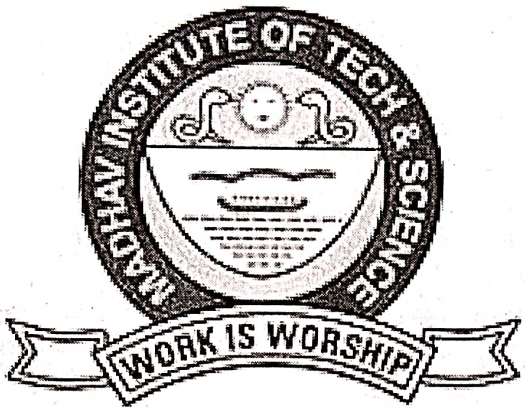
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR
(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)
NAAC Accredited with A++ Grade



Project Report

on

Car Price Prediction Using Deep Learning

Submitted By:

Ankit Sharma

0901AD211006

Faculty Mentor:

Dr. Tej Singh

Assistant Professor

CENTRE FOR ARTIFICIAL INTELLIGENCE
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE
GWALIOR - 474005 (MP) est. 1957

JULY-DEC. 2023

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

CERTIFICATE

This is certified that **Ankit Sharma** (0901AD211006) has submitted the project report titled "**Car Price Prediction Using Deep Learning**" under the mentorship of **Dr. Tej Singh**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in **Artificial Intelligence and Data Science** from Madhav Institute of Technology and Science, Gwalior.

Dr. Tej Singh
Faculty Mentor
Assistant Professor
Centre for Artificial Intelligence

Dr. R. R. Singh
Coordinator
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

DECLARATION

I hereby declare that the work being presented in this project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in **Artificial Intelligence and Data Science** at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of **Dr. Tej Singh**, **Assistant Professor**, Centre for Artificial Intelligence

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Ankit 23/11/2023

Ankit Sharma
0901AD211006
3rd Year,
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

ACKNOWLEDGEMENT

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Centre for Artificial Intelligence,** for allowing me to explore this project. I humbly thank **Dr. R. R. Singh,** Coordinator, Centre for Artificial Intelligence, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Dr. Tej Singh,** Assistant Professor, Centre for Artificial Intelligence, for his continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.

Ankit 23/11/2023

Ankit Sharma
0901AD211006
3rd Year,
Centre for Artificial Intelligence

ABSTRACT

In this work, we have tried to make a machine learning model for predicting the price of cars. Since this is a regression problem, So, to implement this task we can use many machine learning techniques like Linear Regression, Lasso Regularisation, Ridge Regularisation, Elastic Net Regression, e.t.c. But in this work, we will primarily focus on a deep learning (which is a subset of machine learning) model Simple Feed Forward Neural Network (a type of Artificial Neural Network) for Car Price Prediction with various optimizers like Adam, RMSprop, Adagrad, Adadelta, Nadam, e.t.c. and also compare the performance of these optimizers. This model can help us in predicting the price of cars especially in case of purchasing a second hand car according to the brand and features of the car.

Keyword: Machine learning, Regression problem, Linear Regression, Lasso Regularisation, Ridge Regularisation, Elastic Net Regression, Deep Learning, Simple Feed Forward Neural Network, Artificial Neural Network, Car Price Prediction, optimizers, Adam, RMSprop, Adagrad, Adadelta, Nadam.

सार:

इस काम में, हमने कारों की कीमत की भविष्यवाणी के लिए एक मशीन लर्निंग मॉडल बनाने की कोशिश की है। चूँकि यह एक प्रतिगमन समस्या है, इसलिए, इस कार्य को लागू करने के लिए हम कई मशीन लर्निंग तकनीकों का उपयोग कर सकते हैं जैसे लीनियर रिग्रेशन, लैस्सो रेगुलराइजेशन, रिज रेगुलराइजेशन, इलास्टिक नेट रिग्रेशन, आदि। लेकिन इस काम में, हम मुख्य रूप से एडम, आरएमएसप्रॉप, एडाग्राड जैसे विभिन्न ऑप्टिमाइज़र के साथ कार मूल्य भविष्यवाणी के लिए एक गहन शिक्षण (जो मशीन लर्निंग का एक उपसमूह है) मॉडल सिंपल फीड फॉरवर्ड न्यूरल नेटवर्क (एक प्रकार का कृत्रिम न्यूरल नेटवर्क) पर ध्यान केंद्रित करेंगे। एडाडेल्टा, नादम, आदि। और इन अनुकूलकों के प्रदर्शन की तुलना भी करें। यह मॉडल हमें कारों की कीमत का अनुमान लगाने में मदद कर सकता है, खासकर कार के ब्रांड और फीचर्स के अनुसार सेकेंड हैंड कार खरीदने के मामले में।

कीवर्ड: मशीन लर्निंग, रिग्रेशन समस्या, लीनियर रिग्रेशन, लासो रेगुलराइजेशन, रिज रेगुलराइजेशन, इलास्टिक नेट रिग्रेशन, डीप लर्निंग, सिंपल फीड फॉरवर्ड न्यूरल नेटवर्क, आर्टिफिशियल न्यूरल नेटवर्क, कार प्राइस प्रेडिक्शन, ऑप्टिमाइजर्स, एडम, आरएमएसप्रॉप, एडाग्रेड, एडाडेल्टा, नादम।

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Symbol	Description

LIST OF ABBREVIATIONS

Abbreviation	Description
Adam	Adaptive Moment Estimation
RMSprop	Root Mean Square Propagation
Adagrad	Adaptive Gradient
Nadam	Nesterov-Accelerated Adaptive Moment Estimation
MAE	Mean Absolute Error
MSE	Mean Square Error
R2 score	R-squared score(Coefficient of determination)
RSS	Sum of squares of residuals
TSS	total sum of squares

Chapter 1: INTRODUCTION

In today's world, car is a common household thing which can be found in almost every elite and middle class family. It is a dream of every middle class family to have atleast 1 four wheeler but since it is a very expensive commodity, it is difficult for them to afford a brand new car. So, they normally try to purchase a second hand car. But in case of second hand car, it is difficult for them to guess how much should it cost according to its brand, average, fuel type, mileage, engine power and other features. So, for this a technology is required to predict the price of the car according to its brand and concerned features.

So, to solve this problem, I have tried to make a deep learning Simple Feed Forward Neural Network model by training the model on a car dataset by taking various important features as primary concern. In this work, we have used various optimizers and at last we have also discussed the performance of these optimizers.

Chapter 2: Related Work

Many people have done researches on implementing machine learning models for predicting the price of cars. For this they have performed various techniques. Some of the works are given below:

1. Gegic and Isakovic tried to predict the price of cars by using various supervised learning methods. They focused on various features like fuel, power, years e.t.c. and achieved a good accuracy. [1]
2. Sharma and Sharma used simple linear regression method and predicted the price of cars using the various features like distance travelled, fuel type, transmission, years used, e.t.c. and achieved the good r2 score. [2]
3. Pandit and Parekh used various machine learning techniques to predict the car price and concludes that decision tree as the best performer for this task. [3]
4. Rane and Pandaya used various supervised learning techniques like linear, ridge and lasso regression for predicting the prices of used cars. [4]
5. Agrahari and Chaubey considered various features of cars like mileage, power, engine, seats, e.t.c. and used these features to predict the price of cars. [5]
6. Samruddhi and Kumar tried to deal with the task of predicting car price for small datasets by using k nearest neighbour. [6]
7. Kiran and Kala used various machine learning algorithms like decision tree, random forest and voting classifier for price prediction of used cars. [7]
8. Noor and Jan tried to predict the price of different vehicles using multiple linear regression. [8]
9. Nikhade and Borde used various regression algorithms for predicting the price of car by taking year, present price, distance driven e.t.c. as input. [9]

Chapter 3 : About Dataset

In this work, I have used a car dataset which includes all the concerned featured which are necessary for predicting the price of cars. In this dataset, cars of various brands are included.

This dataset have 122 rows and 47 columns. The name of the columns of this dataset are ['Brand_Audi', 'Brand_BMW', 'Brand_Fiat', 'Brand_Ford', 'Brand_Honda','Brand_Hyundai','Brand_Jaguar','Brand_Land','Brand_Mahindra','Brand_Maruti', 'Brand_Mercedes-Benz', 'Brand_Mitsubishi','Brand_Nissan', 'Brand_Renault','Brand_Skoda','Brand_Tata','Brand_Toyota','Brand_Volkswagen','Brand_Volvo','FuelType_Diesel','FuelType_Petrol','TransmissionType_Manual', 'Body Type_Convertible','Body Type_Coupe', 'Body Type_Hatchback', 'Body Type_Minivan','Body Type_Pickup Truck', 'Body Type_SUV', 'Body Type_Sedan','ARAI Mileage', 'Engine Displacement (cc)', 'No. of cylinder','Max Power (kW)', 'Max Torque (Nm)', 'Seating Capacity','Fuel Tank Capacity', 'Power Steering', 'Power Windows Front','Anti Lock Braking System', 'Air Conditioner', 'Driver Airbag','Passenger Airbag', 'Automatic Climate Control', 'Fog Lights - Front','Alloy Wheels', 'Price', 'Ground Clearance Unladen'].

