Infrastructure Cost Reduction Of Municipal Public Transport Using Machine Learning

Radhika Bembalkar, Pravin Game

Abstract: Municipal public transport is the backbone of every city. It provides an economical transport alternative. Effective management of resources for optimizing the revenue is an often discussed issue. This work deals with finding opportunities to increase the revenue of PMPML (Pune Mahanagar Parivahan Mahamandal Ltd.) by reallocating the buses based on the shuttle-wise revenue of individual bus. Here threshold revenue is considered for deciding whether to run that shuttle or reallocate the bus on another busy or overcrowded route. As of now, two routes i.e. route no. 27 and 2 (Source: Bharti Vidyapeeth and Katraj Depot respectively, Destination: Shivajinagar) are considered and one day$'$s run-wise revenue data for these routes has been made available by PMPML which is being used for the analysis. Machine learning algorithms are used for classifying buses based on utilization.
It is observed that the ID3 (Iterative Dichotomiser) gives better accuracy than logistic regression. The results can be used to reduce the infrastructure cost by reallocating the buses for the routes based on revenue generated.

Index Terms: Public Transport, Operational Cost, Customer Satisfaction, Logistic Regression, Resource Management.

1. INTRODUCTION

Municipal public transport is the mobility backbone of every city. Municipal transport is however, always short of resources - men, machine and operating environment. Effective management of these resources is beneficial for the authorities as well as the general public. In most of the Indian cities, buses are run under public and private ownership. The frequency of a transit vehicle depends on the demand for a particular route during peak and off-peak timings. Most of the people opt for a private vehicle for transportation purpose. There is a lack of public transport facilities and therefore Indian citizens are largely dependent on private modes of transport. The reasons behind this are nothing but a large amount of irregularity in the timings of the buses, overcrowded buses or unavailability of public transit for remote areas or for some routes. Hence, PMPML needs to optimize the available resources through various techniques to improve resource availability and reduce infrastructure cost. The main focus of this work lies in the identification of the resource utilization for Pune Municipal Transport which can be divided into two categories – over utilized or underutilized. Machine learning techniques such as logistic regression and decision tree can provide directions for this particular use case. PMPML data for route no. 27 and 2 (Source: Bharti Vidyapeeth and Katraj Depot, Destination: Shivajinagar) has been considered in this work. This data is processed and analysed using machine learning techniques which can aid in the optimization of resource usage.

- *Radhika Bembalkar, persuing master's degree in Computer Engineering, Pune Institute of Computer Technology, Pune, India. radhikabembalkar@gmail.com*
- Pravin Game, *Computer Engineering, Pune Institute of Computer Technology, Pune, India. psgame@pict.edu*

Majority of Indian cities use manual systems for planning and scheduling of public transit resources. This manual planning many a time leads to irregular bus timings, overcrowded buses or unavailability of buses for some routes. Therefore, there is a need to effectively manage the resources from both operating agency as well as commuters' perspective.

1. There is a need to identify best opportunities to improve the efficiency of public transport in terms of men, machine and operating environment.

2. To efficiently use the available infrastructure, there is a need to periodically review and maintain the existing facilities.

2 LITERATURE REVIEW

Researchers have studied various methods for frequency optimization purpose in public transit. Optimization models have been proposed in terms of vehicle cost, capacity and frequency optimizationFatima and Kumar [1] have outlined the public transit sector challenges ranging from technical to operational for Indian cities. The binary logit analysis is used to estimate the variation in modal shift behavior. In order to evaluate the proposed model Statistica and Biogeme environment were used. Log-likelihood was used to predict modal shift behavior. To evaluate the proposed system data was collected for the city of Bardoli, Gujarat, India. The results show that the proposed system significantly improved the current scenario of public transport. Rohani et al. [2] reviewed different bus service types, operations and its quality as well as outlined the role of bus service provider and driver. Rojas et al. [3] explained the transit network planning process. It consists of five different stages: Transit Network Design, Frequency Setting, Transit Network Timetabling, Vehicle Scheduling and Crew Scheduling. The decisions taken at each stage can influence the decision that can be taken at subsequent stages of the planning process, depending on the context i.e. strategic, tactical or operational. Schöbel [4] explained

various transit network design models like cost oriented, passenger-oriented, game-theoretic and location based models as well as outlined various mathematical approaches and algorithms used for line planning purpose.

