APS Fault Detection

A PROJECT REPORT

Submitted by

Krutarth Oza

[190120107077]

In partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

In

Computer Engineering Department

Gandhinagar Institute of Technology

Gandhinagar

Gujarat Technological University, Ahmedabad

May, 2022-23

GUJARAT TECHNOLOGICAL UNIVERSITY

CERTIFICATE FOR COMPLETION OF ALL ACTIVITIES AT ONLINE PROJECT PORTAL

B.E. SEMESTER VIII, ACADEMIC YEAR 2022-2023

Date of certificate generation : 10 May 2023 (11:21:31)

This is to certify that, ***Oza Krutarth Dipakbhai*** (Enrolment Number - 190120107077) working on project entitled with ***APS Fault Detection*** from ***Computer Engineering*** department of ***GANDHINAGAR INSTITUTE OF TECHNOLOGY, GANDHINAGAR*** had submitted following details at online project portal.

Internship Project Report	**Completed**

Name of Student : Oza Krutarth Dipakbhai

Name of Guide : Dr. Rajankumar Ghanshyambhai Patel

Signature of Student :______________________

*Signature of Guide : ______________________

Disclaimer :
This is a computer generated copy and does not indicate that your data has been evaluated. This is the receipt that GTU has received a copy of the data that you have uploaded and submitted as your project work.

*Guide has to sign the certificate, Only if all above activities has been Completed.

291758

Gandhinagar Institute of Technology

Moti Bhoyan Road, Gandhinagar ,Gujarat

(Affiliated with GTU)

COMPUTER ENGINEERING DEPARTMENT

CERTIFICATE

This is to certify that the work of Project/Internship entitled **"APS Fault Detection"** has been carried out by ***Oza Krutarth(190120107077)*** under my guidance in partial fulfilment for the degree of Bachelor of Engineering in ***Computer Engineering, 8th*** Semester in the ***Gandhinagar Institute of Technology***, Moti-Bhoyan, Gandhinagar, Gujarat, during the academic year 2022-2023 and his work is satisfactory. This student has successfully completed all the activity under my guidance related to Project/Internship for 8th semester.

Internal Guide
Dr Rajan Patel

Head of the Department
Prof.(Dr.) Nimisha Patel

Acknowledgement

I have made efforts in this Internship/Project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extendmy sincere thanks to all of them.

I am highly indebted to **Dr Rajan Patel** for their guidance and constant supervision as well as for providing necessary information regarding the Internship. I take this opportunity to thank all my friends and colleagues who started me out on the topic and provided extremely useful review feedback and for their all-time support and help in each and every aspect of the course of my project preparation. I am grateful to my college Gandhinagar Institute of Technology, for providing me all required resources and good working environment. I would like to express my gratitude towards Head of Department, **Prof.(Dr.) Nimisha Patel** and Director Sir for their kind co-operation and encouragement which help me in this
Project.

Thank You

Oza Krutarth

291758

Gandhinagar Institute of Technology

Moti Bhoyan Road, Gandhinagar ,Gujarat

(Affiliated with GTU)

DECLARATION

We hereby declare that the Project report submitted along with the Project entitled APS Fault Detection submitted in partial fulfillment for the degree of Bachelor of Engineering in Computer Engineering to Gujarat Technological University, Ahmedabad, is a bonafide record of original project work carried out by me Gandhinagar Institute of Technology under the supervision of Dr Rajan Patel and that no part of this report has been directly copied from any students' reports or taken from any other source, without providing due reference.

Name of the Student **Sign of Student**

Krutarth Oza ________________

Abstract

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 break 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.

In this project, the system in focus is the Air Pressure System(APS) which generates pressurized air that are utilized in various functions in a truck, such as braking and gear changes. The datasets positive class corresponds to component failures for a specific component of the APS system. The negative class corresponds to trucks with failures for components not related to the APS System.

The problem is to reduce the cost due to unnecessary repairs. So it is required to minimize the false predictions.

Contents

CHAPTER 1

INTRODUCTION

1.1 Project Detail

The dataset consists of data collected from heavy scania trucks in everyday usage. The system Focus is APS which generates pressurized air that is utilized in various functions in a truck, Such as braking and gear changes. The datasets positive class corresponds to component failures for a specific component of the APS system. The negative class corresponds to trucks with failures for components not related to the APS System.

1.2 Project Profile

The Training set contains 60000 examples in total in which 59000 belong to negative class and 1000 belong to positive class. The test set contains 16000 examples. There are 171 attributes per record. The dataset used for this project was imported from UCL ML Repository.

