VEHICLE HEALTH ANALYSIS AND SPARE PARTS E-COMMERCE

A PROJECT REPORT

Submitted by

Avinash reddy Vasipalli [Reg No: RA1911027010007]
Yuvraj Singh Chauhan [Reg No: RA1911027010058]

Under the Guidance of

Dr. T. KARTHICK
(Assistant Professor, Department of Data Science and Business Systems)

In partial fulfillment of the Requirements for the Degreeof

BACHELOR OF TECHNOLOGY

COMPUTER SCIENCE ENGINEERING with specialization in Big Data Analysis

DEPARTMENT OF DATA SCIENCE AND BUSINESS SYSTEMS

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

NOVEMBER 2022
SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR-603203

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that this project report titled "VEHICLE HEALTH ANALYSIS AND INSURANCE ANALYSIS" is the bonafide work of "Avinash reddy Vasipalli [Reg No: RA1911027010007] and Yuvraj Singh Chauhan [Reg No: RA1911027010058]" who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation based on which a degree or award was conferred on an earlier occasion for this or any other candidate.

Dr. T. Karthick
GUIDE
Assistant Professor
Dept. of DSBS

Dr. M.Lakshmi
HEAD OF THE DEPARTMENT
Dept. of DSBS

Signature of Internal Examiner

Signature of External Examiner

ABSTRACT

Vehicles need constant servicing and repairs after a certain interval of time and if not repaired at fixed intervals it can lead to accidents. The objective ofthe project is to give an idea to the owner about the estimated insurance claim which he/she can get in case of certain faults and issues whether the insurance claim exists for the vehicle or not. There are many automobile health monitoring systems which are used to detect and diagnose faults and issues like On-Board Diagnostic (OBD) and Engine Control Units (ECU) which tells the owner where the fault has occurred. In this project, we have used Support Vector Regression, TFIDF, LSTM and a dataset containing the car prices to train a machine learning algorithm that can predict the estimated insurance claim based on the faults which the owner has entered.

Development of automotive industry started in in early 21st century leading to modern solutions to many traditional problems. This development gave rise to data sets that are multidimensional and contain data of automotives starting from a new model to a model dated in 1980's as well. Even with this development, traditional methods are in use to finding solutions to these automotive leading to a lot of wastage of resources like money and time. With increase in automotive usage, better and sophisticated system is required in order to overcome these limitations and provide effective solutions. In this paper, we focus on using Machine algorithms to analysis datasets in predicting the required service in most effective manner with most basic data available. Also, this paper has analysed data to predict insurance premium and its second-hand selling price using machine learning algorithms like SVM, Random Forest and other regression methods. Addition of Ecommerce website to help purchase parts for automobiles has provided all around solution to customers.

Keywords: SVM, Random Forest, Regression, Insurance, Price, Machine Learning, E Commerce

Department of Data Science and Business System

SRM Institute of Science & Technology

Own Work Declaration Form*

Degree/ Course **: Btech/CSE-Big Data Analytics**

Student Name *: Avinash reddy Vasipalli,* Yuvraj singh Chauhan

Registration Number *: RA1911027010007,* RA1911027010058

Title of Work: VEHICLE HEALTH ANALYSIS AND SPARE PARTS E-COMMERCE

We hereby certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism**, as listed in the University Website, Regulations, and the Education Committee guidelines.

We confirm that all the work contained in this assessment is our own except where indicated, and that We have met the following conditions:

- Clearly references / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that I have received from others (e.g.fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook /University website

I understand that any false claim for this work will be penalized in accordance with the University policies and regulations.

DECLARATION:
I am aware of and understand the University's policy on Academic misconduct and plagiarism and I certify that this assessment is our own work, except were indicated by referring, and that I have followed the good academic practices noted above.
Avinash reddy Vasipalli (RA1911027010007) Yuvraj Singh Chauhan (RA1911027010058)

ACKNOWLEDGEMENT

We express our humble gratitude to **Dr C. Muthamizhchelvan**, Vice-Chancellor, SRMInstitute of Science and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to Dean-CET, SRM Institute of Science and Technology, **Dr T.V.Gopal**, for his invaluable support.

We wish to thank **Dr Revathi Venkataraman**, Professor & Chairperson, School of Computing, SRM Institute of Science and Technology, for her support throughout the project work.

We are incredibly grateful to our Head of the Department**, Dr M. Lakshmi** Professor, Department of Data Science and Business Systems, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our program coordinators **Dr.G.Vadivu,** Professor, Department of Data Science and Business Systems, SRM Institute of Science and Technology, for her inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor**, Dr T Kathick,** Assistant Professor, Data Science and Business Systems, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to my guide, **Dr T Kathick,** Assistant Professor, Data Science and Business Systems, SRM Institute of Science and Technology, for providing me with an opportunity to pursue my project under his mentorship. He provided me with the freedom and support to explore the research topics of my interest. His passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank the Data Science and Business Systems staff and students, SRM Institute of Science and Technology, for their help during our project. Finally, we would like to thank parents, family members, and friends for their unconditional love, constant support, and encouragement.

Avinash reddy Vasipalli

Yuvraj Singh Chauhan

TABLE OF CONTENTS

Figure Table

CHAPTER I: Introduction

Modern world highly depends on AI to solve problems that are challenging to human existence and to make human life easy and comfortable. With today's intelligence widely used in automotive industry have helped in solving problems related to production and sales. Present industry is what we call 4th industrial revolution in which physical and digital systems are working hand in hand to provide a better environment that produces data that can be transmuted easily. Maintenance is of many types in automotive industries but majorly exist 3 types: Preventive – when a Fault is occurred, Corrective – replacement of equipment, Predictive – analysis of current vehicle to predict what could be the issue

Automotives are very complex and consist of components like gearbox, brakes, and engines and these have sensors and actuators that connected with Engine Control Unit make it possible for it to run without any issue. There exist many theories and diagnostic tools that can communicate with ECU in which 2 best are UDS and OBD2. Due to increase in complexity, focus moved to data analysis where in wireless communication and android applications provided the most cost-effective way to diagnose automotives. For this, we propose a diagnostic using the most basic data that is available on cars and elements like second hand cost prediction and insurance premium prediction. This paper uses these data and machine learning algorithms and web application of E commers to help customers to buy genuine products. Major advantage of this paper is its all-round service to vehicle owners to ensure their time and money is saved.

1.1Background

Present automotive industry is developing at a rapid pace with more and more electric vehicles getting attention to go green. Traditional methods to work on maintenance and premium has been burden to people who has zero connection to modern world. In present day every household has either one vehicle that runs on fossil fuels but no proper system was presented to help in areas of repair, insurance, and genuine parts purchase. Even with many solutions by companies its still not reachable to customers as many does not trust or understand the work and importance of these systems. Machine learning is one of the major modules in

developing sustainable model. First step in this process comes the premium prediction with the available data on the vehicle. Second step is to predict the cost

of the vehicle when we re sale it. Last step is to provide a E commerce service to search and buy genuine parts from reputed manufactures. Additional proposal is to work on prediction of vehicle maintenance prediction by finding datasets on senser maintenance.

1.1.1 Literature Survey

1. *Mohammed Al-Zeyadi∗, Javier Andreu-Perez∗, Hani Hagras∗, Chris Royce†, Darren Smith†, Piotr Rzonsowski†, Ali Malik*

   "Deep Learning Towards Intelligent Vehicle Fault Diagnosis"

Recently, the rapid development of automotive industries has given rise to large multidimensional datasets both in the production sites and after-sale services. Fault diagnostic systems are one of the services that the automotive industries provide. As a consequence of the rapid development of cars features, traditional rule-based diagnostic systems became very limited. Therefore, more sophisticated AI approaches need to be investigated towards more efficient solutions. In this paper, we focus on 11tilizing deep learning so as to build a diagnostic system that is able to estimate the required services in an efficient and effective way. We propose a new model, called Deep Symptoms-Based Model Deep-SBM, as an approach to predict a wide range of faults by relying on the deep learning technique. The new proposed model is validated through a set of experiments in order to demonstrate how the underlying model runs and its impact on improving the overall performance metrics. We have applied the Deep-SBM on a real historical diagnostic data provided by Cognitran Ltd. The performance of the Deep-SBM was compared against the state-of-the-art approaches and better result has been reported in terms of accuracy, precision, recall, and F-Score. Based on the obtained results, some further directions are suggested in this context. The final goal is having fault prediction data collected online relying on IoT.

2. *Uferah Shafi, Asad Safi, Ahmad Raza Shahid, Sheikh Ziauddin,and Muhammad Qaiser Saleem*

   "Vehicle Remote Health Monitoring and Prognostic Maintenance System"

In many industries inclusive of automotive vehicle industry, predictive maintenance has become more important. It is hard to diagnose failure in advance in the vehicle industry because of the limited availability of sensors and some of the designing exertions. However, with the great development in automotive industry, it looks

feasible today to analyze sensor's data along with machine learning techniques for failure prediction. In this article, an approach is presented for fault prediction of four main subsystems of vehicle, fuel system, ignition system, exhaust system, and cooling system. Sensor is collected when vehicle is on the move, both in faulty condition (when any failure in specific system has occurred) and in normal condition. The data is transmitted to the server which analyzes the data. Interesting patterns are learned using four classifiers, Decision Tree, Support Vector Machine, K – Nearest Neighbor, and Random Forest. These patterns are later used to detect future failures in other vehicles which show the similar behavior. The approach is produced with the end goal of expanding vehicle up-time and was demonstrated on 70 vehicles of Toyota Corolla type. Accuracy comparison of all classifiers is performed based on Receiver Operating Characteristics (ROC) curves.

3. *Ioan virca, Dorel badea*
   "Study on the predictive maintenance of vehicles and its management using the specific "Keep the Machine Running" application"

Due to the advanced equipment for diagnosing the components and technical systems of which they are part, the advanced software techniques and applications for the execution, monitoring and maintenance of management, methods, the tendency to develop the type of predictive maintenance is now being generalized. The importance is given by the possibility of processing the determined results, by estimating the tendency of manifestation of the functional parameters and by establishing the moment of the technical interventions for maintaining the equipment/vehicle in the state of availability. The predictive or conditional maintenance, along with the systematic and predictive maintenance, as ways of planning and executing preventive maintenance, has the advantage that the costs of purchasing diagnostic equipment can be offset by the reduced repair costs, as theoretically the major equipment failure will be greatly reduced. The authors do a study on the possibilities of monitoring and driving automotive predictive maintenance using the Keep the Machine the software application Keep the Machine Running. The research method is to investigate the current level of maintenance and the facilities offered by the specific diagnosis and software application, and to integrate the results in order to achieve an efficient maintenance management

4. *Sebastian Baran, Przemys law Rola*
   "Prediction of motor insurance claims occurrence as an imbalanced machine learning problem"

The insurance industry, with its large datasets, is a natural place to use big data solutions. However, it must be stressed, that significant number of applications for machine learning in insurance industry, like fraud detection or claim prediction, deals with the problem of machine learning on an imbalanced data set. This is due to the fact that frauds or claims are rare events when compared with the entire population of drivers. The problem of imbalanced learning is often hard to overcome. Therefore, the main goal of this work is to present and apply various methods of dealing with an imbalanced dataset in the context of claim occurrence prediction in car insurance. In addition, the above techniques are used to compare the results of machine learning algorithms in the context of claim occurrence prediction in car insurance. Our study covers the following techniques: logistic-regression, decision tree, random forest, xgBoost, feed-forward network. The problem is the classification one.

5. *Thyago P. Carvalho, Fabrízzio A. A. M. N. Soares, Roberto Vita, Roberto da P. Francisco, João P. Basto, Symone G. S. Alcalá*

   "A Proposed Model to Predict Auto Insurance Claims Using Machine Learning Techniques"

The amount of data extracted from production processes has increased exponentially due to the proliferation of sensing technologies. When processed and analyzed, data can bring out valuable information and knowledge from manufacturing process, production system and equipment. In industries, equipment maintenance is an important key, and affects the operation time of equipment and its efficiency. Thus, equipment faults need to be identified and solved, avoiding shutdown in the production processes. Machine Learning (ML) methods have been emerged as a promising tool in Predictive Maintenance (PdM) applications to prevent failures in equipment that make up the production lines in the factory floor. However, the performance of PdM applications depends on the appropriate choice of the ML method. The aim of this paper is to present a systematic literature review of ML methods applied to PdM, showing which are being explored in this field and the performance of the current state-of-the-art ML techniques. This review focuses on two scientific databases and provides a useful foundation on the ML techniques, their main results, challenges and opportunities, as well as it supports new research works in the PdM field

1.2 Purpose

- We try implementing and improving the current working modules with updated machine learning model in future development and try working new technology stack.
- We try improving the accuracy of model by using and comparing it with

other major papers in various aspects to make use of new algorithms

- With use of machine learning, operational efficiency is of model where in the datasets working in different algorithm provided a better integration of data. This is due to the lack on awareness and zero datasets on sensers and present data integration has many unused data which can significantly improve the efficiency.
- This provided a better security and reduce in cost to customers wherein diagnosis, premium, price negotiation in traditional model includes fee and commission to agents providing these services.
- Labour cost such as diagnosis of a problem is avoided in this case and an automatic prediction of premium and price to used cars helped in avoiding agent fee and charges.
- Our aim of automating all this process will help companies in using fewer human resources and provide the best services to the customers.
- This software can be readily used by any personal and avoid any errors as it is user friendly in every aspect of the project.

1.3Scope of the project

Every project has its own value and scope in this modern world and we believe these below listed features are best possible way to describe them:

- Feedback and complaint addressing model to avoid human resource and paper works.
- Accuracy and efficiency increase of feedback model by adding all the running data to the dataset to develop it continuously.
- User friendly and error free

1.4Significance

- Emergency results and feedback
- Complaint addressing chatbot refining
- IOT usage in connectivity of data
- Publishing major problems through real time system alerts

CHAPTER II: Requirements Analysis and Study

2.1 Requirements Analysis

This phase is majorly focus on requirements that are needed to be fulfilled by the system proposed. It gives a detailed plan on how to make use of system to best usage it.

2.1.1 *Information Gathering*

To collect all the information related to the project we are required to specify the requirement and specification of project. We are required to provide functionalities of the project and gather information to understand problem statement and collect useful information. For this project, we collect necessary information in form of various attributes of vehicle like cost, model, price and mileage and many other information.

2.1.2 *Functional Requirement*

2.1.2.1 Dataset:

A dataset is collection of data and these commonly corresponds to contents of single database table, or single statistical data matrix. This matrix has every column of table represents a variable and each row corresponds to given number of datasets. This data set may comprise of various models and types of vehicles with variables like mileage, distance travelled and many others.

2.1.2.2 User:

In terms of user requirements for a e commerce site an account is required. Without user login ordering of parts are not possible but product searching and cart addition is possible. Also, for prediction of premium and price, data entry of at least one variable is required and visualization of variables against each other for entry is possible.

2.1.2.3 Web Module:

This module contains admin, login, and product ordering module.

Admin module – Person handling login credentials, orders details and product listing

Login module – Login of account with login id and password

Product ordering – Order placement and personal info

2.1.3 *Non-Functional Requirements*

2.1.3.1 *Optional Requirement*

- System ability to retrieve data at any time
- System speed on processing request must be high

2.1.3.2 *Software Attribute*

Reliability

It is the measure of ability of system to keep the system to work overtime. It is measure of mean time between failures, expected type systems. Delivery report of activities like going and order placed and

payment related notifications must be done. TCP and IP services provide the user a must better

environment to operate.

Availability

This takes care of safety and dependency for user and the resources that must be provided to him. For a ideal availability corrective maintenance downtown must be present.

Maintainability

Let us understand this with the help of factor that must be considered while modules are changed or added to the application

i. Classes and services of the project must be understood better in context of application framework to avoid any errors in future updates. This also must help in addition of new services with ease.

ii. Roles and access service must be compatible to take care of whole system hierarchy and department included. If there is no flexibility in design then it might become a problem in later steps and architecture might be an issue.

iii. Architecture and framework of the application are must check when a change is made to avoid long terms errors in the application.

2.1.3.3 *Essential Requirements*

- System reliability must be around 95%
- System backup must be done everyday

- Security of data in system must be done
- Special scheme availability must be updated real time
- Chatbot enability

2.2 Feasibility study

This study important in project design as this helps in development stage when in it decides if the project is feasible with the decided input and information, budget, and technologies.

2.2.1 Technical Feasibility

The presented system is made using Xampp, PHP, Big data, JavaScript, MYSQL database. Usage of Machine learning and deep learning in form of Lasso, CNN and SVM to train the module with dataset using various sources. Use of Decision tree to classify and give emergency feedback is under process.

2.2.2 Economic Feasibility

Cost analysis in project making is important factor as many projects might discontinue due to this reason. A depth analysis on project feasibility economically to save both time and money. This can be estimated when benefits outweigh cost making it economically feasible.

2.2.3 Operational Feasibility

Operation is to solve measure of solving problems with the proposed system. This helps in understanding the working of model and takes advantage of all opportunities to fulfill the requirements. The proposed model is a all stop for parts and data prediction of premium and price to the vehicles using the data available. Big data and ML are major player to get this project a success.

2.2.4 Behavioral Feasibility

It is the effort put in education, training the users using staff. Since this is user friendly it needs no effort to educate user on its functionality. This setup once set requires no intervention of user. The easy usage and one time education is major addition for feasibility.

CHAPTER III: System Design and Analysis

3.1 System Design

System design is a process of defining the systems architecture, module, interface, and data of the system. This design is useful from application system to development if product. The purpose of system design process is to provide sufficient detailed data about system and its elements.

3.1.1 Architecture Diagram

Architecture diagram is visual representation of elements to make a project. This helps in understanding the structure of project and its components. Architecture diagram helps in understanding the project in single glance, improves the collaboration to see how efficiency are the modules and boost transparency.

