Article

Using Machine Learning in Predicting the Impact of Meteorological Parameters on Traffic Incidents

Aleksandar Aleksić, Milan Ranđelović and Dragan Ranđelović *

Faculty of Diplomacy and Security, University Union-Nikola Tesla Belgrade, Travnicka 2, 11000 Belgrade, Serbia
* Correspondence: dragan.randjelovic@fdb.edu.rs

Abstract: The opportunity for large amounts of open-for-public and available data is one of the main drivers of the development of an information society at the beginning of the 21st century. In this sense, acquiring knowledge from these data using different methods of machine learning is a prerequisite for solving complex problems in many spheres of human activity, starting from medicine to education and the economy, including traffic as today's important economic branch. Having this in mind, this paper deals with the prediction of the risk of traffic incidents using both historical and real-time data for different atmospheric factors. The main goal is to construct an ensemble model based on the use of several machine learning algorithms which has better characteristics of prediction than any of those installed when individually applied. In global, a case-proposed model could be a multi-agent system, but in a considered case study, a two-agent system is used so that one agent solves the prediction task by learning from the historical data, and the other agent uses the real time data. The authors evaluated the obtained model based on a case study and data for the city of Niš from the Republic of Serbia and also described its implementation as a practical web citizen application.

Keywords: machine learning; regression; classification; prediction; meteorological parameters; traffic incidents; multi-agent architecture

MSC: 68T05; 68T42

1. Introduction

The parameters affecting the occurrence of traffic incidents (TI) comprise three main groups, and they are as follows: human factors, vehicle and environment [1]. However, in [2], the authors divide the main factors into five groups, i.e., in addition to the listed three groups, they add roadway as well as occupants and other road users, and in [3], the authors consider more types of parameters without any groups; therefore, it could be concluded that there is no single taxonomy. In this paper, the authors consider the influence of the mentioned third group, environmental factors, and in it, just the meteorological subgroup which belongs to a wider subgroup of atmospheric parameters from the environment factors group (Dastoorpoor et al. [4]). The prediction of the impact of atmospheric parameters on TI is an important task for solving one global, serious problem because they cause not only human losses but also economic damages. By 2030, TIs are predicted to become the sixth leading cause of death, overtaking cancer [5], i.e., the seventh leading cause of death, and overtaking HIV/AIDS [6] worldwide. TIs also have an economic importance because they cause 3% of the gross domestic product yearly loss globally and roughly double that in lower-middle-income countries [6].

Having in mind the above-mentioned data on the significantly expressed negative consequences of the occurrence of TI on human lives and the economy, it is obvious that the previously mentioned prediction of the influence of various (including meteorological), factors is one important task in preventing their occurrences. This process of predicting

the impact of different factors on the occurrence of a traffic accident has two main tasks: firstly, to help citizens themselves reduce the possibility that they will have any incidents in traffic, and, secondly, to help traffic police increase their control over known locations in specific weather conditions and, in this way, reduce the number of incidents. In this way, with the realization of those two main tasks, using the mentioned prediction, consequently, can reduce human casualties as well as economic damage. As we already mentioned, the subject of this paper is the prediction of the impact of meteorological factors as a subgroup of the group of atmospheric parameters and the wider group of environmental parameters. All those parameters included and others, such as sociological, geographical and so on, are of a different type and can be viewed in different ways. Therefore, for example, Chan et al. in [7] observe those parameters in groups of meteorological variables (temperature, pressure, humidity, wind speed, etc.), pollutant variables (PM, CO, O_3, SO_2, etc.), auxiliary variables (geographical, time, sociological, economic, relation to the type of road, etc.). Jalilian et al. in [8] consider each factor individually without their grouping. It is very important to remark that many of these groups of factors are considered as variables with different values through time historically which enables later learning of knowledge from them. The authors deal, in this paper, only with the first group of meteorological factors, from all the mentioned groups, and their impact on traffic accidents in a day. Therefore, the authors suggest that, in addition to the available data in real life from competent institutions, there are also large amounts of historical data collected in those institutions for the purpose of gaining knowledge using machine learning (ML) to predict the impact of the meteorological factors on the occurrence of TI.

Furthermore, in the paper, the authors propose an ensemble method of aggregation to solve the task of prediction. For this purpose, they used one aggregation of the mentioned two approaches of using historical and real-time data into a methodology that can be effectively implemented in a multi-agent system using modern intelligent technology.

For the evaluation of the impact of these factors on TI, we can find in the literature several classic statistical methods, and the most used among these is regression analysis, followed by factor and discriminant analysis and, on the other hand, some algorithms of data mining and artificial intelligence—between which, the most used are artificial neural networks and different algorithms of classification. Practically, this process means that the choice of the subset from the set of parameters must lead to the problem of the feature selection before creating a prediction [9]. Having in mind these facts, the authors set out to provide the report on one research method as the main objective of this manuscript, and the advantage of aggregation in the prediction model for determining the impact of the meteorological factors on TI in two methods will be discussed. These two methods that the authors proposed are the most used methods from the previously mentioned two groups, both in one ensemble ML prediction model: from the group of statistical methods, binary regression and from the group of ML methods, classification with feature selection. Moreover, as an obligatory part of this research was considering the realization of a developed ensemble model as one multi-agent system (MAS) in a global case, concrete in this paper is a two-agent system in which agent 1 draws knowledge through ML from historically available data, and then agent 2 deals with those same parameters, but in real time. Such two-agent architecture enables decision making using the algorithm which could be a decision matrix from the group of decision makers, and this will be proposed in the paper in the section which deals with the technical solution of the implementation of the proposed MAS including one emergent intelligence technique which enables the integration of MAS into a collaborative whole precisely based on that algorithm. Having in mind the complexity of the considered problem and the already-mentioned fact that the impact of meteorological factors on TI is only one small part of all the groups of factors—such as the presence of human factors, vehicle factors, other environmental parameters from different particles in the atmosphere including air pollution, road factors, geographical factors as well as factors of economic development and so on—it is an obvious need that this model implementation with one MAS must enable its permanent, continuous upgrading. Such a

solution, based on the approach of using MAS in the proposed form, the authors could not find in the literature.

In this paper, for the purpose of evaluating the proposed model, the authors used one case study which observes daily TI data for the city of Niš in the Republic of Serbia in the period from 1999–2009 and data for different meteorological factors for this period for the same city. This study determines individual influence of each of considered meteorological factor on a happening of TIs using for that the model of aggregation different classification algorithms. It is based on an algorithm which was previously pre-processed using different methods of feature selection from ML and binary regression analysis from the group of traditional statistics methodologies in one ensemble method of ML. In this way, the conditions for implementation are ensured in the already-described two-agent system for early warning of interested parties, before everyone else, all ordinary citizens and traffic police, including and using today's most popular social media platforms as can be found in the paper of Lu et al. [10]

Description of influence of different meteorological conditions on TI can be found in papers that use the application of different forms of regression models, from linear and binary regression, than general linearized model to the combination of artificial intelligence and regression and different autoregressive methods for that purpose. This type of statistical models is also often used for the predicting and impact of different mentioned groups as, for example, from an environmental group that could be a location, type of road, date and time and so on, and some individual factors from these groups and their combinations. The global review of possibilities and characteristics of different types of regression methods could be found in Trencevski et al. [11] and Gupta et al. [12]. Too many studies address the impacts of different meteorological parameters on traffic safety [13]. In one of them, a meta-analysis of 34 studies which deal with the effect of precipitation is given and, as a result, gave an average increase in traffic accidents of 71% and 84%, in case of rain and snowfall, respectively [14]. However, in [15], authors considered terms of crash severity and concluded that there is a significant reduction under rainy conditions compared to fine weather. Additionally, the effect of precipitation on traffic accidents is considered on Finnish motorways and it is concluded it could be different for different types of accidents, so that the relative risk for single accidents in relation with multiple accidents in case of snow is 3.37/1.98 [16]. In the paper [17] we find one investigation on the influence of 17 meteorological factors on the number of crashes in the Netherlands during 2002. The impact of a combination of the meteorological factors of temperature and precipitation is given in [18]. Study [19] investigates the impact of weather elements and extreme snow or rain weather changes on seven crash types using five years of data collected for the City of Edmonton in Canada. In [20], authors deal with different vehicle types, from high-sided trucks and buses to vans that are the most affected by strong wind; also, we can find in literature that in general, greater wind speeds increase the severity of traffic accidents caused by single trucks [21].

The effect of the sun glare on traffic accidents is the subject of a small number of studies, but authors of the paper [22] deal with this problem using data from signalized crossroads in the city of Tucson, USA.

They concluded that traffic accidents occur more frequently during glares from the rear-end and sideways and that a sun glare has no effect on the crash severity. However, in paper [23] we can find that traffic accidents in Japan indicate that the sun glare has a strong impact on pedestrian traffic accidents, crashes at crossroads, and bicycle crashes, while there is no indication that the impact of sun glare increases with vehicle speed [23].

In [24], authors deal with determining the impact of snowfall on TI. The different studies mentioned are focused on the impacts of individual or a group of meteorological factors on specific traffic accident types, but these studies could differ with relation to region, time period and methodology, and because of that, they are difficult for comparing the results.

In [25], the study in which correlation and linear regression analysis were conducted to estimate the influence of meteorological factors on road traffic injuries stratified by severity is presented. The study Khan et al. [26] shows that the occurrence of traffic accidents in hazardous weather conditions of wind, rainfall, snowfall and fog broadly follows the patterns for those weather parameters. A paper which deals with weather impacts on various types of road crashes: a quantitative analysis using Generalized Additive Model (GAM) method we can find in [27], but the paper [28] considers the same problem using the combined backward propagation-artificial neural network model (BP–ANN) regression model. In [29] one integer autoregressive model for prediction with four traffic safety categories: vehicle accidents, vehicle fatalities, pedestrian accidents and pedestrian fatalities in Athens was proposed. In [30], we can find the application of one autoregressive integrated moving average (ARIMA) model for determining the impact of weather factors on TI in France and comparing the general linearized model and ARIMA in [31] in the case of considering this problem in France, Greece, and the Netherlands. In [10], we can find the use of the regression model in one modern, conceptualized, complex system for prediction of influence of different whether factors on TI in China with help of data obtained from modern social media, combining these with physically sensed data and also with the help of regression methodology.

On the other hand, using ML algorithms for determining the importance of the individual impact of each of the many meteorological factors on traffic accidents as well as in determining suitable prediction models to solve this problem is today the other frequently used methodology. We can find more and more papers in the existing literature that use these two groups of methods to solve the posed problem discussed in this paper. These methods belong to different individual types of ML: classification, clustering, neural networks, and other standard ML methods; to the aggregations of these standard ML methods mutually or with classic statistical methods as for example regression, and in the end, the newest different ensemble methods is where the solution proposed by the authors belongs.

Thus, in [32], Zheng et al. consider different groups of atmospheric factors: meteorological variables (temperature, pressure, humidity, wind speed, etc.), pollutant variables (PM, CO, O3, SO2, etc.), auxiliary variables (geographical, time, sociological, related to the type of road, etc.), and we could practically find one comprehensive review and taxonomy of different types of ML methods which could be applied in atmospheric environment studies on TI for what is compatible with content presented in the already-cited reference [11], which particularly processed regression models. A similar problem from the standpoint of sensor using in this purpose is presented in [33], but in [34], a review of urban traffic flow prediction techniques with special focus on the literature review is presented. In [35–38], we can also find a similar comprehensive review of different artificial neural network (ANN) methods used for the same purpose. Using ML models based on ANN is a highly effective way to simulate the atmospheric environment, which is very important in the case of time-limited applications [39], and in this group, deep learning has received special research attention [40–42]. Different models of ANN are available, for example, recurrent ANN [43]. Additionally, the ANN predictions of meteorological impact factors on traffic accidents are available for geographically diverse areas across the globe: Switzerland [44], Bangladesh [45], Jordan [46], Iran [47], the USA [48], Australia [49], and are generally considered for developing countries [50].

Moreover, in the literature, we can find an aggregation of the most applicable ML ANN method with other ML methods: for example, with genetic algorithms in [51], with cluster algorithms in [52], using a logit model and factor analysis [53], with the random forest [54,55], and with the second most used ML the decision tree from a classification group of methods in the case study for data from Nigeria in [56]. Additionally, it is used in the case study of Nord England using the UK stats19 data set [57], and in the case study for data from the USA [58]. We can find the application of aggregated different classification

methods in papers such as J48,ID3, classification and regression tree (CART), decision tree and Naive Bayes in [59]; and CHAID, J48 decision tree and Naive Bayes in [60].

In the already-cited reference [32], it is remarked that increasing the model type of ML models for prediction of the impact of atmospheric parameters on different fields of human life were ensemble models; this is the case in the field of traffic accident prediction as well. In [61], one systematic review of ML methods is given where ensemble methods, as most modern types, are considered in separate sections and in [62–65] we could find descriptions of different ensemble methods which deal with ensemble learning for predicting traffic accidents affected by meteorological parameters. Having in mind the model which authors will propose in this paper, it is especially important to remark that the ensemble methods based on aggregation classification and regression tree in one ensemble algorithm for the purpose to solve the considered problem of prediction as an impact of meteorological factors on traffic accidents are very rarely in the literature, but they could be found, as, for example, in [66].

