Uluslararası Sürdürülebilir Mühendislik ve Teknoloji Dergisi
International Journal of Sustainable Engineering and Technology
Sayı 2, Cilt 7, (2023), Sayfa: 152-161

A STUDY ON TRAFFIC CRASH SEVERITY PREDICTION USING MACHINE LEARNING ALGORITHMS

Gül Fatma TÜRKER*[1], Fatih Kürşad GÜNDÜZ[1]

[1]Süleyman Demirel University, Engineering Faculty, Computer Engineering Department, Isparta

Article Info
Received: 05.10.2023
Accepted: 08.12.2023
Published: 29.12.2023

ABSTRACT

Severity of traffic accidents holds significant importance in terms of affecting human life. Predicting the severity of traffic accidents is crucial for accident prevention and ensuring safe driving. The success of predictive methods enables the identification of relevant risk factors and the implementation of countermeasures. The causes of traffic accidents can be attributed to a wide range of variables. In particular, factors such as road disturbances, weather conditions, vehicle speed as visible causes cause traffic accidents. Obtaining real-time information about the immediate situation of an accident is challenging. However, by calculating the severity of a potential accident based on its causes, it is possible to mitigate future accidents. In this study, machine learning algorithms were tested on a dataset of traffic accidents collected in the US states between 2016 and 2023 to predict accident severity. Comparative performance analysis was conducted using machine learning techniques such as Decision Tree, Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Naive Bayes (NB) to identify contributing factors to accidents. The results were evaluated according to Precision, Recall, and F1-Score metrics. It has been determined that the accuracy of the accident severity classification accuracy of the Decision Tree algorithm provides the best performance among others with an accuracy of 99.6%. This outcome signifies the potential of advanced predictive methods in significantly reducing the occurrence of traffic accidents through targeted interventions based on identified risk factors. This study's findings underscore the pivotal role of machine learning algorithms in enhancing the accuracy of traffic accident severity prediction.

Keywords; Classification, crash severity, machine learning, traffic accident

MAKİNE ÖĞRENMESİ ALGORİTMALARI KULLANILARAK TRAFİK KAZASI ŞİDDETİNİN TAHMİNİ ÜZERİNE BİR ÇALIŞMA

Makale Bilgisi
Geliş tarihi: 05.10.2023
Kabul Tarihi: 08.12.2023
Yayın tarihi: 29.12.2023

ÖZET

Trafik kazalarının şiddeti insan hayatını etkilemesi açısından büyük önem arz etmektedir. Trafik kazası şiddetinin tahmin edilmesi, trafik kazalarının önlenmesi ve güvenli bir sürüş sağlanması için önemlidir. Tahmin yöntemlerinin başarısı sayesinde ilgili risk faktörleri çıkarılmakta ve karşı önlemler alınabilmektedir. Trafik kazalarının sebepleri çok fazla değişkene bağlı olabilmektedir. Özellikle görünür sebepler olarak yol bozuklukları, hava şartları, araç hızı gibi faktörler trafik kazalarına neden olmaktadır. Oluşan bir kazada kazanın anlık durumunun ne olduğu bilgisini almak oldukça güçtür. Ancak sebeplerden kaynaklanan bir olası kazanın şiddeti hesaplanarak gelecek zamanda oluşacak kazaların önüne geçilmesi mümkündür. Bu çalışmada kaza şiddeti tahmini için ABD eyaletlerinde 2016-2023 yılları arasında toplanan trafik kaza veri seti üzerinde makine öğrenmesi algoritmaları test edilerek kaza şiddeti tahminleri yapılmıştır. Kazaya sebep olan etkenleri belirlemek için makine öğrenme tekniklerinden Decision Tree, Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbours (KNN) ve Navie Bayes (NB) algoritmaları kullanılarak karşılaştırmalı bir başarı analizi yapılmıştır. Sonuçlar Precision, Recall and F1-Score ölçütlerine göre değerlendirilmiştir. Decision Tree algoritmasının kaza ciddiyetini sınıflandırma accuracy değeri %99.6 doğrulukla diğerleri arasında en iyi performans sağladığı belirlenmiştir.