Luhua and JiangXinkai [5] have addressed the problem of bus service frequency optimization model in terms of minimization of expenses paid by passengers and bus operators as the target of optimization model. Authors have proposed an improved genetic algorithm for optimization purpose and same is illustrated with an example of two buses i.e. Bus No.6 and. Bus No.9 in Changchun City, China. The performance of the model is compared on the basis of departure time interval, total expenses paid by passengers transfer, total expenses of public transit companies with existing model. The outcome shows that the optimization model performs better than existing system. Martĭnez et al. [6] have proposed a frequency optimization model in order to determine the time interval between successive buses on the line (route). The proposed model uses a non-linear bi-level formula based mixed integer linear programming (MILP) method and a metaheuristic approach called Tabu Search for solving smaller and larger instances respectively. The evaluation of the proposed model is performed using two test cases, those are a city of Rivera, Uruguay and Montevideo, capital of Uruguay. An assumption that the bus has sufficient capacity to carry the desired number of passengers may not be always true in a real life scenario. Wagale et al. [7] proposed a Demand and Travel time Responsive (DTR) model to optimize bus schedules in terms of cost and frequency. The proposed model is evaluated using the case study of Jaipur city, India. Sensitivity analysis is used for performance evaluation purpose. The model uses segment-stop based microdata. The system fails in constructing a timetable for stops which require longer riding times. A stochastic optimization model was proposed by Naumann et al. [8] for robust vehicle scheduling purpose. In order to illustrate the proposed model instances of German cities with network layer and service trips are used. The main limitation of the model is it cannot solve the larger city instances within a reasonable amount of time. An evolutionary algorithm is used for timetable optimization model in order to reduce the waiting time of passengers. Tong et al. [9] proposed an optimization model based on network flow to optimize vehicle capacity and bus routing to satisfy user demands. In order to develop a solution algorithm, Lagrangian decomposition method is used and to reduce the solution search space a space-time prism based method is used. In order to evaluate the model, the case studies of Sioux Falls network and Beijing transportation network have been conducted for illustrative and real-world large-scale transit network. The assumption, regarding passenger demand, sinks to provide good results. Sensitivity is used as a measure to evaluate the system performance. Arizti et al. [10] addressed the problem of determining the time interval between successive buses for a public transport line and proposed a bi-level formulation which is then transformed into mixed integer linear programming (MILP) for solving small-sized city instances. Huang et al. [11] addressed the problem of determining optimal bus frequencies by applying bi-level model. Genetic algorithm is used to determine bus frequencies under uncertain demand and evaluated for determining optimal bus frequencies for the city of Liupanshui, China. Sensitivity analysis is used to evaluate the system performance. As a future scope analytical method can be used to obtain optimal frequencies under congested transit network.

