The Project Overview (Dial-A-ride-Problem)

Increasing population in developing country like India is generating the need for more number of public transportation services to provide a convenient commutation to citizens. This area of research is well known as Dial-A-Ride-Problem (further referred as DARP). Increase in use of internet and smart-phones in recent years made a significant increase in service providers in this field. To address the problem of traffic congestion and pollution, they are giving more focus towards providing a shared transportation service. In European and American countries Dial-A-Ride system is well known as a door-t-door commutation service for elderly and disabled persons, which is getting popular among developing countries too. In response to the expanding trend of this system, considerable research has been devoted towards helping the transit agencies planning the capacity for the desired operating conditions.From the inception of research in this field, a variety of analytical, simulation, and statistical models, were developed in order to create a decision support system in the business domain

Statistical and Machine learning models create a functional relation between the capacity planning of a DARP (e.g.fleet size, vehicle operating hours and vehicle traveled kilometers) with other service criteria like the area of service region, demand, the speed of vehicle etc. With the advent of new highly optimized and scalable machine(further referred as ML) learning the software in the twentieth century, it is easier to apply ML algorithms to make accurate predictions for future.

In statistical and ML models it is highly recommended to find the relation and effect of all the predictor/explanatory variables with the response variables to build a robust model.The factors influencing the DAR system is considered from the recent work of (Markovic´ et al. (2013)). They considered 11 predictors as shown in Table-1. They simulated the DARP to get a sufficiently large dataset (nearly 1500 instances) which will be helpful to build statistical and ML model and evaluate them, given randomly generated 11 explanatory variables.The authors build the statistical model on approximately 850 instances in which fleet size ranges from 20 to 90 vehicles.They applied multiple linear regression and artificial neural network algorithm.In the regression model, they calculated the corresponding R^2 and p-values of for the predictors against each of the three response variables(i.e. Fleet size, Vehicle hours and Vehicle km.).

This paper extends the work by applying more advanced machine earning techniques to the data to compare the results between different algorithms.In this project, I make the following contribution:

1. I propose the linear regression model first with one of the independent variable (service area) and application of leave one out cross validation technique to calculate the residual error with an increase of the degree of the variable.
2. The application of the generalized linear model and boosting tree models(with cross-fold validation) and a comparison of residuals for both the models.

2.Data and its analysis

We analyzed the dataset to get the relational effect of different independent variables with three important decision parameters(i.e. response variables).The variables outlined in Table-1 to give a brief overview of the dataset.

Table 1 (Variables considered in the statistical model for planning a dial-a-ride(DAR) system)

Explanatory Variables		Response Variables
(1)Service area(km2)	7.Passenger boarding time(min)	(1) Fleet Size(Vehicles)
2.Demand density(trips/km2-day)	8.Average vehicle speed	(2)Vehicle -hour
3.Peak Demand CV	9.Network circuity factor	(3)Vehicle-km
4.peak-hour Demand(trips/hr)	10 Vehicle Capacity	
5.Pickup time window	11 Maximum ride time ratio	
6.Maximum ride time duration (min)		

- Peak demand CV is the coefficient of variation for the demand in one hour during peak demand period
- Network circuity factor is the average ratio of the shortest route distance between two points and the length of a straight line connecting the same points
- Maximum ride time ratio is the multiple of the ride time and constraints the ride time for a request

The code and analysis you will get from this repository.