**Bike Sharing Demand Prediction**

**By:- Nitin Kumar**

**Abstract**

As a convenient, economical, and eco-friendly travel mode, bike-sharing greatly improved urban mobility. However, it is often very difficult to achieve a balanced utilization of shared bikes due to the asymmetric user demand distribution and the insufficient numbers of shared bikes, docks, or parking areas. If we can predict the short-run bike-sharing demand, it will help operating agencies rebalance bike-sharing systems in a timely and efficient way.

**Key words**

Machine learning, Data mining, Bike sharing demand prediction.

**1.Introduction**

According to recent studies, it is expected that more than 60% of the population in the world tends to dwell in cities, which is higher than 50% of the present scenario. Some countries around the world are practising righteous scenarios, renderings mobility at a fair cost and reduced carbon discharge. On the contrary other cities are far behind in the track. Urban mobility usually fills 64% of the entire kilometres travelled in the world. It ought to be modelled and taken over by inter-modality and networked self-driving vehicles which also provides a sustainable means of mobility. Systems called Mobility on Demand has a vital part in raising the vehicles’ supply, increasing its idle time and numbers.

**Problem Description**

Currently Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. Eventually, providing the city with a stable supply of rental bikes becomes a major concern. The crucial part is the prediction of bike count required at each hour for the stable supply of rental bikes.

Today, bike-sharing systems are blooming across more cities around the world. To complete a short trip renting a bike is a faster way when compared to walking. Moreover, it is eco-friendly and comfortable too compared to driving.

**Problem Statement**

* Maximize: The availability of bikes to the customer.
* Minimize: Minimise the time of waiting to get a bike on rent.

**The main goal of the project is to** Finding factors and cause those influence shortages of bike and time delay of availing bike on rent. Using the data provided, this paper aims to analyse the data to determine what variables are correlated with bike demand prediction. Hourly count of bike for rent will also be predicted.

**2. Data Description**

The dataset contains weather information (Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information

**Attribute Information**

* Date - year-month-day
* Rented Bike count - Count of bikes rented at each hour
* Hour - Hour of the day
* Temperature- Temperature in Celsius
* Humidity - %
* Windspeed - m/s
* Visibility - 10m
* Dew point temperature - Celsius
* Solar radiation - MJ/m2
* Rainfall - mm
* Snowfall - cm
* Seasons - Winter, Spring, Summer, Autumn
* Holiday - Holiday/No holiday
* Functional Day – No Func(Non Functional Hours), Fun(Functional hours)

**Feature Breakdown:**

**Date**: The date of the day, during 365 days from 01/12/2017 to 30/11/2018, formatting in DD/MM/YYYY, we need to convert into date-time format.

**Rented Bike Count**: Number of rented bikes per hour which our dependent variable and we need to predict that

**Hour:** The hour of the day, starting from 0-23 it's in a digital time format

**Temperature (°C):**  Temperature of the weather in Celsius and it varies from -17***°****C to 39.4****°****C*.

**Humidity (%)**: Availability of Humidity in the air during the booking and ranges from 0 to 98%.

**Wind speed (m/s):** Speed of the wind while booking and ranges from 0 to 7.4m/s.

**Visibility (10m):** Visibility to the eyes during driving in “m” and ranges from 27m to 2000m.

**Dew point temperature (°C)**: Temperature At the beginning of the dayand it ranges from -30.6**°**C to 27.2**°**C.

**Solar Radiation (MJ/m2):**  Sun contribution or solar radiation during ride booking which varies from 0 to 3.5 MJ/m2.

**Rainfall (mm):** The amount of rainfall during bike booking which ranges from 0 to 35mm.

**Snowfall (cm):** Amount of snowing in cm during the booking in cm and ranges from 0 to 8.8 cm.

**Seasons:** Seasons of the year and total there are 4 distinct seasons I.e. summer, autumn, spring and winter.

**Holiday:** If the day is holiday period or not and there are 2 types of data that is holiday and no holiday

**Functioning Day:** If the day is a Functioning Day or not and it contains object data type yes and no.

**3.EDA on given Data set**

If we want to explain EDA in simple terms, it means trying to understand the given data much better, so that we can make some sense out of it. we using univariate frequency analysis was conducted to describe key characteristics of each feature including, minimum and maximum value, average, standard deviation and others. It was also used to produce a value distribution and identify missing values, and outliers.