1.3 Project Definition

Our goal in this project is to predict if the truck needs to be serviced or not and minimize the cost associated with:

1) Unnecessary Checks done by mechanic which means negative labeled point is classified as positive.
2) Missing a faulty truck which may cause a breakdown. If positive labeled point is classified as negative.

1.4 Purpose

The purpose of this project is to help companies save finances on maintaining the trucks. With help of this project companies can identify whether the damage done to truck was due to failure in one of the APS components or was it any other machine. It can also be known if mechanics are performing maintenance duties on truck that actually need checks or are they wasting time on already good trucks.

1.5 Scope of Project

The scope of this project is to help multiple companies save money on maintenance of their heavy duty vehicles. We are using dataset from UCL ML repository. This dataset will help us solve this problem. With this project we can predict whether the truck needed maintenance because of failure in APS Components or was it any other machine. We will be using many complex libraries and Machine Learning algorithms to provide most accurate prediction.

1.6 Objective

The primary objective of this project is to provide the most accurate prediction from given dataset to help companies minimize their expenditure on their heavy duty vehicles by using data.

1.7 Literature Review

The APS Component is used in functionalities like braking and gear changes. It is an important part of heavy duty vehicles which need to be properly maintained so to make sure we will use an open source dataset to make an accurate prediction whether it was actually fault of one of APS components or we are wasting money on maintain trucks. This dataset has over 70000 data points which helps us accurately predict our solution which will help the companies using trucks.

CHAPTER 2

ABOUT THE SYSTEM

2.1 System Requirement Specification

Processor	:	(i3) Intel Pentium or more
Ram	:	4 GB
Hard disk	:	512 GB hard diskrecommended

Software Requirements:

Operating system	:	Windows/MacOS

2.2 Gantt chart

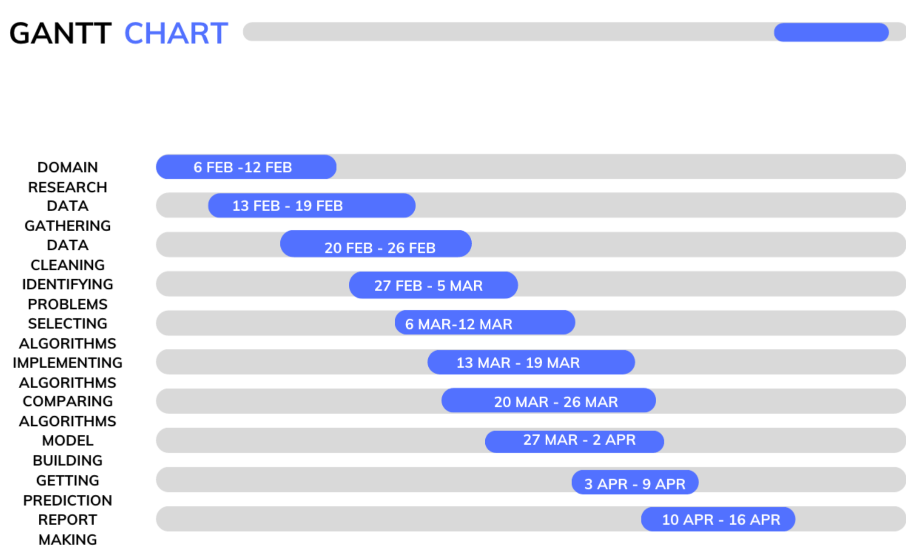


Fig 2.2 Gantt chart

2.3 Project Planning

• For this Project we are following Machine Learning Project planning which consists
Of the following phases in its development

- Phase 1: Requirement gathering andanalysis
- Phase 2: Feasibility study
- Phase 3: Dataset Preparation
- Phase 4: Data splitting
- Phase 5: Modeling
- Phase 6: Model Deployment
- Phase 7: Maintenance

