International Journal of Computer Applications (0975 - 8887)
Volume 165 - No.12, May 2017

Predicting Accidental Locations of Dhaka-Aricha Highway in Bangladesh using Different Data Mining Techniques

Md. Shahriare Satu
Institute of Information Technology
Jahangirnagar University
Dhaka, Bangladesh

Tania Akter
Dept. of CSE
Jahangirnagar University
Dhaka, Bangladesh

Md. Sadrul Arifen
Dept. of CSE
Gono Bishwabidyalay
Dhaka, Bangladesh

Md. Raza Mia
Dept. of CSE
Gono Bishwabidyalay
Dhaka, Bangladesh

ABSTRACT

Road traffic accident is one of the most leading issues which is concerned in many other countries like Bangladesh. Data mining is considered as a reliable technique to analyze traffic accident record and identify factors that provide severity of an accident. The goal of this research to analyze and build classification model that predict an accidental location in the Dhaka-Aricha highway. So, road accidental data is collected from different highway police stations which keep traffic accident record of every road traffic accident on this road. Then, raw dataset is preprocessed and build a classification model with five data mining classification algorithms named Rotation Forest, NBTree, JRip, Naive Bayes and Ridor that analyze traffic accident records to predict risky accidental locations. After classifying this dataset, accuracies of classifiers are compared and the best outcome is showed among them. This results can be used to prevent road accidents in the areas and overcome the number of accidents on the Dhaka-Aricha highway.

Keywords

Road Traffic Accident, Traffic Accident Record, Highway, Classification, Data Mining

1. INTRODUCTION

A road traffic accident refers to any accident that involves at least one road vehicle, open with a public circulation that happens on a road and at least one people died or injured. There are occurred unnatural deaths, disability and property damage for road traffic accident. Bangladesh Jatri Kalyan Samity shows that 17,523 were hurt in road accidents with 1,623 suffering lifelong injuries and 8,589 were killed in 5,928 road accidents between January and December 2014 [1]. Besides, National Road Safety Council(NRSC) in Bangladesh claimed its annual report published in January that over the last five years, road accidents claimed 2,529 lives on average every year, while last year death toll was below 2,000 [1]. But in this circumstances, there are occurred few of research works to prevent and mitigate road traffic accidents. So, it is necessary to build the appropriate model that can prevent road traffic accident in Bangladesh. Data mining is one of the useful techniques to analyze and build a model to prevent this kind of occurrence in the road. Different techniques such as clustering, classification, association rule mining can be used for analyzing traffic accident records. In this work, there a build a model using different classification algorithms on the Dhaka-Aricha highway. Then, we analyze collected data that predicts accidental locations and this result can be used to overcome the accident effects by concerning traffic policies and ensuring traffic safety on roads [2] [3].

In this paper, Section I is introduced about road traffic accident, occurrence of road traffic accident in Bangladesh and a brief overview of our proposed model. Section II describes some related works of this field. Then, we describe a brief discussion of road traffic accident data Section III. Section IV is elaborated our proposed model where we give a brief description about data selection, preprocessing & cleaning, feature selection & extraction, transformation and building a classification model with different classification algorithms. Section V presents and describes the outcomes of this experimental work. Finally, Section VI describes limitations of this work and future plan how to remove difficulties and enhance this model.

2. RELATED WORKS

There are occurred many works with road accidents for improving road safety. Three data mining techniques such as neural network, logistic regression and decision tree are used by Sohn and Shin [4] with a set of influential factors and build up classification models for accident severity. Miao et al [5] considered the performance of four machine learning paradigms which applied to the model considering severity of an injury that occurred during road traffic accidents. Tibebe et al [6] employed classification adaptive regression trees (CART) and random forest approaches which are identified relevant patterns and illustrated the performance of the techniques for the road safety domain where road accident data collected from Addis Ababa traffic office. S Shanthi and R Geetha Ramani are worked on feature relevance analysis and classification of road traffic accident using different data mining techniques [7]. They are implemented in different clustering and association analysis algorithms on the traffic accident records. Humberto Gonzalez tried to find patterns of the road accident in the United Kingdom in 2013. Besides, this paper used association rule mining and apriori algorithm on road accident dataset using R [8]. Sachin

Kumar and Durga Tashniwal used K-means clustering algorithm which takes accident frequency count as a parameter to cluster and then characterized accidental locations [2] [9]. They [5] also proposed a framework that used K-modes clustering technique as a preliminary task for segmentation of 11,574 road accidents on road network of Dehradun between 2009 and 2014. Shi et al [10] have proposed a time series model that was constructed by using Cell Transmission Model to reflect the state of traffic flow by ternary numbers. In this model, a numerical experimentation was carried out and then the result showed the effectiveness of the proposed method.