The first five rows of this dataset are :

	Unnamed: 0	Brand_Audi	Brand_BMW	Brand_Fiat	Brand_Ford	Brand_Honda	Brand_Hyundai	Brand_Jaguar	Brand_Land	Brand_Mahindra	...	Power Windows Front	Bra Sy
0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	
1	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	
2	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	
3	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	
4	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	

5 rows × 48 columns

Hyundai	Brand_Jaguar	Brand_Land	Brand_Mahindra	...	Power Windows Front	Anti Lock Braking System	Air Conditioner	Driver Airbag	Passenger Airbag	Automatic Climate Control	Fog Lights - Front	Alloy Wheels	Price	Ground Clearance Unladen
0.0	0.0	0.0	0.0	...	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	751000.0	689.200000
0.0	0.0	0.0	0.0	...	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1029500.0	204.095238
0.0	0.0	0.0	0.0	...	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1121500.0	204.095238
0.0	0.0	0.0	0.0	...	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	795000.0	174.250000
0.0	0.0	0.0	0.0	...	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1389500.0	210.000000

Fig 3.1

In this dataset, in each row, the value of the brand of the concerned car is 1 and other brands have value 0.

We have tried to find out the relationships between different columns of this data set by using correlation[24].

The correlation Heatmap[25] of this data set is following :

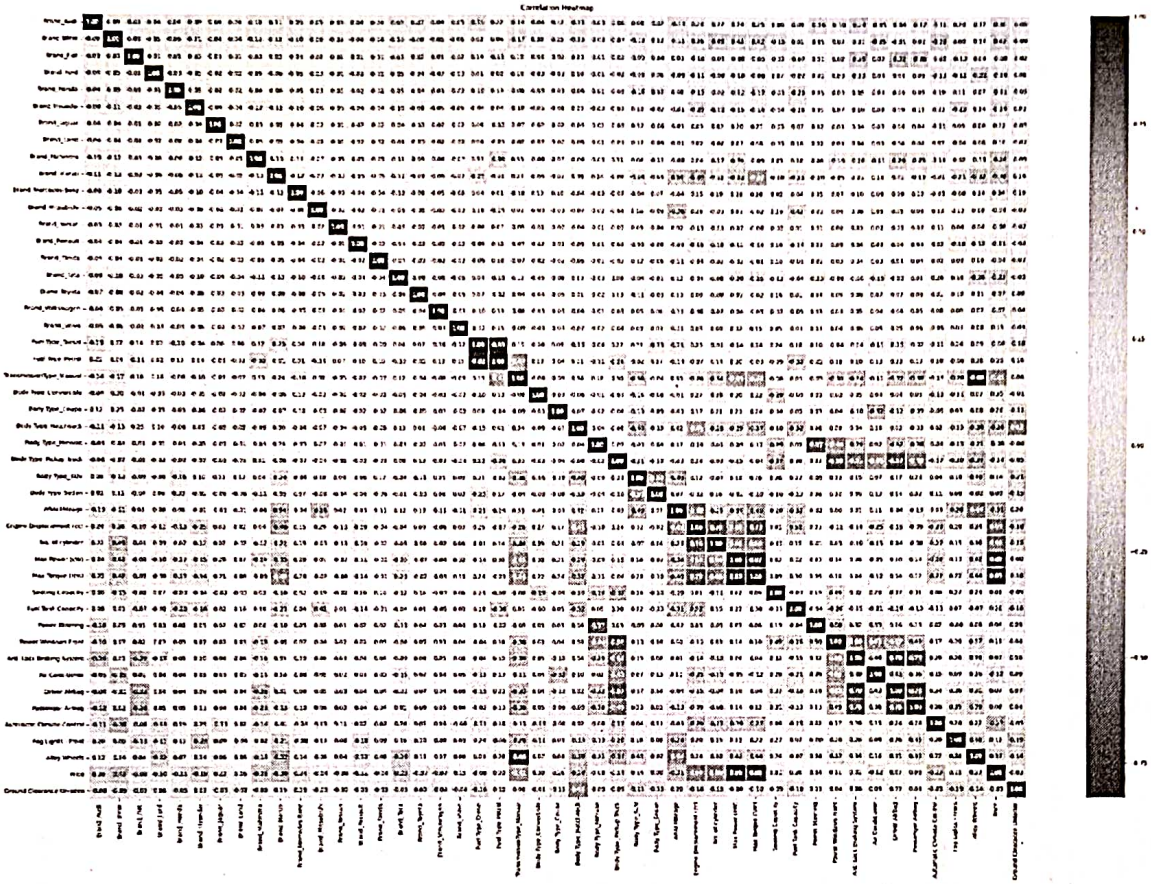


Fig 3.2

The correlation heatmap between different brands of car for this dataset :

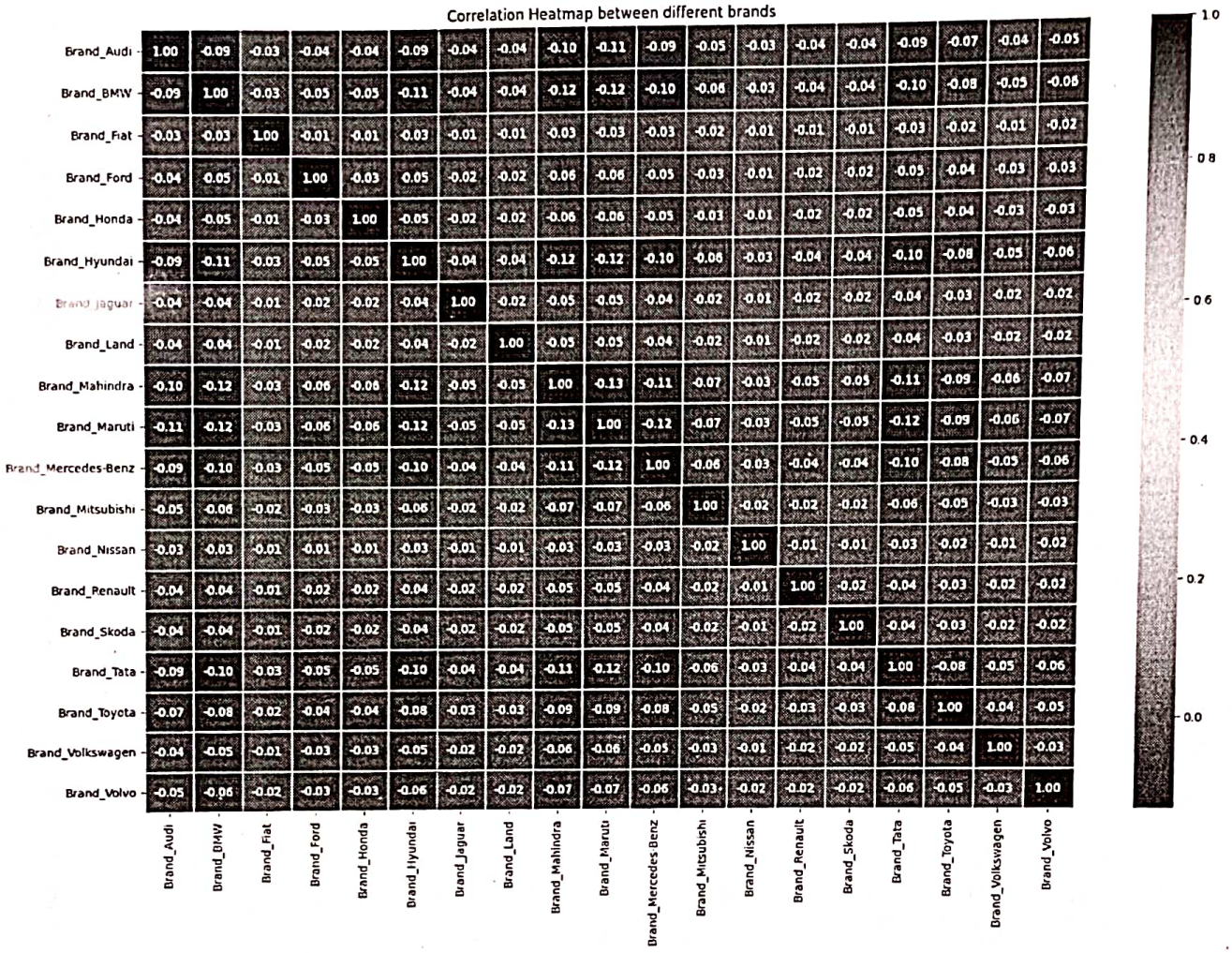


Fig 3.3

The whole description of this dataset is represented though the following figures:

	Brand_Audi	Brand_BMW	Brand_Fiat	Brand_Ford	Brand_Honda	Brand_Hyundai	Brand_Jaguar	Brand_Land	Brand_Mahindra	Brand_Maruti	...	P. Winc F
count	122.000000	122.000000	122.000000	122.000000	122.000000	122.000000	122.000000	122.000000	122.000000	122.000000	...	122.00
mean	0.073770	0.098361	0.008197	0.024590	0.024590	0.098361	0.016393	0.016393	0.114754	0.122951	...	0.95
std	0.262475	0.299030	0.090536	0.155511	0.155511	0.299030	0.127507	0.127507	0.320039	0.329735	...	0.19
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.00
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.00
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.00
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.00

8 rows × 47 columns

Brand_Mahindra	Brand_Maruti	...	Power Windows Front	Anti Lock Braking System	Air Conditioner	Driver Airbag	Passenger Airbag	Automatic Climate Control	Fog Lights - Front	Alloy Wheels	Price	Ground Clearance Unladen
122.000000	122.000000	...	122.000000	122.000000	122.000000	122.000000	122.000000	122.000000	122.000000	122.000000	1.220000e+02	122.000000
0.114754	0.122951	...	0.959016	0.909836	0.934426	0.926230	0.901639	0.409836	0.663934	0.844262	3.848148e+06	258.103437
0.320039	0.329735	...	0.199070	0.287598	0.248556	0.262475	0.299030	0.493831	0.474310	0.364102	3.621703e+06	268.030291
0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	4.200000e+05	100.000000
0.000000	0.000000	...	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	1.000000	1.024000e+06	179.250000
0.000000	0.000000	...	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	1.000000	1.000000	1.991500e+06	204.095238
0.000000	0.000000	...	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	6.328750e+06	204.095238
1.000000	1.000000	...	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.485000e+07	2729.000000

Fig 3.4

Chapter 4 : Model

In this research work, for Car Price Prediction, we have created a Simple Feed Forward Neural Network. For implementing this model, we first split the dataset into training and testing data by taking test_size=0.30 and random_state=42.

Then we standardised the training and training dataset by using StandardScalar() method of sklearn library.

After this, we created a model having 4 layers of neural networks. First layer has 128 neurons along with relu activation function, second layer has 64 neurons along with relu activation function, third layer has 32 neurons along with relu activation function and last(fourth) layer has only one neuron which acts as an output neuron for this regression problem.

After this, we compiled the model by using various optimizers which are following :

1. Adam(Adaptive Moment Estimation) : It is a SGD(Stochastic Gradient Descent) algorithm which is based on adaptive estimations of first and second order moments.
2. RMSprop(Root Mean Square Propagation) : It is an optimization technique based on gradient used for training of Artificial neural networks proposed by Geofrey Hinton(Father of Back Propagation).
3. Adagrad(Adaptive Gradient) : It adapts the learning rate for every parameter according to historic gradient information individually. It performs good for sparse data but it suffers from overly aggressive learning rate decay.
4. Adadelta : It is an extention of Adagrad which deals with the problem of learning rate decay. It takes the running average of the parameter updates for adapting the learning rates and remove the requirement for specifying the initial learning rate.
5. Nadam(Nesterov-Accelerated Adaptive Moment Estimation) : It is an addition to Adam that includes Nesterov-Accelerated Gradient(NAG) into an algorithm. This optimizer is more resistant to noisy gradients as its convergence is improved with the help of NAG.

Here, we will take learning rate=0.01

Now, we trained the model with training and testing data by taking number of epochs=50 and batch size=16.

The plotting of this model is depicted by following figure :

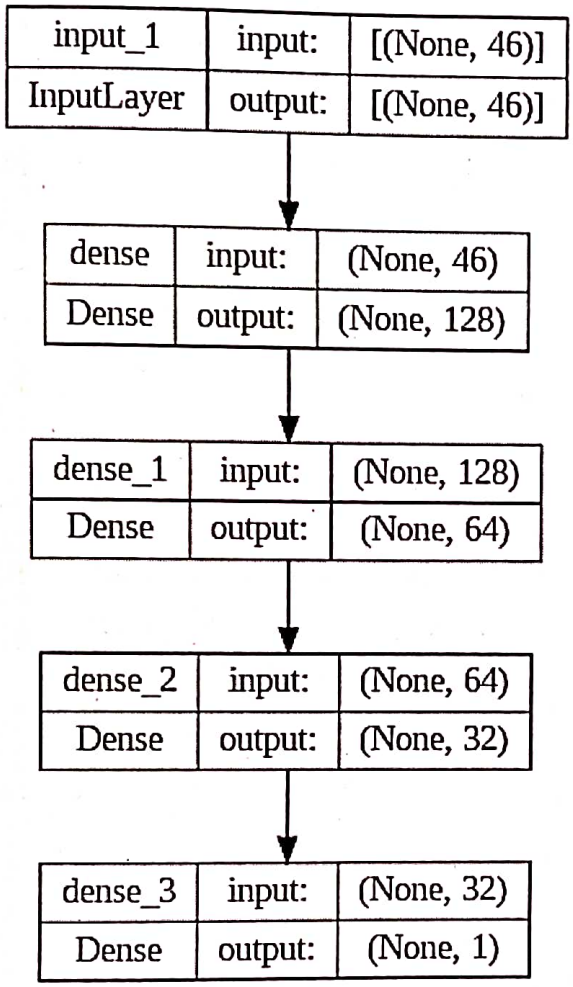


Fig 4.1

The summary of this model is depicted by the following figure :

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 128)               6016

 dense_1 (Dense)             (None, 64)                8256

 dense_2 (Dense)             (None, 32)                2080

 dense_3 (Dense)             (None, 1)                 33

=================================================================
Total params: 16385 (64.00 KB)
Trainable params: 16385 (64.00 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Fig 4.2

Chapter 5 : Results

Since car price prediction is a regression problem. So, we can use the parameters like mean absolute error(mean of the modulus of the difference of y-predicted and y-actual for testing dataset)[33], mean square error(mean of the square of the difference of y-predicted and y-actual for testing dataset)[34], root mean square error (square root of the mean of the square of the difference of y-predicted and y-actual for testing dataset) and r2 score(also known as Coefficient of determination. It is equal to difference of the ratio of RSS(sum of squares of residuals)and TSS(Total sum of squares) from 1) to understand the performance of this deep learning model)[35].

In this model,

For Adam optimizer,
Mean Absolute Error = 1078369.5119562922
Mean Square Error = 3550281808265.53
R2 score = 0.7266976276138666

For RMSprop Optimizer,
Mean Absolute Error = 1278129.1495460304
Mean Square Error = 4748897593830.016
R2 score = 0.6344276176637882

For Adagrad Optimizer,
Mean Absolute Error = 3785143.0123979724
Mean Square Error = 26992325621451.758
R2 score = -1.0778819899273864

For Adadelta Optimizer,
Mean Absolute Error = 3833729.148976664
Mean Square Error = 27687785487375.64
R2 score = -1.1314188192612802

For Nadam optimizer,
Mean Absolute Error = 1361081.8473131335
Mean Square Error = 4728849202328.92
R2 score = 0.6359709523216226

