A Project Report

on

ReCar – A Car re-sale value predictor

Submitted in partial fulfillment of the requirement for the award of the degree of

Bachelors of Technology in Computer Science and Engineering



(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of Mr. S.P Ramesh Assistant Professor

Department of Computer Science and Engineering

Submitted By:

TANMAY GUPTA

19SCSE1010220

Gaurav GUPTA

19SCSE1010460

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING GALGOTIAS UNIVERSITY, GREATER NOIDA INDIA



SCHOOL OF COMPUTING SCIENCE AND ENGINEERING GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled "**ReCar** – **A Car re-sale value predictor**" in partial fulfillment of the requirements for the award of the **Bachelor Of Technology In Computer Science And Engineering** degree submitted in the **School of Computing Science and Engineering** of **Galgotias University**,Greater Noida, is an original work carried out during the period of July – 2021 to December 2021, under the supervision of **Mr. S.P Ramesh**, **Assistant Professor, Department of Computer Science and Engineering**, Galgotias University, Greater Noida.

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

TANMAY GUPTA, 19SCSE1010220 GAURAV GUPTA, 19SCSE1010460

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Mr. S.P Ramesh, Assistant Professor

Abstract

The car manufacturers sets the price of new cars in the industry, with the government incurring some additional costs in the form of taxes. Customers who purchase a new car can be convinced that the amount they spend will be well invested. However, due to rising new car prices and customers' inability to purchase new cars due to a lack of finances, used car sales are on the rise worldwide.

In India alone used car market was valued at USD 27 billion in 2020, and it is expected to reach USD 50 billion by 2026. A extremely user friendly used car price prediction system is needed to accurately estimate the car's value based on a range of factors. Though there are websites that provide this service, their projection approach may not be the most accurate because of the profit margin they seek for themselves. The current system involves a process in which the sellers sets the price at random and the buyer has no knowledge of the car or its current value. In reality, the seller also has no idea what the vehicle is worth or what price he should be asking for it.

To address this issue, we have devised a solution that will be extremely effective. We'll examine the results of numerous machine learning algorithms, such as Linear Regression and Decision Tree Regressor, and pick the best one. Regression Algorithms are applied because they offer us with a continuous number as an output rather than a categorized value, allowing us to anticipate the real price of a car rather than its price range. A user interface will also be created that takes input from any user and displays the price of a car based on their inputs.

List of Tables

Table No.	Table Name	Page Number
1.	Table for Student Data	3
2.	Table for Faculty Data	4

List of Figures

Figure No.	Table Name	Page Number
1.	UML Diagram	7
2.	Data Flow Diagram	6

Table of Contents

Title

Candidate's Declaration Abstract List of Table List of Figures

Chapter 1	Introduction		
	1.1 Introduction		
	1.2 Formulation of Problem		
	1.2.1 Tool and Technology		
	Used		
Chapter 2	Literature Survey/Project Design		
Chapter 3	System Requirement		
Chapter 4	System Architecture		
Chapter 5	Working Of Project:		
	5.1 Algorithm Used		
	5.2 Steps Involved		
	Data Processing		
	-		
	Data FrocessingData Training		
Chapter 6	-		
Chapter 6 Chapter 7	Data Training		

Refrences

INTRODUCTION

The car manufacturers sets the price of new cars in the industry, with the government incurring some additional costs in the form of taxes. Customers who purchase a new car can be convinced that the amount they spend will be well invested. However, due to rising new car prices and customers' inability to purchase new cars due to a lack of finances, used car sales are on the rise worldwide. In India alone used car market was valued at USD 27 billion in 2020, and it is expected to reach USD 50 billion by 2026. A extremely user friendly used car price prediction system is needed to accurately estimate the car's value based on a range of factors. Though there are websites that provide this service, their projection approach may not be the most accurate because of the profit margin they seek for themselves. The current system involves a process in which the sellers sets the price at random and the buyer has no knowledge of the car or its current value. In reality, the seller also has no idea what the vehicle is worth or what price he should be asking for it. To address this issue, we have devised a solution that will be extremely effective.

