



## Area Traffic Control using Predictive Analysis

Sanjat Manohar Mishra  
SCOPE. VIT University  
Vellore, India

Lavanya .K\*  
SCOPE. VIT University  
Vellore, India

**Abstract:** Traffic control in a targeted area is one of the most discussed topics in today's scenario. The increase in number of vehicles in the past few years, has caused a spike in traffic jams at intersections, disrupting entire roadway transportation. These situations demand more efficient methodology to control traffic. This paper proposes a framework to counter the above stated issues using a combination of efficient optimization and decision making algorithms. The framework relies on gathering real time data from commuters, transmitting and analyzing the same to determine vehicle densities on specific routes. The vehicle density on a specific route becomes an important factor in analyzing the traffic conditions in that specific area. The framework also proposes a model to optimize the traffic signal timers at a targeted traffic junction or cross-section, by performing predictive analysis using decision trees.

**Keywords:** Predictive Analysis, Decision Trees, Traffic Optimization, Signal Timer, GPS , Data Analysis , Machine Learning.

### I. INTRODUCTION

The problem of optimizing traffic signal lies at the core of urban area traffic control. Traffic signal control consisting of delay, pollution, queuing, fuel consumption is a multi-objective optimization problem. For a signal-controlled traffic network, use of optimization techniques in determining signal timings has been discussed extensively for past few years, new methods and approaches are being developed to improve the efficiency of traffic signals. Optimization of traffic system is required because they are used to control the vehicle movement which in turn may lead to reduce congestion, improve safety, and enable specific strategies such as minimizing delays, improving pollution etc.

In the present scenario, we have TRANSYT-7F, which is most useful for optimizing traffic signals and is one of the most widely used tools in its category. It consists of two parts:

1.) Traffic flow model and, 2.) Signal time optimizer. The traffic flow model uses a platoon dispersion algorithm, it simulates traffic in a network of signalized intersections to produce a cyclic flow profile of arrivals at each intersection that is used to compute Performance Index(PI). The optimizer uses Hill-Climbing and Genetic Algorithm strategies to reach at an optimal solution.

We propose a different optimization technique for solving the problem. We are trying to solve the problem discussed above using Decision Tree algorithm for predictive analysis of vehicle density. By predicting the vehicle density on each route at different hours of the day, we develop a trend according to which the traffic signal timer is set. This is based on real-time data collected from users which gives us a fairly accurate summary of the number of vehicles travelling on a particular route and about to reach a traffic cross-section.

Furthermore, we propose a reroute mechanism for the users which determines traffic density from source to destination

and proposes an alternate route with lesser traffic congestion. This is achieved by collecting real time vehicle data from the users using GPS. GPS location from each user in real-time helps us determine the vehicle density on a particular route.

The core model of the reroute mechanism is targeted on bypassing traffic congestion in particular areas. This is not only time-saving for the commuters, but may also prove life-saving in case of emergency services. Efficient rerouting mechanism ensures that the user reaches its destination in minimum time. This is particularly useful in case of ambulances, firetrucks, police et cetera. After the real-time collection of vehicle data from the users, this data is analyzed using machine learning or predictive analysis, which in turn gives us an overall idea about how the traffic conditions might be in the next couple of hours. The analysis based on the current traffic trends in a particular area and might not be entirely accurate. The user can view the results of the predictive analysis and therefore make an informed choice of selecting the route to his destination based on both real-time conditions as well as future predictions.

In this paper, we propose to use decision trees for predictive modelling. The real-time data collected from the users contains many variable, which have to be refined and selected for proper predictive analysis. One way to do this is by removing variables which are independent of the input and have little correlation to the target or output. From the received input, a trend is analyzed and extracted by using the data on vehicle densities on a specific route at different times of the day. This analysis concludes in the peak hour timings of the day. The peak hour timing gives us an estimate about the time of day at which the traffic is maximum on a particular route. This peak hour timing is an intermediate output which helps the reroute mechanism. Peak hour finding can also be helpful to the end user as it may help the user to select the route and time of travel more efficiently, thereby avoiding harsh traffic conditions.

Peak hour finding is especially helpful for everyday commuters, for example, people travel every day to work and need to reach by a particular time. Peak hour timings on a particular route can help the commuter avoid traffic and reach the destination under least possible time.

With this paper, we aim at reducing traffic congestion in a targeted area using optimized signal timers, traffic predictive analysis as well as vehicle density analysis. These modules can be further enhanced to provide safest route by analyzing past accidental reports in the target area. This can further help the end users in avoiding areas which are prone to accidents.

