using k-fold cross validation provided the robust evaluation and handled imbalanced data. The proposed CatBoost method established better results including accuracy (0.95), precision (0.78), recall (0.65), RMSE (0.02) when compared with existing Naïve Bayes.**

Keywords—categorical boosting, grid search, improved linear discriminant analysis, random projection paradigm, road traffic accident.

I. INTRODUCTION

Since days, the widespread adoption of Intelligent Transportation Systems (ITSs) had flashed significant interest in traffic flow prediction research over the past few decades [1]. By leveraging traffic flow prediction, transportation management authorities proactively monitored and regulated the road traffic conditions which enabled timely interventions to prevent congestion, optimized traffic signal control [2]. The research on traffic flow faced various challenges which included obtaining accurate and reliable data [3], as historical traffic data which openly contained errors, missing values and inconsistencies [4]. Moreover, to develop robust prediction models, complex interactions between various factors like weather, road conditions made the traffic accidents challenging [5]. Then, prediction accuracy is decreased by handling nonlinear relationships between the variables and traditional models struggled for capturing dynamic nature of traffic flow by resulting in inadequate forecasting [6]. Traffic accidents posed a significant threat to global health by resulting in millions of fatalities and injuries yearly [7]. To identify high risk areas, predictive modeling offered promising results which enabled the authorities by leveraging the advanced analytics and machine learning predictive modeling [8]. The predictive modeling aid in resource allocation, emergency, response planning and post-accident analysis which makes important tool in the pursuit of safer and more efficient transportation systems. The state of art methods includes Random Forest (RF) [9] which effectively handled high dimensional data, robust to noise, interpretable. However, the RF has risk in overfitting, and it is computationally intensive. Simultaneously, Gradient Boosting (GB) [10] effectively handled complex interactions, high accuracy. However, the GB is prone to overfitting, required hyperparameters tuning. Additionally, Neural Networks (NN) handled non-linear relationships, scalable. However, the NN is difficult to interpret.

The main contributions of this research includes:

- The data is preprocessed by using Random Projection Paradigm (RPP) and Improved Linear Discriminant Analysis (ILDA) which effectively reduced dataset dimensions, simplified complex traffic data.
- Data Transformation is done by utilizing One hot encoding which prevented from assuming false relationships between categories.
- The model is selected by using Categorical Boosting (CatBoost) which effectively processed categorical data, the hyperparameters are tuned by using grid search optimization which evaluated reduced the computation time.

This research is structured as: related work of traffic accidents is presented in section 2, proposed CatBoost is explained in section 3, proposed methodology is evaluated in section 4, results are discussed in section 5 and paper ends with conclusion which is present in section 6.

II. LITERATURE REVIEW

Tebogo Bokaba et al. [11] suggested a Random forest classifiers for modelling of road traffic accidents. Initially, the data is taken from Road Traffic Accident (RTA) dataset and this data is preprocessed with the help of Principal Component Analysis (PCA) that is utilized for reducing dimension. The traffic accidents are classified by using Random Forest which incorporated the model interpretability and had efficient processing flexibility. The RF handled high dimensional data, robustness to noise and outliers and has ability to capture nonlinear relationships. However, the RF had computational intensity, and difficulty in interpreting complex models.

Pablo Marcilo et al. [12] demonstrated a review based on the prediction of traffic accidents with heterogonous sources. Initially, data is taken from RTA dataset by taking the research questions, and search strategy is done by removing duplicates, exclusion and inclusion criteria in the methodology. Then, the snowballing technique is selected by the study and evaluated by using Support Vector Machine (SVM) and the quality assessment is done by conducting the review and this data is extracted by using articles and quality instrument. The results are evaluated by using reporting the review. The SVM has high accuracy and effective in high dimensional spaces. However, the SVM had computational intensity, and difficulty in interpreting complex models.

