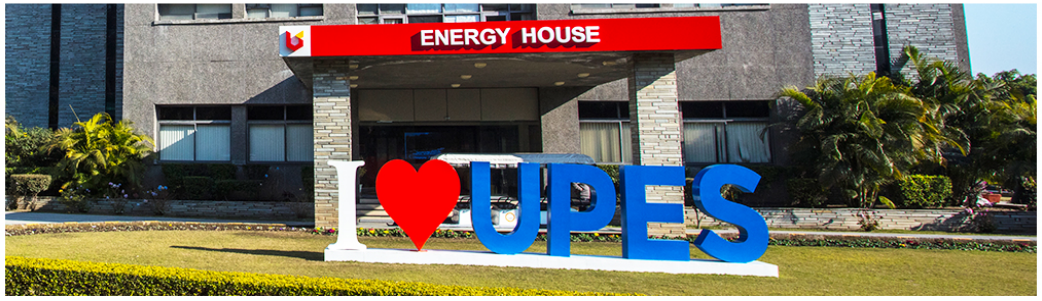


UPES

UNIVERSITY WITH A PURPOSE

MAJOR -2

TITLE :-Prediction Of Selling Prices Using Random Forest Regressor

SUBMITTED BY:-

Shivang Raj	500055050	R114216021
Nitish Kumar	500063395	R114217005
Unnati Gupta	500062488	R114217006

Under the guidance of :-

Ms. Aradhana Kumari singh

Asst. Professor

Dept of Informatics

UPES

UNIVERSITY WITH A PURPOSE

ABSTRACT

- Data Analysis has been a great help in understanding data in several areas like stock market, commercial enterprises, weather, electricity demand, cost and usage of products such as fuel, electricity, etc.
- This project aims to predict used vehicle price and Analysis is done using random forest regressor.

- *Keywords- Data Analysis, data-set, Machine Learning, WebApp.

Introduction

- Approximately 40 million used vehicles are sold each year. Effective pricing strategies can help any company to efficiently sell its products in a competitive market and making profit.

- Whether you're a scientist analysing earthquake data to predict the next "big one", or are in health-care analysing patient wait times to better staff your ER, understanding data is crucial to making better, data informed decisions.

- In the automotive sector, pricing analytics play an essential role for both companies and individuals to assess the market price of a vehicle before putting it on sale or buying it.

Problem Statement

To choose a car ,which is also a hefty investment, a lot of time is wasted by every individual to reduce that time we used random forest regressor to help make better and fast choices, the model makes use of multiple attributes of car to make a differentiated choice.

UPES
UNIVERSITY WITH A PURPOSE

Objectives

- Analysis done on the basis of the data.
- Splitting the dataset for testing and training.
- Predicting the MRP of cars and comparing them to testing data.
- Creating a WebApp that will show Output.

Random Forest Regression

- Random Forest algorithm is Ensemble-Bagging method which operates by constructing multiple decision trees during the training phase.

- Random forest consists of an outsized number of individual decision trees that operate as an ensemble where each tree within the random forest spits out a category prediction then the category with the foremost votes becomes our model's prediction.

Advantages of the Random Forest Algorithm:-

- No Overfitting and takes less training time.
- High accuracy and runs efficiently on large databases.
- Estimate missing values and still maintains high accuracy even having an extensive amount of missing data.