3 PROPOSED METHODOLOGY

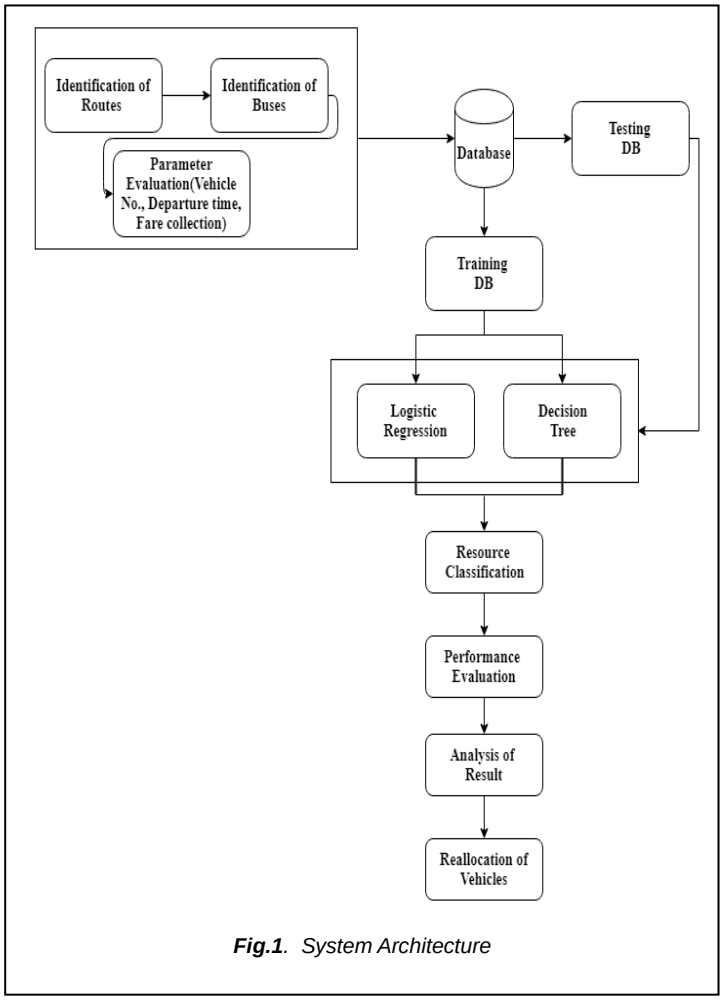

Fig.1. *System Architecture*

In order to promote a feasible, well planned, safe and well maintained transport infrastructure appropriate transport provision in terms of quantity, quality and resource efficiency is important. There is a need to create a system which reduces operational cost, maximizes revenue and provides customer satisfaction. Most of the Indian cities use manual system for resource distribution which may lead to inappropriate usage of the resource. The utilization is classified into four different categories based on revenue collected per shuttle. In order to perform the classification task supervised machine learning logistic regression and decision tree techniques are used. Accuracy has been used as performance measure for the proposed system. Once classification is done, the shifting of buses to other routes, if anyIn order to minimize the cost

overhead of the operating agency and also to provide sufficient amount of vehicles to the commuters for traveling purpose. The proposed system is therefore essential as it tries to optimize the resource utilization. This section describes implementation details of modules in the system.

1. Data Collection

   - The data collection phase requires identification of routes along with buses available for those specific routes.
   - The parameters like vehicle number, departure time and revenue collection of a bus are considered for evaluation purpose.
   - The collected data is stored and pre-processed

2. Authentication for the System

   - The system user needs to be registered to access the system.
   - In order to check system performance the model is tested against test data file

3. Classification Module

   - System admin can apply classification algorithm on loaded data file and train the module.
   - In order to check system performance the model is tested against test data file.

4. Performance Evaluation and Analysis

   The final and major step is performance evaluation.

   - To evaluate the performance of the system various metrics like accuracy, precision and recall are used.
   - Based on the obtained results the reallocation of vehicles takes place.

Algorithms Used:

Decision tree is a flow chart tree like structure where each internal node resembles a test on an attribute, branch represents the test outcome and leaf node depicts the decision or class label.

1. Iterative Dichotomiser (ID3):

ID3 is one of the decision tree method used for prediction purpose. This algorithms works in top-down greedy search manner. It considers all features as a set of attributes and iteratively calculates entropy and information gain. Attribute selection is based on maximum information gain. This helps in partitioning the data and generates subset of the data, the process will continue until the value of information gain becomes zero.

2. Logistic Regression:

It is one of the basic binary classification algorithm used for categorical classification purpose. The logistic model uses a sigmoid function which helps in classifying data into different classes.

4 RESULTS AND DISCUSSIONS

To measure the effectiveness of the system, accuracy is considered as the measuring parameter. Accuracy can be stated with the below mentioned equation 1.

$$Accuracy = (TP + TN)/N \quad \text{--------(1)}$$

Where

TP- True Positive, indicates the properly predicted resource utilization .

TN- True Negative, that indicates the properly predicted negative outcome of the resources utilization.

N- Number of instances.

$$F1\ Score = 2*((precision*recall)/(precision+recall)) \quad \text{--------(2)}$$

