LOW LEVEL DESIGN (LLD)

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 Low Level Document?..............................................5
  2. Scope……………………………………………….………......5
  3. Constraints……………………………………………………...5
  4. Risks………………………………………………………….....5

1. Technical specifications………………………………………………..6
   1. Dataset……………………………………………………...….6
   2. Dataset overview…………………………………………….....6
   3. Data validation, transformation and preprocessing………….....6
   4. Predicting bike demand………………………………………...6
   5. Logging…………………………………………………………6
   6. Exception handling…………………………………………...…6
   7. Database……………………………………………………….6
   8. Clustering………………………………………………………..7
   9. Deployment……………………………………………………..7
2. Technology stack………………………………………………………8
3. Proposed solution…………………………………………………...…9
4. Architecture……………………………………………………………10

6 Conclusion…………………………………………………………...…11

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 Low Level Design Document?

The document gives low level perspective and detailed description of how the product is built. Both stakeholders and developers can benefit from this document and clear all issues and doubts. The objective of this project is to predict the demand for bikes based on certain defining indicators. These indicators include:

1. Temperature
2. Humidity
3. Day of week
4. Whether it is workingday
5. Whether it is holiday
6. Month
7. Season
8. Hour
9. Weather situation and other indicators
   1. Scope

Scope of the project is limited to predicting the demand for bikes.

* 1. Constraints

The data is old and limited to only two years. Many new indicators and changed circumstances are not available like mobility infrastructure, traffic, behavioral change, new technologies etc. The dataset is also small.

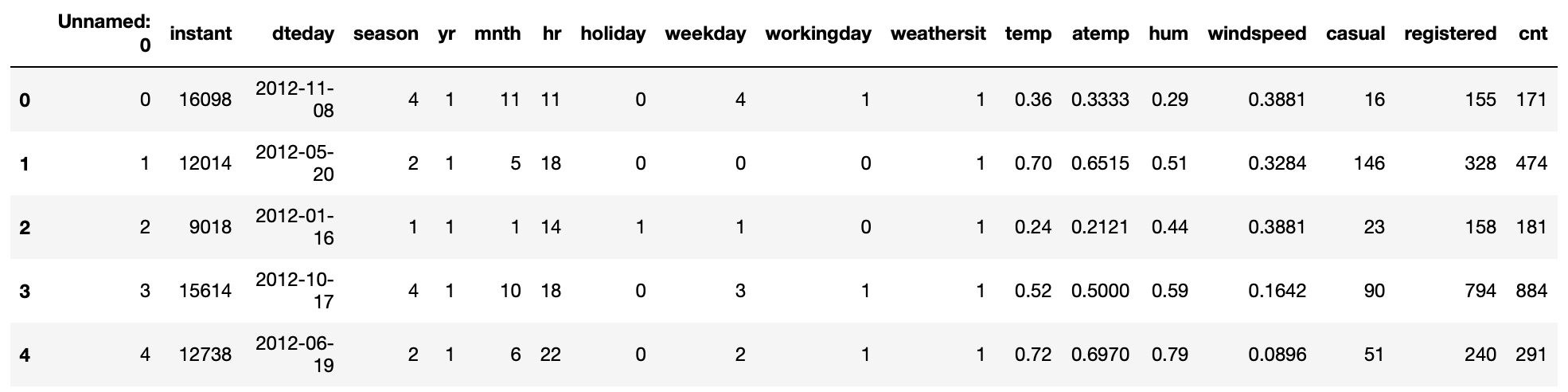
* 1. Risks

Risks in prediction are posed by the above constraints and other challenges that are not yet visible.

1. Technical specifications

|  |  |  |
| --- | --- | --- |
| **Filename** | **Finalized** | **Source** |
| hour.csv | yes | https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset |

* 1. Dataset
  2. Dataset overview



* 1. Data validation, transformation, insertion and preprocessing

Data is validated as per Data Sharing Agreement (DSA). File name, file type, column name, column data type, time and date stamps are checked. Data is separated into good data and bad data based on data validation. Good data is transferred to Cassandra database while bad data is archived. Data is preprocessed to impute null values and to drop unimportant columns.

* 1. Predicting bike demand

Various regression models are applied, trained and best one is selected to predict demand.

2.5 Logging

Logging is done at every step. File logging is used.

* 1. Exception handling

Exception handling is done at every step to address unforeseen and rare events.

* 1. Database

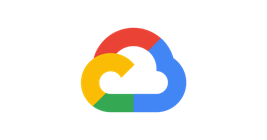
Cassandra database has been used.

* 1. Clustering

Clustering of training data is done and best regression model is selected for each cluster specifically. Prediction data is also clustered to predict using cluster specific best model.

* 1. Deployment

Deployment is done at client preferred cloud service (since multiple deployment options should be given).



1. Technology stack

|  |  |
| --- | --- |
| **Front end** | Postman with API |
| **Back end** | Cloud |
| **Database** | Cassandra |
| **Deployment** | AWS/Azure/Google cloud |

1. Proposed solution

Various regression models would be applied to training dataset and best regression model (based on accuracy score) would be selected to predict bike demand. These regression models include linear, random forest, support vector, lasso, ridge, elasticnet and decision tree.

1. Architecture

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. Conclusion

Key indicators that influence bike demand include temperature, humidity, hour, day of week, whether it is working day or holiday, month, season and weather situation.