LITERATURE REVIEW

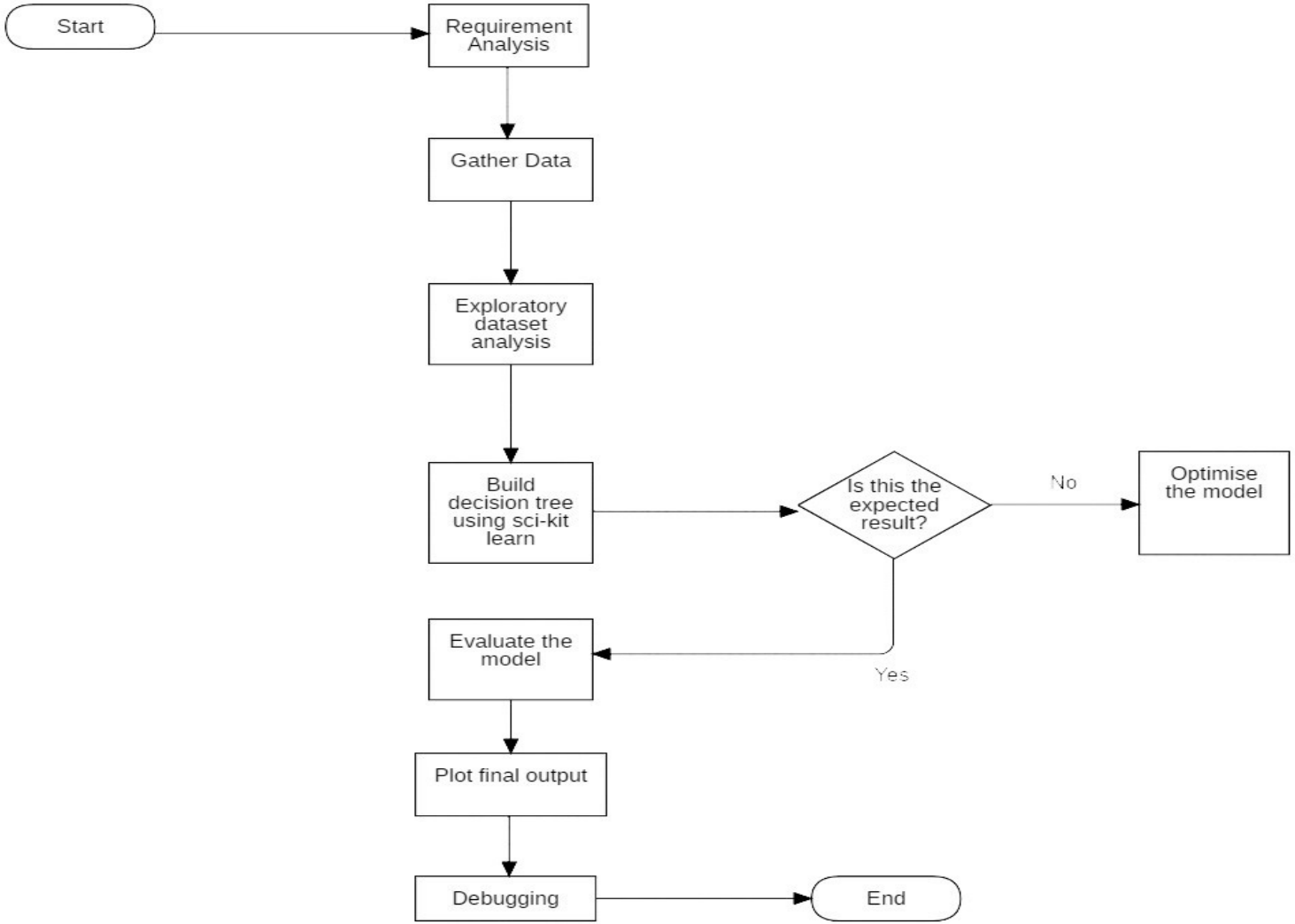
Author	Title	Contribution
Samer Sawalha, Hiba AbdelNabi International Journal of Interactive Mobile Technologies (iJIM)	"Agile Software Development: Methodologies and Trends "	Research paper addressed the software engineering as the application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of soft-ware; that is, the application of engineering to software.[1]It can be seen from the definitions that the final product is important but the approaches and the steps that are needed to produce the software must be done in a systematic, disciplined and quantifi-able way in all of the software life cycles.
Jehad Ali, Nasir Ahmad, Rehanullah Khan	"Random Forests and Decision Trees"	In this paper, They have compared the classification results of two models i.e. Random Forest and the J48 for classifying twenty versatile datasets.[2] They compared the classification results obtained from methods i.e. Random Forest and Decision Tree (J48). The classification parameters consist of correctly classified instances, incorrectly classified instances, F-Measure, Precision, Accuracy and Recall.
Maxwell Olokundun, Hezekiah Falola, Odunayo Salau	"Data set on the effect of training and development"	This article presented data on the effect of training and development on academic staff creativity using Covenant University in Nigeria as the case study. [3] The article was a based on a descriptive quantitative research design using Survey method. Reliability and validity procedures were confirmed. Data was analyzed with the use of Statistical Package for Social Sciences (SPSS).Regression analysis was employed as statistical tool of analysis.

