Classification and Prediction of Road Accidents in Hilly Area

|  |
| --- |
| *Sahil Thakur* |
| Apex Institute of Technology,  Chandigarh University, Gharuan, India  19BCS6103@cuchd.in |

**Abstract -**Road accident prediction is one of the most important research area in traffic safety. The occurrence of road traffic accidents is mainly affected by geometric characteristics of road, traffic flow, characteristics of drivers and environment of road. Many studies have been conducted to predict accident frequencies and analyse the characteristics of traffic accidents, including studies on hazardous location/hot spot identification, accident injury-severities analysis, and accident duration analysis. Some studies focus on mechanism of accidents. Other factors include weather and light conditions of the road.

# INTRODUCTION

With the exponentially increasing number of vehicles, road safety is a matter of huge concern. Road accidents kill 1.2 million people every year. Road crashes cost $518 billion globally, costing individual countries from 1-2% of their economy. In 2017, there have been 2367 accidents with injuries reported in Hyderabad alone. Steps are being taken to combat this issue but they have been ineffective. Some studies focus on mechanism of accidents. Other factors include weather and light conditions of the road. So, ML powered web app which predicts accidents severity based on the conditions is best solution for this problem. It is trained with accident records from the online dataset. More data means greater accuracy.

# BACKGROUND

Lee et al [1] developed a probabilistic model relating significant crash precursors to changes in crash potential. Abdel [2] built a previous crash prediction model with the matched case-control logistic regression technique. No specific approach available for the traffic police to predict which area is accident prone at a specific time. The traffic accident prediction play an important role in the integrated planning and management of traffic, the reason which with much randomness about the traffic accident include some nonlinear elements, such as people, car, road, climate and so on.

The traditional way of linear analyses cannot reveal the really situation since the noise pollution and amount of data are too little, cause the result of prediction cannot satisfactory. Because of the traditional BP network have some defects, such as local minimum, too many iterations, training too slow and so on. The traditional Back propagation network has defects. It has a 7.8% lower accuracy than the proposed model.

An ML powered web app which predicts accidents severity based on the current conditions. It is trained with accident records from the Kaggle dataset. More data means greater accuracy. The purpose of such a model is to be able to predict which conditions will be more prone to accidents, and therefore take preventive measures. We will even try to locate more precisely future accidents in order to provide faster care and precaution service. By using those properties, we train a combination of different machine learning algorithms using various ensemble methods that are not thoroughly explored in the current literature.

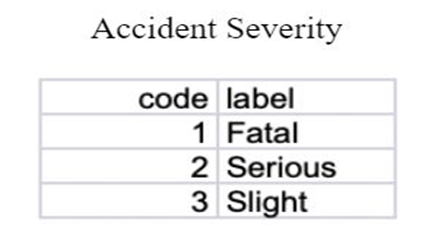
1. **METHODOLOGY**

We have developed a web app for our model. It consists of four components:

**Front-End**: Users input for the prediction factors are taken and sent to the backend server.

**Back-End**: The model is deployed here and the input data is fed into the Machine Learning model.

**Machine Learning Model**: We have used decision tree, random forest and logistic regression and also applied hyper parameter tuning to increase its efficiency. Random Forest algorithm showed the highest accuracy of 86.86% and hence chosen for our model. The model runs and predicts the severity. The severity metrics are 1= Fatal, 2= Serious, 3= Slight.