Anahtar Kelimeler; Sınıflandırma, kaza şiddeti, makine öğrenmesi, trafik kazası.

1. Introduction

Over the past two decades, traffic accidents have emerged as a significant problem, necessitating regular and efficient solutions from various research communities and scientists (WHO, 2023). Several reasons have been identified as contributing factors to road accidents. These include poor lighting conditions, sudden movements or actions, road surface conditions, absence of perception among drivers, and the age of the driver (Rabbani et al. 2020). Each of these factors contributes to increasing the likelihood of accidents occurring on the road. Addressing these issues and finding effective solutions is crucial for improving road safety and reducing the number of accidents (Cologrande, 2022-Kukartsev et al., 2022).

There are many factors in the causes of traffic accidents. Drivers should be able to interpret the information about the environment they are in and adjust their actions accordingly. First of all, it is of great importance to analyze the situations that cause the accident well and to inform the drivers with predictions before possible accident situations occur (Santos et al., 2022; Hou et al., 2022). Studies aiming to minimize the risk of future accidents are carried out thanks to systems that interpret and predict environmental variable information that affects the emergence of possible errors (Marcillo et al., 2022). It is important to make accurate predictions in order to record the data at the time of the accident and to prevent possible accidents (Shaygan et al., 2022). It is clear that traffic congestion will increase the risk, but it is important to ensure traffic safety first of all (Rezaei and Liu, 2019). Deep learning methods and machine learning technics aimed at reducing the risk of accidents in traffic and improving safety provide success in predicting the severity of accidents (Chakraborty et al., 2023; Rahim and Hassan, 2021).

In the modelling of traffic accident severity, studies are carried out with machine learning techniques. Santos etc. applied SVM, Decision Tree, K-Nearest Neighbour, Random Forest in their studies and found that Random Forest was a good approach by evaluating the performance. However, it is clear that unobserved heterogeneity and timing problems should be taken into account as well as the performance of the model. The use of other techniques, which may be most suitable for this purpose, continues in this field (Santos et al., 2022).

Zhang etc. used a Random Forest classifier with a hybrid method of feature selection in machine learning classification approach in single and multi-vehicle accidents and fed to a set of four classifiers (Naive Bayes, K-NN, Binary Logistic Regression and Extreme Gradient Boosting). The gender of the driver, presence of median and presence of shoulder were found to be insignificant. XGBoost surpasses other classifiers in predictive capability. Using the indicated attributes, the accuracy of XGBoost was found to be 82% for single-vehicle crashes and 79% for multi-vehicle crashes (Zhang et al., 2022). Uneven sample distribution is constitutes a significant challenges in most machine learning algorithms. Danesh etc. used two metaheuristic optimization algorithms to achieve a more equitable distribution of data and applied three machine learning algorithms for the dataset, namely Decision Tree, Support Vector Machine and K-Nearest Neighbour. They showed that by editing the unbalanced data set, it provided 30% improvement detection (Danesh et al., 2022).

In the literature, although predictions are made using statistical models, machine learning models are being developed to overcome their emerging limitations. Sattar etc. found that the Keras (MLP) model, which they used as a machine learning approach, outperformed other models with a training time of 3.45 seconds. In the analysis made using TabNet, they found that the variables that increase the severity of the accident are the number of vehicles, the number of injured, the speed limit, the intersection location, the vehicle type and the road type. side etc. They found that using Random Forest and Bayesian optimization as a hybrid model provides higher accuracy than traditional algorithms (Sattar et al., 2023).

Aldhari etc. random forest, XGBoost, and logistic regression classifier, which serves as a technique in machine learning, were used and XGBoost found the highest classification accuracy as 71% in crash severity prediction (Aldhari et al., 2023). Baykal etc. showed that accident severity classification with machine learning algorithms gives more effective results by removing unnecessary variables and preprocessing the data on a large data set. It has been determined that the best algorithm is RF, with an accuracy value of 0.816 (Baykal et al., 2023).

