Bike Sharing Demand Prediction

Rushabh Tikale

Data Science Enthusiast

Almabetter, Banglore.

Content

TITLE PAGE

1. Abstract 0-1

2. Problem Statement 1-3

3. Introduction 3-4

4. How Bike Sharing work 6-7

5. Steps Involved 7-9

6. Types of model used 9-13

9. Evaluation Metrics 13-14

10. Conclusion 14-15

1.Abstract:

The bike sharing is the platform where customer can take a bike on rent for a rupees per hour, it is one of the leading business for helping the people who have not bike so they can rented out for some time. The main moto for the project is to deal with the customer demand , a single need should be fulfil by the consumer. Any customer should not wait for the bike so we have to predict the requirement for bikes that customer cant have to wait.

*Keywords: Data cleaning, Data analysis, Train & test model, Model selection.*

2.Problem Statement

**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.** The dataset contains weather information (Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes

rented per hour and date 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 – The dataset in %

### Windspeed – The dataset in m/s

### Visibility – The visibility by 10m

### Dew point temperature – The dataset is in Celsius

### Solar radiation - The radiation in MJ/m2

### Rainfall – The rainfall in mm

### Snowfall - The snowfall for the day in cm

### Seasons - Winter, Spring, Summer, Autumn

### Holiday - Holiday/No holiday

### Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)

3.Introduction:  
The Bike sharing is one of the leading business now a days so fulfiling the requirement is the most challengeable part for successful run up, so the ml model helps on achieving it. We worked for different model

so the ML model helps on achieving it. We worked for different model such as lasso , ridge and other also so we work now on each and every model simaltaneously. So lets start dealing with them.

4. How the Bike Sharing Work.

The typical bike-share has a number of defining traits and options, together with station-based bikes and cost programs, membership and move charges, and per-hour utilization charges. Programs are usually intuitive sufficient for novice customers to grasp. And, regardless of some variation, the variations are normally sufficiently small to forestall confusion when an everyday consumer of 1 metropolis’s bike-share makes use of one other metropolis’s program for the primary time.

The price of an annual bike-share membership may very well be decrease than what you’d spend every year on upkeep and repairs for a personal trip, relying on the standard of your bike, how onerous you trip it, and the way properly you care for it. If you’ll be able to keep away from or reduce utilization charges, you would come out forward with out sacrificing the mobility and freedom that comes with having two wheels at your disposal.

5.Steps involved:  
Fill Missing value  
 After loading the dataset we performed this method by

replacing nan values with zero. This procedure giving us a

approach to tackle with the null value and making a

foundation strong.

* Data cleaning

In this process we convert all string contained data to numeric data. As instead of this we will unable to do the EDA on the data of play store.

* Exploratory data analysis

In this procedure we simultaneously work with each features from our dataset and clearly visualize each and every point of aspects. This gives a graphically representation of entire dataset.

* Extraction of un-useful data

In entire data some features has un-useful for the data analysis. So it’s better to extract this kind of features from our dataset.

* Remove Outlier   
  This procedure helps us to move out with unrelatable part from the   
  dataset. It gives the very clear data and save the time from playing with unrelatable part.
* Correlation of data

This procedure helps us to know the relation between the features and useful for time saving in analysis duplicates type of data.

* Spliting Data to train and test   
  The model have to fit for training and testing . We split the model as

70% for train and 30% test as it is already decided by the model for under better model act.

* Import Baseline model

As we discussed earlier we have to import model such as ridge, lasso, etc. as it gives us a help to decide the better model for furthure prediction.

* Use Evaluation Metrics

The MSE, RMSE, etc are evaluation used to show the residual error between predicted and actual model.

* Choose the best model

The final predictor model is the last point for our model then after we are confirm to use the model.

6. Types of model use:

\* Lasso

\* Ridge

\* ElasticNet

\* Random Forest

\* Decision Tree

1.Lasso: Least Absolute Shrinkage and Selection Operator.

Lasso regression is one of the regularization methods that create parsimonious models in the presence of a large number of features, where large means either of the below two things:

1. Large enough to enhance the tendency of the model to over-fit. Minimum ten variables can cause overfitting.

2. Large enough to cause computational challenges. This situation can arise in case of millions or billions of features.

As It used to save the model from overfit.

2. Ridge:

Ridge regression or Tikhonov regularization is the regularization technique that performs L2 regularization. It modifies the loss function by adding the penalty (shrinkage quantity) equivalent to the square of the magnitude of coefficients.

3. Elasticnet:

The main purpose of ElasticNet Regression is to find the coefficients that minimize the sum of error squares by applying a penalty to these coefficients. ElasticNet combines L1 and L2 (Lasso and Ridge) approaches. As a result, it performs a more efficient smoothing process. In another source, it is defined as follows:

Elastic Net first emerged as a result of critique on Lasso, whose variable selection can be too dependent on data and thus unstable. The solution is to combine the penalties of Ridge regression and Lasso to get the best of both worlds*.*

## Features of ElasticNet Regression

* It combines the L1 and L2 approaches.
* It performs a more efficient regularization process.
* It has two parameters to be set, λ and α.

