IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 37-41

Classification and Notification of Road Incidents Using Ensemble Approach

Manisha R. Arsule[1], Manjusha R. Gahiwade[2], Varsha S. Hedav[3], Nivedita P. Patil[4], Karuna S. Patil[5]
[1,2,3,4,5]BEIT, Dept. of Information Technology, KKWIEER, Nashik)

Abstract: *Normally many vehicle travels on road and number of traffic incidents occur which influences people journeying on that direction. The task of manual classification of these incidents takes more time which may be minimized the use of the system which classifies these incidents using machine learning algorithms. When the incident takes place administrator fills the information of it for dataset generation. This dataset is used for prediction of incident severity. The incidents are categorized for finding out severity using Naïve Bayes and k-Nearest Neighbour classifiers in an effort to discover and visualize frequent patterns in ancient incidents. The classification model accuracy will be progressed using k-fold cross validation. The ensemble approach of machine learning allows to apply different model which yields better results and reduce problem of overfitting by the combination of model. The system has an android software for consumer who's visiting. Even as traveling at the route, person gets notification concerning incidents like accident prone sector, slippery road beforehand and so on. In order that consumer can take suitable choices and saves time throughout touring.*

I. Introduction

In contemporary cities, traffic conditions are changing each moment and a single abnormality will affect the daily operations of transport and logistics groups. As site visitors information is amassed from commentary from sensors inclusive of digital video cameras and loop detectors on road networks, the records length is substantial and contaminated with noise and errors. Correct and comprehensive accident records are the basis of incident evaluation. The real use of injuries information relies upon on a few elements, just like the accuracy of data, record retention, and data analysis. In India there may be plenty sort of locations like hilly location plateaus, and because of flawed road facilities incidents are greater and death rate because of these incidents is greater. The most number of incidents is reported inside the transport sector, which is on road. In appearance to this, system provides a evaluation on the existing technologies that aim to 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[1].

In an earlier study, the pattern recognition method for road traffic accident severity in Korea was first introduced by Sohn and Shin [2].The work is carried out by doing site survey to identify the accident spots and analyzing the accident data collected from police stations. Analysis of major accidents has concluded that most of the fatal accidents are occurring on the intersections in between Jatra hotel to Adgaon Naka stretch and Mumbai Naka to Pathardi phata stretch, as well as during the month of July and in between 8 am to 8 pm. Due to less number of entry and exit points provided to the newly built over, most of the highway traffic is

shifted to the service roads and that is the main reason of increasing the number of accidents on service roads[3]. The study focuses on developing a regression model to predict the number of accidents along the selected stretch of road in Nashik city of India. Best fit trends for various parameters have been found to arrive at the best model from the analysis using MS Excel 2007 and Origin Software[4].

III. Material And Methods

A. Problem Statement

The objective is to build systems with automatic classifying capabilities. Save time of people in emergency and tell them to choose alternative paths. Predict severity of incident on road.

B. Classifiers:

1. k-Fold Cross-Validation:-

Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure splits the given data sample into k groups. The procedure is called k-fold cross-validation.
Cross validation technique estimate how accurate predictions your model will give in practice.
Cross validation divides original data into training data set into two parts:

- Cross validation training set
- Cross validation testing set or Validation set

Following steps can be used to perform K fold cross validation:
1. Divide the original training data set into k equal subsets. Each subset is called a fold. Let the folds be defined as F1, F2. . .Fk.
2. For i = 1 to i = k
3. Keep the fold _ as Validation set and keep all the remaining k-1 folds in the Cross validation training set.
4. Train your machine learning model using the cross validation training set and find the accuracy of your model by validating the predicted results against the validation set.
5. Estimate the accuracy of your machine learning model by averaging the accuracies resulting in all the k cases of cross validation.

2. Ensemble learning:-

In statistics and machine learning, ensemble methods affect multiple learning algorithms to gain better predictive performance than could be obtain from any of the basic learning algorithms alone. Unlike a statistical ensemble in statistical mechanics, which is usually infinite, a machine learning ensemble consists of only a concrete finite set of substitute models, but typically allows for much more flexible structure to subsist among those alternatives.

Figure 1: Shows the basic architecture of ensemble learning.

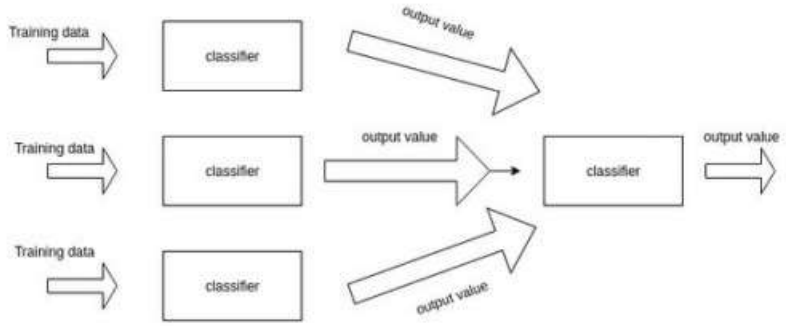

Fig 1: Ensemble learning

An ensemble is a supervised learning algorithm, since it can be trained and then used to make predictions. The trained ensemble, therefore, represents a single hypothesis. This hypothesis, however, is not essentially contained within the hypothesis space of the models from which it is built. Thus, ensembles can be shown to have more flexibility in the functions they can represent. This flexibility can, in theory, allow them to overfit the training data more than a single model would, but in actual fact, some ensemble techniques (especially bagging) be likely to reduce problems related to overfitting of the training data[citation needed].