## II. OVERVIEW OF DECISION TREES

Decision tree is a graphical representation of possible solution to a decision based on certain conditions. It is called decision tree because it starts with a single box (or root), which then branches off into a number of solutions, just like a tree. It is like a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including event outcomes, resource costs, and utility. It is also one way to visualize an algorithm. They help in formulizing a strategy which is most likely to reach a goal efficiently, it is also a popular tool in machine learning. [1]

Decision trees are helpful, not only because they help us visualize what we are thinking, but also because making a decision tree requires a systematic, documented thought process. [2] Often, the biggest limitation of our decision making is that we can only select from the known alternatives. Decision trees help formalize the brainstorming process so we can identify more potential solutions.

Decision Trees are excellent to for helping us to choose between several courses of action. They also help us to form a balanced projection of the risks and rewards associated with each possible course of action.

Decision tree basically has a flowchart like structure, in which internal node represents a test on an attribute, each branch represents the outcome, and each leaf node represents a decision taken after computing all the attributes. The path from root to leaf represent classification rules.

A decision tree mainly has of three types of nodes:

1. Decision nodes
2. Chance nodes
3. End nodes

Advantages:

1. They are simple to understand and interpret. People can understand decision tree models after a brief explanation.

This is because the tree structure remains intact with or without scale changes. Data preparation time is reduced even further as it is not effected by missing values in the input data set. Any missing value doesn't affect the splitting of the trees into branches.

2. Have value even with little hard data. Important insights can be generated based on experts describing a situation (its alternatives, probabilities, and costs) and their preferences for outcomes. Allow the addition of new possible scenarios
3. Help determine worst, best and expected values for different scenarios.
4. Use a white box model. If a given result is provided by a model.
5. Can be combined with other decision techniques.

Decision tree learning uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value.

Classification Tree- Tree models where target variable can take a finite set of values.

Regression Tree- Tree models where target variable can take continuous values. [3]

Algorithms for constructing decision trees usually work top-down, by choosing a variable at each step that best splits the set of items. Different algorithms use different metrics for measuring "best". These generally measure the homogeneity of the target variable within the subsets. [4] These metrics are applied to each candidate subset, and the resulting values are combined (e.g., averaged) to provide a measure of the quality of the split.

## III. DECISION TREES IN AREA TRAFFIC OPTIMIZATION

The basic idea involved while choosing decision trees for solving the area traffic optimization problem is its ability to make accurate predictive analysis based on past data. The ability of decision trees to implicitly perform feature selection or variable screening proves very useful in reducing the time and effort required to manually select features from the received data to create a hypothesis.

The few nodes, where the tree actually splits, is based on the few important parameters shortlisted during feature selection. The data collected during real time contains parameters like vehicle name, vehicle details, GPS real-time location, time at which the vehicle is on a particular route, source address, destination address et cetera.

Another major reason for selecting decision tree for predictive analysis is their ability to reduce efforts for data preparation. To fit a regression model and interpret the coefficients, proper scaling of data has to take place. But such variable transformations are not required with decision trees.

The parameters that we receive from the vehicles may not be in linear relation with each other. For example, the destination address and route do not have any correlation. Decision trees solve the problem as the non-linear relationship between parameters doesn't affect the trees performance.

Another important reason for the selection of decision tree for area traffic is its ease to interpret and explain to users. The

only thing to look out while using decision trees is to prevent overfitting of training data, which may result in poor decision. Overfitting in decision trees can largely be avoided by limiting tree growth by fixing a number of features, or by proper pruning.

#### IV. METHODOLOGY

##### Peak Hour Finding:

The peak hour finding functionality depends on the dataset containing the data on vehicle density trends on particular routes during the previous months, or even years. This will become the basis to analyze and come up with a trend close to finding the peak hours on a specific route.

The peak hour finding functionality depends on the vehicle density trends on every route in the past couple of months. The peak hour finding functionality works on regression analysis and modelling. Here, we choose to perform regression analysis as it helps in determining the relationship between a dependent(target), which is the peak hour timing and the independent (predictor), which in this case is the vehicle density.

After collection of data of past several months, this data is the cleaned to remove any inconsistency or incompleteness. After removal of invalid and irrelevant data, we choose parameters like route details including route name, route length et cetera. We then determine the number of vehicles on a particular route at different hours of the day. In this way, each route present in the data set will be allotted vehicle densities for different hours of the day. The peak timing for a particular route will be decided by choosing the hour of the day with maximum vehicle density on that particular route.