Mohammad Hesam Rashidi et al. [13] suggested traffic crash prediction model by modeling the accuracy. Initially, the data is collected from province crash occurrences for the past 7 years and then the winter is forecasted in the province crash prediction model. Then Root Mean Square Errors are collected from the predictions from the regresses and the indicators of traffic and macroeconomic provides each province and provided the accuracy analysis of multi-layer perceptron. The model enabled safety planners to take proactive measures, helps in identifying attributes by influencing the forecasting accuracy. However, increased root mean square error in provinces with higher variant monthly crashes.

Miaomiao Yan et al. [14] presented traffic accident severity prediction using Bayesian Optimization-Random Forest (BO-RF) used to tune the parameters of traffic accident. The suggested BO-RF employed for classification and regression in traffic accident to predict decision trees of road conditions. The traffic analysis of road accident with weather and time of optimal hyperparameters to maximize prediction of driver behavior using dataset. The robust predictive performance ensemble nature and noisy data by scalability for parallel evaluations. However, the computational complexity caused due to dimensionality with hyperparameters for non-stationary bounds.

Mayura Yeole et al. [15] demonstrated traffic accident on roads for the traffic flow by utilizing Artificial Neural Network (ANN). Initially, the secondary data is collected from existing records such as traffic records, road infrastructure. Then, accident patterns are identified by descriptive statistics and to analyze the relationship between variables. After that, the variables are used by traffic accidents are traffic capacity, road provisions, and has vehicle overload. The maximum accidents are occurred in weather conditions, overload vehicles had a significant cause in accidents. However, this ANN had data limitations and has geographical limitations, overfitting.

III. METHODOLOGY

This research proposes CatBoost with Grid Search for Predictive Modelling of Traffic Accidents which is employed on Road Traffic Accident (RTA) dataset. The data is preprocessed by using Random Projection Paradigm (RPP) and Improved Linear Discriminant Analysis (ILDA) which effectively reduced dataset dimensions, simplified complex traffic data. Then, data is transformed by using one hot encoding which prevented from assuming false relationships between categories.

After that, the model is selected by using Categorical Boosting (CatBoost) which effectively processes the categorical data and ensured robust modeling, the hyperparameters are tuned by using grid search optimization. Finally, the model is evaluated by using cross validation provided the robust evaluation and handled imbalanced data. The overall methodology is shown in Fig. 1.

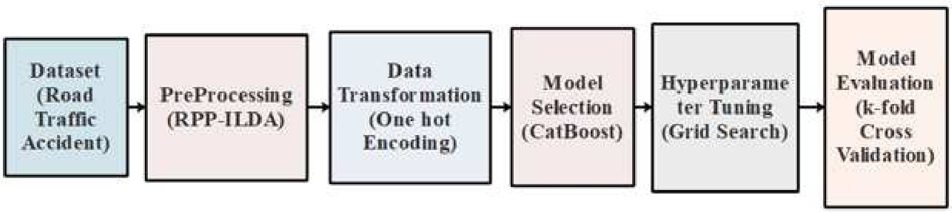

Fig. 1. Block diagram of Proposed CatBoost with Grid Search for traffic accidents predictive modeling

A. Dataset

In this research, CatBoost with Grid Search is proposed for Predictive Modelling of Traffic Accidents which is employed on Road Traffic Accident (RTA) dataset [16]. This RTA dataset is publicly available on kaggle. RTA dataset is illustrated with the road traffic accident records in the span of 2017 to 2020. All delicate data is accurately eliminated during the encoding process. The RTA dataset consists of 32 salient features and 12,316 accident instances. The dataset is optimized for analysis for classification algorithms is employed to identify primary accident causes. The original dataset 'RTA Dataset.csv', endured through cleaning, yielding the refined by 'cleaned.csv' dataset."

B. Preprocessing

The data preprocessed by using various techniques to deal with high dimensional data, Random projection reduces data dimensionality while preserving meaningful information and reduces noise in data by averaging out irrelevant features. ILDA handles high-dimensional data effectively.