For this model, the scatter plot[36] between actual values and predicted values for various optimizers is shown in following graphs :

a. For Adam optimizer,

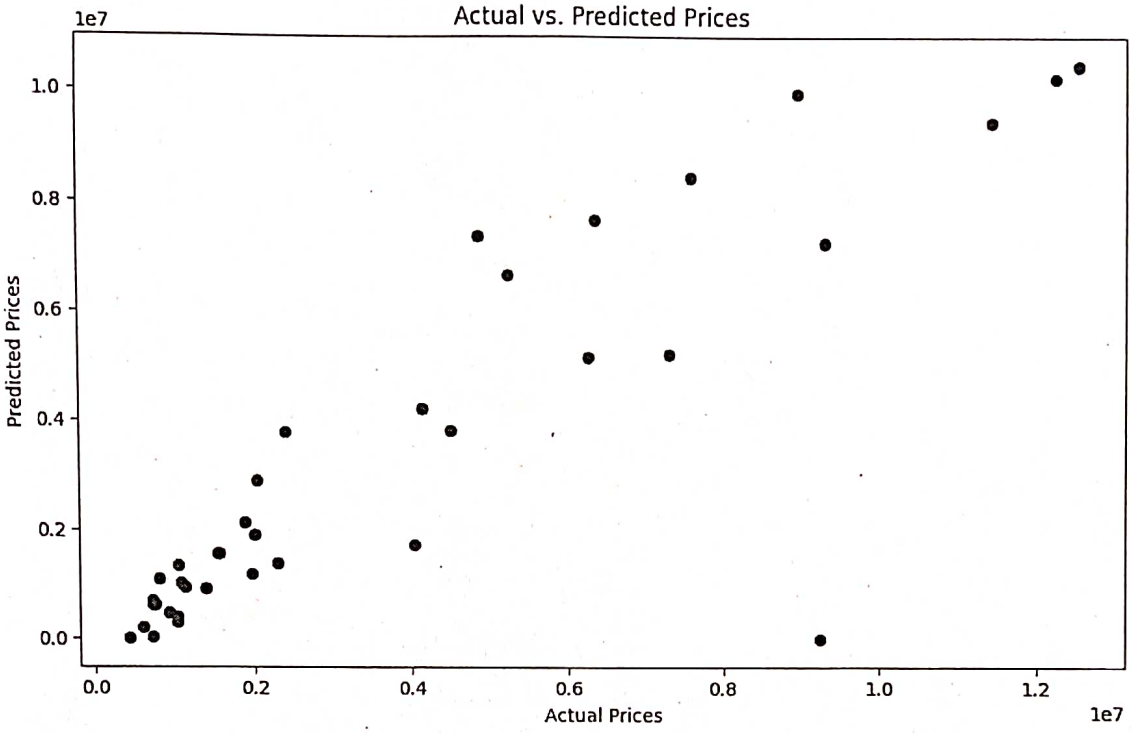


Fig 5.1(a)

b. For RMSprop Optimizer,

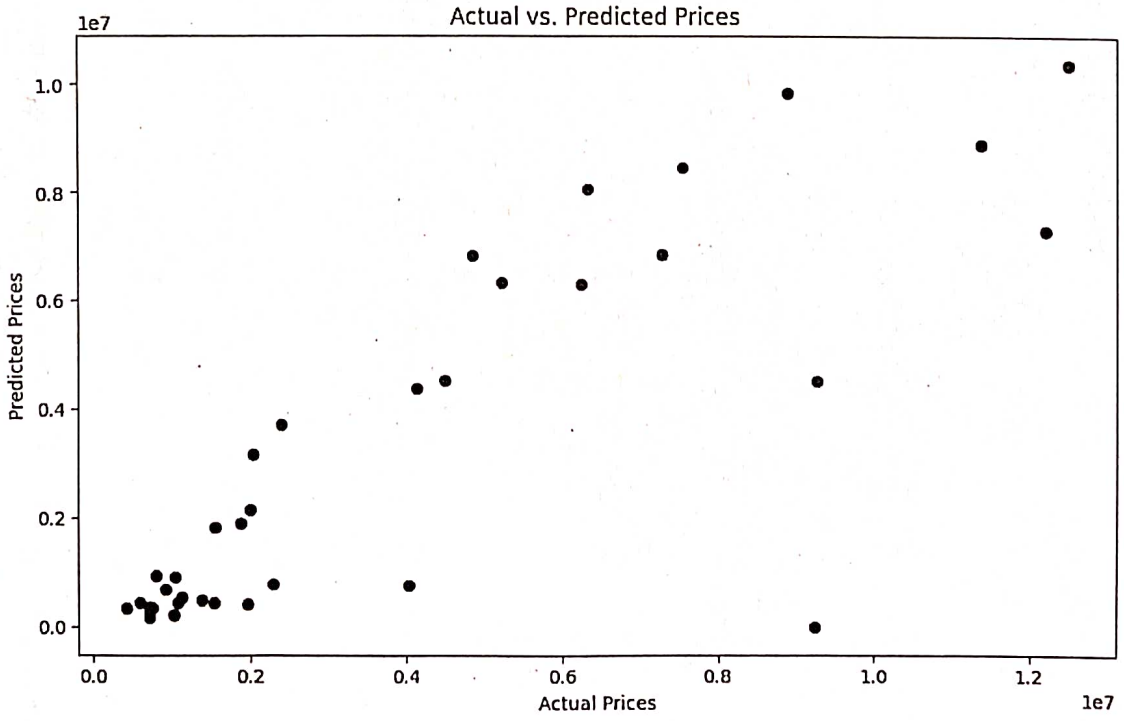


Fig 5.1(b)

c. For Adagrad Optimizer,

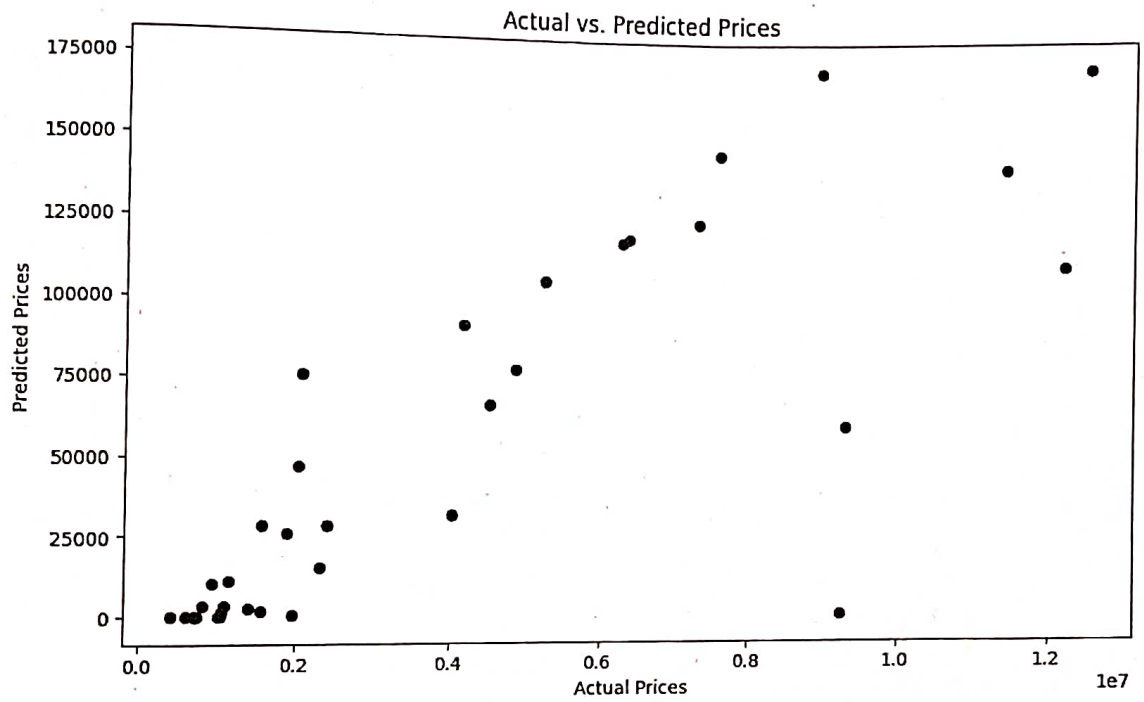


Fig 5.1(c)

d. For Adadelta Optimizer,

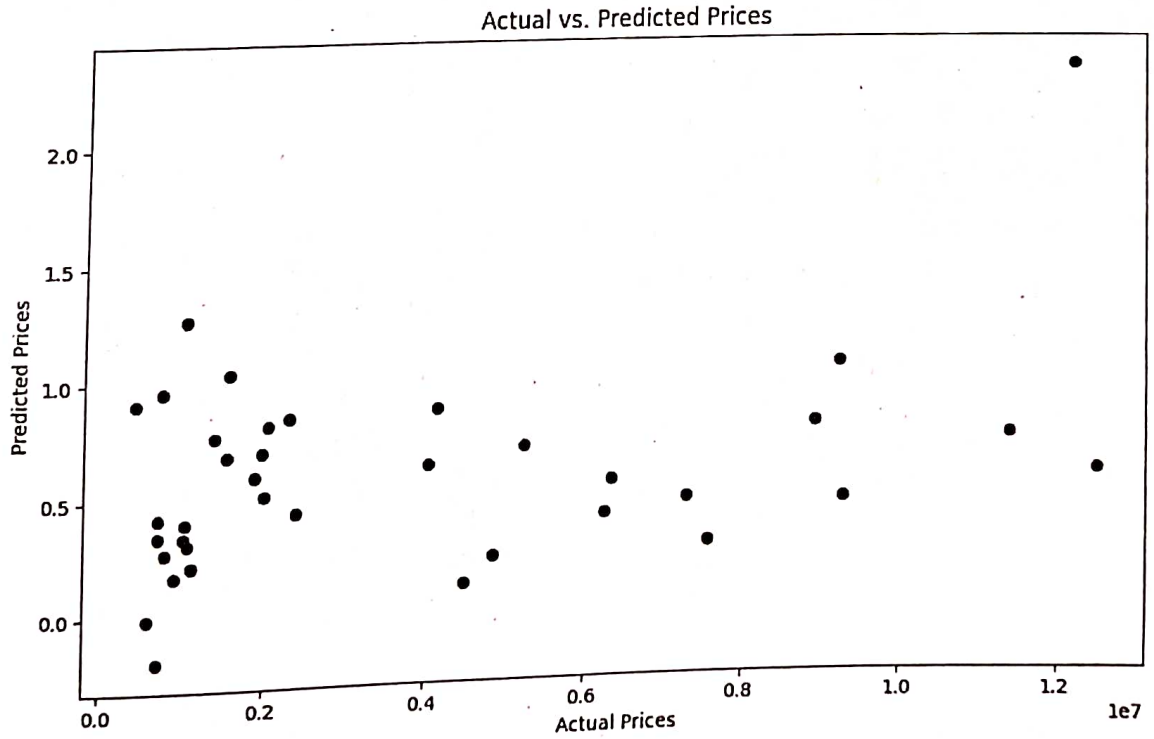


Fig 5.1(d)