Methodology

We will use a combination of iterative and incremental process models (Agile SDLC model) with focus on process adaptability.

- Requirement Analysis
- Building Data Set
- Training Data Set
- Pre processing of data
- Visualizing results
- Applying random forest regressor

PROJECT FLOW



UPES
UNIVERSITY WITH A PURPOSE

TOOLS

- Python version 3.7 (current available)
- Packages : requests, pandas, numpy, matplotlib
- Windows text editor or equivalent software
- Microsoft excel or equivalent spreadsheet software
- JupyterLab

SYSTEM REQUIREMENTS:

Table 1: Hardware Requirements

Hardware Components	Configuration
Processor	Intel i5 7200
Processor Speed	2.5 GHz
RAM Size	8 GB
OS	Windows 10
GPU	NVIDIA GeForce GTX 940M
VRAM	2GB

Data set

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	brand	model	model_ye	list_price	color	configratic	condition	body_type	wheel_cor	transmissi	fuel_type	mileage	carfax_lin	vin_numbe	image_link	dealer_address				
2	Acura	TSX	2006	5400	Blue	w/A-Spec	Used	Sedan	Front-whe	Automatic	Gasoline	167385	https://wv	jh4cl968x6	https://i.e	Old Kennedy Rd, Markham, ON L3R 0L5, Canada				
3	Nissan	Sentra	2019	16495	Black	SV |MOON	Used	Sedan	Front-whe	Automatic	Gasoline	48567	https://reports.carpr		https://i.e	1599 Star Top Road, Ottawa, ON, K1B 5P5				
4	Mercedes-	C-Class	2018	34995	White	C300|4MA	Used	Sedan	All-wheel	Automatic	Other	63851	https://reports.carpr		https://i.e	100 Toro Road, North York, ON, M3J 2A9				
5	Honda	Other	2010	4999	Silver	Sport	Used	Sedan	Front-whe	Manual	Gasoline	136500	https://wv	2HGFA1E6	https://i.e	North York, ON M6A 2X3				
6	Dodge	Grand Car	2014	0	Silver	30th ANNI	Used	Minivan, V	Front-whe	Automatic	Other	157001	https://wv	2c4rdgbg6	https://i.e	56 Martin Ross Ave., North York, ON, M3J 2L4				
7	Toyota	RAV4	2021	44363	Red	Limited	New	SUV, Cross	All-wheel drive (AWD		Gasoline	0			https://i.e	2336 Saint Clair Avenue West, Toronto, ON, M6N 1K8				
8	Ram	1500	2021	79289	Black	Limited Lo	New	Pickup Tru	4 x 4	Automatic	Other	0			https://i.e	212 Lakeshore Road West, Mississauga, ON, L5H 1G6				
9	BMW	3-Series	2013	25000	Black		Used	Convertible				140000			https://i.e	Toronto, ON M1B4K4				
10	Chevrolet	Cruze	2014	5990	Blue	LT ~AUTOI	Used	Sedan	Front-whe	Automatic	Other	148000	https://wv	1G1PC5SB	https://i.e	1113 Finch Ave W., Toronto, ON, M3J 2E5				

