VEHICLE HEALTH ANALYSIS AND SPARE PARTS E-COMMERCE

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

Under the guidance of
Dr. T.Karthick
Assistant Professor
Department of Data Science and Business Systems
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 [15] 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. [2] 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.

The rest of the paper is organised as follows. Section II Contains related work; Section III 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. New Algorithm named "Sequential pattern mining algorithm" in which algorithm learned patterns from warrant data and converted the patterns to rule based system by Choudhary et al. 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 "Consensus Self-Organized Model for Fault Prediction" were presented Byttner et al. 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. "BRACID" (Bottom-up Induction of Rules and Cases for Imbalanced Data) is technique that helped in dealing with data that contains repair records and sensor data which is imbalance to learn rules which outperformed many classifiers like C4.5, PART, CN2 and RIPPER in case of imbalance data. [2]

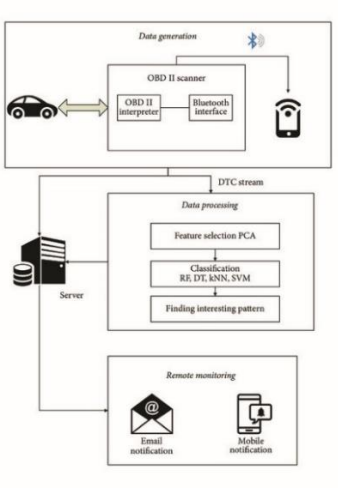

Figure 1 VMMS Architecture

A paper introduced fault detection and diagnosis by using 2 techniques, Auto-Associative Neural Networks (AANN) {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, new framework for online diagnosis and fault detection in which 2 phases are present: first phase is auto-associative neural network where residual between incoming and outgoing data is learnt; Second phase using multi class Support Vector Machine (SVM) to classify the reported faults.[16]

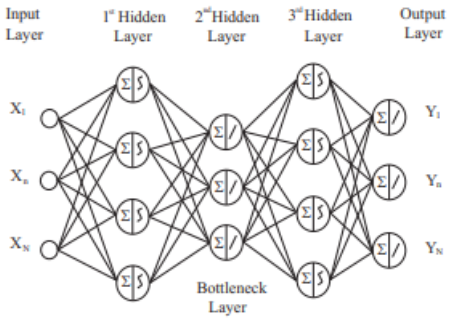

Figure 2 Architecture of AANN

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, authors in [1][16][17] presented new acoustic based fault diagnosis technique of motor. To reduce cost while collecting data Choi et al. presented using different technique. Paper presented by Mohammed Al-Zeyadi using Deep learning architecture (Figure 3) Deep-SBM to predict wide range of faults based on symptoms and was able to classify 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 by Babu et al. in which engine condition, battery condition, and emission system were monitored, and condition was reported by this application. [1]

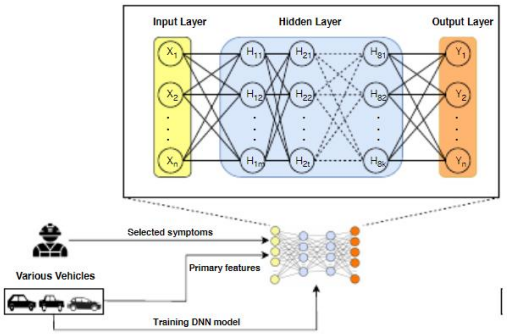

Figure 3 Architecture of DNN Model

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)

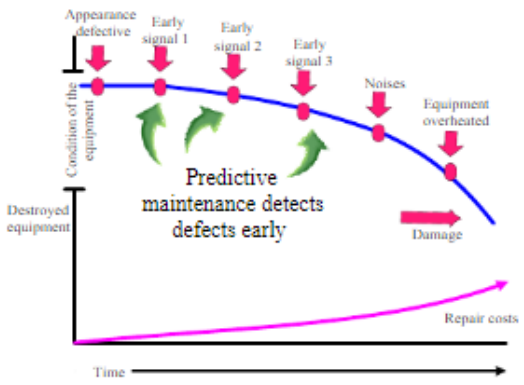

Figure 4 Importance of predictive maintenance

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. [6]

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 [16]:

- 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.