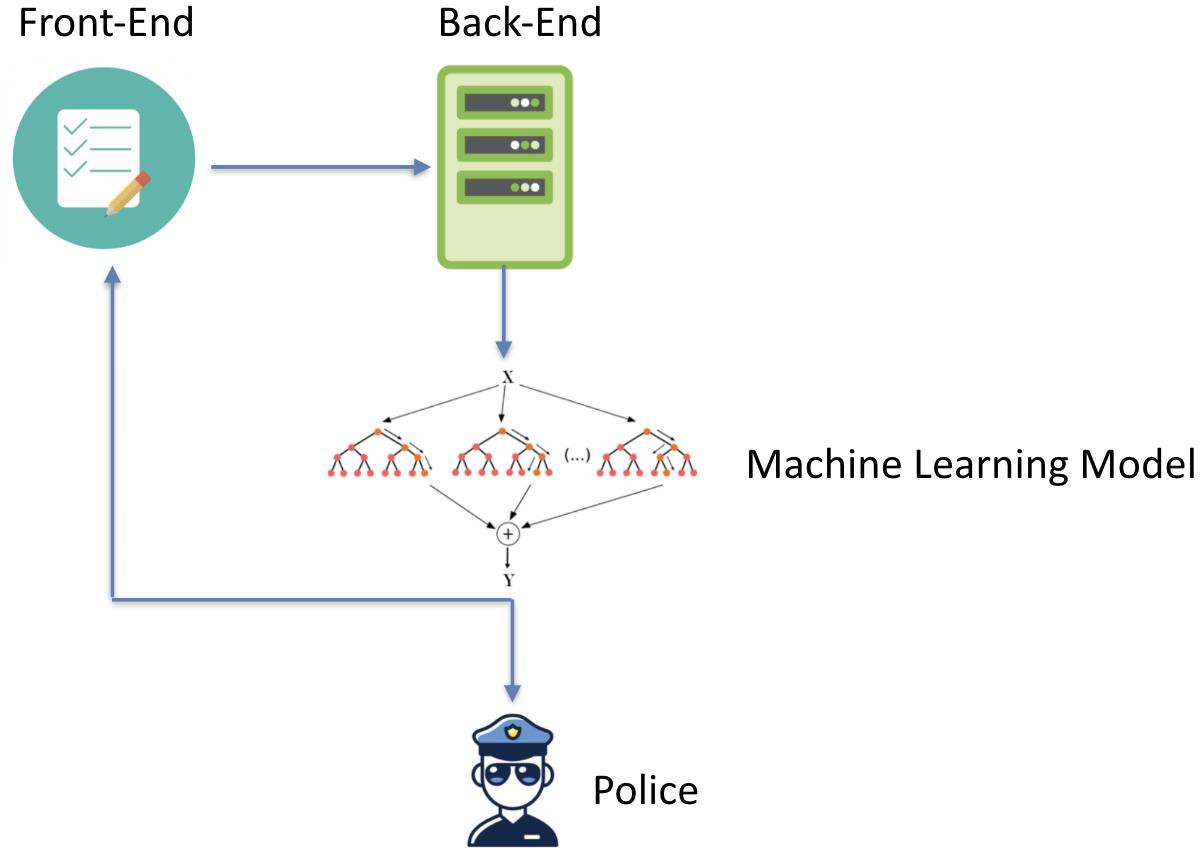
**Fig 3.1 Accident Severity**

The output is sent back to the front-end and displayed to the user.

An sms containing the location coordinates and the severity of accident is sent to the police so that it can take preventive measures at the location**.**

**3.1 System Design**

Describes the data flow in a diagrammatic representation.



**Fig 3.2** System model

**3.2 Modules**

1. **The Virtual Machine**: It has the trained and tested Machine learning algorithm implemented. The frontend and backend server are deployed on it.

2. **The front end (User)**: Geolocation Api takes the location of the user and sends it to the OpenWeatherMap Api which sends geographical conditions. User input is taken for other parameters like age, sex etc. User can view the heatmap of the accidents in the country.

3. **The back end (Admin):** The server is created and maintained. The input details are feeded to the model and severity is predicted. The severity can be sent as a message or email to the police to take preventive measures.

4. **Machine Learning Algorithm:** Classification Algorithms decision tree, random forest and logistic regression have been implemented. Hyper parameter tuning has been applied to find the best accuracy. Random forest has shown the highest accuracy with 86% and has been selected as the model for the web app.

