HIGH LEVEL DESIGN (HLD)

Bike Share Prediction

Final Document Version

Date: 04/08/2022

DOCUMENT VERSION CONTROL

Author:

Jaddumahanthi Vinay

Date:04/08/2022

Final Version

CONTENTS:

Document Version Control………………………………………………2

Abstract……………………………………………………………………...4

1 Introduction…………………………………………………………..5

* 1. Why this High Level Design Document?..........................5
  2. Scope……………………………………………….……….....5

1. General description………………………………………………....6
   1. Product perspective………………………………………….6
   2. Problem statement…………………………………………...6
   3. Proposed solution……………………………………………..6
   4. Further improvements………………………………………..6
   5. Technical requirements………………………………………6
   6. Data requirements……………………………………………6
   7. Tools used……………………………………………………….6
2. Design Details…………………………………………….……………7
   1. Process flow…………………………………………………….7
   2. Logging………………………………………………………….8
   3. Exception handling……………………………………………8
3. Performance…………………………………………………………...9
   1. Resource utilization…………………………………………….9
   2. Deployment………………………………………………….....9

5 Conclusion…………………………………………………………....10

Abstract

There is a rising trend in usage of bikes, particularly sharing bikes with increasing environmental consciousness, health awareness and traffic problems associated with urbanization. Hence, bike sharing or rentals is a highly growing business with immense potential particularly due to ease in renting and handing over, automated process and lack of ownership and maintenance problems. In addition, here is a huge data driven business and research opportunity which itself is very huge in itself, may be bigger than even rental business. This is because departure, arrival, duration of travel, locations of travel etc are recorded in these systems. As such, due to these dual sides of business, predicting exact demand for bikes in necessary to reduce costs related to excess supply as well as to harness full potential of rental and data analysis business. Thus, this project aims to predict demand of users up to the time interval of hours based on previous usage and using end to end regression.

1 Introduction

* 1. Why this High Level Design Document?

The document gives high level perspective, analyses problem and provides solution at high level, clears all kinds of doubts and can be used as reference for project execution and coding.

1.2 Scope

HLD includes:

1. Design aspects
2. User interface
3. Hardware and software details
4. Key performance indicators
5. Product architecture
6. Security features, maintenability, reliability etc.
7. General Description

2.1 Product perspective

Product predicts demand for bikes based on certain prerecorded features that significantly affect bike demand. Machine learning technique, regression is used for prediction.

2.2 Problem statement

Predict the demand for bike sharing.

2.3 Proposed solution

Based on previous data, various regression models are applied to be able to predict demand in future. Best model is selected, trained and is used for demand prediction. Various important indicators necessary for demand prediction are identified and given importance accordingly.

2.4 Further improvements

Future data can be used to retrain the model continuously for better predictions. Research and development can identify further important indicators that can be applied in regression model.

2.5 Technical requirements

Due to small amount of present data availability, simple computers and IDEs with 8GB RAM would be sufficient for machine learning. Any cloud platform based on requirement of user could be used for deployment of model.

2.6 Data requirement

Data is provided by client based on Data Sharing Agreement (DSA). The data provided includes key indicators that predict bike demand and is provided in csv format. Data is provided in one go as batch files.

2.7 Tools used

Numpy, scikit-learn, pandas, flask etc to build the model. Pycharm IDE is used. Model will be deployed on user determined platform.



1. Design Details
   1. Process flow

Start

Data insertion in database

Data transformation

Save model

Export data to CSV

Data insertion in database

Data transformation

Data validation

Get training data through API

Export data to CSV

Data preprocessing

Data clustering

Data validation

Apply best model

Get data for prediction from API

Data preprocessing

Find best model

Data clustering

End

Generate prediction file

Push app to cloud (deployment)

* 1. Logging

Logging is done to let user know of the processes running internally. It also helps in easy maintenance of product.

* 1. Exception handling

Exception handling is done to prevent unusual stoppage of code and to address unexpected and rare shocks.

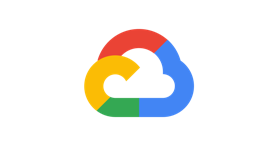
4 Performance

4.1 Resource utilization

All the processors are utilised.

* 1. Deployment

Project will be deployed in user preferred cloud or private server.



1. Conclusion

Client will be able to predict and meet demand for bikes to maximise earning potential.