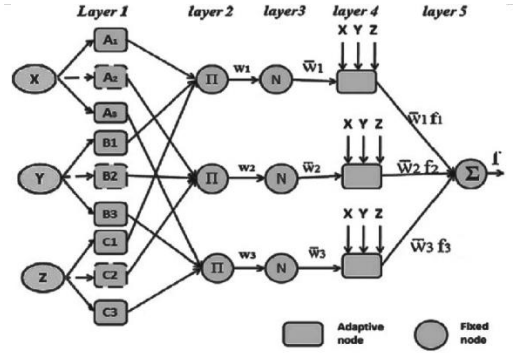

Figure 5 Architecture of ANFIS

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. Let us understand this by example, two input (x, y) and output (z) the common rule set with 2 fuzzy if-then rules is:

- Rule 1
  If (x is A_1) and (y in B_1) → $f_1 = p_1x + q_1y + r_1$
- Rule 2
  If (x is A_2) and (y in B_2) → $f_2 = p_2x + q_2y + r_2$

As shown in Figure 6 below

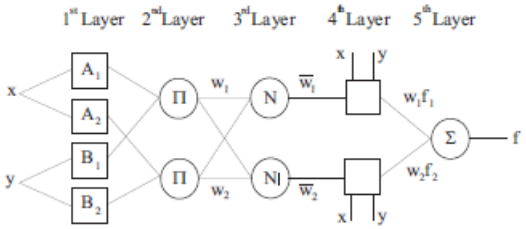

Figure 6

Every node i in ℓ^{th} layer has node function

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

Where $O_{1,i}$ is membership grade of fuzzy set A = {A_1, A_2, B_1, B_2} and this specifies degree in which given input x satisfies the quantifier A. In 2nd layer every node is fixed II – node 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. The i^{th} node calculates ration of i^{th} rules firing strength to sum of all rules firing strength

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

(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 [18]:

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

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.

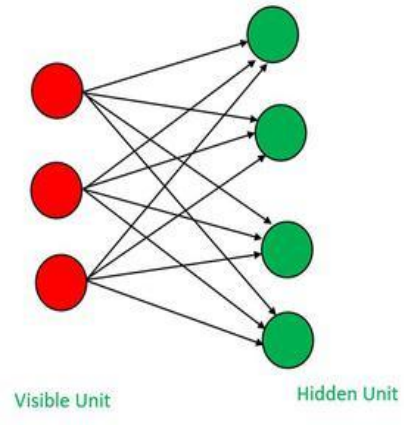

Figure 7 RBM Layers

- 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). [7] [9]

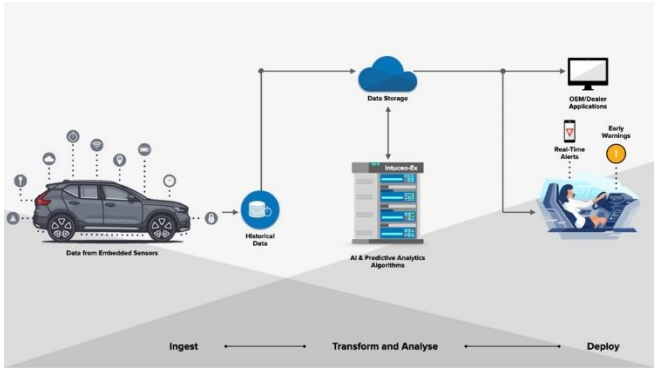

Figure 8 Intuceo method overview

2) IBM: Connected Vehicle Prediction 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. [7]

3) 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. [8]

