RESEARCHERID
THOMSON REUTERS

[NC-Rase 18] ISSN 2348 – 8034
DOI: 10.5281/zenodo.1548499 Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

TRAFFIC INCIDENT CLASSIFICATION AND NOTIFICATION USING MACHINE LEARNING APPROACH

Rupali M. Bora[1] & Manisha R. Arsule[2]
Information Technology Department, KKWIEER, Nashik, Maharashtra, India

ABSTRACT

Everyday many vehicles travel on road and number of traffic incidents occur which affects people travelling on that route. The task of manual classification of these incidents takes more time which can be minimized using the system which classifies these incidents using machine learning algorithms. When the incident occurs administrator fills the details of it in dataset. This dataset is used for prediction of incident severity. The incidents are classified for deciding severity using Naïve Bayes classifier which will discover and visualize frequent patterns in historical incidents. The classification model accuracy will be improved using k-fold cross validation. K-Nearest Neighbor algorithm is used to find nearest incidents for notifications to user. An android application will be used by user who is travelling. While travelling on route, user gets notification regarding incidents like accident prone zone, slippery road ahead etc. so as to take appropriate decisions.

Keywords*: Incidents, Severity, NaiveBayes, K-Nearest Neighbour Classifier, k-fold Cross validation.*

I. INTRODUCTION

In modern cities, traffic conditions are changing every moment and a single abnormality will affect the daily operations of transport. Traffic incidents are detected using multiple sources including communications, video cameras and sensors. Transport systems rely on these diverse information streams. Composite computer software can be used to detect, confirm and clear incidents on the road. Many tasks are involved along with a large amount of manual work like incident classification, traffic management, and emergency services management and information propagation to public [1]. With the support of the modern machine learning (ML) approaches, the classification step can be automatically achieved by looking at the historical incident data. There are two types of detection techniques: the time series analysis (TSA) method that focuses on traffic changes and the machine learning (ML) method that exploits traffic patterns for detection. TSA methods often make use of rules involving limited traffic variables for incident detection. Such rules are generally simple and may not capture various incident scenarios. ML methods are often more flexible in detecting the hidden patterns of incidents. These methods can be used to detect and predict incident severity [2].

Accurate and comprehensive accident records are the basis of incident analysis. The effective use of incidents records depends on some factors, like the accuracy of the data, record retention, and data analysis. In India, there is much kind of places like hilly area plateaus, and due to unsuitable road facilities incidents are more and death rate due to this incidents are more. The maximum number of incidents is reported in the transport sector that is on road. In regard to this, system aims for classification and notification of traffic incidents.

II. RELATED WORK

Classification is a popular task in Machine Learning (ML) where the categories of the instances are automatically assigned based on a train set of historical data. ML approaches train models that represent the data in both a general and more accurate way. The basic advantage of supervised machine learning is that the incident characteristics and classification rules can be automatically learnt through training examples [3].

In an earlier study, the pattern recognition method for road traffic accident severity in Korea was first introduced by Sohn and Shin [4]. One other approach was usage of Shapelet Transform in the field of traffic event detection. The applicability of the algorithm for automatic incident detection was proved where it provides comparable performance to other techniques. Also the Shapelet Transform algorithm can help in improving the detection by guiding the expert input in a cognitive approach.[6]. SVM based approaches were implemented in [5] which focuses on linear separable data only.

Most previous researches tested on data sets covering only few months. There were limitations of specific road type or area. Our system was designed a large data set with many records and it focuses on Nashik area, rather than just a selection of general roads or areas. The model will be trained on larger dataset with cross-validation, which have a number of advantages for real-world incident classification.

III. METHODOLOGY

A. k-Fold Cross-Validation:-

Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a parameter called k that refers to the number of groups that a given data sample is to be split into. A specific value for k is chosen to apply the model. Cross-validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample in order to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model. It is a popular method because it is simple to understand and it generally results in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split.

The general procedure is as follows:[7]

- Shuffle the dataset randomly.
- Split the dataset into k groups
- For each unique group:
    1. Take the group as test data set
    2. Take the remaining groups as a training data set
    3. Fit a model on the training set and evaluate it on the test set
    4. Retain the evaluation score and discard the model
- Summarize the skill of the model using the sample of model evaluation scores

B. Classifiers:

1. Naïve Bayes:-

Naive Bayes classifiers are based on Bayes' Theorem. In this algorithm, every pair of features being classified is independent of other. The dataset is divided into two parts, namely, feature matrix and the response vector. The fundamental Naive Bayes assumption is that each feature makes an independent and equal contribution to the outcome.

Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem is stated mathematically as the following equation:

$$P(A|B)= \frac{P(B|A)P(A)}{P(B)}$$

P(A) is the prior probability of A (i.e. Probability of event before evidence is seen). P(A|B) is a posteriori probability of B.

Naive assumption defines the independence among the features. So now, the evidence is spitted into the independent parts.

Hence, the result is:

$$P(y|x1,\dots xn)=\frac{P(x1|y)P(x2|y)\dots P(xn|y)P(y)}{P(x1)P(x2)\dots P(xn)}$$

2. K-Nearest Neighbours(KNN)

K-Nearest Neighbors is an essential classification algorithm in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection. It is widely

RESEARCHERID
THOMSON REUTERS
[NC-Rase 18] ISSN 2348 – 8034
DOI: 10.5281/zenodo.1548499 Impact Factor- 5.070

used in real-life scenarios as it is non-parametric i.e. it does not make any underlying assumptions about the distribution of data. Some training data is given, which classifies coordinates into groups identified by an attribute. Figure 1 shows algorithm for KNN.

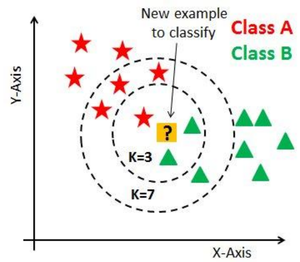

Fig.1 KNN Algorithm

KNN Algorithm:

1. Store the training samples in an array. Each element of this array represents a tuple (x, y).
2. Make set S of K smallest distances. Each of these distances corresponds to an already classified data point.
3. Return the majority label among S.

C. Traffic Incidents Dataset :-

The table I shows incident dataset [4] will be used as input to the system with 13 features with domain values for each feature. The survey of Nashik city was found in [8] & [9], which helps to prepare dataset for incidents related to our project.

Table I- Incident Dataset

Feature Name	**Feature Value/Description**
incident reporter	Traffic police , Member of public
incident type	Accident, breakdown
incident subtype	Bus, car, truck, bike
Incident time	24 Hrs.
day of week	1–7(Mon - Sun)
traffic direction	North, south, east, West
lanes affected	Number of lanes affected
road	Road name, type(highway, flyover, slippery)
incident severity	High, low, medium
Incident area	Name of area
no of vehicles	1/2/3
road surface condition	Speed-breaker, pit-hole, tree
vehicle Number	Vehicle number

D. Architecture of System:

Fig.2 shows the architecture of system. The dataset is given as input to Cross-validation module with which multifold sets are given to improve accuracy of training. After training of Naive Bayes, probabilities can be used for testing. Prediction of results for incidents severity can be used for notifying users travelling on that road.

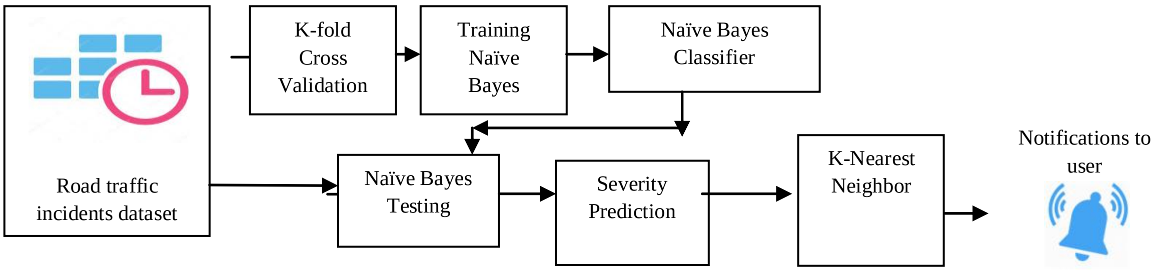

Fig.2. Architecture of system

- *WORKING PRONCIPLS:*

1. Use Case Diagram:

As shown in fig. 3. Mainly three actors are in action. Different use cases are devised for functionality of system for each actor.