CHAPTER 3

ANALYSIS OF THE SYSTEM

3.1 Use Case Diagram

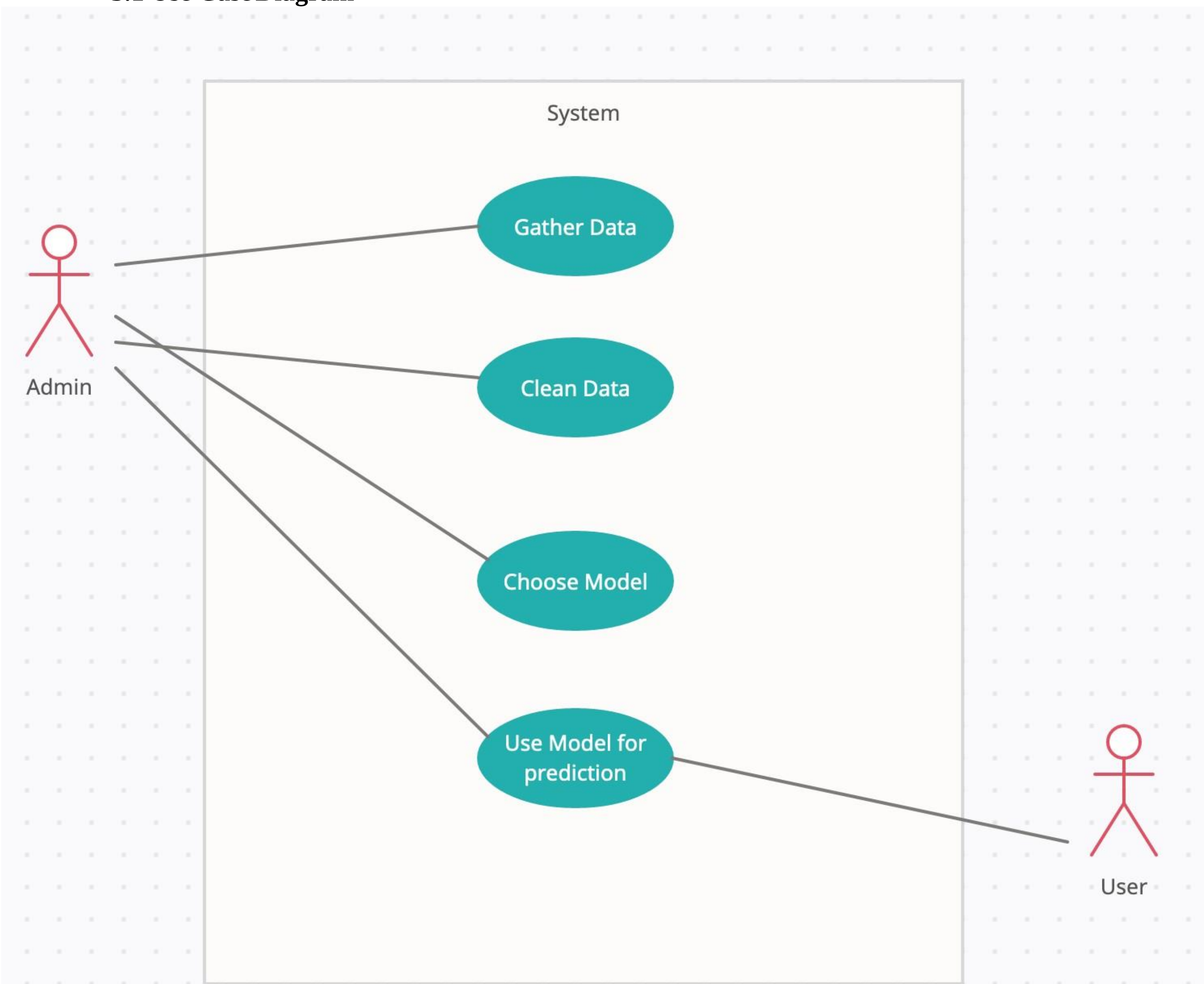


Fig 3.1 Use case Diagram

3.2 Sequence Diagram

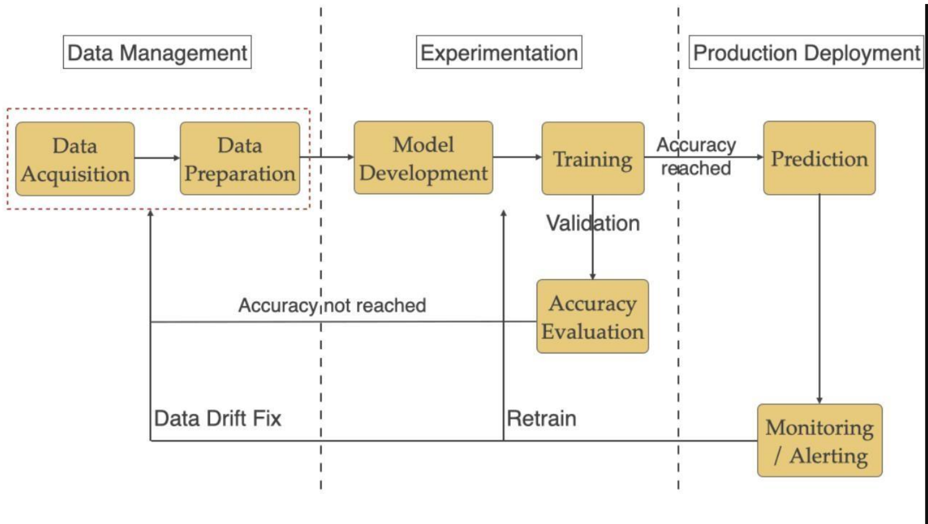


Fig 3.2 Sequence Diagram

3.3 Activity Diagram

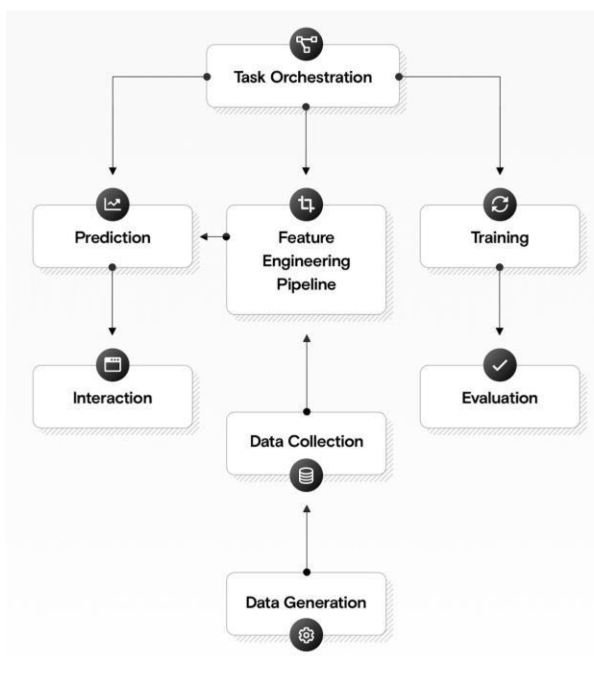


Fig 3.3 Activity Diagram

3.4 Data Flow Diagram

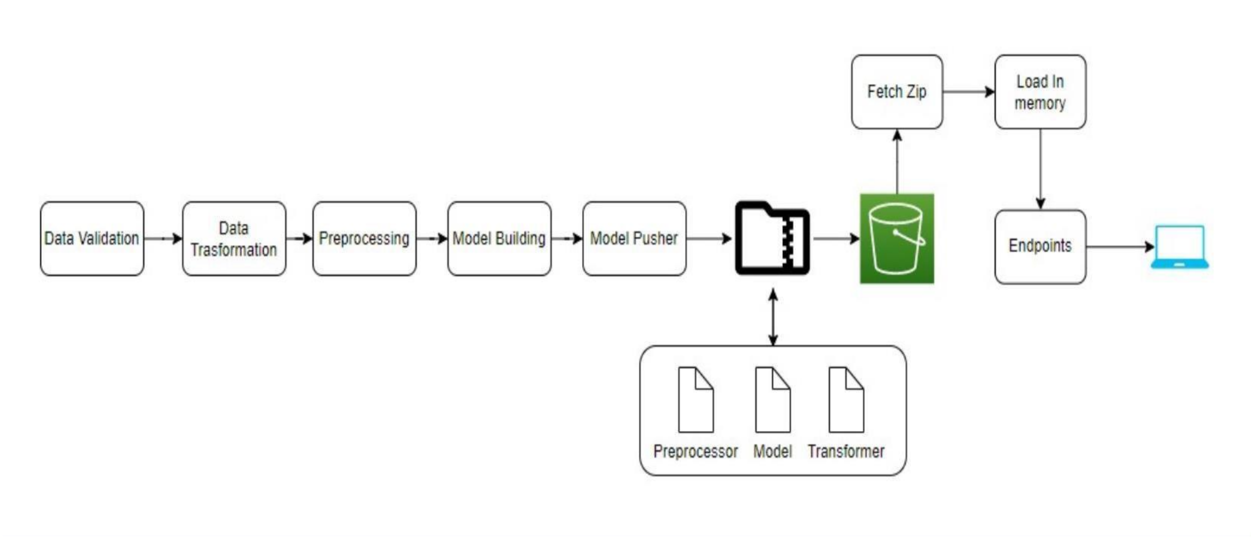


Fig 3.4 Data Flow Diagram

3.5 E R Diagram