1) Random Projection Paradigm (RPP): The Preprocessing technique is an important phase in traffic accident predictive modelling because of the nature of dealing with high dimensional data. The random projection paradigm method was introduced for feature extraction to diminish the high dimensionality of the data. By this feature extraction the high dimensionality of data is changed into low dimensional space without altering much data additionally, the technique attained the distance preserving property. In this process the Euclidean distance among data points in the equivalent lower dimension with superior possibility by using. Johnson–Lindenstrauss lemma equation as given in equation (1) and equation (2). This technique categorized into two points such as the amount of data points in high dimensional space and optional epsilon ϵ . The random projection generates the projection matrix on the basis of number of dimension earlier to the influx of input data. 0.1 is the fixed value of ϵ and it should be in the range of 0 to 1. The measurement for undefined alteration of the distance among correlated low dimensional space is ϵ and high dimensional space. Therefore, the ϵ value needs small amount of dimensions for presenting the data due to the approval of deviation among the distances.

$$A \geq \frac{24}{3e^2 - 2e^3} \log B \tag{1}$$

$$(1-\epsilon)\left\|a_i - a_j\right\|^2 \leq \left\|f(a_i) - f(a_j)\right\|^2 \leq (1+\epsilon)\left\|a_i - a_j\right\|^2 \tag{2}$$

Where B represents the amount of features in the high dimensional input data and N represents the samples in the high input data. Additionally ϵ is the Error progressive value ranges from 0 to 1 and A represents the dimensions need to show the data where $A < B$ and f projection function also $\left\|a_i - a_j\right\|^2$ represents distance between the ith and jth terms.

2) Improved Linear Discriminant Analysis: The ILDA is a dimensionality reduction technique employed in supervised machine learning to simplify preprocessing stage. This LDA transforms the high dimensional data into a lower dimensional space by increasing the class among minimizing and variance. LDA is used a primary preprocessing stage due to its effectiveness to transfer lower dimensional space features. The enhanced LDA involved in computing class means and prior probabilities by calculating the every individual group averages, determining scattering matrix using Eigen vectors and Eigen values, estimated the covariance matrix by computing covariance matrices for each group. The transformation matrix is constructed from eigenvectors are associated with the dominant eigenvalues of optimization criterion. These eigenvectors represent one-dimensional invariant subspaces within the vector space which undergoes the transformation. Specifically, eigenvectors with non-zero corresponding eigenvalues are linearly independent and invariant. Any vector space is expressed as linear combinations of these eigenvectors. By reducing dimensionality, LDA decreased computational complexity and filters out the irrelevant information, by improving the model accuracy. LDA enhanced the model performance by facilitating data visualization in lower dimensional space.

C. Data Transformation

The preprocessed data is transformed by using One-hot encoding process [17]. The One hot encoding is a valuable technique used to convert categorical features into numerical representations. This method extends the original feature vector into multidimensional matrix, where each dimension is corresponded to a specific state. Only one dimension is asserted for given state while others remained zero. For instance, consider a dataset describing driver's characteristics which includes gender, nationality, and grade. The one-hot encoding transforms these features into separate columns by eliminating the influence of differing digitized values on model training. This approach also effectively handled missing data by treating it as a new dimension, ensuring the datasets completeness. The one hot encoding effectively eliminated digitized value biases based on model training, handles missing data without simulation, treating missing values as a distinct class and preserves data structure integrity.

D. Model Selection by CatBoost

The CatBoost is a technique which combines "Categorical" and "Boosting" [18] techniques and widely used in python, by leveraging symmetric decision trees with minimal parameter. This supports class variables and ensures high accuracy for categorical feature processing. CatBoost offered several benefits which include efficient handling of categorical features, reduced overfitting through the gradient bias and prediction shift mitigation which supports for class variables. In decision trees, label means to determine the node splitting using greedy target variable statistics. A way of improving greedy TS is to concatenate distribution relations which reduces the noise effect and frequency data is evaluated as eq. (3), (4)

$$x_k^i = \frac{\sum_{j=1}^{p-1}[x_{j,k} = x_{i,k}].Y_i}{\sum_{j=1}^{n}[x_{j,k} = x_{i,k}]} \tag{3}$$

$$x_k^i = \frac{\sum_{j=1}^{p-1}[x_k = x_p].Y_j + a.p}{\sum_{j=1}^{p-1}[x_j = x_k] + a} \tag{4}$$