Due to the numerous factors that influence the price of a used vehicle on the market, evaluating if the quoted price of a used car is accurate is a challenging task. The goal of this project is to design machine learning models that can properly forecast and determine the price of a used car based on its parameters, so that buyers can make informed choices. The car's pricing will be determined based on a number of factors. We put several learning approaches to the test on a dataset that includes the costs of various brands and models. We'll examine the results of numerous machine learning algorithms, such as Linear Regression and Decision Tree Regressor, and pick the best one. Regression Algorithms are applied because they offer us with a continuous number as an output rather than a categorized value, allowing us to anticipate the real price of a car rather than its price range. A user interface will also be created that takes input from any user and displays the price of a car based on their inputs.

Dealers that sell used cars are one of the most likely groups to be interested in the findings of this study. When the used car dealers have a better understanding of what makes a car desirable and what the most significant qualities are for a used vehicle will be able to apply this knowledge and provide better service. Online pricing systems: There are websites that provide an estimate of a car's value. They might have a good model for forecasting. Having a second model, on the other hand, may aid them in providing a better prediction to their users. As a result, the model established in this study could aid online web services that determine the market price of a used car. Individuals: Many people are interested in the used vehicle market because they wish to sell their car or buy a used car at some time in their lives. It's a major mistake to pay too much or sell for less than the market value in this process.

Because of rising new automobile prices and buyers' inability to afford them, used car sales are on the rise all around the world. As a result, there is a pressing need for a Used Car Price Prediction system that accurately analyses the car's worthiness based on a range of factors. The proposed system will assist in determining the precise price of a secondhand car.

LITERATURE SURVEY

1. <u>Title - Car Price Prediction using Machine Learning Techniques</u>

Author - Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric

A car price prediction has been a high- interest research area, as it requires noticeable effort and knowledge of the field expert. Considerable number of distinct attributes are examined for the reliable and accurate prediction. To build a model for predicting the price of used cars in Bosnia and Herzegovina, we applied three machine learning techniques (Artificial Neural Network, Support Vector Machine and Random Forest). However, the mentioned techniques were applied to work as an ensemble. The data used for the prediction was collected from the web portal <u>autopijaca.ba</u> using web scraper that was written in PHP programming language. Respective performances of different algorithms were then compared to find one that best suits the available data set. The final prediction model was integrated into Java application. Furthermore, the model was evaluated using test data and the accuracy of 87.38% was obtained. Keywords – car price prediction, support vector machines, classification, machine learning

2. <u>Titile - Predicting the Price of Used Cars using Machine Learning Techniques</u>

Author - Sameerchand Pudaruth

In this paper, we investigate the application of supervised machine learning techniques to predict the price of used cars in Mauritius. The predictions are based on historical data collected from daily newspapers. Different techniques like multiple linear regression analysis, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions. The predictions are then evaluated and compared in order to find those which provide the best performances. A seemingly easy problem turned out to be indeed very difficult to resolve with high accuracy. All the four methods provided comparable performance. In the future, we intend to use more sophisticated algorithms to make the predictions.

3. <u>Title - Predicting Used Car Prices</u>

Author - Kshitij Kumbar, Pranav Gadre, Varun Nayak

Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle's price on the market. The focus of this project is developing machine learning models that can accurately predict the price of a used car based on its features, in order to make informed purchases. We implement and evaluate various learning methods on a dataset consisting of the sale prices of different makes and models across cities in the United States. Our results show that Random Forest model and K-Means clustering with linear regression yield the best results, but are compute heavy. Conventional linear regression also yielded satisfactory results, with the advantage of a significantly lower training time in comparison to the aforementioned methods.

System Requirement

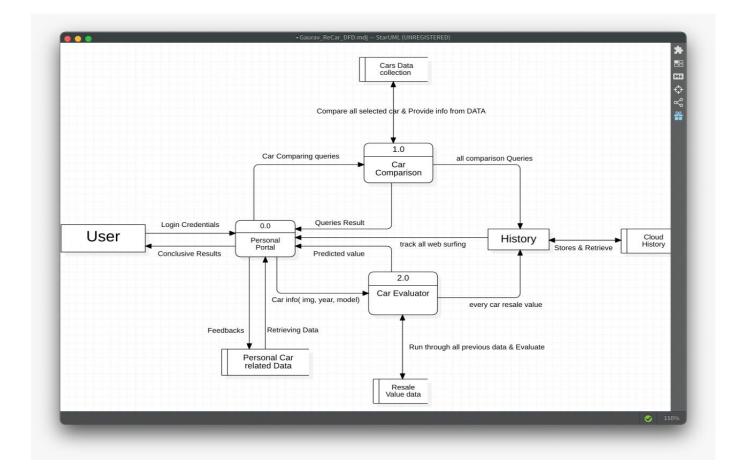
Hardware requirements:

- Operating system- Windows 7,8,10
- Processor- dual core 2.4 GHz (i5 or i7 series Intel processor
- or equivalent AMD)
- RAM-4GB

Software Requirements:

- Python
- Pycharm
- PIP 2.7
- Jupyter Notebook
- Flask
- HTML
- JavaScript
- CSS
- Cloud hosting (AWS/Azure/Heroku)

System Architecture



WORKING OF PROJECT

Algorithm Used :

Random Forest Algorithm –

During the training phase, the Random Forest algorithm constructs many decision trees, which is an Ensemble Bagging approach. The random forest chooses the ultimate option based on the majority of the outputs (trees). The key benefit of Random Forest is that it is a fast algorithm. Regression and Classification are two forms of supervised learning issues that are mixed together.

The Random Forest algorithms are used in many machine learning applications such as :

• For Remote Sensing such as ETM devices used to acquire images of earth's surface, Random Forest is the first choice as it provides Higher Accuracy in a less training time.

• For Multiclass Object Detection, Random Forest is used as it provides better detection in complicated environments.

• This method is used by several gaming consoles to track body movement and replicate it in the game. Random Forest algorithm is trained to identify the body parts and algorithm learns from it. Then it identifies the body parts of the users such as hands, feet, face, eyes, nose etc

Advantages of the Random Forest Algorithm:

1. No Overfitting and takes less training time.

2. High accuracy and runs efficiently on large databases.

3. Estimate missing values and still maintains high accuracy even having an extensive amount of missing data.

Randomised Search CV:

It is a hyper-parameter tuning method of python's scikit learn library. It is used to implement the "fit" and "score" methods, as well as to forecast the best model.

Steps Involved:

1. DATA PROCESSING

Any model that uses any algorithm must be trained before it can be used. Data preprocessing is the most important and will be the most primary phase. There are numerous milestones in the data preprocessing (steps)

Step 1: Import Libraries:

Pandas for data manipulation and analysis, Numpy for numerical analysis, Matplotlib, and Seaborn for better visuals and graphical metrics of the data were the important libraries I utilised for data preprocessing.

Step 2: Import the Dataset:

The dataset has been taken from Kaggle. The name of the dataset was 'Second hand car market'.

In [3]: df.head()

Out[3]:

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
2	ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0

Step 3: Taking care of Missing Data :

After evaluation of this dataset, I found no missing values in the dataset

checking the null values!

In [6]:	df.isnull().sum()	
	Car_Name Year Selling_Price Present_Price Kms_Driven Fuel_Type Seller_Type Transmission	0 0 0 0 0 0 0 0
	Owner dtype: int64	0

Step 4: Encoding categorical data:

This dataset contains categorical values such as fuel type, owner type, and seller type; therefore, we must encode these categorical values into an encoded format in order to better train our model. To do so, I used pandas' get Dummies() method, which converted the entire dataset's categorical values into binary values.

In [17]: fdf=pd.get_dummies(fdf, drop_first=True)

Fuel_Type_Diesel	Fuel_Type_Petrol	Seller_Type_Individual	Transmission_Manual
0	1	0	1
1	0	0	1
0	1	0	1
0	1	0	1
1	0	0	1

Step 5: Splitting the Dataset into the Training set and Test Set:

I utilized the capable machine learning library of python, scikit-learn, or sklearn to partition this dataset into Test and Train datasets to train our machine learning model. Using its model selection approach, or supervised learning, to generate testing data by selecting random values from the given dataset for model prediction.

In [39]: from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y , test_size=0.2)

Step 6: Feature Scaling:

Because all of the data is in a standard format, I don't employ any feature scaling strategies

2. DATA TRAINING AND MODELING :-

First and foremost, we must define the dependent and independent variables in order to train and construct a model. To find these variables, I first used to find the correlation between the output variables, and then I separated my variables into two axes, which we call x and y, with the x-axis containing all the independent variables and the y-axis containing the dependent variable, which in our model is the Used Cars selling price.

This dataset is further distributed in the train-test dataset using Randomized Search CV tweaking of this model to discover the optimum hyperparameters for our model prediction using the sk-learn model selection package and its train test split function.

