* **Abstract :**

Building a machine learning model to predict the number of rented bikes per hour is an important task for many businesses. By predicting the number of bikes rented, businesses can better plan for their rentals, and make better decisions about pricing and inventory management.

This machine learning model will help businesses gain insights into their rental trends and anticipate customer demand. It will also help them optimize their pricing and inventory levels to maximize profits and minimize costs.

Additionally, the model can be used to identify areas of potential growth and help businesses make more informed decisions about future expansion.

* **Table of contents :**
* Abstract
* **Table of contents**
* **List of terms**
* **List of terms**
* **Acknowledgments**
* Problem Statement

#### **Introduction**

#### **Background**

#### **Materials and apparatus**

#### **Procedure**

#### Conclusion

* **List of terms :**

Feature engineering :- the pre-processing step of machine learning, which extracts features from raw data.

Univariate analysis :- Univariate analysis explores each variable in a data set, separately

*Bivariate analysis* :- The bivariate analysis is will measure the correlations between the two variables.

*Multivariate analysis* :- is a Statistical procedure for analysis of data involving more than two types of measurement or observation.

* **Acknowledgments :**

Thank you AImabetter for your help and support in completing my ML project. Your guidance and advice was invaluable and I am grateful for all the help you provided.

* **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 theprediction of the bike count required at each hour for the stable supply of rental bikes.[¶](http://localhost:8889/lab/tree/Almabetter%20Capistone%20Projects/2.%20Supervised%20ML%20-%20Regression/1.Bike%20Sharing%20Demand%20Prediction/Bike%20Sharing%20Demand%20Prediction.ipynb#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.)

### 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 - Humidity in %

#### Windspeed - Windspeed in m/s

#### Visibility - Visibility in 10m

#### Dew point temperature – In Celsius

#### Solar radiation – In MJ/m2

#### Rainfall – In mm

#### Snowfall – per cm

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

#### Holiday - Holiday/No holiday

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

#### ****Introduction :****

#### This project focuses on predicting bike rental demand per hour using machine learning

#### techniques. The data used in this project is from the bike rental system. The goal of this project is to develop a predictive model that accurately predicts the number of bikes that will be rented in a given hour.

#### This predictive model will be used by bike rental services to help them better manage their supply and demand for bikes. The predictive model will be developed using a variety of supervised learning algorithms.

* **Background :**

The model will then be used to make predictions about future bike rental numbers, allowing bike rental services to plan and optimize their services accordingly. By using this model, bike rental services can better understand customer behavior, plan for peak times, and optimize their services to increase customer satisfaction.

* **Materials and apparatus:**

Materials:

1. Computer: A desktop or laptop computer with an operating system capable of running machine learning algorithms is essential for any machine learning project.
2. Data: Datasets are an essential component of any machine learning project. The type of data required for a specific project will depend on the type of machine learning task it is being used for.
3. Software: Machine learning algorithms need to be implemented in a programming language. Popular choices are Python used.
4. Libraries: Libraries are collections of pre-written functions that can be used in the implementation of machine learning algorithms. Popular libraries include scikit-learn used.

Apparatus:

1. GPU: Graphics processing units (GPUs) are specialized hardware used to speed up the training process of machine learning algorithms. GPUs can significantly reduce the time required to train a machine-learning model.

* **Procedure :**

***[1] Import, Loading and Inspection of Data*** *:*

After importing the dataset, we look at its columns and shape. The info() method is used to verify variables and associated datatypes for null values. Using the describe() function, we can determine the fundamental characteristics of each variable, such as the mean, median, count, and so on.

We can better comprehend the meaning of the variable thanks to the supplied variable description. This helped us understand datasets.

***[2] Handling duplicated values :***

Fortunately, there aren't any duplicate values in the dataset, but if there are, you can get rid of them with the drop\_duplicated() method.

***[3] Handling null values :***

Fortunately, the dataset has no null values, but if there are any, you may eliminate them using the dropna() function or by replacing them with the mean, median, or mode.

***[4] Handling outliers :***

Boxplot and distplot are used to detect outliers.The interquartile range approach is used to eliminate outliers from data*.*

***[5] Feature engineering and data wrangling :***

Using feature engineering, we generate new variables from the original one.

***[6] Exploratory data analysis :***

*For exploratory data analysis, we use a count plot, barplot, line plot, heatmap, boxplot, and distribution plot.*

Statistical methods were used to conduct univariate, bivariate, and multivariate analyses, which yielded useful insights.

***[7] Assumptions*** *:*

The log and square root transformations were used to check each variable's distribution and return it to normal.

The correlation between each independent variable and the dependent variable was investigated. To verify multicollinearity, a heatmap and variance inflation factors were used. To eliminate multicollinearity, eliminate any variables having a high VIF.

***[8] Feature Encoding :***

Label encoding is used to convert an object-format variable to an integer that a machine learning model can understand.

#### *****[9] Data Scaling :*****

#### To make it simple for a model to learn and comprehend the problem, data scaling is performed. MinMaxScaler keeps the original distribution's shape. The information present in the original data is not meaningfully changed. Therefore, to scale the dataset, we used MinMaxScaler.

#### *[10] Fitting different ML models :*

Using the Sklearn library, several regression models were fitted, and predictions were made using the test dataset. In order to increase the model's precision and predictability, model fitting is performed using GridSearchCV.

***[11] Evaluation of Model :***

Metrics including R-squared, modified R-squared, mean absolute error, mean squared error, and root mean squared error was used to further assess the model. More R-squared scores and a lower RMSE indicate a model that makes more accurate predictions.

***[12] Model Explainability :***

In order to give additional insight into how an ML model arrived at its final output, SHAP was used to explain the model. It describes which variables affect output favorably or unfavorably, which variables are crucial for generating output, and which variables are useless for the model.

* **Conclusion :**

We tested numerous machine-learning models and assessed them using different metrics. The XGBoost regression model comes the closest to having a residual mean of zero and has the highest R-squared score and lowest root mean squared error (RMSE).

The XGBoost regression model is the perfect and well-trained model for predicting the number of rented bikes needed per hour because to the model's high accuracy, low error, and zero mean of residuals.

.