Particularly, there is a trend of developing forecast models to predict future states in all types of traffic at the beginning of the 21st century. Different taxonomies of those models can be found, for example the division into parametric and non-parametric models depending on the distribution of input values [67], then a division into deterministic models in which the model outputs are fully determined by the input factors values, and probabilistic, i.e., stochastic models [68]. ARIMA is one of the most-used parametric methodologies with its different subtypes: for example, multivariate spatial-temporal autoregressive (MSTAR) model [69], which at the same time belongs into probabilistic methods according to the second mentioned division while time-series analysis and trends belong to deterministic methodology, for example [70].

Bayesian deep learning approach and convolutional neural networks are increasingly present in recently published literature to predict the influence of uncertain environmental parameters, including the meteorological factors considered in the paper on the TI [71]. It is especially expressed in the subfield of so-called short-term predictions of trajectories in different types of public traffic, for example, in aircraft trajectory predicting [71,72] as well as in the prediction of road traffic in general and autonomous driving [73]. Because the authors set as the main goal of this paper that it should give an answer to the two research questions:

(1) Is it possible to construct one ML ensemble method which aggregates ML classification methods and methods of future selection for attribute selection with the binary regression method and which demonstrated better characteristics of prediction than each individually of included in ensemble method?

(2) Can this new ensemble method be implemented in one multi-agent supported technological system?

To give an answer on these two research questions and confirm those two hypotheses, the authors used evaluation of the proposed ensemble model on the case study for the city of Niš, Republic of Serbia, using its meteorological and data for TI for ten years in the beginning of 21 century.

In order to realize the set goal and present that the proposed ensemble model is an effective solution for the considered problem of predicting traffic accidents, the authors realized the rest of this manuscript in the following way: after this first section, the Introduction, the second section follows: Materials and Methods, where the authors gave a description of the material used and a comprehensive review of applied methodologies; then comes the third section—Results and Findings, in which the results of applying the proposed model in the case study are described; in the next section—Technological Implementation of Proposed Ensemble Model, the authors described the implementation of the proposed model as one technical solution and at the end there is a fifth section—Conclusions, in which contributions of this research are given and future work on efficiently solving the problem discussed in this paper is proposed.

2. Materials and Methods

Having in mind the present development of improved solutions for the prediction of the impact of atmospheric parameters on the occurrence of traffic accidents which are computer based and often using the ML techniques existing today, it could be said that its group of mentioned ensemble methods is a trend in solving such a complex problem. Implementation of such solutions using multi-agent solutions follows this trend directly. However, in the literature, there is still not a large enough number of references which integrate more methods of ML, different or of the same type in the ensemble models of prediction, so additional research of such methods is needed; that was the motivation for the authors to develop one such novel method.

In this paper, the authors described not only the new proposed model but also its implementation as one of agents in one multi-agent system of emergent intelligence technique (EIT) for the purpose of one citizens warning system. For the evaluation of the model proposed as such, the authors conduct the material from case study for the City of Niš in the Republic of Serbia which is presented in this paper. In it, the analyzed material is classified so that all data in the period considered is divided into two classes: positive when the daily number of traffic accidents is bigger than the average value for this period, and negative in all other cases. This way, it could be said that the positive class includes the instances when conditions significant enough for the occurrence of traffic accidents on that day are present in the atmosphere.

2.1. Methods

The problem of predicting the impact of meteorological parameters on TI that is the subject of consideration in this paper belongs to the group of classification problems for whose solving two main groups of methods are available: the classic statistical methods of logical regression and ML based classification.

With the logistic regression model, we describe the relationship between predictors that can be continuous, binary, categorical and categorically dependent variables. For example, the dependent variable can be binary-based on some predictors; we predict whether something will happen or not. We actually estimate the probabilities of belonging to each category for a given set of predictors. Depending on the type of dependent variable, we have:

Binary logistic regression—the dependent variable is binary (for example: answer true or false on the questions);

Nominal logistic regression—the dependent variable has three or more categories that cannot be compared in value (for example, colors (white, black, red, green, blue, etc.);

Ordinal logistic regression—the dependent variable has three or more categories that can naturally be compared, but the ranking does not necessarily mean that the "distances" between them are equal (for example: health status (stable, serious or critical).

Logistic regression is used when the dependent variable takes only a finite set of values.

We wonder if we can still use linear regression in classification problems. In the case of binary logistic regression, we consider the dependent variable to be a Bernoulli random variable in notation Y as it is shown in Equation (1). Then, we have two categories that we code with: 0 for failure and 1 for success.

$$Y = \begin{array}{l} 0 - \text{failure} \\ 1 - \text{success} \end{array} \tag{1}$$

Therefore, the dependent variable is a Bernoulli and not some continuous random variable, meaning that errors cannot be normal. Additionally, if we did run a linear regression, we would get some meaningless fitted values—values outside the set {0,1}. In the case of a binary dependent variable, one way to use linear regression for a classification problem can be as follows: for a given set of predictors, if the fitted value by linear regression is greater than 0.5, then we classify that observation as a success, and if not, a failure. This method is then equivalent to the linear discriminant analysis, which we will discuss later.

With this method, we only get a classification for some observation: "success" or "failure". If the fitted values by linear regression are close to 0.5, then we are less confident in our decision. We will also say that, if the dependent variable takes more than two values, then linear regression cannot be used as we described a moment ago, but the linear discriminant analysis must be used instead.

ML is one comprehensive discipline based on statistical analysis and artificial intelligence and it is used for learning of knowledge, i.e., concrete learning of rules, concepts, models, etc. which should be understood and accepted by the people. In the ML process, it is obligatory to have some kind of evaluation of the validity of the knowledge learned in this process, i.e., some kind evaluation of obtained rules, concepts, or models. For this purpose, two evaluation methods are available based on the process in which the available set is divided in different ways into a learning set and a test set:

(1) Evaluation using the test suite-holdout method, whose technique divides the original data set into two disjoint subsets, for training and for classifier testing (e.g., in a ratio of 70:30). Then, the model of classification is obtained on the basis of training data, after which the performance model on test data is evaluated. Thus, the accuracy of the classification can be assessed based on the test data;

(2) K-fold cross-validation is a classification model evaluation technique that is a better choice compared to evaluation using a test set. In general, it is performed by dividing the original data set into k equal subsets (layers). One subset is used for testing and all others for training. The resulting model makes predictions on the current layer. This procedure is repeated for k iterations using each subset exactly once for testing.

One of the most important measures of success of learned knowledge is named predictive accuracy. It is the ratio of the total number of successful classifications to the total number of classifications. For measuring a success of learned knowledge are also often used and precision, recall, F1 measure and receiver operating characteristic curve which will be described in the continuation of this chapter. The basic goal of any predicting process is to obtain one model based on the exact numerically determined combination of independent variables for the dependent variable. It is important to remark that in this process, the choice of variables that will be included in this process from a given data set affects the accuracy and other measures of the obtained prediction model, so because of that, it is necessary to use different techniques for a selection of variables in the data preparation phase, i.e., to apply some method of the so-called feature selection procedure.

One ensemble model of ML is proposed in this paper for predicting the potential risk of traffic accidents caused from meteorological, i.e., atmospheric parameters. As we already mentioned, the proposed method is one aggregation that optimizes more different classification algorithms using attribute reduction and binary regression. Implementation of this algorithm could represent one agent in considered and described two-agent system in this paper in which other agent realizes alarm calculations accordingly to value of meteorological parameters in real time. This two-agent system could be proposed in general as a wider and more complex multi-agent system which could be included and other types of environment parameters beside meteorological and which could be based on different possible forms of emergent intelligence for collective decision making. Such an implementation of an EIT solution as an emergency software tool could be realized as a web application accessible to all stakeholders of human society, starting with citizens and other interested parties.

The subchapters which follow in this paper are devoted for a brief description of these methodologies because the proposed ensemble method aggregates the method of logical regression with ML methods of classification methods and feature selection.

2.1.1. Classification Methodology

Classification algorithms belong to the supervised ML technique and can be used for the task of predictive modelling. Using the classification methodology for this purpose implies the existence of labeled instances in each of more than one class (attribute) of objects

so that it predicts the value of obligatory categorical type of class (attribute) using the values of the remaining predicting attributes [74].

The selection of the appropriate classification algorithm for the concrete considered application is not only the beginning but is also the most important place in the process of ML from big data. For solving the problem which is considered in this paper, in their proposed ensemble model, the authors use a classification which makes the classification into two classes, positive and negative that correspond to true or false in both of them. All possible outcomes of prediction are presented in the confusion matrix shown in Table 1.

Table 1. The confusion matrix for the two-class classifier.

		Predicted Label	
		Positive	**Negative**
Actual label	Positive	TP(true positive)	FN(false negative)
	Negative	FP(false positive)	TN(true negative)

The number of members in the considered set shown in Table 1 is the sum of positive and negative cases and will be classified in notation N, i.e., TP + FN + FP + TN = N. All results that are presented in Table 1, for a considered case of two-class classifier, can be given for the most important measures of classification accuracy, precision, recall and F1 measure with the following formulas:

$$\text{Accuracy} = (\text{TP} + \text{TN})/\text{N} \tag{2}$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{3}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{4}$$

$$\text{F1measure} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{5}$$

In the evaluation of the prediction performance of any classifier, the Receiver Operating Characteristic (ROC) curve is also often used; it represents the value of false positive on the OX axis, and on the OY axis, the value of true positive cases [75,76], so that, for example, point (0, 1) represents perfect prediction, where all the samples are classified correctly, and point (1, 0) represents a classification that classifies all samples incorrectly. Therefore, it is important to known that the output in ROC space produced from naive neural networks or Bayes classifier is a probability which is a score-numeric value and discrete classifiers produce only a single point, but in both cases they represent the degree to which a particular instance belongs to a certain class [77]. The area under the curve (AUC) is the most-used measure of diagnostic accuracy of the model and AUC values greater than 70% have good classification processes.

Practically, classification is the task of ML, but can also be the task of data mining, which performs separation of instances of a considered data set based on the value of the input variables into one of pre-determined ones class of the output variable [78].

The literature review shows that the most commonly applied classifiers include Neural networks, Bayes networks, Decision Trees, K–nearest neighbor, etc. [79].

For the proposed model, the authors used some of the most-used classification algorithms which belong to five different groups of types as it is grouped in one of the most used software for this purpose, Weka [80], i.e., Bayes, meta, trees, rules and functions. Because of that, below this subchapter is a short description of one selected algorithm from each of the mentioned Weka classifiers groups.

The Naive Bayes classifier [81,82] from the Bayes group of Weka belonging to the group of oldest classification algorithms and generates a prediction model using Bayes' theorem. It is called "naive", because of that simplifies the problem of classification by two important assumptions, first that the attributes used in the prediction procedure are conditionally independent and with a known classification, and second, that there are no

hidden attributes that could affect the prediction. In this way, these assumptions allow an efficient classification ML algorithm. For conditionally independent attributes $A_1, \dots, A_k$ probability for class attribute A is calculated using the following rule:

$$P(A_1, \dots, A_k | A) = \prod_{i=1}^{k} (A_i | A) \tag{6}$$

The main advantages of the Naive Bayes classifier in relation to other classifiers are primarily efficiency, simplicty, and convenience for small data sets of data.

The LogitBoost classifier from the meta group of Weka is widely applied in practice because it has very good characteristics, primarily thanks to the boosting algorithm [83]. This classifier uses the principle that finding multiple simple rules could be more efficient than finding a single precise rule, and because of that, usually complex prediction rules. It represents, essentially, one general method for improving the accuracy of ML algorithms.

Decision trees [84] from the trees group of Weka is the most used classification technique, because it includes more possible ways of its construction that are very convenient for interpretation. The trees can be used with all kinds of classification attributes (categorical or numerical). ID3 [85] and C4.5 [86] are the most-used algorithms from this group of classifiers and from the trees in the Weka tool, one of the most known is tree J48.

The PART classifier from the rules group of Weka builds a partial decision tree so that it uses the C4.5 decision tree classifier in each of its iterations and constructs the best sheet and a suitable rule in the tree. This classifier does not belong to the group of oft-used classifiers, but it is useful in binary classification, as applied in this paper.

SMO from the functions group of Weka refers to the specific efficient optimization algorithm used inside the support vector machines (SVM) algorithm implementation. Practically, it solves the quadratic programming problem, which arises during the training of SVM on classification tasks defined on sparse sets of data. Additionally, it is not one of the oft-used classifiers, but it is used in this paper because it is appropriate for binary classification with numerical and binary types of attributes, which is the case in this paper.

2.1.2. Logistic Regression

In ML, in many cases, probabilistic classifiers that return not only the label for the most likely class, but also the probability of that class, are needed. Such a so-called probabilistic classifier is well-calibrated if the predicted probability matches the true probability of the event which is of interest and can be checked using a calibration plot, which demonstrates how good a classifier is in a given set of data with known outcomes that is valid for the binary classifiers considered in this paper (in the case of multi-class classifiers, a separate calibration plot is needed for each of classes).

The authors used the idea of calibration, like many other authors did, as, for example, in [87], and as seen in [83], the univariate calibration using logistic regression for transforming classifier scores into probabilities of class membership for the two-class case.

