Data analytics of bike-sharing system

1st Bhavesh Kumar V D
Vellore Institute of Technology
Bhaveshkumar.vd2019@vitstudent.ac.in

2nd Kshitij Upadhyay
Vellore Institute of Technology
Kshitij.upadhyay2019@vitstudent.ac.in

***Abstract*—**

Applications of data analytics have a vast area and are growing day by day, and still, in a country like India, it is not much popular, especially among small businesses.
By making a system that does data analytics on a bike rental system, we try to explore and use the power of Data Analytics in these businesses, which will result in an exponential growth of these businesses and, in turn, increase their standard of living and provide the users with an excellent experience.

I. Introduction

India is one of the most polluted countries in the world thanks to unplanned industrial development, unplanned residential development, and last but not least the motor vehicles, tons of motor vehicles to be precise, for a country like us it is essential to use all then ways to reduce the pollution using Bikes instead of motor vehicles is one the primary solutions, for bikes bike-sharing systems are very uses

Bike-sharing systems are the new generation of traditional bike rentals where the whole process from membership, rental, and return has become automatic. Through these systems, the user can easily rent a bike from a particular position and return to another position. Currently, there are about over 500 bike-sharing programs around the world which are composed of over 500 thousand bicycles. Today, there exists great interest in these systems due to their essential role in traffic, environmental, and health issues.

Apart from interesting real-world applications of bike-sharing systems, the characteristics of data generated by these systems make them attractive for the research. Opposed to other transport services such as bus or subway, the duration of travel, departure, and arrival position is explicitly recorded in these systems. This feature turns the bike-sharing system into a virtual sensor network that can be used for sensing mobility in the city. Hence, it is expected that the most critical events in the city could be detected via monitoring these data.

II. Problem Statment

Nowadays, many companies are doing data analytics to increase their business by understanding their customers and the market well. Still, in India, most small businesses don't use analytics. By doing data analytics on these kinds of rental stores or small businesses, their business and hence their standard of living can be significantly increased.

In this project, we have combined historical usage patterns with weather data to perform exploratory data analysis and multivariate analysis (hypothesis testing) on bike rental demand and attempt to forecast bike rental demand based on the same variables accurately.

III Literature Survey

In Poland, the most important work on bicycle sharing systems is part of the Warsaw Traffic Survey 2015. As part of the Warsaw traffic survey, a Warsaw agglomeration traffic model was developed. The transport model was divided into three independent models. The third model is for cycling. The bicycle model was created according to the principles of the classic four-stage traffic model. The first step determined the generation and inclusion of the origin and destination of bicycle traffic for the identified travel motifs. Then, in the second step, we split the flow into individual origins and endings for the traffic flow. The final step was to distribute the traffic flow to individual traffic routes. The model also represented the Vertuilo Bicycle Sharing Station, which was able to analyze the mobility behavior of both private bicycle users and users of the Vertuilo bicycle sharing system. Studies show that bicycle traffic is evenly distributed throughout Warsaw's transportation network. Cyclists choose their main route both in the morning and afternoon during rush hours. Comparing the morning and afternoon rush hours, we see an increase in traffic in the selected section of the route and a decrease in traffic in the selected section of the route. In return, the recognized enablers and barriers of Warsaw's bicycle-sharing system have been presented by cyclists and non-cyclists. The results of a poll on the first few years of the existence of Warsaw's bicycle sharing system are B. I flowed into the work of Klepa Kiego and P. Sakowskiego

Due to the fact that bicycle sharing systems became widespread worldwide decades ago, the scientific literature on bicycle sharing systems is relatively new. The scientific papers published so far are mainly related to the structure, function, analysis and modeling of bicycle sharing systems. The works available in literature can be divided into several different disciplines. This includes empirical studies on the function and use of bicycle sharing systems (eg M.Z. Ostwick et al., P. De Maio, A. Faghih Imanietal. -or location and capacity issues with the bike sharing station-e. B. I. Frede and A. Ribero, JR. Rin and T.H. Yang. In addition, you can find studies dedicated to modeling and optimizing relocation problems-eg B. T. Raviv et al., M. Benchimoletal., D. Chemla et al., T. Raviv and O. Kolka -Functions of the bicycle sharing system and its efficiency under various conditions-Example: B.C. Ficker and N. Gast , P. Vogel, D. C. Mattfeld-or using bike sharing-eg B.C. Morency et al.. An important working group is the task of predicting the use of the bicycle rental system. Along. Li and Y. Zheng, Y. Lietal., mainly using time series models (eg A. Kaltenbrunner et al.) and the factors that determine the use of the bicycle sharing system-e. B. A. Faghih Imani and N. Eluru, TD Tran et al., or the positive impact of a bicycle sharing system on people's health and livelihoods, and the natural environment-eg B. Albert metal., R. Jurdak 、Y。Guoua。

