Accident Analysis and Prevention 155 (2021) 106098

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Speed violation analysis of heavy vehicles on highways using spatial analysis and machine learning algorithms

Emre Kuşkapan [a,*], M. Yasin Çodur [a], Ahmet Atalay [b]

[a] Erzurum Technical University, Engineering and Architecture Faculty, Erzurum, Turkey
[b] Ataturk University, Engineering Faculty, Erzurum, Turkey

ARTICLE INFO

Keywords:
Heavy vehicles
Speed violation
Machine learning
Spatial analysis

ABSTRACT

With the development of technology in the world, vehicles that reach high speeds are produced. In addition, with the increase of road width and quality, faster and more comfortable transportation can be provided. These developments also increase the speed violation rates of road vehicles. Drivers who violate speed limits can endanger both their own lives and the lives of others. Speed violations, of especially heavy vehicles, involve much greater risks than that of light vehicles. Heavy vehicles can cause more serious losses of lives and property in accidents, compared to the ones caused by light vehicles, as they can carry much more freight or passengers than light vehicles. In this study, data regarding the speed violations committed by heavy vehicles in Turkey, were used. Speed violations were divided into 10 classes according to the intensity of speed violation rates. After this process, all provinces were classified according to support vector machines (SVM), naive bayes (NB) and k-nearest neighbors (KNN) algorithms. When the accuracy values and error scales of all three algorithms are examined, it has been determined that the algorithm that gives the most accurate results is the NB algorithm. Based on the classification of this algorithm, speed violation density maps of types of heavy vehicles in Turkey were created by using spatial analysis. According to the density maps, the provinces with the highest speed violations were identified. In the results, it was determined that the rate of heavy vehicle speed violation was highest in the cities such as Erzurum, Konya, and Muğla. Later, these cities were examined in terms of heavy vehicle mobility. At the end of this study, measures were proposed to reduce these violations in cities where speeding violations are intense. Material and moral damages can be prevented, to a great extent, with the implementation of recommendations of policymakers which can reduce speed violations.

1. Introduction

Road transport is the most common type of transport. For this reason, road transportation networks are developing day by day and high-quality roads are being built. Also, with the development of technology and smart transportation systems, road vehicles with high technological features and fast access are being produced. Most drivers openly admit that they more or less regularly exceed the speed limit. They state the reasons for these intentional speed limit violations as follows: they were in a hurry, they generally enjoy driving fast and they were bored (Elvik et al., 2004). Speed violations can cause various traffic problems. The biggest of these problems is traffic accidents. Major accidents can occur, causing material and moral damages. Besides, since these vehicles have high weights, it becomes difficult for the drivers to control the vehicle at speeds above the specified level (Anastasopoulos and Mannering, 2016).

Heavy vehicles are generally examined in three categories. Buses, the first category, are frequently used in intercity trips. Transportation by bus is typically regarded as a safe way for people to travel (Morency et al., 2018). However, when accidents occur there is a significant cost in terms of property loss and personal injury. In particular, for accidents that result in fatalities, such events are far from negligible (Bhowmik et al., 2019; Chimba et al., 2010; Zhang et al., 2019). Another type of heavy vehicles is trucks. Trucks are frequently used in heavy cargo transportation. Because they need to travel at slow speeds, they usually travel in the far right lane in traffic. Although their average speed is lower than buses, vehicle control by drivers becomes very difficult when they reach high speeds (de Vries et al., 2017d). If the trucks are out of control, overturning is common. Since trucks carry freight, financial losses occur in most of the accidents. The third category includes truck

* Corresponding author.
E-mail addresses: emre.kuskapan@erzurum.edu.tr (E. Kuşkapan), mycodur@erzurum.edu.tr (M.Y. Çodur), ahatalay@atauni.edu.tr (A. Atalay).

https://doi.org/10.1016/j.aap.2021.106098
Received 31 December 2020; Received in revised form 21 February 2021; Accepted 16 March 2021
Available online 7 April 2021
0001-4575/© 2021 Elsevier Ltd. All rights reserved.

Fig. 1. Map of the network of state roads in Turkey.

trailers and semi-truck trailers. Trailers are also used in freight transportation like trucks. They are longer vehicles than trucks. For this reason, their maneuvers on the highway are very difficult. In case of speed violations, they are very likely to cause chain accidents (Abdel-Aty and Pande, 2005; Anderson and Hernandez, 2017; Behnood and Mannering, 2019; Chen et al., 2021).

Speed violations cause serious problems in all three types of heavy vehicles. Various speed measuring devices are used to detect speed violations. Most of these devices inform the driver of the speed of the vehicle on digital screens and warn the driver to reduce the speed of the vehicle (Hashim et al., 2016). Heavy vehicle drivers can generally be more experienced in traffic than private vehicle drivers. Therefore, the frequency of traffic accidents is lower in heavy vehicles compared to light vehicles. Despite this situation, the amount of loss of life and property may be higher in a heavy vehicle accident compared to a light vehicle accident. In order to reduce these accidents and minimize the losses, the causes of accidents of heavy vehicle drivers should be examined well (Naderi et al., 2018; Wang et al., 2019).

