High Level Design (HLD)

Prediction of failure in APS of Scania Trucks

Revision Number: 1.0

Last date of revision: 01/03/2023

Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
| 01/03/2023 | 1 | HLD | Lovely Patra |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Contents

Document Version Control.....................................................................................................1

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

Introduction421`

1 .1 Why this High-Level Design Document? ..... . . . . . . . . . .5

1 .2 Scope. . . . . .5

1.3 Definitions ... .5

1. General Description. . . . . .6

2. 1 Product Perspective .. . . . . .6

2.2 Problem statement .. . . . . .6

#### 2.3 Business Problem .. . . . . .6

2.4 Machine Learning Problem. .6

2.5 Our metric .

2.6 Tools used........................................................................

 Design Details ..... ..... ..... ................................ . . . . . . 1 0

* 1. Process Flow....................................................................................................... 1 0
     1. Model Training and Evaluation . . . . . . 1 0
     2. Deployment Process 1 1
  2. Event log 1 1
  3. Error Handling . . . . . . . 1 1
  4. Performance........................................................................................................12
  5. Reusability................................................. . . . . . 1 2
  6. Application Compatibility. . . . . . 1 2
  7. Resource Utilization. . . . . . 1 2
  8. Deployment ..... ..... ........................ .. . . . . . . . 1 2

1. Dashboards. . . . . . 1 3
   1. KPls (Key Performance Indicators).............................................. . . . . . 1 3
2. Conclusion .. . . . 1

Abstract

## Air Pressure System is a vital part of heavy vehicles like Scania trucks. The braking system in these vehicles is dependent on-air pressure and hence there is a need for the proper functioning of the air pressure system. Predictive maintenance in automobile industry reduces the maintenance cost and improve the performance of the vehicle. This can be achieved either manually or automatically. Manual predictive maintenance needs interference of the human task and may induce some errors. Automatic predictive maintenance through artificial intelligence techniques explores the hidden cause for failure of air pressure system in the Scania trucks. ln the proposed system, machine learning approaches are investigated for predictive maintenance of the trucks hised on condition of air pressure system. The dataset used in this work consist of 35,188 negative instances and 1,000 positive instances. Hence there is a need to address the class imbalance issue therefore applying machine learning algorithms. Many resampling techniques like under sampling, over sampling and SMOTE are analysed for the efficiency of the classifier. After pre-processing the data, machine learning classification algorithms like Random Forest, Logistic Regression, Gradient Boosting, Decision Tree, K-Neighbors Classifier, XGBClassifier, CatBoosting Classifier and AdaBoost Classifier are implemented and accuracy of the classifiers are analysed. Experimental results show that XGBoost Classifier with 99.6% accuracy and cost of 2950.

# 1 Introduction

## 1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document 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 o Maintainability

 Portability

 Reusability

 Application compatibility o Resource utilization

 Serviceability

## 1.2 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.

### 1.3 Definitions

Term Description

|  |  |
| --- | --- |
| APS  Database  IDE  A WS | Air Pressure System  Collection of all the information monitored by this system  Integrated Development Environment  Amazon Web Services |

# 2 General Description

## 2.1 Product Perspective

The prediction of failures in the air pressure system of Scania Trucks is a machine learning-based object detection model which will help us to detect whether a failure of a Scania Truck component is related to the air pressure system (APS) or not.

2.2 Problem statement

### The Air Pressure System (APS) is a critical component of a heavy-duty vehicle that

### uses compressed air to force a piston to provide pressure to the brake pads, slowing

### the vehicle down. The benefits of using an APS instead of a hydraulic system are the

### easy availability and long-term sustainability of natural air.

### This is a Binary Classification problem, in which the affirmative class indicates that the

### failure was caused by a certain component of the APS, while the negative class

### indicates that the failure was caused by something else.

### Approach: The classical machine learning tasks like Data Exploration, Data Cleaning,

### Feature Engineering, Model Building and Model Testing. To try different machine

### learning algorithms that’s best fit for the above case.

# 2.3 Business problem

It would be great if I know the exact cause of failure.Scania Trucks, one of the largest truck producing company provided the data related to APS(air pressure system), one of the part in trucks. Among all different parts in truck, company especially want to know if the failure is related with APS or not. Basically the motive is to reduce the extra effort and time for unnecessary check of APS.