e. For Nadam Optimizer,

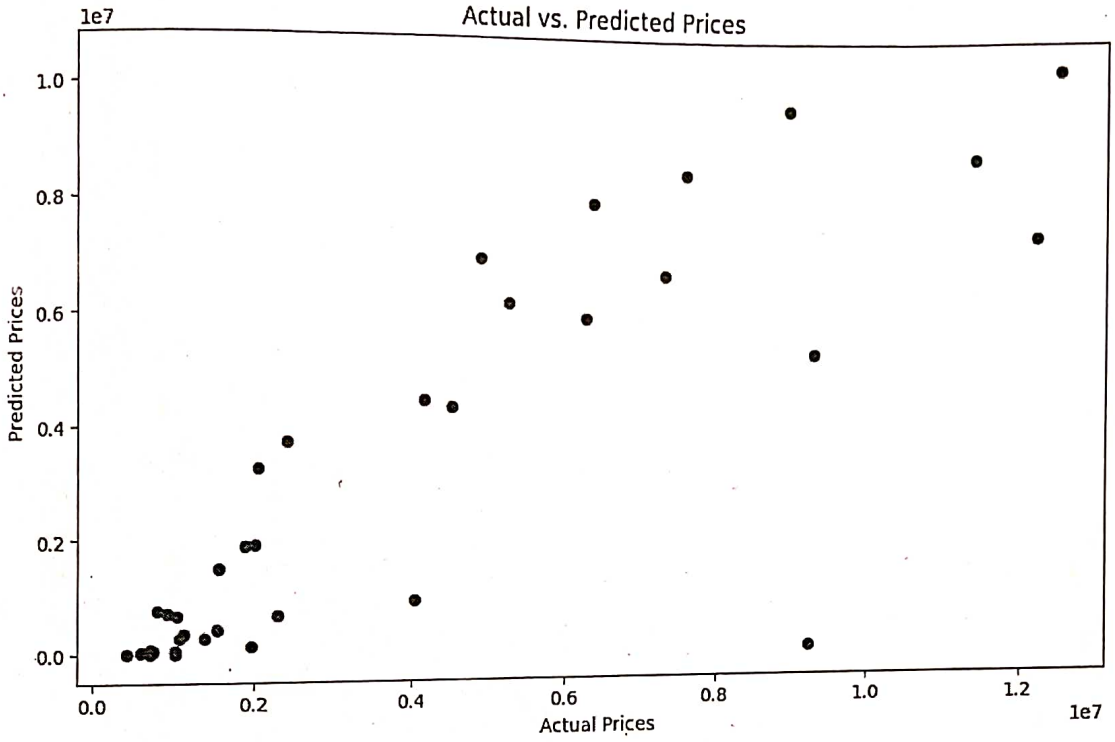


Fig 5.1(e)

Now, the histogram[37] of the residuals(difference between y-test and y-predicted) for various optimizers is shown in following figures :

a. For Adam optimizer,

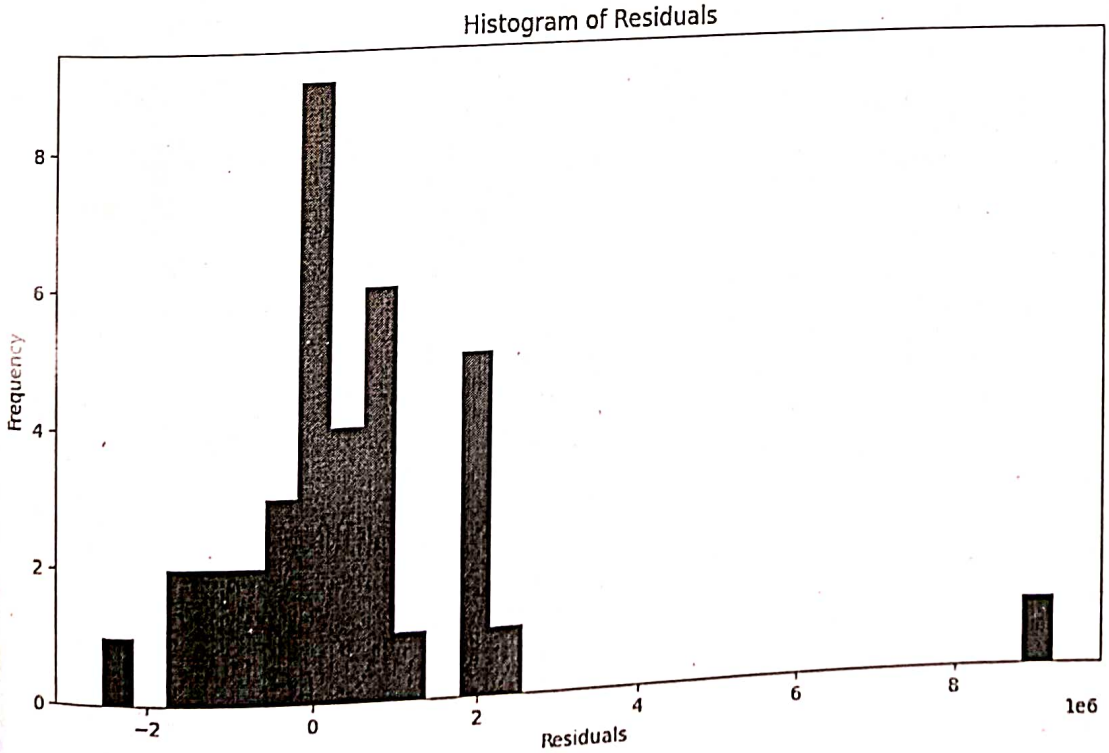


Fig 5.2(a)

b. For RMSprop Optimizer,

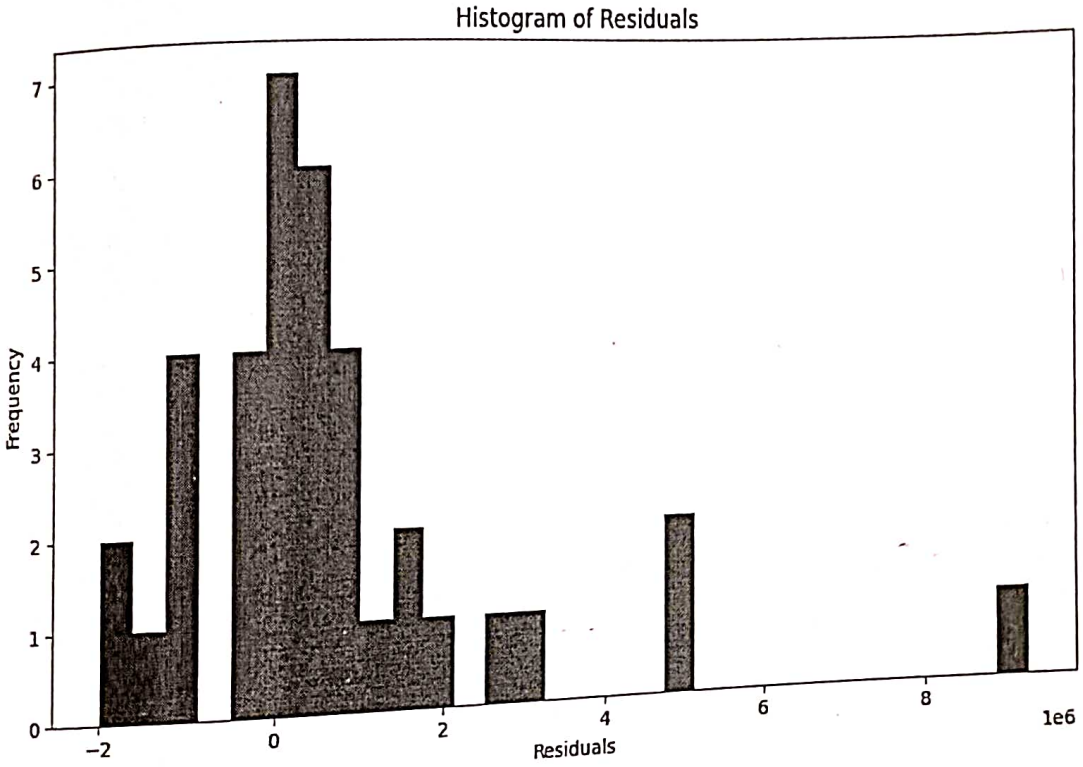


Fig 5.2(b)