Kushwaha etc. compared the accuracy on different ML algorithms on the basis of performance and showed that the RF algorithm outperformed other algorithms (Kushwaha and Abirami, 2022). Sameen etc. constructed a Recurrent Neural Network (RNN) based deep learning model for predicting the injury severity of traffic accidents. They found that accuracy of the RNN model was 71.77%, while the MLP and BLR models reached 65.48% and 58.30% (Sameen and Pradhan, 2017). Vaiyapuri and Gupta employed Machine Learning and Deep Learning algorithms for cognitive assessment and identification of the primary factors contributing to the incidence of traffic accidents. SVM, RF, KNN, MLP and Logistic Regression were applied for accident severity estimation. It demonstrated that the MLP algorithm outperformed the other suggested algorithms with an increasing accuracy of about 88% (Vaiyapuri and Gupta, 2021).

Shao etc. used a multivariate negative binomial model to address unobserved heterogeneity problems in traffic accident prediction (Shao et al., 2018). Kunt etc. similarly in their studies, artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods were compared to understand the severity of traffic accidents (Kunt et al., 2011). Ma etc. stated that using the stacked sparse autoencoder deep learning model to predict the severity of injury in traffic accidents, it provides the best performance compared to the basic models (Ma et al., 2021). Accident detection is a vital part of traffic safety. Machine learning classifiers are frequently used to predict accident severity. However, traffic safety may not be realistic with their incapacity to grasp complicated models owing to their "black box" nature. Khattak etc. In their study, they proposed Meta-Learning for Dynamic Ensemble Selection (META-DES), K-Nearest Oracle Elimination (KNORAE) and Dynamic Ensemble Learning (DES) algorithms. The META-DES model using RF outperformed the other models with accuracy (75%) (Khattak et al., 2022).

Azhar. et al. employed machine learning techniques to identify and estimate road accidents by integrating social media alongside a collection of geotagged data. They introduced an advanced deep learning model for predicting accidents, amalgamating insights gleaned from tweet messages with additional attributes like sentiment analysis, emotional context, weather conditions, geocoded coordinates, and temporal details. The outcomes achieved indicate an enhancement of 8% in accuracy for accident detection, with a resulting test accuracy peaking at 94%. Compared to current state-of-the-art approaches, the suggested algorithm offered enhanced performance, with 2% and 3% increase in accuracy, increasing the accuracy to 97.5% and 90%, respectively (Azhar et al., 2023).

Sameen etc. showed that deep learning-based accident severity estimation models like CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network) exhibit superior performance compared to the traditional NN model in terms of precision and reliability (Sameen et al., 2023). Assi, compared the Deep Neural Network (DNN) with the SVM model in their study and it was found that DNN was superior in estimating the severity of the accident, with a prediction accuracy of 95% and an F1 score of 93% (Assi, 2020). Manzoor etc. offers a collection of ML and DL models by combining Random Forest and Convolutional Neural Network, called RFCNN, to predict the severity of traffic accidents. The results show that RFCNN improves decision making process and outperforms other models with 0.991 accuracy, 0.974 precision, 0.986 recall, and 0.980 F-score (Manzoor et al., 2021).

Formosa etc. have also evaluated data in hardware units to predict real-time traffic collisions using deep learning. It has been determined that the results vary according to the weather and traffic density, and the best DNN model provides 94% accuracy (Formosa, 2020). Predictions developed by deep learning have been used to evaluate the effects of the development of neural network layers, such as different types of injuries, death and property damage (Yang et al., 2022). Li etc. have worked on the utilization of data visualization for enhancing traffic safety research. LightGBM-TPE was added as a hybrid algorithm by incorporating the visualization method into machine learning and it was stated that compared to machine learning algorithms, f1 performed better in terms of accuracy, recall, and precision metrics. Using the LightGBM-TPE, they found that "Longitude", "Latitude", "Hour" and "Day of the Week" were the four primary variables most strongly associated with the accident severity (Li et al., 2022).

As seen in the literature review, various methodologies have been developed for the prediction of traffic accident severity. Employing statistical analysis, machine learning algorithms, neural networks and hybrid approaches, they developed prediction models on existing data and obtained different results by prioritizing different environmental factors.