# 2.4 Machine learning problem

It is given that there is a failure in Truck, the task is to find weather the failure is related to APS or not. Simply in terms of Machine Learning it is a Binary Classification problem where there is only two possibilities:-

* **Failure related with APS.**
* **Failure does not related with APS.**

# 2.5 Our Metric

**Evaluating Factor** of a Machine Learning model is termed as Metric.

The Data contains two class labels:-

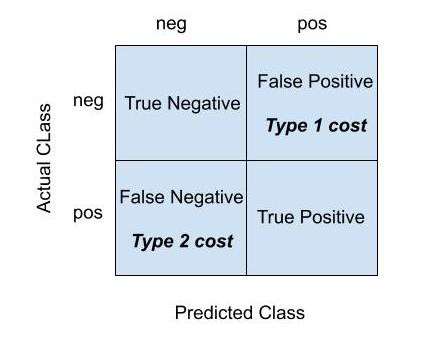
a)**‘neg’** means Failures **not related**with component of APS.

b)**‘pos’** means Failures **related**with components of APS.

Two different type of cost is defined:-

* Cost\_1 = The cost that an unnecessary check in APS needs to be done by a mechanic at workshop.
* Cost\_2 = The cost of missing a faulty truck i.e. the truck had APS failure but it is missed out, which may cause a breakdown.

We can relate it from Confusion matrix so let’s try to understand this from the perspective of confusion matrix.



**False positive** counts those prediction which have**“No APS failure”** but still predicted as “**APS failure.”**

**False negative**counts those prediction which have**“APS failure”** but predicted as “**No APS failure.”**

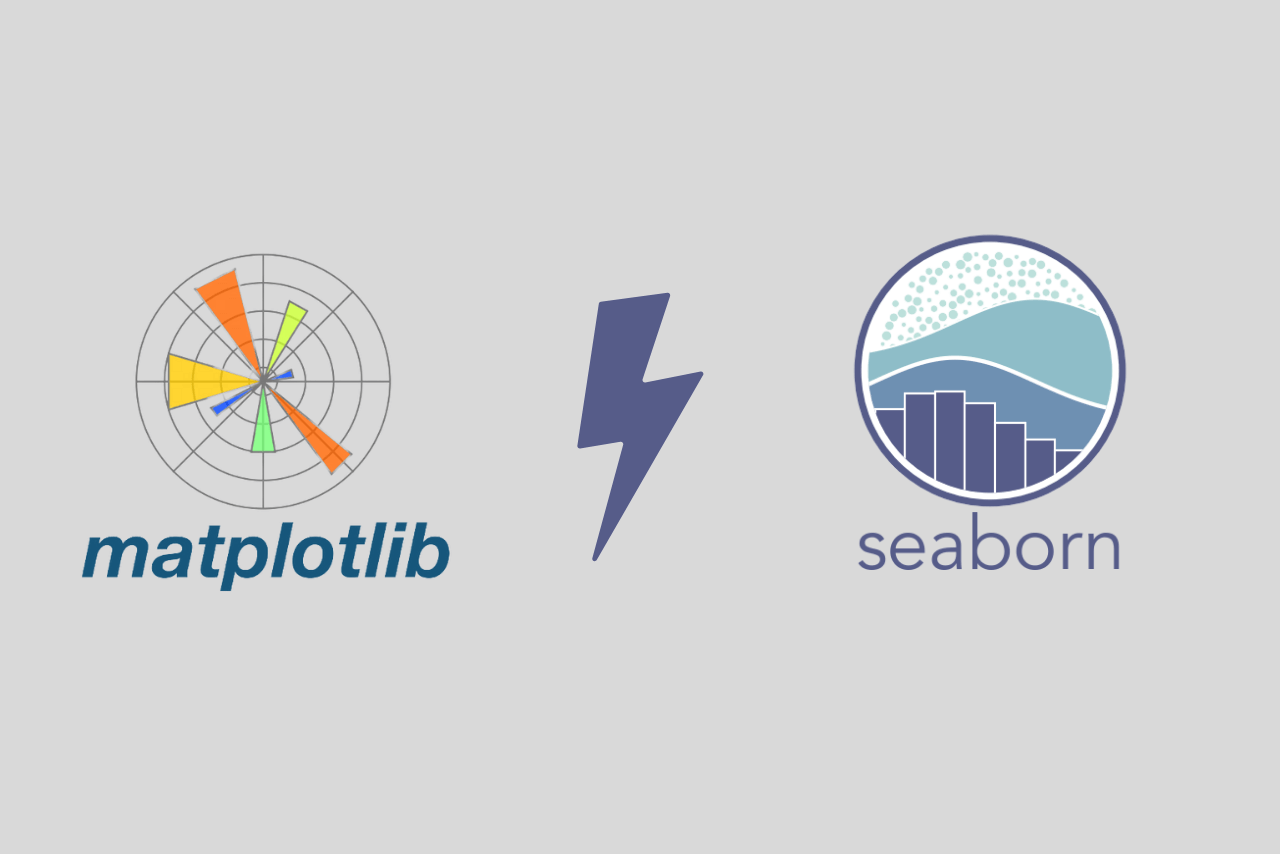
The costs are **“cost\_1 = 10”** and **“cost\_2 = 500”.**This makes sense that cost\_1 is less as unnecessary check need to be done for APS and there is no risk of breakdown where as cost\_2 is very high because of missing fault in APS, which may cause to breakdown.