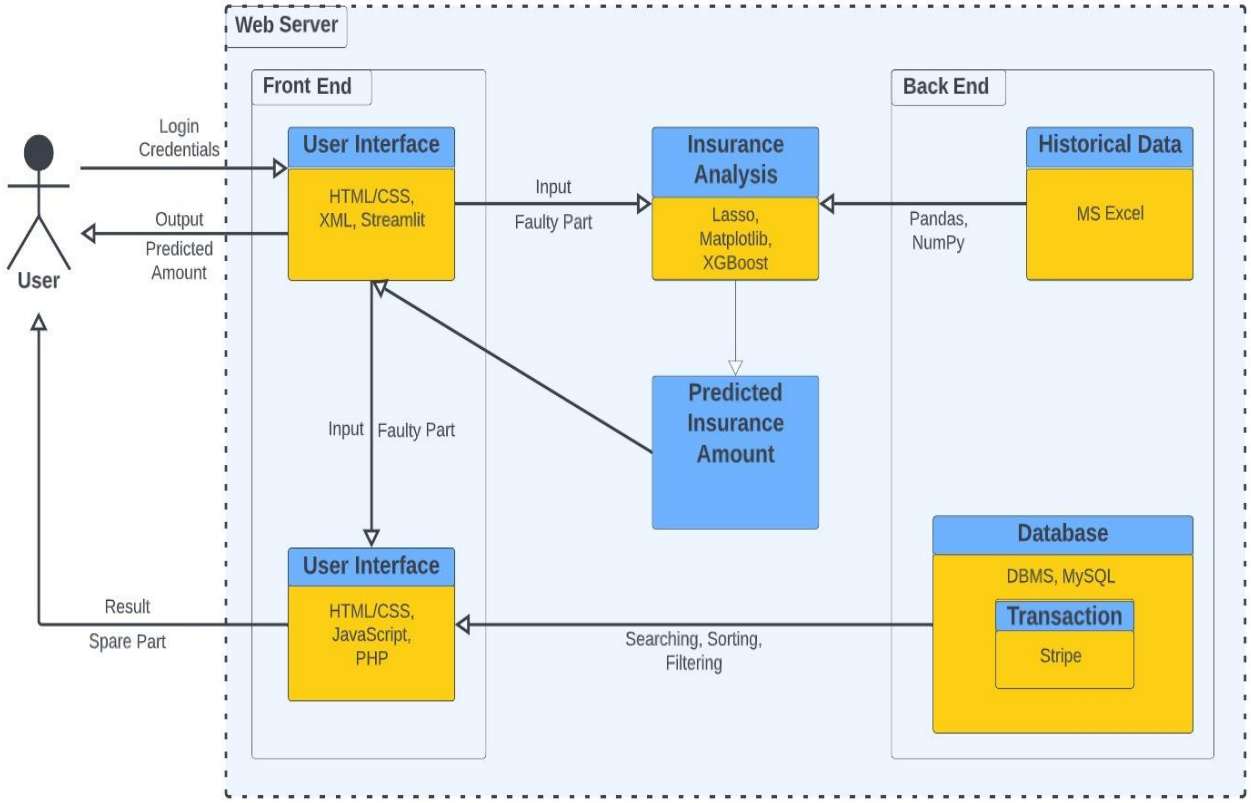


Figure 3.1.1 Architecture Diagram

3.1.2 E-R Diagram

Entity – relationship model is the description of interrelated things of interest in specific domain of knowledge. Most basic model of ER has entities and the

relationships between them. Entities are also known as instances of those entity types. These entities are graphically represented as boxes and connected with the help of line or relationship which describes the association and dependencies. ER diagram is conception and represents model of data in framework. Even though it helps in easy description it is not enough every detail but it helps in its own way for structural and unstructured data. ER diagram has 3 main components Entities (objects or concepts that have data about)

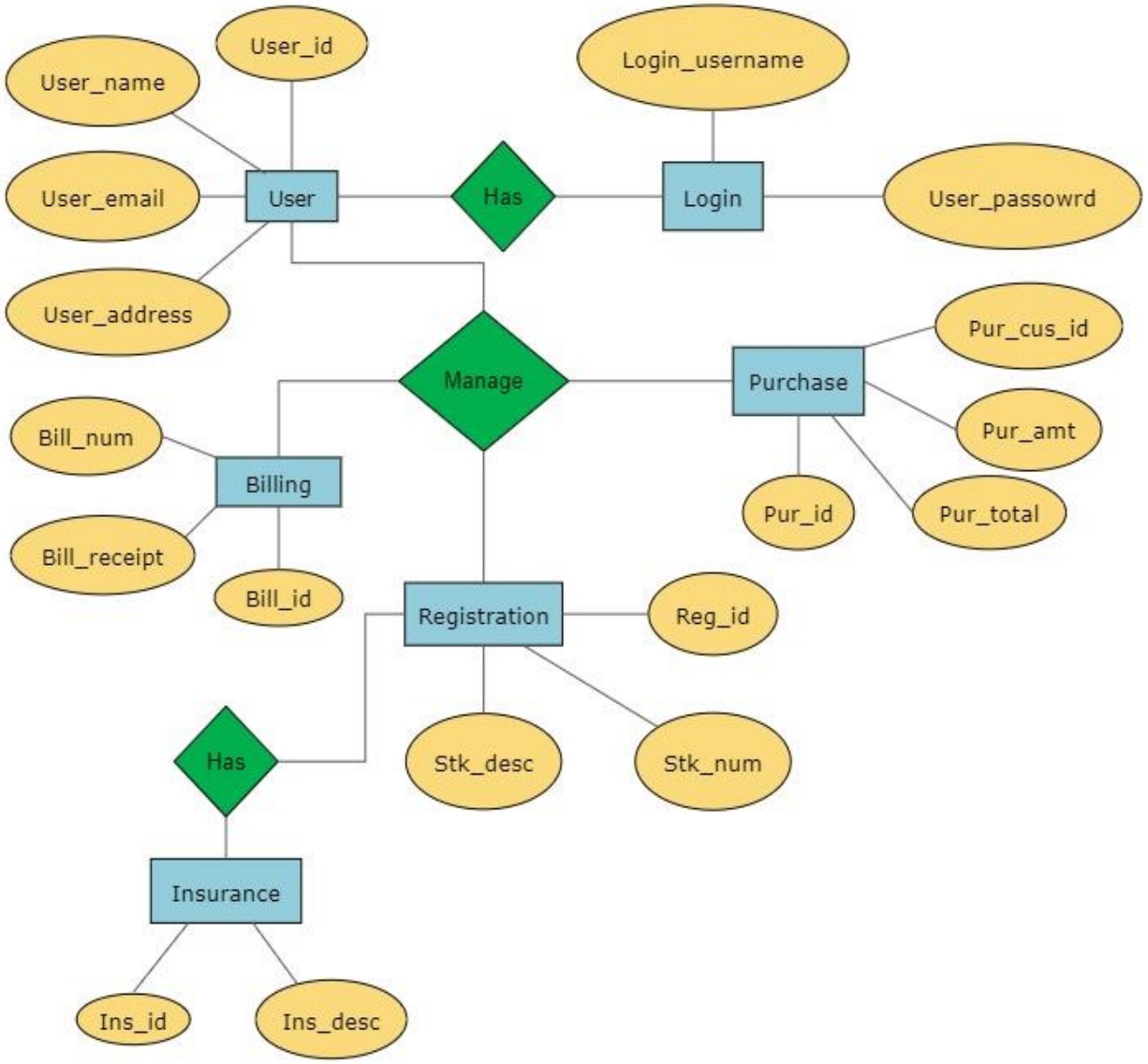


Figure 3.1.2 E-R Diagram

3.1.3 Use Case Diagram

Use case is the basic and simplest representation of the user interface with the system and helps in showing relationship between user and other uses that user can use. Use

case diagram can identify uses and make different use case diagram to accompany the existing structure. Use case are represented by either circle or ellipses and helps in analysis of identity, organize, and clarify system requirements. Use case does not involve complex relations as the moto is to keep it simple.

Figure 3.1.3 Use Case Diagram

3.1.4 Sequence Diagram

By the name of sequence, we can understand it is used in making interaction diagram and process to operate one another and its order. This has time involved making it the object interaction with time and depicts the objects and classes involved and the interaction between them to carry out scenario. This diagram has vertical line parallel to each other representing the different process or objects that exist simultaneously. The connection can be shown with the help of horizontal arrow between them in the order it is going.

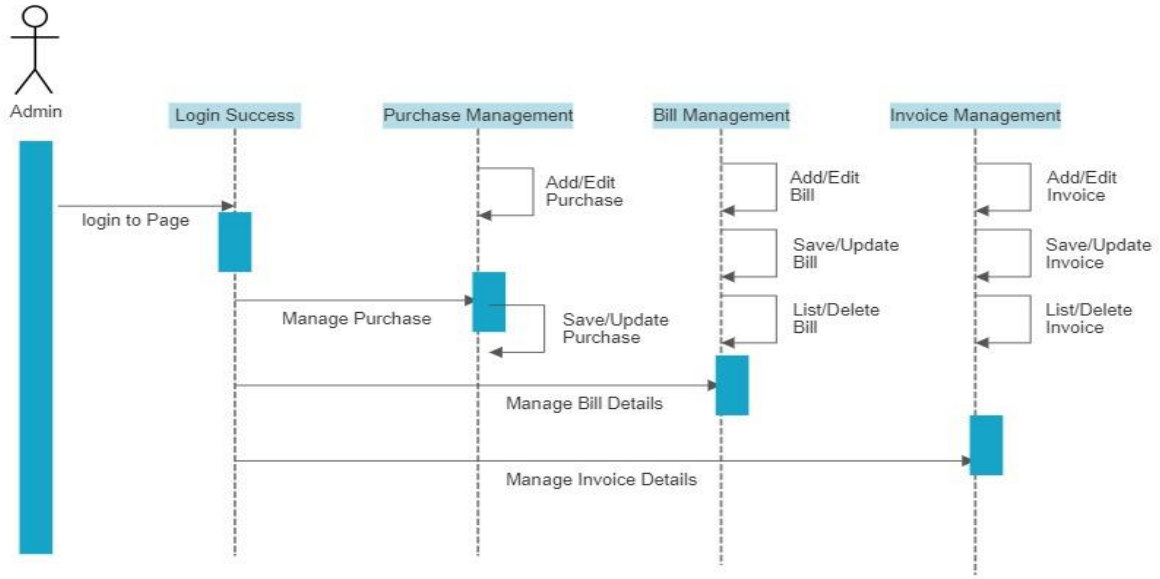


Figure 3.1.4 Sequence Diagram

3.1.5 Data Flow diagram

Data flow is the flow of information for any process or system. This used symbols like rectangle, circle and arrows and text information to show the data, inputs, outputs and storage points and the connections between.

Level 0:

This shows the simplest view know as context diagram which is the top level data flow. As the number suggests this has one process node and that node gives the total idea on the project.

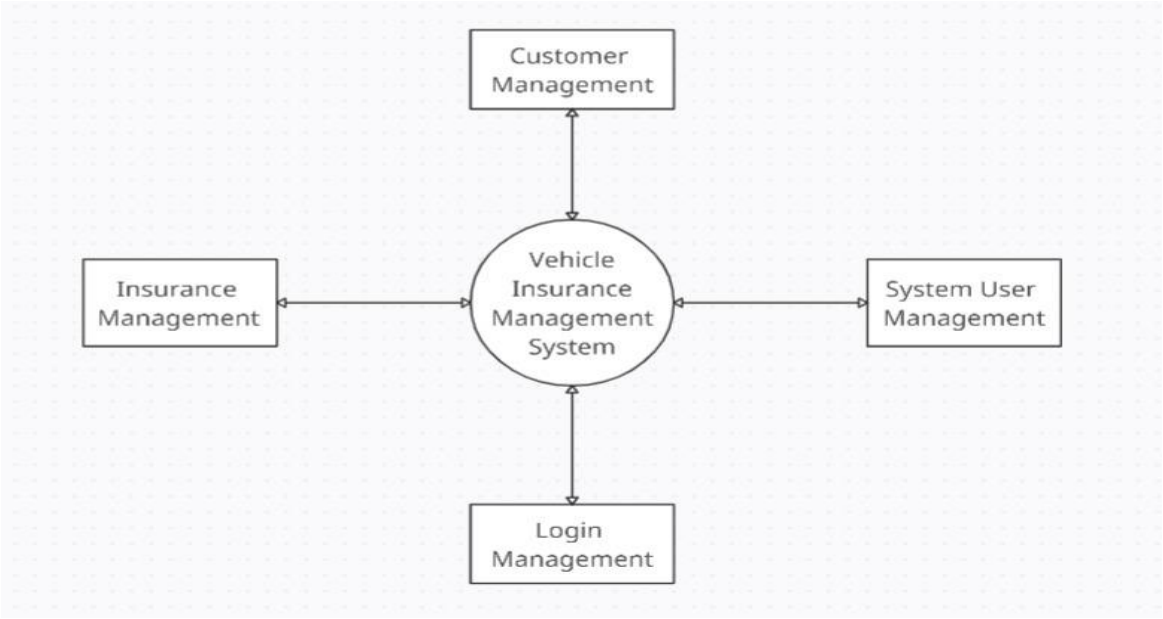


Figure 3.1.5 – A Level 0 of dataflow

Level 1:

Next stage creates dataflow by highlighting only the main functions of the project. This stage considers the different functions but will not go deep i.e., even if two functions have complex relation only an arrow will describe it nothing more. Since it is detail view their might exist multiple external entities and other possible data flows possible in the context diagram.

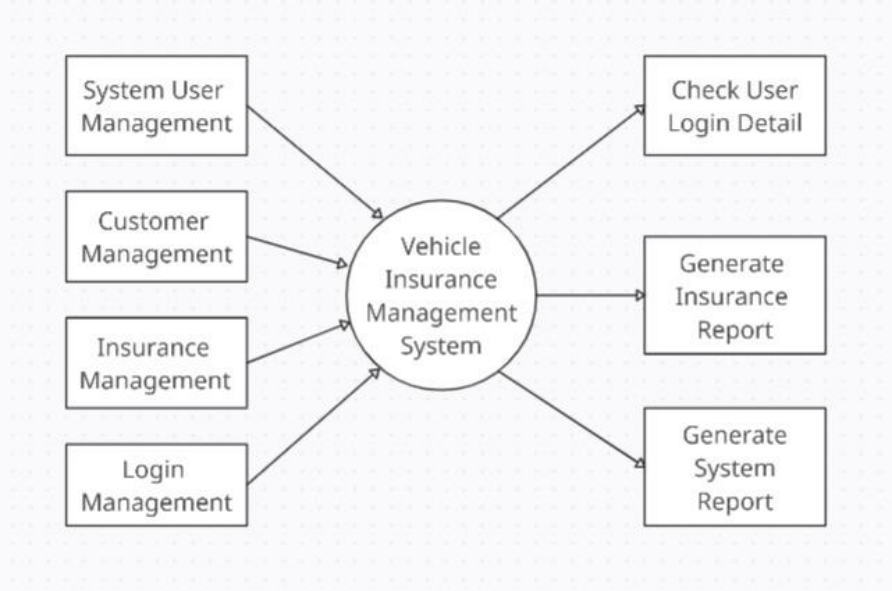


Figure 3.1.5 – B Level 1 of Dataflow

Level 2:

The most detailed explanation of all 3 is Level 2 that gives every detail that is they are about the project. This level has every necessary detail about the project and its functionalities.

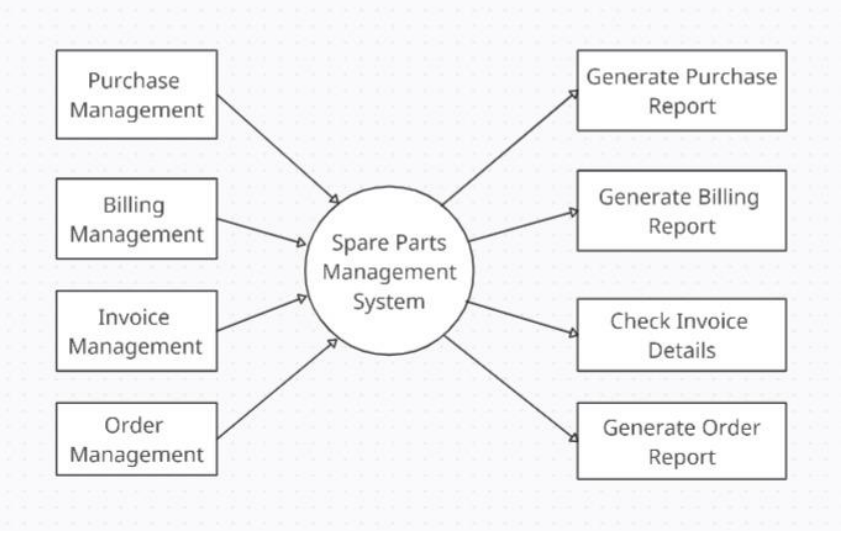


Figure 3.1.5 – C Level 2 of Dataflow

3.2 System Analysis

Use of system analysis is very important as it aids in problem solving and divides the system in component pieces to study and understand the nitration of the project in depth. System analysis is study of business or process to identify the goals and purpose to achieve them effectively. Without analysis making a project is highly risky and impractical as the results expected can always be wrong which out analysis part on projects.

3.2.1 Existing System description

The preset market for automotive is around 25 billion and projected to be 29 by next 4 years. It is impossible to remote monitor every automobile as the no of vehicles is in large number. By this logic existing models to predict and analysis data is never enough as them exist many automotives that are got of sight for these existing systems. Existing models like:

IBM: Connected Vehicle production maintenance solution:

A solution for monitoring connected vehicles to predict analysis wherein this solution provided a mobility service to work even when people present in car or outside car. This model helps in communicating with vehicle via user interface HMI or external application. Additionally, cameras and sensors monitor this health of car components with AI analysing and recommending faults to drivers.

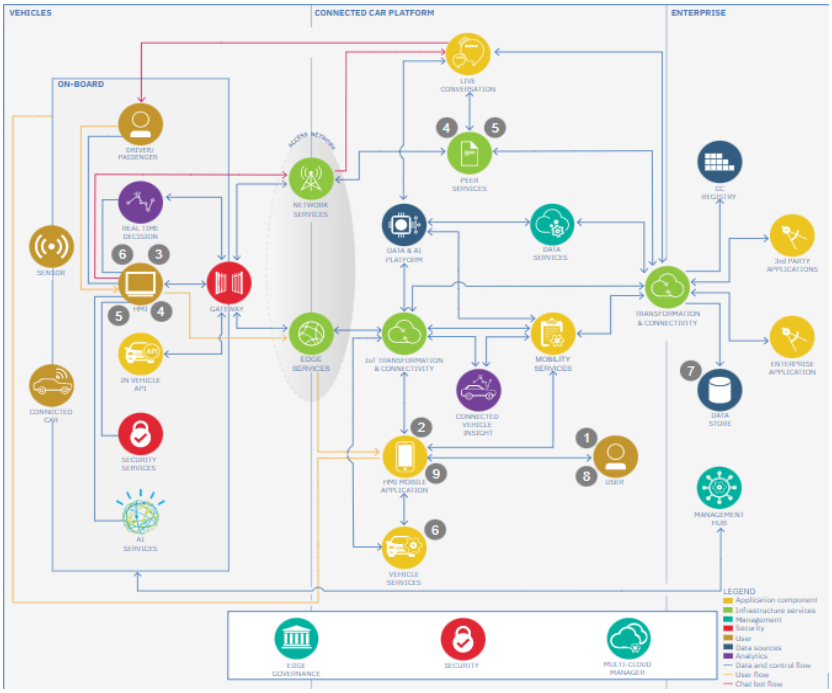


Figure 3.2.1-A IBM Architecture

Intuceo – Predictive Maintenance Solution

Intuceo is wealth of in vehicle sensor data along with machine learning to provide predictive maintenance solution for OEMs and dealers. This solution claims to transform and analysis the in-vehicle sensor data, allowing customers to deploy this solution in manufacturing industry to reduce downtime and cost. A diagram to understand the method is shown below.

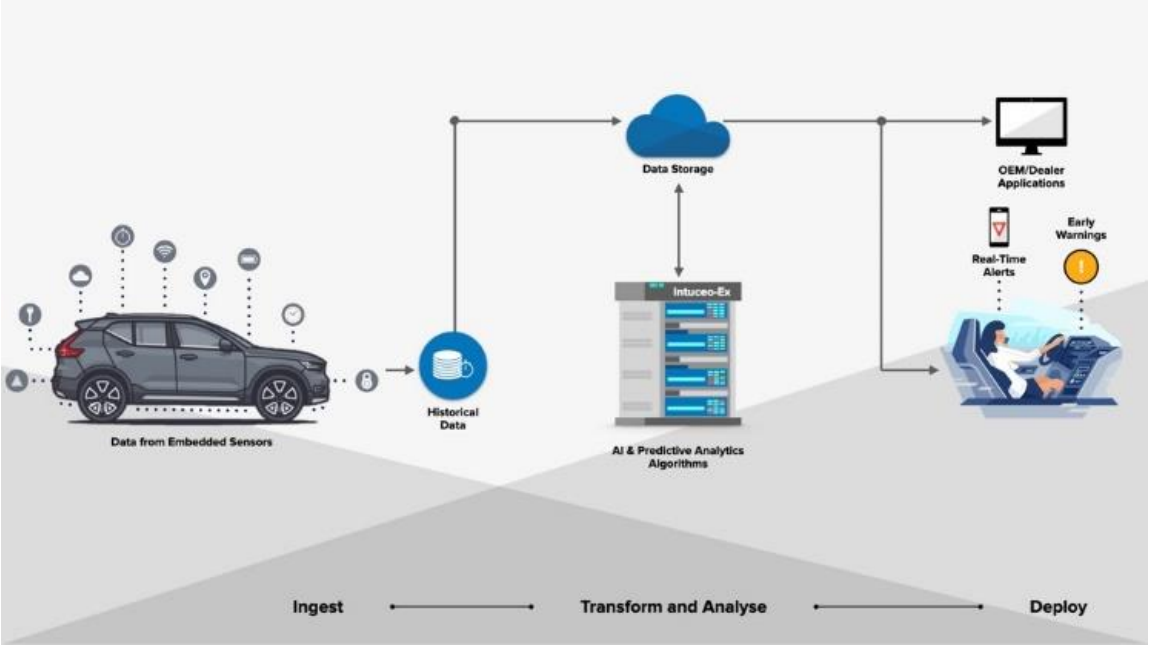


Figure 3.2.1 – B Intuceo Model

HMG: Sound-based Fault Diagnosis and Predictive Maintenance

Research Lab at Hyundai and Kia motors Namyang R&D centre has come up with a AI solution to understand sounds by faulty components. This method is done by extracting various parts of fully functioned engines and train a model. These sound when processed, analysed, and categorized to add them toa database to teach a model. Once the database is grown, we provide the system to test and diagnose the sound and provide a solution for the faulty sound. Figure 9 shows the diagnosis steps followed to detect these faults using AI.

