

## **Application of Recursive Partitioning Methodology in Route Choice**

Deepa M. J.

Anu P. Alex

Manju V. S.

M.Tech student    Assistant Professor    Associate Professor  
College of Engineering Trivandrum

IJERT

## Abstract

*The selection of routes or paths between origins and destinations is known as route choice. The choices usually depend on certain criteria such as minimum travel time, minimum distance and cost. In many transport applications route choice plays an important role. An attempt was made in this study to use a choice decision tool called Recursive Partitioning Methodology (RPM) to develop the route choice model. The study was mainly concentrated on the route choice of the work trips of public transit users. A questionnaire survey was conducted and a single tree Recursive Partitioning model was developed with the help of a software package called D TREG<sup>®</sup>. The variable importance scores for the factors affecting route choice were obtained. This was used to identify the most influential factors that determine the route choice of the public transit users.*

## 1. Introduction

Transportation planning can be defined as “application of planning techniques in the operation, provision and management of facilities and services for any modes of transport to achieve safe, faster, comfortable, convenience, economical and environmentally suitable movement of people and goods.” To make efficient planning decisions, the planner and the engineers need to assess the transport demand of the network. There are mainly four processes involved in transportation demand modelling: trip generation, trip distribution, mode choice and route choice. Among them route choice has immense effect on the traffic volume of a given network compared to others. The route choice can be defined as the selection of routes or paths between origins and destinations based on minimum criterion rules. The route choice model predicts the probability that any given path between origin and destination is selected to perform a trip.

The earliest and most widely used route-selection criterion is the minimization of travel time which was proposed by Wardrop. Wardrop’s first principle states that each used route has the same cost and it is minimal. This leads directly to the consequence that no user can reduce his travel time by switching to other route and hence known as User Optimum or User Equilibrium (UE) principle. Wardrop’s second principle is based on the minimization of the average travel time, and therefore referred to as System Optimization (SO) principle. Many models have been developed based on the above principles. Those models are mainly used for long range planning. They are not used when real time information was provided.

A considerable number of research studies have been conducted on route choice behaviour of an individual. It is been found out that, the drivers’ route choice behaviour is not limited to only the minimization of travel time. He would consider numerous criteria before selecting a particular route to perform a trip. These criteria include travel cost, travel time and its reliability, traffic safety, traffic comfort, roadway characteristics, utility, information supply, drivers’ habits, drivers’ experience, cognitive limits, socio-economic and demographic characteristics, and other behavioural consideration.

A travelers’ decision-making rule is an evaluation system that represents the travelers’ perception and assessment of the attributes of alternatives routes. Together with perception error, network uncertainty, and static/dynamic choice, the type of decision rule must be considered in route choice models of stochastic networks. In general, the route choice models are based on the underlying principles of utility theory and a number of researchers argued that these models are mathematically manageable due to their inherent constraints which in turn compromise their performance.

## 2. Previous Studies

Many research studies have been conducted for years in modelling route choice. Different researchers used different criteria and methods in developing an efficient route choice model. Abdel-Aty (1997) presented a statistical analysis of commuters’ route choice including the effect of traffic information. Adler (2001) investigated the effects of route guidance and traffic advisories on driver's route choice behavior. The study was a two-factor experiment with repeated measures on one factor where the between-subjects factor was the type of traveler information provided and the repeated, within-subjects factor was trips made between a specified origin and destination.

Chen et al. (2001) developed an individual behavioral-based mechanism for exploring the crucial criteria affecting drivers’ route-selection decisions. Levinson (2003) analyzed systems that provide the driver the fastest path between his or her current location and final destination, updated in real-time to consider recurring and nonrecurring congestion. The traveler’s full cost per trip was a bundle comprised of both expected travel time and its reliability. Arslan and Khisty (2006) developed a heuristic way for handling fuzzy perceptions in explaining route choice behaviour from behavioural point of view. A hybrid model where route choice decision making was described in a hierarchy uses concepts from fuzzy logic and the Analytical Hierarchy Process (AHP) was proposed

for making possible a more proper description of route choice behaviour in transportation systems.

