Clusterization algorithm. It helps to ascertain the traffic violation patterns and blackspot of traffic violations. We looked into K-means clustering with some enhancements to aid in the process of identification of patterns and blackspots. We applied these techniques to real traffic data extracted from the Montgomery County of Maryland and validated our results. We also developed a prioritized scheme for attributes here to deal with the limitations of various out of the box clustering methods and ways. This easy to implement data mining framework works with the geo-spatial plot of blackspots and helps to improve the road accidents zones.

Development
With the increasing number of vehicles, traffic violations occur frequently and hence an accident. There are some situations or cases where the traffic violations gets eventually transformed into accidents occurring at the same location or location more than once which could be at a sharp drop or corner. Road traffic accident hazardous locations are called black spots, i.e., the road sections or intersections where traffic accidents are high. Location wise very often one hazardous spot is the first step to improve the road safety state which is indeed a very important state. Therefore, identification of traffic black spot, analysis of the cause of traffic black spot and then reliable operations to improve traffic safety in short duration of time. But, traffic violations are mainly the causes which render traffic accidents. If traffic violations can be controlled efficiently, traffic accidents can be reduced remarkably. Taking the reference of the traffic violations which caused accidents, traffic violation black spots can be obtained. It can reduce traffic violations efficiently through improving the traffic condition of traffic violation black spots.

Data Mining Methods
Many data mining methods can be used to forecast traffic violation, such as linear regression method, exponential smoothing method, Kalman filtering method, BP neural network, Support Vector Machines(SVM) and so on. SVM is popular network used for classification and regression. In the SVM method, it is a burdensome work to fix the kernel parameters. Bayesian interference can improve the process markedly. We will utilize the nonlinear regression SVM whose kernel parameters are fixed with Bayesian interference method to forecast traffic violations. Then, proposed identification method is used to identify traffic violation black spots.

Goal/Objectives
There are several public transportations like bus, shuttles, cab and so on as well which are all summing up to make a huge number of vehicles on the streets and highway. But, even when the road conditions are perfect, a traffic violations could turn into a hazardous accident.

Road traffic safety deals with a complexity of problems as there are some factors that contribute to its disturbance: blind curve, sharp corner, width of roads, driver’s behaviour in traffic, wear and tear of infrastructure, climate scenarios, light conditions and intense traffic. Due to continuous increase of traffic, the amount of accidents rises notably while road traffic congestion safety becomes more hard to maintain. But, out of all, there are most accident prone locations also called black spots where traffic violations turn out to be road accidents. Locating this particular accident hazardous spot is the first step to improve the road safety state. In addition to that, we can also predict a particular age-group of individuals who are mostly making these violations.

And, a specific ethnic group or age-group who mostly cause the traffic violation is also determined.

Data Set
The dataset contains traffic violation information from all electronic traffic violations issued in Montgomery County, Maryland and was extracted from Montgomery County of Maryland, data.montgomerycountymd.gov. It includes all traffic violations since January 1, 2012, so, this analysis is based on 49 months of records. There are 35 attributes and total 819565 instances on this dataset.

The main attributes that’ll be used for data mining are Location, Geo-location, Accident, Alcohol, Work Zone, Year, Violation Type, Charge, Gender, Race and so on. The attribute, geo-location and location will be used to visualize in a geo-spatial data visualization.

Data Mining Algorithms
The procedure of Traffic violation black spot forecasting can be summarized as follows. Firstly, history traffic violation records should be collected. Select the records of a road section or intersection within the analysis time period. The data are further pre-processed using specific filters in Weka. Hidden best rules are determined using Association Rules mining. Then, by using the K-means Algorithm a specific clustered data is obtained. After in-breeding, SVM is used to forecast the future traffic violation number. Every road section and intersection is processed with the procedure. Support Vector Machines(SVM) are managed learning prototype with related learning algorithms that inspects data used for sorting and regression study, seven, a set of train examples, SVM is marked for relating to one of two groups, an SVM training algorithm forms a replica that assigns new samples into one group or the other, making it as a non-probabilistic binary linear classifier.