Studies have also shown in related literature that the number of rental properties increases when stations are located in densely populated areas or where workplaces are concentrated (such as R. Rixey). Similar patterns can occur at train stations near schools, universities, concentration of services, restaurants, bars, cinemas, shops and other recreational facilities (such as A. FaghihImani et al.). In return, F. Gonzales et al., R. Nair et al., A. Faghih Imanietal. analyzed the influence of the occurrence of bikesharing stations at public transport systems such as bus stations, train stations,Sustainability 2020, 12, 3285 6 of 29 metro stations, and bus stops. The results of these analyses show that the location of the bikesharing stations at public transport systems also increases the bikesharing stations usage. Analyzing articles on bikesharing systems, they indicate that the analyses contained in these works are most often based on data obtained from surveying bikesharing system users or managers—e.g., E. Fishman et al., Y. Tang dynasty. , S. Kaplan etc. Or a question to the operator of the bicycle sharing system-z. B. T. Raviv et al., P. S. wierk, A. Kurek or online-Nutzungsquellena Bahnhöfen-z. B. P. Jimenez et al., R. Hampshire, X. Wang et al..

In cities where bicycle-sharing systems have been in operation for a long time, research is being conducted to obtain information on the quality of system functions from users of bicycle-sharing systems. This data can be used by decision makers to improve the delivery and functionality of the bicycle sharing system to best meet user expectations. Studies have been conducted on the factors that influence the use of the bicycle sharing system and the satisfaction with the use of the bicycle sharing system. B. Y. Guo et al. [69]. The purpose of this study was to understand the factors that influence the low utilization of bicycle rental systems in Ningbo, China. Based on their research, some conclusions related to planning, engineering, and public advocacy were discussed to increase the use of bicycle sharing systems in designated cities. A similar study was conducted by L. Caggiani et al. . They have proposed an optimization model that can be used to determine how to improve the bike sharing system using a specific budget to maximize global user satisfaction. Next, C. Etienne and O. Latifa proposed a model of the Paris bicycle rental system and identified the potential factors that shape the geography of the trip, resulting in station-peripheral and station-peripheral types of mobility patterns. Generated. G. Manzi and G. Saibene analyzed their

satisfaction with using the bicycle-sharing service in Milan.

There are various scientific studies in the literature on attempts to determine satisfaction with the use of bicycle sharing systems in different cities around the world. B.F. Xin et al. Conducted this type of study for Shanghai, D. Efthymi ouetal. In the case of Greece, J. Shietal. For China. Analyzing this task reveals that the research approach differs significantly in terms of the type of data acquisition for analysis, the number of test samples, how the analysis is performed, and the conclusions reached.

IV DATASET

This dataset consists of 2 years of Data

Data Fields

datetime - hourly date + timestamp

season - 1 = spring, 2 = summer, 3 = fall, 4 = winter

holiday - whether the day is considered a holiday

workingday - whether the day is neither a weekend nor holiday weather - 1: Clear, Few clouds, Partly cloudy, Partly cloudy

2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist

3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

temp - temperature in Celsius

atemp - "feels like" temperature in Celsius

humidity - relative humidity

windspeed - wind speed

casual - number of non-registered user rentals initiated

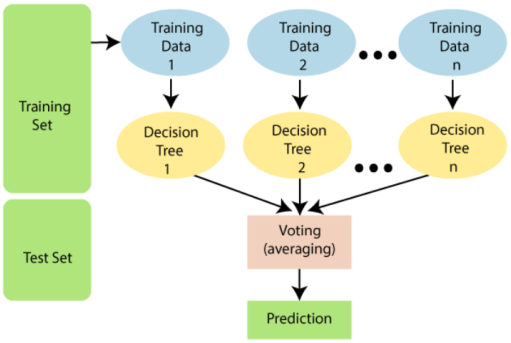
registered - number of registered user rentals initiated