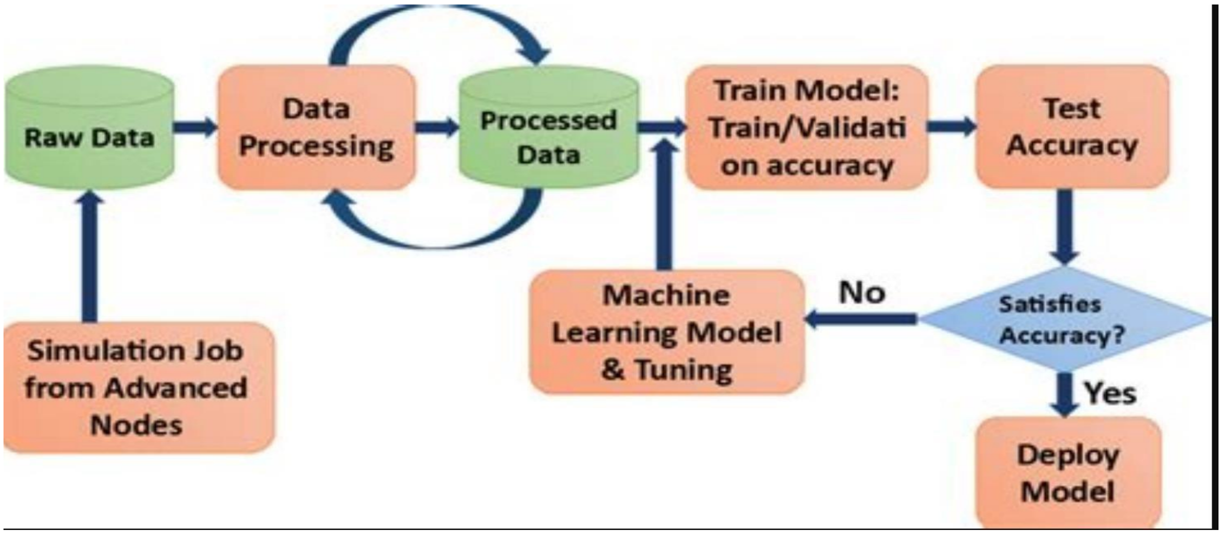


Fig 3.5 E R Diagram

3.6 Class Diagram

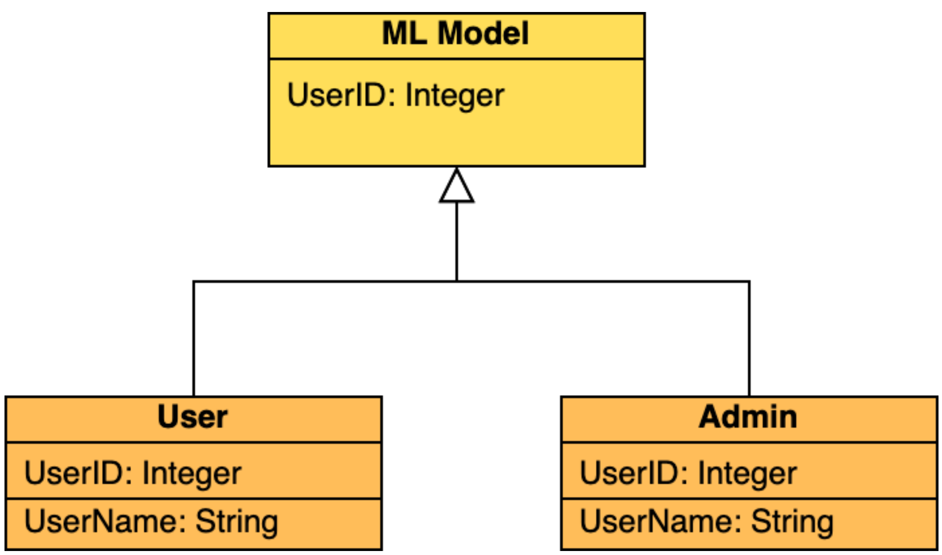


Fig 3.6 Class Diagram

CHAPTER 4

DESIGN

4.1 System Flow Diagram

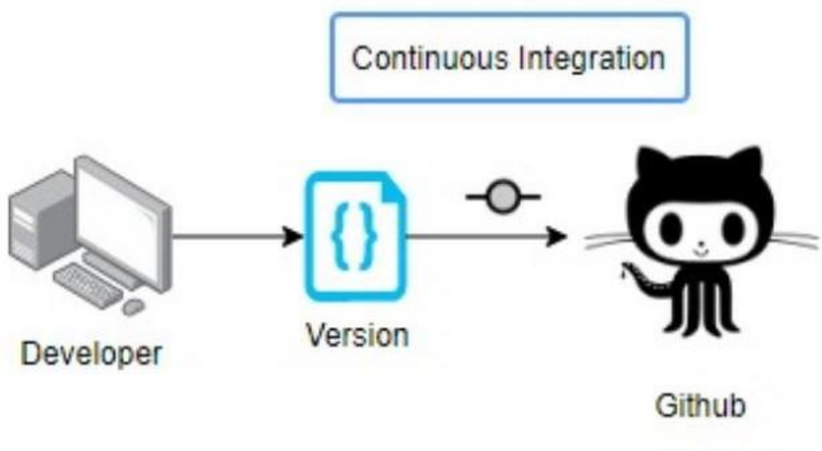


Fig4.1.1 SYSTEM WITH LEVEL ZERO

4.2 Data Dictionary

- The attribute names of the data have been anonymized for proprietary reasons. It consists of both single numerical