Knoop et al. (2010) investigated about the extent to which travelers change their route when faced with unexpected traffic situation. Shifan et al. (2010) presented a learning-based model of route-choice behavior when information was provided in real time. Grange et al. (2011) presented a route choice model for public transit networks that incorporates variables related to network topology, complementing those found in traditional models based on service levels and users' socioeconomic and demographic characteristics. Higuchi et al. (2011) proposed new type of combined mode and route choice network equilibrium model where the travellers were assumed to choose their mode considering whole of their trip-chain and the common lines problem was considered in the public transportation assignment. Zhou et al. (2011) developed a general travel decision-making rule utilizing Cumulative Prospect Theory (CPT). They investigated the mechanism of travellers' behaviour, examined the probability of applying CPT as a measure of commute utility, and established a general utility measurement system, the results of which were found to be more consistent with the experimental data than those of Expected Utility Theory (EUT)-based route choice models.

Non-parametric classification tree and regression techniques are most widely used in various scientific fields. Many research studies had been conducted in it. Wolf et al. (1997) had conducted a study on modeling hot –stabilized emissions from motor vehicles using Binary Recursive Partitioning Method. The study deals with an alternate modeling approach, Hierarchical Tree Based Regression (HTBR) for hydrocarbon (HC) emissions from motor vehicles. Washington (2000) had developed an iterative modeling method that combines desirable properties of Ordinary Least-Squares (OLS) with Hierarchical Tree-Based Regression (HTBR). This combined approach, named Iteratively Specified Tree-Based Regression (ISTBR), provided insight into data structure provided by hierarchical tree-based regression, while retaining the desirable parametric properties of OLS.

Karlaftis and Golias (2002) studied the relationship between rural road geometric characteristics, accident rate and their prediction using a rigorous non-parametric statistical methodology called Hierarchical Tree-Based Regression. Karlaftis (2004) developed a model for predicting mode choice through Multivariate Recursive Partitioning. They extended prior research by developing a methodology for predicting individual mode choice based on a

nonparametric classification methodology that imposes very few constraining assumptions in yielding mode choice predictions.

Bouack et al. (2009) conducted a study on mapping and clustering technology literatures in solid-state lighting using recursive process. Pittou et al. (2009) developed a nonparametric binary recursive partitioning for deterioration prediction of infrastructure elements. It was used for estimating bridge deck deterioration and treated it as a classification and decision problem.

Antoine (2010) used recursive partitioning method for indexing trajectories in the unrestricted space. A trajectory is a time-varying spatial phenomenon. They developed a method called Recursive Partitioned Trajectory Index (RPTI).

It can be seen from the above studies that RPM is an efficient tool for modelling decision making problems. It has been noticed that not much studies have been conducted in transportation field using RPM. As route choice is a decision making problem, in this study an attempt was made to check the application of RPM in developing the route choice model. Developing a route choice model is a very tedious procedure. A large data set is required for the model development and sometimes the developed model has a very complex structure. Using RPM an individual route choice is obtained, which uses very few constraining assumptions to yield route choice prediction.

### 3. Methodology

The study was conducted in Thiruvananthapuram city, the capital of state Kerala, a medium sized city lying in the southern most part of India. The study was concentrated on work trips of the public transit users.

A stated preference survey was conducted to identify the main factors affecting the route choice of the work trips and the respondents were asked to prioritize the factors. The prioritized factors include both travel and traveller information of the public transit users. The travel information factors are given below:

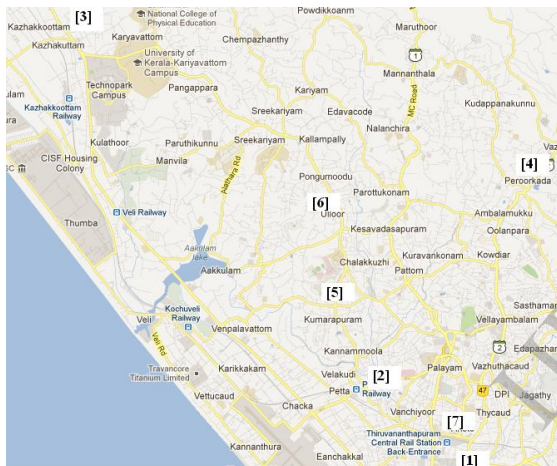
- i. Journey time
- ii. Fare
- iii. Number of transfers
- iv. Traffic congestion
- v. Condition of road
- vi. Frequency
- vii. Crowding in the bus

The traveller information collected included the age and sex of the commuters. For the development of

route choice model, five main Origin Destination (OD) pairs of the study area were selected. The five OD pairs are given in Figure 1. The routes between the OD pairs are: [1] – [6], [7] – [5], [2] – [6], [1] – [4], [1] – [3].