This study underscores the critical importance of leveraging machine learning algorithms for predicting traffic accident severity, as evidenced by the meticulous analysis of accident records spanning 49 states in the USA from 2016 to 2023. Through the systematic application of Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, K-Nearest Neighbours, and Naive Bayes algorithms, the research aimed to not only discern contributing factors but also determine the algorithm that best excels in predicting accident severity. The outcomes, evaluated against Precision, Recall and F1-Score criteria, highlight the profound significance of employing advanced computational models for enhancing our understanding of traffic accident dynamics and, ultimately, formulating effective preventive measures.

2. Proposed Methodology

Prediction of traffic crash severity is an important area of research that aims to develop methodologies for accurately assessing the potential severity of traffic accidents. Various approaches and techniques have been employed in this field. The use of statistical analysis, machine learning algorithms, deep learning, and hybrid approaches can vary depending on the research objectives and computational resources (Niyogisibizo et al., 2023-Kong et al., 2023). Machine learning algorithms learn to classify or predict the severity levels of traffic accidents by using various input parameters. They utilize these input parameters to classify or predict the severity levels of traffic accidents (Hossain et al., 2022).

Determining the severity levels of road accidents is crucial for traffic accident analysis and safety. Classifying datasets into different severity levels such as minor, moderate, major, and fatal can be beneficial in understanding the impact of traffic accidents and identifying various risk factors (Santos et al. 2022).

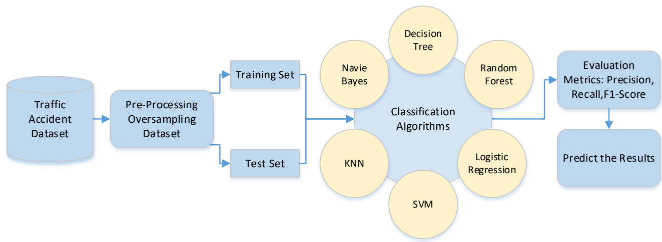

Figure 1. Developed model using various classification algorithms

The developed model shows our proposed study in Figure 1. nominal, numeric and mixed type data in the dataset are used. The selected data is processed; different machine learning classification techniques (such as Decision Tree, RF, LR, SVM, KNN and NB) are applied, predictions categorized according to varying levels of severity (minor, moderate, major, and fatal) are compared.

2.1. Dataset

The records of traffic accidents were gathered from traffic data captured by various agencies, including transportation agencies and road traffic sensors in road networks located in 49 states of the United States. The data includes 7.7 million accident records spanning from February 2016 to March 2023 (US Accidents, 2023).

Predictive studies of accident severity were conducted on input parameters such as Severity, Start Time, End Time, Description, Street, City, County, State, Zipcode, Country, Timezone, Airport Code, Source, Weather Timestamp, Wind Direction, and Weather Condition. The selected data is preprocessed, and then machine learning techniques are utilized for the purpose of comparison and calculate predictions according to the levels of severity.

2.2. Technics and Algorithms

In this study, 6 Machine Learning classification models, namely Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, KNN and Naive Bayes were used to determine the predominant factors contributing to the majority of traffic accidents.

Decision Tree is a machine learning algorithm used to classify data with decisions in a series. It divides the data into sections using a tree-like structure and performs the classification with the decisions in each section.

The goal is to divide the dataset into homogeneous subsets in the best way possible. This splitting process starts at the top of the dataset and at each step the dataset is split into subsets according to the best split criterion. Decision Tree uses different splitting criteria to split data. For classification tasks, criteria such as Gini impurity or Entropy are often used. These partitioning criteria are chosen to maximize the homogeneity of the subsets.

Random Forest Classifier is an ensemble method that builds decision trees on data samples and then combines the predictions from each decision tree to choose the best solution. Random Forest generates numerous decision trees on different subsets of the data and aggregates their predictions using majority voting to produce the final classification result. This ensemble approach improves the accuracy and robustness of the model (Yan and Shen, 2022).

Logistic Regression is a machine learning model used in classification problems. It provides an efficient method for estimating the likelihood of a binary outcome based on at least one factor. The sigmoid function (∂) and input (x) are found using equations (1) and (2).

$$\partial (x) = 1/(1+e(-x)) \quad (1)$$

$$x = w_0 z_0 + w_1 z_1 + \ldots.. w_n z_n \quad (2)$$

It calculates the classification probability using the sigmoid function and assigns the object to a class by deciding with the threshold value. It combines inputs and coefficients with specified independent variables, using the sigmoid function to predict the probability of a categorical dependent variable. It is an efficient method for both binary and multi-class classification, the estimated probability value takes a value between zero and one, and provides good performance for linearly separable classes.