Special Features of this application is:

- Improves availability to 10+%
- Increases life of vehicle and parts by 15%
- Reduces TCO by 20+%

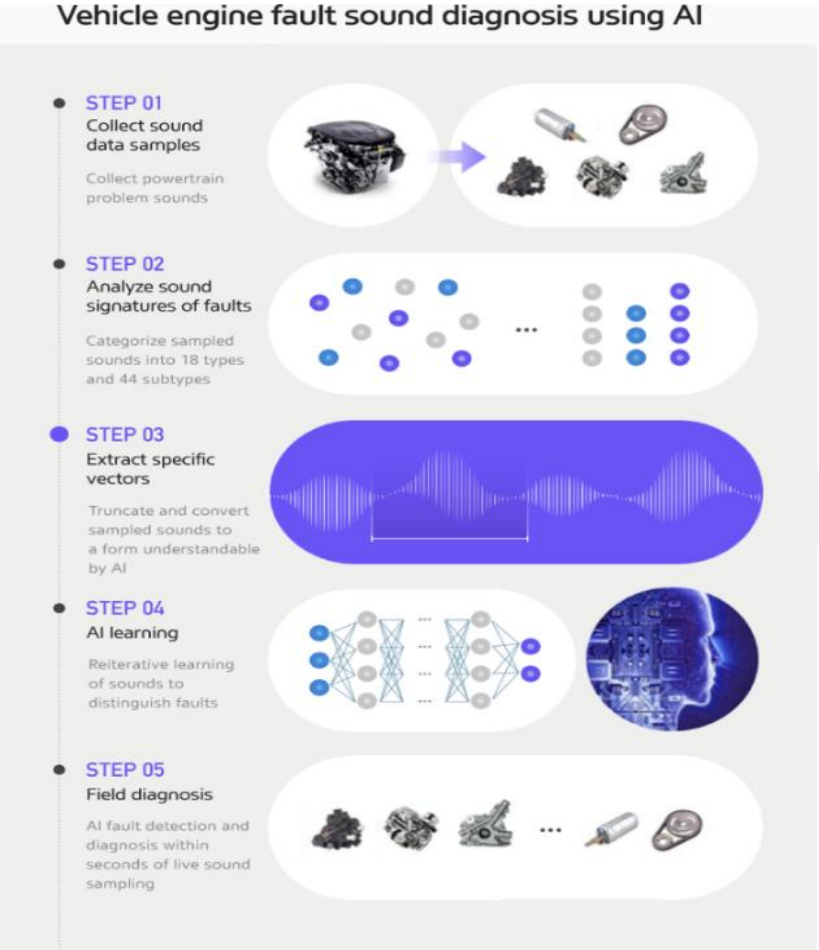


Figure 3.2.1-C Hyundai Sound Diagnosis

3.2.2 Proposed System

This project aims to reduce the complexity of analysis of large datasets. In existing datasets limitation on model, company or attributes like sound is used. With the help of this project, we wish to integrate the dataset better with the help of ANFIS and Restricted Boltzman Machine. In machine learning, ANFIS is a network of 5 layers with hidden layers integrating better than CNN. The main concept of summation is used with complex rule. Restricted Boltzman Machine doesn't give a output but helps in filtering, classification, feature learning of dataset. These components are very important as they play a role in output module but won't provide a specific output..

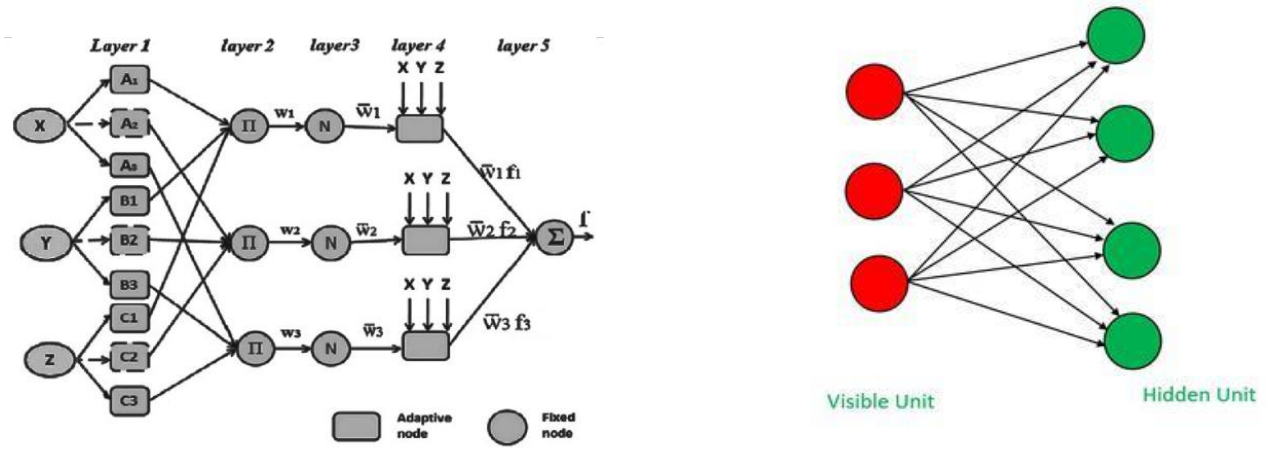
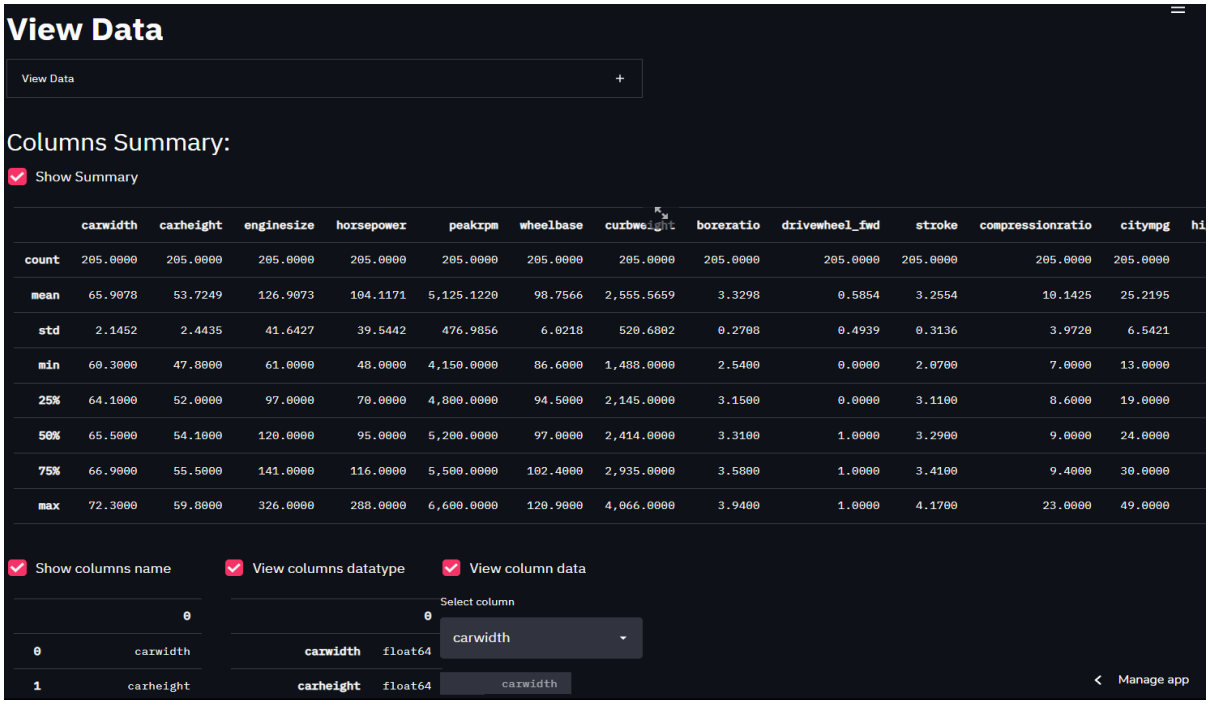


Figure 3.2.2 Architecture of ANFIS and RBM

3.3 Screenshots

3.3.1 Database

SQL is called Structural Query Language that helps in accessing multiple datasets and bases. MYSQL helps in Oracle-backed pen source RDBMS (Relational Database Management System) based on SQL. The main advantage of MYSQL is its flexibility to run on any platform from windows to Linux. The main practical usage of MYSQL is its association with web application and online publishing.



	carwidth	carheight	enginesize	horsepower	peakrpm	wheelbase	curbweight	boreratio	drivewheel_fwd	stroke	compressionratio	citympg
count	205.0000	205.0000	205.0000	205.0000	205.0000	205.0000	205.0000	205.0000	205.0000	205.0000	205.0000	205.0000
mean	65.9078	53.7249	126.9073	104.1171	5,125.1220	98.7566	2,555.5659	3.3298	0.5854	3.2554	10.1425	25.2195
std	2.1452	2.4435	41.6427	39.5442	476.9856	6.0218	520.6802	0.2708	0.4939	0.3136	3.9720	6.5421
min	60.3000	47.8000	61.0000	48.0000	4,150.0000	86.6000	1,488.0000	2.5400	0.0000	2.0700	7.0000	13.0000
25%	64.1000	52.0000	97.0000	70.0000	4,800.0000	94.5000	2,145.0000	3.1500	0.0000	3.1100	8.6000	19.0000
50%	65.5000	54.1000	120.0000	95.0000	5,200.0000	97.0000	2,414.0000	3.3100	1.0000	3.2900	9.0000	24.0000
75%	66.9000	55.5000	141.0000	116.0000	5,500.0000	102.4000	2,935.0000	3.5800	1.0000	3.4100	9.4000	30.0000
max	72.3000	59.8000	326.0000	288.0000	6,600.0000	120.9000	4,066.0000	3.9400	1.0000	4.1700	23.0000	49.0000

