Chapter 13
A Comparative Analysis of Urban Transport Using K-Means Clustering and Multi-Class Classification

Aswani Kumar Cherukuri
https://orcid.org/0000-0001-8455-9108
Vellore Institute of Technology, Vellore, India

Karan Bhowmick
Vellore Institute of Technology, Vellore, India

Firuz Kamalov
Candian University, Dubai, UAE

Chee Ling Thong
https://orcid.org/0000-0002-5138-883X
UCSI University, Malaysia

ABSTRACT

The transportation planning process requires a comprehensive study of the regions that need development. This study is an extension of the methodology of transportation planning. The authors use real-time data from Foursquare API to map out the number of transportation facilities and infrastructure available for each city. This study will shed light on areas that need the most development in terms of intra-neighbourhood and inter-neighbourhood transportation. We use k-means clustering to organize and visualize clusters based on a calculated metric called "Availability Factor" that they have defined, and the number of transportation facilities available in each neighbourhood. Finally, they use the data at hand to create a model for multiclass classification to segregate new data into the predefined classes produced by the unsupervised learning model. The information procured in this work can be used to assess the quality of transportation available in the neighbourhoods of a location and help identify key areas for development.

DOI: 10.4018/978-1-6684-5250-9.ch013

INTRODUCTION

Mumbai is the commercial hub of India, the recent population explosion has been a boon and a bane for the residents of Mumbai. With the increasing traffic volumes as a consequence of the increasing population, the transportation infrastructure in Mumbai needs to be augmented. The problems in transportation infrastructure range from a lack of availability, increased hours of commuting, and rush hours. The congestion of traffic in Mumbai has been steadily increasing according to recent trends. Furthermore, the peak-hour crowding has also shown increasing trends. This causes major discomfort in planning trips using public transportation, and especially in increasing the difficulties faced by daily commuters, since this adds an element of unpredictability to their routines. And, longer commuting times, not to mention the environmental efficacy of the modes of transportation being used are also significant hurdles that need to be circumvented.

To tackle these issues, we plan to introduce two comprehensive models that categorize the neighbourhoods of Mumbai on the basis of the transportation infrastructure available. This will be achieved through data mining and machine learning techniques that are elaborated upon below. We use data mining and clustering techniques for this scope because of its relatively unexplored potential in targeted development for transportation infrastructure. We create two models developed by K-Means Clustering algorithm. We propose this model for analysing trends in the transportation amenities of Mumbai on a geographic scale. Data Visualization is used to present two different clustering algorithms based on the total number of transportation infrastructure and the availability factor, which is a metric introduced in the scope of this paper. The first clustering algorithm can be used as a metric to assess the inter-neighbourhood transportation. The second clustering algorithm is used as a metric to assess the fit for intra-neighbourhood transportation. We use folium maps to project the geographical trends using these metrics.

Our contributions aim to fill salient gaps in the literature survey. We use the guidelines explored in S. Na et al. (2010) and M. A. Syakur et al. (2017), and we combine elements of the methodology proposed in Jieh-HaurChen et al. (2020), Omar Elmansouri et al. (2020) and M. A. Syakur et al. (2017) to our functionality. With this exploration of various techniques in K-Means Clustering and Urban transportation analysis, we craft a new methodology which explores the classification of transportation infrastructure across two unexplored directions, namely intra-neighbourhood and inter-neighbourhood parameters. We explore the insights produced by the study in the following sections, starting with the background on which we base this study which includes the literature survey, problem statement and the objectives. Next, we explicate the proposed model through the dataset description, system architecture and methodology. We, then, explore the experimental analysis of the study. Finally, we sum up our findings in the Discussion and Conclusion sections.

BACKGROUND

In this section, we present the findings of the literature survey, research gaps and problem statements that helped inspire this study. We take a look at the Literature surveyed (2.1), the work done in the field and how it can be augmented to produce efficacious results. Next, we elaborate on the goals we have set out for this study and the problems we address (2.2).