2. Related work

There are various studies examining speed violations of vehicles. Some of these studies are directly and some are indirectly related to the speed violations of heavy vehicles. Afghari et al. (2018) analyzed the proportion of speed limit violations across highway segments using a panel mixed logit fractional split model. The results of the model suggested a tendency among drivers to commit minor speed limit violations irrespective of causal factors. Imaninasab et al. (2016) developed a prediction model using heavy vehicles accident. Poisson regression and negative binomial regression models were employed for modeling purposes. After assessment of the models, factors including natural light condition, heavy vehicles daily volume, speed limit violation and vehicle defects, especially in lighting system during darkness, were found to be the most effective factors on heavy vehicle accidents. Saifizul et al. (2011) analyzed the effect of gross vehicle weight (GVW) and category of heavy vehicle on free-flow speed and their interactions using empirical statistical techniques. The study results showed that the current speed limit is relatively high for a heavy vehicle with a GVW of over 20 t. Therefore, the authors proposed a new concept of setting the speed limit for heavy vehicles by incorporating the weight parameter.

Still, studies examining speed violations with inverse distance weighting (IDW) interpolation, a spatial analysis method, are less documented. Galgamuwa et al. (2020) estimated driver and environmental characteristics into the cross-sectional model. Generalized linear mixed models based on the Poisson distribution were used to develop cross-sectional models that incorporate driver and environmental characteristics as well as conventional cross-sectional models incorporating only road geometric and traffic-related characteristics. IDW was used for the interpolation of data. Arvin et al. (2019) quantified variations in vehicular movements and studied the impact of instantaneous driving behavior on the type of crashes at intersections. In order to capture variations in vehicle control and movement, they quantified and used 30 measures of driving volatility by using speed, longitudinal and lateral acceleration, and yaw-rate. By applying IDW interpolation, they mapped estimated lateral acceleration volatilities.

The number of studies examining speed violations with machine learning algorithms is also limited. Kong et al. (2020) investigated the hidden rules that associated trip/driving/roadway features with speeding behavior. A data mining algorithm was adopted to explore the hidden rules from two perspectives of speeding: speeding duration and speeding pattern. Samerei et al. (2020) investigated the relationship between the bus involved crash severity and chains of the effective factors (environment, humans, and vehicles) that lead to a fatality in bus involved crashes using a combination of two data mining methods. Wu and Hsu (2021) aimed to develop effective approaches for predicting at-fault crash driver frequency using only city-level traffic enforcement predictors. A fusion deep learning approach combining a convolution neural network (CNN) and gated recurrent units (GRU) was used to compare predictive performance with one econometric approach, two machine learning approaches, and another deep learning approach. In a study, both IDW interpolation and machine learning methods were used together.

As seen above, the literature regarding vehicle violation analyses using IDW or machine learning is acutely limited. Subsequently, the current study seeks to expand the literature on heavy-vehicle speed violation analysis by using IDW interpolation and machine learning algorithms. Previous works have generally only IDW or machine learnings algorithms. But the present study analyzes heavy-vehicle speed violations utilizing roadway using both IDW interpolation and machine learning algorithms. Speed violation rates of heavy vehicles in the state roads in Turkey were examined. Using data from all provinces, 10 different classes were determined according to the rate of speed violations. Performance values and error scales of SVM, NB and KNN algorithms for these 10 classes were calculated with WEKA Software. The results obtained in the algorithm, speed violation density map was created for each heavy vehicle type with the IDW interpolation in ArcGis

Table 1
A sample is taken from data including AADT values and speed violations according to road sections.

Region no	City	Road length (km)	Total AADT (Vehicles/ Day)	Bus		Truck		Trailer		Heavy vehicle (%)
				(Vehicles /Day)	Speed violation (%)	(Vehicles /Day)	Speed violation (%)	(Vehicles /Day)	Speed violation (%)	
7	Samsun	9	35,999	320	47	1749	13	3313	10	14
7	Samsun	13	66,696	326	90	2292	61	3481	68	9
7	Samsun	6	35,710	311	38	1578	9	2711	8	12
7	Samsun	21	20,797	333	68	1137	20	2316	14	17
7	Samsun	18	18,395	318	54	1053	21	2312	15	18
7	Çorum	28	11,529	273	74	501	20	1538	19	18
7	Çorum	7	13,594	252	60	572	21	1585	17	16
7	Çorum	8	2542	19	11	119	13	105	2	9
7	Tokat	19	8124	102	69	412	20	671	15	13
7	Tokat	43	4052	131	76	217	24	855	12	26
7	Tokat	24	3344	110	74	211	34	878	18	33
7	Ordu	26	22,942	310	25	1078	7	2311	7	15
7	Ordu	6	23,302	285	21	965	9	2168	7	13
7	Sinop	11	3453	35	64	133	17	229	18	10
7	Sinop	22	3417	40	11	244	6	248	8	14
7	Amasya	21	12,699	393	51	599	27	1882	19	20
7	Amasya	12	24,491	542	57	1077	21	3760	13	20
7	Amasya	24	13,738	241	73	546	27	1383	19	14
7	Amasya	14	18,621	159	23	905	9	859	10	9

Software. In the study, the accuracy of the classification was increased by using the machine learning algorithm. High accuracy maps were obtained by using the best performance classification algorithm. IDW interpolation was used while creating maps.

3. Material method

3.1. Study area

Turkey is an important transit point between Asia and Europe. It is like a bridge between the two continents in passenger and freight transport. Due to this situation, heavy vehicle traffic is quite high in the country. Approximately 90 % of freight transportation and approximately 88 % of passenger transportation in the country are provided by the highway. For this reason, the transportation network has a developed structure. As of 2020, there is about 31 thousand km of state roads in Turkey (General Directorate of Highways, 2020). State roads also play an important role in the transition between provinces. Speed limits are applied to ensure safe transportation on these roads. A map of the network of state roads in Turkey is given in Fig. 1 below. The speed limit for buses and trucks is 80 kph and for trailers is 50 kph, on undivided state roads. The speed limit for buses is 90 kph, for trucks is 85 kph, and for trailers is 50 kph, on divided state roads.