In Bangladesh, ARI (Accident Research Institute), BUET is collected road traffic accidental data from different highway police stations throughout this country and worked on it to ensure road safety [11] [12].

3. DATASET DESCRIPTION

There are existing 96 (3812.78 km) national highways, 126 (4246.97 km) regional highways and 654 (13242.33 km) zilla roads in Bangladesh. Dhaka-Aricha highway has consisted of 75.4 km which is started from 11.9 km reference point at Aminbazar Bridge to 87.3 km at Aricha Ferry Ghat. It is built in 1960 with six Upazilas or subdistricts which are Savar, Dhamrai, Saturia, Manikgonj, Ghior and Shibalaya of Dhaka and Manikgonj districts. It is very important part of the national highway network and connecting Dhaka with the ferry routes at Aricha. It is also a section of Asian Highway route (AH1) [13]. Then, 104 accidental records have been collected from Savar highway police station, Golora highway police station, Barangail highway police station and Savar police station that occurred from January 2015 to December 2016. In Table 1, a brief description of attributes of this dataset is given below:

Table 1. Road Traffic Accident Attributes

S/N	Attribute Name	Type	Values
01	Month Name	Nominal	Number of months
02	Region Name	Nominal	Where the accident is occurring
03	Vehicle Type	Nominal	Which type of vehicle are occurred in the accident
04	Number of Vehicles	Ratio	How many vehicles are damaged in the accident
05	Victims Injured	Ratio	How many peoples are injured
06	Victim Death	Ratio	How many peoples are dying
07	Victim Gender	Nominal	What type of Gender
08	Class Type	Nominal	Where the accident is occurring

4. PROPOSED MODEL

There are considered several steps to build a classification model that analyze road accidental records and predict accidental locations in the Dhaka-Aricha highway. There are considered several steps to analyze and manipulate required dataset. First, we have collected road traffic accident data that preprocess and extract some features for analyzing further manipulation of it. Then we used different classification algorithms such as Rotation Forest, NBTree, JRip, Naïve Bayes, Ridor to classify them into different accidental locations and visualize predicted model with the appropriate figure. In figure 1, there are represented several steps how to implement this model. Those steps are described briefly as follows:

4.1 Data Selection

In this work, several field studies have done to collect raw road traffic accident data from different highway police stations of Dhaka-Aricha highway. After completing this study, there are collected 104 road traffic accident data from January 2015 to December 2016.

4.2 Data Preprocessing & Cleaning

Data preprocessing is the primary task to prepare data of traffic accident records for further analysis and getting good results about road traffic accident. Data quality is explained in terms of accuracy, consistency, completeness, believability, interpretability and timeliness. These qualities are assessed by the usage of the data. In this study, there are removed several tuples from which have multiple unclear, duplicate and missing values from existing data.

4.3 Feature Selection & Extraction

Feature selection and extraction is the principle step in any machine learning algorithm to select most relevant attributes and combine attributes into a new reduced set of features. For this reason, unnecessary records are filtered and selected relevant eight features that impacts on road accident such as month name, region name, vehicle type, number of vehicles, victims injured, victim death, victim gender and class types.

4.4 Data Transformation

Data transformation is the process of converting data one format to the another for manipulating different tasks. In this work, different attribute have transformed such as month name, region name, vehicle type and victim gender to convert data string to nominal in our work . As a result, different visualization and mining algorithm can execute and represent data efficiently in this work.

4.5 Classification

Classification is the process to find the function or model that explains the classes whose label is unknown for the intension to predict the class of objects. It is the formation of data analysis which extracts the models that describe important classes of records. So, this analysis of data is called classification where a classifier is used for predicting classes.