##### Reroute Mechanism:

The reroute mechanism module is a dependent module. It depends on the peak hour finding to find the peak hour timings of various routes, before it can suggest alternate routes.

The working of reroute mechanism is based on the decision tree algorithm for predictive analysis.

##### Algorithm for Creation of Decision Tree:

*Step 1.* Place the best feature or attribute at the root of the tree. In this case, the best attribute from the dataset will be vehicle density.

*Step 2.* The training dataset is split into subsets in such a way that each subset contains the same value for a data for a specific attribute.

*Step 3.* Repeat step 1 and step 2 on each subset until leaf nodes of all branches of the tree are obtained.

##### Assumptions to be made during creation of Decision Tree:

1. The whole training set is considered to be the root at the beginning.

2. Attribute or feature values determine the distribution of records. The distribution is made by recursion.
3. All feature values are preferred to be categorical. Any continuous value is discretized prior to building of the model.
4. Statistical approaches are used to order and place attributes as internal nodes or the root node.

The final output of the decision tree is either “Yes” or “No”. The decision tree will help us predict whether the route the user chooses to reach his destination will have traffic in the next couple of hours or not.

The primary hurdle faced while using decision tree is the selection of an attribute which will become a baseline to consider as the root and each level. The attribute selection can be done using the following measures:

1. Information Gain
2. Gini Index

The above-mentioned criteria will calculate the values for every attribute present in the dataset. [5] The values are then sorted, and the attributes are placed in the tree by following the order, i.e., any attribute with high value is placed at the root. It is to be taken care that while considering Information Gain as the criterion, the attributes should be assumed categorical, whereas, in case of Gini Index, the attributes are assumed to be continuous.

**Information Gain:** By selecting Information gain as a criterion, we estimate the information contained in each attribute. [6] We basically measure the randomness or the uncertainty of a random variable ‘X’, which is also known as the Entropy.

Since we are dealing with a binary classification problem, i.e., only two output classes are present:

1. If all training examples are positive, or all negative, the entropy is zero or low.
2. If half the training examples are positive, and the other half negative, then, the entropy is equal to one, or high.

$$H(X) = \mathbb{E}_X[I(x)] = - \sum_{x \in X} p(x) \log p(x).$$

**Gini Index:** By selecting Gini Index as a criterion, we measure, how often a randomly selected element is incorrectly identified. In short, we select the feature with lower Gini Index. [7]

$$\text{Gini Index} = 1 - \sum_j p_j^2$$

The main problem we need to avoid during predictive analysis is the problem of overfitting. The decision tree algorithm continues to go deeper and deeper to reduce the error during training, ending with more set errors than before. [8] This problem can be corrected by pruning and reducing irregularities in data. Two types of pruning can be done:

1. Pre-Pruning

## 2. Post-Pruning

### *Traffic Signal Timers Optimization:*

The traffic signal timer module is one of the large-scale implementations of this research framework. The traffic signal timer optimization has to be implemented on every traffic cross-section for full-fledged utilization.

The aim of this module is to strategically set timings of the traffic signals at various traffic junctions, which will result in optimized traffic control and lower congestion due to traffic.

The traffic system currently implemented in most parts of India, set a static time limit on the traffic signals. This doesn't help while facing extreme traffic conditions. The module being implemented here, uses predictive analysis to obtain an idea about the traffic conditions that may take place in the next couple of hours, and sets the timers accordingly to create the best possible scenario during heavy congestions.

Pseudocode:

*Step 1.* Collect the real-time data sets from users.

*Step 2.* Obtain the results from peak-hour analysis module.

*Step 3.* Decision Tree creation using Information Gain criterion for feature selection.

*Step 4.* Solving overfitting using post-pruning, due to large datasets.

*Step 5.* Collect results of Vehicle Density on each route in real-time.

*Step 6.* Obtain the results of predictive analysis of vehicle density at various roots.

*Step 7.* Segregate groups of routes leading up to different traffic junctions.

