**Bike-Sharing Rental**

**Problem Statement :** This dataset aims to address is the prediction of bike rental demand based on various external factors such as weather, season, and holiday information. Accurately predicting bike rental demand can help bike sharing companies optimize their operations, including bike availability, pricing strategies, and marketing efforts.

1. **Exploratory Data Analysis (EDA):**

* Describe the dataset: Understand the structure, columns, and data types.
* Clean the data: Check for invalid records, missing values, duplicated records, and outliers.
* Handle missing values: Detect missing values and apply appropriate imputation techniques.
* Detect outliers: Identify outliers and decide whether to remove or transform them.

1. **Data Visualization:**

* Creating visualizations such as scatter plots, line plots, and bar plots to explore relationships between variables.
* Plotting time series data to analyze trends and seasonality.
* Using heatmaps or correlation matrices to understand the correlation between different features.
* Utilizing seaborn, matplotlib, or other libraries for generating informative and visually appealing plots.

1. **Feature Engineering:**

* Generating new features from existing ones to improve model performance.
* Handling categorical variables through techniques like one-hot encoding or label encoding.
* Scaling or normalizing numerical features to ensure uniformity in their ranges.
* Incorporating domain knowledge to create meaningful features that capture important aspects of the data.

1. **Model Building:**

* Selecting appropriate machine learning algorithms such as Decision Tree, Random Forest, and Gradient Boosting Regression for predicting bike rental demand.
* Splitting the dataset into training and testing sets for model evaluation.
* Training various models using the training data and evaluating their performance using metrics like RMSE, MAE, or R-squared.
* Fine-tuning model parameters to optimize performance.

1. **Hyperparameter Tuning:**

* Conducting hyperparameter tuning for Decision Tree, Random Forest, and Gradient Boosting Regression using techniques like grid search or random search to find the best combination of model parameters.
* Evaluating model performance with different hyperparameter values using cross-validation techniques.
* Balancing model complexity and performance to prevent overfitting or underfitting.

1. **Model Evaluation:**

* Evaluation metrics for regression models: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R^2), etc.
* Compare the performance of different models to select the best one.

1. **Model Deployment**

* Once the best model is selected, deploy it in a production environment where it can be used to make predictions.
* Monitor the performance of the deployed model over time and update it as needed.