HIGH LEVEL DESIGN (HLD)

APS FAILURE AT SCANIA TRUCKS

Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 1-04-2024 | 1 | Initial HLD-V1.0 | Prince Francis |
|  |  |  |  |
|  |  |  |  |

CONTENTS

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

Abstract……………………………………………………………………………………………………………………4

1. Introduction………………………………………………………………………………………………………5
   1. Why this Hight-Level Design Document?........................................................5
   2. Scope…………………………………………...................................................................5
   3. Definitions…………………………………………………………………………………………………..5
2. General Description……………………………………………………………………………………………6
   1. Product Perspective…………………………………………………………………………………….6
   2. Problem Statement……………………………………………………………………………………..6
   3. Proposed Solution……………………………………………………………………………………….6
   4. Further Improvements…………………………………………………………………………………6
   5. Hardware Requirements………………………………………………………………………………6
   6. Data Requirements………………………………………………………………………………………6
   7. Tools Used……………………………………………………………………………………………………7
   8. Constraints…………………………………………………………………………………………………..8
   9. Assumptions…………………………………………………………………………………………………8
3. Design Details……………………………………………………………………………………………………..9
   1. Model Training and Evaluation……………………………………………………………………..9
   2. Deployment Process…………………………………………………………………………………..10
   3. Event log…………………………………………………………………………………………………….10
   4. Error Handling……………………………………………………………………………………………10
4. Performance…..…………………………………………………………………………………………………10
   1. Reusability………………………………………………………………………………………………….11
   2. Application Compatibility……………………………………………………………………………11
   3. Resource Utilization……………………………………………………………………………………11
   4. Deployment……………………………………………………………………………………………….11
5. Conclusion…………………………………………………………………………………………………………11
6. References…………………………………………………………………………………………………………11

Abstract

The dataset consists of data collected from heavy Scania trucks in everyday usage. The system in focus is the Air Pressure system (APS) which generates pressurised air that are utilized in various functions in a truck, such as braking and gear changes. The datasets positive class consists of component failures for a specific component of the APS system. The negative class consists of trucks with failures for components not related to the APS.

**Challenge metric**:

Cost-metric of miss-classification:

Predicted class | True class |

| pos | neg |

-----------------------------------------

pos | - | Cost\_1 |

-----------------------------------------

neg | Cost\_2 | - |

-----------------------------------------

Cost\_1 = 10 and cost\_2 = 500

The total cost of a prediction model is the sum of "Cost\_1" multiplied by the number of Instances with type 1 failure and "Cost\_2" with the number of instances with type 2 failure, resulting in a "Total\_cost". In this case Cost\_1 refers to the cost that an unnecessary check needs to be done by an mechanic at an workshop, while Cost\_2 refer to the cost of missing a faulty truck, which may cause a breakdown.

Total\_cost = Cost\_1\*No\_Instances + Cost\_2\*No\_Instances.

1. Introduction
   1. Why this High-Level Design Document?

The purpose of this High-Level Document (HLD) is to add the necessary detail 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 the modules interact at a high level.

The HLD will:

* Present all of 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 design features and the 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 presents 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 be understandable to the administrators of the system.

* 1. Definitions
* AWS- Amazon Web Services
* VS – Visual Studio

1. GENERAL DESCRIPTION
   1. Product Perspective

The Aps failure at Scania trucks solution system is a Machine Learning based solution system which predicts the class to which the fault in the components of the truck is related to. This will further help to reduce the total maintenance cost related with the truck.

* 1. Problem Statement

To predict the class to which the fault of the component belongs to.

* 1. Proposed Solution

After collection of the data we have to analyze the data and check the various relationships present in the data. We can then select the most suitable features to build our model on. Lastly, using various Supervised Machine Learning Techniques we can build the model on top of the selected features and then evaluate their performance.

* 1. Further Improvements

In order to further improve the model we can use hyperparameter tuning . This can be conducted for various Machine Learning models which are performing well in our initial analysis.

* 1. Hardware Requirements

This document addresses the hardware requirements for the implementation of this project. This includes:

* A system which is capable of doing basic Machine Learning tasks such as visualizing the data, and training the models.
  1. Data Requirements

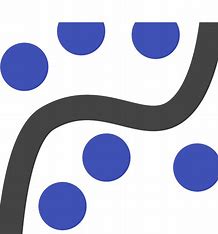
Data requirement completely depend on our problem statement.