	0
0	carwidth
1	carheight

	0	
carwidth	float64	carwidth
carheight	float64	

Figure 3.3.1- A Database

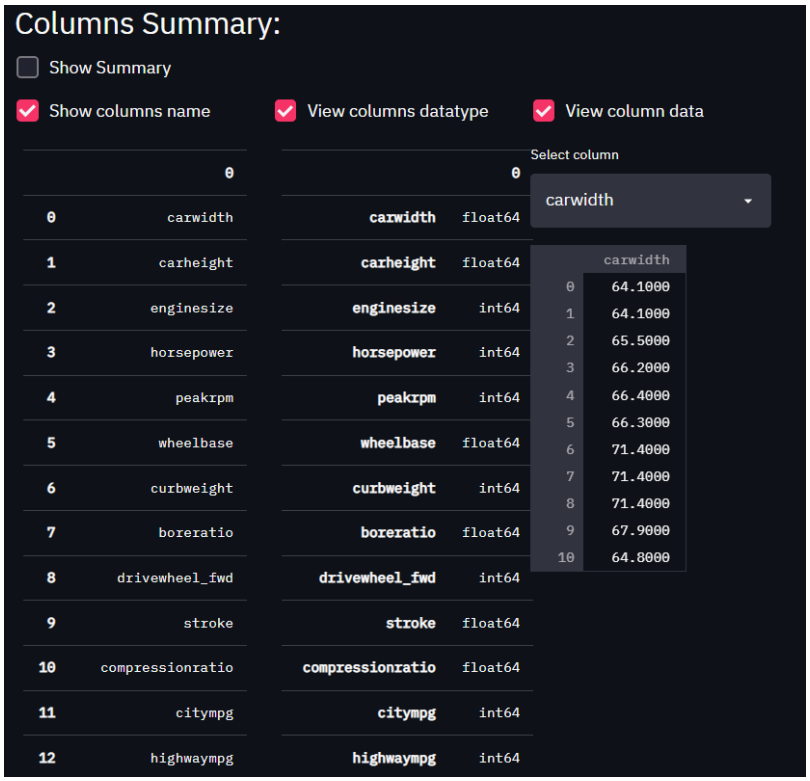


Figure 3.3.1- B Database

3.3.2 *Website Module*

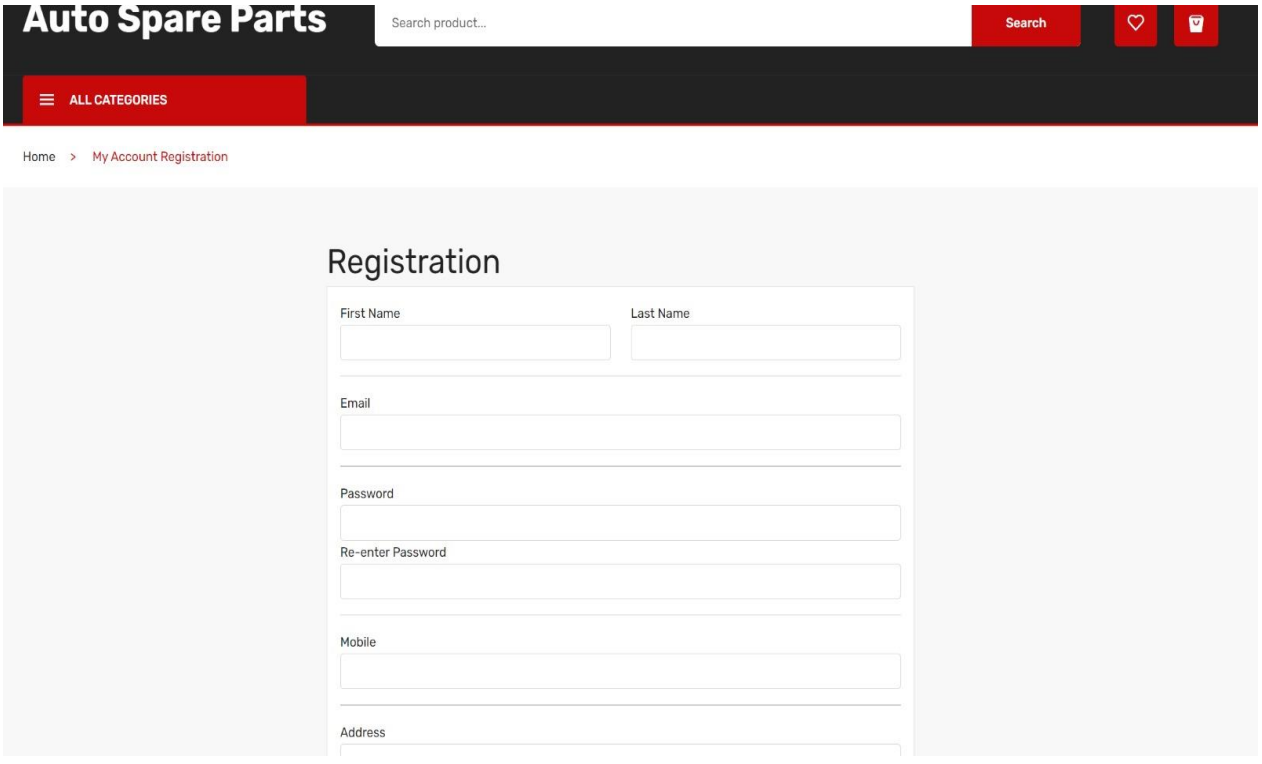


Figure 3.3.2.1 Customer Registration

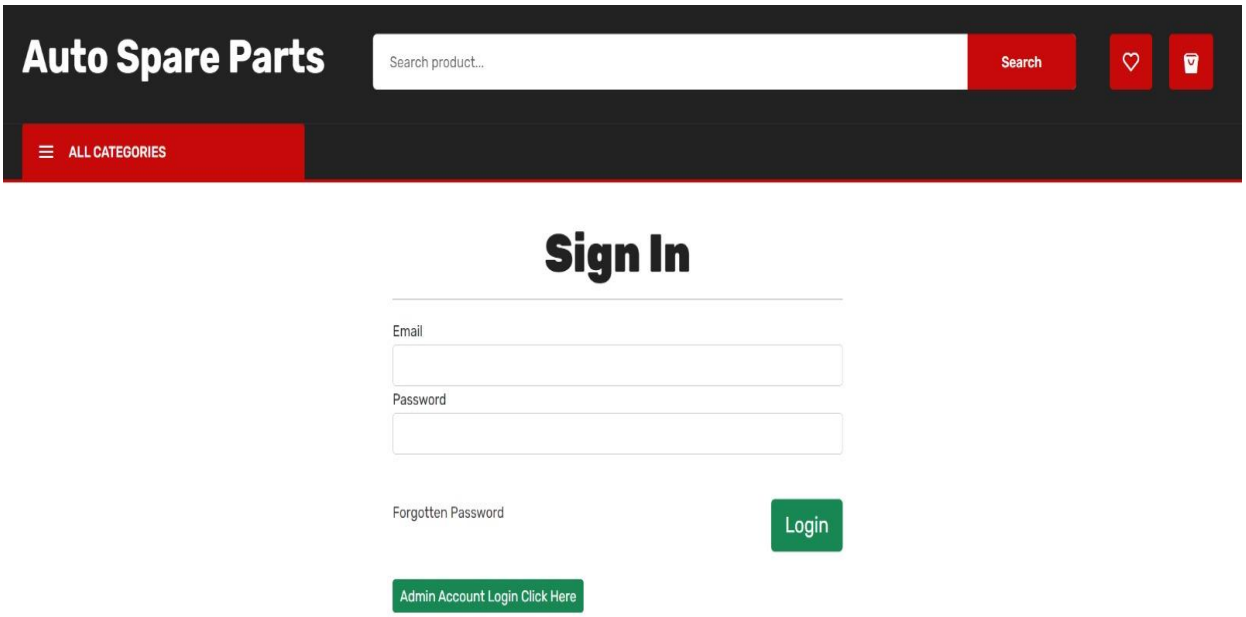


Figure 3.3.2.2 Website login

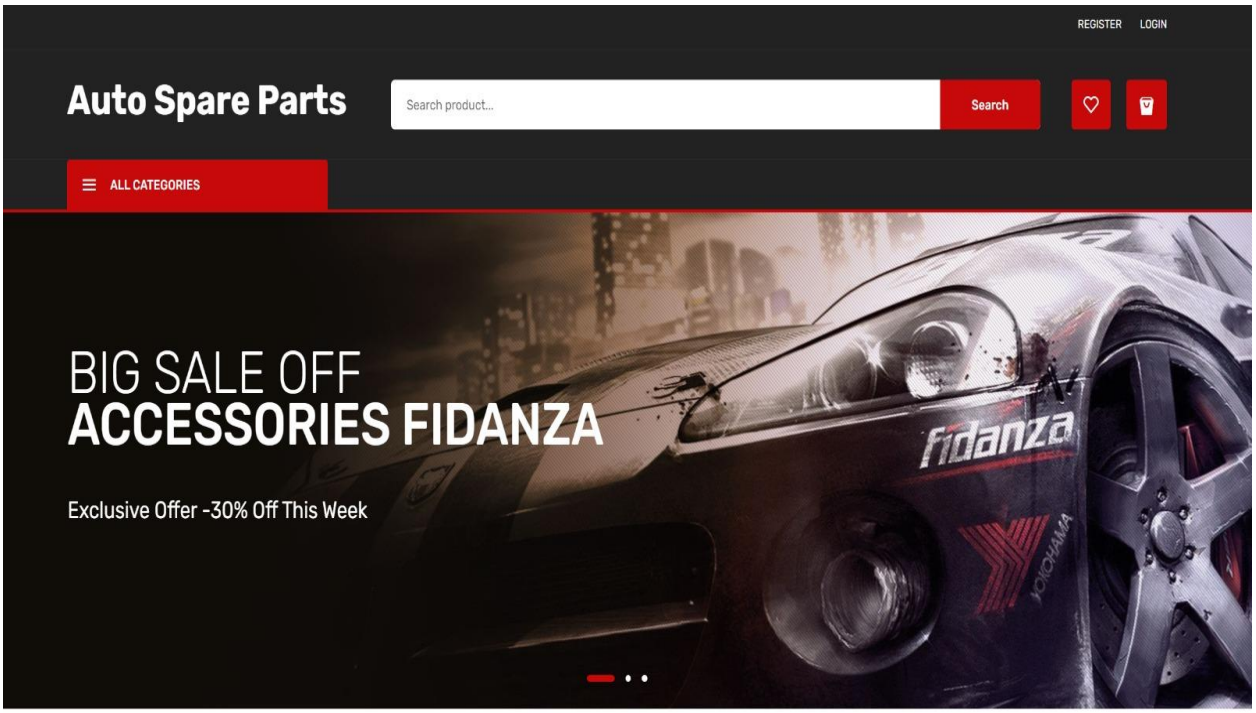


Figure 3.3.2.3 Home Page

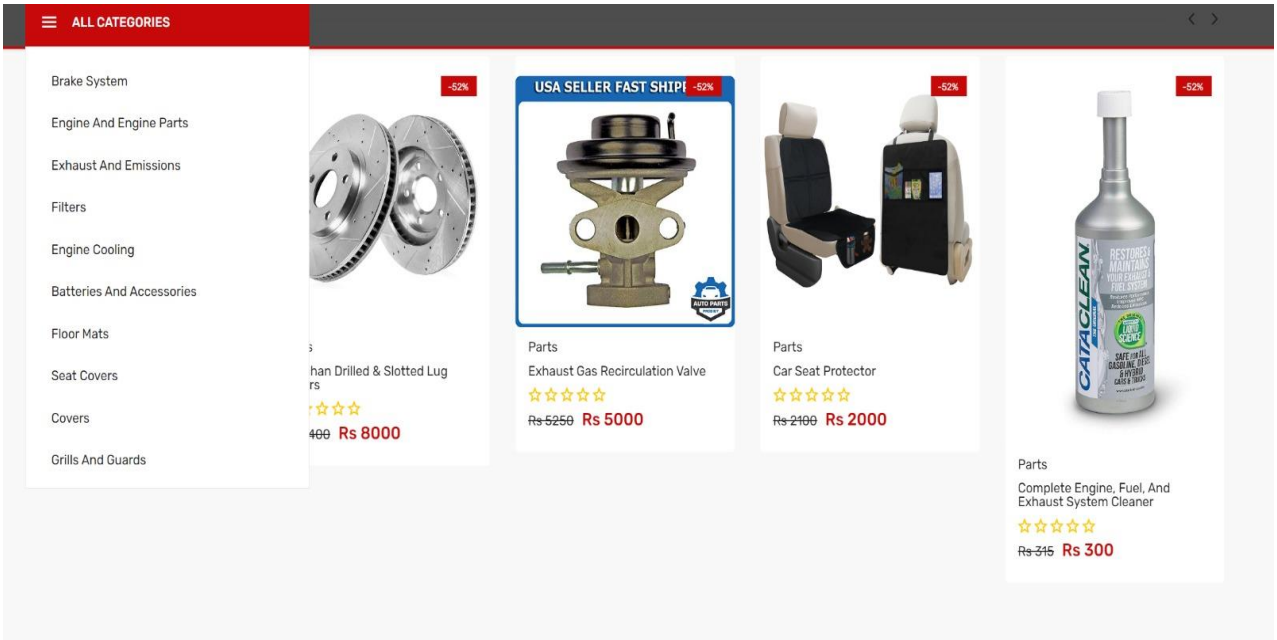


Figure 3.3.2.4 Product Listing

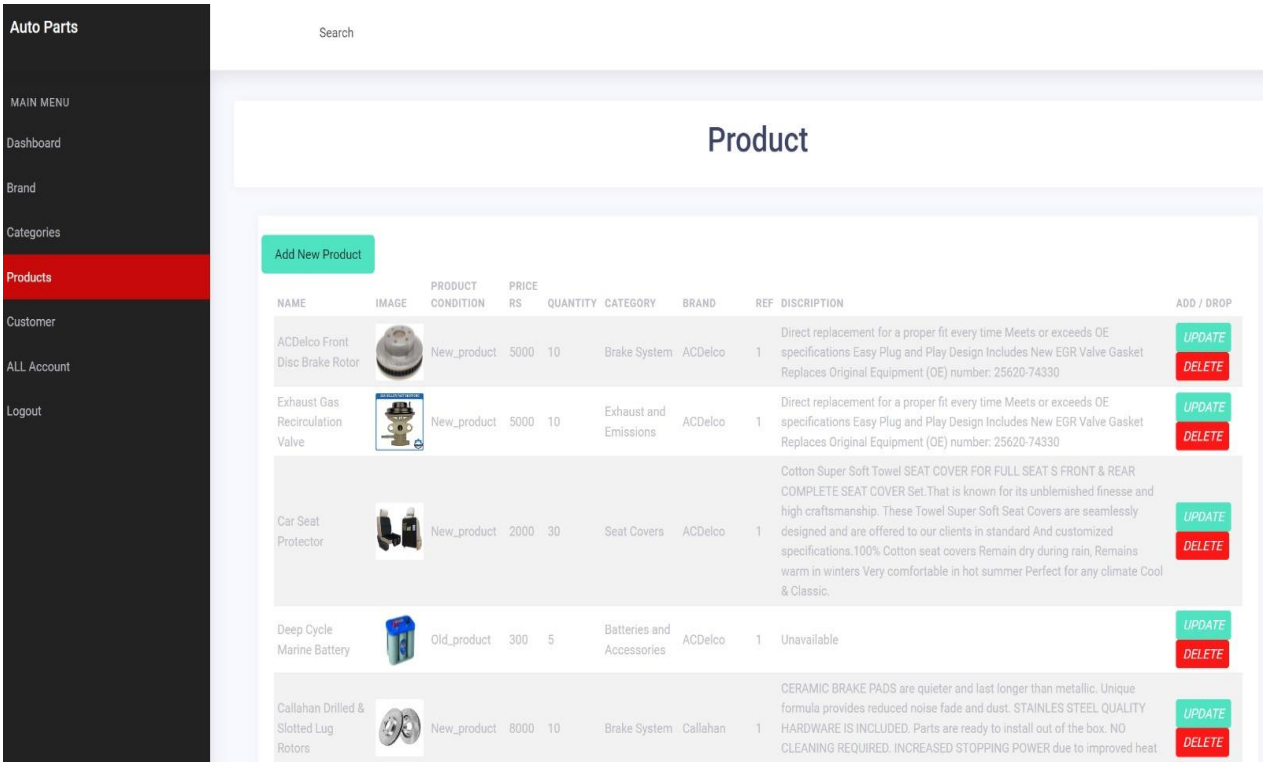


Figure 3.3.2.5 Seller side Add/Drop products

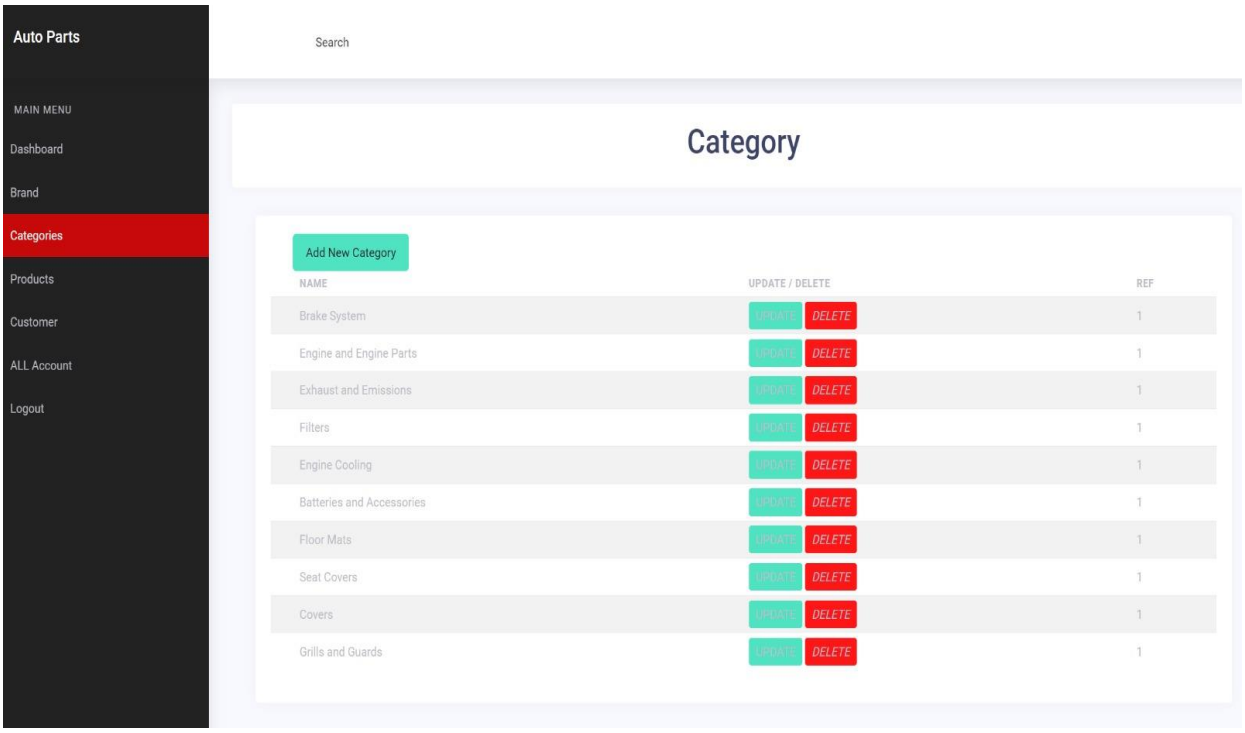


Figure 3.3.2.6 Category Page

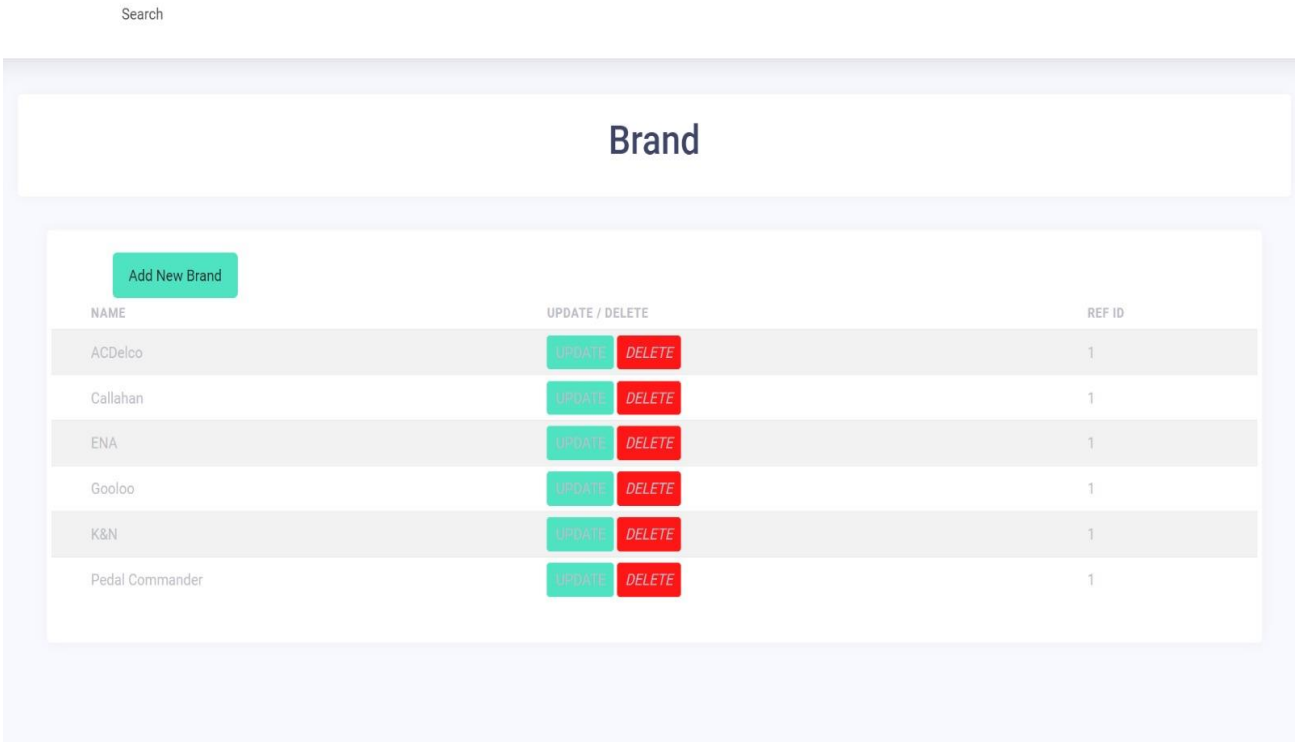


Figure 3.3.2.7 Brand Addition Page

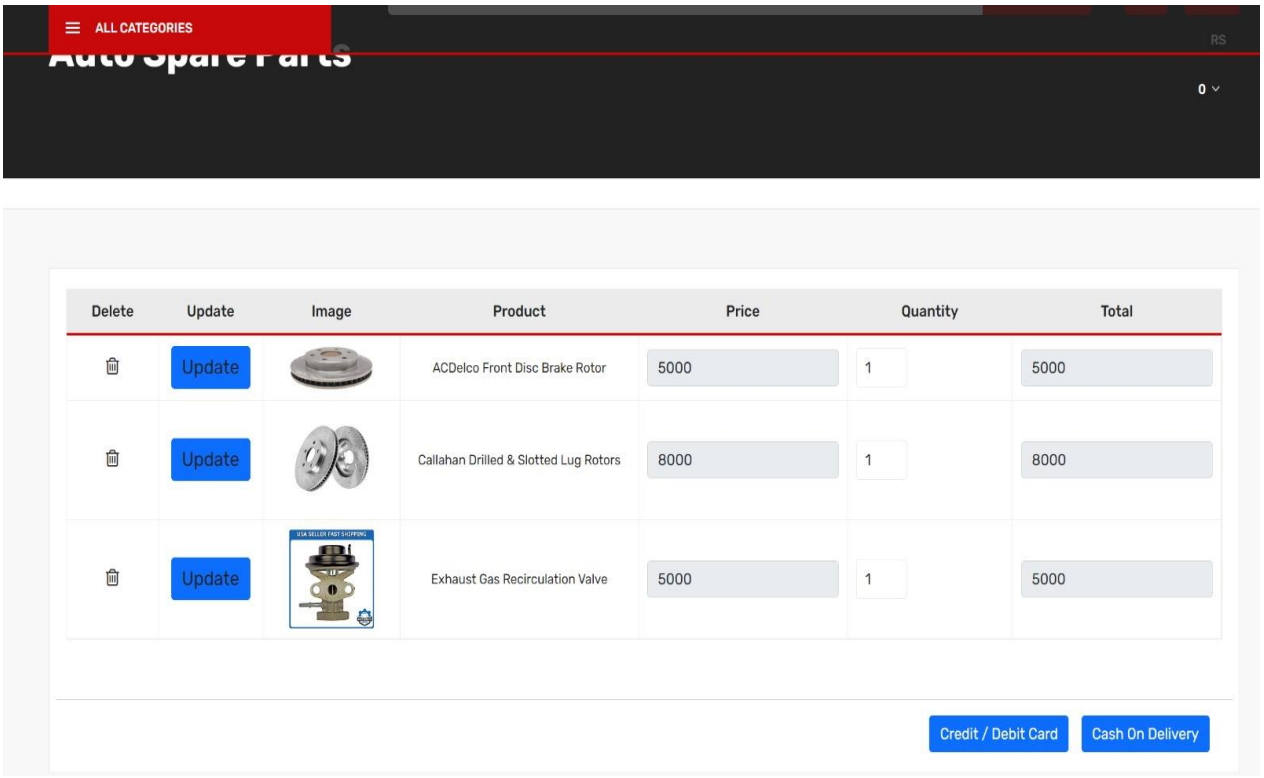


Figure 3.3.2.8 Cart Page

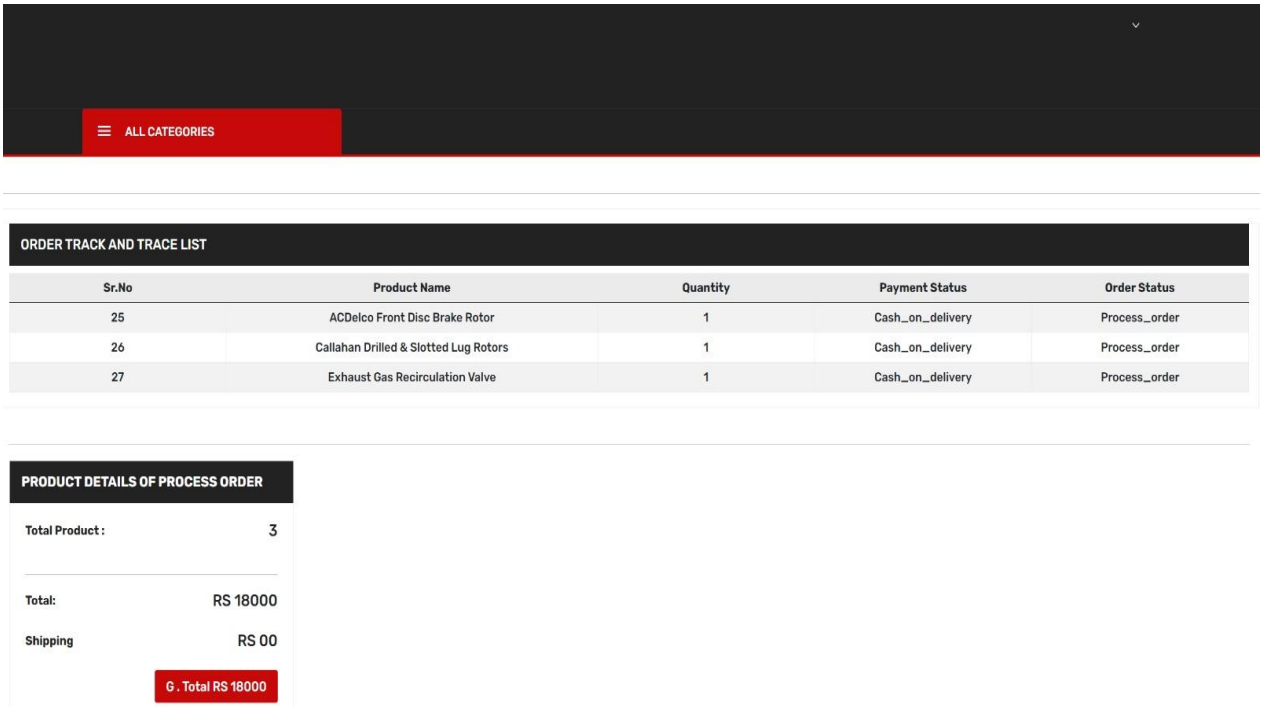


Figure 3.3.2.9 Order Review Page

3.3.3 Data Analysis Module

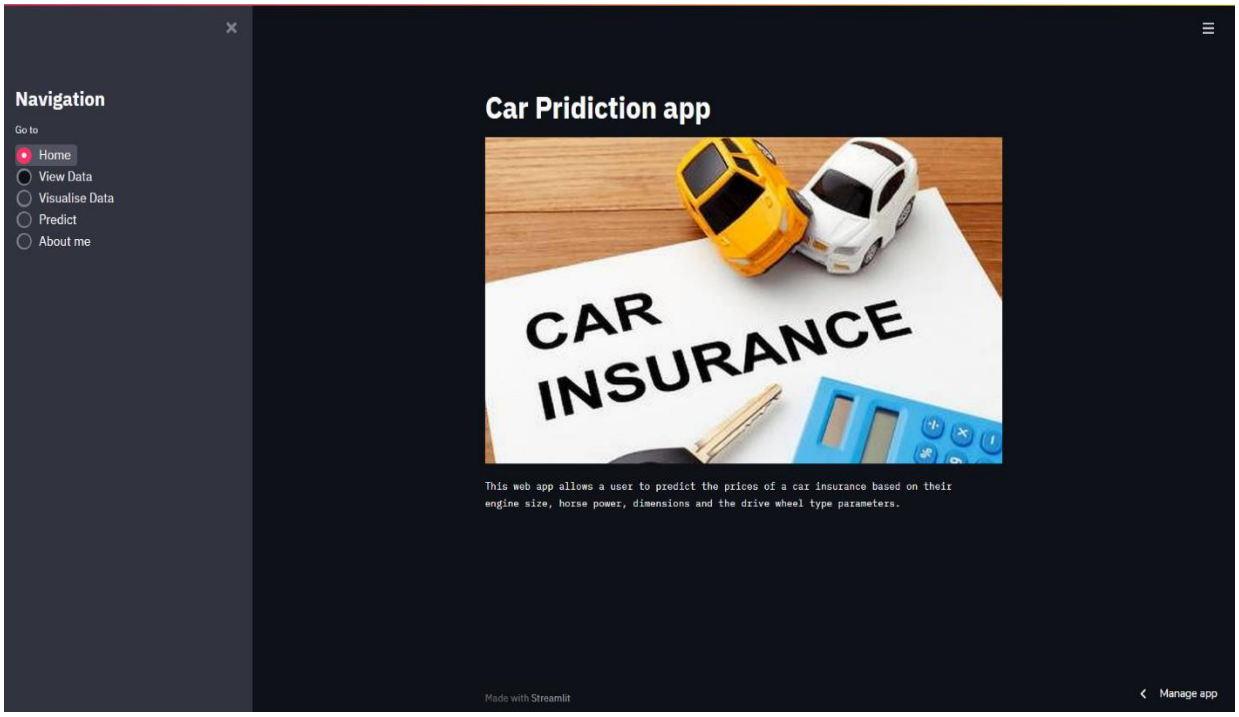


Figure 3.3.3.1 Home page

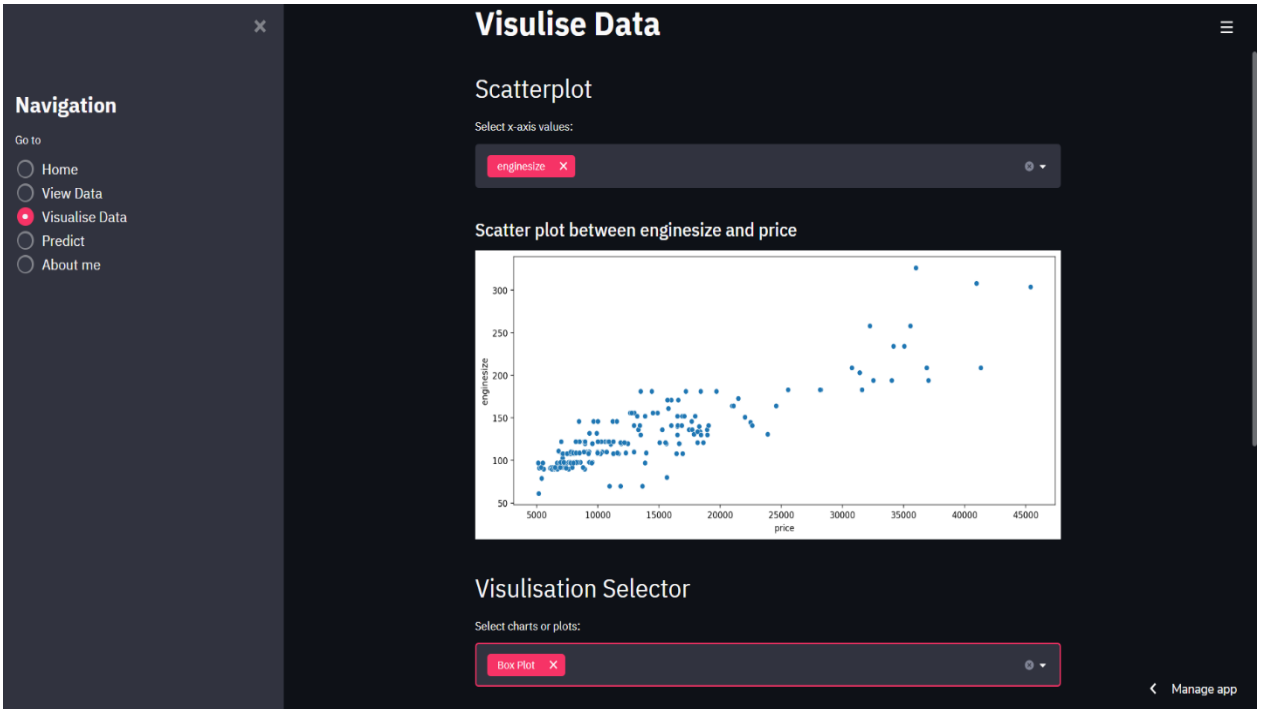


Figure 3.3.3.1 Visualization data -1

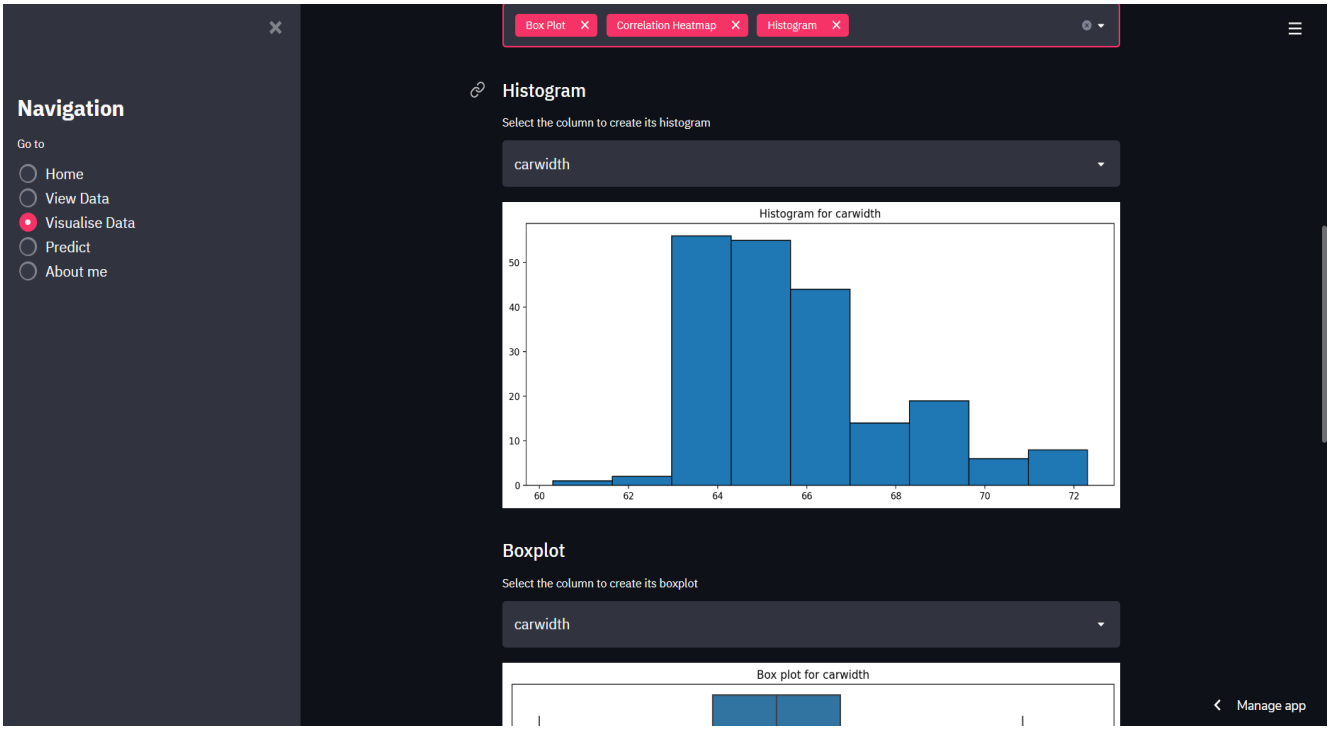


Figure 3.3.3.3 Visualization data -2

Figure 3.3.3.4 Prediction

CHAPTER IV: Coding and Testing

4.1 Coding

For this chapter let's understand the technologies used:

4.1.1 Pandas

An open-source python package builds on Numpy that is most used for data cleaning and analysis tool. This is fast, flexible, and expresses data effectively. Pandas deal with three data structures Series, Data frame. Panel. Series is one D array with homogenous data where size cane 't be changed but values are editable. Data frame is two D array with heterogenous data where data is in form of rows and columns and size can be changed. Lastly panel is three D with Heterogenous data making it difficult to represent in graphics. Pandas is capable of reading files from Excel, CSV and many other files to analysis data.

Some Listed features are:

- It provides time-series functionality
- Size mutability: Data Frame and higher-dimensional object columns can be added and deleted
- Represents the data in tabular form
- Merges and joins two datasets easily

4.1.2 Machine Learning and Deep Learning

Algorithms like ANFIS (Adaptive Network based Fuzzy Inference system), Restricted Boltzmann Machine, Lasso Regression, Gradient Boost has helped a lot in this project. Making use of these along with pandas, Seaborn, Matplotlib and Scikit Learn has brough the project to a better position. Deep learning in form of ANFIS has really helped.

i. *Adaptive Network based Fuzzy Inference system (ANFIS*)

This network used 5 layers to analysis and give output. As shown in Figure 5 the first hidden layer maps input variable to each membership function. In second layer T-norm is applied to calculate antecedents of rules and third layer is used in normalization of rules strengths. In fourth layer consequents of rules are determined. The last layer or output layer gives out global output after calculating by summation of signals taken in. ANFIS uses backpropagation learning to determine input membership functions and least mean square method to determine the consequent parameters. These are class of adaptive networks that are equivalent to Fuzzy Inference system in function.

ii. Restricted Boltzmann Machine

A variant of Boltzman Machine in which neurons are present in input layer and hidden layer encompasses symmetric connection. In restricted Machine connection to same layer type is not allowed so no two neurons in input layer or hidden layer cannot connect to each other. This rule is not applicable in case of hidden and visible layer. This Machine does not have any output layer but adjects the weight and see how our prediction is accurate. By using this machine, we can verify the output we received from ANFIS. The energy function of this machine is:

$$E(v, h) = -a^T v - b^T h - v^T W h$$

Some major applications include:

- Filtering
- Classification
- Risk Detection
- Feature Learning
- Economic analysis

iii. Gradient Boost

Like random forest we propose the use of Gradient Boosting which ensembles method that creates many decisions trees. This model helps in adjusting the value of the vehicle if them exist an error or defect in the prediction. This algorithm is trained by the error difference it faces from previous results This helps in building a better output from small datasets. The major drawback of this model is the time it takes as making a tree is time consuming. With addition of ANN and SVM to this model we believe it is possible to predict the cost with most accuracy. After through study we believe the better the data cleaning and no of inputs prediction performance increases. Even with complex cases with this model we believe 70% accuracy is possible. Therefore, multi algorithm approach is always the best approach in this prediction module as the better the refinement is the better, we can expect the outcome.

4.2.3 *Matplotlib*

A visualization library in python for plot in 2D. Matplotlib was built on Numpy array and has the biggest be benefit in visualization by allowing visual access to huge amount of data in easy manner. Matplotlib is flexible and can be used on windows, macOS and Linux with dependencies only. The library has various plots to help understand trends, patterns and make corrections also.

4.1.4 Seaborn

Similarly, to that matplotlib Seaborn also helps in visualization based on matplotlib. Seaborn helps in making high level interface for drawing attractive and informative graphics. Compared to Matplotlib the major advantage is its close integration to pandas. Major features of Seaborn are:

- Plotting of statistical time series data

- Comes with built themes to style matplotlib
- Univariate and bivariate data visualization

4.1.5 Streamlit

A free and open-source framework to build and share machine learning and data science web apps. It is python-based library specifically for machine learning engineers. The biggest advantage is that prior knowledge on web development is not required and can deploy models with easy with few lines of code. Major feature is Streamlit is:

- Experience in front end is not required
- It is compatible with major libraries of python like pandas. Keras, pyTorch, seaborn etc.
- The computation is fast and simply.

4.1.6 Xampp

Xampp Is a open source Web server solution used for web application testing on local host. This mainly consist of Apache HTTP Server, MariaDB database which interprets the scripts in PHP and peri language

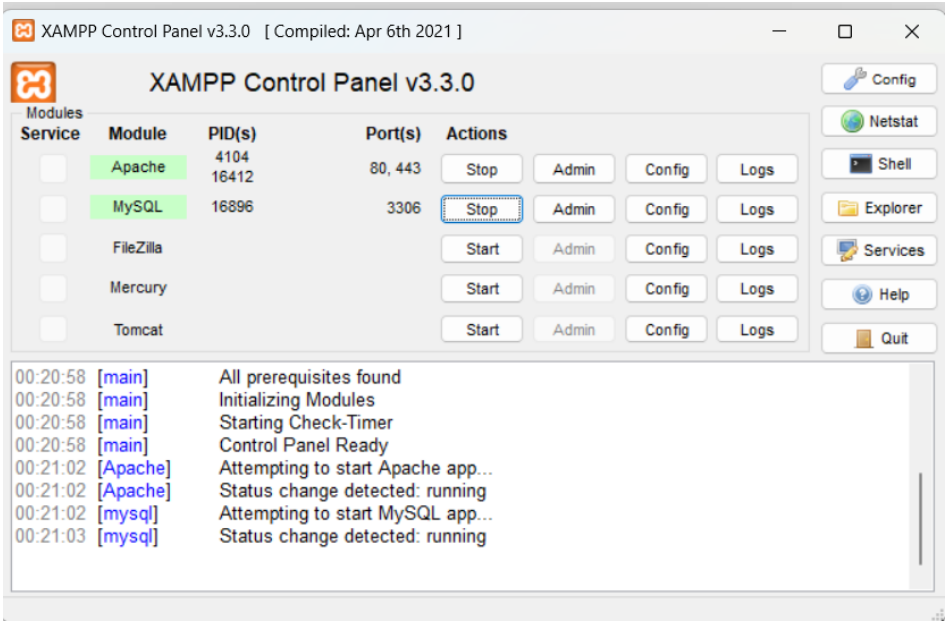


Figure 4.1.6 Xampp Control Panel

4.2 Testing

The main aim of testing is to find errors in project since a program is subjected to set of inputs and various observation were made based on observation to determine how a program behaves. There are 3 types of testing:

1. Unit Testing
2. Integrated testing
3. Regression testing

4.2.1 Unit Testing

It is done when a module is made and review correctly. To do unit testing we need to provide complete environment along with module.

- Non local data that model access must be available
- Process to call functions under test with parameters

Unit testing is done on the following modules:

1. Test for Login module for web application