Based on the factors identified earlier, a questionnaire was designed. A survey was carried out among the working commuters in the city. The commuters were interviewed at their work place and at the bus stops. The main advantage of doing this was that the people from various part of the city were covered.

A total of 100 commuters were interviewed and from this 1200 route choices were obtained. During survey, the commuters were also asked to rank the factors based on their priority. This was done to identify the factors that a particular group of commuter would prefer while selecting a route.



**Figure 1. Study area**

Among the commuters interviewed about 54% came under the age group of 18 – 30 years, 24% were under the age group of 30 – 50 and 22% were under the age group of greater than 50. Commuters surveyed included 52% female and 48% male.

The RPM model was developed using predictive modeling software D TREG<sup>®</sup>. Here the accuracy of the model was checked in three ways. The first check was done by the software itself by means of cross-validation. The second check was done by comparing the commuters' preferences which was obtained during the questionnaire survey and the variable importance scores obtained from the model. Then the predictive accuracy of RPM model was compared with another model developed by Discriminate Analysis.

## 4. Data analysis

### 4.1. RPM model

The RPM model developed using DTREG software was a single tree model and classification and regression tree analysis were carried out. For the development of a good decision tree there must be a good splitting algorithm. In this study, Gini splitting algorithm was used to obtain a good decision tree. The whole sample was divided into training set (80%) and testing set (20%). The validation was done by cross validation.

The main problems faced while developing a decision tree is overfitting. The term overfitting refers to the fact that a classifier that adapts too closely to the learning sample will not only discover the systematic components of the structure that is present in the population, but also the random variation from this structure that is present in the learning data due to random sampling. When such an overfitted model is later applied to a new test sample from the same population, its performance will be poor because it does not generalize well (Strobl et al. (2009)).

In recursive partitioning pruning technique is adopted to reduce the effect of overfitting. The principle behind the pruning is to remove the branches that add little to the predictive value of the tree. The pruning relies on complex parameter. Here the tree pruning criteria is minimum cost complexity which is zero standard error.

D TREG<sup>®</sup> accepts a dataset containing number of rows with a column for each variable. One of the variables is the "target variable" whose value is to be modeled and predicted as a function of the "predictor variables". The Predictor variables and Target variables were given as below:

1. Predictor variables
  - Sex – Male, Female
  - Age – 18-30, 30-50, >50
  - Travel time – Low, Moderate, High
  - Condition of road – Good, Moderate
  - Travel cost – Low, Moderate, High
  - Traffic congestion – Moderate, High
  - Crowding – Moderate, High
  - Number of transfers – 0, 1, 2 ....
  - Frequency – Low, Moderate, High
2. Target variable
  - Route – Whether the route is selected or not

D TREG<sup>®</sup> analyzes the data and generates a model in which the target variable can be best predicted based on values of the predictor variables. A single tree model was developed and

the maximum tree depth obtained was 9. The tree underwent 55 group split. The saturation tree or the perfect tree contained a total of 44 terminal nodes. The tree was pruned to avoid the overfitting phenomenon. The minimum validation relative error occurred with 15 nodes. Therefore the tree got pruned from 44 terminal nodes to 15 terminal nodes.

The single tree model developed is shown in the Figure 2. The model is relatively easy to understand. As we can see the Node 1 is a parent node. The first split was based on the travel time, i.e.; whether the travel time is low or high/moderate. Thus two child nodes (Node 2 and Node 3) were formed from the parent node. The same procedure was done for each of the child nodes. Thus the entire tree was split. The relative error of the model obtained was 0.7920 and the standard error was 0.0069.

## 4.2. Variable importance score

During the development of a single tree model, some of the variable may appear explicitly as splitters, which may be interpreted that these variables are more important than others in predicting the dependent variables. Unlike a simple linear regression model, in single tree model, a variable can be considered as highly important even if that variable never appears as node splitter. The variable importance score measures a variable's ability to perform in a specific tree of a specific size either as a primary splitter or as a surrogate splitter. The scores reflect the contribution that each variable makes in classifying or predicting the target variable, with the contribution stemming from both the variable's role as a primary splitter and its role as a surrogate to any of the primary splitters. It is seen from Figure 3 that the frequency of buses has higher importance followed by travel time and sex of the commuter.