* We require as much as relevant train data for our problem statement.
* The rows of the dataset can go upto 1,00,000.
* Data should be represented in the table format where each column represents features and a row represents an instance towards these features.
* Data can be present in the csv or excel format.
* Besides we would need a Test set to predict the Prices on.
* csv - CSV is a plain text format used to represent tabular data, where each line in the file represents a row, and columns are separated by a delimiter, typically a comma.
* Xlsx - XLSX is a file format used by Microsoft Excel to store spreadsheet data. It is part of the Microsoft Office Open XML format introduced in Excel 2007.

**NOTE**: We can also import data if it is in the JSON file format. However the data should be present in such a manner that it can be converted to a tabular form.

* 1. Tools Used

Python programming language and frameworks such as NumPy, Pandas , Scikit-learn, Statsmodels are used to build the whole model.



* VS code is used as IDE
* For visualization of the plots Matplotlib, Seaborn and Plotly are used.
* AWS is used for deployment of the model.
* MongoDB is used to insert and retrieve the data from the database.
* Github is used a version control system.
* Apache airflow is used to schedule the task.
  1. Constraints

The whole system solution must be user friendly, as automated as possible and users should not be required to know any of the workings.

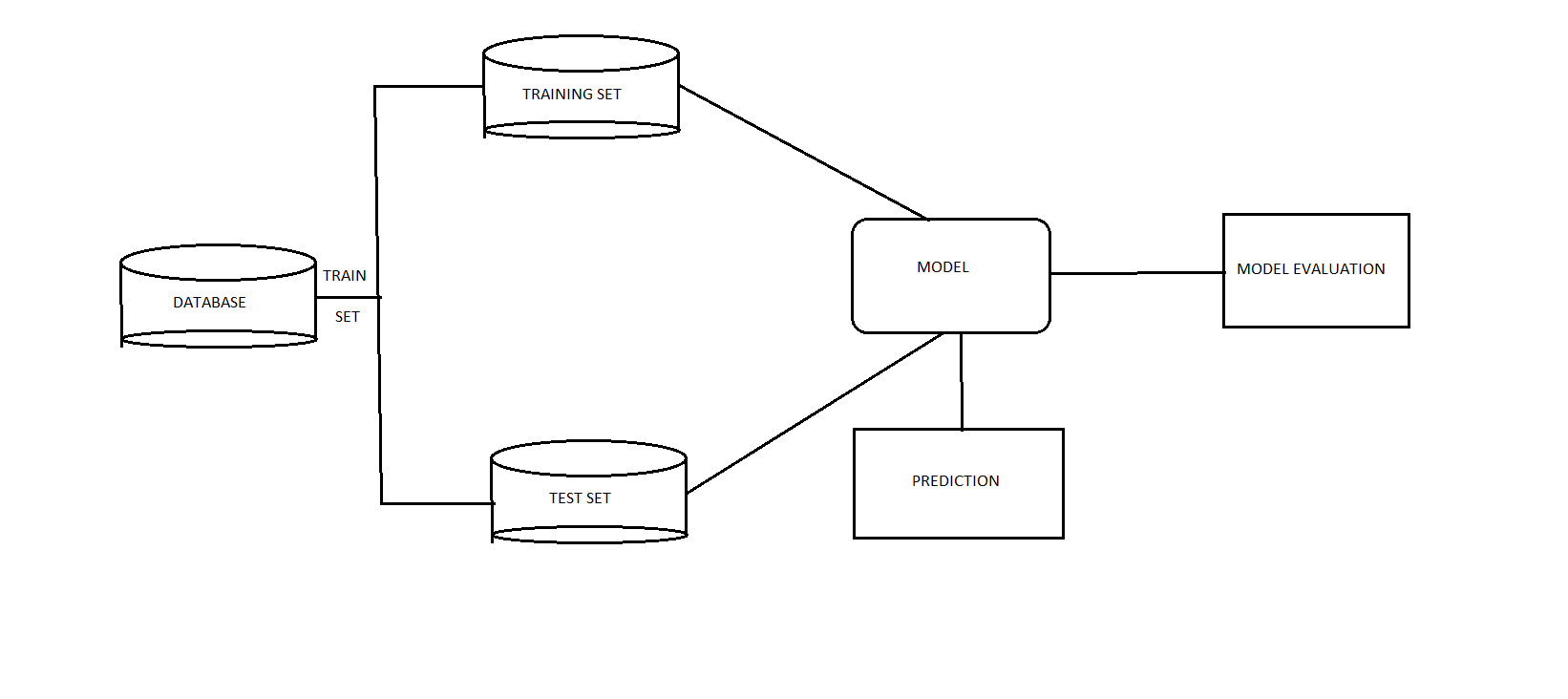
* 1. Assumptions

The main objective of the project is to implement the make predictions of the flight prices as previously mentioned (2.2 Problem Statement) for dataset that is coming through from the client end. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