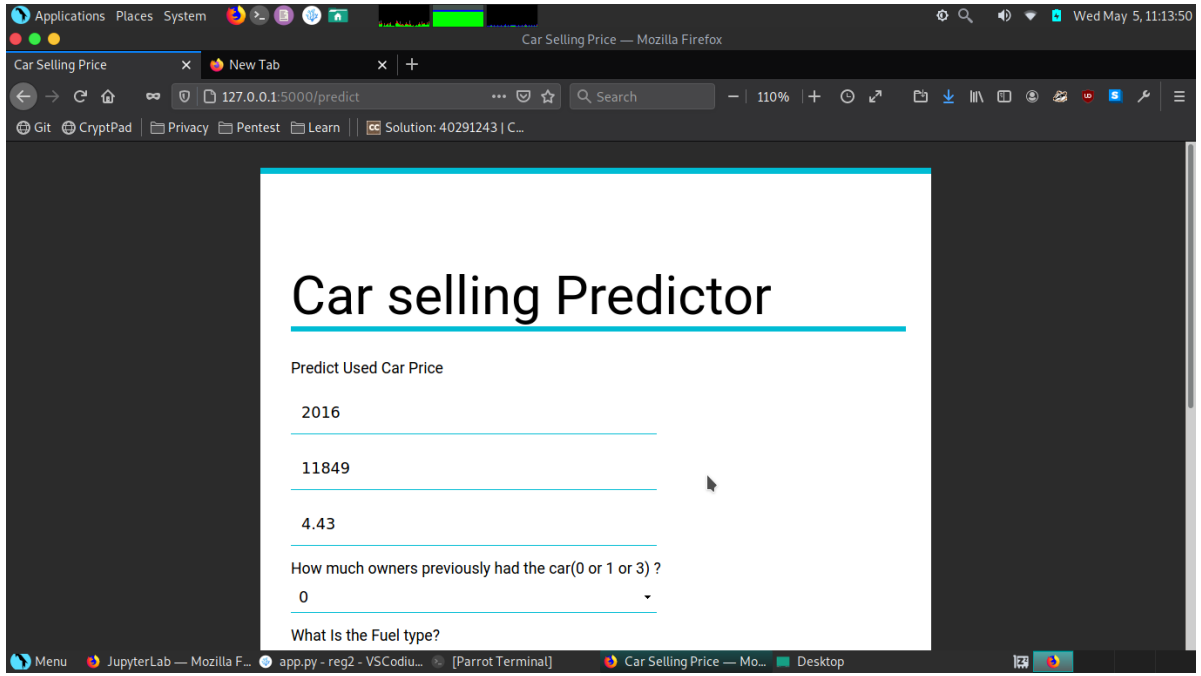
Figure 1: Data set after cleaning

Data set after required modification to be used as input for algorithms

Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
swift	2014	4.6	6.87	42450	Diesel	Dealer	Manual	0
vitara brezza	2018	9.25	9.83	2071	Diesel	Dealer	Manual	0
ciaz	2015	6.75	8.12	18796	Petrol	Dealer	Manual	0
s cross	2015	6.5	8.61	33429	Diesel	Dealer	Manual	0
ciaz	2016	8.75	8.89	20273	Diesel	Dealer	Manual	0