EDA is a process of examining the available dataset to discover patterns, spot anomalies, test hypotheses, and check assumptions using statistical measures. In this chapter, we are going to discuss the steps involved in performing top notch exploratory data analysis

**3.1 Data Analysis:**

This is one of the most crucial steps that deals with descriptive statistics and analysis of the data. The main tasks involve summarizing the data, finding the hidden correlation and relationships among the data, developing predictive models, evaluating the models, and calculating the accuracies. Some of the techniques used for data summarization are summary tables, graphs, descriptive statistics, inferential statistics, correlation statistics, searching, grouping, and mathematical models.

**3.2 Data Sourcing**

Data Sourcing is the process of finding and loading the data into our system. Broadly there are two ways in which we can find data.

1. Private Data
2. Public Data

Data collected from several sources must be stored in the correct format and transferred to the right information technology personnel within a company. As mentioned previously, data can be collected from several objects on several events using different types of sensors and storage tools.

**3.3 Data Pre-processing**

A dataset may contain noise, missing values, and inconsistent data; thus, pre-processing of data is essential to improve the quality of data and time required in the data mining.

**3.4 Data Cleaning**

After completing the Data Sourcing, the next step in the process of EDA is Data Cleaning. It is very important to get rid of the irregularities and clean the data after sourcing it into our system.

Irregularities are of different types of data.

Missing Values

1. Incorrect Format
2. Incorrect Headers
3. Anomalies/Outliers

**3.5 Data Deduplication**

It is very likely that your dataset contains duplicate rows. Removing them is essential to enhance the quality of the dataset.

**3.6 Missing Values**

There is a representation of each service and product for each customer. Missing values may occur because not all customers have the same subscription. Some of them may have a number of service and others may have something different. In addition, there are some columns related to system configurations and these columns may have null values but in our orange telecom data set there are no null values present

If there are missing values in the Dataset before doing any statistical analysis, we need to handle those missing values.

There are mainly three types of missing values.

1. MCAR (Missing completely at random): These values do not depend on any other features.
2. MAR (Missing at random): These values may be dependent on some other features.
3. MNAR (Missing not at random): These missing values have some reason for why they are missing.

**3.7 Dropping Missing Values**

One of the ways to handle missing values is to simply remove them from our dataset. We have known that we can use the is null() and not null() functions from the pandas library to determine null values

* 1. **Handling Outliers**

Outliers are data points that diverge from other observations for several reasons. During the EDA phase, one of our common tasks is to detect and filter these outliers. The main reason for this detection and filtering of outliers is that the presence of such outliers can cause serious issues in statistical analysis.

There are two types of outliers:

**3.9 Univariate Outliers**

Univariate outliers are the data points whose values lie beyond the range of expected values based on one variable.

* 1. **Multivariate Outliers:**

While plotting data, some values of one variable may not lie beyond the expected range, but when you plot the data with some other variable, these values may lie far from the expected value.

**3.11 Measures of Central Tendency**

The measure of central tendency tends to describe the average or mean value of datasets that is supposed to provide an optimal summarization of the entire set of measurements. This value is a number that is in some way central to the set. The most common measures for analysing the distribution frequency of data are the mean, median, and mode.

**3.12 Measures of Dispersion**

The second type of descriptive statistics is the measure of dispersion, also known as a measure of variability. If we are analysing the dataset closely, sometimes, the mean/average might not be the best representation of the data because it will vary when there are large variations between the data. In such a case, a measure of dispersion will represent the variability in a dataset much more accurately.

Multiple techniques provide the measures of dispersion in our dataset. Some commonly used methods are standard deviation (or variance), the minimum and maximum values of the variables, range, kurtosis, and skewness.

* 1. **Standardizing Values:**

To perform data analysis on a set of values, we have to make sure the values in the same column should be on the same scale. For example, if the data contains the values of the top speed of different companies’ cars, then the whole column should be either in meters/sec scale or miles/sec scale.