Support vector machine is a classification machine learning algorithm, represents data points in a multidimensional space. It handles multiple continuous and categorical variables. It classifies datasets to determined the Maximum Marginal Hyperplane (MMH). In the SVM model, iterative hyperplanes that optimally differentiate the classes are generated and the hyperplane that accurately segregates the classes is selected. This hyperplane is defined as the best parser plane that separates the data points into different classes.

SVM uses support vectors to optimally separate data points on a hyperplane. Support vectors are the data points closest to the classification plane and are critical points that affect the classification decision. The SVM performs the classification by trying to maximize the margin between these support vectors and the hyperplane.

K-NN classification algorithm is a supervised machine learning algorithm. It is also used for regression problems. In classification tasks, the K-NN algorithm looks at the K nearest neighbour in the training set to classify an unknown data point. "K" is an adjustable parameter set by the user and determines how many neighbours are to be considered. The unfamiliar data point is assigned to a class, according to the class most commonly seen in these K neighbours.

Naive Bayes is a classification algorithm based on Bayes' theorem. It is used both in natural language processing applications like textual categorization and in a variety of other classification problems. It also works well with data whose observation values are considered independent of each other. This classification method uses probability calculations to calculate the effect of features in the dataset on classes. It uses Bayes' theorem to calculate the probabilities of the classes, under the assumption that the features are independent of each other.

3. Experimental Results and Analysis

In this study, we showcase experimental findings using a dataset collected from the USA. First of all, 75% of the data are trained for training purposes and 25% for testing purposes. Various machine learning algorithms are applied on the data to predict possible accidents by analyzing traffic accident severity. This approach uses Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, KNN and Naive Bayes, for example, predictions are generated based on Precision, Recall, and F1-Score criteria.

3.1. Performance Metrics

The confusion matrix illustrates the discrepancies in the machine learning classification model's predictions.

True Positives (TPs): Instances where the actual outcome is positive and is correctly predicted as positive.

False Negatives (FNs): Instances where the actual outcome is positive but is incorrectly predicted as negative.

True Negatives (TNs): Instances where the actual outcome is negative and is correctly predicted as negative.

False Positives (FPs): Instances where the actual outcome is negative but is incorrectly predicted as positive.

Bottom of Form

In order to thoroughly evaluate the model's performance, a systematic analysis has been carried out, categorizing its results into four distinct categories. Accuracy is a crucial and primary metric for evaluating a model's performance prior to practical application. It represents the proportion of correct predictions among the overall count of cases tested, providing a clear picture of how well the model aligns with the actual results (Rodionova et al., 2022).

Precision is a metric employed to assess the effectiveness of a classification model by measuring how precise or exact it is in predicting positive instances. It focuses on the accuracy of the positive predictions made by the model, specifically the ratio of true positive predictions to the total number of predicted positive instances. (Pillai et al. 2021). Precision = True Positive / (True Positive + False Negative)

Recall, often referred to as sensitivity or the true positive rate, is a metric utilized to assess the performance of a classification model, especially when there is a significant cost associated with false negatives. Recall focuses on the proportion of actual positive instances that the model correctly identifies as positive. Recall= True Positive / (True Positive + False Negative).

F1-Score represents an evaluative metric for assessing classification performance models by taking the harmonic average of the precision and recall metrics. It aims to achieve a balanced performance by minimizing both false positives and false negatives of the model. F1= 2*((Precision*Recall)/(Precision+Recall))

The F1-score is particularly useful for evaluating models in cases of imbalanced datasets or imbalanced classification problems, as it provides a comprehensive evaluation of the model's performance. The severity levels of accidents are described using the terms "minor" (insignificant/low severity), "moderate" (medium level/moderate severity), "major" (significant/high severity), and "fatal" (deadly).

Minor: Refers to accidents or incidents with low severity and minimal damage.