**3.3 Implementation of Proposed solution**

There are four important steps:

1. Preprocessing

2. Training

3. Testing

4. Web App Integration

**3.3.1 Data Importing**

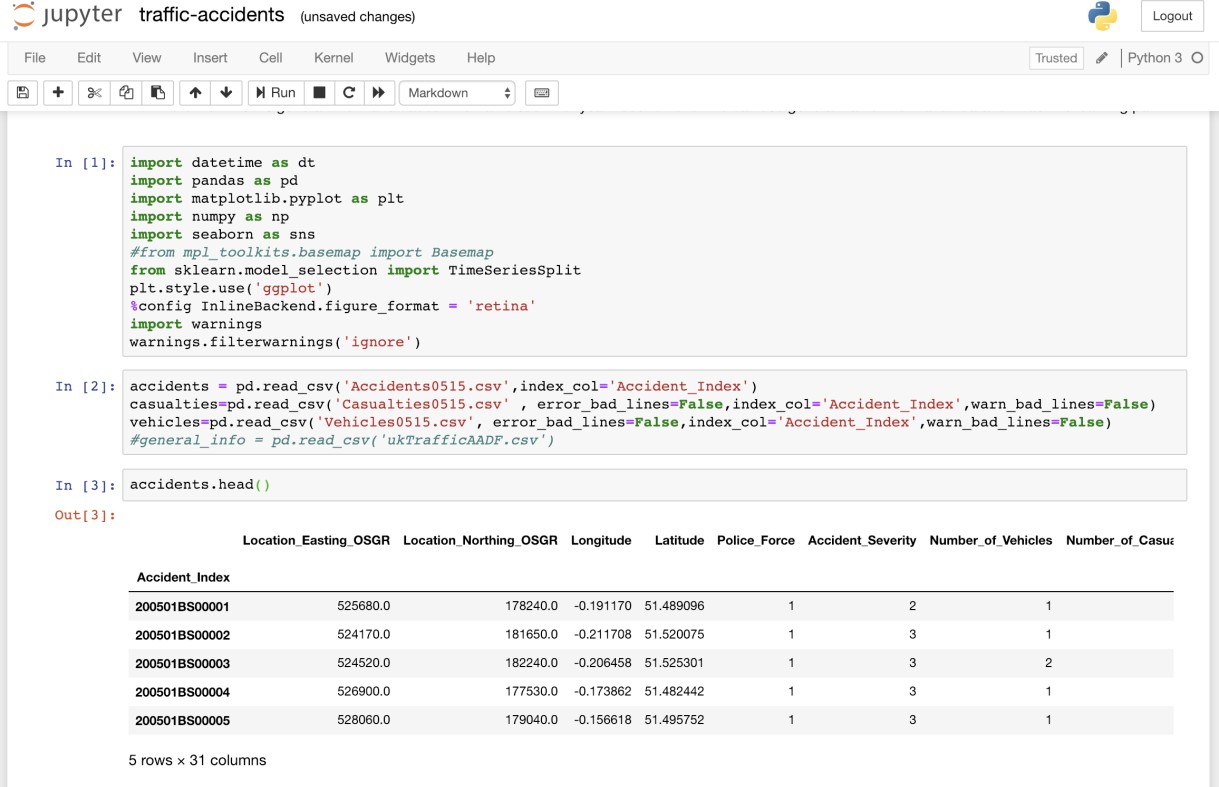
We import three files to perform analysis on this data. This data is consist of three files that are accidents, casualties and vehicles. However, we have one more file which is general information about the traffic count for year 2000 to 2015. We can use general traffic information data for machine learning part.

● Importing of packages needed is done.