3.2. Data processing

The General Directorate of Highways (GDH) is a state agency in charge of the construction and maintenance of all public roadways outside of cities and towns in Turkey. Also, it has activities such as improving and repairing the roads and keeping them under constant maintenance to ensure their safe use. Another important task is to collect, press and publish the information about activities of GDH.

GDH publishes the annual average daily number of traffic for all vehicle types passing through road sections in all provinces and speed violation rates for all vehicle types. Thanks to these data, it is easier to make traffic analyses. Highways in Turkey are divided into 17 regions with respect to their location. Data are calculated separately for these 17 regional directorates. Each regional directorate covers several cities. The data generated by the regional directorates include detailed information. It is possible to get down to the center of the problems thanks to the creation of data, especially considering the vehicle types. One of the tables including traffic transportation information prepared by GDH is shown in Table 1. In this table, as an example, traffic information on some road sections of the state roads in a part of the 7th Regional Directorate can be seen (General Directorate of Highways, 2020). The speed violation rate given in the table is the ratio of the number of vehicles that violate speed on a road section to the total number of vehicles passing that road. For example, in the first line of the table, the annual average daily traffic value on a 9 km road section in Samsun is 35,999 vehicles/day. Approximately 14 % of the vehicles passing through this road section are heavy vehicles. An average of 320 buses, 1749 trucks and 3313 trailers per day pass through this road section. 47 % of the passing buses, 13 % of the passing trucks and 10 % of the passing trailers commit speed limit violations.

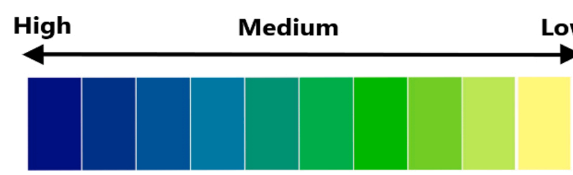


Fig. 2. Classification according to speed violation intensity.

When the table is examined, it is seen that heavy vehicles exceed the speed limit at a serious rate. It is a very risky situation for buses to speed over 90 % on some road sections. Because buses are heavy vehicles with high passenger capacity. The speed limit exceeding in trucks is less than in buses. Trailers exceed the speed limit less than trucks. Since these vehicles have heavier loads and larger dimensions than trucks, the rate of exceeding the speed limit is lower. However, serious material and moral damages can be caused by accidents by all these heavy vehicle types. To minimize the accidents caused by heavy vehicles, it is very important to determine the regions where speed violations are intense and to take some precautions in these regions. In order to do this, it is necessary to classify the speed violation rates according to the intensity. In this study, speeding violation rates are divided into 10 different classes according to their intensity. Fig. 2 below shows the classes created according to the velocity violation intensity and the coloring to be used in the ArcGis Software for these classes.

As can be seen in the figure, the classes with high density of speed violation are shown in bold colors. Trials have been made to determine the optimum number in classification. When the number of classes increased, the error rates of machine learning algorithms increased, too. When the number of classes was reduced, the maps created were not detailed enough. For this reason, the classification number recommended by Arcgis Software was preferred.

3.3. Machine learning

3.3.1. Support vector machine (SVM)

In machine learning, SVM is a supervised learning algorithm. It can be used for classification as well as regression. They are very efficient for many applications in science and engineering, especially for classification problems (pattern recognition). The main objective is to identify a hyper-plane that can classify the data points distinctively. The SVM classifier can be made non-linear by mapping the input variables to a higher dimension space and then applying the algorithm to the processed parameters (Burges, 1998). This mapping is done using the SVM kernels. Two kernels are evaluated; radial basis function (RBF) and polynomial. The equation for radial basis function (RBF) and polynomial in Scikit-learn is shown in Eqs. (1) and (2), respectively (Buitinck et al., 2013).

$$K(x,z) = exp\left(-y\|x-z\|^2\right) \tag{1}$$

$$K(x,z) = \left(\mathrm{y}\left(\mathrm{x}^T z\right) + c\right)^d \tag{2}$$

In Eqs. (1) and (2), K is the kernel function, x and z are vectors in the input space, c is the coefficient, d is the order of the polynomial, and γγ is the kernel parameter. The kernel parameter γγ gives the influence of a single training data point (Karandikar, 2019; Smola and Scholkopf, 2004).

3.3.2. Classification via Naive Bayes (NB)

NB is an algorithm that performs operations based on probability calculations. It processes the found train data according to its formula and extracts a percentage for each case and performs the classification of the test set according to these probabilities (Atmaca, 2020). P (A/B); The probability that event A occurs when event B occurs (see conditional probability), P (B/A); It is the likelihood that event B will occur when event A occurs. P (A) and P (B); A and B are the preliminary probabilities of events (Benhassine et al., 2020). The algorithm explained Eq. (3).

$$P(A\ /\ B) = [P(B/A) \times P(A)]/P(B) \tag{3}$$

3.3.3. K-nearest neighbors (KNN)

KNN is also known as sample-based learning. This algorithm is a useful data mining technique that allows past data samples to be used with known output values to estimate an unknown output value of a new data sample. The algorithm compares new problem examples with those seen in education and stored in memory, instead of making clear generalizations. The most important advantage of the neighbor algorithm is that it can adapt its model to unprecedented data. KNN, also known as memory-based learning, estimates a value or class for a new sample while calculating distances or similarities with previous training examples for this example (Machines, 2019). It is found by calculating the distances from each point in the KNN master data set to a point in the test data of which the core value is unknown. Thus, neighbors are calculated by selecting the k number of observations with the closest distance. This method uses Euclidean distance, which is formulated in Eq. (4) for points i and j when calculating distances (Çavuşoğlu and Kaçar, 2019).