4.5.1 ***Classification Algorithms***. There are used five classification algorithms such as Rotation Forest, NBTree, JRip, Naïve Bayes, Ridor to classify this dataset. These algorithms are mentioned briefly.

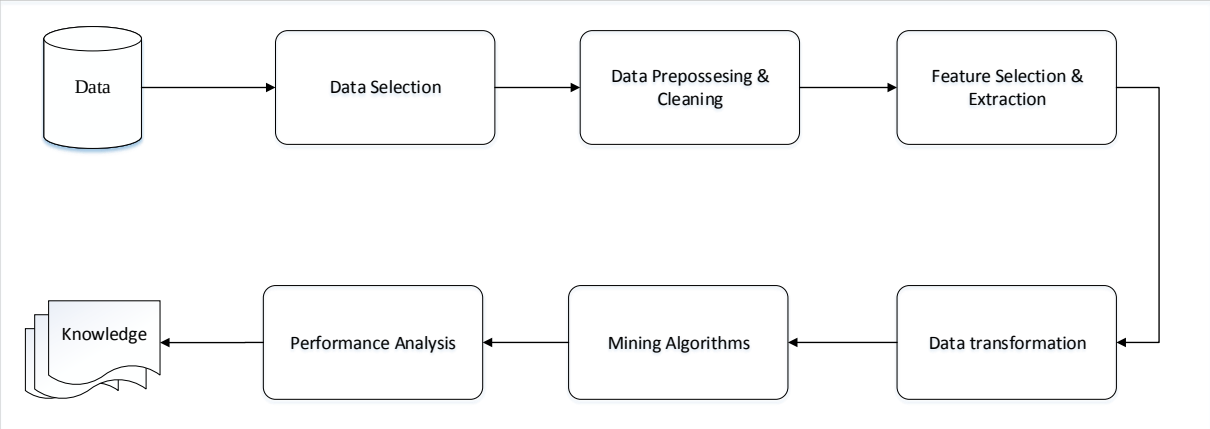

Fig. 1. Working flow diagram of proposed model

—**Rotation Forest Classifier:** Rotation Forest (RTF) is meta-algorithm that is generated classifier ensembles based on feature extraction [14]. It use J48 classifier in rotation forest which is The feature set is randomly split into K subsets and Principal Component Analysis (PCA) is applied to each subset. All principal components have preserved the variability of information. Diversity is promoted to the ensemble by applying Principal Component Analysis (PCA) to extract features from the dataset. The main idea of rotation forest is to use Principal Component Analysis (PCA) to rotate K axis in order to obtain different training sets for classification or regression [15].

—**Naïve Bayes Classifier:** Naïve Bayes classifier is a simple probabilistic classifier by applying Bayesian theorem (from Bayesian statistics). It is fast highly scalable model building and scoring. It is more acceptable when the dimension of input is high. Numeric precision values of estimators are taken from training data analysis. It accomplishes as well in many complex real-world situations in spite of oversimplified expectations. It utilizes the maximum likelihood to estimate parameters of Naïve Bayes model. It requires a small amount of training data to predict the parameters. [16].

—**NBTree Classifier:** NBTree (Naïve Bayesian tree) is tree-based classifier that consists of Naïve Bayesian classification and decision tree learning model. It is organized with the example of a leaf and then assigns a class label by applying a Naïve Bayes classification process on that leaf. By using Naïve Bayes classifier for each leaf node, the instances are classified. This process repeated until no example is left. So, NBTree habitually achieves higher accuracy either Naïve Bayesian classifier or decision tree learning algorithm to classify required dataset [17].

—**JRip Classifier:** JRip is a rule-based classification algorithm that produces a set of rules to classify data. Classes are assessed by growing size and a set of rules for the class which is generated using incrementally reduced error. By providing all the samples of a particular decision on the training data set and finding a set of rules which cover all the records of this class. Subsequently, it executes to the next class and does the same process, repeating this until all classes have been covered [18].

—**Ridor Classifier**: Ridor is a rule-based classification algorithm that generates the default rule with exceptions. For finding the smallest error rate, an incrementally reduced error pruning is used to find the best exception with iteration processes for each exception. This exception generates a default rule for working data. Therefore, it accomplishes a tree-like expansion of exceptions [16] [19] [20] [21].