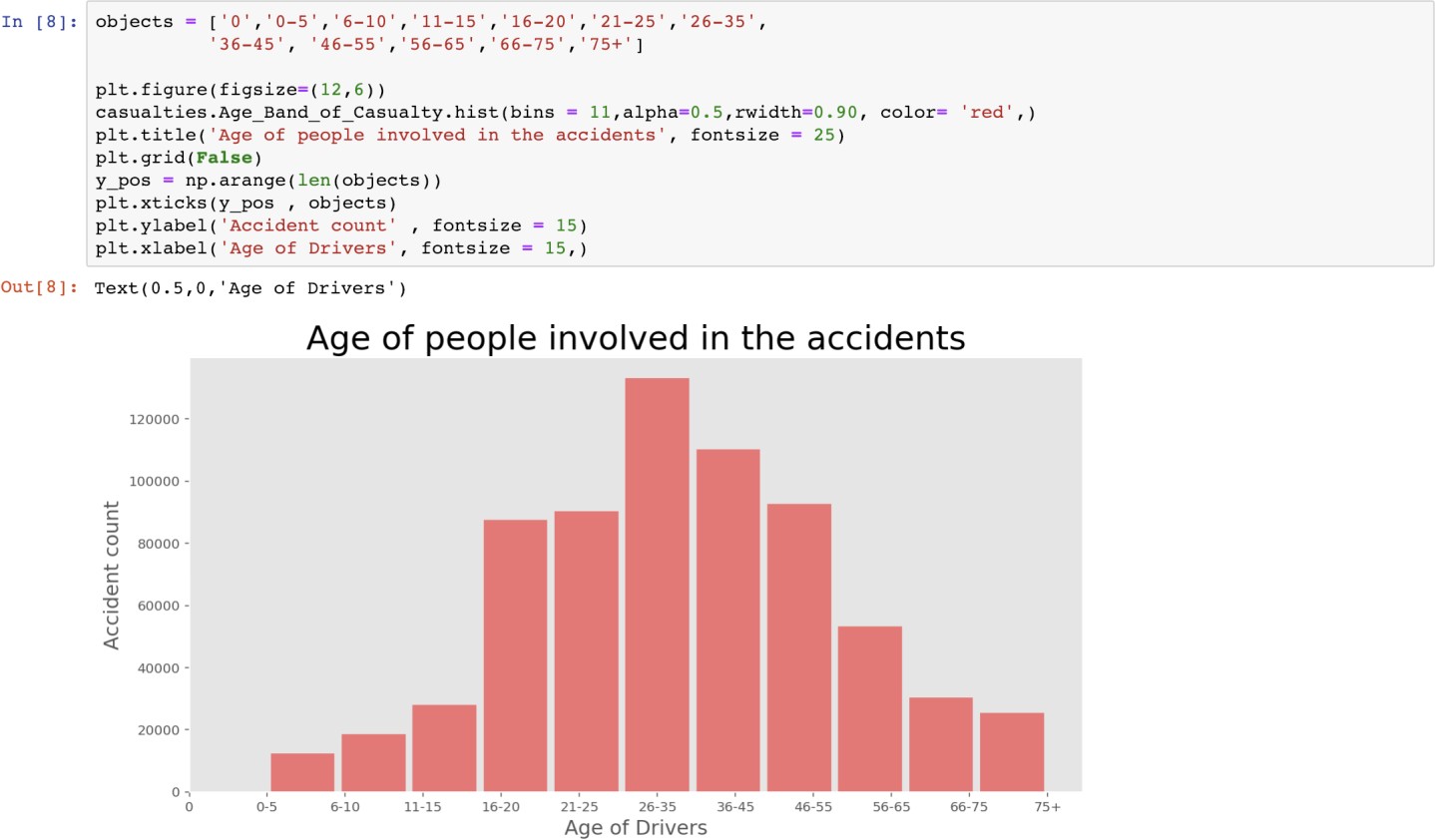
● 3 CSV files Accidents.csv Casualties.csv Vehicles.csv

● Using pandas to import data into dataframe.

● accident.head() views top 5 rows of dataframe.