**Total\_cost = False\_Negative\*500 + False\_positive\*10**

So, our **metric** is **total\_cost.**We will try to reduce it as much as possible which will certainly reduce the risk of failure as well as time for mechanic.

#### 2.6 Tools used

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



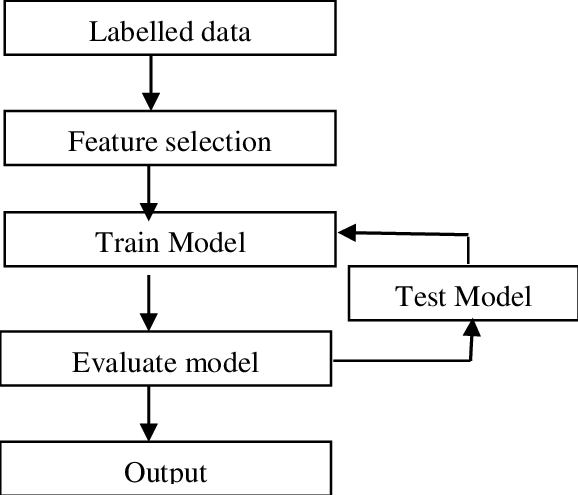
* VSCode is used as IDE.
* Pandas is used for data manipulation and data cleaning
* For visualization of the plots, Matplotlib, Seaborn and Plotly are used.
* AWS is used for deployment of the model.
* MySQL/MongoDB is used to retrieve, insert, delete, and update the database.
* GitHub is used as version control system

Design Details

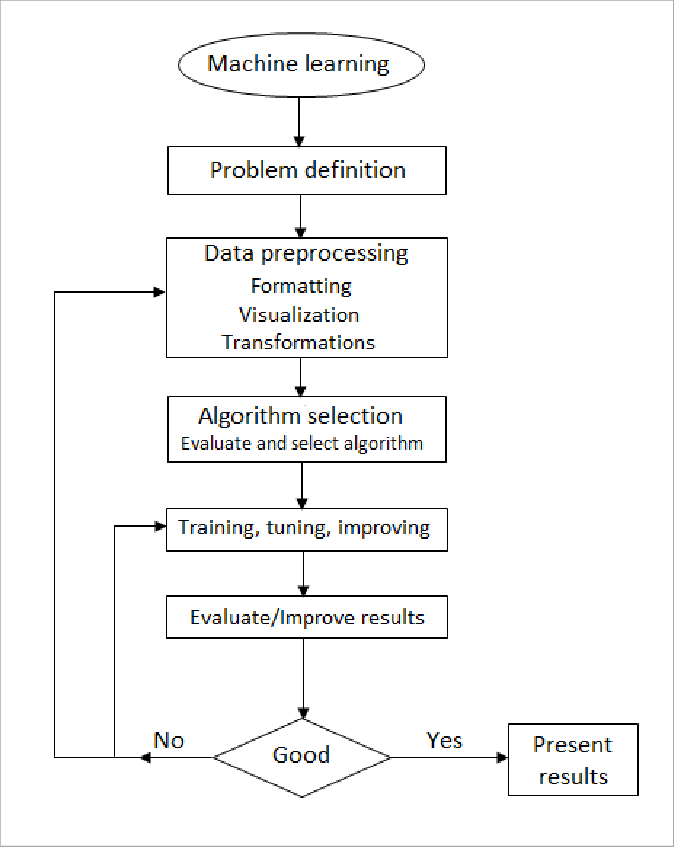
3.1 Process Flow

For identifying the different types of anomalies, we will use a deep learning base model. Below is the process flow diagram is as shown below.