The main goal of logistic regression is to obtain the best-fitting model for describing the relationship between the dichotomous characteristic of interest, which is a dependent variable (response or outcome variable) with a set of independent variables (predictor or explanatory variables).

Logistic regression generates the coefficients of a formula to predict a logit transformation of the characteristic of interest presence probability which can be notated as p (with determined standard error and significance level):

$$\text{logit}(p) = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k \tag{7}$$

The logit transformation is defined as the logged odds:

$$\text{odds} = \frac{p}{1 - p} = \frac{\text{probability of characteristics presence}}{\text{probability of characteristics absence}} \tag{8}$$

and

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) \tag{9}$$

In ordinary regression, the choosing of parameters that minimize the sum of squared errors is present, while in logistic regression it chooses parameters that maximize the likelihood of observed sample values. The regression equation coefficients are the coefficients b_0, b_1, b_2, $\ldots$ b_k. The logistic regression coefficients show increasing (when $b_i > 0$), and decreasing (when $b_i < 0$) in the predicted logged odds for the independent variables. In the case the independent variables X_a and X_b are dichotomous, then the impact of these on the dependent variable is simply determined by comparing their coefficients of regression b_a and b_b. By taking the exponent for both sides in the regression equation as it is shown above, the equation can be given as one of form of logistic regression:

$$\text{odds} = \frac{p}{1-p} = e^{b_0} \cdot e^{b_1 X_1} \cdot e^{b_2 X_2} \cdot e^{b_s X_s} \cdot \ldots \cdot e^{b_k X_k} \tag{10}$$

From the given formula, it is evident that when a variable Xi increases by 1 unit, and all other parameters remain unchanged, then the odds will increase by a parameter e^{b_i}.

$$e^{b_t(1+X_t)} - e^{b_t X_t} = e^{b_t X_t} = e^{b_t(1+X_t)-b_t X_t} = e^{b_t + b_t X_t - b_t X_t} = e^{b_t} \tag{11}$$

This factor e^{b_i} is the odds ratio (O.R.) for the independent variable X_i, and it gives the relative amount by which the odds of the outcome increase (O.R. greater than 1) or decrease (O.R. less than 1) when the value of the independent variable is increased by one unit.

Implementation of several methods for performing logistic regression can be found in statistical programs, of which IBM SPSS [88] is the most famous. This tool realizes three basic methods of binary regression and that is the enter method, stepwise method and hierarchical method. The enter method includes all the independent variables in the regression model together, stepwise methods include two categories of regression procedures-forward selection and backward elimination, and in the hierarchical method, the researcher themself determines the order of inclusion of independent variables in the model. Otherwise, all of the three methods are used to remove independent variables that are weakly correlated with the dependent variable. The authors use the standard enter method for the model proposed in this paper.

2.1.3. Future Selection Techniques

Classification methods of ML are sensitive from data dimensionality and it is showed evidently that application of dimensionality reduction enables them giving better results. Selecting a suitable subset before the application of these methods finds a set of attributes which together achieve the best result.

Algorithms for feature subset extraction perform a space search based on candidate evaluation [89]. The optimal subset is selected when the search is complete. Some of the existing evaluation measures that have been shown to be effective in removing irrelevant and redundant features include the consistency measure [90] and the correlation measure [91]. The consistency measure seeks to find the minimum number of features that consistently separate the class labels into a complete set. An inconsistency is defined for two instances that have different class labels for the same feature values.

- Future selection methods can be realized using three groups of methods [92]:
- Filter, where the most known are Relief, Infogain, Gainratio, and so on.
- Wrapper, among which the most well-known are BestFirst, RankSearch, GeneticSerch, and so on.

Embedded, which combine the qualities of the filter and wrapper methods and where, among others, ridge regression (as one technique for analyzing multiple regression data

that suffer from multicollinearity) and different types of decision tree based algorithms as BoostedTrees, RandomForest, NBTree, and so on, belong.

One of the free-to-use software that has an option that performs feature selection, reducing the amount of included attributes by applying different type algorithms, is the already-mentioned software tool Weka [80]. Because of that, this software was used to evaluate the proposed model on a selected case study. Practically, this evaluation results in determining the importance of factors that influence the risk of a traffic accident as well as for determining one prediction model using techniques such as regression and/or classification for this tasks.

Because the first two groups of methods are used in model which authors proposed in this paper these are described in short hereinafter.

Filter-Ranker Methods

Filter models rely on the general characteristics of the data to estimate the exclusion features of the learning algorithm. For some data set D, the filter algorithm starts the search by initializing a subset S1 (the empty set, the full set, or a randomly selected subset) and searches the feature value space using a specific search strategy. Each generated subset S is evaluated against an independent measure and compared to the previous best. If it is found to be better than the previous best, it is considered the current best subset. The search continues until a previously defined stopping criterion is met. The output of the algorithm is the last best subset and that is the final result. By changing the search strategy and evaluation measure, different algorithms can be implemented within the filter model. The feature selection process often uses the entropy measure as one characterization of the purity of an arbitrary collection of examples, and considers a measure of the system's unpredictability.

The entropy of Y is:

$$H(Y) = -\sum_{y \subset Y} p(y) \cdot \log_2(p(y)) \tag{12}$$

At the same time, feature selection methods differ in how they treat the problems of irrelevant as well as redundant attributes [93].

For the proposed model, authors used the following five shortly described filter algorithms.

Having in mind that the entropy could be a criterion of impurity in a training set S, it is possible to define a measure reflecting additional information about each Attribute which is generated by Class, and that is the amount by which the entropy of Attribute decreases [94]. This measure is named the information gain and, in abbreviation, is notated as InfoGain and favors variables with more values.

InfoGain evaluates the worth of an Attribute according to Class using the following formula:

$$\text{InfoGain}(\text{Class}, \text{Attribute}) = H(\text{Class}) - H(\text{Class}|\text{Attribute}) \tag{13}$$

where H is the entropy of information. The information gained about an attribute after observing class is equal to the information gained using observation in the reverse direction.

The information gain ratio, noted as GainRatio, is one so-called non-symmetrical measure that was introduced in the theory of feature selection to compensate for the bias of the already-described measure InfoGain [95]. GainRatio is one modification of the InfoGain that reduces its bias on different attributes and it is given with the following formula:

$$\text{GainRatio} = \frac{\text{InfoGain}}{H(\text{Class})} \tag{14}$$

As it is given in Formula (13), when it is needed to predict some variable-Attribute, the InfoGain is normalized so that it is divided by the entropy of Class, and in vice versa. This normalization enables that the GainRatio values must be ever in the range [0, 1]. GainRatio = 1 means that the knowledge of Class completely predicts variable-Attribute, but GainRatio = 0 indicates that there is no relation between variable-Attribute and Class.

The GainRatio favors variables with fewer values. Thus, for example, the decision tree classification algorithms C4.5 [96] and ID3 [97] use the GainRatio criterion to select the attributes that should be at every node of the tree.

FilteredAttributeEval is a classifier class for running an arbitrary evaluator on data that has been passed through an arbitrary filter which are structured based exclusively on training data. This classifier executes nominal and binary classifications with nominal, string, relational, binary, unary, as well as missing attributes.

SymmetricalUncertAttributeEval is a classifier which evaluates the worth of an attribute by measuring the symmetrical uncertainty with respect to the class.

$$\text{SymmU}(\text{Class}, \text{Attribute}) = 2 * (\text{H}(\text{Class}) - \text{H}(\text{Class}|\text{Attribute}))/\text{H}(\text{Class}) + \text{H}(\text{Attribute}) \quad (15)$$

This classifier executes nominal, binary, and classification of missing classes with nominal, binary, unary, as well as attributes.

ChiSquaredAttributeEaval is a classifier based on the chi-square test used to test the independence of two events so that, for the given data of two variables, we can obtain the observed count O and the expected count E and, using the Chi-Square measure, how expected count E and observed count O deviate from each other, which is shown in Equation (16):

$$\chi_c^2 = \sum_i \frac{(\text{O}_i - \text{E}_i)^2}{\text{E}_i} \quad (16)$$

In Equation (16) c is degrees of freedom, O_i is observed value and E_i is expected value whereby degrees of freedom refer to the total number of observations reduced by the number of independent constraints which are imposed with the observations, and having in mind definitions that the random variable follows chi-square distribution only if it can be written in the form of the sum of squared standard normal variables like it is given in Equation (17):

$$\chi^2 = \sum_i Z_i^{\,2} \quad (17)$$

where Z_i are standard normal variables.

Degrees of freedom refer to the maximum number of logically independent values, which have the freedom to vary. In simple words, it can be defined as the total number of observations minus the number of independent constraints imposed on the observations.

Wrapper Methods

In the case of these learning methods, certain modeling algorithms are used in order to evaluate subsets of attributes in relation to their classification or predictive power. It is a computationally very demanding procedure due to the frequent execution of the ML algorithm. It is practically necessary to evaluate the performance of the corresponding model for each subset of attributes, and the total number of subsets grows exponentially when the number of attributes increases. For these reasons, different search techniques are used from the group of greedy techniques, which represent an approach to solving the problem based on the best selected option available at that moment [98].

According to some of the classification frameworks [99], wrapper methods can be broadly classified according to the method of searching a set of attributes into deterministic and randomized wrapper methods. The first subgroup of wrapper methods—deterministic wrapper methods, use a complete strategy of attribute space search in one sub-subgroup and certainly give the best results with a very demanding time and sequential strategies or heuristic search in the second subgroup of deterministic wrapper methods. Another subgroup of wrapper methods consists of randomized methods, which in turn rely on stochastic search approaches. The authors chose and used five methods from this wrapper group of methods for the proposed model in this paper which are implemented in the Weka software, and those are: from the deterministic subgroup and sub-subgroup of complete strategy search the ExhaustiveSearch and sub-subgroup sequential strategies or heuristic

search—three of them—Best First, LinearForvardSelection, and GreediStepvise, and from another subgroup of wrapper methods named stohastic-GeneticSearch, and all of them with CfsSubsetEval classifier.

I. Algorithms from the group of deterministic search wrapper methods

 I.1 The first subgroup are those with full search, and these algorithms usually showed the good results.

ExhaustiveSearch is the most well-known algorithm from this subgroup; it conducts an exhaustive search through the complete space of attribute subsets starting from the empty set of attributes. On end reports the best subset found.

 I.2 The second subgroup of deterministic search wrapper methods is the group of algorithms with sequential search techniques which are the most-used wrapper algorithms, and because of that, the authors use primarily different algorithms from this subgroup in the proposed algorithm.

The BestFirst algorithm as a basic algorithm from this subgroup searches the space of attribute subsets by greedy hill climbing augmented with a backtracking facility. This algorithm may start with the empty set of attributes and search forward, or start with the full set, i.e., all attributes and search backward, or start at any point between those, and search in both directions.

The LinearForwardSelection algorithm is one Extension of the BestFirst algorithm. Takes a restricted number of k attributes into account. Fixed-set selects a fixed number k of attributes, whereas k is increased in each step when fixed-width is selected. The search uses the initial ordering to select the top k attributes, but can also use the ranking. The search direction is forward or floating forward selection with using optional backward search steps.

The Subset Size Forward Selection algorithm is one Extension of the LinearForwardSelection algorithm.

GreedyStepwise performs a greedy search in both directions, forward or backward, through the space of attribute subsets. It may start with no or all attributes, or from an arbitrary point in the space. It stops in the moment when the addition, i.e., deletion of any remaining attributes results in a decrease in evaluation. It can also produce a ranked list of attributes by traversing the space from one side to the other and recording the order that attributes are selected in.

II. Algorithms from the group of stochastic search of wrapper methods

The most-known from this group is the genetic algorithm which the authors use in the proposed algorithm as representative of that subgroup of methods. This algorithm belongs to a wider class of so-called population methods, i.e., evolutionary algorithms that use stochastic optimization. Genetic algorithms only select the initial population at random; in later steps, the selection procedure is strictly defined. The steps of the genetic algorithm are iteratively repeated until the desired target is reached value, i.e., the stopping criterion of the algorithm.

As we already mentioned, all of the wrapper methods used are applied in the proposed model with using the CfsSubsetEval classifiers (Correlation-based feature selection).This method ranks and selects the attribute sets with biases towards to subsets containing features that are highly correlated with the class, and at the same time, they are uncorrelated with each other. Measuring the significance of attributes in this method is on the basis of predictive ability of attributes and their redundancy degree.

2.1.4. Ensemble Method for Prediction of Meteorological Impact on Occurrence of TI

It is known that, in ML, methods that use several individual aggregated algorithms to achieve better results than those that would be achieved with any of the algorithms individually aggregated into it are called ensemble ML methods. To solve the predictive problem which is considered in this paper, the authors proposed an ensemble algorithm

showen with the procedure, which is given in Algorithm 1 and showed as the block schema in Figure 1.