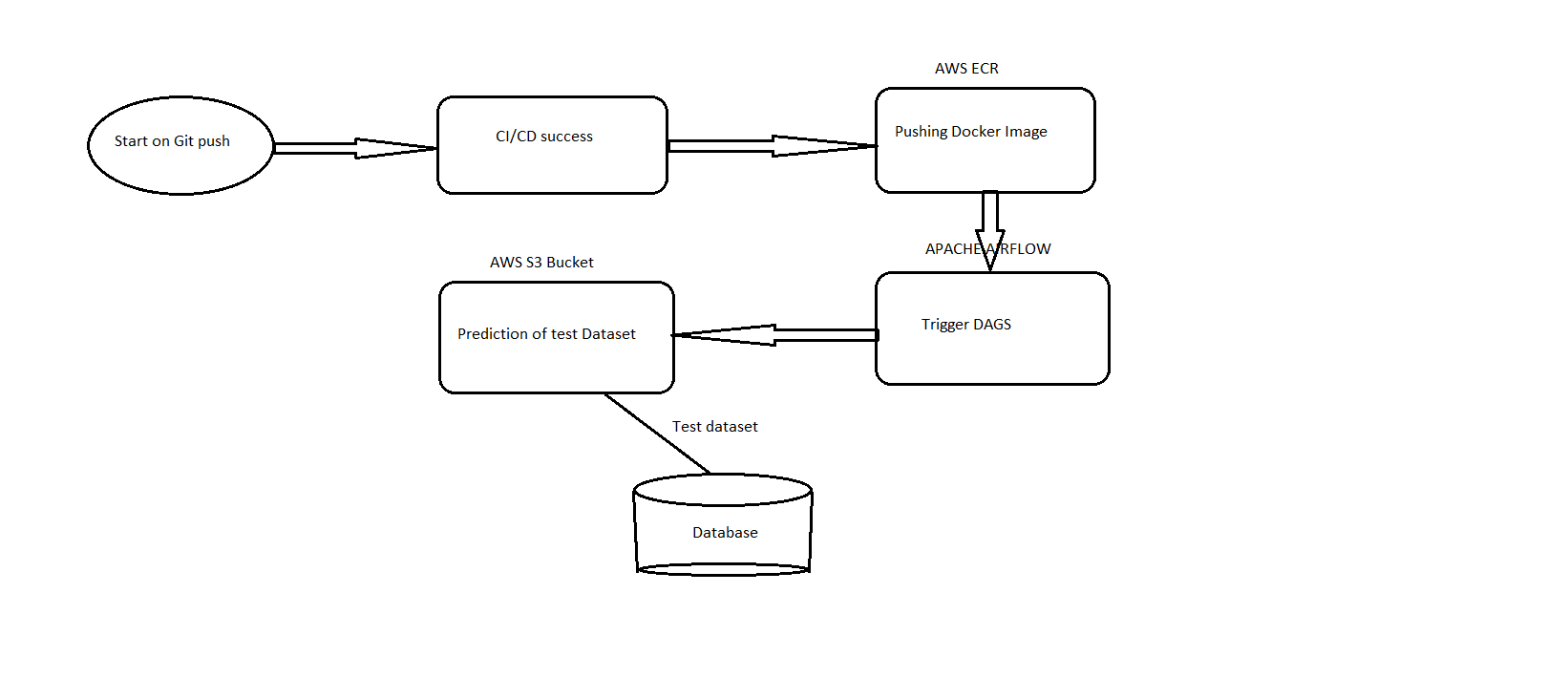
1. DESIGN DETAILS

Below is the process flow diagram as shown below:

* 1. Model Training and Evaluation



* 1. Deployment Process



* 1. Event Log

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

Initial Step-By-Step Description:

* The system identifies at what step logging required.
* The system should be able to log each and every system flow.
* Developer can choose logging method. You can choose database logging/File logging as well.
  1. Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage.

1. Performance

We are trying to improve the performance of our model so that it is generalizable on unseen datasets.

* 1. Reusability

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

* 1. Application Compatibility

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

* 1. Resource Utilization

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

* 1. Deployment



1. Conclusion

The designed solution for flight price prediction will predict Prices/Fares given the input datasets from the client-end.

1. References

* Google images