Here, added prior term is p, weight coefficient is $a.x$ is independent term, j is term of index, k is weight matrix. In decision trees, label means to determine the node splitting using greedy target variable and this approach is introduced by conditional bias when training and test datasets differ in the structure and distribution. The CatBoost approach incorporated priori distribution for minimizing the noise impact. CatBoost enhanced the model robustness by adding prior terms and weight coefficients. The CatBoost is suitable for categorical dataset by enabling effective learning and improved model expressiveness.

E. Hyperparameter tuning by Grid Search

By optimizing hyperparameters for traffic accidents, Grid Search (GS) [19] remained a widely employed method. GS explores the hyperparameter configuration space by resting all combinations within a predefined grid. This approach assess the product of user defined values. This GS required manual refinement to identify global optima to start with a broad search space and iteratively narrow it based on previous findings. The GS has high dimensional spaces due to exponential growth in each added hyperparameter and increases the dimensionality curse. GS is efficient for limited configuration spaces. In traffic accident prediction, efficient hyperparameter optimization is important. 'GridSearchCV' facilitated GS implementation. While grid search prevails due to its ease of implementation and effectiveness in low dimensional spaces, the hyperparameter tuning significantly impact the model prediction accuracy. The global optimization technique explores the alternative hyperparameter optimization enhanced the model performance and improved traffic safety.

F. Model Evaluation – K- fold Cross Validation

Evaluating the performance of traffic accident prediction model required robust methodologies. The k-fold cross-validation to optimize the weights for averaging predictions, which enhanced the accuracy in traffic accident risk assessment. By dividing the dataset into K groups, each serves as validation sample, the model predictive capabilities are evaluated. The k-fold cross validation process involved in five steps they are Firstly, divide the dataset into k groups of observation in each group with $2 \leq k \leq n$. Secondly, model estimators are calculated for each $k = 1, \ldots, k$ using remaining observations for each group. Thirdly, the predictions are calculated for all the observations in each model as in eq. (5)

$$Y_m = (Y_{m,1}^{-1}, \ldots, Y_{m,j}^{-1}, \ldots, Y_{m,1}^{-k}, \ldots, Y_{m,j}^{-K})^T \quad (5)$$

Where, Y is the average prediction, m is Euclidean norm, for the excluded observations. Then, k-fold cross validation is constructed by minimizing the k fold validation shown in eq. (6)

$$w = argminCV(w) \quad (6)$$

Where, w is the weight matrix, via quadratic programming. This approach ensured robust prediction accuracy by minimizing sum of squared prediction errors across all groups. The quadratic nature of k-cross validation facilitated the efficient numerical computation in the traffic accident modeling where, common choices for K included 5 and 10, with leave one out cross validation when k equals the sample size. The sensitivity to k's choice occurs for small values. The k-cross validation offered numerous advantages for evaluating the predictive modeling of traffic accidents. By leveraging the models are evaluated the ensures accurate reliable predictions. Overfitting is reduced by k-cross validation and improved generalizability and optimized the hyperparameters for better performance. This cross validation facilitated the model comparison provided the robust evaluation and handled imbalanced data, maximized data utilization.

IV. Experimental Results

In this study, CatBoost with Grid Search is proposed for predictive modelling of traffic accidents which is employed on Road Traffic Accident dataset. This CatBoost is implemented on ample memory, 16 gigabytes of RAM, 512 gigabytes of SSD dedicated with NVIDA GPU, 64-bit operating system. The proposed CatBoost is evaluated by using accuracy, precision, recall and RSME which is calculated by using below eq. (7), (8), (9) and (10)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(x_{obs,t} - x_{model,t})^2}{n}} \quad (10)$$

Here, TP , TF , FP , FN signifies True Negative, True Positive, False Positive and False Negatives $x_{obs,t}$ is the actual value at t, $x_{model,t}$ is the value predicted by model at time t.