counters and histograms consisting of bins with different conditions.
- Typically the histograms have open-ended conditions at each end. For example if we measuring the ambient temperature "T" then the histogram could be defined with 4 bins where: bin 1 collect values for temperature T < -20 bin collect values for temperature T >= -20 and T < 0 bin 3 collect values for temperature T >= 0 and T < 20 bin 4 collect values for temperature T > 20

- Cost 1 = 10 and Cost 2 = 500

  The total cost of a prediction model 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 unnessecary 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.`

4.3 Relationship of the Table

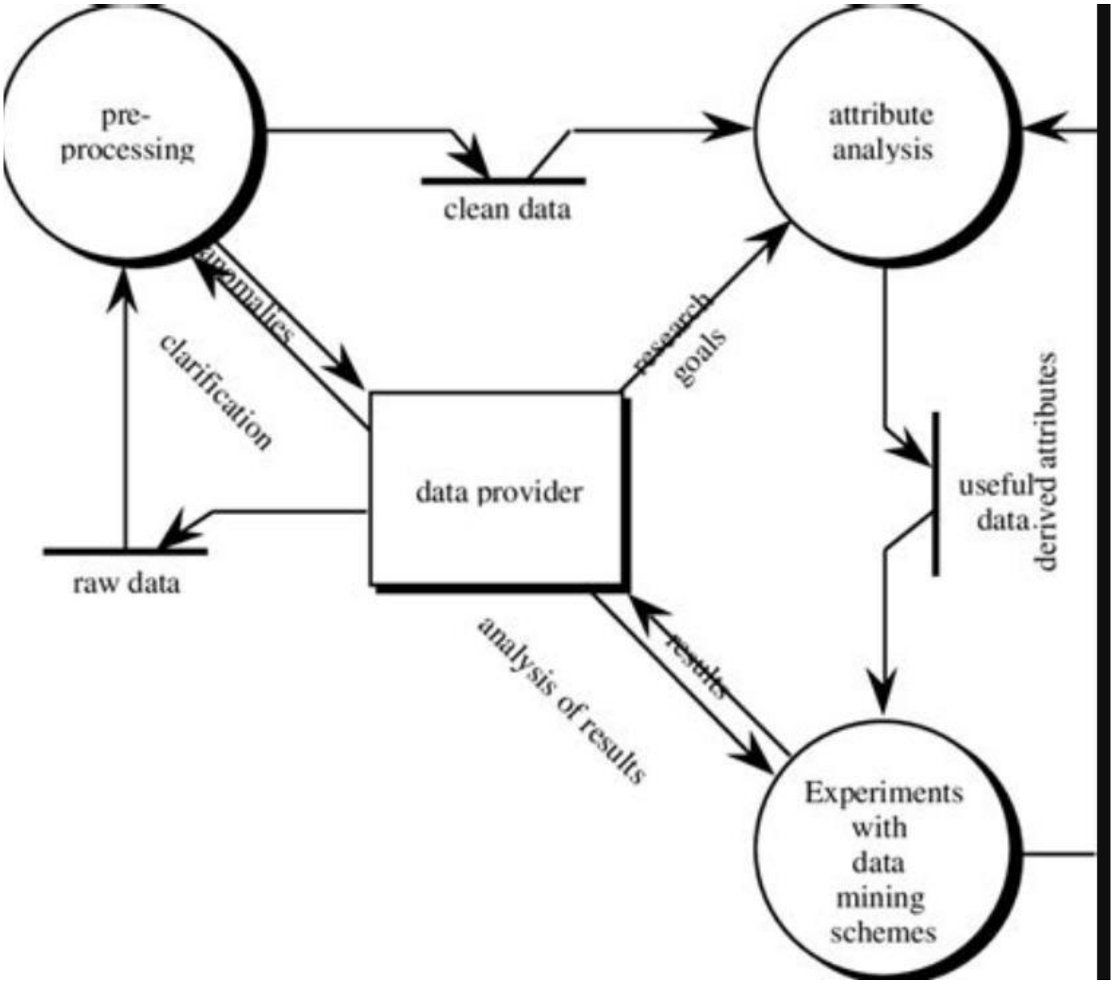


Fig 4.3 Relationship of the Table

4.4 User Interface

Sensor Component Failure Prediction

]:

1) Problem statement.