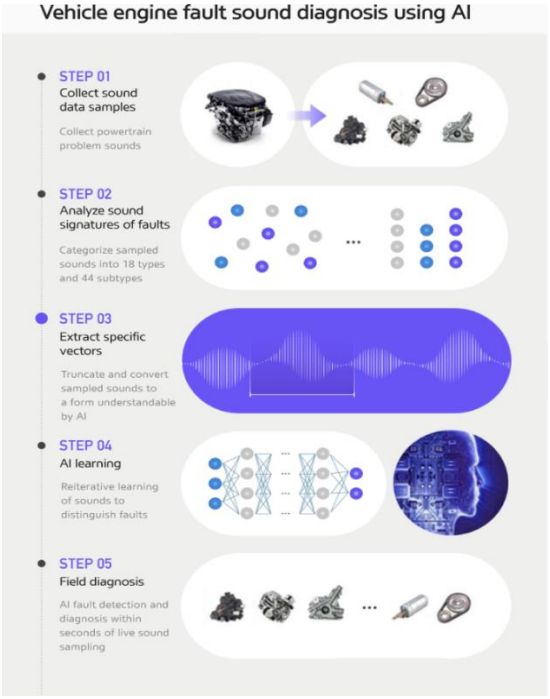

Figure 9 Sound Base Fault diagnosis

(Source: Hyundai)

Special Features of this application is:

- Improves availability to 10+%
- Increases life of vehicle and parts by 15%
- Reduces TCO 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 [14].

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. [19]

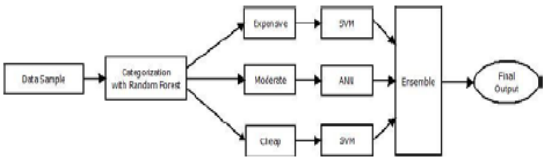

Figure 10 Prediction Model

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. [13]

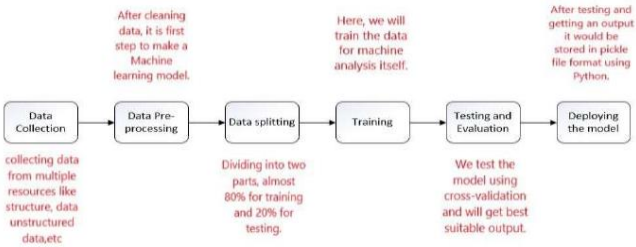

Figure 11 Model Proposed

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. [10] 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.

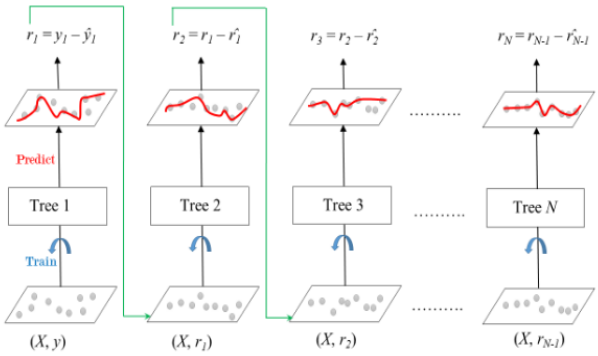

Figure 12 Gradient Boost trees for regression

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.

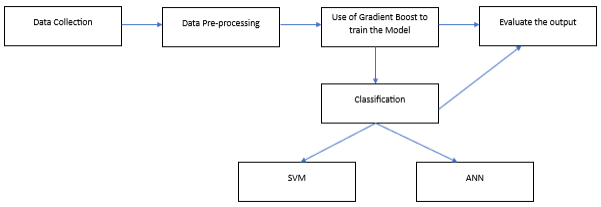

Figure 13 Structure of Proposed model

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 [10-13] 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).

