High Level Design (HLD)

**Bike Share Prediction**

**Revision Number:**

**Last date of revision:**

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| Data Issued | Version | Description | Author |
|  |  |  |  |
|  |  |  |  |

**Contents**

Document Version Control

Abstract

1. Introduction
   1. Why this High-Level Design document?
   2. Scope
   3. Definitions
2. General Description

**Abstract**

Bike sharing systems are new generation of traditional bike rentals where whole process from membership, rental and return back has become automatic. Through these systems, user is able to easily rent a bike from a particular position and return back at another position. Currently, there are about over 500 bike-sharing programs around the world which is composed of over 500 thousand bicycles. Today, there exists great interest in these systems due to their important role in traffic, environmental and health issues. Apart from interesting real-world applications of bike sharing systems, the characteristics of data being generated by these systems make them attractive for the research. The work discusses the implementation a Machine learning backend to predict the availability of bikes for sharing, that can be best picked by the users to save time for their usual works and make them likely to choose this mode of transportation for helping the greater benefits of health, fuel consumption and many more.

1. **Introduction**
   1. **Why this High-Level Design document?**

The purpose of this High-Level Design (HLD) document is to add necessary details to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how modules interact at a high level.

The HLD will:

* Present all the design aspects and define them in detail
* Describe the user interface being implemented
* Describe the hardware and software interfaces
* Describe the performance requirements
* Include the design features and architecture of the project
* List and describe the non-functional attributes like:
* Security
* Reliability
* Maintainability
* Portability
* Reusability
* Application compatibility
* Resource utilization
* Serviceability
  1. **Scope**

The HLD documentation present the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly technical terms which should the understandable to the administrator of the system.

* 1. **Definitions**

|  |  |
| --- | --- |
| Term | Description |
| Database | Collection of all the information monitored by this system |
| IDE | Integrated Development Environment |
| AWS | Amazon Web Service |

1. **General Description**

**2.1 Product Perspective**

The Bike Sharing Prediction system is a machine learning based regression model which will help us to predict the availability of the sharable rental bikes from a particular position and return back at the user’s destination.

**2.2** **Problem Statement**

To crate an AI solution to understand the factors on which the demand for these shared bikes depend upon.

**2.3** **Proposed Solution**

The solution proposed here is to build a Machine learning Regression model to perform the above used cases. In first case, the model will predict the bookings of the in the given weather conditions, week-days, months and timings and will try to maintain the availability of bikes to be rent on the peak timings. Further, in the second, management can try to manage the bikes properly according to the customers feasibility.

**2.4 Further Improvements**

The same model can also be used to process real time streaming data which can be received from GPS (Global Positioning System) to provide the best possible location of the bikes with-in a place to rent a bike easily. Also, it can also be integrated with the location of parking lot dataset for proper parking of bikes after users finish their ride. Further if the model is also trained to predict the traffic with relation to the hours of the day and weekdays, we can provide a better option to the users to where and when to pick a rent.

**2.5 Data Requirements**

Data requirements are completely depended on our problem statement.

* We need the office holiday and working day data for the whole year.
* We will also need weather report for the whole year such as temperature, air temperature, humidity and windspeed.

**2.6 Tools used**

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn.

* PyCharm is used as IDE
* MySQL/Cassandra is used to retrieve, insert, delete and update the database.
* AWS is used for deployment of the model.
* GitHub is used for version control
* For front end development HTML/CSS is used.
* Python FLASK is used for backend development.

**2.7 Constraints**

The Bike share prediction model should be user friendly, as automated as possible and user shouldn’t be required to know any of the workings.

**2.8 Assumptions**

The main objective of the project is to implement the use cases as previously mentioned (2.2 Problem Statement) for new dataset that comes through with reference to bike share prediction. Machine Learning based regression model is trained with the available dataset to predict the bike share with the new dataset provided by the user.

**3 Design Details**

**3.1 Process Flow**

For identifying the different types of anomalies, we will use a Machine Learning based regression model. Below is the process flow diagram is as shown below.

**Proposed Methodology**

DATA COLLECTION

ANNONATION CHECK

DATA PREPROCESSEING

BUILD MACHINE LEARNING MODEL FOR BIKE SHARE PREDICTION

PREDICT ANOMALIES (use cases)

TAKE NECESSARY ACTION

**3.1.1 Model training and evaluation**

(7) Prediction

New Data Input

Training set

(3) Data validation check

(5) Training

Result of the evaluation

**Model**

Annotated Dataset

dataset

Dataset

(2) (4)

(1) Annotations Split

(6)

Test set

Evaluation

Prediction

**3.1.2 Deployment Process**

**Load Model**

**Get frames from ML**

**Preprocess Data**

**Make Prediction**

**Predict Result**

**3.2** **Event Log**

The system should log every event so that the user will know what process is running internally.

**Initial step-by-step Description:**

1. The system identifies what step logging is required.
2. The System should be able to log each and every system flow
3. Developer can choose logging method, either database logging or file logging as well
4. System shouldn’t hang even after using so many loggings. Logging is done so that we can debug issues at process level, so it is mandatory.

**3.3 Error Handling**

System should be able to handle the detected error and an explanation should be displayed is to what went wrong? An error will be defined as anything that falls outside the normal and intended usage.

**4 Performance**

The Bike share prediction model is used for predicting the bike share demand in the customers in different hours, days and weather conditions in a year, so it should be as accurate as possible for giving most probable prediction. Also, model retraining is also very important to improve the performance.

**4.1 Reusability**

The code written and the components used should have the ability to be reused with no problems.

**4.2 Application Compatibility**

The different components for this project will use python as an interface between them. Each component and module will have its own task to perform, it is the job of the python to ensure the proper transfer of information.

**4.3 Resource Utilization**

When any task is performed, it will likely use all the processing power available until that function is finished.

**4.4 Deployment**

* Microsoft Azure
* Google Cloud Platform (GCP)
* Amazon Web Service (AWS)

**5 Conclusion**

The designed Bike share prediction model will predict and give a relation between the factors and demand for the shared bikes based on the various anomalies data used to train our algorithm. So that, the company will conclude the following points:

* Which variable are significant in predicting the demand for shared bikes
* How well those variables describe the bike demands

**6 References**

1. **Dataset-** <https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>