A. Performance Analysis

The proposed CatBoost with Grid Search is compared with existing methods such as Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost) which is described in below Table 1.

TABLE I. Performance Analysis of Proposed CatBoost with Grid Search

Performance Analysis	Accuracy	Precision	Recall	RMSE
RF	0.81	0.71	0.59	0.51
GB	0.83	0.73	0.61	0.45
XGBoost	0.86	0.75	0.63	0.43
Proposed CatBoost with Grid Search	0.95	0.78	0.65	0.02

From above Table 1, the proposed CatBoost with grid search method established better results including accuracy (0.95), precision (0.78), recall (0.65) and RMSE (0.02) when compared with existing RF, GB and XGBoost.

B. Comparative Analysis

The comparison among the proposed CatBoost with Grid Search and existing models Naïve Bayes (NB), Support Vector Machine (SVM) which is described in below given Table 2.

TABLE II. Comparative Analysis

Comparative Analysis	Accuracy	Precision	Recall	RMSE
NB [11]	0.80	0.50	0.43	0.51
SVM [12]	0.87	0.75	NA	0.03
Proposed CatBoost with Grid Search	0.95	0.78	0.65	0.02

From Table 2, the proposed CatBoost with Grid search method established better results including accuracy (0.95), precision (0.78), recall (0.65), RMSE (0.02) when compared with existing NB [11] accuracy (0.80), recall (0.43), precision (0.50), RMSE (0.51), SVM [12] accuracy 0.87, precision 0.75, RMSE 0.03.

V. Discussion

In this research, CatBoost with Grid Search is proposed for predictive modelling of traffic accidents which is employed on Road Traffic Accident (RTA) dataset. The data is preprocessed by using Random Projection Paradigm (RPP) and Improved Linear Discriminant Analysis (ILDA) which effectively reduced dataset dimensions, simplified complex traffic data. Then, data is transformed by using one hot encoding which prevented from assuming false relationships between categories. After that, the model is selected by using CatBoost which effectively processes the categorical data and ensured robust modeling, the hyperparameters are tuned by using grid search optimization. Finally, the model is evaluated by using k-fold cross validation provided the robust evaluation and handled imbalanced data. The proposed CatBoost with Grid Search method established better results including accuracy (0.95), precision (0.78), recall (0.65), RMSE (0.02) when compared with existing NB [11] accuracy (0.80), precision (0.50), recall (0.43), RMSE (0.51), SVM [12] accuracy 0.87, precision 0.75, RMSE 0.03. This proposed CatBoost is time consuming process when, the model is trained by large datasets.

VI. CONCLUSION

This research proposes CatBoost Grid Search for predictive modelling of traffic accidents. Initially, the model is employed on Road Traffic Accident dataset, this data is preprocessed by using RPP and ILDA which effectively reduced dataset dimensions, simplified complex traffic data. Then, data is transformed by using one hot encoding which prevented from assuming false relationships between categories. After that, the model is selected by using CatBoost which effectively processes the categorical data and ensured robust modeling, the hyperparameters are tuned by using grid search optimization which evaluates all the combinations of hyperparameters and easily reduced the computation time. Finally, the model is evaluated by using k-fold cross validation provided the robust evaluation and handled imbalanced data. The proposed CatBoost with Grid Search method established better results including accuracy (0.95), precision (0.78), recall (0.65), RMSE (0.02) when compared with existing NB, SVM. In the future, CatBoost will evolve by integrating real time data to explore cut edge machine learning techniques, and prioritize interpretability and scalability.

REFERENCES

[1] Alkaabi, K., 2023. Identification of hotspot areas for traffic accidents and analyzing drivers' behaviors and road accidents. Transportation research interdisciplinary perspectives, 22, p.100929.

[2] Liu, Y., Wu, C., Wen, J., Xiao, X. and Chen, Z., 2022. A grey convolutional neural network model for traffic flow prediction under traffic accidents. Neurocomputing, 500, pp.761-775.