Moderate: Describes accidents or incidents with a moderate level of severity and moderate damage.

Major: Indicates accidents or incidents with high severity and significant damage.

Fatal: Denotes accidents or incidents that result in fatalities or death.

These are helpful in understanding the severity of incidents and classifying significant events, particularly in the context of traffic accidents or other incidents.

3.2. Prediction of Traffic Crash Severity and Analysis

Experiments were performed on Machine learning algorithms; Decision Tree, Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naive Bayes (NB). The accuracy values of the classification algorithms used are given in Table 1. Decision Tree classifier algorithm with an accuracy value % 99.6 was as the most suitable model. 99.6% accuracy rate is quite high and suggests that the model is performing very well on the given dataset. Decision Tree classifiers are known for their interpretability and ability to capture complex relationships in data, making them a popular choice for various classification tasks.

Tablo 1. Classification algorithms and accuracy values

Algorithms	AccuracyScore
Decision Tree	%99.6
Random Forest	%99.3
Logistic Regression	%62
Support Vector Machine	%64
K-Nearest Neigbour	%77
Navie bayes	%42

Confusion matrix is a measurement tool that offers insights into the accuracy of predictions (Yas et al., 2018). Confusion matrix for the Decision Tree classifier algorithm is given in Figure 2. Using the confusion matrix, accuracy, precision, recall, and f1-score values were calculated and their performances were compared.

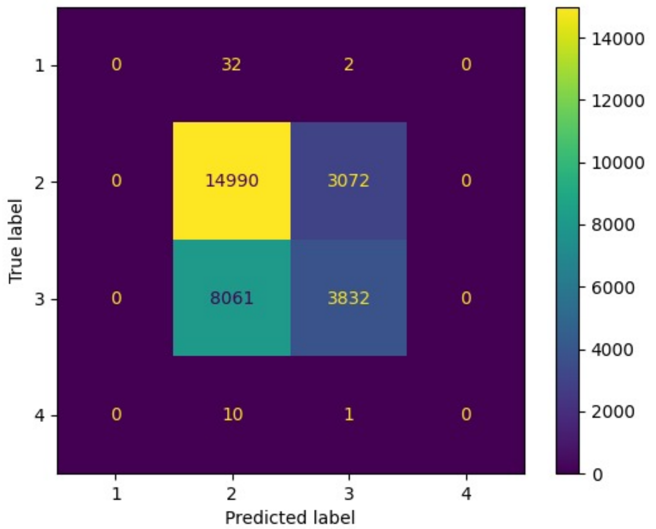

Figure 2. Confusion matrix for Decision Tree model

The classification algorithms utilized in the study will be assessed based on criteria including accuracy, precision, recall. The concept known as F1-measure is employed as a performance metric in the field of data sciences. Each severity class was rated as minor, moderate, major and fatal levels. An example of traffic crash severity classification levels is given in Table 2.

Tablo 2. Evaluation criteria of K-NN model

Severity		Precision	Recall	F1 Score
Minor	1	0.63	0.43	0.51
Moderate	2	0.78	0.94	0.85
Major	3	0.64	0.35	0.45
Fatal	4	0.70	0.41	0.52

4. Conclusions

It is important to examine datasets that contain information about road, vehicle and human factors that affect the occurrence and severity of accidents. In order to provide a comprehensive and precise analysis of traffic accident severity, studies have been conducted to identify the main factors that cause traffic accident severity. A comprehensive literature review has shown that predictive methods are used for decision making, improving road safety and reducing the severity of future accidents. In this article, within the context of road accidents, variable data obtained from various sources are evaluated to classify accidents into different severity levels: minor, moderate, major, and fatal. Machine learning algorithms including Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, K-Nearest Neighbors and Naive Bayes classifier techniques are employed to understand the causes of traffic accidents. As a result of the comparative evaluation, the best classifier with accuracy value was found. The performance of different classification techniques in the classification of traffic crash severity was evaluated at four different levels. According to the determined performance criteria, Decision Tree showed the best performance in classifying accident severity with 96.8% accuracy. In the literature studies, the data collected in previous years of the similar dataset were evaluated with various classification algorithms, but in this study, the classification algorithms tested on the increasing dataset as of 2023 gave successful results. Based on tests using real traffic accident data from the USA, our proposed model predicted traffic crash severity risks with good accuracy. Thanks to the prediction of the crash severity in the study conducted with a large dataset, measures can be taken before the accident occurs, and in the event of an accident, the technical team will be able to guide the relevant units according to the severity level of the accident.