In [47]: **from** sklearn.model selection **import** RandomizedSearchCV

```
In [48]: #Randomized Search CV
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 100, stop = 1200, num = 12)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(5, 30, num = 6)]
# max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10, 15, 100]
# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2, 5, 10]
```

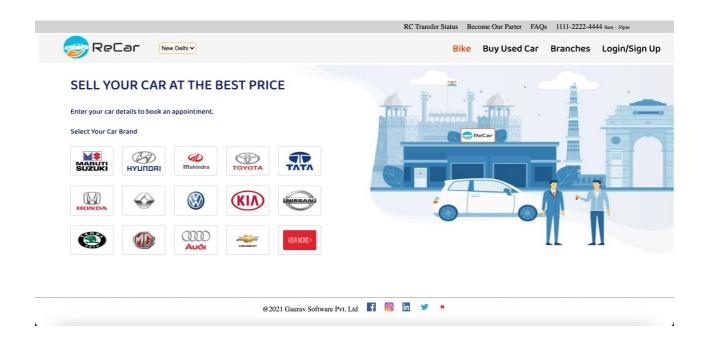
Training the model using Train test.

In [53]: rf_random.fit(x_train,y_train)

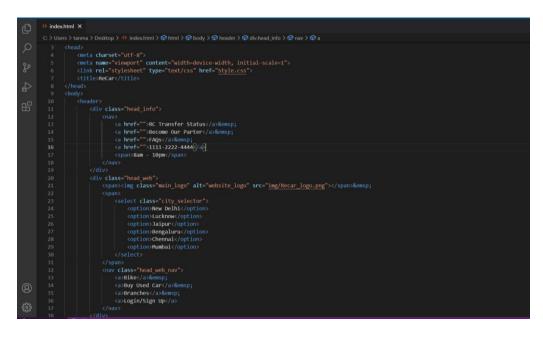
[CV] n estimators=900, min samples split=5, min samples leaf=5, max features=sqrt, max depth=10, total= 0.8s [CV] n estimators=900, min samples split=5, min samples leaf=5, m ax features=sqrt, max depth=10 [CV] n estimators=900, min samples split=5, min samples leaf=5, max features=sqrt, max depth=10, total= 0.8s [CV] n estimators=900, min samples split=5, min samples leaf=5, m ax features=sqrt, max depth=10 [CV] n estimators=900, min samples split=5, min samples leaf=5, max features=sqrt, max depth=10, total= 0.9s [CV] n estimators=900, min samples split=5, min samples leaf=5, m ax features=sqrt, max depth=10 [CV] n estimators=900, min samples split=5, min samples leaf=5, max_features=sqrt, max_depth=10, total= 0.9s [CV] n estimators=1100, min_samples_split=10, min_samples_leaf=2, max features=sqrt, max depth=15 [CV] n estimators=1100, min samples split=10, min samples leaf= 2, max_features=sqrt, max_depth=15, total= 1.1s [CV] n estimators=1100, min samples split=10, min samples leaf=2, upoc-capt may donth-1E

IMPLEENTATION

A very interactive web application has been built using JavaScript and CSS. The user interface has been designed for easy understanding and extremely user friendly.



HTML code for the front page:



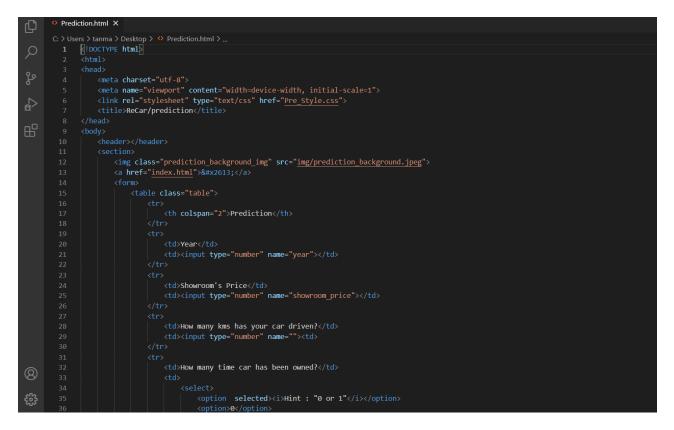
பு	index.	tml X
	C: > User	s > tanma > Desktop > 🜼 index.html > 🔗 html > 🄗 body > 🔗 section > 🖓 div.body_logos > 🔗 div.select_brand > 🔗 nav
Q		<div class="body_3">Select Your Car Brand</div>
		<div class="select_brand"></div>
ဒိုစ		<nav></nav>
6		<a href="<u>Prediction.html</u>"><img height="45px" src="<u>img/1.jpeg</u>" width="75px"/>
~		
_a⊳		
		