Empirically, ensembles tend to defer better results when there is a significant diversity among the models. Many ensemble methods, as a result, seek to promote diversity among the models they merge. Even though non-intuitive, more casual algorithms can be used to create a stronger ensemble than very purposeful algorithms (like entropy reducing decision trees). Using a diversity of strong learning algorithms, however, has been shown to be more effective than using techniques that attempt to dumb-down the models in order to promote diversity.

3. Naïve Bayes:-

Naive Bayes classifiers are a collection of classification algorithms based on probability theory and Bayes' Theorem. They are probabilistic, which means they predict membership probabilities for each class such as the probability that given record or data point belongs to a particular class. The class with highest probability is most likely class. They share a common principle, i.e. all the features are unrelated to each other.

Bayes' Theorem:-
Bayes' Theorem calculates the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem is stated mathematically using following equation:
$P(A|B)= \frac{P(B|A)P(A)}{P(B)}$
Basically, to find how often event A happens, given the event B happens. .P(A|B) is a posteriori probability of B. Event B is also termed as evidence. P(B|A)is likelihood. P(A) is the priori probability of A .

4. Support Vector Machine:-

Support Vector Machine (SVM) is mostly used for classification. SVM works on the principle of margin calculation. It mainly, draws margin between the classes. The margins are drawn in such a way that the distance between the margin and the classes is maximum which minimize the classification error.
Figure 2: Shows the basic support vector machine.

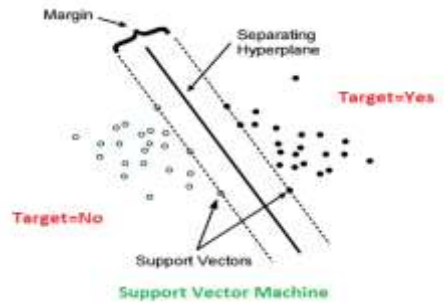

Fig 2: Support Vector Machine

Support Vector Machines makes use of decision planes that define decision boundaries. A decision plane is one that differentiate between a set of objects having different class memberships.
Figure 2: Shows the classic example of a linear classifier, i.e., a classifier that separates a set of objects into their respective groups (GREEN and RED in this case) with a line.

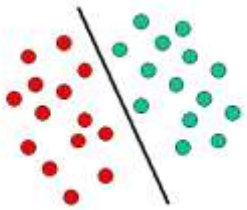

Fig 3: A schematic example of linear classifier

5. ***K-Nearest Neighbours***

K-Nearest Neighbours is the classification algorithms in Machine Learning. It is supervised learning algorithm and finds intense application in pattern recognition, data mining and intrusion detection. K-Nearest Neighbour performs classification by matching the test instance with training examples and decides its class based on the similarly to its nearest neighbors.
Steps of KNN Algorithm:-

1. Compute the distance from the test record to other training records.
2. Identify k nearest neighbours based on distance metric.
3. Make use of class labels of nearest neighbours to determine the class label of test record. If there is one class label in the neighbours, then the test record is assigned to the same class label. If the neighbours have more than one class labels, then the class label of test record is determined by taking majority vote.

C. Data Dictionary:-
The table I shows data dictionary for input to the system with 13 features with value domain for each feature.

Feature Name	Feature Value
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	Speedbraker, pithole, tree
Vehical Number	Vehicle number
Date	dd-mm-yyyy
Lighting Condition	Night light on, Night light off,day(NOL,NOF,day)

Table I- Data dictionary

D. Architecture of System:

Fig.1 shows the architecture of system. The dataset is given as input to Cross-validation module with which accuracy of training can be improved. After ensemble learning can be used for prediction of results. Those predictions of incidents severity can be used for notifying users travelling on that road.