Algorithm 1 Prediction of Accidental Locations of Dhaka-Aricha highway

Input: Set of attributes of A_{all}, set of all Classifier C
Output: Find best classifiers on v fold cross-validation.

Begin
$A \leftarrow 0$
for each attribute $a \in A_{all}$ **do**
    $A \leftarrow A \cup \{a\}$
end for
for each classifier $c_i \in C$ **do**
    for each cross validation $v_j \in V$ **do**
        $Accuracy_{ij} \leftarrow$ accuracy of c_i with j^{th} fold.
    end for
    Select top value of $Accuracy_{ij}$ list
    Return i^{th} classifier for j^{th} fold cross validation.
end for
End

4.5.2 ***Working Process***. In this section, we represent our working process of building a prediction model that classifying accidental location of road traffic accident in the Dhaka-Aricha highway that is explained as follows.

—As the first step of this research, traffic accident records have collected from different highway police stations in the Dhaka-Aricha highway and preprocessed it. Then, some attributes/features are identified and extract from the data set that are related to finding accidental locations which are listed in Table 1.

—After generating an accidental data set, there are existing a large number of classification techniques that are used for classification tasks. So, some classification algorithms are used such as Rotation Forest, NBTree, JRip, Naïve Bayes, and Ridor with 10 fold cross-validation to analyze road accident dataset.

—The classifier is evaluated that how we can analyze and predict a set of instances of classes of loaded from a file. After this execution, different classifiers are evaluated by comparing the accuracies among all classification algorithms. Then, the best classifier can be determined by the evaluation of accuracy of algorithms.

5. RESULT AND DISCUSSION

Weka is a data mining tool that is implemented in JAVA and developed by the University of Waikato in New Zealand. It consists of different machine learning algorithms to accomplish data mining tasks [22]. After data preprocessing, it can implement different types of algorithms for classification, clustering, regression, association rule mining and visualization. So, this tool is helpful to develop new learning model for different purposes and activities. In this work, we collect 104 samples of road traffic accident that are converted into .arff file that loads into weka explorer. There are defined three categories which are high-frequency accidental location (HFAL), moderate-frequency accident location (MFAL) and low-frequency accidental location (LFAL) to classify in this dataset.

There are used several classifiers such as Rotation Forest, NBTree, JRip, Naïve Bayes,and Ridor form Weka, execute existing dataset using those classifiers. Besides, we have used 10 fold cross-validation model evaluate the performance of classifiers in this experiment. Evaluation is manipulated based on precision(P), recall(R) and F1-Score [23]. We used equation 1, 2 and 3 to calculate precision, recall and F1-score respectively.

$$precision = \frac{tp}{tp + fp} \tag{1}$$

$$recall = \frac{tp}{tp + fn} \tag{2}$$

$$f1 - score = 2\frac{precision * recall}{precision + recall} \tag{3}$$

Where, tp equals to the true positive, fp equals to the false positive and fn equals to the false negative.

Table 2 represents the performance of every classifier to compare them based on the precision (P), recall (R) and F1-Score. From this table, weighted average of precision, recall and F1-score value are 0.907, 0.904 and 0.904 for Rotation Forest, 0.907, 0.904 and 0.904 for NBtree, 0.892, 0.894 and 0.893 for JRip, 0.876, 0.875 and 0.871 for Naïve Bayes and 0.881, 0.875 and 0.877 for Ridor. So, to compare the value of precision, recall and F1-score of existing classifiers, Rotation Forest and NBTree algorithm outperforms all classifiers within the data set in this experiment.

Cohen's kappa coefficient (kappa statistics) is a statistic that evaluates inter-rater agreement for qualitative (categorical) items.