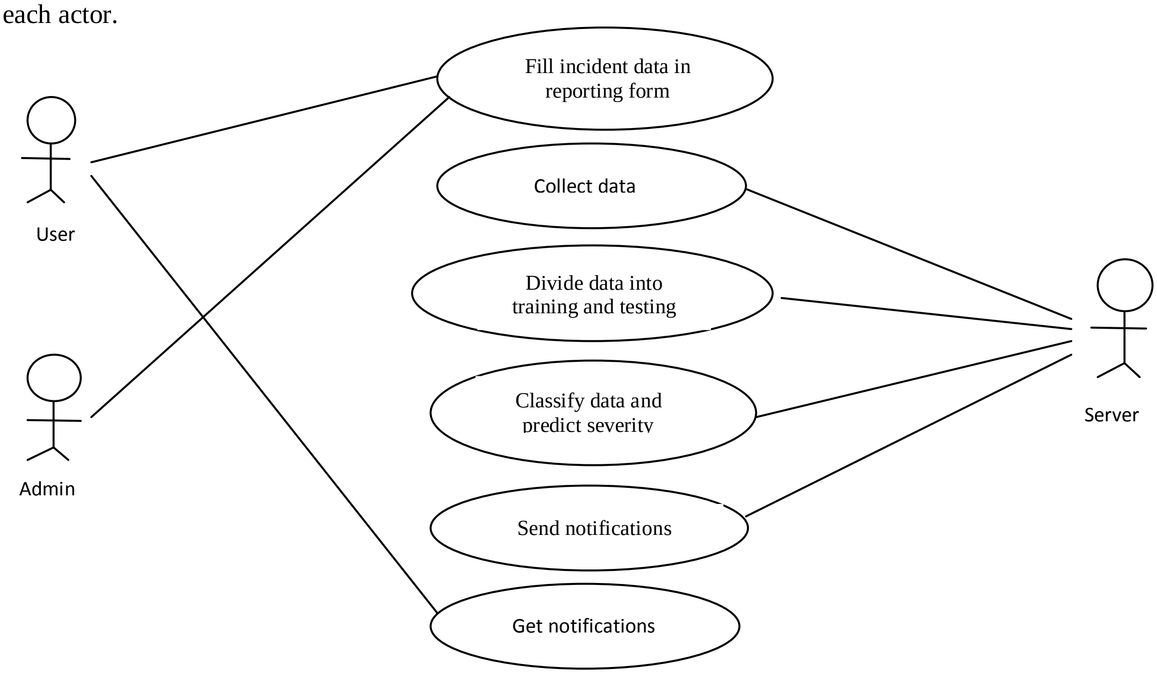

Fig. 3. Use case Diagram

IV. EXPERIMENTAL RESULTS

Evaluation Parameters:-

The primary evaluation metrics are precision (P), recall (R) and Accuracy. Precision is the rate of retrieved instances that are correct while recall is the proportion of relevant instances that are retrieved. Accuracy will give correct classification of incidents. They are calculated based on the true positive (TP), true negative, false positive (FP) and false negative

Precision=$\frac{TP}{TP+FP}$ Recall=$\frac{TP}{TP+FN}$ Accuracy=$\frac{TP+TN}{TP+TN+FP+FN}$

Fig. 4 shows workflow of the system with necessary modules of applications.

