



Final Project Report

Project Title: Predicting Bike Rental Demand - Linear Regression Analysis

Course Title: WM-ASDS04: Introduction to Data Science with Python

Section: A

Batch:09

Submitted to

Farhana Afrin Duty

Submitted by

Name

ID

Jayanta Sarker Shuva

20229022

Sujan Kumar Bhowmik

20229026

Md Sakhawat Hossain

20229041

HASIN AHMED MALLICK

20229029

Title: Predicting Bike Rental Demand - Linear Regression Analysis

Objective

The objective of this analysis is to develop a linear regression model to predict bike rental demand based on various features such as season, weather, temperature, humidity, windspeed, weekday, month, and holiday/working day status. The dataset used for this analysis contains historical data of bike rentals

Data Preprocessing: The dataset was initially explored and cleaned by handling missing values, converting categorical variables into binary representations, and scaling numeric features using min-max scaling.

Exploratory Data Analysis (EDA): Descriptive statistics and visualizations were used to gain insights into the distribution and relationships of the variables. It was observed that temperature and humidity have a significant impact on bike rental demand, with higher demand during warmer months and lower demand during rainy or humid conditions.

Feature Engineering: Categorical variables such as season, weather, weekday, and month were converted into dummy variables to be used as predictors in the linear regression model.

.

Model Building: We have built a Linear Regression model using the stats model API in Python. We have added a constant variable and fitted the model on the training set. We have calculated the p-values and t-values to check the significance of the coefficients. The model summary shows that all the variables are significant with p-values less than 0.05.

Model Evaluation: We have used the model to make predictions on the testing set. We have evaluated the model's performance using the following metrics:

- MAE (Mean Absolute Error)
- MSE (Mean Squared Error)
- RMSE (Root Mean Squared Error)
- R-squared

The scatter plot between y_{test} and y_{pred} indicates that the predicted values and original values have a linear relationship. The evaluation metrics for the Linear Regression model are as follows:

- MAE : 0.042
- MSE : 0.004
- RMSE : 0.062
- R-squared : 0.793

Conclusion

The Linear Regression model built on the Bike Sharing dataset has an R-squared value of 0.793, which indicates that the model explains 79.3% of the variance in the target variable. This model can be used to predict the demand for bikes in the bike-sharing system based on the given set of independent variables.