Table 2. Class Level Accuracy for Classifiers

Classifier	Precision	Recall	F1-Score	Class
Rotation Forest	0.938	0.789	0.857	HFAL
	0.778	0.84	0.808	MFAL
	0.951	0.967	0.959	LFAL
Weighted Average	**0.907**	**0.904**	**0.904**	
NBTree	0.882	0.789	0.833	HFAL
	0.786	0.88	0.83	MFAL
	0.966	0.95	0.958	LFAL
Weighted Average	**0.907**	**0.904**	**0.904**	
JRip	0.889	0.842	0.865	HFAL
	0.792	0.76	0.776	MFAL
	0.935	0.967	0.956	LFAL
Weighted Average	**0.892**	**0.894**	**0.893**	
Naive Bayes	0.941	0.842	0.889	HFAL
	0.85	0.68	0.756	MFAL
	0.866	0.967	0.913	LFAL
Weighted Average	**0.876**	**0.875**	**0.871**	
Ridor	0.944	0.895	0.919	HFAL
	0.714	0.8	0.755	MFAL
	0.931	0.9	0.915	LFAL
Weighted Average	**0.881**	**0.875**	**0.877**	

It is calculated by using following equation [24]:

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e} \tag{4}$$

where p_o is relative observed agreement among raters and p_e is the hypothetical probability of chance or expected agreement using observed data.

Table 3. Cohen's kappa coefficient measurement for classifiers

Evalution Criteria	Rotation Forest	NB Tree	JRip	Naive Bayes	Ridor
Kappa Statistics	0.8316	0.8337	0.8141	0.7736	0.7852

There is shown some measurement of kappa statistics of different classifiers in Table 3. In this experiment, NBTree(0.8337) and Rotation Forest(0.8316) shows highest values of kappa coefficient rather than other classifiers.

Besides, we use equation 4,5,6 and 7 to calculate the error between predicted and true result to get Mean Absolute Error(MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE) and Root Relative Square Error (RRSE) in this work. Now, those equations are given as follows [25]:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} \mid \hat{\theta}_i - \theta_i \mid \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{\theta}_i - \theta_i)^2} \tag{6}$$

$$RAE = \sqrt{\frac{\sum_{i=1}^{N} \mid \hat{\theta}_i - \theta_i \mid}{\sum_{i=1}^{N} \mid \bar{\theta} - \theta_i \mid}} \tag{7}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{\theta}_i - \theta_i)^2}{\sum_{i=1}^{N}(\bar{\theta} - \theta_i)^2}} \quad (8)$$

where $\hat{\theta}_i$ equals to estimated value, θ_i equals to true value, N equals to the number of samples and $\bar{\theta}$ is a mean value of θ_i.

Table 4. Error Measurement for Classifiers

Evaluation Criteria	Rotation Forest	NB Tree	JRip	Naïve Bayes	Ridor
Mean absolute error	0.1252	0.1263	0.0996	0.1183	0.0833
Root mean squared error	0.2298	0.2465	0.2514	0.269	0.2887
Relative absolute error	32.42%	32.72%	25.79%	30.65%	21.59%
Root relative squared error	52.40%	56.21%	57.33%	61.35%	65.84%

There are showed several error rate calculation of different classifiers in Table 4. There are considered different kinds of measurement to calculate errors in this experiment. Mean absolute error(MAE) of Ridor (0.0833) is minimum and root mean squared error (RMSE) of Rotation Forest (0.2298) is minimum compared to others classifiers. Besides, relative absolute error (RAE) of Ridor(21.59%) is minimum and root relative square error (RRAE) of Rotation Forest(52.40%) is the minimum rather than others.

MAE and RMSE are represented average difference between those two values that can interpret to compare the scale of variable. On the other hand, RAE and RRSE have divided those differences by the variation of θ and they have a scale from 0 to 1.

Table 5. Performance Analysis of Classifiers

Evaluation criteria	Rotation Forest	NB Tree	JRip	Naïve Bayes	Ridor
Timing to build the model	0.26	27	0.02	0	0.01
Correctly classified instances	94	94	93	91	91
Incorrectly classified instances	10	10	11	13	13
Accuracy by class	90.39%	90.39%	89.43%	87.50%	87.50%

In Table 5, it is referred as an accuracy of the particular classifier. The accuracy of Rotation Forest is 90.39% , the accuracy of NBTree is 90.39%, accuracy of JRip is 89.43% , accuracy of Naïve Bayes is 87.50% and accuracy of Ridor is 87.50% that is determined to consider the percentage of the ratio between correctly classified instances and total instances in this experiment. In this case, we can say Rotation Forest and NBTree both are the best classifiers, but if we consider execution time that needs to manipulate each classifier, then Rotation Forest is needed 0.26s and NBTree is needed 27s to bring the same outcome.

