**Bike Sharing Demand Prediction**

**Submitted by:- Subhasis Chattopadhyay**

**E-mail id:-** [**subhasischattopadhyaydbb@gmail.com**](mailto:subhasischattopadhyaydbb@gmail.com)

**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.

**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 - NoFunc(Non Functional Hours), Fun(Functional hours)

**Data Pre-processing**

### 1. Getting the dataset

### 2. Importing libraries

### 3. Importing datasets

### 4. Finding Missing Data

### 5. Label Encoding of certain features

## 6. Data Cleaning and Feature Engineering

**Exploratory Data Analysis(EDA)**

1. Check how Hour of the day affects rented bike counts
2. Check how Temperature affects rented bike counts
3. Check how Humidity affects rented bike counts
4. Check how Windspeed affects rented bike counts
5. Check how Visibility affects rented bike counts
6. Check how Solar radiation affects rented bike counts
7. Check how rainfall and snowfall affects rented bike counts
8. Check how Seasons affect rented bike counts
9. Check how Month affects rented bike counts
10. Check how Quarter affects rented bike counts
11. Check for Multi Co-linearity

**Machine Learning Algorithm implementation**

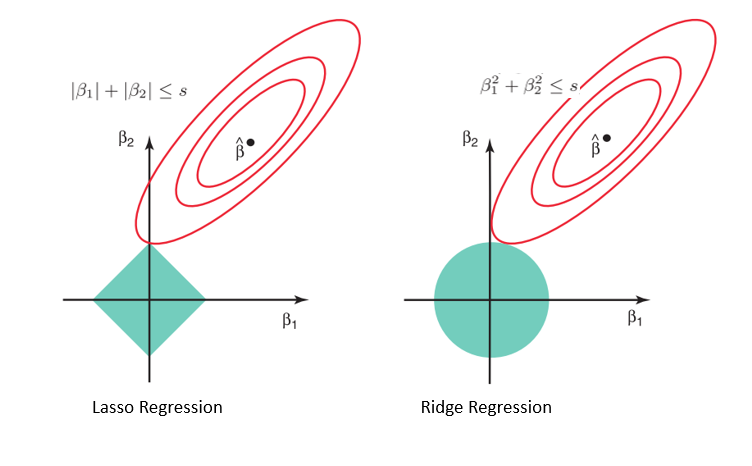
**Linear Regression**

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.



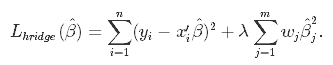
**Lasso Regression**

**Lasso Regression** is a popular type of regularized linear regression that includes an L1 penalty. This has the effect of shrinking the coefficients for those input variables that do not contribute much to the prediction task. This penalty allows some coefficient values to go to the value of zero, allowing input variables to be effectively removed from the model, providing a type of automatic feature selection.



**Ridge Regression**

Similar to the lasso regression, ridge regression puts a similar constraint on the coefficients by introducing a penalty factor. However, while lasso regression takes the magnitude of the coefficients, ridge regression takes the square.



Ridge regression is also referred to as L2 Regularization.

**Random Forest Regression**

Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance



**XGBoost**

**Gradient boosting** refers to a class of ensemble machine learning algorithms that can be used for classification or regression predictive modeling problems.

Ensembles are constructed from decision tree models. Trees are added one at a time to the ensemble and fit to correct the prediction errors made by prior models. This is a type of ensemble machine learning model referred to as boosting.

Models are fit using any arbitrary differentiable loss function and gradient descent optimization algorithm. This gives the technique its name, “gradient boosting,” as the loss gradient is minimized as the model is fit, much like a neural network. 

**Conclusion**

As we can see Hyper Parameter Tuned XGBoost ML model performs best with an accuracy of 90%. Total amount of bike rentals increases with the temperature per month. Whereas it seems that the rentals are not much affected by wind speed and humidity, because they are almost constant over the months. This also confirms on the one hand the high correlation between rentals and temperature and on the other hand that nice weather could be a good predictor .So people mainly rent bikes on nice days and optimum temperature. This could be important of planning new bike rental stations.

**References**

* **GeekforGeeks**
* **Stack\_overflow**
* **Scikit-learn**

**GitHub link-**

[**https://github.com/SubhasisChattopadhyay/Almabetter-Capstone-Projects/blob/main/Bike\_Sharing\_Demand\_Prediction\_Capstone\_Project.ipynb**](https://github.com/SubhasisChattopadhyay/Almabetter-Capstone-Projects/blob/main/Bike_Sharing_Demand_Prediction_Capstone_Project.ipynb)