5. Kaynaklar

Aldhari, I., Almoshaogeh, M., Jamal, A., Alharbi, F., Alinizzi, M., Haider, H., 2022. Severity Prediction of Highway Crashes in Saudi Arabia Using Machine Learning Techniques. Applied Sciences, 13, (1), 233.

Assi, K., 2020. Prediction of traffic crash severity using deep neural networks: a comparative study. In 2020 International Conference on Innovation and Intelligence for Informatics, December, Computing and Technologies (3ICT) (pp. 1-6). IEEE.

Azhar, A., Rubab, S., Khan, M. M., Bangash, Y. A., Alshehri, M. D., Illahi, F., Bashir, A. K., 2023. Detection and prediction of traffic accidents using deep learning techniques. Cluster Computing, 26, (1), 477-493.

Baykal, T., Ergezer, F., Eriskin, E., & Terzi, S., 2023. Accident Severity Prediction in Big Data Using Auto-Machine Learning. Scientia Iranica.

Chakraborty, M., Gates, T. J., Sinha, S., 2023. Causal analysis and classification of traffic crash injury severity using machine learning algorithms. Data science for transportation, 5, (2), 12.

Colagrande, S., 2022. A methodology for the characterization of urban road safety through accident data analysis. Transportation research procedia, 60, 504-511.

Danesh, A., Ehsani, M., Moghadas Nejad, F., Zakeri, H., 2022. Prediction model of crash severity in imbalanced dataset using data leveling methods and metaheuristic optimization algorithms. International journal of crashworthiness, 27, (6), 1869-1882.

Dong, C., Shao, C., Li, J., Xiong, Z., 2018. An improved deep learning model for traffic crash prediction. Journal of Advanced Transportation, 2018, 1-13.

Formosa, N., Quddus, M., Ison, S., Abdel-Aty, M., Yuan, J., 2020. Predicting real-time traffic conflicts using deep learning. Accident Analysis & Prevention, 136, 105429.

Hossain, M. J., Ivan, J. N., Zhao, S., Wang, K., Sharmin, S., Ravishanker, N., Jackson, E., 2023. Considering demographics of other involved drivers in predicting the highest driver injury severity in multi-vehicle crashes on rural two-lane roads in California. Journal of Transportation Safety & Security, 15, (1), 43-58.

Hou, Q., Huo, X., Leng, J., Mannering, F., 2022. A note on out-of-sample prediction, marginal effects computations, and temporal testing with random parameters crash-injury severity models. Analytic methods in accident research, 33, 100191.

Hussain, S. F., Ashraf, M. M., 2023. A novel one-vs-rest consensus learning method for crash severity prediction. Expert Systems with Applications, 228, 120443.

Khattak, A., Almujibah, H., Elamary, A., Matara, C. M., 2022. Interpretable Dynamic Ensemble Selection Approach for the Prediction of Road Traffic Injury Severity: A Case Study of Pakistan's National Highway N-5. Sustainability, 14, (19), 12340.

Kong, J. S., Lee, K. H., Kim, O. H., Lee, H. Y., Kang, C. Y., Choi, D., Sung, T. E., 2023. Machine learning-based injury severity prediction of level 1 trauma center enrolled patients associated with car-to-car crashes in Korea. Computers in biology and medicine, 153, 106393.

Kukartsev, V., Mikhalev, A., Stashkevich, A., Moiseeva, K., Kauts, I., 2022. Analysis of Data in solving the problem of reducing the accident rate through the use of special means on public roads. In 2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), June, 1-4. IEEE.

Kunt, M. M., Aghayan, I., Noii, N., 2011. Prediction for traffic accident severity: comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods. Transport, 26, (4), 353-366.