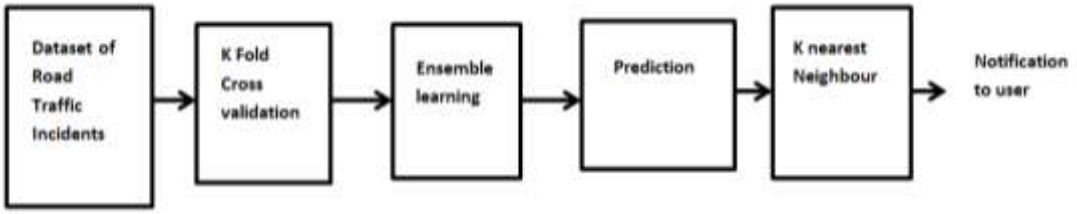

Fig. 1. Architecture of system

E. Working Principles*:*

1. Use Case Diagram:

As shown in fig. 2.As shown in fig. 2. The use case diagram contains 3 actors. And their functionalities are shown below:

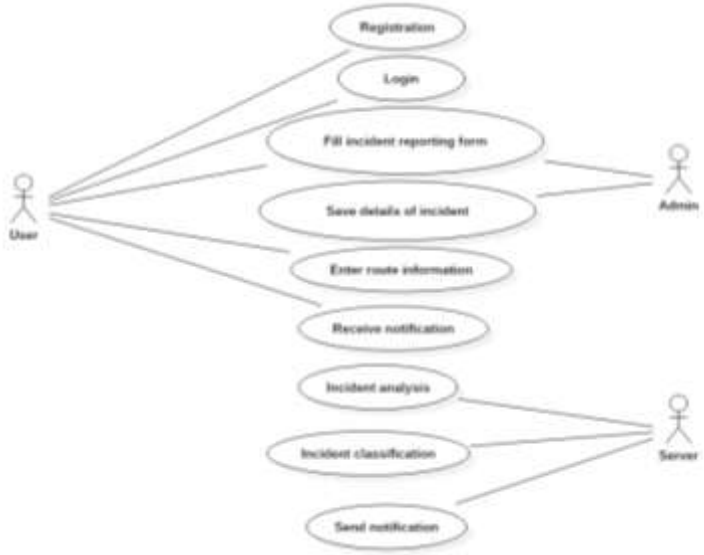

Fig.2 Use Case Diagram

IV. Result

Evaluation Parameters:-

The primary evaluation metrics are precision (P), recall (R) and F1-score (F). Precision is the rate of retrieved instances that are correct while recall is the proportion of relevant instances that are retrieved. Harmonic mean of precision and recall can be given using F1 score. They are calculated based on the true positive (TP), true negative, false positive (FP) and false negative

Precision=$\frac{TP}{TP+FP}$ Recall=$\frac{TP}{TP+FN}$ F1=2*$\frac{Precision*Recall}{Precision+Recall}$

Fig. 3 shows workflow of the system with necessary modules of applications.

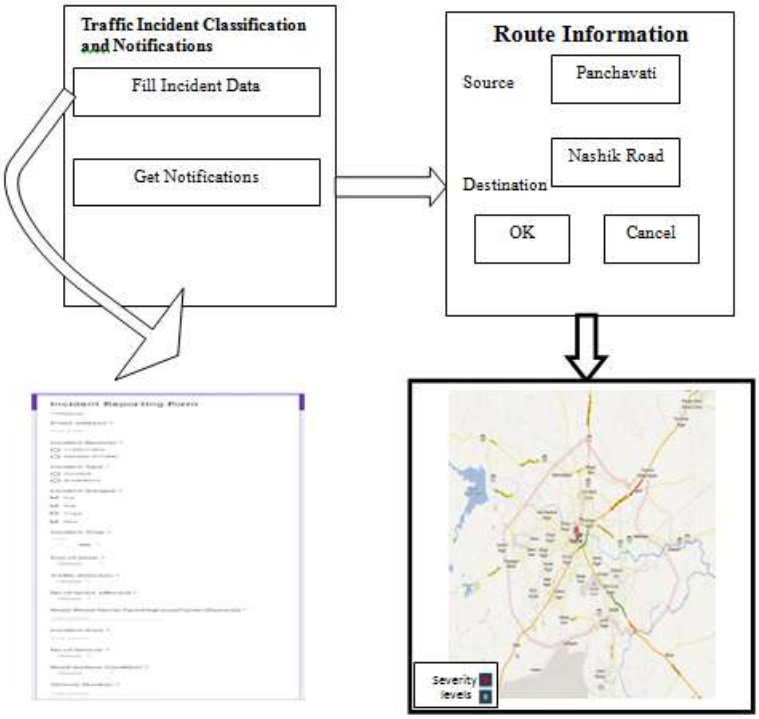

Fig.3 System Workflow

V. Conclusion

This project carried out particular analyses, visualizations and category of the incidents said from nashik road network. It's miles able to imparting beneficial statistics in a actual-time way. This can assist the administrator to respond more quick to incidents that would doubtlessly affect the ordinary flow of site visitors, and to make higher management. The ideal ml technique for incident severity type is implemented. The model is successful to study from incident records and make prediction in actual time. Ml algorithms specifically uncertainty sampling can be applied to optimize the training set or pick the maximum informative incident record to update the model's knowledge.

References

[1]. 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.

[2]. 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.

[3]. Anuja H. Vanjari, Madhura C. Aher, Shrikant R. Baviskar.: Road safetyaudit of service roads adjacent to national highway- 3 passing through nashik.

[4]. V.S. Aher1, K.H. Ahirrao2, M.R. Gaikar3, S.S. Surve4, R.R. Kshatriya5 .:Road Accident Prediction for Nashik City.