2. Searching of products
3. Integration of machine learning code

4.2.2 Integration testing

Second step is integration testing where we test various integration of project by providing input. The primary objective is to test module interface to ensure no errors when modules are working together. This is done in group in which modules are code, applications, client and server applications and networking. The steps followed in this testing are:

1. Bottom-up testing where it starts from lowest or innermost unit and moves up. This is done until modules are integrated and application is tested as single unit.
2. Top-down testing is done from the top of application to the lowest point of module. In this the application is test from top to bottom to see the interaction of web application tested as single unit.

4.2.3 *Regression Testing*

This is very important when u is making any changes to the module as it helps in rerunning existing tests against modified code to determine if the changes disturb any other models and it also helps in writing new test. To do regression we need to focus on time and adequate coverage of module. Never test already passed modules as it is waste of time and makes the module slow.

Some factors and strategies to consider are:

- For every bug fixed we must do regression test
- Tracing the effect of change on program
- If multiple tests are similar then determine the less effective and remove it
- Review must be periodical as it helps in elimination of unnecessary tests.

CHAPTER V: Maintained Feature

5.1 Maintenance features dependency

The best use of this software is it requires less maintenance and all we need is to keep using software to improve the quality. Let's understand some feature:

1. Providing training dataset for machine to learn before we use it in real life scenarios. This dataset will further improve accuracy of software in which larger the dataset the higher the accuracy
2. Code must be updated with newest version of libraries
3. Next check is on the smooth function of software by keeping it up to date as per user interface. This is important to parsing of extracted data.

It entails setting up the website on the computer system and starting the main operations. In this stage, users are introduced to the website for the first time and given the opportunity to interact with it. It entails keeping the website constantly updated to make sure it complies with the most recent updates to the same and current information needs. The developer evaluates the modifications and incorporates them as well. The following tasks are included in the implementation phase. the creation of websites, limiting which people may make changes to the website, and tracking logs to maintain tabs on data and website activity.

5.2 Web Development Life Cycle

The WEB DEVELOPMENT LIFE CYCLE includes these phases, which take into account the following 10 guidelines:

- Rule No. 1: Website Goals To get the intended results, the primary goal for the website's construction must be determined and evaluated.
- Rule 2: Website Map Prepare the overall website's physical look as well as its modularized structure. This might serve as a blueprint for future work.
- Rule 3: Gathering Requirements Include all client requirements and requirements criteria. The development process would be more succinct the more explicit the requirements are.
- Rule 4: Verification Before launching the website, the needs that were gathered and the design criteria needed to be checked against intellectual property laws to prevent privacy concerns.
- Rule 5: Website designs should be straightforward, clear of ambiguity, and uncomplicated so that the general public may readily understand the material.

- Rule 6: Avoid shortcuts Offer readers navigational tools to make it easier for them to browse the website quickly and get the information they need.
- Rule 7: Unification After the structure is created, the website may be evaluated against other similarly functional websites to provide the user a special application experience.
- Rule 8: Multilingualism The extra features can be added to give multilingual, handicapped consumers with technical help.
- Rule 9: Testing After doing multiple rounds of testing, the developed web application should be published. The testing verifies that the completed system meets the initial objectives set out.
- Rule 10: Upkeep The website must be updated often to keep it accurate and of high caliber. Additionally, the consumer may receive current information from this

CHAPTER VI ADVANTAGES AND CHALLENGES OF DEVELOPED SYSTEM

6.1 Advantages of proposed system

There are several advantages of using a Vehicle Health Analysis and Spare Parts E-Commerce website over traditional methods. One of the main advantages is convenience. With this website, customers can easily access information about their vehicle's health and order spare parts online from the comfort of their own home. This saves time and effort compared to visiting a physical store.

Another advantage is the ability to easily compare prices and products from multiple suppliers. This helps customers find the best deals and ensures that they are getting high-quality spare parts. Additionally, the website can provide personalized recommendations based on the customer's vehicle make and model, which can further streamline the process of finding the right spare parts.

6.1.1 Insurance price optimization

The presence of cars on the road exposes them to potential minor or major accidents. Insurance companies typically pay out claims to approximately 8% of their clients every year, which can range from a few hundred to thousands of dollars. Although only 1% of the claims are generally large, they can result in pay-outs of $10,000 or more. Machine learning has the potential to predict if a driver is at risk of causing a significant loss, which is excellent news for insurers since predictive analysis can achieve an accuracy rate of 78% or higher. Insurers can now build dynamic pricing models for each client that are still marketable. This allows insurance companies to provide real-time pricing at the point of sale, enabling them to optimize their pricing and create new insurance services that are beneficial to good drivers and prevent them from being charged the same premium as poor drivers. Predictive analytics for insurance are a win-win situation for both the insurer and the policyholder.

6.1.1.1 Reduce fraudulent claims

Let’s understand by stats

- In US and Canada 10% of claims are fraud and caused huge lass to the industry
- A company stated that 80 billion dollars were paid to false claims

To avoid this and reduce the fault we try this system. With the help of this project, we can find proper value that a user should pay and also in future when developed we can also include AI in detection of fraudulent claim and reduce risk score.

6.1.1.2 Marketing and buyers’ behaviour forecast

Predictive analytics can benefit car insurance companies in multiple ways. From optimized pricing and reducing fraudulent claims to utilizing forecasting, these companies can optimize the use of their time and resources. Not only that, these benefits extend to the customers as well.