B		<pre>  </pre>
ш		
		<pre>\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\</pre>
	56	<pre>< information and informa</pre>
		<pre>cimg height="45px" width="75px" src="img/8.jpeg">,/a>  cimg height="45px" width="75px" src="img/8.jpeg">,/a> </pre>
		<pre>kemsp;</pre>
		
		<pre></pre> <pre></pre>
		<nav></nav>
		
		<pre> </pre>
		<a href="<u>Prediction.html</u>"><img height="45px" src="<u>img/13.jpeg</u>" width="75px"/>
		
		<a href="<u>Prediction.html</u>"><img height="45px" src="<u>img/17.png</u>" width="75px"/>
	71	<pre></pre>
		<nav></nav>
		@2021 Gaurav Software Pvt. Ltd
		
		
Ø		
		
sur s		<img <="" a="" class="utube" height="20px" src="img/utube_logo.png" width="20px"/>
583		
	80	

Form in which the user will fill the required information.

ĸ

Prediction		
Year		
Showroom's Price		
How many kms has your car driven?		
How many time car has been owned?	Hint : "0 or 1" 🗸	
Select Fuel Type	Petrol ~	
Are you Dealer or Individual	Dealer 🗸	
Select Transmission Type	Automatic ~	
Get Car Price		

Code for the form



Ð	• Prediction.html ×				
	C: > Use	s > tanma > Desktop > • Prediction.html >			
ρ		Select Fuel Type			
		<pre></pre>			
ço					
		<pre><option>Petrol</option></pre>			
		<pre><pre><pre><pre><pre><pre><pre><pre></pre></pre></pre></pre></pre></pre></pre></pre>			
å					
ß	51 52				
		Are you Dealer or Individual			
		<pre><select></select></pre>			
		<pre><option selected="">Dealer</option></pre>			
		<option>Individual</option>			
		Select Transmission Type			
		<pre></pre>			
		<pre><select <option="">Automatic</select></pre>			
		<pre><pre>coption/Hanual/option></pre></pre>			
		<button>Get Car Price</button>			
	73 74				
8	74				
\diamond	76				
~~					
5 223					

COLCLUSION

This model is built on machine learning algorithms, and we were attempting to forecast the selling price of used cars using the Kaggle dataset. We utilised machine learning techniques to predict this dataset: Random Forest and Randomized Search CV. The model's forecast is then compared to a test dataset constructed by randomly selecting values from the original dataset, and the prediction is then evaluated using several approaches. Due to the numerous factors that influence the price of a used vehicle on the market, evaluating if the quoted price of a used car is accurate is a challenging task. We have achieved the goal of this project that is to design machine learning models that can properly forecast and determine the price of a used car based on its parameters, so that buyers can make informed choice.

REFRENCES

[1] SAS Academy website -"https://www.sas.com/en_in/insights/analytics/machine learning.html"

[2] Dr M. J. Garbade – "Clearing the Confusion: AI vs Machine Learning vs Deep Learning Differences" Available: https://towardsdatascience.com/clearingthe confusion-ai-vs-machine-learning-vs-deep learningdifferencesfce69b21d5eb

[3] A. Wilson – "A Brief Introduction to Supervised Learning" Available: https://towardsdatascience.com/a-brief introduction-to-supervised-learning-54a3e3932590

[4] A. Wilson – "A Brief Introduction to Supervised Learning" Available: https://towardsdatascience.com/a-brief introduction-to-supervised-learning-54a3e3932590

[5] J. Rocca – "Ensemble methods: bagging, boosting and stacking" Available: https://towardsdatascience.com/ensemble methods-bagging-boosting-and-stacking-c9214a10a205

[6] J. Brownlee – "Bagging and Random Forest Ensemble Algorithms for Machine Learning" Available:

https://machinelearningmastery.com/bagging and-random-forest-ensemblealgorithms-for machinelearning/

[7] T. Yiu –" Understanding Random Forest How the Algorithm Works and Why It Is So Effective" Available: https://towardsdatascience.com/understanding random-forest-58381e0602d2