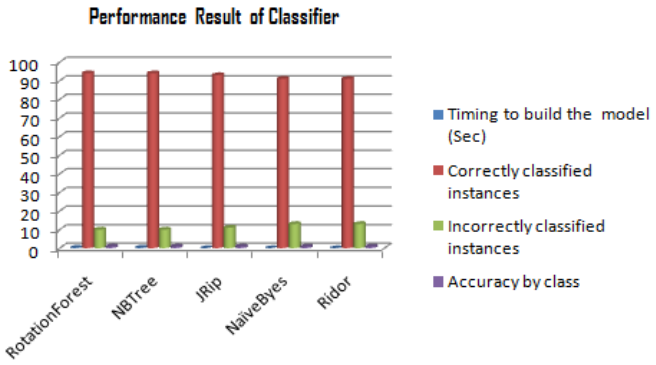

Fig. 2. Efficiencies and accuracies of classifiers

The Figure 2 shows that the graphical representation of efficiency and accuracy of correctly classifying instances of road-accident data. It also shows the best classifier that can classify data according to the requirement.

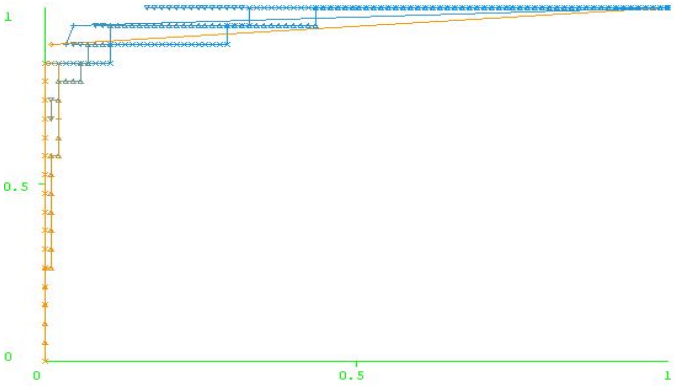

Fig. 3. Model Performance ROC Curve of Road traffic Accidental data in Dhaka-Aricha Highway

In Figure 3, $\times$, $+$, $\diamond$, $\triangle$, ∇ are indicated as Naïve Bayes, JRip, Ridor, NBtree and Rotation Forest. Both diagrams are represented TPR (True Positive Rate) in Y-axis and FPR (False Positive Rate) in X-axis for showing ROC Curve of road traffic accident data in Dhaka-Aricha highway. In this work, the classification result of rotation forest is better than other classifiers because its curve is more responsive to the TPR.

So, according to the experimental result of different perspective, it can be noticed that Rotation Forest is the best classifier to find accidental locations of road traffic accident in the Dhaka-Aricha highway.

6. CONCLUSION AND FUTURE WORK

Road accidents are serious issues that can bear death, disabilities, injuries and further fatalities. In order to decrease the number of accidents, we need to understand and analyze them [26]. As the previous discussion, we have used different classifiers to analyze the datasets and evaluation performance. Although this data mining approach is quite sufficient to uncover reasonable information from the selected data set, the results remain at a very general level as source data does not contain other accident related information such as the speed of vehicles at the time of the accident, weather information, road surface condition. The data with more number of attributes which can reveal more information using our approach. The overall performance of the Rotation Forest algorithm is acceptable due to it has shown more accurate outcome than other techniques. This report represents the real-world accident training dataset.

Acknowledgment

We are thankful to the Savar highway police station, the Golora highway police station, the Barangail highway police station and the Savar Thana for providing data for our research work. Besides, we are also thankful to Farha Farida Sathi, Ahadduzaman Ahad and Fahad Ebne Mostafa to help us for collecting traffic accident records form different police stations.

7. REFERENCES

[1] Study: Road accidents killed one per hour in 2014. "http://archive.dhakatribune.com/bangladesh/2015/apr/03/study-road-accidents-killed-one-hour-2014", March 2017.

[2] Sachin Kumar and Durga Toshniwal. Analysing road accident data using association rule mining. In *Computing, Communication and Security (ICCCS), 2015 International Conference on*, pages 1–6. IEEE, 2015.