$$d(i,j) = \sqrt{\sum_{k=1}^{p} \left(x_{ik} - x_{jk}\right)^2} \tag{4}$$

3.3.4. Performance and error scales analysis of algorithms

While performing performance analysis in machine learning, basic success criteria concepts are used. These concepts are precision, sensitivity, F-measure, and ROC area. When calculating the values of these concepts, the comparison of the estimated and available data is taken

		Predictive Values	
		Class =1	Class =0
Actual Values	Class=1	TP	FN
	Class =0	FP	TN

Fig. 3. Confusion matrix.

into account (Zhou, 2016). In the comparison process, TP (true positive-right) means TN (true negative-right means false), FP (false positive-false means), and FN (false negative-false means wrong) values are used.

Using the confusion matrix given in Fig. 3, the accuracy values of the classification algorithms can be calculated. The equations used in these calculations are given below. All of these performance values take values between 0 and 1.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{5}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{6}$$

$$\text{F-measure} = (2\times \text{ Recall } \times \text{ Precision})\ /\ (\text{Recall} + \text{Precision}) \tag{7}$$

$$\text{Accuracy} = (\text{TP} + \text{TN})\ /\ (\text{TP} + \text{FP} + \text{FN} + \text{TN}) \tag{8}$$

The error scales of the model are determined by the accuracy rate, mean square error (MAE), root mean square error (RMSE), and Kappa statistics. Kappa value is a term expressed to measure the mismatch between observations. The closer this value is to 1, the better the agreement between observations. The Kappa statistic is frequently used to test interrater reliability. The Kappa statistic is formulated as Eq. (9).

$$K = \frac{p_o - p_e}{1 - p_e} \tag{9}$$

Where p_o and p_e are expectation and observation, respectively. The meaning of this calculation has come into question, and ranges for the measure vary. However, an example would include the following: K < 0.20 = poor agreement; K = 0.21 to 0.40 is fair; K = 0.41 to 0.60 is moderate; K = 0.61 to 0.80 is substantial; and K > 0.81 is good (McHugh, 2012).

MAE error of a model concerning a test set is the mean of the absolute values of the individual prediction errors over all instances in the test set. It can be expressed in Eq. (10).

$$MAE = \frac{1}{N}\sum_{\mathrm{i}=1}^{N}\left|x_{f,i} - x_{o,i}\right| \tag{10}$$

Where $x_{f,i}$ and $x_{o,i}$ are the ith expectation and observation, respectively (Sammut and Webb, 2010).

RMSE is the square root of mean squared error. RMSE measures the differences between values predicted by a hypothetical model and the observed values (Profile, 2010). It can be expressed in Eq. (11).

$$RMSE = \sqrt{\frac{1}{N}\sum_{\mathrm{i}=1}^{N}\left(x_{f,i} - x_{o,i}\right)^2} \tag{11}$$

Where $x_{f,i}$ and $x_{o,i}$ are the ith expectation and observation, respectively.

(c) KNN

Fig. 4. Performance values of the algorithms.

3.3.5. WEKA

WEKA Software contains many machine learning techniques, which were developed using Java, for carrying out numerous data mining processes and was developed at the University of Waikato in New Zealand. WEKA includes various algorithms for analysis and predictive modeling. These algorithms can be applied directly to the data set. This tool consists of many tools for data mining activities such as classification, data preprocessing, clustering, regression, association rules, or visualization (Smith and Frank, 2016). WEKA helps to develop many supplemental machine learning algorithms. It also contains many classes that can be easily accessed by other WEKA classes.

3.4. Spatial analysis

Spatial analysis is sampling, manipulation, exploratory and confirmatory analysis of data into spatial modeling which encompasses a large and diverse set of models. IDW is one of the most widely used methods in spatial analysis. In the IDW interpolation, the cell values are determined by a linear weighted combination of the dataset of sample points. The weight represents a function of inverse distance (Nistor et al., 2020). The IDW interpolation assumes the proportionality of correlations and similarities between neighbors and the distance between them. The interpolated surface represents a locally dependent variable (Chen et al., 2018). In the standard IDW method, the height of the interpolation point at the P (x, y) position in the region covered by the N = {X, Y, Z} point set is calculated by the equation,

$$z_e = \left(\sum_{i=1}^{n} S_i * Z_i\right) \Big/ \left(\sum_{i=1}^{n} S_i\right) \tag{12}$$

In the equation, x, y, z denote the height value of the point, Z the height values of the fulcrums, S the weight values, n the number of fulcrum points. S weight values in Eq. (12) as a function of the distance (d) between the fulcrum point and the interpolation point is calculated from the equation,

$$Si = 1/d_i^p, \mathrm{i} = 1, 2, 3, 4\ldots \quad \mathrm{p} = 1, 2, 3, 4 \tag{13}$$

As the value of the p power parameter is increased in the function, the effect of distant points on the calculation decreases in Eq. (13). In other words, it is aimed to minimize the negative effects in modeling the surface where the interpolation point of the data obtained from distant points is located (Shukla et al., 2020).

Table 2
Error scales of the algorithms.