[8] N. Donges –" A COMPLETE GUIDE TO THE RANDOM FOREST ALGORITHM" Figure 7: Deployment of the model using Heroku Electronic copy available at: https://ssrn.com/abstract=3702236 Available: https://builtin.com/data-science/random-forest algorithm

[9] N. Donges –" A COMPLETE GUIDE TO THE RANDOM FOREST ALGORITHM" Available: https://builtin.com/data-science/random-forest algorithm

[10] A. Dey – "Data Pre-processing for Machine Learning" Available: https://medium.com/datadriveninvestor/data preprocessing-for-machinelearning-188e9eef1d2c

[11] "Cross-Validation" – Available: https://www.techopedia.com/definition/32064/cross validation

[12] Tomar, Ravi, Hanumat G. Sastry, and Manish Prateek. 2020. "A Novel Protocol for Information Dissemination in Vehicular Networks." Pp. 1–14 in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 11894 LNCS. Springer, Cham.

[13] Bansal, Parnika, Bhawna Aggarwal, and Ravi Tomar. 2019. "Low-Voltage Multi-Input High TransConductance Amplifier Using Flipped Voltage Follower and Its Application in High Pass Filter." Pp. 525–29 in 2019 International Conference on Automation, Computational and Technology Management, ICACTM 2019. IEEE.

[14] Tomar, Ravi, Rahul Tiwari, and Sarishma. 2019. "Information Delivery System for Early Forest Fire Detection Using Internet of Things." Pp. 477–86 in Communications in Computer and Information Science. Vol. 1045. Springer, Singapore.

[15] Tomar, Ravi, Manish Prateek, and Hanumat G. Sastry. 2017. "A Novel Approach to Multicast in VANET Using MQTT." Pp. 231–35 in Ada User Journal. Vol. 38. Ada-Europe.

[16] Tomar, Ravi, Hanumat Sastry, and Manish Prateek. 2020. "Establishing Parameters for Comparative Analysis of V2V Communication in VANET." Journal of Scientific and Industrial Research (JSIR) 79(01):26–29.

[17] Tomar, Ravi and Sarishma. 2019. "Maintaining Trust in VANETs Using Blockchain." Ada User Journal 40(4):236–41.

[18] Kumar, Shiwanshu and Ravi Tomar. 2018. "The Role of Artificial Intelligence In Space Exploration." Pp. 499–503 in 2018 International Conference on Communication, Computing and Internet of Things (IC3IoT). IEEE.

[19] Sharma, S., Aggarwal, A., & Choudhury, T. (2018). Breast Cancer

Detection Using Machine Learning Algorithms. 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), 114–118.

[20] Rohit, Sabitha, S., & Choudhury, T. (2018). Proposed approach for book recommendation based on user kNN. In Advances in Intelligent Systems and Computing (Vol. 554). <u>https://doi.org/10.1007/978-981-10-3773-3_53</u>

[21] Mehta, I. S., Chakraborty, A., Choudhury, T., & Sharma, M. (2018).
Efficient approach towards bitcoin security algorithm. 2017 International Conference on Infocom Technologies and Unmanned Systems: Trends and Future Directions, ICTUS 2017, 2018-Janua.
https://doi.org/10.1109/ICTUS.2017.8286117

[22] Chhabra, A. S., Choudhury, T., Srivastava, A. V., & Aggarwal, A. (2018). Prediction for big data and IoT in 2017. 2017 International Conference on Infocom Technologies and Unmanned Systems: Trends and Future Directions, ICTUS 2017, 2018-Janua. <u>https://doi.org/10.1109/ICTUS.2017.8286001</u>

[23] Kashyap, N., Choudhury, T., Chaudhary, D. K., & Lal, R. (2016). Mood based classification of music by analyzing lyrical data using text mining. Proceedings - 2016 International Conference on Micro-Electronics and Telecommunication Engineering, ICMETE 2016. https://doi.org/10.1109/ICMETE.2016.65

[24] Choudhury, T., Kaur, A., & Verma, U. S. (2017). Agricultural aid to seed cultivation: An Agribot. Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016. https://doi.org/10.1109/CCAA.2016.7813860

[25] Khunger, M., Choudhury, T., Satapathy, S. C., & Ting, K.-C. (2019). Automated detection of glaucoma using image processing techniques. In Advances in Intelligent Systems and Computing (Vol. 814). https://doi.org/10.1007/978-981-13-1501-5_2