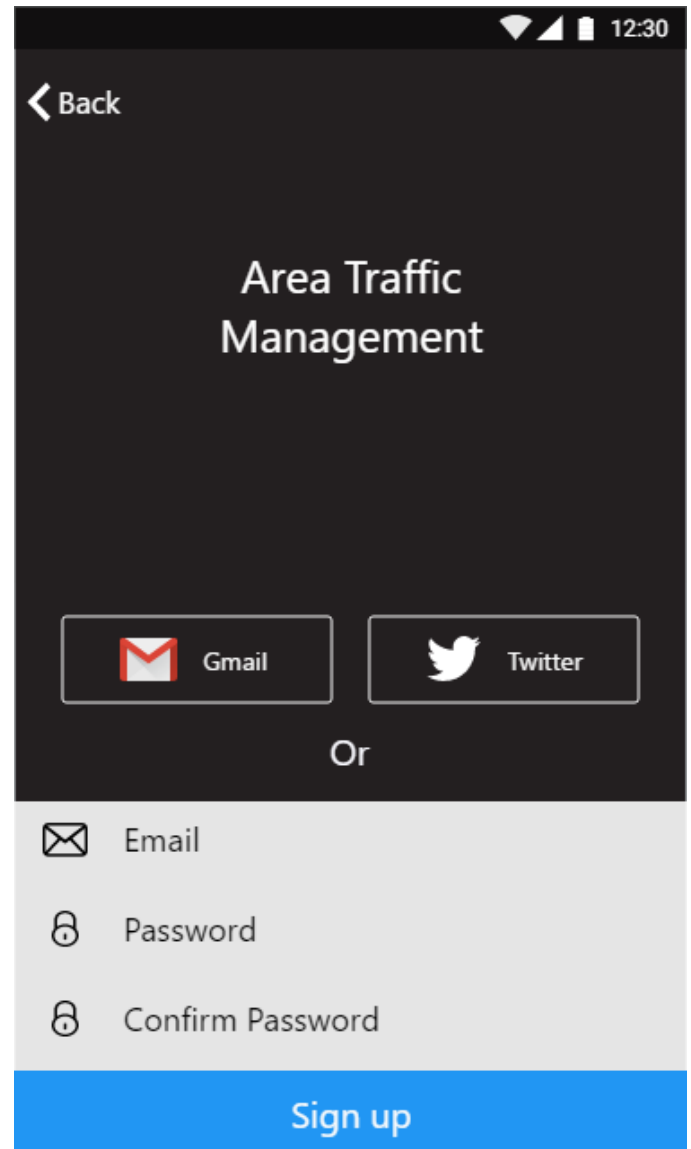
*Step 8.* Set traffic timers at different junctions according to predictive traffic density analysis.

If predicted traffic density at a route is high in the next couple of hours, short timers are set for the route in advance to clear up any incoming traffic before congestion.

If the predicted traffic density of a route is low, comparatively longer timers are set.

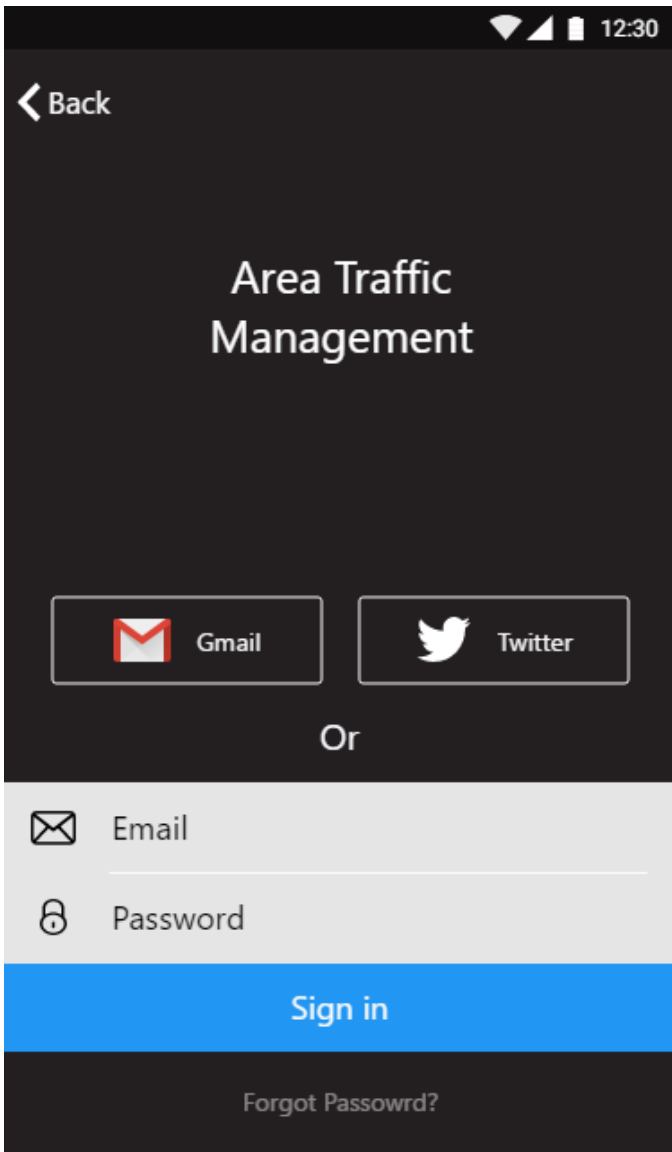
This prediction based traffic timer can reduce traffic congestion to a large extent and ensure efficient traffic control system.