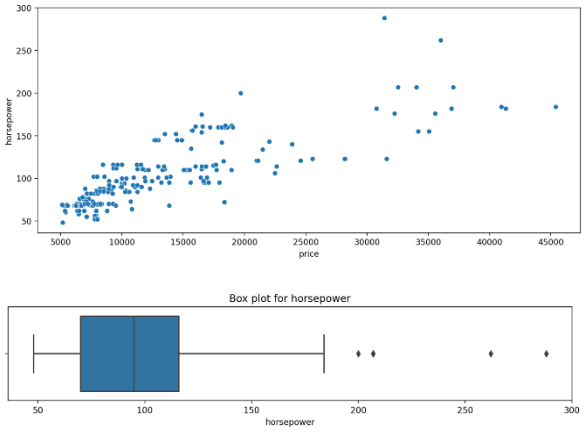

Figure 14 Plot representation of horsepower

Lasso algorithm or *Least Absolute Shrinkage and Selection Operator* 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)

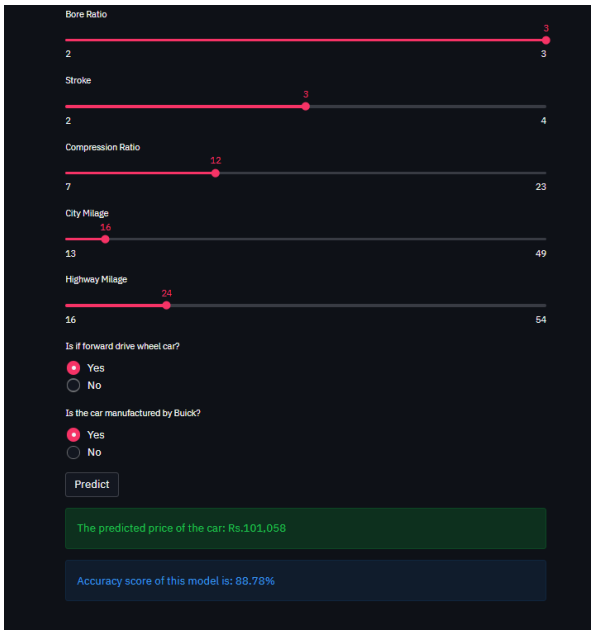

Figure 15 Premium Prediction

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.

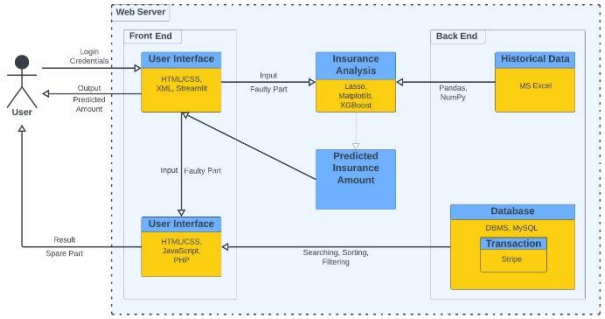

Figure 16 Architecture of E commerce

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.

X. References

1. Deep Learning Towards Intelligent Vehicle Fault Diagnosis 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. Car Price Prediction using Machine Learning Techniques *Gegic, Enis; Isakovic, Becir; Keco, Dino; Masetic, Zerina; Kevric, Jasmin*
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. A proposed model to predict auto insurance claims using machine learning techniques; shady Abdelhadi, Khaled Elbahnasy, Mohamed Abdelsalam
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 Using Telematics Data—XGBoost versus Logistic Regression. Risks 2019,7, 70
13. Car Value Prediction Using Machine Learning Prediction Using Random Forest Algorithm *Junaid Mandviwala, Mohd Mubin Girach, Yash Darekar, Yash Patel, Sarang Pandit*
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. Vehicle Fault Detection and Diagnos is combining AANN and ANFIS *⋆Juan Pablo Nieto Gonzalez ⁎Luis E Garza Casta˜non ⁎Abdelhamid Rabhi ⁎⁎ Ahmed El Hajjaji ⁎⁎Ruben Morales-Menendez ⁎*
17. Artificial neural network-based fault diagnosis in the AC–DC converter of the power supply of series hybrid electric vehicle *Seyed Saeid Moosavi, Abdesslem Djerdir, Youcef Ait-Amirat, Davood Arab Khaburi, Abdoul N'Diaye*
18. https://www.geeksforgeeks.org/restricted-boltzmann-machine/