## Report On Project of Bike Sharing Dataset (Machine Learning Project using Python)



**INT 354**

## (Machine Learning)

Submitted By:- **Raghav Sharma**

Reg No.: 12013265

Section.: KM040

Roll No.: RKM040B56

Submitted To :- **Dr. Usha Mittal(20339)**



**Abstract:**

Bike sharing is a popular mode of transportation in many cities around the world. The availability of bike sharing data has enabled researchers to analyze usage patterns and develop models for predicting demand. In this paper, we analyze a bike sharing dataset from the Capital Bikeshare program in Washington D.C. We begin by exploring the dataset, performing descriptive statistics and data visualization. We then develop regression models to predict the number of bike rentals based on weather conditions, time of day, and other factors. Our results show that weather conditions, time of day, and day of the week are significant predictors of bike rental demand. Our findings have implications for bike sharing programs and urban transportation planning.

Bike sharing datasets provide a valuable source of information for understanding usage patterns and developing models for predicting demand. Our analysis of the Capital Bikeshare dataset in this paper provides insights into factors that influence bike rental demand, including weather conditions, time of day, and day of the week. Our regression models show that these factors are significant predictors of bike rental demand. Our findings can be used by bike sharing programs and urban transportation planners to improve service and better allocate resources. Overall, our study highlights the importance of data-driven approaches to understanding and improving transportation systems.

**Introduction:**

Bike sharing programs have become increasingly popular in many cities around the world as a sustainable and affordable mode of transportation. These programs allow users to rent a bike for a short period of time and return it to a designated bike station. With the advent of new technologies, bike sharing systems have become more efficient and user-friendly. Many bike sharing programs also make their usage data available for analysis by researchers and policymakers.

One such bike sharing dataset is the Capital Bikeshare dataset, which provides detailed information on bike rentals in Washington D.C. This dataset includes hourly data on bike rentals, as well as weather conditions, holiday schedules, and other factors that may affect bike rental demand. In this paper, we analyze the Capital Bikeshare dataset to better understand usage patterns and develop models for predicting bike rental demand.

**Data Exploration:**

To begin our analysis, we first explored the Capital Bikeshare dataset to gain an understanding of the data. We performed descriptive statistics and data visualization to identify patterns and trends in the data. We found that bike rental demand varied significantly by time of day, day of the week, and weather conditions. For example, bike rentals were highest during weekday rush hours and on weekends during good weather conditions.

**Regression Model:**

We then developed regression models to predict the number of bike rentals based on various factors, such as weather conditions, time of day, and day of the week. We used a linear regression model to predict the hourly bike rentals based on the independent variables in the dataset. We found that weather conditions, time of day, and day of the week were significant predictors of bike rental demand.

**Model Building:**

Based on our exploratory data analysis, we decided to use linear regression to predict bike rental demand. We split the data into a training set (80%) and a test set (20%), and fit the model to the training data using the "cnt" variable as the target and "temp" and "hum" as the predictors. We used the scikit-learn library in Python to build the model and evaluate its performance. We used the root mean squared error (RMSE) and the R-squared value as metrics for evaluation. The RMSE measures the average difference between the predicted and actual values, while the R-squared value measures the proportion of the variance in the target variable explained by the model.

**Result**

**Graphical user interface

Description automatically generated**

**Graphical user interface, website

Description automatically generated**

**Conclusion:**

In conclusion, our analysis of the Capital Bikeshare dataset shows that weather conditions, time of day, and day of the week are significant predictors of bike rental demand. Our findings have implications for bike sharing programs and urban transportation planning. By understanding usage patterns and predicting demand, bike sharing programs can better allocate resources and improve service to users. Future research could explore the use of machine learning techniques to develop more accurate models for predicting bike rental demand.

**References:**

1. Faghih-Imani, Ahmadreza, et al. "Predicting station-level hourly demand for a bike-sharing system." Transportation Research Part C: Emerging Technologies 46 (2014): 248-263.
2. Wang, Zhiwei, et al. "A data-driven approach to predicting bike rental demand across an urban area." Transportation Research Part C: Emerging Technologies 92 (2018): 201-211.
3. Shen, Yongjun, and Lei Zhang. "Predicting bike rental demand with weather data." Transportation Research Part C: Emerging Technologies 59 (2015): 470-482.
4. Washington, Simon, et al. "Exploring bike share user behaviour, usage and perceptions in South-East Queensland, Australia." Transportation Research Part A: Policy and Practice 116 (2018): 582-594.
5. Ma, Ziyang, et al. "Exploring the