Data: Sensor Data

Problem statement :

- The system in focus is the Air Pressure system (APS) which generates pressurized air that are utilized in various functions in a truck, such as braking and gear changes. The datasets positive class corresponds to component failures for a specific component of the APS system. The negative class corresponds to trucks with failures for components not related to the APS system.
- The problem is to reduce the cost due to unnecessary repairs. So it is required to minimize the false predictions.

True class	Positive	Negative
Predicted class		
Positive	-	cost_1
Negative	cost_2	

Cost 1 = 10 and Cost 2 = 500

- The total cost of a prediction model 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 unnessecary 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.`
- From the above problem statement we could observe that, we have to reduce false positives and false negatives. More importantly we have to **reduce false negatives, since cost incurred due to false negative is 50 times higher than the false positives.**

Fig.4.4 Jupyter Doc

CHAPTER 5

IMPLEMENTATION

5.1 Implementation Environment (tools & Technology)

Programming Languages: Python.

SOFTWARE REQUIREMENTS:

Operating system	:	Windows Family/MacOS

HARDWARE SPECIFICATIONS

Processor	:	(i3) Intel Pentium or more
Ram	:	4 GB
Hard disk	:	512 GB hard disk recommended

CHAPTER 6

TESTING

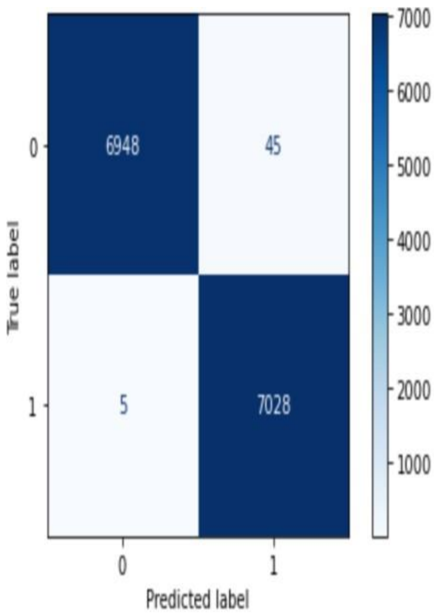
6.1 Testing Plan

Testing phase is a very important for a successful system. In this phase before implementing the new system into operation, for eliminating bugs a test run of the system is done. After completing codes for the whole programs of the system, a test plan should be developed and run one given set of test data.

Using the test data subsequent test run are carried out:
We split the whole dataset into two sections where one section will be used only for training Data and another will be used for training data. This will not affect the accuracy of the model But it will ensure that model is performing at its peak efficiency. Here we will implement multiple models and algorithms to make sure we get the most accurate prediction. To eliminate the null values we will also use multiple functions and choose the best one.

6.2 Test results & analysis

After completing multiple tests and implementing multiple models we found which method and model will give us best prediction for our problem statement.

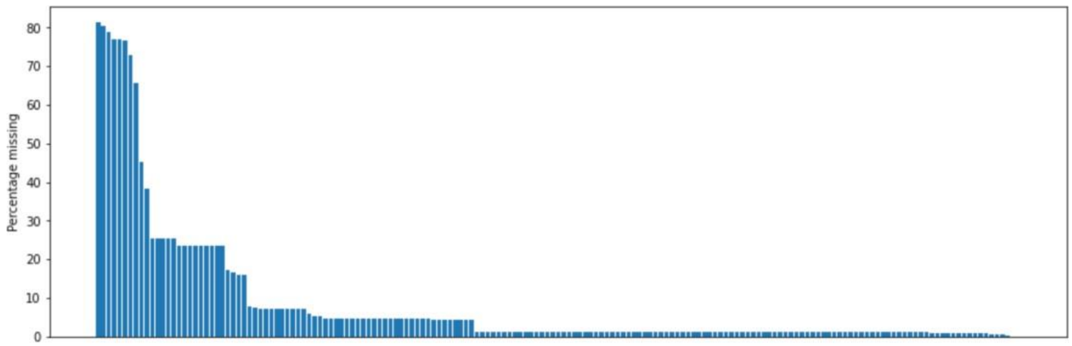


The best Model is XGBoost Classifier with 99.6% accuracy and cost of 2950

CHAPTER 7

IMPLEMENTATION

7.1 Prototype with Results / Screenshot



Dropping Columns which has more than 70% of missing values.

