**Capstone Project-2 Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. Name: - Vikash Kumar Diwakar 2. Email ID:- diwakarvikash32@gmail.com  * Contributed In notebook helped with Google diver data connectivity and data cleaning, data manipulation, and in EDA Visualization * Contributed for the contents of PPT. * Contributed in Technical Documentation in content of problem statement goal of project and steps involved. * Contributed in notebook for data cleaning, data manipulation, and in EDA Visualization , feature engineering, training, testing, deployment and finalizing the conclusion. * Contributed in PPT by making sure all the points to be covered. * Contributed in Technical Documentation in content of problem statement goal of project and steps involved. * Did feature engineering thoroughly. Extracted Important features, removed multicollinearity. * Trained and tested with different linear regression models and selected the best. |
| **Please paste the GitHub Repo link.** |
| Github Link: - <https://github.com/Vikash2009/Bike-Sharing-Demand-Prediction> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| Bike sharing demand prediction is an exciting machine learning project that aims to forecast the number of bikes that will be rented from a bike-sharing system at a given time. With the rise of urbanization and the increasing popularity of sustainable transportation options, bike-sharing programs have become prevalent in many cities around the world. Accurately predicting demand is crucial for these systems to optimize bike availability and enhance user experience.  Machine learning algorithms can analyze historical data on bike rentals, incorporating factors such as time of day, weather conditions, day of the week, and even special events to make accurate predictions. By understanding patterns and correlations in the data, these models can estimate the number of bikes needed at specific locations, allowing bike-sharing companies to efficiently allocate their resources.  To build a bike sharing demand prediction model, one would typically start by collecting and preprocessing the relevant data, which includes bike rental records and additional features like weather data. Next, a suitable machine learning algorithm, such as linear regression, decision trees, or neural networks, can be trained on the data to learn the underlying patterns and make predictions.  The performance of the model can be evaluated using metrics such as mean absolute error or root mean square error, comparing the predicted rental counts with the actual values. Fine-tuning the model and incorporating additional features can help improve its accuracy over time.  Overall, bike sharing demand prediction using machine learning offers valuable insights for bike-sharing system operators, enabling them to optimize bike availability, anticipate demand fluctuations, and provide a seamless user experience for riders.  With the cleaned data, we have performed Exploratory Data Analysis to understand our dataset like The dependent variable - rented bike counts is positively skewed. Normally distributed attributes: temperature, humidity. Positively skewed attributes: wind, solar radiation, snowfall, rainfall. Negatively skewed attributes: visibility.  Our motive in whole project was to analyze the data and find out main components that can seamlessly predict the bike demand at any given point of time. After completion of analysis I predictive models that can predict the demand for rental bikes based on different weather conditions and other factors and, they were evaluated using RMSE. The XG Boost prediction model had the lowest RMSE. We developed Shapely value plots to understand the predictions obtained from the XG Boost model. The final choice of model for deployment depends on the business needs; if high accuracy in results is necessary, we can deploy XG Boost model. If the model interpretability is important to the stakeholders, we can choose deploy the decision tree model.  XGBoost, or Extreme Gradient Boosting, has emerged as a powerful algorithm for various machine learning tasks, including bike sharing demand prediction. It has demonstrated superior performance compared to other algorithms in several aspects, making it a popular choice for predictive modeling tasks.  XGBoost can effectively capture interactions between features and make accurate predictions. This makes it particularly suitable for bike sharing demand prediction, where the relationships between factors such as time, weather, and rental counts can be intricate.  This bike sharing demand prediction model, based on the Linear Regression algorithm, has exhibited exceptional performance and reliability in forecasting the number of bikes rented from our bike-sharing system. The model's accuracy and predictive power have been validated through rigorous testing and evaluation.  In terms of accuracy, our model consistently achieves low error rates, as measured by metrics such as root mean square error. This indicates that the predicted rental counts closely align with the actual values observed in our historical data. We have also conducted thorough cross-validation to ensure the model's stability and generalization capability.  Additionally, our model has demonstrated robustness in capturing complex relationships and non-linear patterns within the data. By incorporating features such as time of day, weather conditions, day of the week, and other relevant factors, the model effectively captures the dynamic nature of bike rental demand.  Furthermore, the model's scalability and efficiency enable it to handle large volumes of data and make predictions in real-time. This ensures that our bike-sharing system can respond promptly to fluctuations in demand and optimize bike availability for our users.  Overall, the performance of our bike sharing demand prediction model based on XGBoost has proven to be highly accurate, reliable, and scalable. By leveraging the model's insights, we can efficiently allocate our resources, enhance user experience, and maximize the utilization of our bike-sharing system. |