High Level Design (HLD)

Extended Warranty Division of Automobile Parts.

Revision Number: 1.0
Last date of revision: 02/06/2019

Our Project Team

Document version control

Date Issued	Version	Description	Author
5th oct 2019	1.0	First Version	Full Team

Contents

Abstract:

Warrenty devision project is machine learning project in which we are working on solurtion for the detection of fraud claims and genuine claims of warranty on parts of automobile. Machine Learning architecture will be implemented for solution. As well as finding out inferences required by industry based on data gathered. Inferences will be used for decision making authority.

1. Introduction:

Why Warranty Analysis?

After-sales, automobiles get post-sale services from dealers. A warranty analysis is mainly based on the data collected from those services, claims over a certain period.

The model analyses the performance of the dealers by considering the customers feedback, churn rate (ie customers moving out or stop doing business) of that location, preference to the other locations by the customers, change of base location etc.

The model helps the company to take some regulatory decisions based the performance of dealer. It also helps the company to decide, whether to introduce new schemes, offers etc. to the affected location or to change the dealer.

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
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - 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.

1.2 Definitions

Term	Description
SMPS	Switch Mode Power Supply
Database	Collection of all the information monitored by this system
IDE	Integrated Development Environment
AWS	Amazon Web Services

2 . General Description

1.3 Product Perspective

Currently, if the vehicle is secured by extended warranty assurance, then after the one year from the sale i.e. after the base warranty period, if any failure happens then the affected part/parts is/are either repaired or replaced under the warranty terms of the company and all the expenses will be done by the company. To reduce fraud claims and predict the product lifespan this prediction and classification model will become more efficient industrial practice.

2.2 Tools used

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

- PyCharm is used as IDE.
- For visualization of the plots, Matplotlib, Seaborn are used.

- AWS is used for deployment of the model.
- Tableau/Power BI is used for dashboard creation.
- MySQL/MongoDB is used to retrieve, insert, delete, and update the database.
- Front end development is done using HTML/CSS.
- Python Django is used for backend development.
- GitHub is used as version control system.

2.3 Constraints

- Typical challenges would be the accuracy and timeliness of data, the length of historical data, the number of parts to be modeled, and model selection based on accuracy on testing data.
- Regular data generation in bulk amount from multiple resources so that retraining can be done with good amount of prediction accuracy.

2.4 Assumptions

- The main objective of the project is to predict the onset of the PCB's for new cases based on the information in the Image given datasets by using Machine Learning and Deep Learning techniques. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

2 Design Details

3.1 Functional Architecture

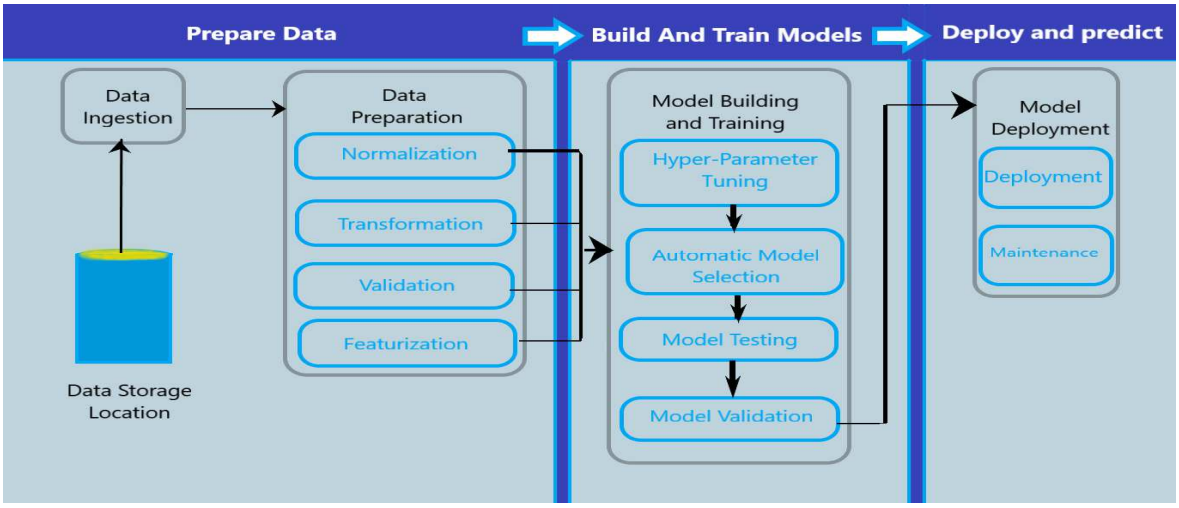

Figure 3.1

3.2 Dataset Used For Training :

- A warranty analysis is the analysis of **time-to-event/failure data**. In our example, the individual part is followed from the Automobile sold time to its failure.
- As in typical model building, we split the data into train and test datasets. With the training data, we first estimate the parameters of the distribution, and then using test data, we see if the model fitted works well on that data or not.
- We get a large amount of Raw data from the industry for training a machine learneing to binary classify as well as for other inference purposes. The records of the warranty of parts are taken .Data can into different formats like CSV files, XML etc which can be converted into required formats.

- By creating rich datasets from various sources provided by manufacturing company, parts suppliers, vendors and major outlets are combmined so we can trained our models, to work exceptionally well on such raw data, allowing our API to become part of a **comprehensive data processing pipeline**

3.3 Applications:

Industrial Application:

1. Monthly report of top Dealers and parts that has been failed:

   We need to generate the list of top ten dealers that are processing the extended warranty of parts

   The list of top ten dealers helps the company to analyze and understand the performances of dealers.
   Monthly report of top ten failed parts helps the

2. company to understand following:
   a) Quality of the parts
   b) Pattern of failure
   c) Mostly reported failed part in the quarter/year

   Data visualization libraries like **seaborn and matplotlib** are used to visualize the data.