6.2 Limitation of system

Most significant limitation is the availability of data set for vehicle health prediction. This is because no company has given open-source access to the sensor data of maintenance this is the major limitation and reason for the project to be hard to complete. Let's understand some more:

- No data set available and companies not providing any data sets on sensor data
- With evolution of modern technologies, the machine learning and deep learning algorithms used in this project will be outdated
- New technologies with better algorithms are coming in play every year making it hard to have a stable project

6.3 Challenges of system

- Data Accuracy: One of the biggest challenges in developing a website is ensuring that the data used in the analysis and prediction models is accurate and reliable.
- Data Privacy and Security: With the collection of sensitive vehicle and customer data, the website must have robust security measures in place to ensure the protection of personal information.
- Technical Complexity: The development of such a website requires a deep understanding of multiple technologies, including machine learning, web development, and database management, which can pose technical challenges.
- Integration of Multiple Modules: The website includes multiple modules, such as vehicle health analysis and spare parts e-commerce, which require seamless integration and compatibility to ensure the website's proper functioning.

CHAPTER VII: CONCLUSION AND IMPROMENT IN FURTURE

7.1 Conclusion of work

The vehicle monitoring system offers accurate information on all fundamental factors influencing a vehicle's performance. Due to the early detection of damage made possible by this information, the performance of the vehicle will improve. As a result, it gives the owner accurate information so that during car servicing, the service centers cannot present incorrect information and demand additional payment. The vehicle insurance app aids the owner in obtaining the most insurance payout possible if the vehicle develops flaws or malfunctions as a result of normal wear and tear. According to the company's insurance policy, the owner can obtain the greatest claim possible, and that too without deceit or fraud. This method of auto insurance aids in eradicating such offensive behaviors.

The website aims to provide a comprehensive solution to the customers by offering insurance analysis and spare parts e-commerce services. The website employs machine learning algorithms to predict the insurance amount for a vehicle and has a spare parts e-commerce platform for customers to purchase genuine spare parts for their vehicles. The development of such a website can pose various challenges, including data accuracy, data privacy and security, technical complexity, and integration of multiple modules. However, with proper planning, analysis, and testing, these challenges can be overcome, and a high-quality website can be developed. The website has immense potential for the automobile industry and can provide a significant value proposition to the customers. It can provide customers with personalized and accurate insurance predictions and a convenient platform to purchase genuine spare parts, making it a one-stop-shop for their vehicle needs.

7.2 Suggestions to project

An automatic tool that can provide diagnosis analysis service and responds to user diagnosis in real time. The data analysis approach can also be used to analysis other production system that is based on frequent vehicle failure. There exist N number of algorithm and with slightest change we can find change in accuracy of the model. Models where previous mistakes are considered when prediction in unsupervised learning is one different approach to make this project work.

Reference

[1] Deep Learning Towards Intelligent Vehicle FaultDiagnosis Mohammed Al-Zeyadi∗, Javier Andreu-Perez∗, Hani Hagras∗, Chris Royce†, Darren Smith†, Piotr Rzonsowski†, Ali Malik‡

[2] Vehicle Remote Health Monitoring and Prognostic Maintenance System Uferah Shafi, Asad Safi, Ahmad Raza Shahid, Sheikh Ziauddin,and Muhammad Qaiser Saleem

[3] Machine learning methods for vehicle predictive maintenance using off-board and on-board data, Rune Prytz

[4] Z. Cheng, M.-Y. Chow, D. Jung, and J. Jeon, "A big data based deep learning approach for vehicle speed prediction," in 2017 IEEE 26thInternational Symposium on Industrial Electronics (ISIE). IEEE, 2017, pp. 389–394

[5] Machine learning methods for vehicle predictive maintenance using off-board and on-board data, Rune Prytz

[6] Study on the predictive maintenance of vehicles and its management using the specific "Keep the Machine Running" application ioan virca, dorel badea

[7] https://www.analyticsvidhya.com/blog/2022/07/application-of-ai-in-predictive-maintenance-of-vehicles/

[8] https://tech.hyundaimotorgroup.com/article/worlds-first-ai-fault-detection- and-diagnostics-by-the-hmg/

[9] https://www.intuceo.com/blog/how-to-get-started-with-predictive-maintenance-using-machine-learning/

[10] Prediction of motor insurance claims occurrence as animbalanced machine learning problem; Sebastian Baran, Przemys law Rola

[11] Hanafy, M; Ming, R. Machine Learning Approaches for Auto Insurance Big Data. Risks2021,9(2) 42

[12] Pesantez-Narvaez, J.; Guillen, M.; Alca˜niz, M. Predicting Motor Insurance Claims UsingTelematics Data—XGBoost versus Logistic Regression. Risks 2019,7, 70

[13] A proposed model to predict auto insurance claims using machine learning techniques; shady Abdelhadi, Khaled Elbahnasy, Mohamed Abdelsalam

[14] https://www.kaggle.com/datasets/nehalbirla/vehicle-dataset-from-cardekho

[15] A systematic literature review of machine learning methods applied to predictive maintenance Author links open overlay panel *Thyago P. Carvalho, Fabrízzio A. A. M. N. Soares, Roberto Vita, Roberto da P. Francisco, João P. Basto, Symone G. S. Alcalá*

[16] A study on car insurance purchase prediction using machine learning Shwetha Shree K*1, Mohan Murali P*2 (2022)

[17] Machine Learning Approaches for Auto insurance Big Data, Mohamed Hanafy, Ruixing Ming (2021)

[18] Vehicle Health Monitoring and Analysis May 2016, Parmesh K. R

[19] A Vehicle Health Monitoring System Evaluated Experimentally on a Passenger Vehicle October 2006, Hok K. Ng, Robert H. Chen, Jason Speyer

[20] Motor Insurance Claim Status Prediction using Machine Learning Techniques, Endelew Alamir, Teklu Urgessa, Ashebir Hunegnaw, Tiruveedula Gopikrishna

[21] Prediction of motor insurance claims occurrence as an imbalanced machine learning problem Sebastian Baran, Przemys law Rola (2022)

[22] A proposed model to predict auto insurance claims using machine learning techniques, shady abdelhadi, khaled A. Elbahnasy, mohamed M. Abdelsalam

[23] A Review on Vehicle Health Monitoring System Gaffar Gulag Momin (2020)

[24] Vehicle Health Monitoring System E. Chandrashekhar (2021)

[25] Research of the System of Remoting Vehicle Insurance Estimation Ming-Song LIN, Cun-Xiang LIU (2016)

[26] A review of the motor vehicle insurance industry in India Uma Shankar Yadav, Ravindra Tripathi, Mano Ashish Tripathi (2022)

[27] Prediction of Customers' Interests in Vehicle Insurance Jacob Mehlman, Jiahui Wu, Hanshen Jing, Jingyi Qi

Appendix (Source code and materials)

Appendix – A Source code

- **fyp.sql**

```
SET SQL_MODE = "NO_AUTO_VALUE_ON_ZERO";
START TRANSACTION;
SET time_zone = "+00:00";


/*!40101 SET
@OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
/*!40101 SET
@OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULT
S */;
/*!40101 SET
@OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION
*/;
/*!40101 SET NAMES utf8mb4 */;


--
-- Database: `fyp`
--


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


--
-- Table structure for table `admin`
--

CREATE TABLE `admin` (
  `id` int(11) NOT NULL,
  `name` varchar(50) NOT NULL,
  `email` varchar(100) NOT NULL,
```

```
  `password` varchar(255) NOT NULL,
  `seller_or_admin` varchar(100) NOT NULL,
  `status` varchar(255) NOT NULL,
  `seller_id` int(11) NOT NULL,
  `cnic` varchar(100) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;


--
-- Dumping data for table `admin`
--


INSERT INTO `admin` (`id`, `name`, `email`, `password`, `seller_or_admin`,
`status`, `seller_id`, `cnic`) VALUES
(1, 'yuvraj', 'c.yuvisingh2002@gmail.com',
'$2y$10$kFC5/Zluyy17I7UuoOe/y.lWjf8dwAHcQ/pHX8Ts6yAgxZZzzOymy',
'admin', '', 0, '21231234'),
(35, 's', 's@s.com',
'$2y$10$nqy30lU6enPk3t75pNMace5pj4m0wFNo93YtTR8yXmnJ5w7t4FRFC
', 'seller', '', 17, '35202-4003746-5'),
(36, 'manger', 'm@m.com',
'$2y$10$MlqQHRa6CVV2i4rXKSWuA.lNbhtGRthYkGlGG35ww4RidrE1Nul
lC', 'manager', '', 0, '232'),
(37, 'add', 'aaa@a.com',
'$2y$10$NtM9pUkrZXGM6JFjjRhRi.hK0j1F4WCO6IK8lw3qyKUpUDX8MQ
Qua', 'clerk', '', 0, '2123'),
(38, 'aaaa', 'aaaa@a.com',
'$2y$10$cRc/DRYDICSOXYfQuK2IEu3Xii6TC2O5D6NmsrxIMW8Y6EZ5qq
DIS', 'vender', '', 0, '2123'),
(39, 'add', 'usman@u.com',
'$2y$10$laleR.uZ1NevU1BE8q4lx./SuhLkpxJidXTtumJs8kBTPajkjjQ8.',
'vender', '', 0, '2123'),
(40, 'aaaaaaa', 'ab@ab.com',
'$2y$10$r/GKHJmr5I/gBm39ygPG.Oqgr8/l5H3/aRJS3/qwqbXtNWbvXgoLK',
'vender', '', 19, '35202-4493736-4');


-- --------------------------------------------------------
```

```
--
-- Table structure for table `brands`
--

CREATE TABLE `brands` (
  `brand_id` int(100) NOT NULL,
  `brand_title` text NOT NULL,
  `admin_id` varchar(100) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

--
-- Dumping data for table `brands`
--

INSERT INTO `brands` (`brand_id`, `brand_title`, `admin_id`) VALUES
(1, 'BMW', '3'),
(2, ' Ferrari', '3'),
(3, 'Alfa Romeo', '3'),
(4, ' Aston Martin', '3'),
(8, 'Lamborghini', ''),
(11, 'Maserati', ''),
(14, 'Porsche', ''),
(15, 'Honda', ''),
(16, 'Mercedes-Benz', ''),
(17, 'Toyota', ''),
(19, 'XYZZ', '1'),
(21, 'Pagani', '1');

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

--
-- Table structure for table `cart`
--
```

```
CREATE TABLE `cart` (
  `id` int(10) NOT NULL,
  `p_id` int(10) NOT NULL,
  `ip_add` varchar(250) NOT NULL,
  `user_id` int(10) DEFAULT NULL,
  `qty` int(10) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

--
-- Dumping data for table `cart`
--

INSERT INTO `cart` (`id`, `p_id`, `ip_add`, `user_id`, `qty`) VALUES
(20, 45, '', 18, 1);

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

--
-- Table structure for table `categories`
--

CREATE TABLE `categories` (
  `cat_id` int(100) NOT NULL,
  `cat_title` text NOT NULL,
  `admin_id` varchar(100) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

--
-- Dumping data for table `categories`
--

INSERT INTO `categories` (`cat_id`, `cat_title`, `admin_id`) VALUES
```

```
(1, 'Tyres and Wheels', '3'),
(2, 'Tools', '3'),
(3, 'Gadgets', ''),
(4, 'Oil', '3'),
(6, ' Brakes', ''),
(7, 'Lights', '1'),
(9, 'Audio OR Video', '');

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

--
-- Table structure for table `orders`
--

CREATE TABLE `orders` (
  `order_id` int(100) NOT NULL,
  `user_id` int(100) NOT NULL,
  `product_id` int(100) NOT NULL,
  `payment_status` varchar(200) NOT NULL,
  `qty` int(100) NOT NULL,
  `order_status` varchar(100) NOT NULL,
  `edit_status` varchar(100) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

--
-- Dumping data for table `orders`
--

INSERT INTO `orders` (`order_id`, `user_id`, `product_id`, `payment_status`,
`qty`, `order_status`, `edit_status`) VALUES
(8, 16, 40, 'cash_on_delivery', 2, 'process_order', 'uncomplete'),
(9, 16, 42, 'cash_on_delivery', 1, 'process_order', 'uncomplete'),
(10, 16, 43, 'cash_on_delivery', 1, 'process_order', 'uncomplete'),
```

```
(11, 16, 46, 'cash_on_delivery', 1, 'process_order', 'uncomplete'),
(12, 16, 46, 'cash_on_delivery', 1, 'process_order', 'uncomplete'),
(13, 16, 43, 'cash_on_delivery', 1, 'process_order', 'uncomplete'),
(14, 16, 44, 'payment', 1, 'process_order', 'uncomplete'),
(15, 16, 44, 'cash_on_delivery', 1, 'process_c', '*AD*usman gujjar1'),
(16, 16, 48, 'cash_on_delivery', 1, 'process_order', 'uncomplete'),
(17, 16, 57, 'cash_on_delivery', 2, 'process_order', 'uncomplete'),
(18, 20, 59, 'cash_on_delivery', 1, 'process_order', 'uncomplete');

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

--
-- Table structure for table `payment`
--

CREATE TABLE `payment` (
  `id` int(100) NOT NULL,
  `user_id` varchar(100) NOT NULL,
  `account_name` varchar(100) NOT NULL,
  `card_no` varchar(100) NOT NULL,
  `card_H_N` varchar(100) NOT NULL,
  `exp_date` varchar(100) NOT NULL,
  `cvc` varchar(100) NOT NULL,
  `address` varchar(300) NOT NULL,
  `Total_Bill` int(11) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

--
-- Dumping data for table `payment`
--

INSERT INTO `payment` (`id`, `user_id`, `account_name`, `card_no`,
`card_H_N`, `exp_date`, `cvc`, `address`, `Total_Bill`) VALUES
```

```
(1, '12', 'VISA', '1234567891231', 'ali', '12-2020', '123', 'lhr-OR-lhr', 0),
(2, '16', 'VISA', '1234567890123', 'ussss', '12-12-2022', '1231', 'qqqqqq-OR-sfa', 0);

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

--
-- Table structure for table `products`
--

CREATE TABLE `products` (
  `product_id` int(100) NOT NULL,
  `product_cat` int(11) NOT NULL,
  `product_brand` int(100) NOT NULL,
  `product_title` varchar(255) NOT NULL,
  `product_price` int(100) NOT NULL,
  `product_qty` int(11) NOT NULL,
  `product_desc` text NOT NULL,
  `product_image` text NOT NULL,
  `product_keywords` text NOT NULL,
  `product_condition` varchar(100) NOT NULL,
  `pro_user_add_id` varchar(100) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

--
-- Dumping data for table `products`
--

INSERT INTO `products` (`product_id`, `product_cat`, `product_brand`, `product_title`, `product_price`, `product_qty`, `product_desc`, `product_image`, `product_keywords`, `product_condition`, `pro_user_add_id`) VALUES
(57, 6, 1, 'XYZ', 150, 10, 'ASDCY IUFVGVC', '1676733312.JPG', 'BRK HKJH KJLK LKJL', 'New_product', '1'),
```

```
(58, 2, 1, 'AC', 13000, 10, 'AC OF TYPE ...xyz', '1212402198.26', 'AC tools ', 'New_product', '1'),
(59, 6, 3, 'ABC PRODUCT NAME OF EXAMPLE', 150, 12, 'assadaXX', '821114285.png', 'AKDKK BR NFSAKF', 'Old_product', '38');

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

--
-- Table structure for table `tbl_whish_list`
--

CREATE TABLE `tbl_whish_list` (
  `id` int(11) NOT NULL,
  `member_id` int(11) NOT NULL,
  `product_id` int(11) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;

--
-- Dumping data for table `tbl_whish_list`
--

INSERT INTO `tbl_whish_list` (`id`, `member_id`, `product_id`) VALUES
(7, 16, 43),
(8, 16, 46),
(9, 16, 42);

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

--
-- Table structure for table `user_info`
--

CREATE TABLE `user_info` (
```

```
  `user_id` int(10) NOT NULL,
  `first_name` varchar(300) NOT NULL,
  `last_name` varchar(300) NOT NULL,
  `email` varchar(300) NOT NULL,
  `password` varchar(300) NOT NULL,
  `mobile` varchar(300) NOT NULL,
  `address1` varchar(300) NOT NULL,
  `address2` varchar(300) NOT NULL,
  `user_type` varchar(300) NOT NULL,
  `cnic` varchar(300) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

--
-- Dumping data for table `user_info`
--

INSERT INTO `user_info` (`user_id`, `first_name`, `last_name`, `email`,
`password`, `mobile`, `address1`, `address2`, `user_type`, `cnic`) VALUES
(16, 'us', 'us', 'c@c.com', '25f9e794323b453885f5181f1b624d0b',
'0032448XXX', 'qqqqqq', 'sfa', 'customer', '35202-4293748-5'),
(17, 's', 's', 's@s.com', '25f9e794323b453885f5181f1b624d0b', '0032448XXX',
'aaaaa', '', 'seller', '35202-4393746-5'),
(18, 'aaaa', 'aaaa', 'aaaa@a.com',
'$2y$10$iJmBcWnCJJ3Mbvh.xn0djOzItULSkeli4cDG4v05Be9KvtNSEpozy',
'0032448XXX', 'aaaaaaa', 'aaajaskj', 'vender', '35202-4444974-6'),
(19, 'aaaaaaa', 'bbbbb', 'ab@ab.com',
'$2y$10$r/GKHJmr5I/gBm39ygPG.Oqgr8/l5H3/aRJS3/qwqbXtNWbvXgoLK',
'0032448XXX', 'dsafasfa asdfas', 'sadfa', 'vender', '35202-4493736-4'),
(20, 'usman', 'gujjar', 'usman@customer.com',
'25f9e794323b453885f5181f1b624d0b', '03004444444', 'lahhhh pkkk jsakjska',
'sa', 'customer', '35200-4299999-7');

--
-- Indexes for dumped tables
--
```

```
--
-- Indexes for table `admin`
--
ALTER TABLE `admin`
  ADD PRIMARY KEY (`id`),
  ADD UNIQUE KEY `email` (`email`);

--
-- Indexes for table `brands`
--
ALTER TABLE `brands`
  ADD PRIMARY KEY (`brand_id`);

--
-- Indexes for table `cart`
--
ALTER TABLE `cart`
  ADD PRIMARY KEY (`id`);

--
-- Indexes for table `categories`
--
ALTER TABLE `categories`
  ADD PRIMARY KEY (`cat_id`);

--
-- Indexes for table `orders`
--
ALTER TABLE `orders`
  ADD PRIMARY KEY (`order_id`);

--
```

```
-- Indexes for table `payment`
--
ALTER TABLE `payment`
  ADD PRIMARY KEY (`id`);

--
-- Indexes for table `products`
--
ALTER TABLE `products`
  ADD PRIMARY KEY (`product_id`),
  ADD KEY `fk_product_cat` (`product_cat`),
  ADD KEY `fk_product_brand` (`product_brand`);

--
-- Indexes for table `tbl_whish_list`
--
ALTER TABLE `tbl_whish_list`
  ADD PRIMARY KEY (`id`);

--
-- Indexes for table `user_info`
--
ALTER TABLE `user_info`
  ADD PRIMARY KEY (`user_id`);

--
-- AUTO_INCREMENT for dumped tables
--

--
-- AUTO_INCREMENT for table `admin`
--
ALTER TABLE `admin`
```

```
  MODIFY `id` int(11) NOT NULL AUTO_INCREMENT,
AUTO_INCREMENT=43;


--

-- AUTO_INCREMENT for table `brands`

--

ALTER TABLE `brands`

  MODIFY `brand_id` int(100) NOT NULL AUTO_INCREMENT,
AUTO_INCREMENT=22;


--

-- AUTO_INCREMENT for table `cart`

--

ALTER TABLE `cart`

  MODIFY `id` int(10) NOT NULL AUTO_INCREMENT,
AUTO_INCREMENT=28;


--

-- AUTO_INCREMENT for table `categories`

--

ALTER TABLE `categories`

  MODIFY `cat_id` int(100) NOT NULL AUTO_INCREMENT,
AUTO_INCREMENT=13;


--

-- AUTO_INCREMENT for table `orders`

--

ALTER TABLE `orders`

  MODIFY `order_id` int(100) NOT NULL AUTO_INCREMENT,
AUTO_INCREMENT=19;


--

-- AUTO_INCREMENT for table `payment`

--
```

```
ALTER TABLE `payment`
  MODIFY `id` int(100) NOT NULL AUTO_INCREMENT,
AUTO_INCREMENT=3;


--
-- AUTO_INCREMENT for table `products`
--
ALTER TABLE `products`
  MODIFY `product_id` int(100) NOT NULL AUTO_INCREMENT,
AUTO_INCREMENT=60;


--
-- AUTO_INCREMENT for table `tbl_whish_list`
--
ALTER TABLE `tbl_whish_list`
  MODIFY `id` int(11) NOT NULL AUTO_INCREMENT,
AUTO_INCREMENT=10;


--
-- AUTO_INCREMENT for table `user_info`
--
ALTER TABLE `user_info`
  MODIFY `user_id` int(10) NOT NULL AUTO_INCREMENT,
AUTO_INCREMENT=21;


--
-- Constraints for dumped tables
--


--
-- Constraints for table `products`
--
ALTER TABLE `products`
  ADD CONSTRAINT `fk_product_brand` FOREIGN KEY (`product_brand`)
```

```
  REFERENCES `brands` (`brand_id`),
   ADD CONSTRAINT `fk_product_cat` FOREIGN KEY (`product_cat`)
  REFERENCES `categories` (`cat_id`);
  COMMIT;


  /*!40101 SET
  CHARACTER_SET_CLIENT=@OLD_CHARACTER_SET_CLIENT */;
  /*!40101 SET
  CHARACTER_SET_RESULTS=@OLD_CHARACTER_SET_RESULTS */;
  /*!40101 SET
  COLLATION_CONNECTION=@OLD_COLLATION_CONNECTION */;
```