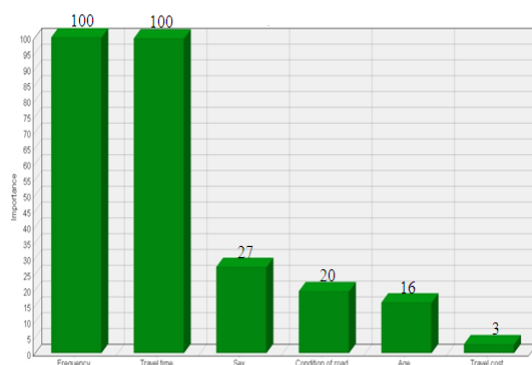


Figure 3. Variable importance score

## 4.3. Model validation

### 4.3.1. By D TREG® software

In order to evaluate the accuracy of the prediction yielded by the proposed methodology, cross-validation was used. In this type of evaluation, the classification algorithm was computed from one part of the data set, called the learning sample, and its predictive accuracy was tested by applying it to predict outcome in the remaining part of the data set called the test sample.



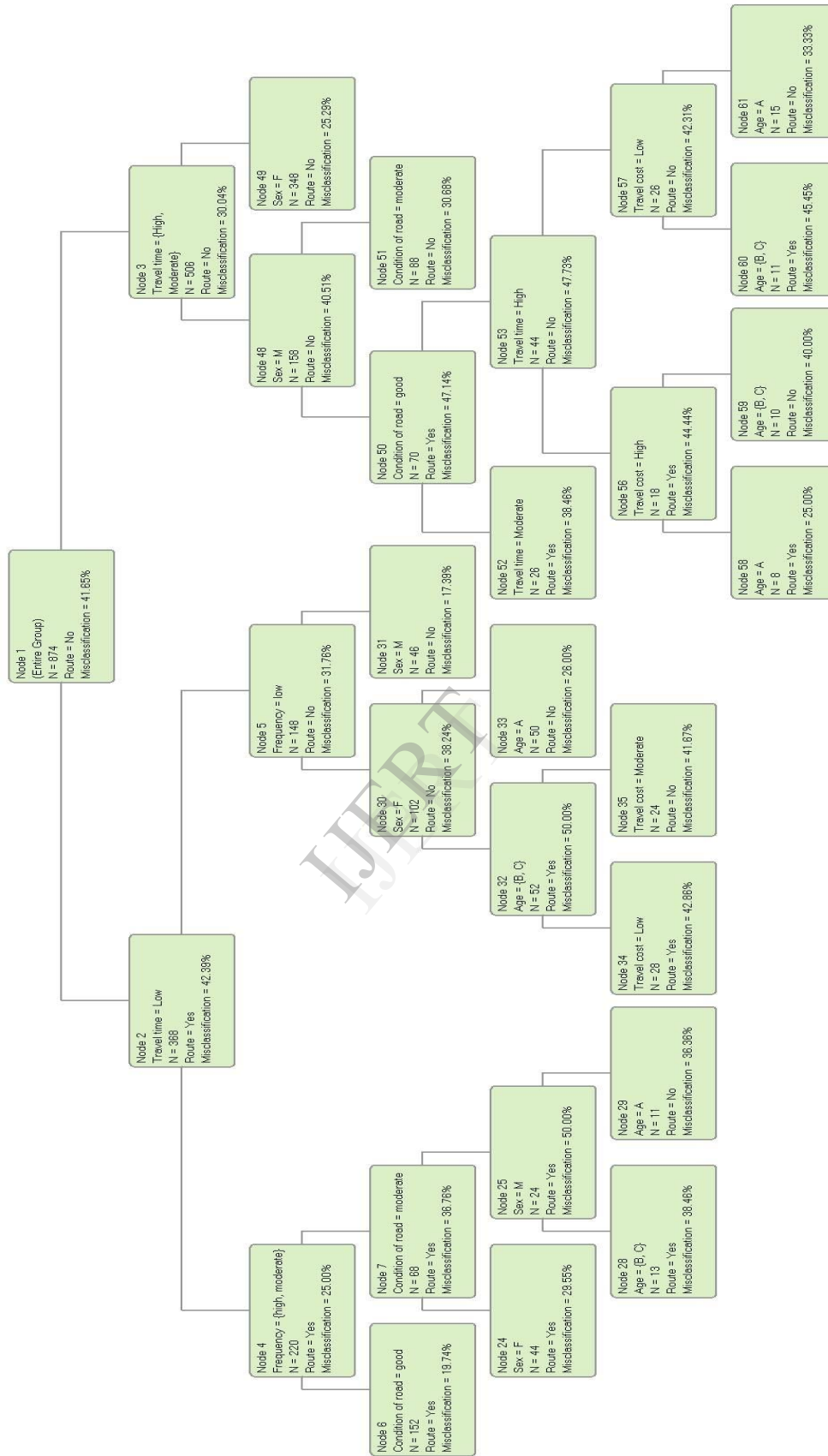


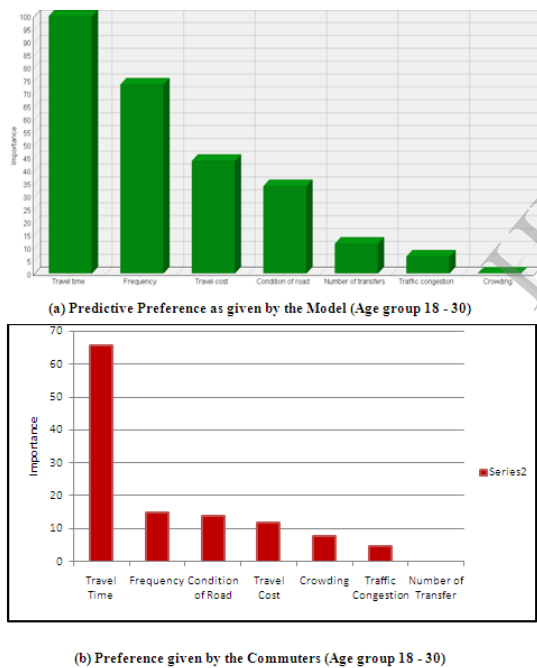
Figure 2. RPM Model developed

The learning and test samples were created by splitting the initial data set through simple random sampling. 80% of the initial data were reserved for learning and 20% for testing. The predictive accuracy of the model obtained was 71.79%, which is satisfactory.

**4.3.2. Survey data**

The ranked factors of the commuters’ were tabulated based on the age and sex. The predicted preference of the factors affecting the route choice was obtained from the model and these two were compared. The results are discussed below.

It can be seen from Figures 4 to 8 that the predicted preference and commuters’ preference for all age group and sex group showed almost similar trend. It was seen that for the commuters’ of age group 18 – 30 and 30 – 50 the most important factor affecting the route choice as obtained from the model is travel time which is similar to the commuters’ preferences. This can be understood from the Figure 4 and 5.