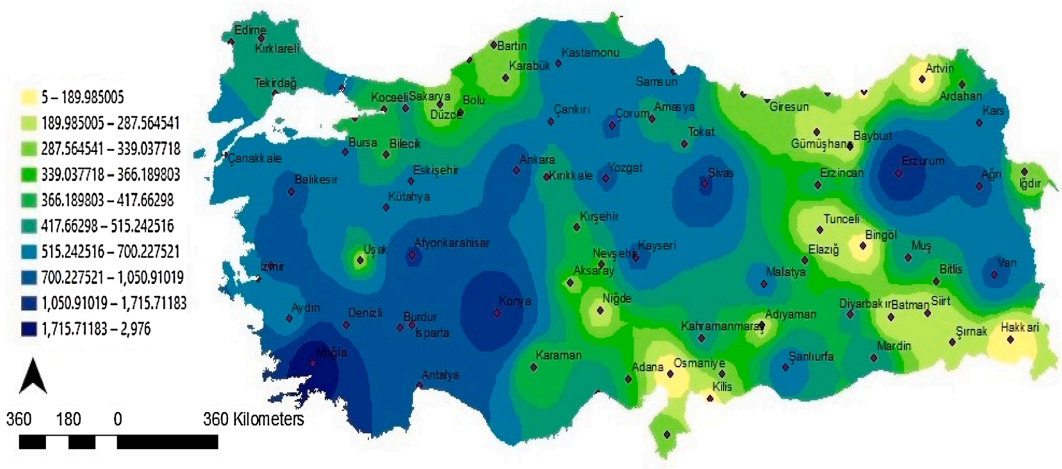
Algorithms	Correctly Classified Instances (%)	MAE	RMSE	Kappa
SVM	64.65	0.263	0.443	0.546
NB	88.73	0.129	0.316	0.827
KNN	76.12	0.163	0.332	0.739

4. Results and discussions

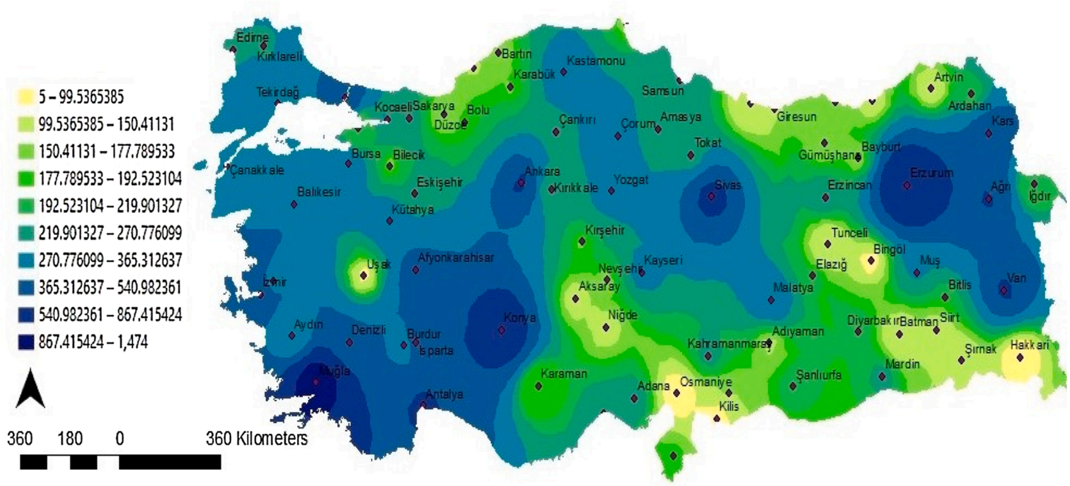
The classification cases of SVM, NB and KNN algorithms were examined for the classes determined according to the rate of speed violations. Because each algorithm classifies with different methods, performance values and error scales are different. When classified according to speed violation rates, the algorithm with the highest performance values has high accuracy. However, a higher correctly classified instance value and Kappa statistic value of an algorithm implies that this algorithm yielded better results. Since MAE and RMSE express error values, these values are expected to be small in an algorithm that performs better. The algorithm that gives the best results in terms of these criteria has a higher fit to the used data set. By choosing the most suitable algorithm for the data set, it will be possible to create maps with high accuracy while creating a density map of the speed violation rate. Performance values obtained by each algorithm are given in Fig. 4 below.

When the performance values of all three algorithms are examined, it is seen that the NB algorithm gives the best results. In second place is the KNN algorithm. Among these three algorithms, the SVM algorithm showed the lowest performance for the current data set. It is also necessary to look at the error scale values to determine which of these three algorithms is preferred. Error scales of each algorithm are given in Table 2 below.