**Fig 3.3 Importing**

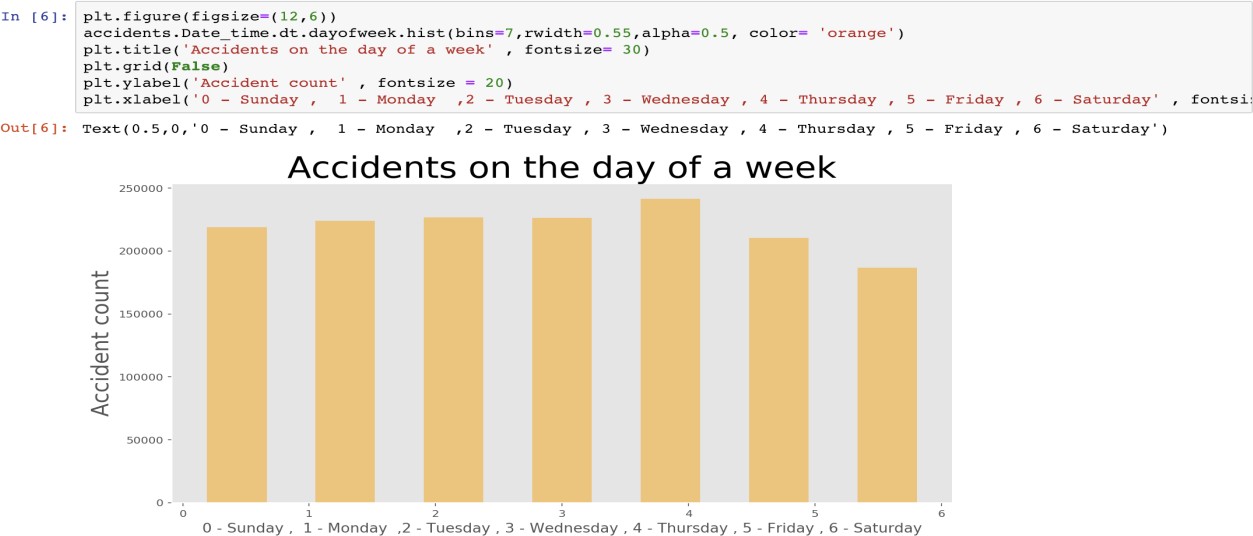
**3.4 Data Visualization**

The first thing we can do is to find out about accidents time to get intuition and some driver's age who are involved in the accident.

● We can find out the number of accidents on the days of a week.

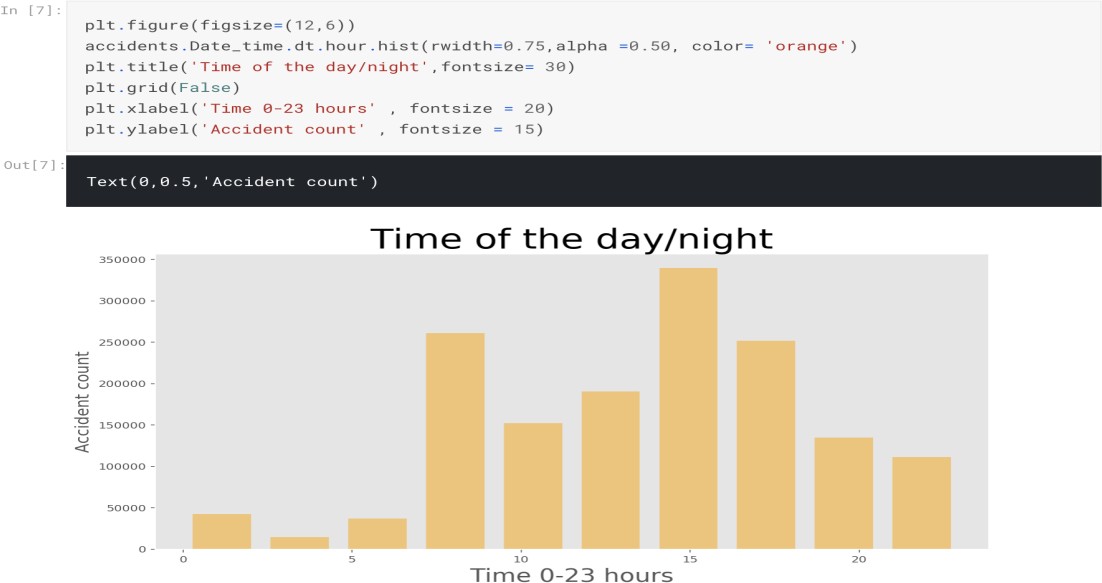
● We can find out about the accidents number using hours of the day.