**3.14 Univariate analysis**

Univariate analysis is the simplest form of analysing data. It means that our data has only one type of variable and that we perform analysis over it. The main purpose of univariate analysis is to take data, summarize that data, and find patterns among the values. It doesn't deal with causes or relationships between the values. Several techniques that describe the patterns found in univariate data include central tendency (that is the mean, mode, and median) and dispersion (that is, the range, variance, maximum and minimum quartiles (including the interquartile range), and standard deviation).

**3.15 Bivariate Analysis**

If we analyse data by taking two variables/columns into consideration from a dataset, it is known as Bivariate Analysis.

**a) Numeric-Numeric Analysis:**

Analysing the two numeric variables from a dataset is known as numeric-numeric analysis. We can analyse it in three different ways.

* Scatter Plot
* Pair Plot
* Correlation Matrix

**b) Numeric - Categorical Analysis:**

Analysing the one numeric variable and one categorical variable from a dataset is known as numeric-categorical analysis. We analyse those mainly using mean, median, and box plots.

**3.16 Multivariate Analysis**

Multivariate analysis is the analysis of three or more variables. This allows us to look at correlations (that is, how one variable changes with respect to another) and attempt to make predictions for future behaviour more accurately than with bivariate analysis.

One common way of plotting multivariate data is to make a matrix scatter plot, known as a pair plot. A matrix plot or pair plot shows each pair of variables plotted against each other. The pair plot allows us to see both the distribution of single variables and the relationships between two variables

**3.17 Correlation Among Variables**

In words, the statistical technique that examines the relationship and explains whether, and how strongly, pairs of variables are related to one another is known as correlation. Correlation answers questions such as how one variable changes with respect to another. If it does change, then to what degree or strength? Additionally, if the relation between those variables is strong enough, then we can make predictions for future behaviour

**3.18 Graphical Representation of The Results**

This step involves presenting the dataset to the target audience in the form of graphs, summary tables, maps, and diagrams. This is also an essential step as the result analysed from the dataset should be interpretable by the business stakeholders, which is one of the major goals of EDA. Most of the graphical analysis techniques include Line chart, Bar chart, Scatter plot, Area plot, and stacked plot Pie chart, Table chart, Polar chart, Histogram, Lollipop chart etc.

**4.Algorithms**

**1. Linear Regression**

Linear regression is a supervised machine learning model majorly used in forecasting. Supervised machine learning models are those where we use the training data to build the model and then test the accuracy of the model using the loss function.

Linear regression is one of the most widely known time series forecasting techniques which is used for predictive modelling. As the name suggests, it assumes a linear relationship between a set of independent variables to that of the dependent variable (the variable of interest).

We’re going to fit a line

**y = β0 + β1x**

to our data. Here, x is called the independent variable or predictor variable, and y is called the dependent variable or response variable. Before we talk about how to do the fit, let’s take a closer look at the important quantities from the fit:

• β1 is the slope of the line: this is one of the most important quantities in any linear regression analysis

• β0 is the intercept of the line.



Fig: Linear Regression

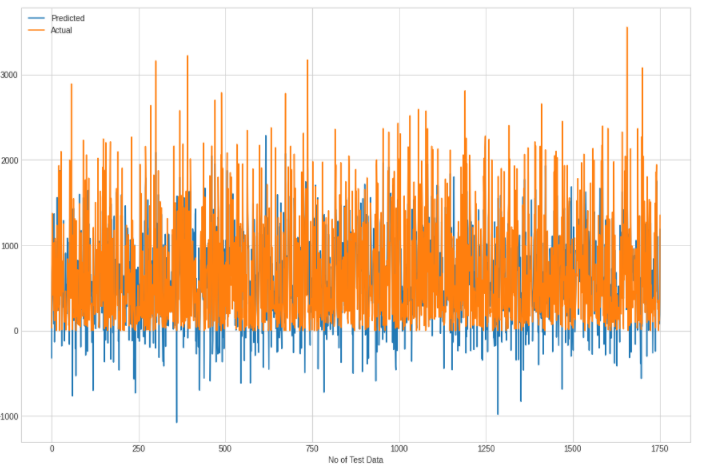


Fig: Linear Regression (Actual results on data)