Algorithm 1: Obtaining significant predictors of TI caused by atmospheric factors

---

1. Perform a logistic-binary regression Enter method for a model in which n atmospheric factors are predictors and the dependent variable is the number of TI logically determined by a threshold, which could be a value greater than 150% of the average value of daily TI for considered case study, and has a nominal value 1 in that case and 0 in all others. We start the algorithm in first cycle i = 1 with referent value which represents the number of attributes which is in start step number noted as n1–in concrete case study n1 = 27. In the Enter method of binary regression used, all of the predictors will be included in the prediction; only in the possible presence of impermissible collinearity of certain predictors, they will be excluded from the model. After that, using the Cox and Snell R Square and Nagelkerke R Square test, the algorithm will determine the value of the percentage of the variance that is explained, i.e., the connection between the tested factors and the dependent variable, and using the Hosmer and Lemeshow tests, the algorithm will determine its goodness-of-fit, i.e., the adaptation of the model to the given data, i.e., calibration which will evaluate the goodness of the proposed ensemble model in this and in the later steps, including the most important last step of the proposed algorithm in order to use the AUC to determine the quality measure of the classification binary regression analysis model.
2. Apply a set of at least five methods of classification which belong to different types of classification (for example, how it is already mentioned in Weka software, so any five, each from different types—Decision trees, Bayes, Meta, Rules, Functions, MI, etc.) and find two classification algorithms from this set that has the highest value of AUC among other algorithms used (also other parameters such as precision, recall, and F-measure which are with good values). That classification algorithm will be used in the step that follows in which attribute selection is carried out to select the best of several used attribute selection algorithms from two different types of groups.
The values of Hosmer and Lemeshow test and even more significant AUC values that determine the threshold of whether the desired level of goodness of the model has been reached—take the values determined in steps 1, i.e., step 2 of this algorithm, respectively.
3. Using five algorithms from each of both groups of feature selection methods is with the basic aim to use in this ensemble classification algorithms that are good and eliminate bad characteristics:
3.1. Using at least five of the mentioned attribute selection algorithms from both the wrapper and the filter groups more broadly explained in Section 2.1.3. of this paper, perform attribute classification in one class of the two possible classes of instances which are defined in step 1 of this algorithm and according to the criterion of whether the value of this attribute exceeds or does not exceed the daily TI threshold.
3.1.1. Classifiers for filter attribute selection could be any five different algorithms: for example Information-Gain Attribute evaluation, Gain-Ratio Attribute evaluation, Symmetrical Uncertainty Attribute evaluation, Chi-Square Attribute evaluation, Filtered Attribute Eval, Relief Attribute, Principal Components, etc. The authors used the first five of these in this paper. Those chosen algorithms are used to determine the feature subset of attribute $A' = \{ \ldots , a_{i-1}, a_i \}$ and their ranks from the starting set $A = \{a_1, a_2, \ldots , a_n\}$, $i \leq n$. It is necessary to remark that n is the starting number of attributes in such a way that the decision to exclude a particular attribute is made by the majority of exclusion decisions made individually by each of the algorithms.
3.1.2. Classifiers for wrapper attribute selection can be any five from this group of algorithms: for example Best First, Linear Forward Selection, Genetic Search, Greedy Stepwise, Subset Size Forward Selection, etc. The authors used the first five of these in this paper. Those chosen algorithms are used to compute a subset $A'' = \{ \ldots , a_{j-1}, a_j \}$ from the starting set $A = \{a_1, a_2, \ldots , a_n\}$, $j \leq n$. It is necessary to remark that n is the starting number of attributes in such a way that the decision to exclude a particular attribute is made by the majority of exclusion decisions made individually by each of the algorithms.
3.2. Determine a subset $A''' = A' \cap A'' = \{ \ldots , a_{m-1}, a_m \}$ from the starting set $A = \{a_1, a_2, \ldots , a_n\}$, $m \leq i, j, n$, where n is the starting number of attributes and i and j values determined in the previous steps of the algorithm 3.1.1 and 3.1.2

Algorithm 1: *Cont.*

We could have, at the end of this step, not only a different number of selected attributes using both groups of attribute selection algorithms considered as it is given in 3.1.1. and 3.1.2., possibly different notated attributes as well, and that is why we use the intersection operation for these obtained subsets A′ and A″, which determines only common attributes as those that will be removed from the initial, i.e., in later cycles from the observed set A.

3.3. If m < n exists, which is determined in the previous step 3.2., and Hosmer and Lemeshow test determined the goodness of the algorithm as positive, the algorithm continues with the next step 4 using set $A''' = \{ \dots, a_{m-1}, a_m \}$ attributes; otherwise, finish with the prediction which determined the existing number of parameters which was in the observed set.

4. Choose one from five filter classifiers with the smallest number of attributes l_i which has the highest AUC value using for that already determined two classification algorithms in step 2 of this algorithm.

5. Perform the binary regression Enter method again now with a smaller number of attributes li selected in step 4 of this algorithm, and if the values of Hosmer and Lemeshow tests are worse than those obtained in the previous test executed in step 3 of this algorithm or the obtained number of attributes satisfied value preset in advance, the procedure is finished; otherwise the procedure continues cyclically with step 3 of this algorithm with new set referent value. Preset value of the number of selected attributes on the specific need for each case separately and for the case study in this paper, the authors chose it at less than 15% from the starting number of attributes.
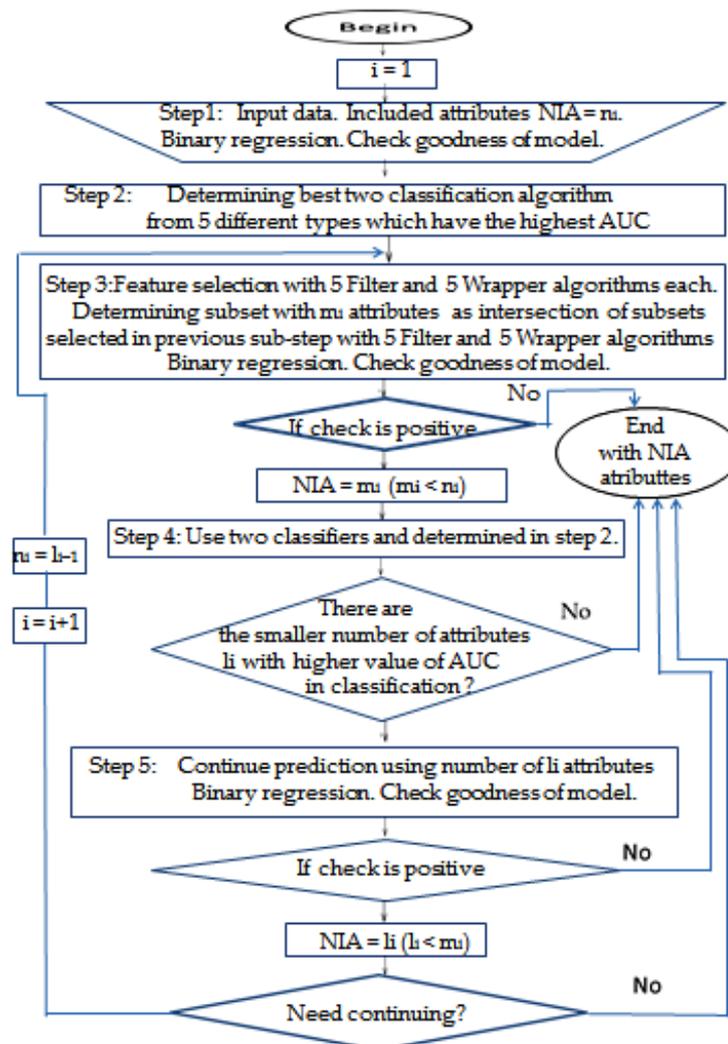

Figure 1. Block schema for the procedure which is described with Algorithm 1.

2.2. Materials

The weight coefficients determination applied in this study used the data covering the period from 1992 up to 2009 of atmospheric factors and daily traffic accidents related to the City of Niš, Republic of Serbia. The atmospheric data used in this case study is for twenty-seven variables. Data used in this study was derived from several sources. Atmospheric data was obtained from the Republic Hydro-meteorological Institute for 1992–2009, and the database of the number of daily traffic accidents for the same period was supplied by the Ministry of Interior of the Republic of Serbia. All of this data is given as a Supplementary File in which the dependent variable is given in the excel table as twenty eighths, which is shown in the table—Table 2. In order to conduct the case study more efficiently, the dates were organized on daily level in the period of eighteen years which the authors consider in the case study of this paper.

Table 2. Atmospheric parameters used in case study.

Variable	Parameter
1-V1	Air pressure at 7 o'clock (mbar)
2-V2	Air pressure at 14 o'clock (mbar)
3-V3	Air pressure at 21 o'clock (mbar)
4-V4	Mean daily air pressure (mbar)
5-V5	Maximum daily temperature (°C)
6-V6	Minimum daily temperature (°C)
7-V7	Daily temperature amplitude (°C)
8-V8	Temperature at 7 o'clock (°C)
9-V9	Temperature at 14 o'clock (°C)
10-V10	Temperature at 21 o'clock (°C)
11-V11	Mean daily temperature (°C)
12-V12	Relative humidity at 7 o'clock -percent
13-V13	Relative humidity at 14 o'clock -percent
14-V14	Relative humidity at 21 o'clock -percent
15-V15	Mean daily relative humidity-percent
16-V16	Water vapour saturation at 7 o'clock (mbar)
17-V17	Water vapour saturation at 14 o'clock (mbar)
18-V18	Water vapour saturation at 21 o'clock (mbar)
19-V19	Mean daily water vapour saturation (mbar)
20-V20	Mean daily wind speed (m/sec)
21-V21	Insolation (h)
22-V22	Cloudiness at 7 o'clock (in tenths of the sky)
23-V23	Cloudiness at 14 o'clock (in tenths of the sky)
24-V24	Cloudiness at 21 o'clock (in tenths of the sky)
25-V25	Mean daily cloudiness (in tenths of the sky)
26-V26	Snowfall (cm)
27-V27	Rainfall (mm)
28-V28	Number of daily traffic accidents

3. Results and Findings

Prediction of the impact of the meteorological factors on the appearance of traffic accidents is realized in this paper using the meteorological and the traffic factor data related to the City of Niš, Republic of Serbia. The data is for the period from 1992 up to 2009 from which twenty-seven variables are used for meteorological and one variable was available which represents the number of daily traffic accidents. Meteorological data used in this study was derived from the Republic Hydro-meteorological Institute and the database of daily traffic accidents was supplied by the Ministry of Interior of the Republic of Serbia. All variables are given in Table 2 and the case study is realized with the data attached in the excel table Mathematics-NovoTrafficAccidentsNaj1.

The authors had in mind that the basic aim of each prediction process is to create a model that, using a suitable combination of independent variables, draws conclusions for

the dependent variable. Bearing in mind the task set in the research which is the subject of this paper, we prepare the data for daily traffic accidents in binary form. As it is mentioned in this paper, the value of the dependent variable take value logic-exactly, i.e., binary-1 in the case that the number of daily traffic accidents is greater than 10, which is about 150% of the mean value for the considered period.

3.1. Application of Proposed Algorithm of Ensemble Learning

In the first step according to the steps from algorithm 1, a binary regression procedure using SPSS 17 tool [88] was carried out on the available data. All 27 meteorological parameters are used as predictors and the dichotomous variable of daily traffic accidents is used as dependable variable.

The results of applied binary regression obtained are shown in Table 3.

Table 3. Results of applied binary regression—all 27 parameters.

			Binary Regression			
	B	**S.E.**	**Wald**	**Df**	**Sig.**	**Exp (B)**
1-V1	−0.076	0.061	1.583	1	0.208	0.927
2-V2	−0.057	0.073	0.610	1	0.435	0.945
3-V3	−0.129	0.061	4.508	1	0.034	0.879
4-V4	0.265	0.148	3.199	1	0.074	1.303
5-V5	0.078	0.095	0.661	1	0.416	1.081
6-V6	−0.045	0.097	0.215	1	0.643	0.956
7-V7	−0.020	0.092	0.050	1	0.824	0.980
8-V8	−0.018	0.067	0.077	1	0.782	0.982
9-V9	−0.094	0.065	2.088	1	0.148	0.910
10-V10	−0.137	0.085	2.572	1	0.109	0.872
11-V11	0.229	0.143	2.557	1	0.110	1.257
12-V12	0.083	0.057	2.072	1	0.150	1.086
13-V13	0.092	0.058	2.527	1	0.112	1.096
14=V14	0.068	0.058	1.368	1	0.242	1.070
15-V15	−0.220	0.170	1.672	1	0.196	0.802
16-V16	0.018	0.078	0.053	1	0.817	1.018
17-V17	−0.069	0.069	1.005	1	0.316	0.933
18-V18	0.072	0.082	0.759	1	0.384	1.074
19-V19	−0.031	0.149	0.045	1	0.832	0.969
20-V20	−0.162	0.075	4.640	1	0.031	0.850
21-V21	−0.020	0.031	0.402	1	0.526	0.981
22-V22	−0.023	0.064	0.132	1	0.716	0.977
23-V23	−0.050	0.065	0.584	1	0.445	0.951
24-V24	0.028	0.063	0.194	1	0.660	1.028
25-V25	−0.004	0.182	0.000	1	0.984	0.996
Constant	0.057	0.033	3.098	1	0.078	1.059

		Classification Table [a,b]			
			Predicted		Percentage Correct
	Observed		Number of daily traffic accidents > 10		
			0	1	
Step 0	Number of daily traffic accidents > 10	0	3522	0	100.0
		1	468	0	0.0
	Overall Percentage				88.3

a. Constant is included in the model. b. The cut value is 0.500.