[3] I.H. Witten, E. Frank, and M.A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2011.

[4] So Young Sohn and Hyungwon Shin. Pattern recognition for road traffic accident severity in korea. *Ergonomics*, 44(1):107–117, 2001.

[5] Miao M Chong, Ajith Abraham, and Marcin Paprzycki. Traffic accident analysis using machine learning paradigms. *Informatica (Slovenia)*, 29(1):89–98, 2005.

[6] Tibebe Beshah, Dejene Ejigu, Ajith Abraham, Vaclav Snasel, and Pavel Kromer. Pattern recognition and knowledge discovery from road traffic accident data in ethiopia: Implications for improving road safety. In *Information and Communication Technologies (WICT), 2011 World Congress on*, pages 1241–1246. IEEE, 2011.

[7] S Shanthi and R Geetha Ramani. Feature relevance analysis and classification of road traffic accident data through data mining techniques. In *Proceedings of the World Congress on Engineering and Computer Science*, volume 1, pages 24–26, 2012.

[8] Humberto Gonzalez. Finding patterns in 2013 road accident data in united kingdom. 2015.

[9] Sachin Kumar and Durga Toshniwal. A data mining approach to characterize road accident locations. *Journal of Modern Transportation*, 24(1):62–72, 2016.

[10] An Shi, Zhang Tao, Zhang Xinming, and Wang Jian. Evolution of traffic flow analysis under accidents on highways using temporal data mining. In *Intelligent Systems Design and Engineering Applications (ISDEA), 2014 Fifth International Conference on*, pages 454–457. IEEE, 2014.

[11] SM Sohel Mahmud, Md Shamsul Hoque, and QA Shakur. Road safety research in bangladesh: constraints and requirements. In *The 4th Annual paper meet (APM) and the 1st Civil Engineering Congress, organized by Civil Engineering Division Institution of Engineers, Bangladesh (IEB), Session V: Transportation Engineering-II*, pages 22–24, 2011.

[12] SM Sohel Mahmuda, Ishtiaque Ahmedb, and Md Shamsul Hoquec. Road safety problems in bangladesh: Achievable target and tangible sustainable actions. 2014.

[13] Md Shamsul Hoque, Shah Md Muniruzzaman, and SN Ahmed. Performance evaluation of road safety measures: a case study of the dhaka-aricha highway in bangladesh. *Transport and communications bulletin for Asia and the Pacific*, 74, 2005.

[14] Juan José Rodriguez, Ludmila I Kuncheva, and Carlos J Alonso. Rotation forest: A new classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence*, 28(10):1619–1630, 2006.

[15] Tadeusz Lasota, Tomasz Łuczak, and Bogdan Trawiński. Investigation of rotation forest method applied to property price prediction. In *International Conference on Artificial Intelligence and Soft Computing*, pages 403–411. Springer, 2012.

[16] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.

[17] Yumin Zhao, Zhendong Niu, and Xueping Peng. Research on data mining technologies for complicated attributes relationship in digital library collections. *Applied Mathematics & Information Sciences*, 8(3):1173, 2014.

[18] Vaishali S Parsania, NN Jani, and Navneet H Bhalodiya. Applying naïve bayes, bayesnet, part, jrip and oner algorithms on hypothyroid database for comparative analysis.

[19] V Veeralakshmi and D Ramyachitra. Ripple down rule learner (ridor) classifier for iris dataset. *Issues*, 1(1):79–85.

[20] SR Kalmegh and SN Deshmukh. Categorical identification of indian news using j48 and ridor algorithm.

[21] A Sudha, P Gayathri, and N Jaisankar. Effective analysis and predictive model of stroke disease using classification methods. *International Journal of Computer Applications*, 43(14):26–31, 2012.

[22] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

[23] Cyril Goutte and Eric Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European Conference on Information Retrieval*, pages 345–359. Springer, 2005.

[24] Cohen's kappa. "https://en.wikipedia.org/wiki/Cohen%27s_kappa", April 2017.

[25] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. *Data Mining: Practical machine learning tools and techniques*. Elsevier, 2017.

[26] Eyad Abdullah and Ahmed Emam. Traffic accidents analyzer using big data. In *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 392–397. IEEE, 2015.