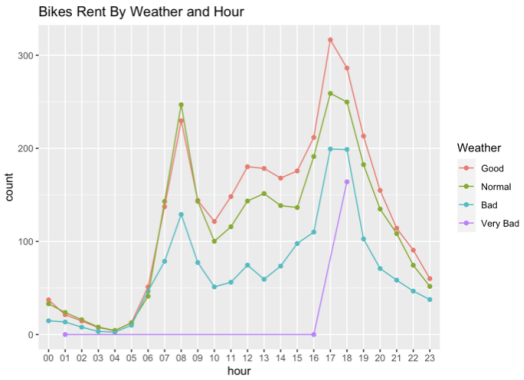
count - number of total rentals

V.METHODOLOGY

The most fundamental and widely used type of predictive analysis is linear regression. The goal of regression is to look at two things: (1) Is it possible to forecast an outcome (dependent) variable using a set of predictor variables? (2) Which variables in particular are significant predictors of the outcome variable, and how do they influence the outcome variable (as indicated by the size and sign of the beta estimates)? These regression estimations are used to illustrate how one dependent variable interacts with one or more independent variables. The simplest version of the regression equation with one dependent and one independent variable is y = c + b*x, where y represents the estimated dependent variable score, c represents the constant, b represents the regression coefficient, and x represents the independent variable score.

Naming the Variables. There are many names for a regression's dependent variable. It may be called an outcome variable, criterion variable, endogenous variable, or regression. The independent variables can be called exogenous variables, predictor variables, or regressors.
Three major uses for regression analysis are (1) determining the strength of predictors, (2) forecasting an effect, and (3) trend forecasting.
Random Forest is a well-known machine learning algorithm that uses the supervised learning method. In machine learning, it can be utilized for both classification and regression issues. It is based on ensemble learning, which is a method of integrating several classifiers to solve a complex problem and increase the model's performance.
"Random Forest is a classifier that contains a number of decision trees on various subsets of a given dataset and takes the average to enhance the predicted accuracy of that dataset," according to the name. Instead of relying on a single decision tree, the random forest collects the forecasts from each tree and predicts the final output based on the majority votes of predictions.
The bigger the number of trees in the forest, the more accurate it is and the problem of overfitting is avoided.



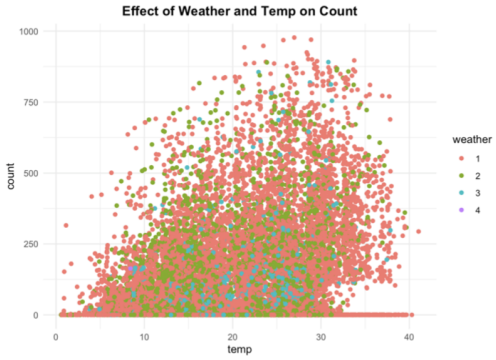


The Working process can be explained in the below steps and diagram:

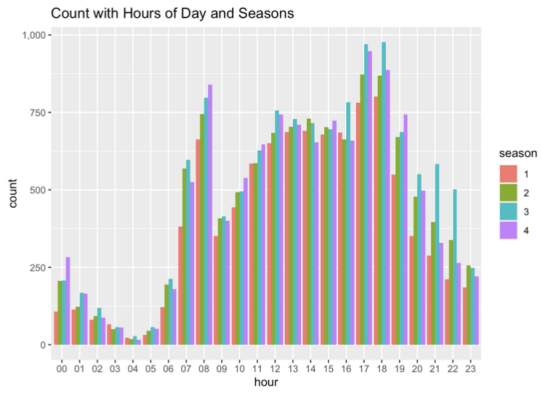
Step 1: Pick K data points at random from the training set.