	Model Summary		
Step	−2 Log likelihood	Cox–Snell R Square	Nagelkerke R Square
1	2833.054 [c]	0.013	0.025

c. Estimation terminated at iteration 5 because parameter estimates changed by less than 0.001.

	Hosmer and Lemeshow Test.		
Step	Chi-square	Df	Sig.
1	12.187	8	0.143

Sig > 0.05 indicates that the data fit the model.

The result shows that the model of logistic regression using all the 27 meteorological factors monitored explains the considered problem with the 1.3 percent of variance by Cox and Snell and 2.5 by Nagelkerke, which indicates its insignificant connection with the data (bigger than 0 and less than 0.3) [100], the Hosmer and Lemeshow test value

0.143 indicates that the data fit with the model (because Sig. > 0,05); what this means is that the model is well calibrated [101] and also that the model is without excluding any of these 27 parameters because of correlation. Given that 468 instances that cause an increased number of TI and 3522 that did not are identified in the examined sample, the accuracy of the classification by random selection is (468/3990)2 + (3522/3990)2 = 0.7929, which is 79.29%, so it can be seen that the model of binary logistic regression analysis with 88.3% has a higher classification accuracy than random selection models [102]. As the quality of the model significantly determines the value of the AUC [103], the value of that measure is determined in a separate following step of the proposed model.

In the second step of the proposed Algorithm 1, five classification algorithms applied each from different types that were chosen by the authors for this purpose in this paper are Naive Bayes, J48 Decision Trees, SMO, LogitBoost, and PART algorithms. The method of 10 folds cross-validation test was applied in the model estimation. The performance indicators of five classification algorithms are given in Table 4, which shows that the LoogitBoost and classifiers achieved the most accurate prediction results especially having in mind that the most important measure is AUC value.

Table 4. Performance indicators—classification using all 27 parameters.

	Accuracy	Recall	F1 Measure	ROC
J48	0.794	0.881	0.828	0.496
Naive Bayes	0.809	0.827	0.817	0.541
Logit Boost	0.779	0.881	0.827	0.547
PART	0.815	0.882	0.829	0.524
SMO	0.779	0.883	0.828	0.500

As presented in Table 4, the LogitBoost and Naive Bayes classifiers achieved the two highest values for AUC at 0.547 and 0.541, respectively, and also the next similar values for other measures of classification, i.e., accuracy of 77.9 and 80.9%, recall 88.1 and 82.7% and F1 measure of 82.7%, and 81.7%, respectively, which implies that between these two classification algorithms, there will be one which will order predictors with highest value of AUC for the smaller number of attribute subset.

In step 3 of the proposed algorithm, the process of attribute selection by searching the attribute subsets using evaluation with two types of this method and that filter and wrapper type is realized.

3.1.1. Filter

Filter feature subset evaluation methods were conducted with a rank searching to determine the best attribute subset, and they are listed as follows:

(1)   Information-Gain Attribute evaluation(IG),
(2)   Gain-Ratio Attribute evaluation (GR),
(3)   SymmetricalUncertAttributeEval (SU),
(4)   Chi-Square Attribute evaluation (CS),
(5)   Filtered Attribute Eval (FA).

The ranks of considered parameters obtained by the above three methods on the training data are given in Table 5 where the four attributes that are selected are presented: V7, V13,V-15 and V-20.

Table 5. Feature selection using five filter ranker classifiers (smaller serial number represents bigger rank of factor).

	SU	GR	IG	CS	FA
13-V13	1/0.0063	1/0.0054	1/0.0038	1/23.13	1/0.0038
7-V7	2/0.0055	2/0.0051	2/0.0031	2/19.29	2/0.0031

Table 5. *Cont.*

	SU	GR	IG	CS	FA
20-V20	3/0.0039	3/0.0034	4/0.0023	4/11.89	4/0.0023
15-V15	4/0.0038	4/0.0029	3/0.0029	3/16.32	3/0.0029
4-V4	5/0	5/0	5/0	5/0	5/0
10-V10	6/0	6/0	9/0	9/0	9/0
3-V3	7/0	7/0	14/0	14/0	14/0
11-V11	8/0	8/0	7/0	7/0	7/0
9-V9	9/0	9/0	8/0	8/0	8/0
8-V8	10/0	10/0	10/0	10/0	10/0
2-V2	11/0	11/0	6/0	6/0	6/0
5-V5	12/0	12/0	11/0	11/0	11/0
6-V6	13/0	13/0	12/0	12/0	12/0
12-V12	14/0	14/0	13/0	13/0	13/0
27-V27	15/0	15/0	15/0	15/0	15/0
14-V14	16/0	16/0	16/0	16/0	16/0
26-V26	17/0	17/0	17/0	17/0	17/0
24-V24	18/0	18/0	18/0	18/0	18/0
25-V25	19/0	19/0	19/0	19/0	19/0
22-V22	20/0	20/0	20/0	20/0	20/0
23-V23	21/0	21/0	21/0	21/0	21/0
21-V21	22/0	22/0	22/0	22/0	22/0
16-V16	23/0	23/0	23/0	23/0	23/0
17-V17	24/0	24/0	24/0	24/0	24/0
18-V18	25/0	25/0	25/0	25/0	25/0
19-V19	26/0	26/0	26/0	26/0	26/0
1-V1	27/0	27/0	27/0	27/0	27/0

3.1.2. Wrapper

Wrapper feature subset evaluation methods were conducted without rank searching to determine the best attribute subset, and they are listed as follows:

(1)   Best First (BF),
(2)   Linear Forward Selection (LF),
(3)   Genetic Search (GS),
(4)   Greedy Stepwise (GST),
(5)   Subset Size Forward Selection (SSFS).

The obtained results presented in Table 6 shown that the same four attributes, V7,V13, V-15 and V-20, were selected using five wrapper algorithms as it was the case with five filter classifiers.

Table 6. Results of feature selection using five wrapper classifiers (symbol $\sqrt{}$ notates selection of attribute).

	BF	LF	GS	GST	SSFS
7-V7	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
13-V13	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
20-V20	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
15-V15	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

In substep 3.2. of the proposed algorithm, we determine the selected attributes as a set operation, the intersection of a subset of the selected attributes using the filter and wrapper methodology, and based on the obtained results, we notice that in our case study, we are talking about the same four attributes: V7, V13,V-15 and V-20, i.e., m = 4.

We examine the next subset 3.3 and determine that there are four selected which is less than the initial 27 attributes in the model and at the end of this substep, we check the goodness of the model used, whose results are given in Table 7.

Table 7. Results of the binary regression Enter method using the 4 selected attributes.

Binary Regression Enter Method						
	B	S.E.	Wald	Df	Sig.	Exp (B)
V7	0.025	0.017	2.039	1	0.153	1.025
V13	0.013	0.008	3.002	1	0.083	1.013
V15	0.003	0.010	0.068	1	0.794	1.003
V20	−0.184	0.078	5.601	1	0.018	0.832
Constant	0.003	0.010	0.068	1	0.794	1.003

Classification Table [a,b]				
		Predicted		
Observed		Number of daily traffic accidents > 10		Percentage Correct
		0	1	
Step 0	Number of daily traffic accidents > 10	0	3522	0	100.0
		1	468	0	0.0
	Overall Percentage				88.3

a. Constant is included in the model. b. The cut value is 0.500.

Model Summary		
Step	−2 Log likelihood	Cox–Snell R Square	Nagelkerke R Square
1	2860.403 [c]	0.006	0.012

c. Estimation terminated at iteration number 4 because parameter estimates changed by less than 0.001.

Hosmer and Lemeshow Test		
Step	Chi-square	Df	Sig.
1	11.234	8	0.189

Sig > 0.05 indicates that the data fit the model.

The result shows that the model of logistic regression is taking into consideration feature of the selected 4 meteorological parameters to explain the considered problem with the same accuracy of classification value of 88.3%, with the 0.6 percent of variance by Cox and Snell The result shows that the model of logistic regression taking into consideration selected 4 meteorological parameters explain considering problem with same accuracy of classification value of 88.3% as when uses all 27 parameters, i.e., 1.2 by Nagelkerke without excluding any of this parameters because of correlation and with Hosmer and Lemeshow test value 0.189, which is evidently better than the results obtained in step 1 when it was used for all 27 attributes in regression model.

Additionally, we can see in Table 8 that the classification measure values for determined two best classification algorithms have better characteristics than results obtained in step 2 of this algorithm presented in Table 4.

Table 8. Performance indicators obtained by the classification algorithms using 4 parameters.

	Accuracy	Recall	F1 Measure	ROC
Naive Bayes	0.809/0.779	0.827/0.883	0.817/0.828	0.541/0.565
Logit Boost	0.779/0.897	0.881/0.883	0.827/0.828	0.547/0.610

Results given in Tables 7 and 8 clearly show that the applied two groups of filter and wrapper methodologies with five specific algorithms each with reduced dimension from 27 parameters to only four attributes, i.e., variables show good results of correctness of such a reduced model, and because of that, we can continue with step 4 of the proposed algorithm 1; otherwise, that would be the end and exit from the procedure with undone dimensionality reduction.

In step 4 of proposed algorithm 1, we generate a diagram with AUC values depending on the number of attributes used, for the best classification algorithm Loogitboost which is determined on the basis of the results given in Tables 4 and 8 and on the basis of results for each from the chosen five filter classifiers given in Table 5. The x-axis shows the number of attributes, and the y-axis shows the AUC value of feature subset generated for each of the five filter classifiers. In this way, we determine if the best results for the AUC measure we

can obtain with a decreasing number of attributes, in our case the number of four attributes determined in step 3 and that using the ranking of the subset of attributes obtained with SymmetricalUncertAttributeEval classifier where it should be noted that the GainRatio classifier gives the same ranking of attributes. The rank of each of the 27 attributes which is obtained with SU and GR classifier determines the order of elimination of each one individually starting attribute and begins with the one with the lowest 27th rank and a suitable value of AUC determined using the LogitBoost classification algorithm. At the end in the diagram shown in Figure 2., it is clearly presented that three is the minimal number of used attributes with the maximal value of AUC for the number of attributes smaller than the four determined in the previous third step of the algorithm, also taking into account other classification measures, and in this way, determines the definitively chosen feature subset.
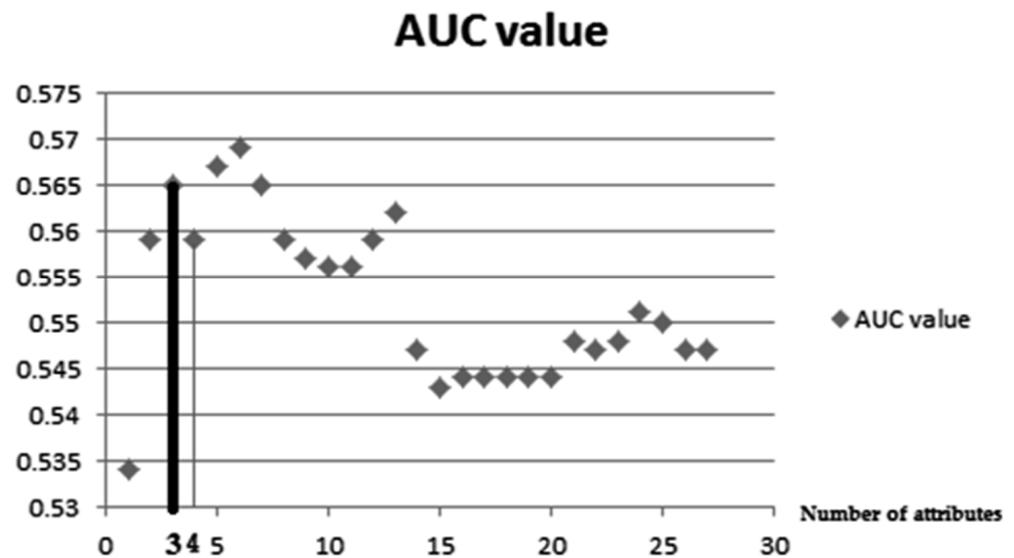
Figure 2. Diagram for determining maximum AUC value of classification for minimum number of attributes.

As we can conclude using results from the Diagram presented in Figure 2 and results determined with the best filter classifiers SU, i.e., GR given in Table 5, in this step of the algorithm, we obtain an added decrease of the selected attributes which will be included in the prediction formula in the following three: V13-Relative humidity at 14 o'clock in percent, V7-Daily temperature amplitude in °C and V20-Mean daily wind speed in m/sec.

The LogitBoost algorithm of classification shows, evidently, the best results in each of the measures including the AUC value for a reduced number of the three attributes mentioned as it is given in Table 9.

Table 9. Result evaluation of LogitBoost classification using all 27, 4 and 3 parameters.

	Accuracy	Recall	F1 Measure	ROC
27 parameters	0.779	0.881	0.827	0.547
4 parameters	0.897	0.883	0.828	0.610
3 parameters	0.897	0.883	0.828	0.613

In the last step 5 of Algorithm 1, a logistic regression is carried out, as in steps 1 and 3, to check the goodness of the model with 3 parameters selected in the previous step 4, and the results are given in Table 10.

Table 10. Results of applied logic regression with the selected subset of 3 parameters.