**2. Ridge Regression:**

Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values to be far away from the actual values.

we have concluded that we would like to decrease the model complexity, that is the number of predictors. We could use the forward or backward selection for this, but that way we would not be able to tell anything about the removed variables' effect on the response. Removing predictors from the model can be seen as settings their coefficients to zero. Instead of forcing them to be exactly zero, let's penalize them if they are too far from zero, thus enforcing them to be small in a continuous way. This way, we decrease model complexity while keeping all variables in the model. This, basically, is what Ridge Regression does.



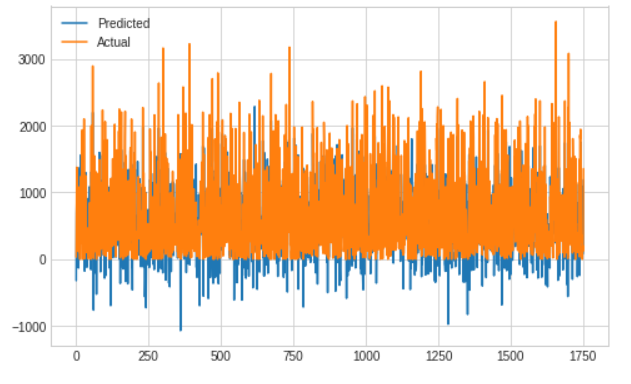
****

Fig: Ridge Regression (Actual results on data)

**3. Lasso Regression**

Lasso, or Least Absolute Shrinkage and Selection Operator, is quite similar conceptually to ridge regression. It also adds a penalty for non-zero coefficients, but unlike ridge regression which penalizes sum of squared coefficients (the so-called L2 penalty), lasso penalizes the sum of their absolute values (L1 penalty). As a result, for high values of λ, many coefficients are exactly zeroed under lasso, which is never the case in ridge regression.

The only difference in ridge and lasso loss functions is in the penalty terms. Under lasso, the loss is defined as:





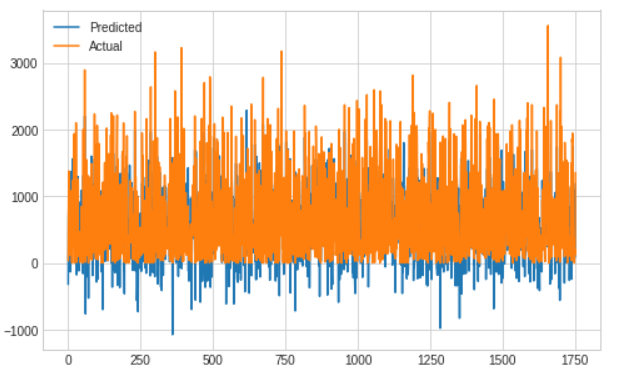


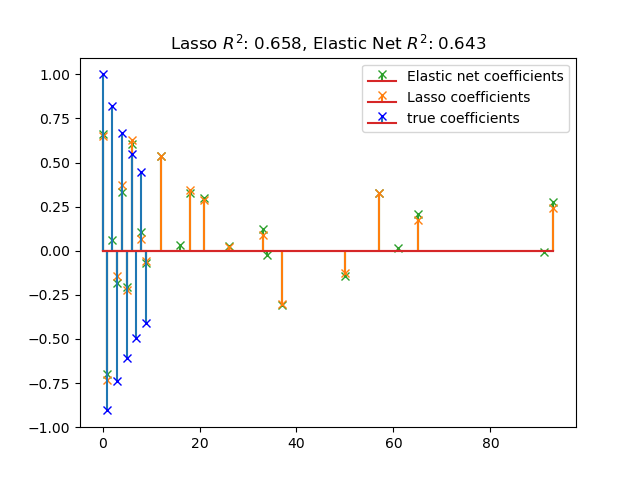
Fig: Lasso Regression (Actual results on data)

**4.Elastic Net**

The elastic net method overcomes the limitations of the [LASSO](https://en.wikipedia.org/wiki/Lasso_(statistics)) (least absolute shrinkage and selection operator) method which uses a penalty function based on

{\displaystyle \|\beta \|\_{1}=\textstyle \sum \_{j=1}^{p}|\beta \_{j}|.}Use of this penalty function has several limitations. For example, in the "large *p*, small *n*" case (high-dimensional data with few examples), the LASSO selects at most n variables before it saturates. Also, if there is a group of highly correlated variables, then the LASSO tends to select one variable from a group and ignore the others.