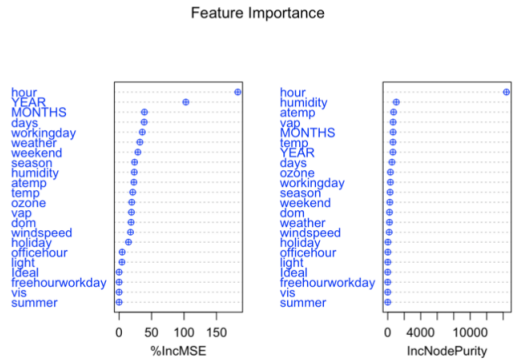
Step 2: Create decision trees for the data points you've chosen (Subsets). Step 3: Decide on the number N for the decision trees you wish to create. Step 4: Repetition of Steps 1 and 2.

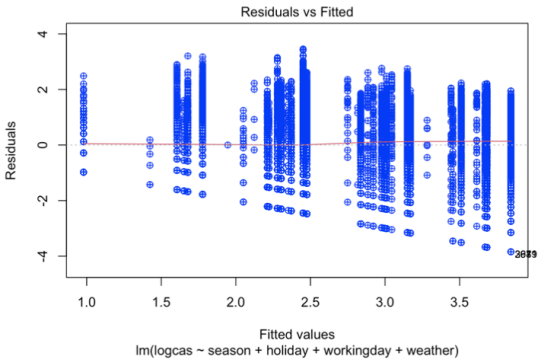
Step 5: Find the forecasts of each decision tree for new data points, and allocate the new data points to the category with the most votes.



VI. Results and Observations







Conclusion:

We performed the thorough analysis of the dataset and have visualized a lots of features and relations between them and we tried to predict the casual and registered users and the important features of our dataset using linear regression and random forest model

Limitation

The dataset is a bit old and not of India, and this is highly location specific . this also needs lots of data to all these visualisations and models accuratel

References

Bicycle sharing systems demand Inês Fradea,*, Anabela Ribeiroa

The Bike-Sharing System as an Element of Enhancing Sustainable Mobility—A Case Study based on a City in Poland

Elz˙bieta Macioszek 1,* , Paulina S´wierk 2 and Agata Kurek 1

AASHTO Executive Committee. (1999). *Guide for the Development of Bicycle Facilities*. Washington, DC: American Association of State Highway and Transportation Officials.

An, M., & Chen, M. (2006). Estimating Non-motorized Travel Demand. *Transportation Research Record: Journal of the Transportation Research Board*, *07-2410*, 18–25.

ConBici, C. en D. de la B. (2007). *Guía metodológica para la implantación de sistemas de bicicletas públicas en España*. Madrid (Spain). Daddio, D. W. (2012). MAXIMIZING BICYCLE SHARING : AN EMPIRICAL ANALYSIS OF CAPITAL.

dell'Olio, L., Ibeas, A., & Moura, J. L. (2011). Implementing bike-sharing systems. *Municipal Engineer*, *000*, 1–13.

DeMaio, P. (2009). Bikesharing Its History, Models of Provision, and Future.pdf. *Velocity 2009 Conference*. Brussels, Belgium.

Dill, J., & Voros, K. (2007). Factors affecting bicycling demand: initial survey findings from the Portland, Oregon, region. *Transportation Research Record: Journal of the Transportation Research Board*, *2031*, 9–17. Retrieved from http://trb.metapress.com/index/rlq1710417402412.pdf

Duthie, J., Brady, J. F., Mills, A. F., & Machemehl, R. B. (2010). Effects of On-Street Bicycle Facility Configuration on Bicyclist and Motorist Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, *2190*(-1), 37–44. doi:10.3141/2190-05 European Commission. (1999). *Cycling: the way ahead for towns and cities*. Luxembourg: Publications Office of the European Union. European Commission. (2009). *A sustainable future for transport - Towards an integrated, technology-led and user-friendly system*. *Publications Office of the European Union*. Luxembourg: Publications Office of the European Union.

Gregerson, J., Hepp-buchanan, M., Rowe, D., Sluis, J. Vander, Wygonik, E., Xenakis, M., & McCormack, E. (2010). *Seattle Bicycle Share - Feasibility Study*.