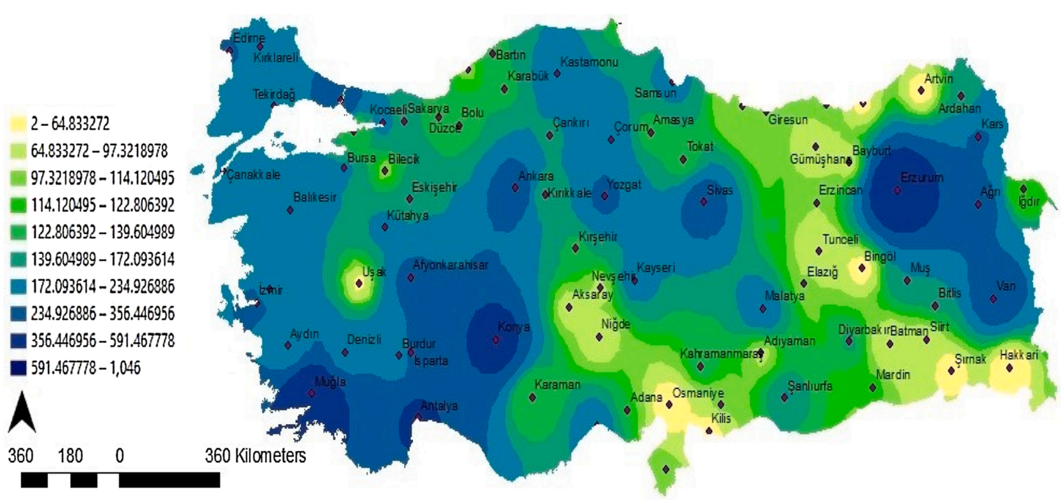
When the table is examined, the NB algorithm gave better results in error scales as well as in performance values. Accordingly, the algorithm has an 88.73 % correct classification, revealing that it is more compatible with the data set. Similarly, a higher Kappa statistic value of the algorithm indicates that this algorithm has a better performance. For this reason, MAE and RMSE values are also very low. The second KNN algorithm has 76.12 % correct classification and the last SVM algorithm has 64.65 % correct classification. As with the error scales, it has been shown that the NB algorithm gives better results in performance values. Therefore, the results of this algorithm were taken as the basis when creating the heavy vehicle speed violation density map. Density maps



(a) Bus



(b) Truck



(c) Trailer

Fig. 5. Speed violation density maps by heavy vehicle types.

containing speed violation rates for each heavy vehicle type were obtained with ArcGis Software using the IDW analysis method. Intensity maps of speed violations occurring according to buses, trucks and trailers are given respectively in Fig. 5 below.

The common feature of all three cities is that they are medium-sized and centrally located. There are also different features that increase the volume of heavy vehicles for cities. Erzurum ranks first in cattle breeding throughout the country. Due to this situation, cattle are transported from Erzurum to other cities via trucks and trailers. Another important feature of the city is that it has the largest ski centers in the country. For this reason, many local tourists come to Erzurum from the surrounding cities via buses during the winter months. Both of these conditions increase the city's heavy vehicle mobility at certain periods. Because the city is cold during the winter months, frost and icing may occur on the roads. That is why many serious accidents occur during the winter months. In case of snow and icing on the road surface, it becomes difficult to control vehicles at high speeds. Measures to be taken to reduce speed violations, especially in winter, can significantly reduce accidents in this city.

Konya is the first in the country in the grain trade. For this reason, truck and trailer mobility is high in the city. Also, due to the fact that there are 5 different universities in the city, bus mobility increases at certain times of the year. Horizontal and vertical curves are rare on the roads because the city of Konya is established on a very wide flat plain. Roads that are straight for very long distances can lead drivers to speed violations. In order to prevent this, large intersections can be placed on straight road sections with certain intervals, or curves with large radius can be added. In this way, the driver's attention loss can be prevented and speed violation tendencies can be prevented on long straight roads.

Since Muğla is one of the most popular holiday cities, many tourists come from other cities throughout the year by bus. Also, seafood trade is common in this city. For these reasons, heavy vehicle mobility is high in the city. Despite this, radar systems that control speed violations are very few in the city. For this reason, many accidents due to speed violations occur on the road networks in the city.

In addition to all these situations, the wide road networks and the high number of lanes in these cities can cause high density of speed violations. On the other hand, speed violations are lower in other cities with similar characteristics to those of Erzurum, Konya and Muğla.

Cities with low speed violations are almost the similar in all three types of heavy vehicles. These cities are Uşak, Osmaniye, Şırnak, Hakkari, Artvin and Bingöl. The common feature of these cities is that they generally have low population density and the transportation network is not developed enough. On the other hand, these cities are cities where the traffic load is lower than the large cities located near them.

The value of speed is also important in analyzing speed violations. For example, if the speed limit is exceeded on a road where the speed limit is 30 kph, the seriousness of an accident that will occur is less. However, the severity of the accident that will occur as a result of the speed violation is much higher in the road section where the speed limit is 70 kph. In this respect, it would be more correct to evaluate the roads having the same speed limit together when examining speed violations. In this study, state roads where the speed limit is higher than other roads are taken as the basis. When the speed limits on state roads are exceeded, serious material and moral damages can occur. For this reason, it is important to reduce speed violations in areas shown in bold on speed violation maps created to reduce accidents. Warning signs can be useful in reducing speed violations in road sections where road networks are developed and the number of lanes is high. These problems can be reduced by speed controls, radar applications and implementation of speed violation sanctions, especially during periods of increased mobility of heavy vehicles.

5. Conclusion

In this study, all the cities in Turkey are divided into 10 different classes depending on the rate of speed violations committed by heavy vehicles. Then SVM, NB and KNN algorithms were analyzed according to these classes. As a result of the analysis, it was observed that the NB algorithm gave better results than SVM and KNN algorithms in terms of both performance values and error scales. Using the data obtained according to the NB algorithm, a speed violation density map was created separately for trucks, buses and trailers. IDW interpolation was used while creating these maps. When the maps are examined, the cities with the highest rate of speed violations for all three types of heavy vehicles are approximately similar to each other. According to these maps, it was observed that speed violations were mostly in the cities of Erzurum, Muğla and Konya, which are medium-sized and located in more central locations than the surrounding provinces. Other cities that are close to dark colors, like these cities, are of critical importance for speed violation rates. Since the speed violation rates are high in these cities, some measures can be taken by policymakers. Speed inspections can be increased and various sanctions can be applied to drivers who violate speed limits in these cities. In this way, speed violations will be reduced and possible traffic accidents will be prevented. Thus, the material and moral damage that may occur can be minimized.