Binary Regression						
	B	S.E.	Wald	Df	Sig.	Exp (B)
V7	0.025	0.017	2.120	1	0.145	1.025
V13	0.015	0.004	11.002	1	0.001	1.015
V20	−0.191	0.073	6.877	1	0.009	0.826
Constant	−2.896	0.457	40.170	1	0.000	0.055

Classification Table [a,b]

			Predicted		Percentage Correct	
	Observed		Number of daily traffic accidents > 10			
			0	1		
	Number of daily	0	3522	0		100.0
Step 0	traffic accidents > 10	1	468	0		0.0
	Overall Percentage					88.3

a. Constant is included in the model. b. The cut value is 0.500.

Model Summary

Step	−2 Log likelihood	Cox–Snell R Square	Nagelkerke R Square
1	2860.472 [c]	0.006	0.012

c. Estimation terminated at iteration number 4 because parameter estimates changed by less than 0.001.

Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	10.469	8	0.234

Sig > 0.05 indicates that the data fit the model.

The result shows that the model of logistic regression taking in consideration the three selected meteorological parameters explain the considered problem with the 0.6 percent of variance by Cox and Snell, i.e., 1.2 percent of variance by Nagelkerke without excluding any of these parameters because of correlation and with Hosmer and Lemeshow test value 0.234 which is evidently better than the results obtained in step 3 when it was used with four attributes in the regression model. Because of that, we can continue with step 3 of the proposed algorithm 1 to check eventual further decreasing of attributes; otherwise, it would be the end and exit from the procedure with dimensionality reduction done to this moment. However, in the end of this last step of the proposed algorithm, before continuing the algorithm with step 3, it is obligatory to check the case that it is obtained value preset in advance. This is the case in our paper, because the reduced number on three important attributes is smaller than the preset threshold value which is four attributes, so this fact implies the exit from the procedure in our case study.

For the concrete considered case study in this paper, the predictive formula is as follows:

$$-2.896 + 0.025V7 + 0.015V13 - 0.191V20 > 10 \tag{18}$$

3.2. The Model of Emergent Intelligence as One Implementation of the Proposed Ensemble Method

As the authors had already mentioned in the introduction of Section 2. Materials and methods, in this paper they described not only the new proposed model but also its implementation as one of the agents in one multi-agent system of the emergent intelligence technique (EIT) for the purpose of one citizens warning system—Figure 3. In this respect, let us mark the task of giving a warning to those from one region or big city interested in the meteorological parameters that has reached the existence of conditions which affect the increased possibility of traffic accidents in this concrete place with T. The task is performed on the basis of measuring the values of all parameters included in the proposed model in this paper with real time and obtaining historical data of those values from specialized electronic data sources. In carrying out the set task, it is obligatory to use suitable prediction models as well as the proposed model in Section 2.1.4 of this paper and the data in real time.
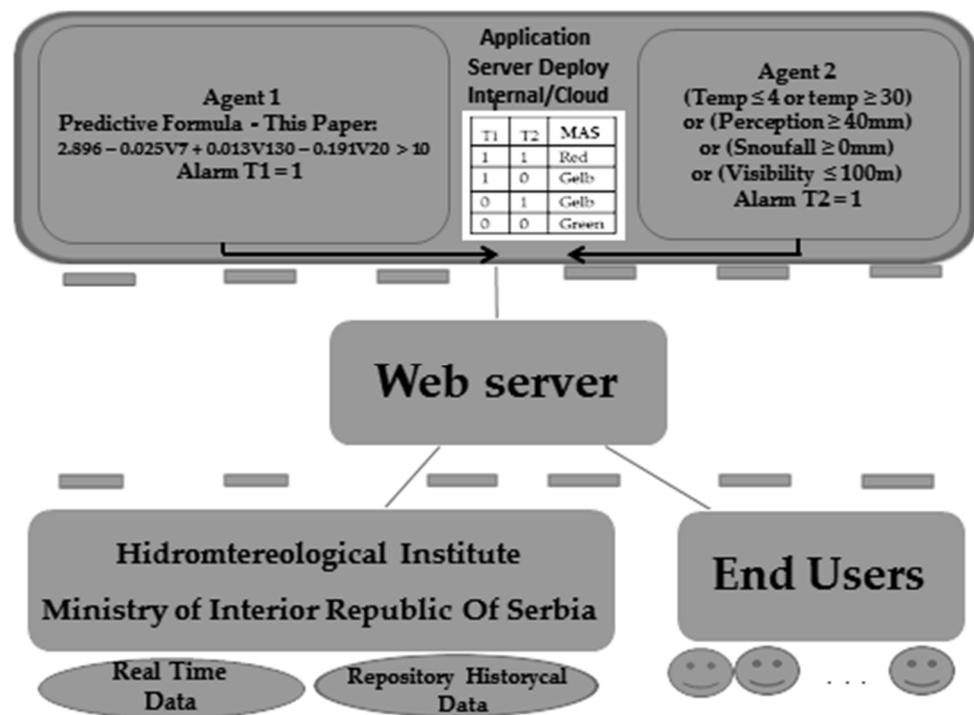
Figure 3. EIT two-agent system for generating a warning of the possibility of traffic accidents.

That is why we divide the set task T into 2 subtasks for the model of the two-agent EIT system, and these would be the tasks: T1, which determines the warning of the existence and possibilities of increased traffic accidents based on a prediction from historical data, already described in Section 2.1.4 using the proposed ensemble model of ML and prediction formula given with Equation (18), and subtask T2, which determines the existence of that possibility based on the given exceeding or undershooting pre-set values for some of the most important meteorological parameters in real time like Temperature (≤ 4 or ≥ 30), Precipitation (≥ 40 mm), Snowfall (≥ 0 mm), and Visibility (≤ 100 m). In the proposed EIT, the two-agent system in Figure 3, the decision matrix realizes one warning alarm which is in the node of EIT where the main task T is solving, using the already-solved agents tasks T1 and T2, and this matrix is given in Table 11 on a way to generate the red alarm in the case that both agents T1 and T2 give a warning; the yellow alarm is generated if only one of them gives a warning, while there is no warning if neither of them gives a warning.

Table 11. Decision matrix of EIT for generating a warning of the possibility of traffic accidents.

T1	T2	EITalarm
1	1	Red
1	0	Gelb
0	1	Gelb
0	0	Green

4. Technological Implementation of the Proposed Ensemble Model

The technical implementation of the proposed solution implies the implementation of the considered two-agent system EIT with additional indication of the possibility in the future for different implementations in a more complex and multidimensional agent system, and some specific types of parameters that could and should be included in such a system are listed. The proposed technical solution is considered through two subsections in this paper—architecture and implementation.

4.1. The Architecture of the Proposed Technical Solution

The proposed technical solution is client server architecture which uses Firebase as the cloud messenger service in the proposed solution, and can also be used as a real-time database in the Backend-as-a-Service application development platform. In this architecture, Firebase connects user applications from client side with the server application on the server side consisting of four modules noted as Agent 1, Agent 2, EIT and notification module.

The user application works on the client side in this solution. The user application, which is the client application, works with different mobile operating platforms such as IOS and android, android auto, Google assistant driving mode; the same story with Apple devices and car-play systems, and the authors realized it in the proposed solution on IOS. During the installation, the application requests permission from the user to track the location in the background. Then, the server application, specifically the notification module, requests a list of hydro-meteorological stations with their geo locations, as well as data on topic names for the defined alarms. Since the topic, among other mentioned parameters, is made up of the name of the hydro-meteorological station, the nearest station is determined based on the current geolocation of the user and the geolocation of the hydro-meteorological stations. After selecting a hydro-meteorological station, the user fills in information about the type of alarm he is interested in, and more precisely, what type of vehicle they drive and whether they wear glasses. Based on this data, a topic is created to which the user application logs on to Firebase. Furthermore, the application monitors the change of location, and with each change, determines whether there is a station that is closer than the currently selected one; when this happens, the application logs out of the previous topic and logs in to the new one. Additionally, in the case when the notification module sends a notification for a topic for which the client has registered, that notification is displayed to the user.

The notification module serves to provide the client application with data about hydro-meteorological stations and their locations, as well as other options for determining the topic. Additionally, when the EIT module from the server determines this, this module addresses Firebase and forwards a notification to all users logged in to the topic defined.

Agents 1 has a database of historical data that it uses to generate an alarm according to the prediction formula that is generated by the proposed prediction model from this paper that takes meteorological conditions into account. In this way, it decides whether to raise an alarm notated T1. Thereby, the historical data is updated by the Hydro-meteorological Institute of the Republic of Serbia and from the Ministry of Interior of Serbia, the number and place of the accident-city, i.e., the number of roads. The data is given to clients and official members of the MUP, in which case the EIT generates a report that includes cases of binoculars and not both truck and car and gives such a report to an official person—that is, in all four variants.

Agents 2 decides whether to raise the alarm based on the defined rules and the current situation. Agent 2 generates an alarm T2 in the logical function of meeting the meteorological conditions of the given conditions in the image for temperature, rain, snow, wind, and fog, as well as the type of vehicle (truck or car) and visibility.

The EIT module addresses the agents at a defined time and takes answers from them. Based on the answers received, it forwards an alarm to the notification module for groups that need it, which then forwards notifications to Firebase.

4.2. The Implementation of Proposed Technical Solution

The implementation of the proposed technical solution, which is based on the diagram given in Figure 4, is realized with attached program codes for each part included in the proposed EIT system separately with server-TrafficIncidents-master and clients application-TrafficAccidentPrevention. Software implementation of the proposed technical solution is realized using Python as a widespread software platform (see Algorithm 2).
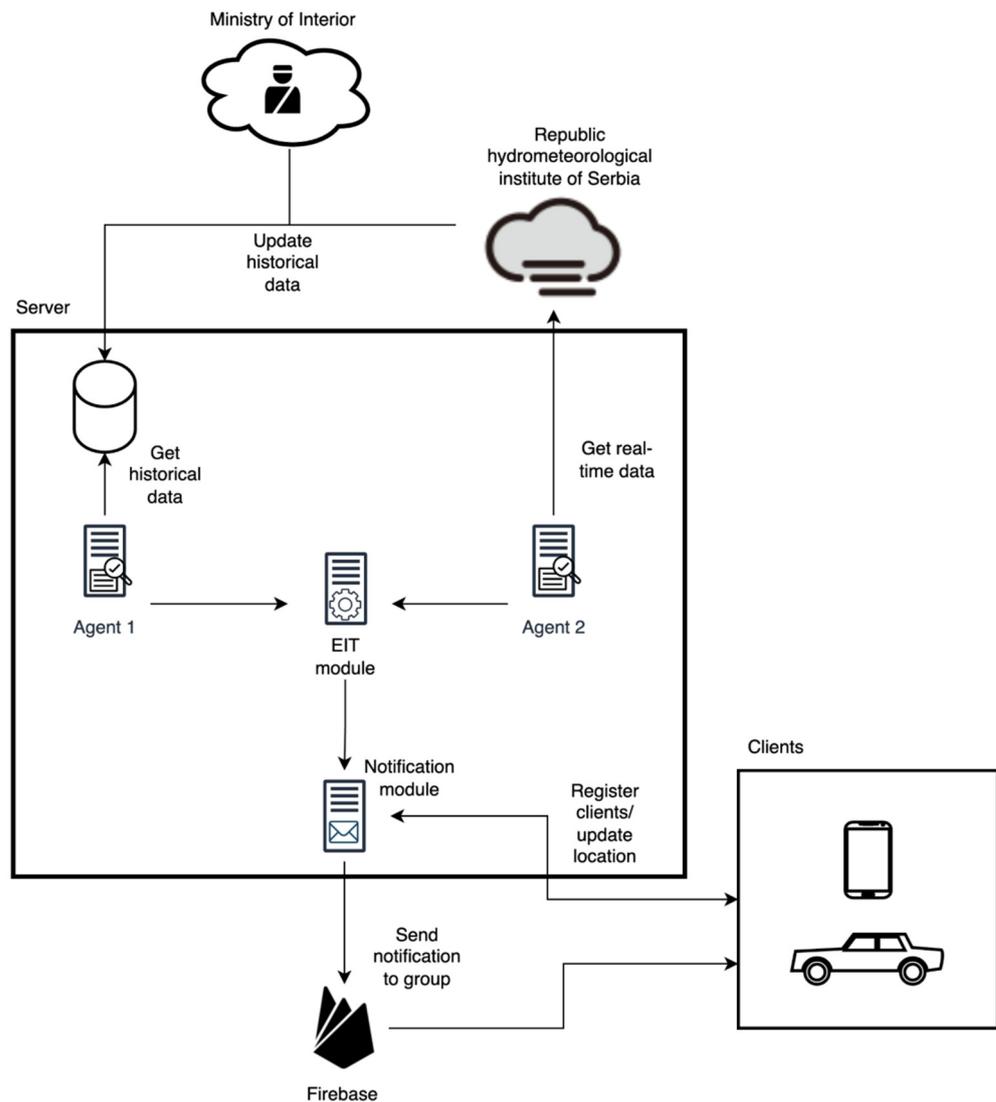
Figure 4. EIT two-agent system implementation—generating a warning of the possibility of traffic accidents.