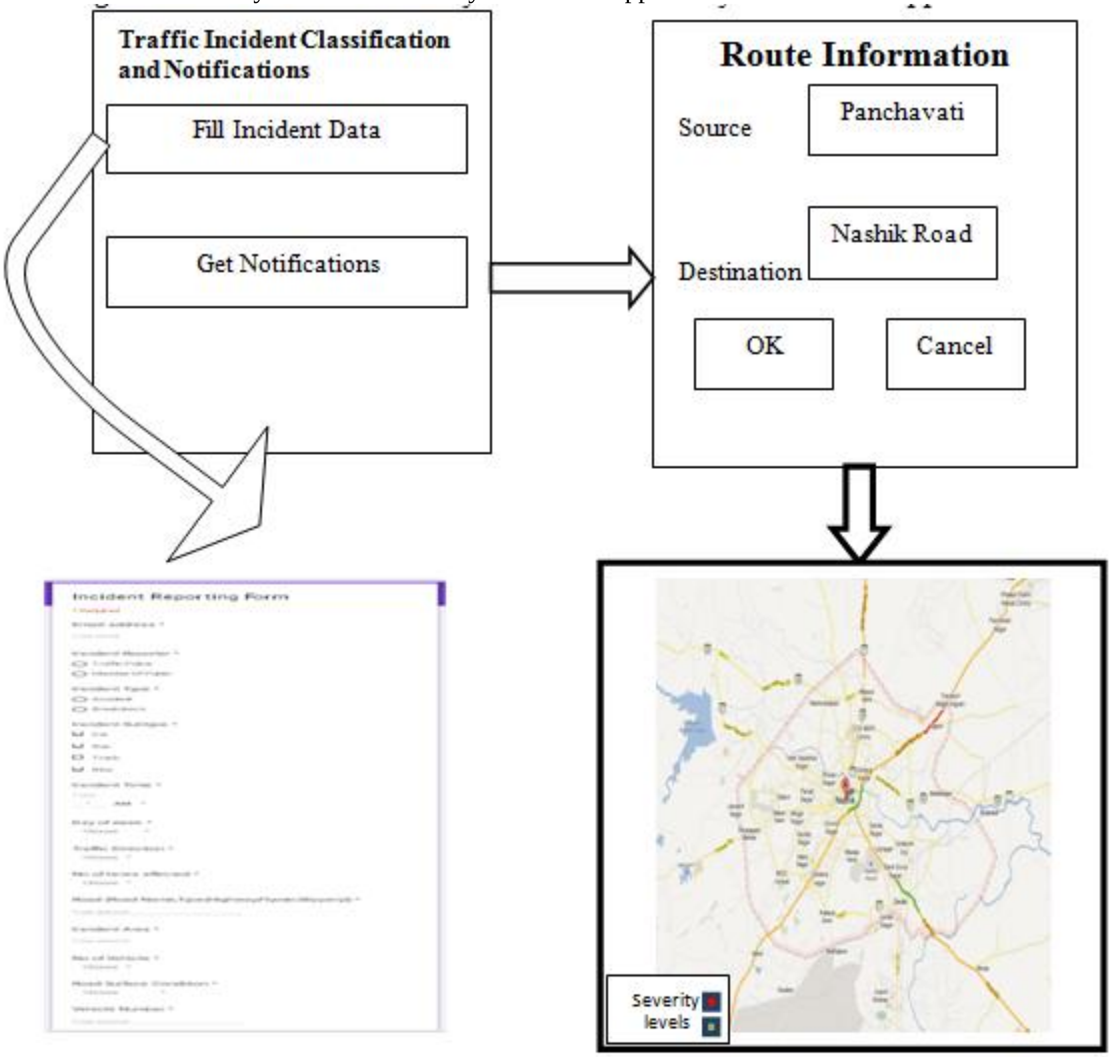

Fig. 4. System Workflow

V. CONCLUSION

This project carried out detailed analyses, visualizations and classification of the incidents reported from Nashik routes network. It is capable of providing useful information in a real-time manner. This can help the administrator to respond more quickly to incidents that could potentially affect the normal flow of traffic, and to make better management. The appropriate ML method for incident severity classification is implemented. The model is capable to learn from incident data and make prediction in real time. ML algorithms like k-fold cross-validation and can be utilized to optimize the train set or select the most informative incident record to update the model's knowledge.

REFERENCES

1. *Taib, R., Yee, D., Fang, F., et al.: "Improved incident management through anomaly detection in historical records". Proc. of the 21th ITS World Congress, Detroit, USA, 2014, pp. 11–23*
2. *Wang, Jiawei, et al. "A hybrid approach for automatic incident detection." IEEE Transactions on Intelligent Transportation Systems 14.3 (2013): 1176-1185.*

[NC-Rase 18] **ISSN 2348 – 8034**
DOI: 10.5281/zenodo.1548499 **Impact Factor- 5.070**

3. *Hoang Nguyen, Chen Cai, Fang Chen Data61 – CSIRO, "Automatic classification of traffic incident's severity using machine learning approaches", 13 Garden Street, Eveleigh NSW 2015, Australia*
4. *So Young Sohn, Sung Ho Lee, "Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea", Department of Computer Science and Industrial Systems Engineering, Yonsei University,134 Shinchon-dong, Sudaemon-ku, Seoul, South Korea*
5. *Liu, Haihong, et al. "Study on traffic information fusion algorithm based on support vector machines." Intelligent Systems Design and Applications, 2006. ISDA'06. Sixth International Conference on. Vol. 1. IEEE, 2006.*
6. *AlDhanhani, Ahmed, et al. "Analysis of Shapelet Transform Usage in Traffic Event Detection." 2018 IEEE International Conference on Cognitive Computing (ICCC). IEEE, 2018.*
7. *Ref: https://machinelearningmastery.com/k-fold-cross-validation/*
8. *Vanjari, Anuja H., Madhura C. Aher, and Shrikant R. Baviskar. "ROAD SAFETY AUDIT OF SERVICE ROADS ADJACENT TO NATIONAL HIGHWAY-3 PASSING THROUGH NASHIK."*
9. *Aher, V. S., et al. "Road Accident Prediction for Nashik City."*