Overall, this study emphasizes that speeding violations cause serious loss of life and property. Especially since heavy vehicles carry loads or passengers, losses increase in accidents involving these vehicles. Future research on the topic can be based on reducing speed violations with the help of technological advances. In addition, future lines of research can measure how speeding violation rates can be reduced and the overall impact of road controls on road safety. Finally, the effect of the mental state of heavy vehicle drivers on speed violations can be investigated.

Author statement

Category 1

Conception and design of study: Emre Kuşkapan, M. Yasin Çodur, Ahmet Atalay

Acquisition of data: Emre Kuşkapan, M. Yasin Çodur, Ahmet Atalay

Analysis and/or interpretation of data: Emre Kuşkapan, M. Yasin Çodur, Ahmet Atalay

Category 2

Drafting the manuscript: Emre Kuşkapan, M. Yasin Çodur, Ahmet Atalay

Revising the manuscript critically for important intellectual content: Emre Kuşkapan, M. Yasin Çodur, Ahmet Atalay

Category 3

Approval of the version of the manuscript to be published: Emre Kuşkapan, M. Yasin Çodur, Ahmet Atalay

CRediT authorship contribution statement

Emre Kuşkapan: Conceptualization, Methodology, Formal analysis, Writing - original draft, Funding acquisition. **M. Yasin Çodur:** Investigation, Data curation, Writing - original draft. **Ahmet Atalay:** Conceptualization, Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgment

The authors would also like to gratefully acknowledge the General Directorate of Highways (GDH) for providing access to vehicle speed information.