Implementation of the Agent 1 that generates alarm T1

Data:

V7 is the daily temperature amplitude in degrees Celsius

V13 Relative humidity at 14 o'clock in percent

V20 is Mean daily wind speed m/sec

if ($-2.896 - 0.025v7 + 0.013v13 - 1.191v20 > 10$)

Alarm T1 = 1

else

Alarm T1 = 0

Implementation of the Agent 2 algorithm that generates alarm T2

Data on the current hydro-meteorological situation:

temp-current temperature INT

fog-presence of fog BOOLEAN

wind-wind speed in m/s INT

snowfall-is it snowing BOOLEAN

rain-is it raining BOOLEAN

cloudiness-is it cloudy BOOLEAN

User data:

tracks-whether they drive a truck or a bus BOOLEAN

cars-do they drive a car BOOLEAN
farsightedness-whether they wear glasses while driving BOOLEAN
if (
(temp ≤ 4)
or (temp ≥ 30)
or (fog and farsightedness)
or (rain)
or (wind ≥ 50 and tracks)
or (wind ≥ 65 and cars)
or (snowfall)
or (cloudiness and farsightedness)
)
Alarm T2 = 1
else
Alarm T2 = 0

Algorithm 2: Implementation of the EIT algorithm that generates alarm EITalarm

---

T1 and T 2 agent alarms
f (T1 = 1 and T2 = 1)
Red alarm
else
if (T1 = 1 or T2 = 1)
Yellow alarm
else
Green alarm-no alarm

---

5. Conclusions

The authors had two main aims in this paper which was directly connected with proving two set hypotheses. The results of the research with the proposed ensemble method of aggregation of five methods from different classification groups of algorithms and binary regression algorithm confirmed the first set hypothesis. It could be concluded that it is possible to aggregate several classification methods and include several feature selection methods into one ensemble method with better characteristics than each individually installed method when it is applied alone to solve the same task. Thereby, each used classification methods of ML belongs to a different type of classification algorithms, and also, each algorithm of attribute reduction belongs to different types of feature selection algorithms. The authors also gave an answer on the second hypothesis set and the question: Is it possible for such potential obtained ensemble method to be implemented in one multi agent system? They did it in a way that they proposed one technological system supported with emergence intelligence as one good framework for the implementation of the proposed model defined with the algorithm described.

The authors confirmed those two hypotheses using the results obtained in the case study conducted for the data for the City of Niš in the Republic of Serbia and these were evaluated using a 10-fold cross validation for each of the applied algorithms in Weka software.

The authors have claimed that the proposed model has not demonstrated significant limitations. The authors will deal with it by examining the inclusion of a greater number of types of classification groups and feature selection algorithms and the inclusion of n-modular redundancy into the construction of the proposed ensemble algorithm in their future work related to this topic. Moreover, the authors will also consider the implementation of the proposed model in multi agent systems with more than two included agents based on the emergence of intelligence technology and for obtaining better prediction models for TI for solving similar prediction problems in different fields of human life.

References

1.　Gao, J.; Chen, X.; Woodward, A.; Liu, X.; Wu, H.; Lu, Y.; Li, L.; Liu, Q. The association between meteorological factors and road traffic injuries: A case analysis from Shantou city, China. *Sci. Rep.* **2016**, *6*, 37300. [CrossRef]

2.　Verster, T.; Fourie, E. The good, the bad and the ugly of South African fatal road accidents. *S. Afr. J. Sci.* **2018**, *114*. [CrossRef] [PubMed]

3.　Lankarani, K.B.; Heydari, S.T.; Aghabeigi, M.R.; Moafian, G.; Hoseinzadeh, A.; Vossoughi, M.J. The impact of environmental factors on traffic accidents in Iran. *Inj Violence Res.* **2014**, *6*, 64–71. [CrossRef]

4.　Dastoorpoor, M.; Idani, E.; Khanjani, N.; Goudarzi, G.; Bahrampour, A. Relationship Between Air Pollution, Weather, Traffic, and Traffic-Related Mortality. *Trauma Mon.* **2016**, *21*, e37585. [CrossRef]

5.　Chekijian, S.; Paul, M.; Kohl, V.P.; Walker, D.M.; Tomassoni, A.J.; Cone, D.C.; Vaca, F.E. The global burden of road injury: Its relevance to the emergency physician. *Emerg. Med. Int.* **2014**, *2014*, 139219. [CrossRef] [PubMed]

6.　Xie, S.H.; Wu, Y.S.; Liu, X.J.; Fu, Y.B.; Li, S.S.; Ma, H.W.; Zou, F.; Cheng, J.Q. Mortality from road traffic accidents in a rapidly urbanizing Chinese city: A 20-year analysis in Shenzhen, 1994–2013. *Traffic Inj. Prev.* **2016**, *17*, 3943. [CrossRef]

7.　Chan, C.T.; Pai, C.W.; Wu, C.C.; Hsu, J.C.; Chen, R.J.; Chiu, W.T.; Lam, C. Association of Air Pollution and Wheather Factors with Traffic Injuri Severity: A Study in Taiwan. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7442. [CrossRef] [PubMed]

8.　Jalilian, M.M.; Safarpour, H.; Bazyar, J.; Keykaleh, M.S.; Malekyan, L.; Khorshidi, A. Environmental Related Risk Factors to Road Traffic Accidents in Ilam, Iran. *Med. Arch.* **2019**, *73*, 169–172. [CrossRef]

9.　Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.

10.　Lu, H.; Zhu, Y.; Shi, K.; Yisheng, L.; Shi, P.; Niu, Z. Using Adverse Weather Data in Social Media to Assist with City-Level Traffic Situation Awareness and Alerting. *Appl. Sci.* **2018**, *8*, 1193. [CrossRef]

11.　Trenchevski, A.; Kalendar, M.; Gjoreski, H.; Efnusheva, D. Prediction of Air Pollution Concentration Using Weather Data and Regression Models. In Proceedings of the 8th International Conference on Applied Innovations in IT, (ICAIIT), Koethen (Anhalt), Germany, 9 March 2020; pp. 55–61.

12.　Gupta, A.; Sharma, A.; Goel, A. Review of Regression Analysis Models. *Int. J. Eng. Res. Technol.* **2017**, *6*, 58–61.

13.　Theofilatos, A.; Yannis, G. A review of the effect of traffic and weather characteristics on road safety. *Accid. Anal. Prev.* **2014**, *72*, 244–256. [CrossRef]

14.　Qiu, L.; Nixon, W.A. Effects of adverse weather on traffic crashes: Systematic review and meta-analysis. *Transp. Res. Rec.* **2008**, *2055*, 139–146. [CrossRef]

15.　Edwards, J.B. The relationship between road accident severity and recorded weather. *J. Saf. Res.* **1998**, *29*, 249–262. [CrossRef]

16.　Malin, F.; Norros, I.; Innamaa, S. Accident risk of road and weather conditions on different road types. *Accid. Anal. Prev.* **2019**, *122*, 181–188. [CrossRef] [PubMed]

17.　Brijs, T.; Offermans, C.; Hermans, E.; Stiers, T. The Impact of Weather Conditions on Road Safety Investigated on an HourlyBasis. In Proceedings of the Transportation Research Board 85th Annual Meeting, Washington, DC, USA, 22–26 January 2006.

18.　Antoniou, C.; Yannis, G.; Katsochis, D. Impact of meteorological factors on the number of injury accidents. In Proceedings of the 13th World Conference on Transport Research (WCTR 2013), Rio de Janeiro, Brazylia, 15–18 July 2013; Volume 15.

19.　El-Basyouny, K.; Barua, S.; Islam, M.T. Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. *Accid. Anal. Prev.* **2014**, *73*, 91–99. [CrossRef]

20.　Baker, C.; Reynolds, S. Wind-induced accidents of road vehicles. *Accid. Anal. Prev.* **1992**, *24*, 559–575. [CrossRef]

21.　Naik, B.; Tung, L.W.; Zhao, S.; Khattak, A.J. Weather impacts on single-vehicle truck crash injury severity. *J. Saf. Res.* **2016**, *58*, 57–65. [CrossRef] [PubMed]

22.　Mitra, S. Sun glare and road safety: An empirical investigation of intersection crashes. *Saf. Sci.* **2014**, *70*, 246–254. [CrossRef]

23.　Hagita, K.; Mori, K. The effect of sun glare on traffic accidents in Chiba prefecture, Japan. *Asian Transp. Stud.* **2014**, *3*, 205–219. [CrossRef]

24. Buisán, S.T.; Earle, M.E.; Collado, J.L.; Kochendorfer, J.; Alastrué, J.; Wolff, M.; Smith, C.D.; López-Moreno, J.I. Assessment of snowfall accumulation underestimation by tipping bucket gauges in the Spanish operational network. *Atmos. Meas. Tech.* **2017**, *10*, 1079–1091. [CrossRef]

25. Lio, C.F.; Cheong, H.H.; Un, C.H.; Lo, I.L.; Tsai, S.Y. The association between meteorological variables and road traffic injuries: A study from Macao. *PeerJ* **2019**, *7*, e6438. [CrossRef] [PubMed]

26. Khan, G.; Qin, X.; Noyce, D. Spatial Analysis of Weather Crash Patterns. *J. Transp. Eng.* **2008**, *134*, 191–202. [CrossRef]

27. Becker, N.; Rust, H.W.; Ulbrich, U. Predictive modeling of hourly probabilities for weather-related road accidents. *Nat. Hazards Earth Syst. Sci.* **2020**, *20*, 2857–2871. [CrossRef]

28. Song, X.; Zhao, X.; Zhang, Y.; Li, Y.; Yin, C.; Chen, J. The effect of meteorological factors on road traffic injuries in Beijing. *Appl. Ecol. Environ. Res.* **2019**, *17*, 9505–9514. [CrossRef]

29. Matthew, G.K.; Yannis, G. Weather Effects on Daily Traffic Accidents and Fatalities: Time Series Count Data Approach. In Proceedings of the Transportation Research Board 89th Annual Meeting, Washington, DC, USA, 10–14 January 2010; p. 17.

30. Bergel-Hayat, R.; Depire, A. Climate, road traffic and road risk—An aggregate approach. In Proceedings of the 10th WCTR (World Conference on Transport Research Society), Istanbul, Turkey, 4–8 July 2004.

31. Bergel-Hayat, R.; Debbarh, M.; Antoniou, C.; Yannis, G. Explaining the road accident risk: Weather effects. *Accid. Anal. Prev.* **2013**, *60*, 456–465. [CrossRef]

32. Zheng, L.; Lin, R.; Wang, X.; Chen, W. The Development and Application of Machine Learning in Atmospheric Environment Studies. *Remote Sens.* **2021**, *13*, 4839. [CrossRef]

33. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [CrossRef]

34. Medina-Salgado, B.; Sánchez-DelaCruz, E.; Pozos-Parra, P.; Sierra, J.E. Urban traffic flow prediction techniques: A review. *Sustain. Comput. Inform. Syst.* **2022**, *35*, 100739. [CrossRef]

35. Shaik, M.E.; Islam, M.M.; Hossain, Q.S. A review on neural network techniques for the prediction of road traffic accident severity. *Asian Transp. Stud.* **2021**, *7*, 100040. [CrossRef]

36. Moghaddam, F.R.; Afandizadeh, S.; Ziyadi, M. Prediction of accident severity using artificial neural networks. *Int. J. Civ. Eng.* **2011**, *9*, 41–49.

37. Pradhan, B.; Sameen, M.I. Review of traffic accident predictions with neural networks. In *Laser Scanning Systems in Highway and Safety Assessment, Technology & Innovation (IEREK Interdisciplinary Series for Sustainable Development)*; Springer: Cham, Switzerland, 2020; pp. 97–109.

38. Profillidis, V.A.; Botzoris, G.N. Chapter 8—Artificial intelligence—Neural network methods. In *Modeling of Transport Demand Analyzing, Calculating, and Forecasting Transport Demand*; Elsevier: St. Louis, MO, USA, 2019; pp. 353–382. [CrossRef]

39. Yuan, J.; Abdel-Aty, M.; Gong, Y.; Cai, Q. Real-time crash risk prediction using long short-term memory recurrent neural network. *Transport. Res. Rec. J. Transport. Res. Board* **2019**, *2673*, 1–13. [CrossRef]

40. Rezapour, M.; Nazneen, S.; Ksaibati, K. Application of deep learning techniques in predicting motorcycle crash severity. *Eng. Rep.* **2020**, *2*, e12175. [CrossRef]

41. Sameen, M.I.; Pradhan, B.; Shafri, H.Z.M.; Hamid, H.B. Applications of deep learning in severity prediction of traffic accidents. In *Global Civil Engineering Conference*; Springer: Singapore, 2019; pp. 793–808.

42. Zheng, M.; Li, T.; Zhu, R.; Chen, J.; Ma, Z.; Tang, M.; Cui, Z.; Wang, A.Z. Traffic accident's severity prediction: A deep-learning approach-based CNN network. *IEEE Access* **2019**, *7*, 39897–39910. [CrossRef]

43. Sameen, M.I.; Pradhan, B. Severity prediction of traffic accidents with recurrent neural networks. *Appl. Sci.* **2017**, *7*, 476. [CrossRef]

44. Soto, B.G.; Bumbacher, A.; Deublein, M.; Adey, B.T. Predicting road traffic accidents using artificial neural network models. *Infrastruct. Asset Manag.* **2018**, *5*, 132–144. [CrossRef]

45. Ebrahim, S.; Hossain, Q.S. An Artificial Neural Network Model for Road Accident Prediction: A Case Study of Khulna Metropolitan City, Bangladesh. In Proceedings of the Fourth International Conference on Civil Engineering for Sustainable Development (ICCESD 2018), Khulna, Bangladesh, 9–11 February 2018; KUET: Khulna, Bangladesh, 2018.