[3] Swetha, G., Datla, R. and Mohan, C.K., 2024. Federated learning to categorize air quality using satellite images. IEICE Proceedings Series, 81(S4-1).

[4] Hazaymeh, K., Almagbile, A. and Alomari, A.H., 2022. Spatiotemporal analysis of traffic accidents hotspots based on geospatial techniques. ISPRS International Journal of Geo-Information, 11(4), p.260.

[5] Sunitha, D., Balmuri, K.R., de Prado, R.P., Divakarachari, P.B., Vijayarangan, R. and Hemalatha, K.L., 2023. Congestion centric multi - objective reptile search algorithm - based clustering and routing in cognitive radio sensor network. Transactions on Emerging Telecommunications Technologies, 34(11), p.e4629.

[6] Canese, L., Cardarilli, G.C., Dehghan Pir, M.M., Di Nunzio, L. and Spanò, S., 2024. Design and Development of Multi-Agent Reinforcement Learning Intelligence on the Robotarium Platform for Embedded System Applications. Electronics, 13(10), p.1819.

[7] Satpathy, R.B., Ramesh, G.P., Alzubaidi, L.H., Chanti, Y., Hussein, A.H.A. and Theja, B.S., 2023, December. A Hexagonal Fractal Antenna with Switchable Radiation Pattern for Network Enabled On-Body Wireless Communications. In 2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC) (pp. 1-5). IEEE.

[8] Canonica, A.C., Alonso, A.C., da Silva, V.C., Bombana, H.S., Muzaurieta, A.A., Leyton, V. and Greve, J.M.D.A., 2023. Factors contributing to traffic accidents in hospitalized patients in terms of severity and functionality. International journal of environmental research and public health, 20(1), p.853.

[9] Zhou, Z., Dong, X., Li, Z., Yu, K., Ding, C. and Yang, Y., 2022. Spatio-temporal feature encoding for traffic accident detection in VANET environment. IEEE Transactions on Intelligent Transportation Systems, 23(10), pp.19772-19781.

[10] Zhang, L., Zhang, M., Tang, J., Ma, J., Duan, X., Sun, J., Hu, X. and Xu, S., 2022. Analysis of traffic accident based on knowledge graph. Journal of advanced transportation, 2022(1), p.3915467.

[11] Bokaba, T., Doorsamy, W. and Paul, B.S., 2022. Comparative study of machine learning classifiers for modelling road traffic accidents. Applied Sciences, 12(2), p.828.

[12] Marcillo, P., Valdivieso Caraguay, Á.L. and Hernández-Álvarez, M., 2022. A systematic literature review of learning-based traffic accident prediction models based on heterogeneous sources. Applied Sciences, 12(9), p.4529.

[13] Rashidi, M.H., Keshavarz, S., Pazari, P., Safahieh, N. and Samimi, A., 2022. Modeling the accuracy of traffic crash prediction models. IATSS research, 46(3), pp.345-352.

[14] Yan, M. and Shen, Y., 2022. Traffic accident severity prediction based on random forest. Sustainability, 14(3), p.1729.

[15] Yeole, M., Jain, R.K. and Menon, R., 2023. Road traffic accident prediction for mixed traffic flow using artificial neural network. Materials Today: Proceedings, 77, pp.832-837.

[16] RTA dataset: https://www.kaggle.com/datasets/saurabhshahane/road-traffic-accidents

[17] Yu, L., Zhou, R., Chen, R. and Lai, K.K., 2022. Missing data preprocessing in credit classification: One-hot encoding or imputation?. Emerging Markets Finance and Trade, 58(2), pp.472-482.

[18] Chang, W., Wang, X., Yang, J. and Qin, T., 2023. An improved CatBoost-based classification model for ecological suitability of blueberries. Sensors, 23(4), p.1811.

[19] Belete, D.M. and Huchaiah, M.D., 2022. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. International Journal of Computers and Applications, 44(9), pp.875-886.