To overcome these limitations, the elastic net adds a quadratic part to the penalty ({\displaystyle \|\beta \|^{2}}which when used alone is [ridge regression](https://en.wikipedia.org/wiki/Ridge_regression) (known also as [Tikhonov regularization](https://en.wikipedia.org/wiki/Tikhonov_regularization)). The estimates from the elastic net method are defined by

 {\displaystyle {\hat {\beta }}\equiv {\underset {\beta }{\operatorname {argmin} }}(\|y-X\beta \|^{2}+\lambda \_{2}\|\beta \|^{2}+\lambda \_{1}\|\beta \|\_{1}).}

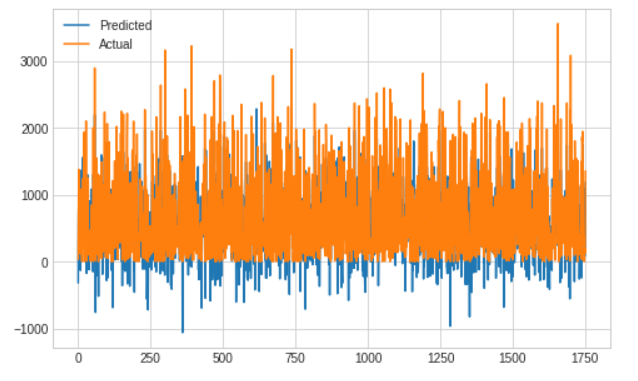
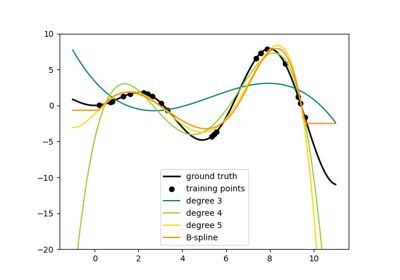


Fig: Elastic Net Regression (Actual results on data)

**5.Polynomial Features**

[Polynomial](https://en.wikipedia.org/wiki/Polynomial) features are those features created by raising existing features to an exponent. For example, if a dataset had one input feature X, then a polynomial feature would be the addition of a new feature (column) where values were calculated by squaring the values in X, e.g. X^2. This process can be repeated for each input variable in the dataset, creating a transformed version of each.

As such, polynomial features are a type of feature engineering, e.g. the creation of new input features based on the existing features.

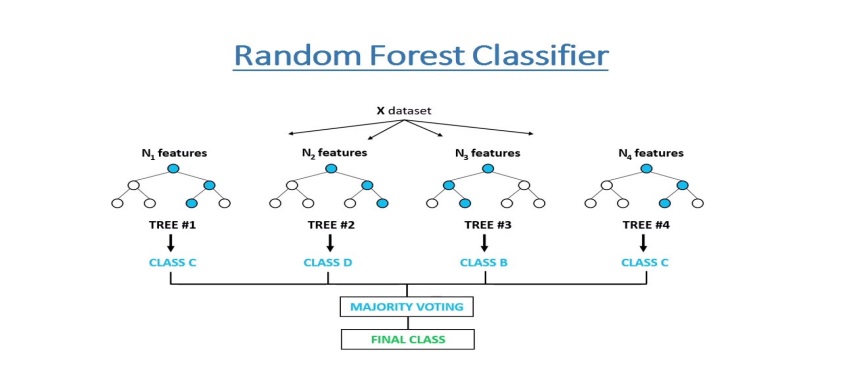


**6.Decision Tree:**

Decision tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. A tree can be “learned” by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called*recursive* partitioning. Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. An instance is classified by starting at the root node of the tree, testing the attribute specified by this node, and then moving down the tree branch corresponding to the value of the attribute as shown in the above figure. This process is then repeated for the subtree rooted at the new node.