c. For Adagrad Optimizer,

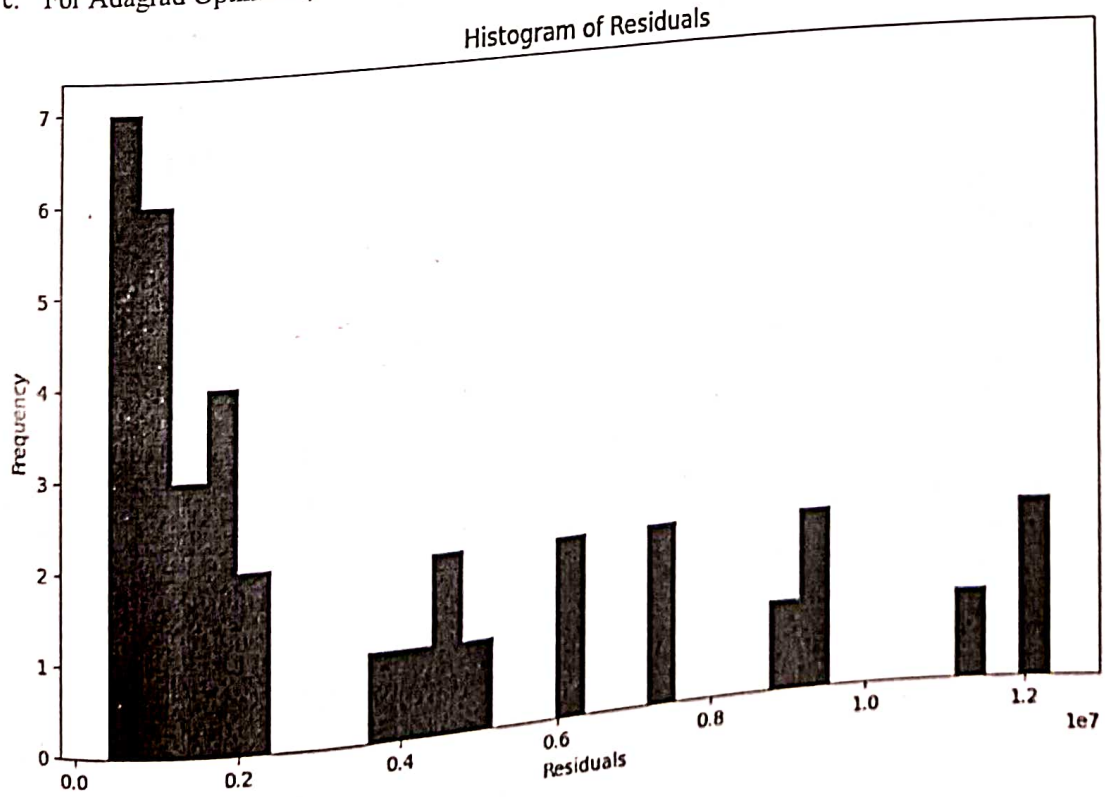


Fig 5.2(c)

d. For Adadelta Optimizer,

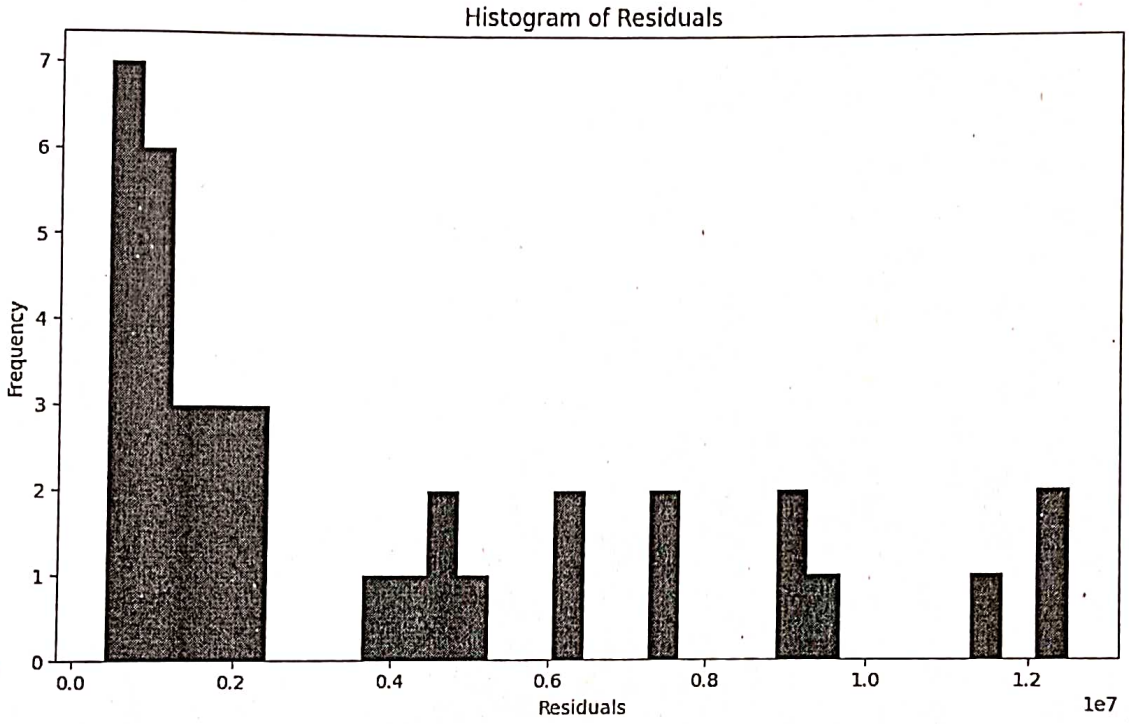


Fig 5.2(d)

e. For Nadam Optimizer,

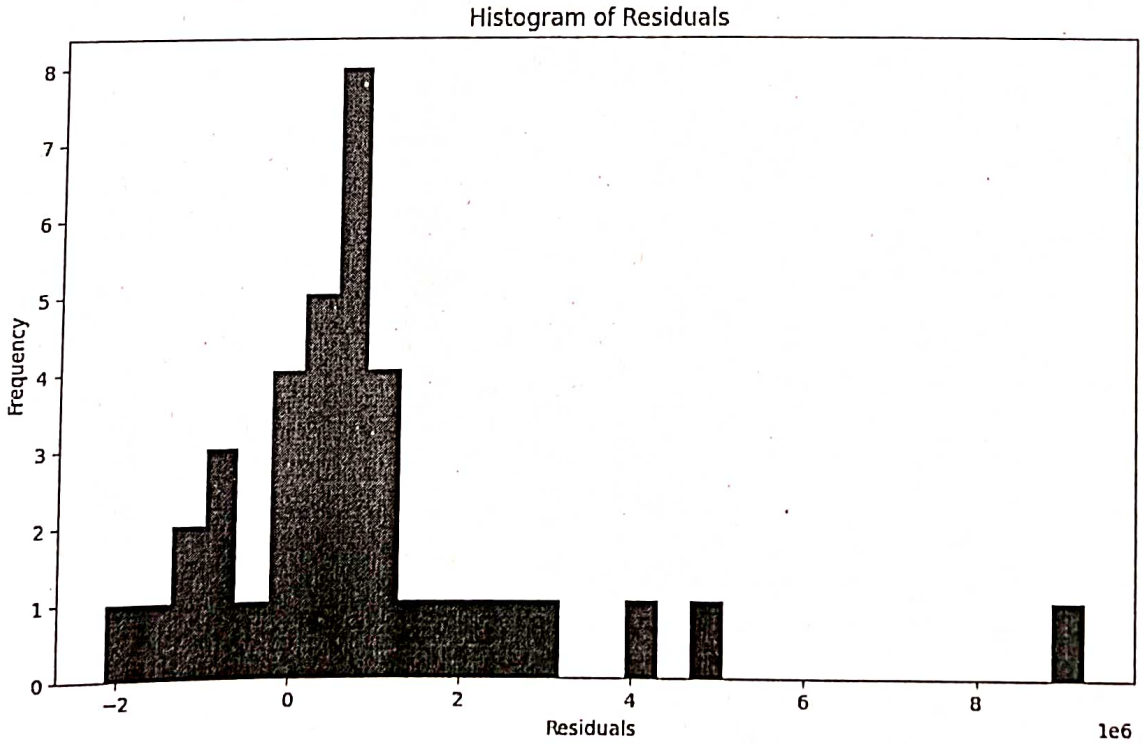


Fig 5.2(c)