Kushwaha, M., Abirami, M. S., 2021. Comparative Analysis on the Prediction of Road Accident Severity Using Machine Learning Algorithms. In International Conference on Micro-Electronics and Telecommunication Engineering, September, 269-280. Singapore: Springer Nature Singapore.

Li, K., Xu, H., Liu, X., 2022. Analysis and visualization of accidents severity based on LightGBM-TPE. Chaos, Solitons & Fractals, 157, 111987.

Ma, Z., Mei, G., Cuomo, S., 2021. An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors. Accident Analysis & Prevention, 160, 106322.

Manzoor, M., Umer, M., Sadiq, S., Ishaq, A., Ullah, S., Madni, H. A., Bisogni, C., 2021. RFCNN: Traffic accident severity prediction based on decision level fusion of machine and deep learning model. IEEE Access, 9, 128359-128371.

Marcillo, P., Valdivieso Caraguay, Á. L., Hernández-Álvarez, M., 2022. A Systematic Literature Review of Learning-Based Traffic Accident Prediction Models Based on Heterogeneous Sources. Applied Sciences, 12, (9), 4529.

Niyogisubizo, J., Liao, L., Sun, Q., Nziyumva, E., Wang, Y., Luo, L., Murwanashyaka, E., 2023. Predicting Crash Injury Severity in Smart Cities: a Novel Computational Approach with Wide and Deep Learning Model. International Journal of Intelligent Transportation Systems Research, 21, (1), 240-258.

Rabbani, M. B. A., Musarat, M. A., Alaloul, W. S., Ayub, S., Bukhari, H., Altaf, M., 2022. Road accident data collection systems in developing and developed countries: a review. International Journal of Integrated Engineering, 14, (1), 336-352.

Rahim, M. A., Hassan, H. M., 2021. A deep learning based traffic crash severity prediction framework. Accident Analysis & Prevention, 154, 106090.

Rezaei, S., Liu, X., 2019. Deep learning for encrypted traffic classification: An overview. IEEE communications magazine, 57, (5), 76-81.

Sameen, M. I., Pradhan, B., 2017. Severity prediction of traffic accidents with recurrent neural networks. Applied Sciences, 7, (6), 476.

Sameen, M. I., Pradhan, B., Shafri, H. Z. M., Hamid, H. B., 2019. Applications of deep learning in severity prediction of traffic accidents. In GCEC 2017: Proceedings of the 1st Global Civil Engineering Conference 1 (pp. 793-808). Springer Singapore.

Santos, K., Dias, J. P., Amado, C., 2022. A literature review of machine learning algorithms for crash injury severity prediction. Journal of safety research, 80, 254-269.

Sattar, K., Chikh Oughali, F., Assi, K., Ratrout, N., Jamal, A., Masiur Rahman, S., 2023. Transparent deep machine learning framework for predicting traffic crash severity. Neural Computing and Applications, 35, (2), 1535-1547.

Shaygan, M., Meese, C., Li, W., Zhao, X. G., Nejad, M., 2022. Traffic prediction using artificial intelligence: review of recent advances and emerging opportunities. Transportation research part C: emerging technologies, 145, 103921.

Vaiyapuri, T., Gupta, M., 2021. Traffic accident severity prediction and cognitive analysis using deep learning. Soft Computing, 1-13.

World Health Organization, "Road Traffic Injuries", 2023. Date of Access https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Yang, Z., Zhang, W., Feng, J., 2022. Predicting multiple types of traffic accident severity with explanations: A multi-task deep learning framework. Safety science, 146, 105522.

Zhang, S., Khattak, A., Matara, C. M., Hussain, A., Farooq, A., 2022. Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents. PLoS one, 17, (2), e0262941.

Dataset, "US Accidents 2016-2023. Date of Access:https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents

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

Yas, Q. M., Zaidan, A. A., Zaidan, B. B., Rahmatullah, B., Karim, H. A., 2018. Comprehensive insights into evaluation and benchmarking of real-time skin detectors: Review, open issues & challenges, and recommended solutions. Measurement, 114, 243-260.

Rodionova, M., Skhvediani, A., Kudryavtseva, T., 2022. Prediction of crash severity as a way of road safety improvement: The case of SaintPetersburg, Russia. Sustainability, 14, (16), 9840.