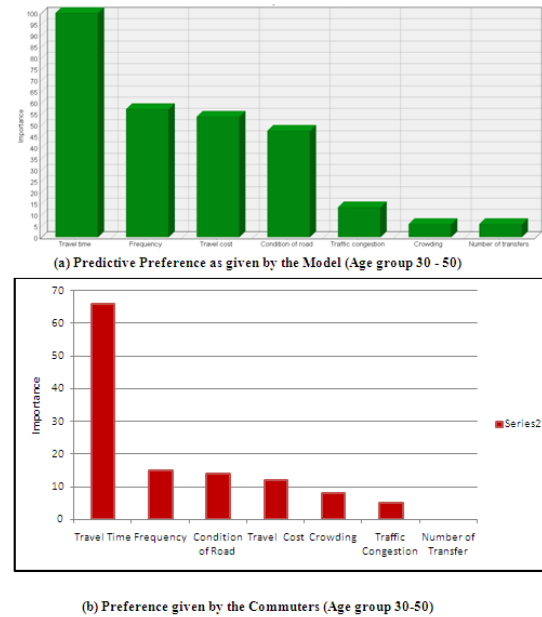


**Figure 4. Comparison of commuters’ preference with preference given by the model for age group of 18-30**

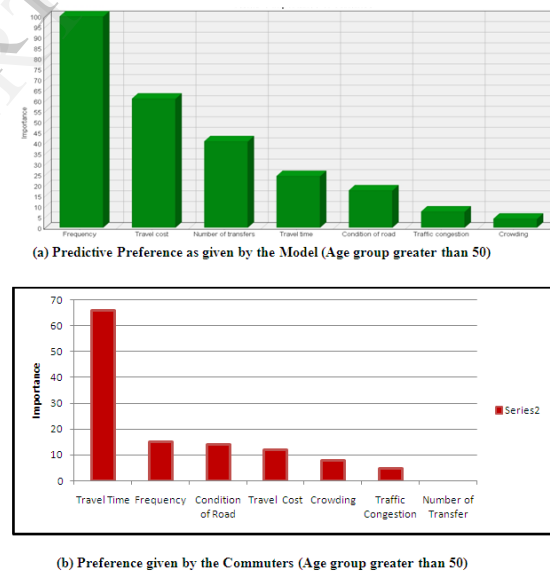
Except for the age group greater than 50, all the other age group shows almost similar trend between the predicted preference and commuters’ preference (Figure 6).

The predicted preference given by the model and sex wise commuters’ preference shows a similar trend (Figure 7 and 8). It can be seen that the predicted factors that affect the route choice

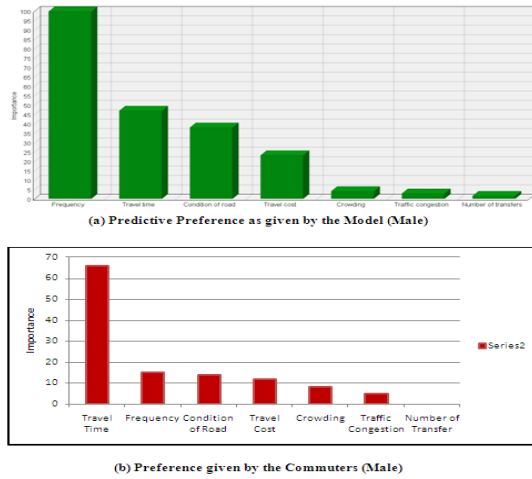
decision obtained from the model are frequency and travel time for both male and female group, which matches with the preference by commuter.



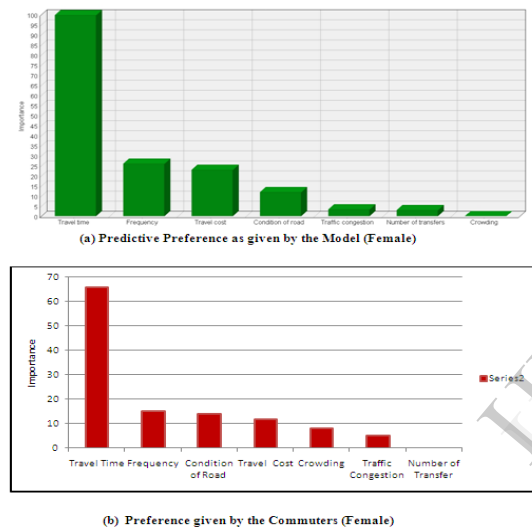
**Figure 5. Comparison of commuters’ preference with model preference for age group of 30-50**



**Figure 6. Comparison of commuters’ preference with model preference for age group > 50**



**Figure 7. Comparison of commuters' preference with model preference for male**



**Figure 8. Comparison of commuters' preference with model preference for female**

#### 4.3.3. Discriminate analysis

Discriminate analysis represents a procedure for determining the group to which an individual belongs, based on the characteristics of that individual. The model was developed using SPSS software. The obtained model is given in equation (1).

$$D = 0.013x S + 0.057 x A + 1.272 x TT - 0.351 CR + 0.223 TCO - 0.125 TC - 0.126 C + 2.381 NT - 1.507 F - 0.894$$

(1)

Where,

D – Discriminate function      NT – Number of transfer

S – Sex                                      CR – Condition of road

C – Crowding

TT – Travel time                      F – Frequency

TC – Traffic congestion              C – Crowding

The predictive accuracy of the model obtained is 67.60%, which is less than the accuracy obtained for the RPM model, which is 71.79%. Therefore it can be concluded that the model developed by Recursive Partitioning Methodology is more reliable.

## 5. Conclusion

Route choice plays an important role in many transport applications. Developing a route choice model is an important step in traffic assignment process. The usual practice is to use the criteria of shortest distance between the origin and destination for the route choice. The route choice behaviour of the driver also depends on many factors like travel time and its reliability, safety, comfort, travel cost, condition of road, information supply etc.. and hence such assumptions will not be realistic.

