Architecture Design on

Rental BikeShare Demand Prediction

MD RIAN UR RAHAMAN

Contents

Abstract ……………………………………………………………………………...

1. Introduction ………………………………………………………………………….
   1. Why this Architecture Design Document ? ……………………………..
2. Architecture …………….……………………………………………………………
3. Architecture Design ………………………………………………………………
   1. Data Collection ……….……………………………………………………
   2. Data Description …..…………………………………………………
   3. Loading dataset ……………………………………………..
   4. Data Preprocessing ………………………………………………..….
   5. Modeling Process ………………....……………………...…….…..
   6. UI Integration ……………………………………………………..……..
   7. Data from User …………………………………………………………..…
   8. Data Validation …….………………………………………………….…...
   9. Rendering the Results …………….………………………………….
   10. Deployment ………………………………………………………...

# Abstract

The project "Rental Bike Share Prediction" focuses on developing a machine learning solution to predict bike rental demand within bike-sharing systems. These systems automate the rental process, allowing users to rent and return bikes at different locations. With the rise of bike-sharing programs globally, the data generated by these systems offers valuable insights for traffic management, environmental planning, and public health.

This project will involve building an end-to-end regression model to predict bike rental counts. The solution will be evaluated based on code quality, including modularity, safety, testability, maintainability, and portability. The project will use a Cassandra database for data storage, and the solution will be hosted on a cloud platform like AWS, Azure, or GCP. The project will also require detailed documentation of system architecture, high-level and low-level design, and response time analysis.

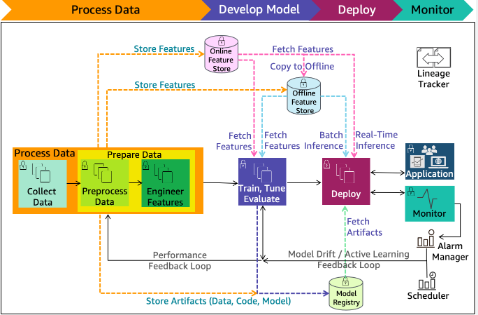
The final deliverable will include a fully optimized solution with deployment options, either as an API or a user interface, and will be maintained on a public GitHub repository. The project will adhere to industry best practices, including logging, coding standards, and possibly integrating an AI ops pipeline for continuous integration and deployment.

# Introduction

## Why this Architecture Design Document ?

The main objective of the Architecture design documentation is to provide the internal logic understanding of the Rental Bike share demand prediction code. The Architecture design documentation is designed in such a way that the programmer can directly code after reading each module description in the documentation.

# Architecture

****

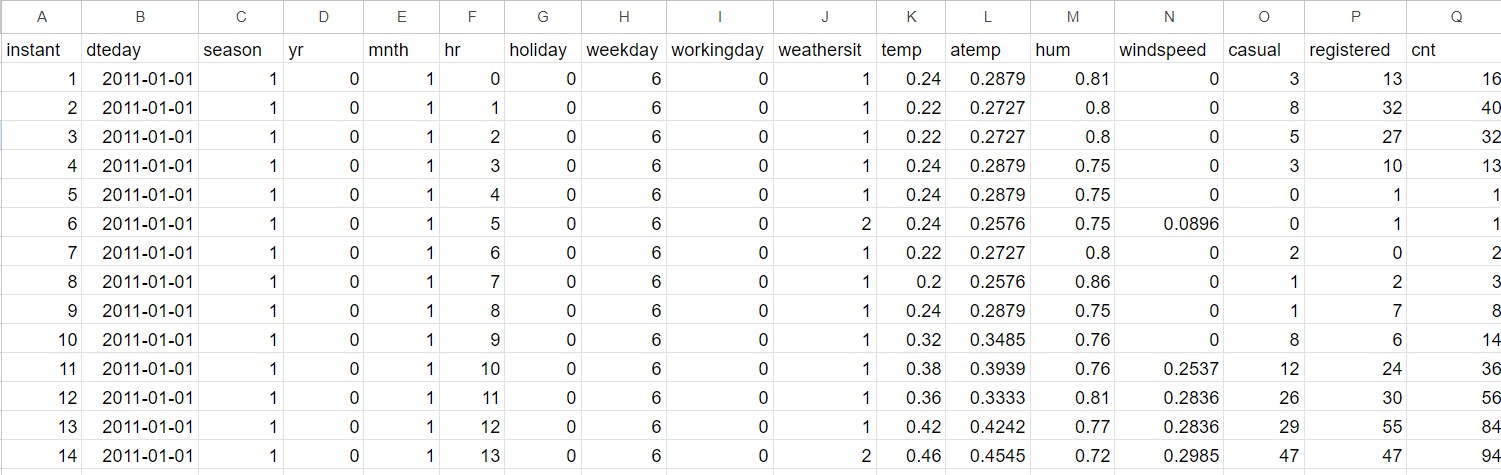
1. **Architecture Design**

## Data Collection

The dataset was taken from the UCI Machine Learning Repository . Dataset hour.csv is available in this link [https://archive.ics.uci.edu/dataset/275/bike+sharing+dataset](https://archive.ics.uci.edu/dataset/275/bike%2Bsharing%2Bdataset)

## Data Description

Rental Bikeshare Demand Prediction is 17K+ dataset publicly available on UCI Repository. The information in the dataset is present in a csv file named hour.csv. Dataset contains 17379 rows which shows the information such as index instant, Dateday, season, year, month, hour, holiday, weekday, workingday, weathersit, temperature, atemperature, humidity, windspeed, casual, registered, count. The glance of the Dataset is :



## Loading dataset

Local file system was used for loading the dataset (hour.csv) using read\_csv function of Pandas Library

## Data Preprocessing

* + - All the necessary libraries were imported first such as Numpy, Pandas, Matplotlib, Seaborn, sk-learn.
    - Checking the basic profile of the dataset. To get a better understanding of the dataset.
      * Using Info method
      * Using Describe method
      * Checking for unique values of each column.
    - Checking for info of the Dataset, to verify the correct datatype of the Columns.
    - Checking for Null values, because the null values can affect the accuracy of the model.
    - Doing Feature selection and dropping those columns which are not needed.
    - Converting the integer columns [‘season’, ‘yr’, ‘mnth’, ‘hr’ , ‘holiday’ , ‘weekday’ , ‘workingday’, ‘weathersit’] into categorical columns.
    - Performing standard scaling on numerical and categorical columns. And, One Hot Encoding on Categorical columns.

Now, the info is prepared to train a Machine Learning Model.

## Modeling Process

* After preprocessing the data, the data will be split into 2 sets X and y. X contains all the columns except the target column in our case ( Count ), y contains only the Target column.
* Using train test split we first split the dataset into X\_train,X\_test, y\_train, y\_test .
* Then use following Regression Algorithms like: GradientBoosting, Random Forest, Bagging, Decision Tree, Gradient Boost, KNeighbors, Linear Regression to predict the count of Rental Bikes required based on certain conditions and achieve 91% accuracy.
* After creating the following normal without any hyper tuned models, find the best performing model out of all.
* Then use Grid Search CV to find the best suiting parameters for all the models and again find out the best performing model after Hyperparameter tuning.

## UI Integration

Both CSS and HTML files are being created and are being integrated with the created machine learning model. All the required files are then integrated to the app.py file and tested locally.

## Data from User

The data from the user is retrieved from the created HTML web page.

## Data Validation

The data provided by the user is then being processed by app.py file and validated. The validated data is then sent to the prepared model for the prediction.

## Rendering the Results

The data sent for the prediction is then rendered to the web page.

## Deployment

The tested model is then deployed to Heroku.