● Finding out about the age of driver can tell us more about the accidents.



**Fig 3.4 Accidents on the days of a week.**

We can find out the number of accidents on the days of a week. As we can see that Thursday has the highest amount of accidents in this dataset from 2005 to 2015. We have to keep in mind that accidents numbers could be depending on traffic amount on particular day.



**Fig 3.5 Accidents number using hours of the day.**

He we found out that the most of accidents happened around after noon. We can assume that this time of the day has the most traffic moving such as people leaving from work.

**Fig 3.6 Age Band of Casualties.**

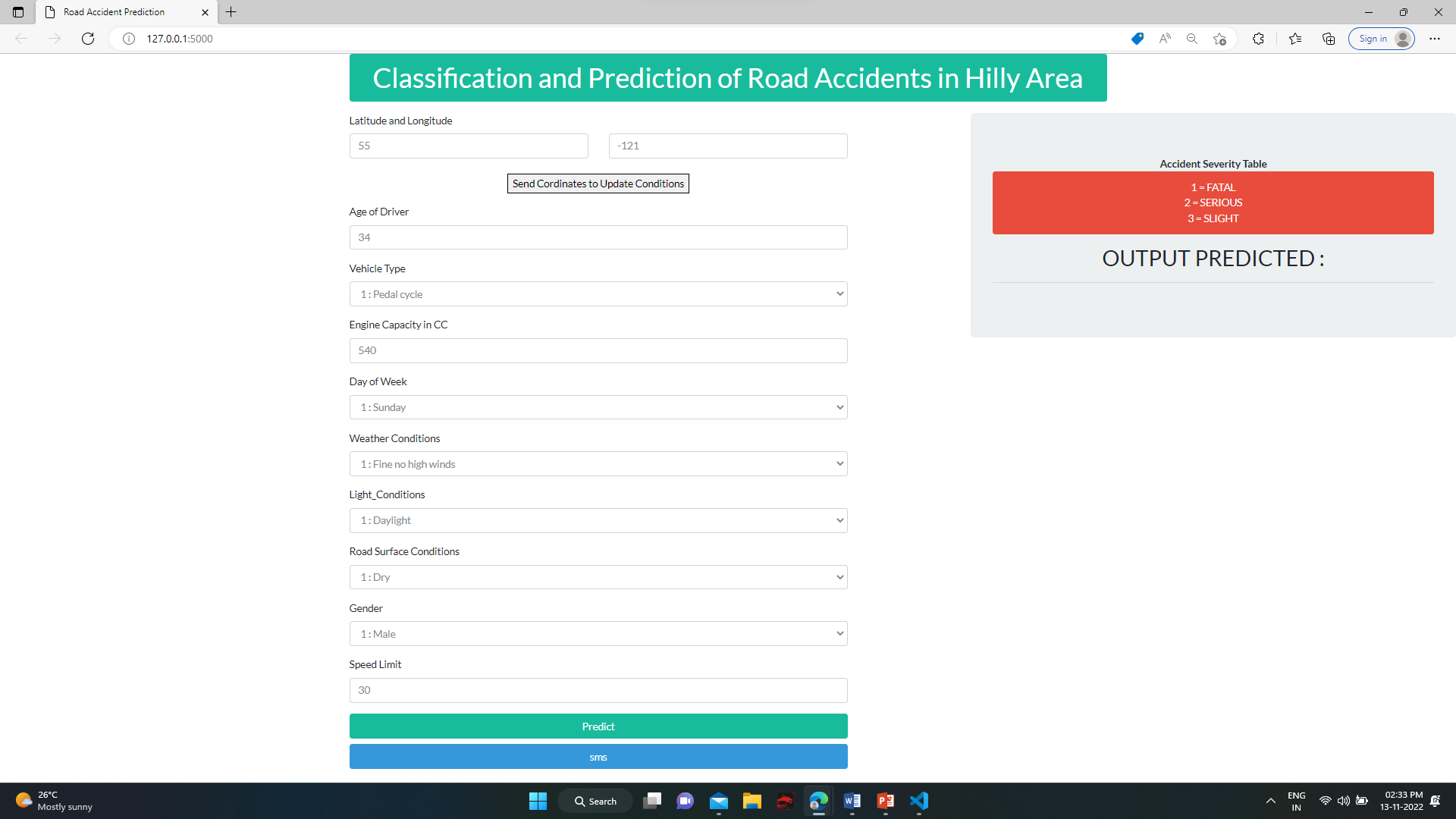
In this dataset, age band is grouped in 11 different codes. We will create the labels and pass it to the plot so we can have idea about the bins representation this is very interesting fact about this dataset. Most of the driver’s age is around 225 to 35 who are involved in the accident. However, we do not know the number of drivers with age 25 to 35 on the road compare to other ages. Intuitively, I would assume that the driver with age 25 to 35 are more in the number of drivers with different age.