Figure 2: Data set after required modification

RESULTS



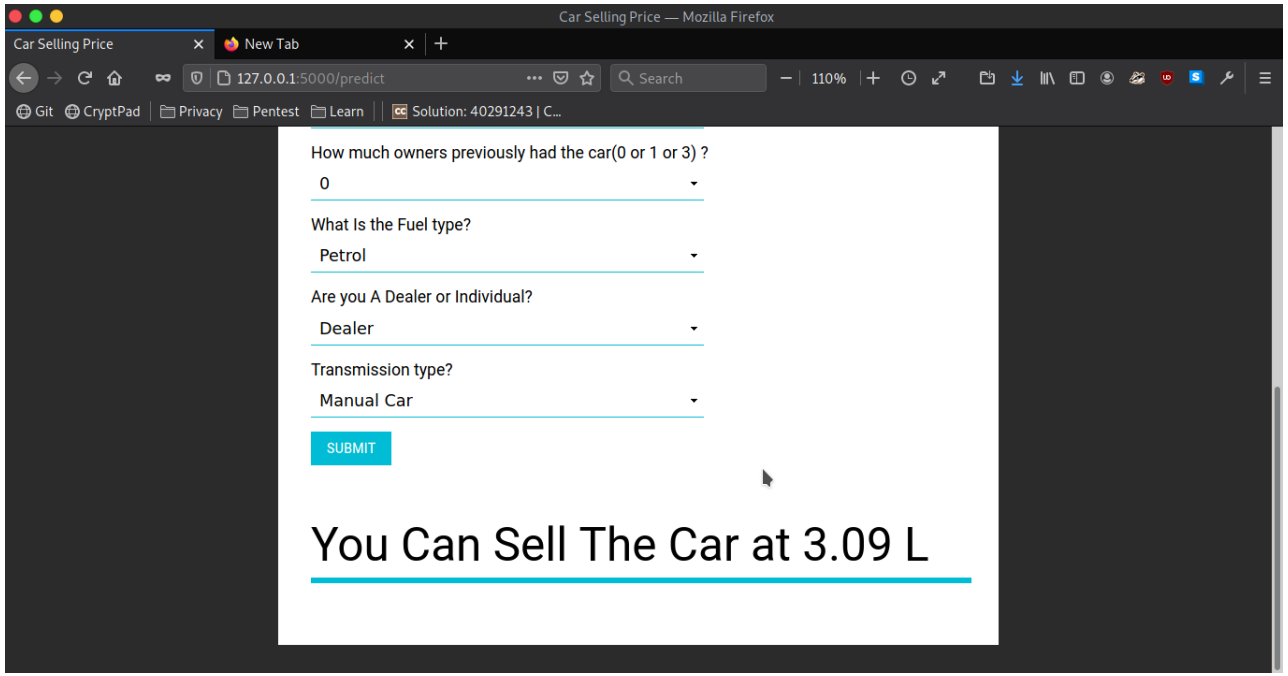


Figure 3: Prediction of Selling Price after applying regression.