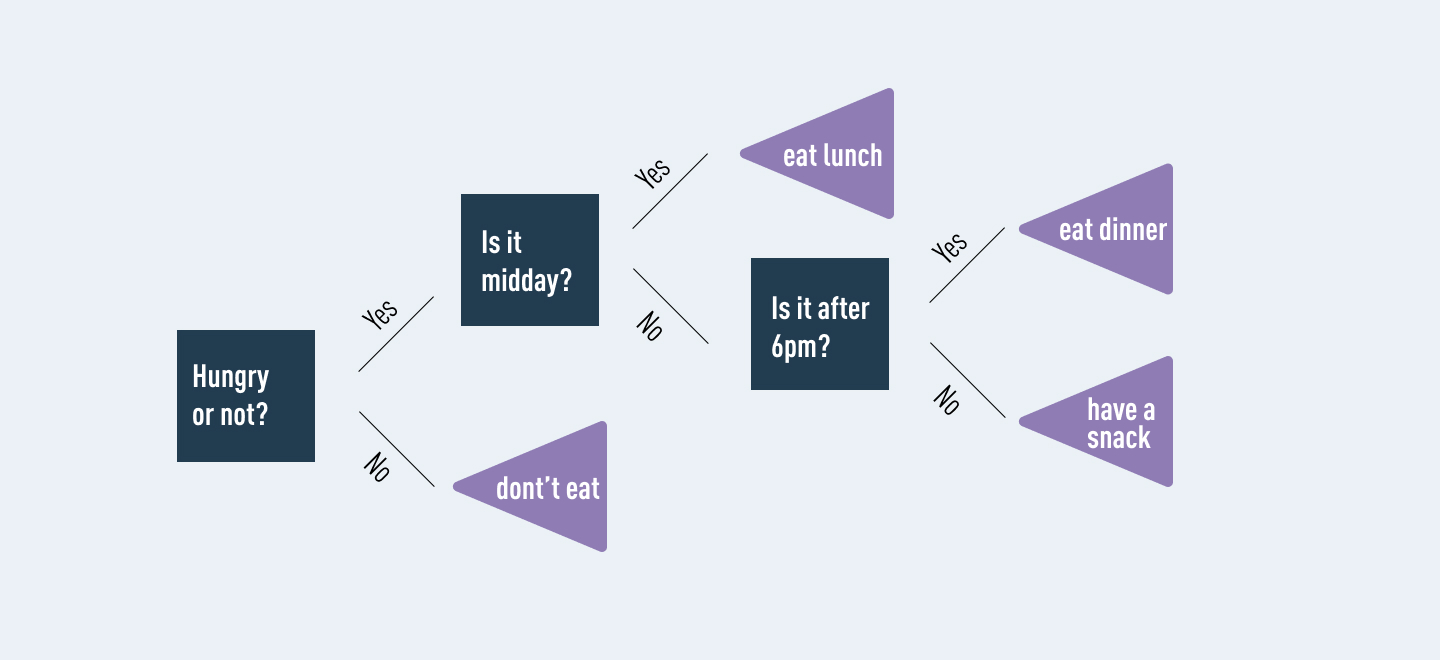
4.Random forest :

The sample taken by randomly for training and testing data. The random forest has work like the forest has so many trees so we have to choose the tree randomly by this we can overid wth the residual error. Random Forest grows multiple decision trees which are merged together for a more accurate prediction.

The logic behind the Random Forest model is that multiple uncorrelated models (the individual decision trees) perform much better as a group than they do alone.

5. Decision Tree:  
 This model classified according the extraction by conditionaly. The data has been sorted according to the condition and from the end calculation the final data is selected as it very clean and clear with no robust entry from the data.

It Works according to this chart.



7.Evaluation Metrics

Here are three common evaluation metrics for regression problems:

**Mean Absolute Error (MAE)** is the mean of the absolute value of the errors:

1n∑i=1n|yi−y^i|

**Mean Squared Error (MSE)** is the mean of the squared errors:

1n∑i=1n(yi−y^i)2

**Root Mean Squared Error (RMSE)** is the square root of the mean of the squared errors:

1n∑i=√1n(yi−y^i)2

Comparing these metrics:

MAE is the easiest to understand, because it's the average error.  
MSE is more popular than MAE, because MSE "punishes" larger errors, which tends to be useful in the real world.  
RMSE is even more popular than MSE, because RMSE is interpretable in the "y" units.  
All of these are loss functions, because we want to minimize them.

8.Conclusion

Here we come with the final conclusion from the business environment that the toughest part is to deal with the large data.The whole model depends on the data cleaning so the maximum part we spent on cleaning the data.We mostly get stuck in the preparation The data consists of large features so we have to deal with each and every subsets.

From the feature engineering the temperature denotes the important predictor variable. We first deal with the data sorting make the data more convenient and easy to use. Then we reomove some outliers which gives us the major look for predicting the future model. Moved with **multicollinearity** which is very essential part for predicting the model.

We train and test the data of regression by taking the sample of data. Then we used different baseline models such as lasso, ridge, ElasticNet which gives us the view for model good or bad.From this algorithm we did not get any good prediction for model by which we cant develop the business model. Then we used Random Forest and Decision Tree model. From this we get a good model by Random Forest.From this we can go with upnext business problem, and we can used as good predictor model.

Let us Keep calm and get relaxed the discussion is over here....

Abstract:

CCo