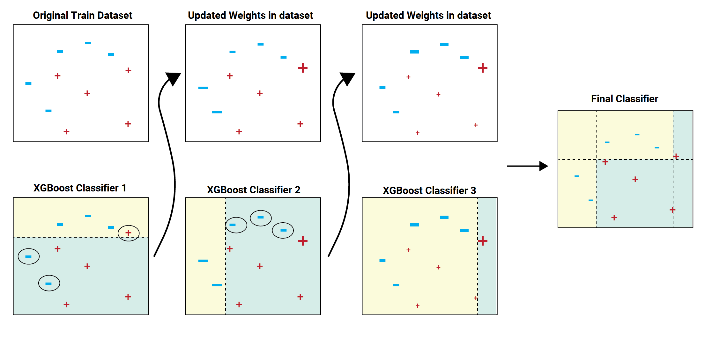


**7. Random Forest:**

Random Forest is a bagging type of Decision Tree Algorithm that creates a number of decision trees from a randomly selected subset of the training set, collects the labels from these subsets and then averages the final prediction depending on the greatest number of times a label has been predicted out of all.

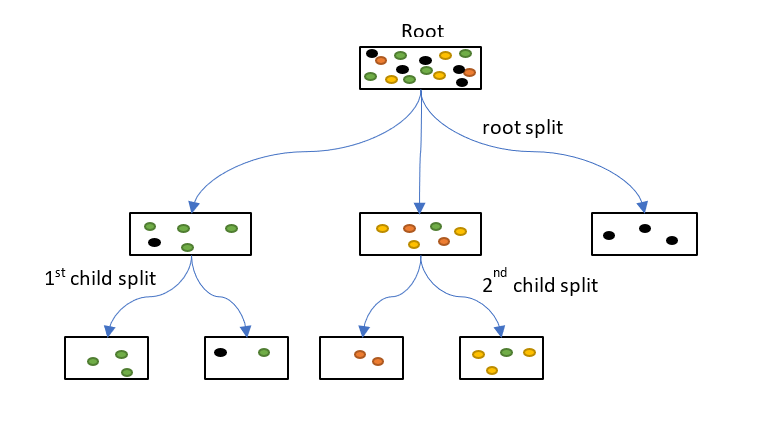
**8. XG Boost**

**XGBoost** is an optimized distributed gradient boosting library designed to be highly **efficient, flexible**and**portable**. It implements machine learning algorithms under the [Gradient Boosting](https://en.wikipedia.org/wiki/Gradient_boosting) framework. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurate way. The same code runs on major distributed environment (Hadoop, SGE, MPI) and can solve problems beyond billions of examples.



**9. Cat Boost**

CatBoost is an algorithm for gradient boosting on decision trees. It is developed by Yandex researchers and engineers, and is used for search, recommendation systems, personal assistant, self-driving cars, weather prediction and many other tasks at Yandex and in other companies, including CERN, Cloudflare, Careem taxi. It is in open-source and can be used by anyone.



**5.Conclusions**

Bicycle sharing systems can be the new boom in India, with use of various prediction models the ease of operations will be increased. The Nine algorithms are applied on the bike share dataset for predicting the count of bicycles that will be rented per hour. We got some good results and accuracy with random forest and XG boost. The accuracy and performance have been compared between the models using Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), R2 and Adjusted R2. If these systems include the use of analytics the probability of building a successful system will increase

**6.References**

[1] “Short-Term Prediction of Bike-Sharing Demand Using Multi-Source Data: A Spatial-Temporal Graph Attentional LSTM Approach” by Xinwei Ma 1 , Yurui Yin 1 , Yuchuan Jin 2 , Mingjia He 3 and Minqing Zhu 4,\*

[2]https://www.computerscijournal.org/vol10no1/prediction-of-bike-sharing-demand/ [3]https://link.springer.com/chapter/10.1007/978-3-030-94751-4\_25