References

Abdel-Aty, M., Pande, A., 2005. Identifying crash propensity using specific traffic speed conditions. J. Safety Res. 36 (1), 97–108. https://doi.org/10.1016/j.jsr.2004.11.002.
Afghari, A.P., Haque, M.M., Washington, S., 2018. Applying fractional split model to examine the effects of roadway geometric and traffic characteristics on speeding behavior. Traffic Inj. Prev. 19 (8), 860–866. https://doi.org/10.1080/15389588.2018.1509208.
Anastasopoulos, P.C., Mannering, F.L., 2016. The effect of speed limits on drivers' choice of speed: a random parameters seemingly unrelated equations approach. Anal. Methods Accid. Res. 10, 1–11. https://doi.org/10.1016/j.amar.2016.03.001.
Anderson, J., Hernandez, S., 2017. Roadway classifications and the accident injury severities of heavy-vehicle drivers. Anal. Methods Accid. Res. 15, 17–28. https://doi.org/10.1016/j.amar.2017.04.002.
Arvin, R., Kamrani, M., Khattak, A.J., 2019. How instantaneous driving behavior contributes to crashes at intersections: extracting useful information from connected vehicle message data. Accid. Anal. Prev. 127 (August), 118–133. https://doi.org/10.1016/j.aap.2019.01.014, 2018.
Atmaca, K., 2020. Naive Bayesian Algorithm [WWW Document]. https://kenanatmaca.com/naive-bayesian-algoritmasi/.
Behnood, A., Mannering, F., 2019. Time-of-day variations and temporal instability of factors affecting injury severities in large-truck crashes. Anal. Methods Accid. Res. 23, 100102 https://doi.org/10.1016/j.amar.2019.100102.
Benhassine, N.E., Boukaache, A., Boudjehem, D., 2020. Classification of mammogram images using the energy probability in frequency domain and most discriminative power coefficients. Int. J. Imaging Syst. Technol. 30 (1), 45–56. https://doi.org/10.1002/ima.22352.
Bhowmik, T., Yasmin, S., Eluru, N., 2019. A multilevel generalized ordered probit fractional split model for analyzing vehicle speed. Anal. Methods Accid. Res. 21 (July), 13–31. https://doi.org/10.1016/j.amar.2018.12.001, 2018.
Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B., Varoquaux, G., 2013. API Design for Machine Learning Software: Experiences From the Scikit-Learn Project, pp. 1–15.
Burges, Christopher J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. 2, 121–167.
Çavuşoğlu, Ü., Kaçar, S., 2019. Performance analysis of data mining algorithms for abnormal traffic detection. Acad. Platf. J. Eng. Sci. 7, 2. https://doi.org/10.21541/apjes.418519.
Chen, C., Peng, F., Yan, R., Fan, Z., Li, Y., Wei, D., 2018. Posture-dependent stability prediction of a milling industrial robot based on inverse distance weighted method. Procedia Manuf. 17, 993–1000. https://doi.org/10.1016/j.promfg.2018.10.104.
Chen, G.X., Sieber, W.K., Collins, J.W., Hitchcock, E.M., Lincoln, J.E., Pratt, S.G., Sweeney, M.H., 2021. Truck driver reported unrealistically tight delivery schedules linked to their opinions of maximum speed limits and hours-of-service rules and their compliance with these safety laws and regulations. Saf. Sci. 133 (June), 105003 https://doi.org/10.1016/j.ssci.2020.105003, 2020.
Chimba, D., Sando, T., Kwigizile, V., 2010. Effect of bus size and operation to crash occurrences. Accid. Anal. Prev. 42 (6), 2063–2067. https://doi.org/10.1016/j.aap.2010.06.018.
de Vries, J., de Koster, R., Rijsdijk, S., Roy, D., 2017d. Determinants of safe and productive truck driving: empirical evidence from long-haul cargo transport. Transp. Res. Part E Logist. Transp. Rev. 97, 113–131. https://doi.org/10.1016/j.tre.2016.11.003.
Elvik, R., Christensen, P., Amundsen, A., 2004. Speed and Road Accidents. An Evaluation of the Power Model. TØI Report 740/2004. Institute of Transport Economics TOI, Oslo.
Galgamuwa, U., Vahl, C.I., Dissanayake, S., 2020. Role of driver behaviors and environmental characteristics in evaluating safety effectiveness of roadway countermeasures: a novel approach of estimating crash modification factors. J. Transp. Saf. Secur. 12 (6), 764–781. https://doi.org/10.1080/19439962.2018.1545714.
General Directorate of Highways, 2020. Traffic and Transportation Information [WWW Document]. https://www.kgm.gov.tr/Sayfalar/KGM/SiteTr/Trafik/TrafikveUlasimBilgileri.aspx.
Hashim, W., Arshad, A.K., Mustaffa, M., Kamaluddin, N.A., 2016. Heavy vehicles speed profiling on urban expressway: The case of federal highway. J. Teknol. 78 (7–2), 19–23. https://doi.org/10.11113/jt.v78.9467.
Imaninasab, R., Sekhavati, E., Hajihoseinloo, M., 2016. Modeling crash frequency of heavy vehicles in rural freeways. J. Traffic Logist. Eng. 4 (2), 98–102. https://doi.org/10.18178/jtle.4.2.98-102.
Karandikar, J., 2019. Machine learning classification for tool life modeling using production shop-floor tool wear data. Procedia Manuf. 34, 446–454. https://doi.org/10.1016/j.promfg.2019.06.192.
Kong, X., Das, S., Jha, K., Zhang, Y., 2020. Understanding speeding behavior from naturalistic driving data: applying classification based association rule mining. Accid. Anal. Prev. 144 (June), 105620 https://doi.org/10.1016/j.aap.2020.105620.
Machines, I.B., 2019. Nearest Neighbor and Server-Side Library [WWW Document]. https://www.ibm.com/developerworks/library/os-weka3/index.html.
McHugh, M.L., 2012. Lessons in biostatistics interrater reliability: the kappa statistic. Biochem. Medica 22 (3), 276–282.
Morency, P., Strauss, J., Pépin, F., Tessier, F., Grondines, J., 2018. Traveling by bus instead of Car on urban major roads: safety benefits for vehicle occupants, pedestrians, and cyclists. J. Urban Health 95 (2), 196–207. https://doi.org/10.1007/s11524-017-0222-6.
Naderi, H., Nassiri, H., Sahebi, S., 2018. Assessing the relationship between heavy vehicle driver sleep problems and confirmed driver behavior measurement tools in Iran. Transp. Res. Part F Traffic Psychol. Behav. 59, 57–66. https://doi.org/10.1016/j.trf.2018.08.007.
Nistor, M.M., Rahardjo, H., Satyanaga, A., Hao, K.Z., Xiaosheng, Q., Sham, A.W.L., 2020. Investigation of groundwater table distribution using borehole piezometer data interpolation: case study of Singapore. Eng. Geol. 271 (March), 105590 https://doi.org/10.1016/j.enggeo.2020.105590.
Profile, S.E.E., 2010. Farklı Yöntemler Kullanılarak GeliÇtirilen Trafik Kaza Tahmin Modelleri ve Analizi. Uluslararası Mühendislik Araştırma ve Geliştirme Derg. 2 (1), 16–22.
Saifizul, A.A., Yamanaka, H., Karim, M.R., 2011. Empirical analysis of gross vehicle weight and free flow speed and consideration on its relation with differential speed limit. Accid. Anal. Prev. 43 (3), 1068–1073. https://doi.org/10.1016/j.aap.2010.12.013.
Samerei, S.A., Aghabayk, K., Mohammadi, A., Shiwakoti, N., 2020. Data mining approach to model bus crash severity in Australia. J. Safety Res. https://doi.org/10.1016/j.jsr.2020.12.004.
Sammut, C., Webb, G.I. (Eds.), 2010. Mean Absolute Error, in: Encyclopedia of Machine Learning. Springer US, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_525, p. 652.
Shukla, K., Kumar, P., Mann, G.S., Khare, M., 2020. Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi. Sustain. Cities Soc. 54 (December), 101997 https://doi.org/10.1016/j.scs.2019.101997, 2019.
Smith, T.C., Frank, E., 2016. Introducing machine learning concepts with WEKA. Stat. Genom. Met. Prot. 353–378. https://doi.org/10.1007/978-1-4939-3578-9_17.
Smola, A.J., Scholkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222.
Wang, L., Abdel-Aty, M., Lee, J., Shi, Q., 2019. Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. Accid. Anal. Prev. 122 (December), 378–384. https://doi.org/10.1016/j.aap.2017.06.003, 2016.
Wu, Y.W., Hsu, T.P., 2021. Mid-term prediction of at-fault crash driver frequency using fusion deep learning with city-level traffic violation data. Accid. Anal. Prev., 105910 https://doi.org/10.1016/j.aap.2020.105910, 150 May 2020.
Zhang, X., Hu, W., Zhou, J., Zuo, Q., Wu, R., Tang, Z., 2019. The failure route between active and latent error in bus accident. IEEE Access 7, 164941–164951. https://doi.org/10.1109/ACCESS.2019.2949858.
Zhou, Z.H., 2016. Learnware: on the future of machine learning. Front. Comput. Sci. China 10 (4), 589–590. https://doi.org/10.1007/s11704-016-6906-3.