- **main.py**

```
"""This module creates the web page"""

# Import necessary modules.
import streamlit as st

# Import pages.
import home
import data
import plots
import predict
import about

# Import other necessary things.
from prepro import load_data

# Configure the web page.
st.set_page_config(
   page_title = 'Car Insurance Price Prediction',
   page_icon = 'car',
   layout = 'centered',
   initial_sidebar_state = 'auto'
)
```

```
# Create a dict for pages.
pages_dict = {
        "Home": home,
        "View Data": data,
        "Visualise Data": plots,


        "Predict": predict,
        "About me": about
    }

# Load the dataset.
df = load_data()

# Create navbar in sidebar.
st.sidebar.title("Navigation")
user_choice = st.sidebar.radio('Go to', ("Home", "View Data", "Visualise Data",
"Predict", "About me"))

# Open the page selected by the user.
if (user_choice == "Home" or user_choice == "About me"):
   selected_page = pages_dict[user_choice]
   selected_page.app()
else:
   selected_page = pages_dict[user_choice]
   selected_page.app(df)
```

- **home.py**

```
"""This module creates the home page."""

# Import necessary modules.
import streamlit as st
```

```
def app():
    st.title("Car Pridiction app")
    st.image("./welcome.jpg")
    st.text(
        """
        This web app allows a user to predict the prices of a car insurance based on their
        engine size, horse power, dimensions and the drive wheel type parameters.
        """
```

- **plot.py**

```
"""This create visulaise data page"""

# Import necessary module
import streamlit as st
import matplotlib.pyplot as plt
import seaborn as sns

def app(df):
    # Remove deprecation warning.
    st.set_option('deprecation.showPyplotGlobalUse', False)

    # Give title
    st.title("Visulise Data")

    # Creat a section for scatter plot
    st.header("Scatterplot")

    # Creat a mulit-select option to get x-axis from the user.
    feature_list = st.multiselect("Select x-axis values:", ('carwidth', 'enginesize', 'horsepower', 'drivewheel_fwd', 'car_company_buick'))

    for feature in feature_list:
        fig = plt.figure(figsize=(12, 5))
        st.subheader(f"Scatter plot between {feature} and price")
        sns.scatterplot(x='price', y=feature, data=df)
        st.pyplot(fig)

    # Create a section for Visualisation Selector
    st.header("Visulisation Selector")

    # Create a multiselect option to create plots or charts.
    plot_type = st.multiselect("Select charts or plots:", ('Histogram', 'Box Plot', 'Correlation Heatmap'))
```

```
    # Create plot for histogram.
    if ("Histogram" in plot_type):
        st.subheader("Histogram")
        # Take column from user.
        hist_column = st.selectbox("Select the column to create its histogram", ('carwidth', 'enginesize', 'horsepower'))
        # Plot the chart.
        fig = plt.figure(figsize=(12, 5))
        plt.title(f"Histogram for {hist_column}")

        plt.hist(x=df[hist_column], bins = 'sturges', edgecolor = 'black')
        st.pyplot(fig)

    # Create plot for boxplot.
    if ("Box Plot" in plot_type):
        st.subheader("Boxplot")
        # Take column from user.
        box_column = st.selectbox("Select the column to create its boxplot", ('carwidth', 'enginesize', 'horsepower'))
        # Plot the chart.
        fig = plt.figure(figsize=(12, 2))
        plt.title(f"Box plot for {box_column}")
        sns.boxplot(df[box_column])
        st.pyplot(fig)

    # Create plot for boxplot.
    if ("Correlation Heatmap" in plot_type):
        st.subheader("Correlation Heatmap")
        # Plot the chart.
        fig = plt.figure(figsize=(12, 10))
        ax = sns.heatmap(df.corr(), annot=True)
        bottom, top = ax.get_ylim() # Getting the top and bottom margin limits.
        ax.set_ylim(bottom + 0.5, top - 0.5) # Increasing the bottom and decreasing the bottom margins respectively.
        st.pyplot(fig)
```

- **predict.py**

```
"""This create prediction page"""

from math import sqrt
import streamlit as st
from sklearn.model_selection import train_test_split
from sklearn.linear_model import Lasso
from sklearn.metrics import r2_score, mean_absolute_error, mean_squared_log_error, mean_squared_error
```

```
def app(df):

    # Create a section for user to input data.
    st.header("Select Values:")

    # Create sliders.
    car_width = st.slider("Car Width", float(df["carwidth"].min()), float(df["carwidth"].max()))
    car_height = st.slider("Car Height", float(df["carheight"].min()),

float(df["carheight"].max()))
    engine_size = st.slider("Engine Size", float(df["enginesize"].min()), float(df["enginesize"].max()))
    horse_power = st.slider("Horse Power", float(df["horsepower"].min()), float(df["horsepower"].max()))
    peakrpm = st.slider("Peak RPM", int(df["peakrpm"].min()), int(df["peakrpm"].max()))
    wheelbase = st.slider("Wheel Base", int(df["wheelbase"].min()), int(df["wheelbase"].max()))
    curbweight = st.slider("Curb Weight", int(df["curbweight"].min()), int(df["curbweight"].max()))
    boreratio = st.slider("Bore Ratio", int(df["boreratio"].min()), int(df["boreratio"].max()))
    stroke = st.slider("Stroke", int(df["stroke"].min()), int(df["stroke"].max()))
    compressionratio = st.slider("Compression Ratio", int(df["compressionratio"].min()), int(df["compressionratio"].max()))
    citympg = st.slider("City Milage", int(df["citympg"].min()), int(df["citympg"].max()))
    highwaympg = st.slider("Highway Milage", int(df["highwaympg"].min()), int(df["highwaympg"].max()))
    # Creat two radio selection for 0 1 input.
    drivewheel_fwd = st.radio("Is if forward drive wheel car?", ("Yes", "No"))
    car_company_buick = st.radio("Is the car manufactured by Buick?", ("Yes", "No"))

    # Modify radio data.
    if (drivewheel_fwd == "Yes"):
        drivewheel_fwd = 1;
    else:
        drivewheel_fwd = 0;

    if (car_company_buick == "Yes"):
        car_company_buick = 1;
    else:
        car_company_buick = 0;
```

```
    # Create a list of all input.
    feature_list = [[car_width,car_height, engine_size, horse_power,peakrpm, wheelbase,citympg,highwaympg,compressionratio,stroke,boreratio,curbweight, drivewheel_fwd, car_company_buick]]

    # Create a button to predict.
    if st.button("Predict"):
        # Get the all values from predict funciton.
        score, pred_price = predict(df, feature_list)

        # Display all the values.
        st.success(f"The predicted price of the car: Rs.{int(pred_price):,}")
        st.info(f"Accuracy score of this model is: {score:.2%}")


@st.cache()
def predict(df, feature_list):
    # Create feature and target variable
    X = df.iloc[:, :-1]
    y = df["price"]

    # Split the data in train test.
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

    # Create the regression model
    model = Lasso()
    model.fit(X_train, y_train)

    # Store score and predicted price in a variable.
    score = model.score(X_train, y_train)
    pred_price = model.predict(feature_list)
    pred_price = pred_price[0]

    # Calculate statical data from the model.
    y_test_pred = model.predict(X_test)

   #lasso value
    k = -0000.1
    # Return the values.
    return score, pred_price*k
```

Prepro.py

```
"""This module helps to preprocess the data."""
```

```
# Import necessary modules.
import pandas as pd
import streamlit as st

# Load the dataset.
@st.cache()
def load_data():
    """This function perform preprocessing on dataset and return that"""
    # read the dataset.
    df = pd.read_csv("./car-prices.csv")

    # Extract the name of the car manufactures.
    car_companies = pd.Series([car.split(" ")[0] for car in df['CarName']], index = df.index)

    # Create new column for car company names.
    df['car_company'] = car_companies

    # Replace the misspelled car company names.
    df.loc[(df['car_company'] == "vw") | (df['car_company'] == "vokswagen"), 'car_company'] = 'volkswagen'
    df.loc[df['car_company'] == "porcshce", 'car_company'] = 'porsche'
    df.loc[df['car_company'] == "toyouta", 'car_company'] = 'toyota'
    df.loc[df['car_company'] == "Nissan", 'car_company'] = 'nissan'
    df.loc[df['car_company'] == "maxda", 'car_company'] = 'mazda'
    df.drop(columns= ['CarName'], axis = 1, inplace = True)
    cars_numeric_df = df.select_dtypes(include = ['int64', 'float64'])
    cars_numeric_df.drop(columns = ['car_ID'], axis = 1, inplace = True)

    # Map the values for the doornumbers and cylindernumbers.
    df[['cylindernumber', 'doornumber']] = df[['cylindernumber', 'doornumber']].apply(num_map, axis = 1)

    # Create dummy for car_body.
    car_body_dummies = pd.get_dummies(df['carbody'], dtype=int)

    # Create dummy for car_body and drop first column
    car_body_new_dummies = pd.get_dummies(df['carbody'], drop_first=True, dtype=int)

    # Creating a dataframe for non-numerical values.
    car_catagorical_df = df.select_dtypes(include=[object])

    # Get dummy for catagorical columns.
    cars_dummy_df = pd.get_dummies(car_catagorical_df, drop_first=True, dtype=int)
```

```
    # Drop catagorical columns from the dataset.
    df.drop(list(car_catagorical_df.columns), axis=1, inplace=True)

    # Concat df with car_dummy_df
    df = pd.concat([df, cars_dummy_df], axis=1)

    # Drop the car_ID.
    df.drop('car_ID', axis=1, inplace=True)

    # Create list of final columns.
    final_col =
['carwidth','carheight','enginesize','horsepower','peakrpm','wheelbase','curbweight','
boreratio','drivewheel_fwd','stroke','compressionratio','citympg','highwaympg','car_
company_buick', 'price']

    # Return the processed dataset.
    return df[final_col]

def num_map(series):
    """This function changes nominal data to numerical data."""
    word_dict = {"two": 2, "three": 3, "four": 4, "five": 5, "six": 6, "eight": 8,
"twelve": 12}
    return series.map(word_dict)
```

Website Module

- **login.py**

```
<?php


session_start();
if(isset($_SESSION["uid"])){
    header("location:Customer_profile.php");
exit;
}
?>



<!DOCTYPE html>
<html lang="en">
    <head>
```

```
    <script src="assets/B/js/jquery2.js"></script>

    <script src="assets/functio/A_function.js"></script>

    </head>
    <body class="gradl">

    <?php
include'top_head.php';
include'header_area_menu.php';

?>
      <p><br/></p>

       <div class="container-fluid ">
         <div class="row justify-content-center" >
                <div class="col-md-4"></div>
                <div class="col-md-4">


                 <h1 ><center><b>Sign In</b></center></h1>
                   <hr>
                   <div id="e_msg"></div>
                   <div id="e_msg2"></div>
                                  <form onsubmit="return false" id="login">
                                        <label for="email">Email</label>
                                        <input type="email" class="form-control"
name="email" id="email" required/>
                                        <label for="email">Password</label>
                                        <input type="password" class="form-control"
name="password" id="password" required/>
                                        <p><br/></p>
```

```
                                        <a href="forgot.php" style="color:#333; list-
style:none;">Forgotten Password</a>

                    <input type="submit" class="btn btn-lg btn-success"
style="float:right;" Value="Login" onclick="reloadPage();">
                                        <!---
                                        <div><a href="Registration_C_S.php">Create a
new account</a></div>
                    --->
                                    </form>

                <br>
                <br>
                <p>
                <a  href="login3231xp27w.php">
                        <button  class="btn btn-sm btn-success" > Admin Account Login
Click Here </button>
                    </a>
                </p>


        </div>

      <div class="col-md-4"></div>

    </div>
  </div>

      <!------
      <div class="container-fluid ">
        <div class="row">
              <div class="col-md-2"></div>
              <div class="col-md-8" id="signup_msg">
```

```
            </div>
            <div class="col-md-2"></div>
        </div>
        <div class="row">
            <div class="col-md-4"></div>
            <div class="col-md-4">
                <div class="panel panel-primary">
                    <div class="panel-heading">Customer Login Form</div>
                    <div class="panel-body">
                <div id="e_msg"></div>

                        <form onsubmit="return false" id="login">
                            <label for="email">Email</label>
                            <input type="email" class="form-control"
name="email" id="email" required/>
                            <label for="email">Password</label>
                            <input type="password" class="form-control"
name="password" id="password" required/>
                            <p><br/></p>

                            <a href="#" style="color:#333; list-
style:none;">Forgotten Password</a>
                    <input type="submit" class="btn btn-success" style="float:right;"
Value="Login" onclick="reloadPage();">

                            <div><a href="Registration_C_S.php">Create a
new account</a></div>
                        </form>
                </div>

            </div>
        </div>
        <div class="col-md-4"></div>
```

```
  </div>
</div>
   ---------->

</body>
```

report

ORIGINALITY REPORT

7%	5%	1%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	seventablets.com Internet Source	2%
2	docplayer.net Internet Source	1%
3	Submitted to King's Own Institute Student Paper	1%
4	www.researchgate.net Internet Source	1%
5	statistikhessen-blog.de Internet Source	<1%
6	Submitted to Multimedia University Student Paper	<1%
7	www.coursehero.com Internet Source	<1%
8	Submitted to Hatboro-Horsham High School Student Paper	<1%
9	Submitted to University College Birmingham Student Paper	<1%

10	Submitted to Metropolia Ammattikorkeakoulu Student Paper	<1 %
11	Submitted to University of Queensland Student Paper	<1 %
12	baseballblowout.com Internet Source	<1 %
13	www.slideshare.net Internet Source	<1 %
14	Submitted to University of Greenwich Student Paper	<1 %
15	Submitted to 87988 Student Paper	<1 %
16	trap.ncirl.ie Internet Source	<1 %
17	dspace.univ-tlemcen.dz Internet Source	<1 %
18	Ravi Suppiah. "Motor-imagery analysis without the use of sensorimotor cortex signals", 2015 10th International Conference on Information, Communications and Signal Processing (ICICS), 2015 Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

VEHICLE HEALTH ANALYSIS AND SPARE PARTS E-COMMERCE

Avinash reddy Vasipalli[1] Yuvraj Singh Chauhan[2]

SRM University, Kattankulathur, 603 203

Abstract

Development of automotive industry started in in early 21st century leading to modern solutions to many traditional problems. This development gave rise to data sets that are multidimensional and contain data of automotives starting from a new model to a model dated in 1980's as well. Even with this development, traditional methods are in use to finding solutions to these automotive leading to a lot of wastage of resources like money and time. With increase in automotive usage, better and sophisticated system is required in order to overcome these limitations and provide effective solutions. In this paper, we focus on using Machine algorithms to analysis datasets in predicting the required service in most effective manner with most basic data available. Also, this paper has analysed data to predict insurance premium and its second-hand selling price using machine learning algorithms like SVM, Random Forest and other regression methods. Addition of Ecommerce website to help purchase parts for automobiles has provided all around solution to customers.

Keywords: SVM, Random Forest, Regression, Insurance, Price, Machine Learning, E Commerce

I. Introduction

Modern world highly depends on AI to solve problems that are challenging to human existence and to make human life easy and comfortable. With today's intelligence widely used in automotive industry have helped in solving problems related to production and sales. Present industry is what we call 4th industrial revolution in which physical and digital systems are working hand in hand to provide a better environment that produces data that can be transmuted easily. Maintenance is of many types in automotive industries but majorly exist 3 types:

1. Preventive – when a Fault is occurred
2. Corrective – replacement of equipment
3. Predictive – analysis of current vehicle to predict what could be the issue

Automotives are very complex and consist of components like gearbox, brakes, and engines and these have sensors and actuators that connected with Engine Control Unit make it possible for it to run without any issue. There exist many theories and diagnostic tools that can communicate with ECU in which 2 best are UDS and OBD2. Due to increase in complexity, focus moved to data analysis where in wireless communication and android applications provided the most cost-effective way to diagnose automotives. For this, we propose a diagnostic using the most basic data that is available on cars and elements like second hand cost prediction and insurance premium prediction. This paper uses these data and machine learning algorithms and web application of E commers to help customers to buy genuine products. Major advantage of this paper is its all-round service to vehicle owners to ensure their time and money is saved.

Proposed machine learning algorithm; Section IV Existing Application; Section V Second hand cost prediction; Section VI Premium prediction for insurance; Section; Section VII E Commerce Architecture; Section VIII Result and discussions; Section IX Challenges and Conclusion; Section X References

II. Related Work

Automotive industry has used likes of technology in many areas in last few years to improve vehicle health and its run time. With existing technology along with Machine learning, analysis and computing has increased this possibility. Another paper used sensor data, Machine learning algorithms and android applications to analyse data to detect faults where in VMMS (Figure 1) heled in in prediction by Uferah Shafi. With increase in data, new models like Unsupervised techniques were used in papers by Prytz et al. in which relations between sensor data were found in two rounds; first round, good relation (MSE); second round used LASSO and least error. Liner Regression method by Alzghoul et al. on fault detection and shown regression Performance is better than classification. helped in dealing with data that contains repair records and sensor data which is imbalance to learn rules which outperformed many classifiers.

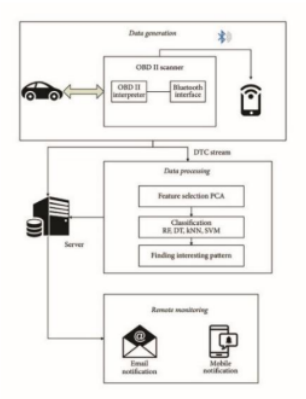


Figure 1 VMMS Architecture

A paper introduced {Figure 2} and Auto-Associative Neural Networks (ANFIS) but could test on 10 vehicles only and the same team with a new approach by detecting faults using historical data in online fashion. Furthermore, where residual between incoming and outgoing data is learnt; Second phase using

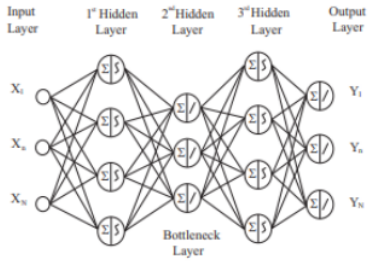


Figure 2 Architecture of AANN

Generative adversarial network (GAN) a technique to prevent failures of vehicle components. This showed a way to develop Predictive Maintenance. Models using mathematics and deep learning in automotives that outperformed KNN and ANN-BP in fault diagnosis. By using ANN method fault detection and diagnose is done based on AC-DC converter. Also, To reduce cost while collecting data Choi et al. presented using different technique. Paper presented by Mohammed Al-Zeyadi using 3000 different diagnostic fault types. Furthermore, papers by Manakov et al, Ruddle et al, Hodge et al, Ganesan and Mydhile on various aspects of vehicle health diagnosis has helped the research process. Android application condition was reported by this application.