1. **RESULTS**

A web app has been developed for my model. It can be easily accessed through the custom domain name https://road-accident-predict.herokuapp.com/. The model is deployed in the back-end. The input data from the front-end is fed into the Machine Learning model. We have used Random Forest algorithm which showed the highest accuracy of 86.86% as our model. The model runs and predicts the severity. The severity metrics are 1= Fatal, 2= Serious, 3= Slight.

The output is sent back to the front-end and displayed to the user.

An SMS containing the location coordinates and the severity of accident is sent to the police so that it can take preventive measures at the location.



## Fig. 5: GUI

**V. CONCLUSION AND FUTURE WORK:**

This project aims at using Machine Learning classification techniques to predict severity of an accident at any particular location. Machine Learning has enabled us to analyze meaningful data to provide solutions with a greater accuracy than with humans. We have built a model with an accuracy greater than 10% of the conventional system [1]. A web-based app using the most accurate algorithm has been developed which can be accessed through the domain name https://www.accidentprediction.com:4000.

This project can be used by governments to prevent accidents.

With more resources, continuous prediction and alerts can be sent to the police for every location at regular intervals of time to take preventive measures. The web app can be incorporated with Google Maps which can be live tracked by the police. A fully-fledged web app for user and police interaction can be published for use in real-time. It can be used for Indian states or cities, if proper data of accidents is provided by the Indian Government.

**VI. REFERENCES:**

1. Lu Wenqi, Luo Dongyu & Yan Menghua, “A Model of Traffic Accident Prediction”
2. INSPEC Accession Number: 17239218 DOI: 10.1109/ICITE.2017.
3. Thineswaran Gunasegaran Yu-N Cheah, “Evolutionary Cross validation”
4. INSPEC Accession Number: 17285520 DOI: 10.1109/ICITECH.2017.
5. Simon Bernard, Laurent Heutte and Sebastien Adam, “On the Selection of Decision Trees in Random Forests” INSPEC Accession Number: 10802866 DOI: 10.1109/IJCNN.2009.
6. Rafael G.Mantovan,, Ricardo Cerri, Joaquin Vanschoren, “Hyper-parameter Tuning of a Decision Tree Induction Algorithm” INSPEC Accession Number: 16651860 DOI: 10.1109/bracis.2016.018
7. Fu Huilin, Zhou Yucai, “The Traffic Accident Prediction Based on Neural Network”, 2011.
8. Lin, L., Wang, Q., Sadek, A.W., 2014. Data mining and complex networks algorithms for traffic accident analysis. In: Transportation Research Board 93rd Annual Meeting (No. 14-4172).
9. Gunasegaran, T., & Cheah, Y.-N. (2017). Evolutionary cross validation. 2017 8th International Conference on Information Technology (ICIT). doi:10.1109/icitech.2017.
10. Bernard, S., Heutte, L., & Adam, S. (2009). On the selection of decision trees in Random Forests. 2009 International Joint Conference on Neural Networks. doi:10.1109/ijcnn.2009.