In this study, the application of a choice decision tool called Recursive Partitioning Methodology was investigated. The predictive accuracy of the model obtained was 71.92%. The model showed that the frequency of the transit is the most influential factor affecting the route choice.

Comparison of the predictive preference of the model with the actual commuters' preferences showed similar trends. The model developed by Discriminate Analysis showed a predictive accuracy of 67.60%, which was less than the predictive accuracy of RPM model. Hence it can be concluded that the RPM model can be successfully used for developing Route choice model.

## 6. Reference

- [1] A.M. Abdel – Aty, R. Kitamura, and P.P Jovanis, "Using Stated Preference Data for Studying the Effect of Advance Traffic Information on Drivers' Route Choice" *Transportation Research Part C*, Vol 5, 1997, pp. 39 – 50.
- [2] J.L. Adler, "Investigating the Learning Effects of Route Guidance and Traffic Advisories on Route Choice Behavior" *Transportation Research Part C*, Vol 5, 2001, pp. 1 – 14.



- [3] T. Arslan, and J.C. Khisty, "A rational approach to handling fuzzy perceptions in route choice", *European Journal of Operational Research*, Vol 168, 2006, pp. 571 -583.
- [4] H.L. Chang, T.Y. Chen, and G.H. Tzeng, "Using a Weight-Assessing Model to Identify Route Choice Criteria and Information Effects" *Transportation Research*, Vol 35, 2001, pp .197 – 224.
- [5] L.Y. Chang, and H.W. Wang, "Analysis of Traffic Injury Severity: An Application of Non-parametric Classification Tree Techniques", *Accident Analysis and Prevention*, Vol 38, 2006, pp 1019-1027.
- [6] H. Contrino, N. McGukin, and D. Banks, "Exploring the Full Continuum of Travel " Data Fusion by Recursive Partitioning Regression" *International Association of Travel Behaviour Research Conference*, 2000.
- [7] L.D. Grange, J.C. Munoz, and S. Raveau, "A topological route choice model for metro" *Transportation Research*, Vol 45, 2011, pp. 138–147.
- [8] G.M. Karlaftis, and I. Golias, "Effect of Road Geometry and Traffic Volume on Rural Roadway Accidents Rates", *Accident Analysis and Prevention* , Vol 34, 2002, pp 357-365
- [9] M.G. Karlaftis, "Predicting Mode Choice Through Multivariate Recursive Partitioning", *Journal of Transportation Engineering* © ASCE, Vol 30, 2004, pp 245-250
- [10] V.L. Knoopa, and H.V. Zuylena, "Rerouting behaviour of travellers under exceptional traffic conditions – an empirical analysis of route choice" *Procedia Engineering* 3, 2010, pp. 113–128.
- [11] D. Levinson, "The value of advanced traveler information systems for route choice", *Transportation Research Part C*, Vol 11 2003, pp. 75 – 87.
- [12] M. Pittou, "Nonparametric Binary Recursive Partitioning for Deterioration Prediction of Infrastructure Elements", *Advances in Civil Engineering*, vol. 2009, 2009. 12 pages
- [13] Y. Shifan, and Ben – Elia, "Which road do I take? A learning-based model of route-choice behavior with real-time information", *Transportation Research Part A*, Vol 44, 2010, pp. 249 - 264.
- [14] C. Strobl, J. Malley, and G. Tutz, "An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests", *American Psychological Association*, Vol 14, Issue 4, 2009, pp 323-348.
- [15] S. Washington, "Iteratively specified tree-based regression: theory and trip generation example" *Journal of Transportation Engineering*, Vol 126, 2000, pp 482-491.
- [16] J. Wolf, S. Washington, and R. Guensler, "Binary Recursive Partitioning Method for Modeling Hot-Stabilized Emission from Motor Vehicles", *Research Record*, Vol 1587, 1997, pp 96-105.
- [17] J. Zhou, H. Xu, and W. Xu, "A decision-making rule for modeling travelers' route choice behavior based on cumulative prospect theory" *Transportation Research*, Vol 19, 2011, pp. 218–228.
- [18] P.H. Sherrod, "*D TREG*<sup>®</sup> Predictive modeling software", 2003, <http://www.nlreg.com>.