## V. RESULTS



The above screenshot is taken from the signup page of the mobile android application. The user has to sign up using his credentials, as well as, vehicle details, like vehicle number, vehicle type, vehicle capacity.

The vehicle type attribute contains categories like two-wheeler, 3-wheeler, 4-wheeler, car-pool service et cetera. These attributes make a part of the dataset that will be required for predictive analysis.



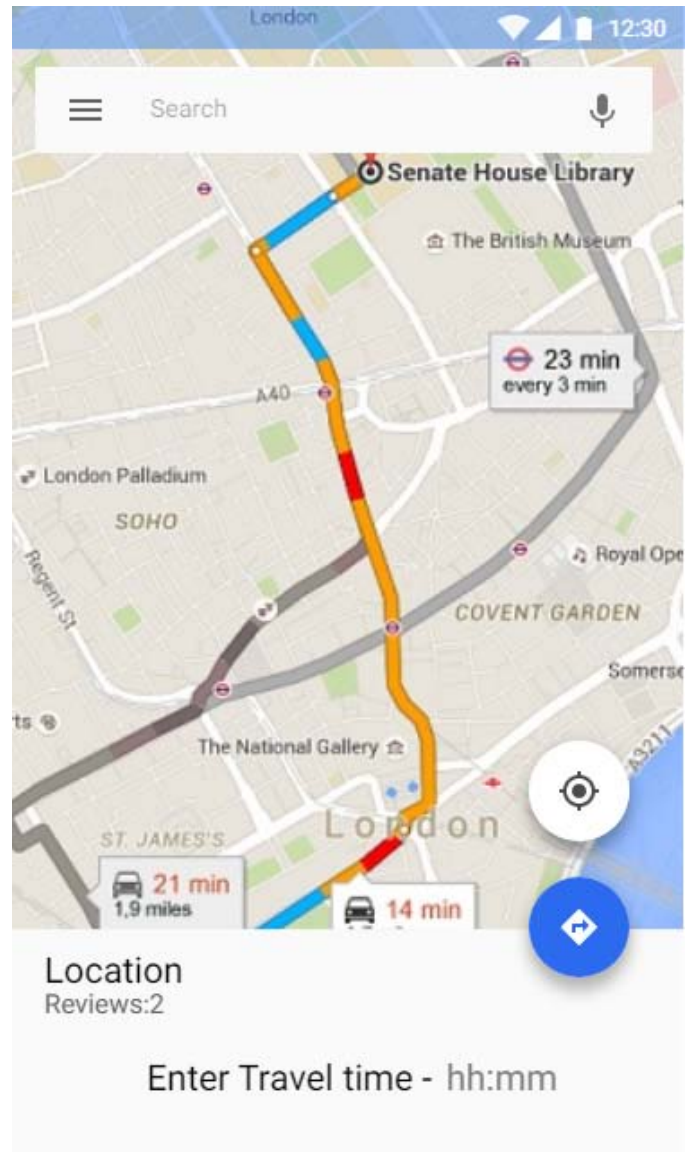
The above image is the screenshot of the Log-in page of the application. The user must have an account in order to log into the application. After logging in, the GPS activates, sending data about the users' location to the server.

The above screenshot shows after the user has selected his source point and destination address. The maps API gives us details about estimated travel time and current traffic scenarios.

The user can further enter his/her time of travel in order to obtain results of the predictive analysis of traffic on that route, at that time of the day.

The above screenshot shows after the user has selected his source point and destination address. The maps API gives us details about estimated travel time and current traffic scenarios.