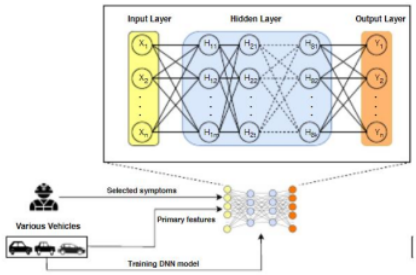


Figure 3 Architecture of DNN Model

Paper on application data of "Keep the Machine running" has used all the data of vehicles that used this application for maintenance and other usage to and concluded that predictive maintenance gains increasing importance in ensuring vehicle availability and stated that superior results were made possible in

maintenance process. It also started the importance of predictive maintenance in terms of equipment and repair cost relation (Figure 4)

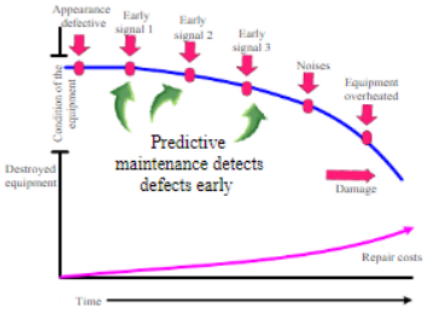


Figure 4 Importance of predictive maintenance

Although these papers and studies provided important insights in fault diagnostics assessment, but these does not consider multiple types of faults. This paper focus on premium prediction, second vale and diagnosis of fault parts and e commerce support to automotive.

III. Proposed Methodology

Many papers suggested different type of methodologies and made applications that show good accuracy. In this paper we propose using 2 learning algorithms from deep learning. Idea to use these came from a paper "Deep Learning Towards Intelligent Vehicle Fault Diagnosis" where deep learning techniques AANN and ANFIS were used. Similarly, we propose the use of ANFIS and Restricted Boltzmann Machine wherein we believe they give better results to the input data. Let us understand the proposed algorithms:

- Adaptive Network based Fuzzy Inference system (ANFIS)

This network used 5 layers to analysis and give output. As shown in Figure 5 the first hidden layer maps input variable to each membership function. In second layer T-norm is applied to calculate antecedents of rules and third layer is used in normalization of rules strengths. In fourth layer consequents of rules are determined. The last layer or output layer gives out global output after calculating by summation of signals taken in.

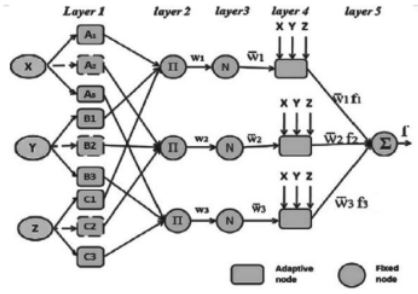


Figure 5 Architecture of ANFIS

whose output is product of incoming signals

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y) \quad i = 1, 2 \tag{3}$$

Each node output gives out firing strength of rule. Every node in 3rd layer is fixed N node. strength

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1, 2 \tag{4}$$

In 4th −layer every node i is an adaptive node with a node function

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i(p_i x + q_i y + r_i) \tag{5}$$

these are the consequent parameters. The single node in 5th−layer is a fixed _-node, which computes the overall output as the summation of all incoming signals

$$f = O_{5,i} = \Sigma_i \overline{w}_i f_i = \frac{\Sigma_i w_i f_i}{\Sigma_i w_i} \tag{6}$$

- Restricted Boltzmann Machine

A variant of Boltzman Machine in which neurons are present in input layer and hidden layer encompasses symmetric connection. In restricted Machine connection to same layer type is not allowed so no two neurons in input layer or hidden layer can not connect to each other. This rule is not applicable in case of hidden and visible layer. This Machine does not have any output layer but adjects the weight and see how our prediction is accurate. By using

this machine, we can verify the output we received from ANFIS. The energy function of this machine is:

$$E(v, h) = -a^T v - b^T h - v^T W h$$

Some major applications include:

- Filtering
- Classification
- Risk Detection
- Feature Learning
- Economic analysis

To understand RBM we need to look at 2 phases:

- Phase 1

By activation of hidden layer and concepts of weights we take in input layers and this is called Feed Forward Pass. In this we identify both positive and negative associations. A positive is when visible and hidden unit is positive and negative is vice versa.

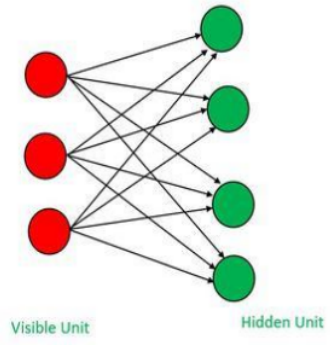


Figure 7 RBM Layers

- Phase 2

Since no output we reconstruct the input layer using activated hidden state and this is process is Feed Backward Pass in which we backtrack input layer through activated hidden neuron. This helps in reconstruction of input through activated hidden state.

Equation:

Error = Reconstructed input layer – actual input layer

Adjust Weight = Input*error*learning rate (0.1)

This process will get pattern that is responsible to activate hidden neurons as displayed in Figure 7.

IV. Existing Applications

For this section let us consider some already existing solutions in the market.

1) Intuceo – Predictive Maintenance Solution

Intuceo is wealth of in vehicle sensor data along with machine learning to provide predictive maintenance solution for OEMs and dealers. This solution claims to transform and analysis the in-vehicle sensor data, allowing customers to deploy this solution in manufacturing industry to reduce downtime and cost. A diagram to understand the method is shown in Figure 8.

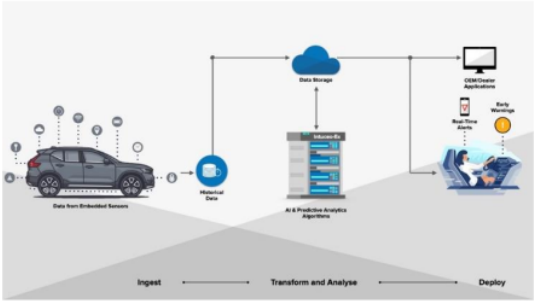


Figure 8 Intuceo method overview

to predict analysis wherein this solution provided a mobility service to work even when people present in car or outside car. This model helps in communicating with vehicle via user interface HMI or external application. Additionally, cameras and sensors monitor this health of car components with AI analysing and recommending faults to drivers.

centre has come up with a AI solution to understand sounds by faulty components. This method is done by extracting various parts of fully functioned engines and train a model. These sound when processed, analysed, and categorized to add

them toa database to teach a model. Once the database is grown, we provide the system to test and diagnose the sound and provide a solution for the faulty sound. Figure 9 shows the diagnosis steps followed to detect these faults using AI.

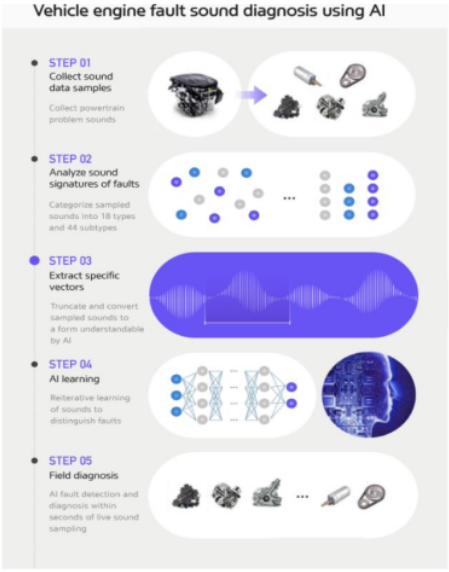


Figure 9 Sound Base Fault diagnosis

(Source: Hyundai)

Special Features of this application is:

- Improves availability to 10%
- Life of vehicle and parts by increase by 15%
- TCO reduction by 20+%

V. Cost Prediction

Vehicle cost is set by companies based on features provided in the car and market demand. In a situation where an old car should be sold, with traditional methods the involvement of middle men and agents' companies has a drastic effect on getting the correct price. The service fee to these agents and their way of dealing is something we never know and must adjust even if it causes a lose to our pockets. To avoid this, we propose a AI system that helps predict cost to a car to help resell. This process is done with the help of based data taken from vehicle. This step helps us approximate the selling price of vehicle based on No of previous owners, Kilometres driven, engine capacity, horsepower, fuel type, year it was brough.

For this part we referred to papers online and found some reference to them. The first paper we referred is "Car Price Prediction using Machine Learning Techniques" in which they named few SVM models present as related work and compared SVM and ANN models for cost prediction. After analysis they concluded that SVM is better for Cheap and expensive while ANN for Moderate (Figure 10). And the use of GUI and Java for User application made it possible for prediction to other inputs and stated an 87.38% accuracy.

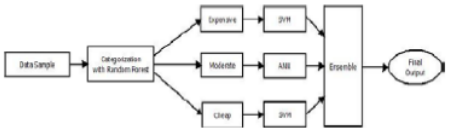


Figure 10 Prediction Model

The next paper is "Car Value Prediction Using Machine Learning (Prediction Using Random Forest Algorithm)" where the paper compared the accuracy for algorithms like Liner, Lasso, Ridge, XGBoost and Random Forest Regression. The proposed model is done in 6 steps as described in Figure 11. The dataset used was from open source named Kaggle. They stated that among all the algorithms used Random Forest regression has the best accuracy with 97.04%. After a clear mapping and plotting they believed random forest is best for this and made a Web application to provide an estimated cost for the inputs we provide.

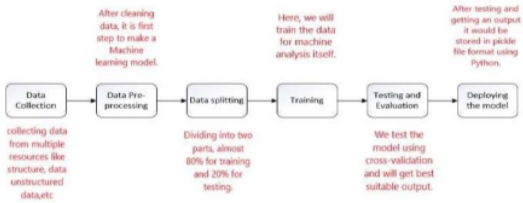


Figure 11 Model Proposed

Like random forest we propose the use of Gradient Boosting which ensembles method that creates many decisions trees. This model helps in adjusting the value of the vehicle if them exist an error or defect in the prediction. This algorithm is trained by the error difference it faces from previous results (Figure 12). This helps in building a better output from small datasets. The major drawback of this model is the time it takes as making a tree is time consuming.

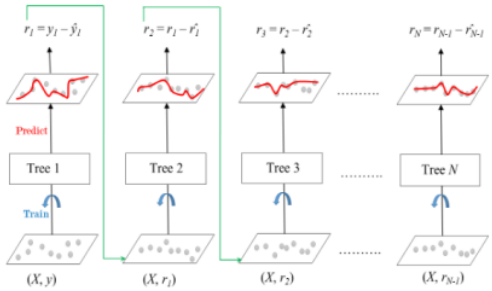


Figure 12 Gradient Boost trees for regression

With addition of ANN and SVM to this model we believe it is possible to predict the cost with most accuracy. (Figure 13). After through study we believe the better the data cleaning and no of inputs prediction performance increases. Even with complex cases with this model we believe 70% accuracy is possible. Therefore, multi algorithm approach is always the best approach in this prediction module as the better the refinement is the better, we can expect the outcome.

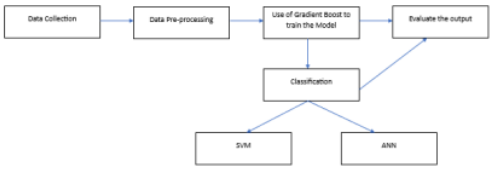


Figure 13 Structure of Proposed model

VI. Premium prediction for insurance

Insurance helps people in many ways these with growing population health insurance is very important for people. Similarly, to human insurance to vehicles I much needed to help us save in crisis. Un Insurance vehicle are fined in many parts of the world making it one of the most profitable markets for companies. Many players in modern era have used insurance as a business and still making profits. Importance of insurance has increased rapidly and spread to every item used by mankind from appliances to vehicles and as well. Even with this importance many rural areas use agent system to get insurance costing them a lot more than the actual amount. These people pay more in fear of agents who trap them in their weakness. To avoid this and provide a fair vale to people we made this project.

With the help of papers like [] we were able to understand the importance and the need for this project. For this project we propose a model in which premium prediction for car can be done with most basic inputs possible. Even though there exit many solutions the accuracy of these models is just around 86%. Even with methods like hyper tuning the accuracy has not improved. Some papers have accuracy for CNN and decision tree as well and the maximum attainted is 92%. For this problem we propose using 2 algorithms which are CNN and Lasso (Figure 14).

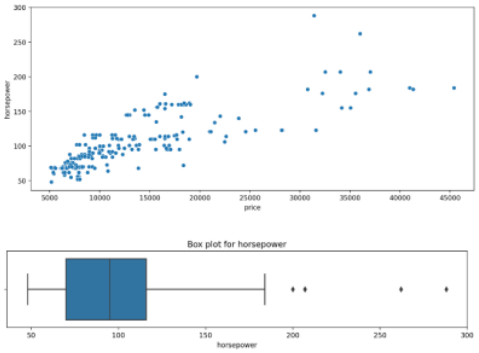


Figure 14 Plot representation of horsepower

Lasso algorithm is a derivate of Linear regression which considers all features equally important. When we have limited data and which to fully use it then we use lasso regression.

Similarly, we use of CNN has provided an 92% accuracy but their exist cases where some data is neglected or unused in this prediction to avoid this, we combine it with Lasso. This model has provided as better result compared to only CNN and the error percentage is comparatively low. For this model we developed a web application to predict the value and plot the data to visualize the data in a much better manner. (Figure 15)

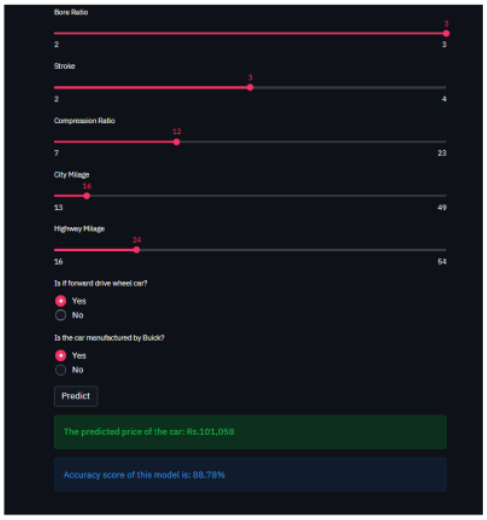


Figure 15 Premium Prediction

VII. E Commerce

The e-commerce component of the website is designed using PHP and MySQL. It includes a user-friendly interface that allows customers to browse, search, and purchase genuine spare parts for their vehicles.

The e-commerce component of the website comprises of the following components:

1. Product management module: This module allows administrators to add, edit, and delete products, as well as manage product categories and attributes.

2. Shopping cart module: This module allows customers to add products to their shopping cart, view their cart, and complete the checkout process.

3. Payment gateway module: This module integrates with a payment gateway to facilitate secure online payments.

4. Order management module: This module allows administrators to manage orders, process payments, and track shipping and delivery.

5. User account management module: This module allows customers to create and manage their user accounts, including viewing order history and updating their account information.

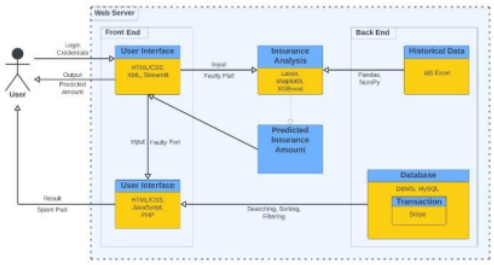


Figure 16 Architecture of E commerce

The architecture of the insurance analysis (Figure 16) component is designed using Python and Streamlit. It includes the following modules:

1. Data Collection Module: This module collects data from various sources, like customer details, vehicle details, etc.

2. Machine Learning Module: This module trains machine learning models on the pre-processed data to predict the insurance amount for a particular vehicle.

3. User Interface Module: This module provides a user-friendly interface for customers to enter their vehicle details and obtain an insurance prediction.

The architecture of the website is designed to be scalable and secure, with a modular structure that allows for easy maintenance and updates. The front-end of the website is built using Streamlit, a Python library for building web applications,

which provides a simple and intuitive interface for users to interact with the website.

The back-end of the website is built using a combination of PHP and MySQL. PHP is used to handle server-side processing and interact with the MySQL database, while MySQL is used to store product and customer information.

VIII. Result and Discussion

- The machine learning algorithms used for predicting insurance amounts provide accurate estimates based on the vehicle's health analysis.
- The spare parts E-commerce component offers a wide range of options for purchasing high-quality and authentic spare parts for vehicles.
- The website provides a user-friendly interface, making it easy for users to navigate and access the various features.

Discussions:

- The use of machine learning algorithms for insurance prediction analysis can help vehicle owners make informed decisions regarding maintenance and repair, leading to cost savings and increased safety.
- The spare parts E-commerce component offers convenience and accessibility for vehicle maintenance, which can lead to increased customer satisfaction and loyalty.
- The website's modern technology stack, including Python, Streamlit, PHP, and MySQL, offers scalability, performance, and security for handling large amounts of data and user traffic.
- The website's success depends on the quality of the data used for analysis and prediction, as well as the accuracy and reliability of the machine learning algorithms.
- The user experience and interface design of the website should be continuously evaluated and improved based on user feedback to ensure that it remains user-friendly and effective in meeting user needs.

The website would be the ultimate solution for vehicle owners. The website would be super intuitive, user-friendly, and allow customers to browse through the latest automotive parts and accessories. The vehicle health analysis feature would provide a detailed report on the health of the vehicle, and the spare parts e-commerce feature would make it easy for customers to purchase what they need, without leaving the comfort of their homes. This would save them time, money, and hassle. The website would have a state-of-the-art security system in place, ensuring that customer information and transactions are protected from any unauthorized access. All in all, this website would be a game-changer in the automotive industry, and set the standard for the future of vehicle health analysis and spare parts e-commerce.

IX. Challenges and Conclusion

Challenges:

- Data Accuracy: One of the biggest challenges in developing a website is ensuring that the data used in the analysis and prediction models is accurate and reliable.
- Data Privacy and Security: With the collection of sensitive vehicle and customer data, the website must have robust security measures in place to ensure the protection of personal information.
- Technical Complexity: The development of such a website requires a deep understanding of multiple technologies, including machine learning, web development, and database management, which can pose technical challenges.
- Integration of Multiple Modules: The website includes multiple modules, such as vehicle health analysis and spare parts e-commerce, which require seamless integration and compatibility to ensure the website's proper functioning.

Conclusion:

The website aims to provide a comprehensive solution to the customers by offering insurance analysis and spare parts e-commerce services. The website employs machine learning algorithms to predict the insurance amount for a vehicle and has a spare parts e-commerce platform for customers to purchase genuine spare parts for their vehicles.

The development of such a website can pose various challenges, including data accuracy, data privacy and security, technical complexity, and integration of multiple modules. However, with proper planning, analysis, and testing, these challenges can be overcome, and a high-quality website can be developed.

The website has immense potential for the automobile industry and can provide a significant value proposition to the customers. It can provide customers with personalized and accurate insurance predictions and a convenient platform to purchase genuine spare parts, making it a one-stop-shop for their vehicle needs.

RP

ORIGINALITY REPORT

7%	5%	3%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.geeksforgeeks.org Internet Source	1%
2	coek.info Internet Source	1%
3	ijates.com Internet Source	1%
4	Submitted to Majan College Student Paper	1%
5	Submitted to HCUC Student Paper	<1%
6	S. Hore, S. K. Das, S. Banerjee, S. Mukherjee. "An adaptive neuro-fuzzy inference system-based modelling to predict mechanical properties of hot-rolled TRIP steel", Ironmaking & Steelmaking, 2016 Publication	<1%
7	www.ijastems.org Internet Source	<1%
8	hdl.handle.net Internet Source	

		<1%
9	Submitted to Franklin University Student Paper	<1%
10	Ravi Suppiah. "Motor-imagery analysis without the use of sensorimotor cortex signals", 2015 10th International Conference on Information, Communications and Signal Processing (ICICS), 2015 Publication	<1%
11	baseballblowout.com Internet Source	<1%
12	www.researchgate.net Internet Source	<1%
13	AJITH KUMAR R. "Experimental and Numerical Simulation of Neuro Fuzzy Based Cartesian Robot for Soft Material Cutting", Journal of Applied Research and Technology, 2021 Publication	<1%
14	es.scribd.com Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		