**Capstone Project Summary**

Bike Sharing Demand Prediction

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)** |
| A growing number of metropolitan areas are now offering bike rentals as a means of enhancing mobility convenience. The public must have access to the rental bike at the appropriate time so that it reduces the amount of time people have to wait. Providing the city with a consistent supply of rental bikes becomes a major concern at some point. The most important part is the expected hourly bicycle count for the constant supply of rental bicycles.  In this dataset, we predict the demand for the Bike Sharing Program in Seoul based on historical usage patterns about temperature, time, and other data.  We began our analysis by performing EDA on all of our datasets. First, we looked at and changed our dependent variable, "Rental Bike Count." After that, we looked at categorical variables and eliminated those that represented the majority of one class. We also looked at numerical variables and discovered their correlation, distribution, and connection to the dependent variable. Additionally, we hot-encoded the categorical variables and removed some numerical features that primarily had 0 values.  Linear Regression, Lasso, Ridge, Elastic net, Decision Tree, Random Forest, and XGBoost were the next eight machine learning algorithms we used. To enhance the performance of our model, we performed hyperparameter tuning.  Here are some solutions to manage Bike Sharing Demand:   * The majority of rentals are for daily commutes to workplaces and colleges. Therefore, open additional stations near these landmarks to reach their primary customers. * While planning for extra bikes to stations the peak rental hours must be considered, i.e., 7–9 am and 5–6 pm. * Start a new renting program for premium customers to increase business. * Maintenance activities for bikes should be done at night due to the low usage of bikes during the night-time * We see 2 rental patterns across the day in bike rentals count - first for a Working Day where the rental count is high at peak office hours (8 am and 5 pm) and the second for a non\_working day where the rental count is more or less uniform across the day with a peak at around noon. * Hour of the day: Bike rental count is mostly correlated with the time of the day. As indicated above, the count reaches a high point during peak hours on a working day and is mostly uniform during the day on a non-working day. * Temperature: People generally prefer to bike at moderate to high temperatures. We see the highest rental counts between 32 to 36 degrees Celcius * Due to less no. of data in the dataset, the training R2 score is around 99% and the test R2 score is 92%. Once we get more data, we can retrain our algorithm for better performance.   **Team Member’s Name, Email, and Contribution:** |
| **Contributor Role:**   1. **Navneet Keshri –** 2. Problem Statement 3. Motivation 4. Know Your Data  * Import Libraries * Dataset Loading * Dataset View and Information  1. Understanding Variables  * Variable Description * Checking Unique Values  1. Data Wrangling  * Checking Missing and Duplicate values * Changing Complex Column Names * Creating some New Columns * Changing Data Type of Categorical Columns  1. Data Visualization  * Univariate Analysis on Dependent Variable * Bivariate and Multivariate Analysis  1. Outlier Analysis  * Z score >4 Pruning * Correlation Analysis * Heatmap Plot  1. Feature Engineering and Data Pre-Processing  * Categorical Encoding * Normalize Dependent Variable * Data Splitting * Setting Evaluation Metrics  1. ML Model Implementation  * Linear Regression * Lasso Regression * Ridge Regression * Elastic Net Regression * Decision Tree * Random Forest * Gradient Boosting * Hyperparameter Tuning  1. Conclusion |
| **Please paste the GitHub Repo link.**  GitHub Link: - <https://github.com/Navneet2409/bike-sharing-demand-prediction> |
| **Please paste the drive link to your deliverables folder. Ensure that this folder consists of the project Collab notebook, project presentation, and videos.**  Drive Link: - <https://drive.google.com/drive/folders/1uFV7IEZGLyZP8hoRhaCnCe4xiTsRzH3L?usp=sharing> |