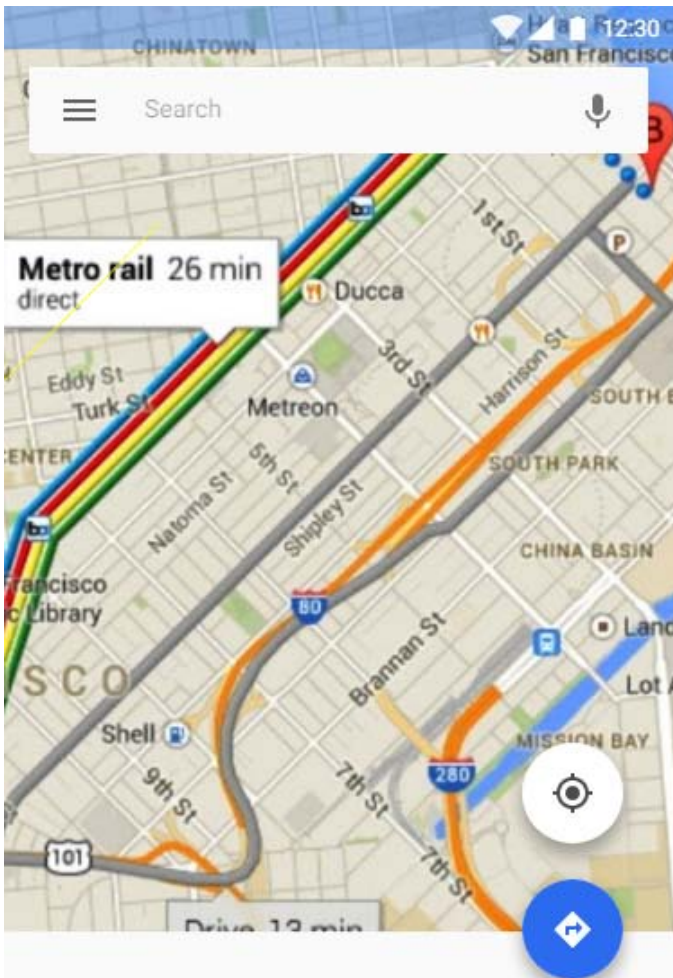
The user can further enter his/her time of travel in order to obtain results of the predictive analysis of traffic on that route, at that time of the day.



The above screenshot shows after the user has selected his source point and destination address. The maps API gives us details about estimated travel time and current traffic scenarios.

The user can further enter his/her time of travel in order to obtain results of the predictive analysis of traffic on that route, at that time of the day.





The above screenshot shows the predicted traffic analysis on all nearby routes at the time the user wants to travel. This predictive analysis is made using the decision tree algorithm stated in the methodology.

## VI. CONCLUSION

Traffic congestion has become one of the major problems we face in our day to day lives. Extreme traffic conditions not only increase commuting time, but also increase the possibility of casualties due to traffic. In many countries across the world, bad traffic conditions have led to delayed response of Emergency services like Ambulance, Fire Truck et cetera

leading to more number of casualties. Such situations can be minimized by optimizing the traffic management system. In this paper, we have proposed Traffic optimization using predictive analysis.

The problem of traffic congestion can be eased by a large extent by implementing predictive analysis methods in real-time. The decision tree methodology proposed in this paper, if implemented on a large scale, can result in optimized traffic with less congestion. The peak hour timing and rerouting modules implemented in this paper can help the users make informed decisions on what route to take, in order to reach their destination facing least possible traffic. This can reduce the commuters travel time by a large extent leading to efficient travel. The traffic timer optimization module, if implemented, can use predictive analysis to efficiently time the signals, which will reduce congestion during peak hours. The traffic timer is set by predicting the vehicle density in real-time for the next couple of hours. This helps in clearing the inbound traffic as soon as it arrives on the traffic junction.

All the modules in this paper can be implemented using predictive analysis using decision trees and hold the capacity to solve the problem of traffic congestion due to increasing vehicles and ill-managed traffic systems.

## VII. REFERENCES

- [1] Data Mining: Practical Machine Learning Tools and Techniques (The Morgan Kaufmann Series in Data Management Systems)
- [2] Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die – Thomas H. Davenport
- [3] Applied predictive modelling- Max Kuhn and Kjell Johnson – [www.appliedpredictivemodelling.com](http://www.appliedpredictivemodelling.com)
- [4] Introduction to Statistical Learning – Trevor Hastie and Robert Tibshirani , December 2013
- [5] Introduction of Decision Trees – JR Quinlan - Springer ,1986
- [6] Decision Tree Methods: Application for classification and prediction - Yan Yan Song and Ying Lu – NCBI article
- [7] Decision Tree for predictive modelling – Padraic G Neville – SAS Institute 1999
- [8] Decision Trees – [www.simafore.com](http://www.simafore.com)