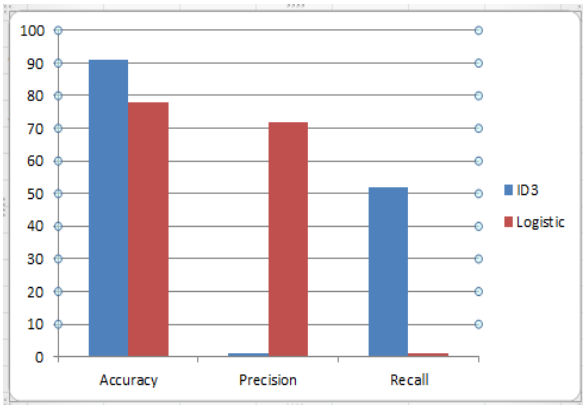

Fig. 2: *Accuracy Measurement Results for Route 27*

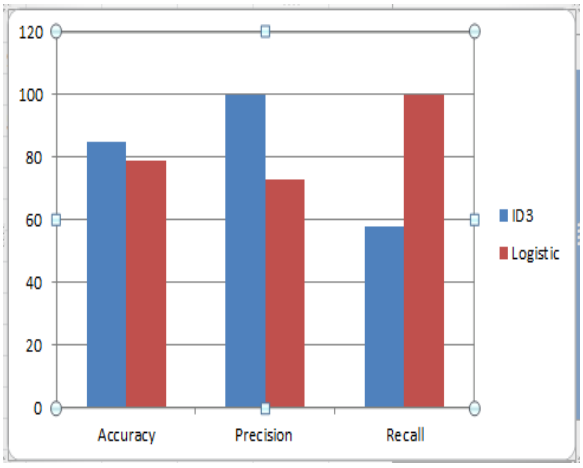

Fig. 3: *Accuracy Measurement Results for Route 2*

The above figures show the performance of ID3 and logistic regression model. The X-axis represents the performance measures used and Y-axis represents the percentage. The accuracy achieved by ID3 and logistic model is 91% ,79% and 85%, 79% for route 27 and route 2 respectively and F1 score for ID3 and logistic regression are 0.74, 0.89 and 0.87, 0.77 for route 27 and route 2.

5 CONCLUSION

The proposed work is beneficial as it provides an approach to the public transport administrator to analyze the resource utilization (machine utilization). The decision tree provides better accuracy as compared with logistic regression. The results of the work show that there exists opportunities to reallocate underutilized buses to the overutilized routes. This will help to increase the revenue as operating cost will be leveraged. This will also help to increase the customer satisfaction by having more buses on overutilized routes.

6 FUTURE WORK

As of now the proposed work considers two routes i.e. route no. 27 and 2 (Source: Bharti Vidyapeeth and Katraj Depot, Destination: Shivajinagar). In future it can be enhanced to incorporate other routes as well. Care needs to be taken that while reallocating the buses, the bus should be made available to the original route for the optimally used shuttle.

REFERENCES

[1] E. Fatima, R. Kumar, ``Introduction of public bus transit in Indian cities,'' International Journal of Sustainable Built Environment, vol. 3, Issue 1, pp. 27-34, 2014.

[2] M. Rohani, D. C. Wijeyesekera, A. T. Abd. Karim,``Bus operation, quality service and the role of bus provider and driver,'' Procedia Engineering, vol.53, pp.167-178, 2013.

[3] O. J. Ibarra-Rojas, F. Delgado, R. Giesen, J. C. Muñoz, ``Planning, operation, and control of bus transport systems: A literature review,'' Transportation Research Part B, vol. 77, pp. 38–75, 2015.

[4] A. Schöbel, ``Line planning in public transportation: Models and methods,'' OR Spectrum, vol. 34, Issue 3, pp. 41-51, 2012.

[5] ShenLuhua, H. JiangXinkai, ``Study on method of bus service frequency optimal model based on genetic algorithm,'' Procedia Environmental Sciences, vol. 10, Part A, pp. 869-874, 2011.

[6] H. Martĭnez, A. Mauttone, M. E. Urquhart, ``Frequency optimization in public transportation systems: Formulation and metaheuristic approach,'' European Journal of Operational Research, vol. 236, Issue 1, pp. 27-36, 2014.

[7] M. Wagale, A. P. Singh, A. K. Sarkar, S. Arkatkar, ``Real-time optimal bus scheduling for a city using a DTR model,'' Procedia-Social and Behavioral Sciences, vol. 104, pp. 845-854, 2013.

[8] M. Naumann, L. Suhl, S. Kramkowski,``A stochastic programming approach for robust vehicle scheduling in public bus transport,'' Procedia Social and Behavioral Sciences, vol. 20, pp. 826-835, 2011.

[9] L. Tong, L. Zhou, J. Liu, X. Zhou, ``Customized bus service design for jointly optimizing passenger-to vehicle assignment and vehicle routing,'' Transportation Research Part C, vol. 85, pp. 451-475, 2017.

[10] A. Arizti, A. Mauttone, M. E. Urquhart, ``A bilevel approach to frequency optimization in public transportation systems,'' 18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization and Systems, vol. 65, pp. 7:1–7:13, 2018.

[11] Z. Huang, G. Ren, H. Liu, ``Optimizing bus frequencies under uncertain demand: Case study of the transit network in a developing city,” Mathematical Problems in Engineering, vol. 2013, Article ID 375084, pp. 1-10, 2013.