Plot of the training loss(Mean Square error) with respect to number of epochs for various optimizers is depicted through following graphs :

a. For Adam Optimizer,

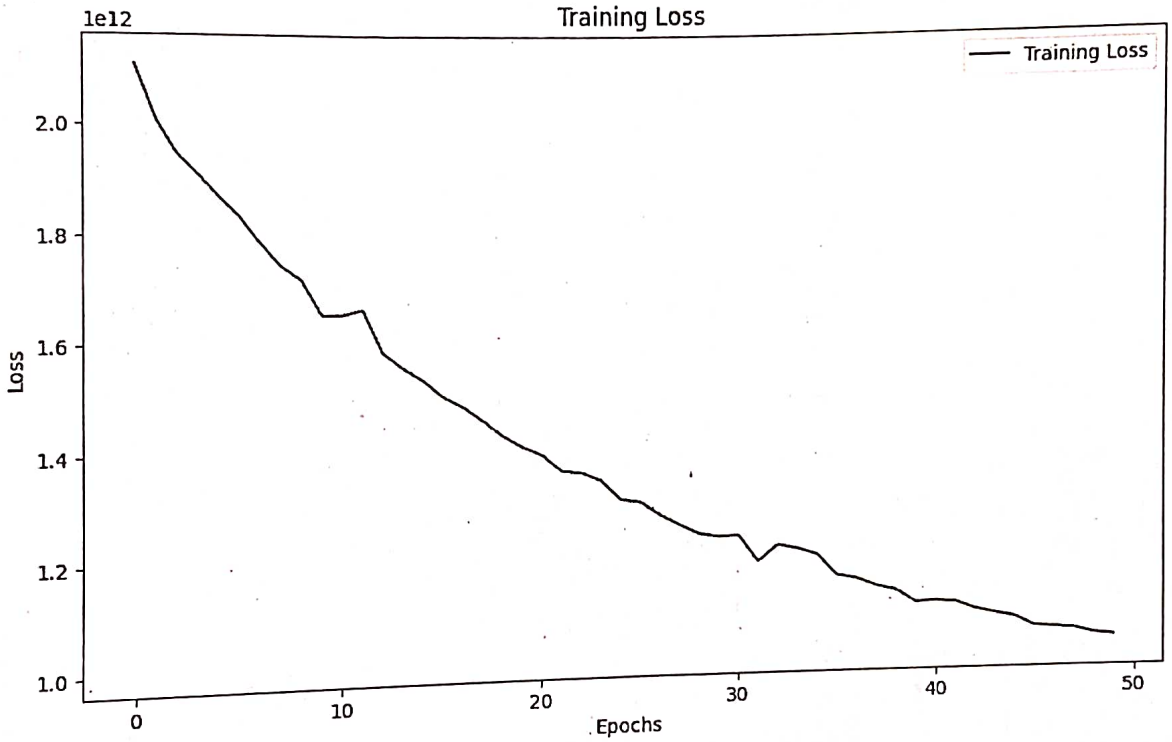


Fig 5.3(a)

b. For RMSprop Optimizer,

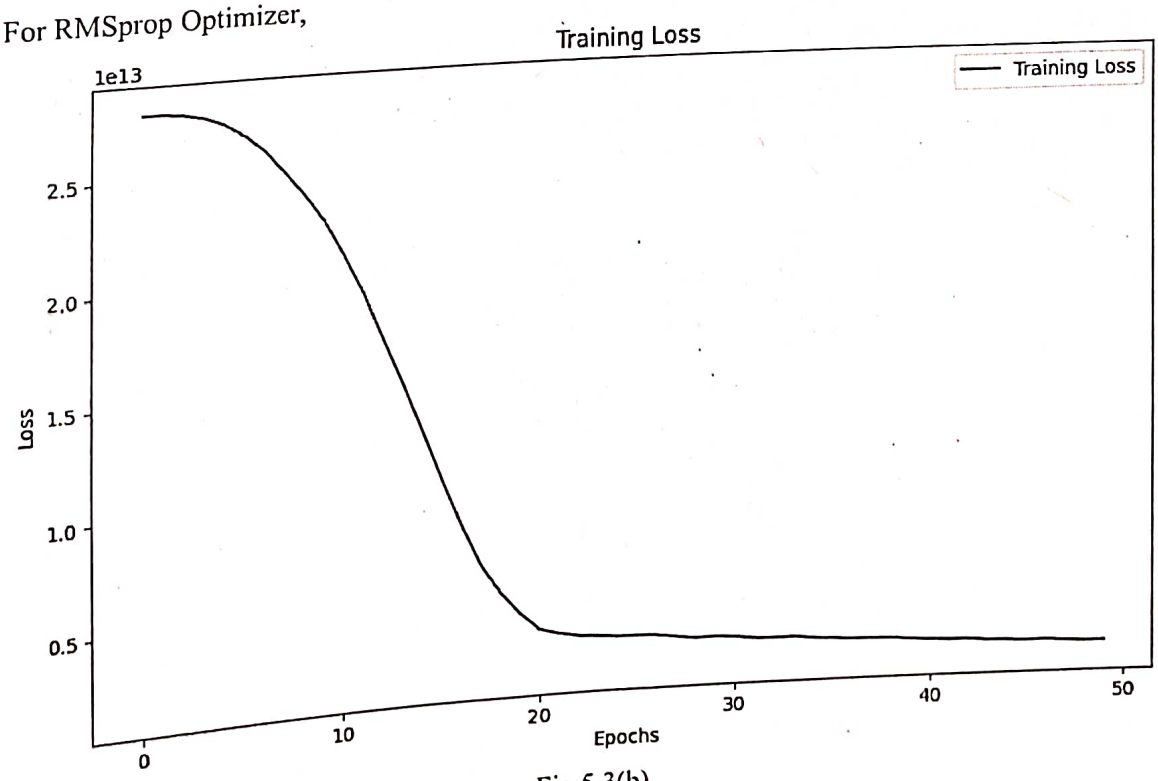


Fig 5.3(b)

c. For Adagrad Optimizer,

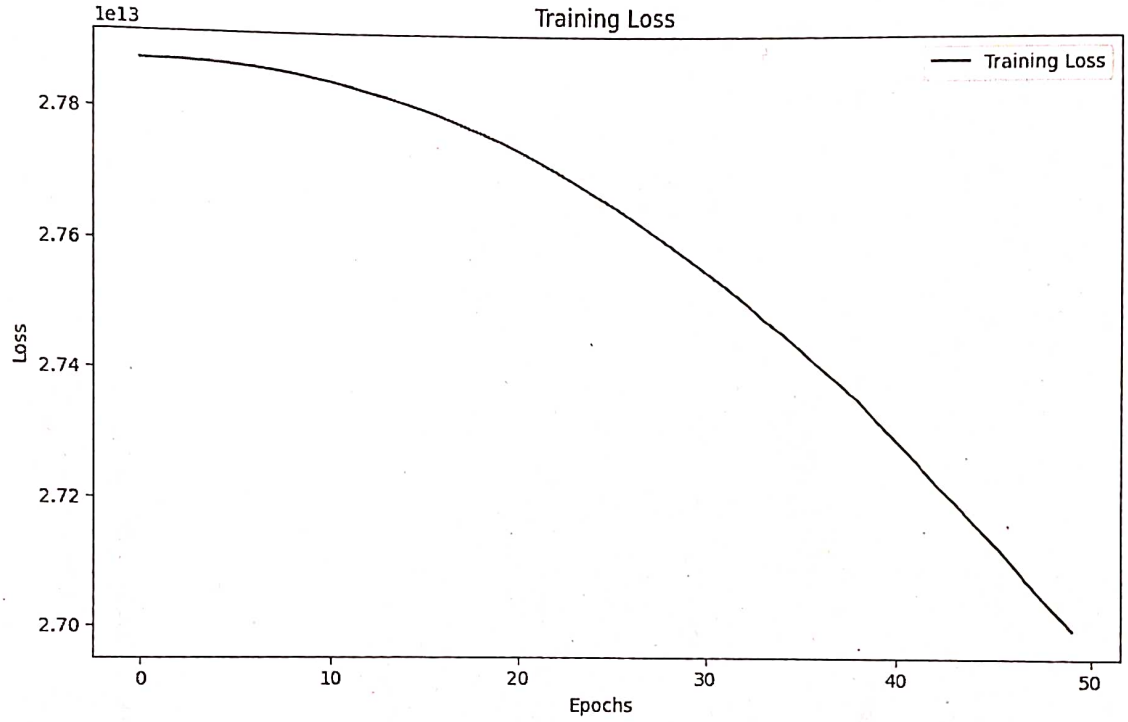


Fig 5.3(c)