UPES
UNIVERSITY WITH A PURPOSE

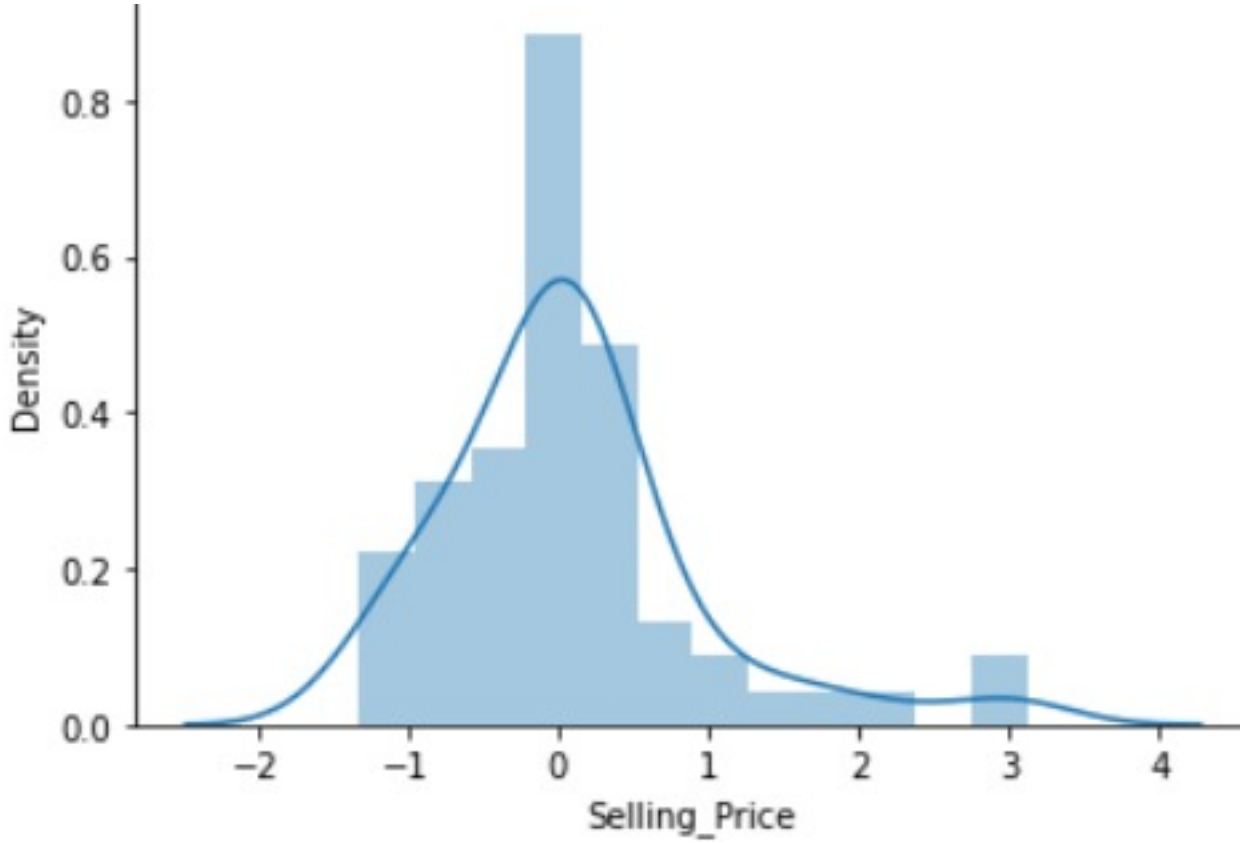


Figure 4: Correctness in mean price

[20]:
```
from sklearn import metrics
print('MAE:', metrics.mean_absolute_error(Y_test, predictions))
print('MSE:', metrics.mean_squared_error(Y_test, predictions))
print('RMSE:', np.sqrt(metrics.mean_squared_error(Y_test, predictions)))
rmse=np.sqrt(metrics.mean_squared_error(Y_test, predictions))
```

```
MAE: 0.5780238176907727
MSE: 0.7524035143630823
RMSE: 0.8674119634655049
```

[24]:
```
print("Accuracy= {}".format(100*max(0,rmse)))
```

```
Accuracy= 86.74119634655048
```

Figure 5: Accuracy

REFERENCES

[1]. Agile Methodology & Model: Guide for So ware Development & Testing. 1 9.

[2]. Data Training: Current Approachesand Issues1Vaishali Chandrakant Wangikar and 2Ratnadeep R. Deshmukh1MCA Department, Maharashtra Academy of Engi- neering, Alandi,

[3]. Pune (MS), India,2Deptartment of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada

[4]. University, Aurangabad (MS), India E-mail: vaishali.wangikar@gmail.com, ratnadeep_deshmukh@yahoo.co.in

[5]. Srivastava, A.K. and Shalabh (1995): "Predictions in Regression Models With Measurement Errors", Indian Journal of Applied Economics, Vol. 4, No. 2, pp. 1-14.

[6]. Exploring filtering based approach to web advertising Eloisa Vargiu1, 2, Mirko Urru11. Dipar mento di Matema ca e Informa ca, Università di Cagliari, Italy. 2. Barcelona Digital Technology Centre, SpainCorre- spondence: Eloisa Vargiu. Address: Barcelona Digital Technology Center, Italy.
Email: evargiu@bdigital.org.

[7]. Toutenburg, H. and Shalabh (1999) : "Estimation of Regression Models with Equi-correlated Responses when some Observations on the Response Variable are Missing", SFB Discussion Paper 174, University of Munich, Munich, Germany.

UPES
UNIVERSITY WITH A PURPOSE

THANK YOU

UPES

UNIVERSITY WITH A PURPOSE