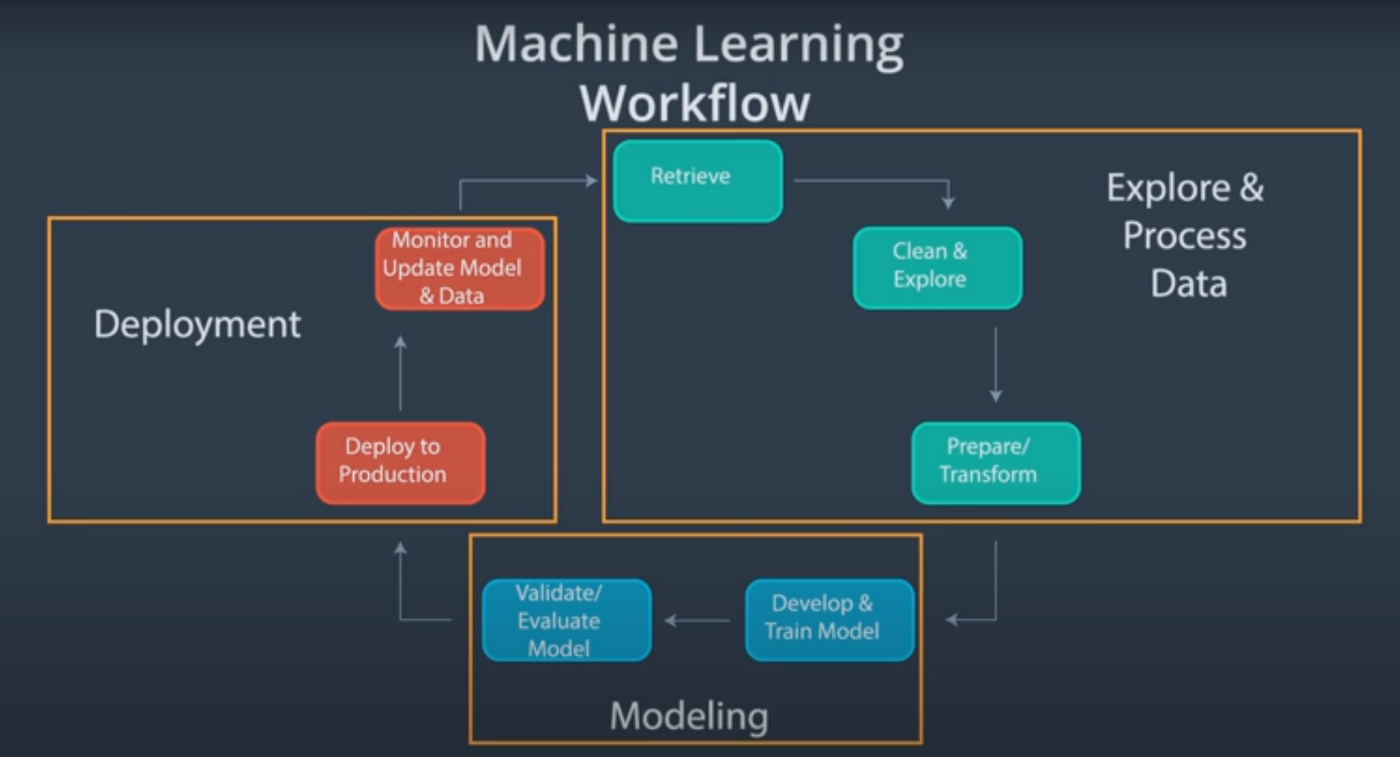
#### Proposed methodology



##### 3.1.1 Model Training and Evaluation



##### 3.1.2 Deployment Process



#### 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 at what step logging required
2. The System should be able to log each and every system flow.
3. Developer can choose logging method. You can choose database logging/ File logging as well.
4. System should not hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

#### 3.3 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.

#### 4.1 Reusability

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

#### 4.2 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.

#### 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

****

**AWS(AMAZON WEB SERVICES)**

Conclusion

An early detection of a failure in an Air Pressure System in trucks can save the

company a lot money. The prediction of a fault can be performed even if the

meaning of the measured values is unknown or only histograms are available.

We demonstrated how meaningful features of histograms can be computed to

improve the prediction. Also, we showed how the forecasts can be adapted to a

cost function using a threshold on the conﬁdence of a Random Forest. Finally, a signiﬁcantly lowered main cost compared to the naive approaches was achieved.

References

1. <https://towardsdatascience.com/predicting-a-failure-in-scanias-air-pressure-system-aps-c260bcc4d038>
2. https://www.researchgate.net/publication/309195602\_Prediction\_of\_Failures\_in\_the\_Air\_Pressure\_System\_of\_Scania\_Trucks\_Using\_a\_Random\_Forest\_and\_Feature\_Engineering