d. For Adadelta Optimizer,

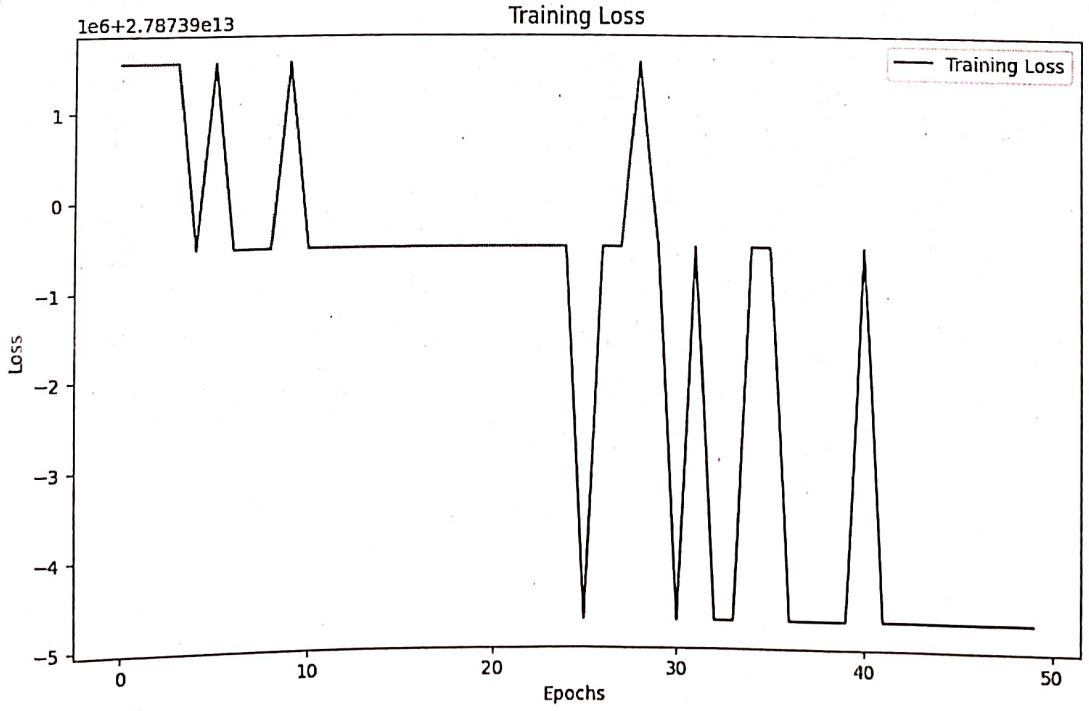


Fig 5.3(d)

e. For Nadam Optimizer,

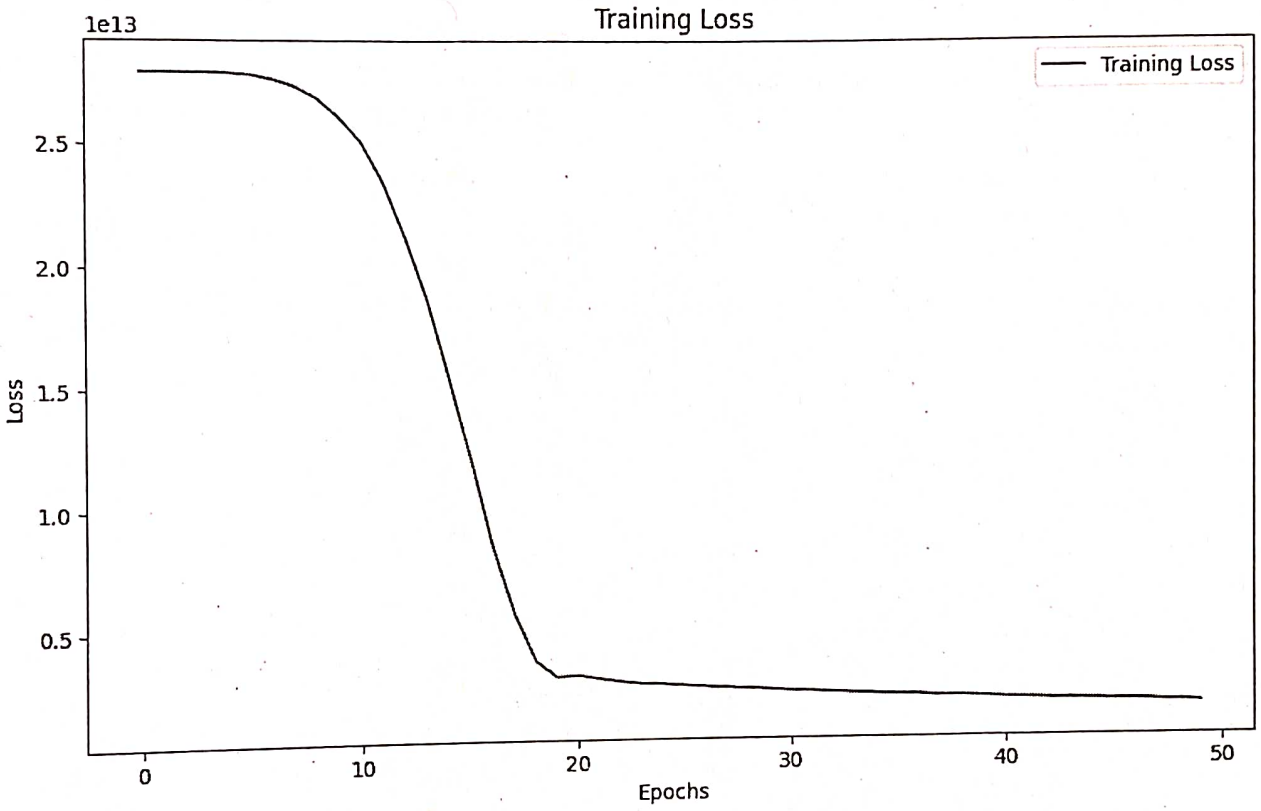


Fig 5.3(e)

Chapter 6 : Conclusion

In this car price prediction model, after seeing performance of various optimizers, we can conclude that Adam optimizer has given the best results, RMSprop and Nadam optimizers are also giving good results but Adagrad and Adadelta optimizers are giving poor results.

The comparision between the performance of these optimizers is:

Adam > RMSprop > Nadam > Adagrad > Adadelta

Hence, we can use this model with Adam, RMSprop or Nadam optimizer for predicting the price of cars and we can also change the number of layers of neurons, activation function, density(number of neurons in a layer) of a layer and learning rate to get better results.

eferences

] E. Gegic, B. Isakovic, D. Keco, Z. Masetic and J. Kevric, "Car Price Prediction using Machine Learning Techniques," *TEM,* vol. 8, no. 1, pp. 113-118, 2019.

] A. D. Sharma and V. Sharma, "USED CAR PRICE PREDICTION USING LINEAR REGRESSION MODEL," *IRJMET,* vol. 2, no. 11, pp. 946-953, 2020.

] E. Pandit, H. Parekh, P. Pashte and A. Natani, "Prediction of Used Car Prices using Machine Learning Techniques," *IRJET,* vol. 09, no. 12, pp. 355-360, 2022.

] P. Rane, D. Pandaya and D. Kotak, "USED CAR PRICE PREDICTION," *IRJET,* vol. 08, no. 04, pp. 1486-1488, 2021.

] K. Agrahari, A. Chaubey, M. Khan and M. Srivastava, "Car Price Prediction Using Machine Learning," *IJIRT,* vol. 8, no. 1, pp. 572-575, 2021.

] K. Samruddhi and R. Kumar, "Used Car Price Prediction using K-Nearest Neighbour Based Model," *IJIRASE,* vol. 4, no. 3, pp. 686-689, 2020.

] V. Kiran, R. kala, V. Nagesh and S. Navdeep, "USED CAR PRICE PREDICTION," *JES,* vol. 13, no. 06, pp. 387-393, 2022.

] K. Noor and S. Jan, "Vehicle Price Prediction System using Machine Learning Techniques," *International Journal of Computer Applications,* vol. 167, no. 09, pp. 27-31, 2017.

] A. Nikhade and R. Borde, "Car Price Prediction using Machine Learning," *IARJSET,* vol. 9, no. 4, pp. 377-382, 2022.