46. Jadaan, K.S.; Al-Fayyad, M.; Gammoh, H.F. Prediction of road traffic accidents in Jordan using artificial neural network (ANN). *J. Traffic Log. Eng.* **2014**, *2*, 92–94. [CrossRef]

47. Moslehi, S.; Gholami, A.; Haghdoust, Z.; Abed, H.; Mohammadpour, S.; Moslehi, M.A. Predictions of traffic accidents based on wheather coditions in Gilan provice using artificial neuran network. *J. Health Adm.* **2021**, *24*, 67–78.

48. Liu, Y. Weather Impact on Road Accident Severity in Maryland. Ph.D. Thesis, Faculty of Graduate School, Maryland University, College Park, MD, USA, 2013.

49. Zou, X. Bayesian network approach to causation analysis of road accidents using Netica. *J. Adv. Transp.* **2017**, *2017*, 2525481. [CrossRef]

50. Ogwueleka, F.N.; Misra, S.; Ogwueleka, T.C.; Fernandez-Sanz, L. An artificial neural network model for road accident prediction: A case study of a developing country. *Acta Polytech. Hung.* **2014**, *11*, 177–197.

51. Kunt, M.M.; Aghayan, I.; Noii, N. Prediction for traffic accident severity: Comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods. *Transport* **2011**, *26*, 353–366. [CrossRef]

52. Taamneh, M.; Taamneh, S.; Alkheder, S. Clustering-based classification of road traffic accidents using hierarchical clustering and artificial neural networks. *Int. J. Inj. Control Saf. Promot.* **2017**, *24*, 388–395. [CrossRef]

53. Ghasedi, M.; Sarfjoo, M.; Bargegol, I. Prediction and Analysis of the Severity and Number of Suburban Accidents Using Logit Model, Factor Analysis and Machine Learning: A case study in a developing country. *SN Appl. Sci.* **2021**, *3*, 13. [CrossRef]

54. Mondal, A.R.; Bhuiyan, M.A.; Yang, F. Advancement of weather-related crash prediction model using nonparametric machine learning algorithms. *SN Appl. Sci.* **2020**, *2*, 1372. [CrossRef]

55. Liang, M.; Zhang, Y.; Yao, Z.; Qu, G.; Shi, T.; Min, M.; Ye, P.; Duan, L.; Bi, P.; Sun, Y. Meteorological Variables and Prediction of Road Traffic Accident Severity in Suzhou city of Anhui Province of China. 2020. Available online: https://www.researchgate.net/publication/340197416MeteorologicalVariables_and_Prediction____Road_Traffic_Accident_Severity_in_Suzhou_city_of_Anhui_Province_of_China (accessed on 20 November 2022). [CrossRef]

56. Olutayo, V.A.; Eludire, A.A. Traffic accident analysis using decision trees and neural networks. *Int. J. Inf. Technol. Comput. Sci.* **2014**, *6*, 22–28.

57. Silva, H.C.E.; Saraee, M.H. Predicting road traffic accident severity using decision trees and time-series calendar heat maps. In Proceedings of the 6th IEEE Conference on Sustainbility Utilization and Development in Engineering and Technology, Penang, Malaysia, 7–9 November 2019.

58. Chong, M.; Abraham, A.; Paprzycki, M. Traffic Accident Analysis Using Decision Trees and Neural Networks. *arXiv* **2004**, arXiv:cs/0405050.

59. Bahiru, T.K.; Kumar Singh, D.; Tessfaw, E.A. Comparative study on Data Mining Classification Algorithms for Predicting Road Traffic Accident Severity. In Proceedings of the 2018 Second Inernational Conference on Inventive Communication and Computational Technologies ICICCT, Coimbatore, India, 20–21 April 2018; pp. 1655–1660. [CrossRef]

60. Al-Turaiki, I.; Aloumi, M.; Aloumi, N.; Alghamdi, K. Modeling traffic accidents in Saudi Arabia using classification techniques. In Proceedings of the 2016 4th Saudi International Conference on Information Technology (Big data aNALYSIS) KACSTIT, Ryadh, Saudi Arabia, 6–9 November 2016; pp. 1–5. [CrossRef]

61. Lepperod, A.J. Air Quality Prediction with Machine Learning. Master's Thesis, Norwegian University of Science and Technology, Oslo, Norway, 2019.

62. Dong, S.; Khattak, A.; Ullah, I.; Zhou, J.; Hussain, A. Predicting and Analyzing Road Traffic Injury Severity Using Boosting-Based Ensemble Learning Models with SHAPley Additive exPlanations. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2925. [CrossRef] [PubMed]

63. Kim, J.H.; Kim, J.; Lee, G.; Park, J. Machine Learning-Based Models for Accident Prediction at a Korean Container Port. *Sustainability* **2021**, *13*, 9137. [CrossRef]

64. Gutierrez-Osorio, C.; González, F.A.; Pedraza, C.A. Deep Learning Ensemble Model for the Prediction of Traffic Accidents Using Social Media Data. *Computers* **2022**, *11*, 126. [CrossRef]

65. Yuexu, Z.; Wei, D. Prediction in Traffic Accident Duration Based on Heterogeneous Ensemble Learning. *Appl. Artif. Intell.* **2022**, *36*, 2018643. [CrossRef]

66. Chang, L.Y.; Wang, H.W. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accid. Anal. Prev.* **2006**, *38*, 1019–1027. [CrossRef] [PubMed]

67. Yang, G.; Wang, Y.; Yu, H.; Ren, Y.; Xie, J. Short-Term Traffic State Prediction Based on the Spatiotemporal Features of Critical Road Sections. *Sensors* **2018**, *18*, 2287. [CrossRef]

68. Li, G.; Knoop, V.L.; Van Lint, H. Estimate the limit of predictability in short-term traffic forecasting: An entropy-based approach. *Transp. Res. Part C Emerg. Technol.* **2022**, *138*, 103607. [CrossRef]

69. Min, W.; Wynter, L. Real-time road traffic prediction with spatio-temporal correlations. *Transp. Res. Part C Emerg. Technol.* **2011**, *19*, 606–616. [CrossRef]

70. Paz, A.; Veeramisti, N.; De la Fuente-Mella, H. Forecasting Performance Measures for Traffic Safety Using Deterministic and Stochastic Models. In Proceedings of the IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain, 15–18 September 2015; pp. 2965–2970. [CrossRef]

71. Pang, Y.; Zhao, X.; Yan, H.; Liu, Y. Data-driven trajectory prediction with weather uncertainties: A bayesian deep learning approach. *Transp. Res. Part C Emerg. Technol.* **2021**, *130*, 103326. [CrossRef]

72. Pang, Y.; Zhao, X.; Hu, J.; Yan, H.; Liu, Y. Bayesian spatio-temporal graph transformer network(b-star) for multi-aircraft trajectory prediction. *Knowl. Based Syst.* **2022**, *249*, 108998. [CrossRef]

73. Pang, Y.; Guo, Z.; Zhuang, B. Prospectnet: Weighted conditional attention for future interaction modeling in behavior prediction. *arXiv* **2022**, arXiv:2208.13848.

74. Romero, C.; Ventura, S.; Espejo, P.; Hervas, C. Data mining algorithms to classify students. Proceedings for the 1st IC on Educational Data Mining (EDM08), Montreal, QC, Canada, 20–21 June 2008; pp. 20–21.

75. Fawcett, T. *ROC Graphs: Notes and Practical Considerations for Data Mining Researchers*; Technical Report HP Laboratories: Palo Alto, CA, USA, 2003.

76. Vuk, M.; Curk, T. ROC curve, lift chart and calibration plot. *Metod. Zv.* **2006**, *3*, 89–108. [CrossRef]

77. Dimić, G.; Prokin, D.; Kuk, K.; Micalović, M. Primena Decision Trees i Naive Bayes klasifikatora na skup podataka izdvojen iz Moodle kursa. In Proceedings of the Conference INFOTEH, Jahorina, Bosnia and Herzegovina, 21–23 March 2012; Volume 11, pp. 877–882.

78. Witten, H.; Eibe, F. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Morgan Kaufmann Series in Data Management Systems; Elsevier: Cambridge, MA, USA, 2005.

79. Benoit, G. Data Mining. *Annu. Rev. Inf. Sci. Technol.* **2002**, *36*, 265–310. [CrossRef]
80. Weka (University of Waikato: New Zealand). Available online: http://www.cs.waikato.ac.nz/ml/weka (accessed on 20 November 2022).
81. Berrar, D. Bayes' Theorem and Naive Bayes Classifier. *Encycl. Bioinform. Comput. Biol.* **2018**, *1*, 403–412. [CrossRef]
82. Zhang, H. *The Optimality of Naive Bayes, FLAIRS Conference*; AAAI Press: Miami Beach, FL, USA, 2004.
83. Friedman, J.; Hastie, T.; Tibshirani, R. Additive Logistic Regression: A Statistical View of Boosting. *Ann. Stat.* **2000**, *28*, 337–407. [CrossRef]
84. Rokach, L.; Maimon, O. *Decision Trees. In The Data Mining and Knowledge Discovery Handbook*; Springer: Boston, MA, USA, 2005; pp. 165–192. [CrossRef]
85. Xiaohu, W.; Lele, W.; Nianfeng, L. An Application of Decision Tree Based on ID3. *Phys. Procedia* **2012**, *25*, 1017–1021. [CrossRef]
86. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: San Mateo, CA, USA, 1993.
87. Bella, A.; Ferri, C.; Hernández-Orallo, J.; Ramírez-Quintana, M.J. Calibration of machine learning models. In *Handbook of Research on Machine Learning Applications*; IGI Global: Hershey, PA, USA, 2009.
88. SPSS Statistics 17.0 Brief Guide. Available online: http://www.sussex.ac.uk/its/pdfs/SPSS_Statistics_Brief_Guide_17.0.pdf (accessed on 20 November 2022).
89. Liu, H.; Motoda, H. *Feature Selection for Knowledge Discovery and Data Mining*; Kluwer Academic: Boston, MA, USA, 1998.
90. Dash, M.; Liu, H.; Motoda, H. Consistency based feature selection. In Proceedings of the Fourth Pacific Asia Conference on Knowledge Discovery and Data Mining, Kyoto, Japan, 18–20 April 2000; pp. 98–109.
91. Hall, M.A. Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning. In Proceedings of the 17th IEEE Int'l Conf. Machine Learning, Orlando, FL, USA, 17–20 December 2000; pp. 359–366.
92. Novaković, J. Rešavanje klasifikacionih problema mašinskog učenja. In *Bussines Process Reeingineering*; Faculty of Technical sciences Čačak, University of Kragujevac: Kragujevac, Serbia, 2013; Volume 4.
93. Daelemans, W.; Hoste, V.; Meulder, F.D.; Naudts, B. Combined Optimization of Feature Selection and Algorithm Parameter Interaction in Machine Learning of Language. In Proceedings of the 14th European Conference on Machine Learning (ECML-2003), Lecture Notes in Computer Science 2837, Cavtat-Dubrovnik, Croatia, 22–26 September 2003; pp. 84–95.
94. Hall, M.A.; Smith, L.A. Practical feature subset selection for machine learning. In Proceedings of the 21st Australian Computer Science Conference, Perth, Australia, 4–6 February 1998; pp. 181–191.
95. Moriwal, R.; Prakash, V. An efficient info-gain algorithm for finding frequent sequential traversal patterns from web logs based on dynamic weight constraint. In Proceedings of the CUBE International Information Technology Conference (CUBE '12), Pune, India, 3– 6September 2012; ACM: New York, NY, USA, 2012; pp. 718–723.
96. Salzberg, L.S. Book Review: C4.5: By J. Ross Quinlan. Inc., 1993. Programs for Machine Learning Morgan Kaufmann Publishers. *Mach. Learn.* **1994**, *16*, 235–240. [CrossRef]
97. Thakur, D.; Markandaiah, N.; Raj, D.S. Re optimization of ID3 and C4.5 decision tree. In Proceedings of the 2010 International Conference on Computer and Communication Technology (ICCCT 2010), Allahabad, Uttar Pradesh, India, 17–19 September 2010; pp. 448–450.
98. Available online: https://www.programiz.com/dsa/greedy-algorithm (accessed on 15 November 2022).
99. Girish, S.; Chandrashekar, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28.
100. Moore, S.; Notz, I.; Flinger, A. *The Basic Practice of Statistics*; W.H. Freeman: New York, NY, USA, 2013.
101. Ilin, V. The Models for Identification and Quantification of the Determinants of ICT Adoption in Logistics Enterprises. Ph.D. Thesis, Faculty of Technical Sciences University Novi Sad, Novi Sad, Serbia, 2018.
102. Hair, J.F.; Anderson, R.E.; Tatham, R.L.; Black, W.C. *Multivariate Data Analysis*; Prentice-Hall, Inc.: New York, NY, USA, 1998.
103. Yang, T.; Ying, Y. AUC Maximization in the Era of Big Data and AI: A Survey. *ACM Comput. Surv.* **2022**, *37*. [CrossRef]