3. Predict the parts that may fail according to the location, mileage and age:

In this case, by analyzing the historical data, we need to predict the parts that can be expected to fail according to the age, km driven and location of the vehicle.

This data helps the company to understand the manufacturing requirements of the part i.e. which part is to be manufactured in what quantity so as to avoid unavailability of the parts when claim is raised for that part.

4. To determine whether the defective part is to be replaced or repaired:

   The model was developed to decide whether the failed part has to be repaired or replaced with the new one.

3.5 Workflow :

Part I:

1. Finding out Top dealers and top failed part. Data analysis and visualizations.

Part II:

Warranty division project to predict the proposed claims are either fraud or genuine.

1. Data augmentation is used to increase the variety of data in order to avoid overfitting issue.

2. Random forest algorithm is used to train the prepared dataset. Pre-trained model.

3. Using the testing dataset, the accuracy of the model is evaluated.

Several parameters are configured prior to the training and it affects the performance of the output model. For this project, both the training data size size70% and Testing data size are set to 30%.

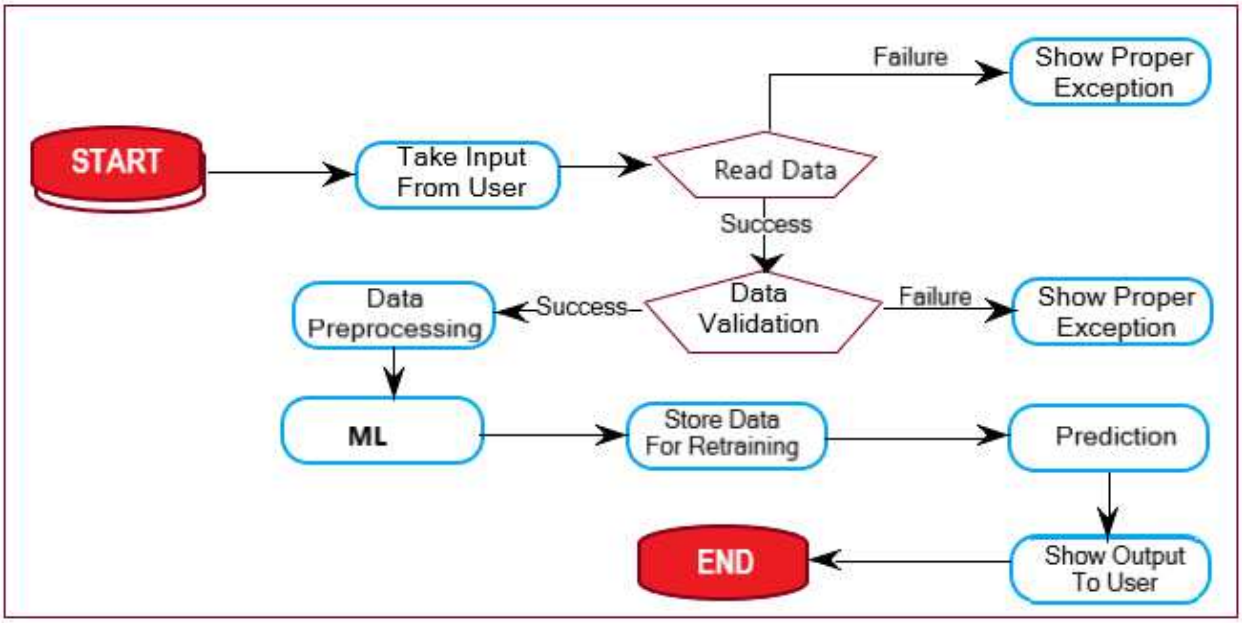

Figure 3.5

3.6 Web Application Architecture

The user interface is a very simple plain layout with little graphics. It will display information very clearly for the user and will primarily output information to the user through HTML pages. Also, all the details for the user input will be provided.

API will be on Flask Framework. Deployment of the project will be on heroku platform

3.7 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.

3.8 Help

The 'Help' option is provided in web application for guiding users regarding maximum range of valid inputs required for predicting a particular disease.

3.9 Performance

To understand whether the claim is a genuine claim or a fraudulent claim based on different independent variables. Warranty division claims detection system is used for predicting the onset, it should be as accurate as possible. So that it will not mislead the user. Also, model retraining is very important to improve the performance.

3.10 Reusability

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

3.11 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.

3.12 Resource utilization :

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

3.13 Deployment

Deployment of machine learning models, or simply, putting models into production, means making your models available to other systems within the organization or the web, so that they can receive data and return their predictions.

we will be deploying our project in either of the cloud platform.

4.1 KPIs (Key Performance Indicators)

Key indicators delivering accurate prediction of parts failure and detection of genuine warranty claims.

1. Time and workload reduction using this model.
2. Comparison of accuracy of model prediction and Engineer's prediction.
3. Number of times data provided it should detect fraud and genuine claims and other predictions in small interval of time.

5.Conclusion:

This project is playing very crucial role into the cost cutting on fake claims, and imrpovement in the quality of product manufacturing.

Approach to counter difficulties occurred in human's manual inspection of data in claim settlement which will has a lot of human errors when the quantity of data inspected increases in bulk.

We can counter it with, Cleaned data and arranging relevant feature selection for required target feature so that the prediction becomes accuracy rich. So it will affect the industry quantitatively as well as qualitatively.