Fig 7.1 Null Data

Positive: 1000, Negative: 35188

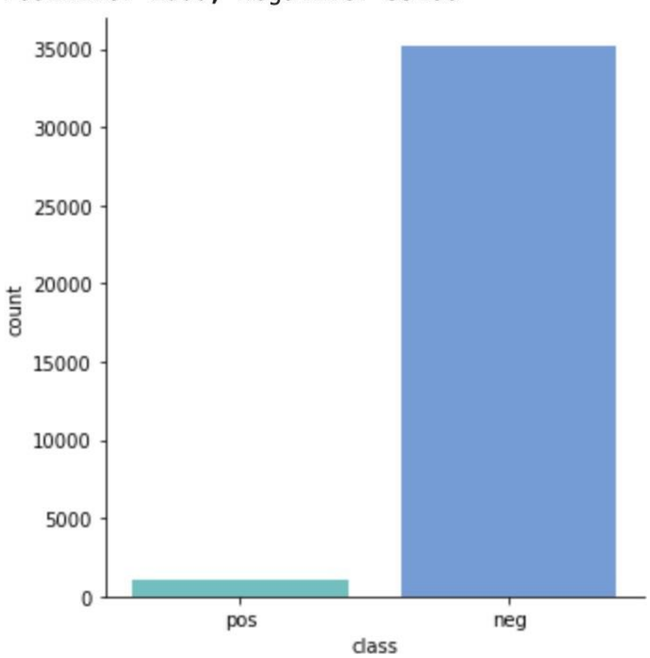


Fig 7.2 Classification

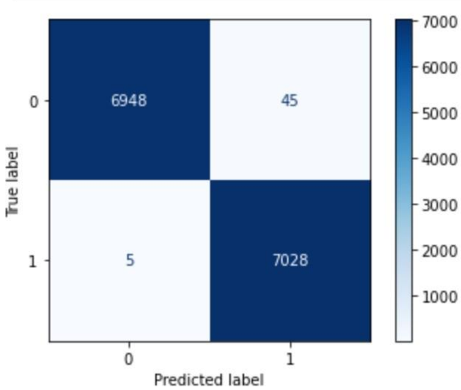
```python
import pandas as pd
import seaborn as sns
import numpy as np
from statistics import mean
import matplotlib.pyplot as plt
import warnings
from sklearn.preprocessing import PowerTransformer
import numpy as np
from sklearn.preprocessing import LabelEncoder
from sklearn.pipeline import Pipeline
from sklearn.utils import resample

from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier, GradientBoostingClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report,ConfusionMatrixDisplay, \
                            precision_score, recall_score, f1_score, roc_auc_score,roc_curve,confusion_matrix


from sklearn import metrics
from sklearn.model_selection import  train_test_split, RepeatedStratifiedKFold, cross_val_score
from sklearn.preprocessing import OneHotEncoder, MinMaxScaler
from sklearn.compose import ColumnTransformer
from sklearn.impute import SimpleImputer, KNNImputer
from xgboost import XGBClassifier
from sklearn.preprocessing import StandardScaler, MinMaxScaler,RobustScaler
from sklearn.compose import ColumnTransformer
from catboost import CatBoostClassifier

warnings.filterwarnings("ignore")
%matplotlib inline
```

Fig 7.3 Required Libraries



The best Model is XGBoost Classifier with 99.6% accuracy and cost of 2950

Fig 7.4 Confusion Matrix

CHAPTER 8

CONCLUSION AND DISCUSSION

8.1 Conclusion

In conclusion after training multiple models with multiple values we found that XG Boost
Will give the best prediction for our requirement . This will help multiple companies to minimize maintenance cost for their heavy duty vehicles.

8.2 Summary of Project Work

The main purpose to develop this model has been achieved by getting the most accurate prediction. We have used languages like python and implemented multiple libraries and models to achieve this prediction.

8.3 Problem Encountered and Possible Solutions

- Need to Handle many Null values in almost all columns
- No low-latency requirement.
- Interpretability is not important.
- misclassification leads the unecessary repair costs.

8.4 Limitation and Future Work

The limitations of ML model are:

- The target classes are highly imbalanced
- Class imbalance is a scenario that arises when we have unequal distribution of class in a dataset i.e. the no. of data points in the negative class (majority class) very large compared to that of the positive class (minority class)
- If the imbalanced data is not treated beforehand, then this will degrade the performance of the classifier model.
- Hence we should handle imbalanced data with certain methods.

The future work of our ML model:

The APS Fault detection model should be published on any deployment services like ASW EC2 where it can be updated in real time and can be viewed by users in a more convenient manner

REFERENCES

https://www.kaggle.com/datasets/uciml/aps-failure-at-scania-trucks-data-set