

Introduction to Machine Learning Project Report

Mobile price prediction and Explainability

Team members : Ansh Bansal (B22ME012) ,Ojasvi Pandey (B22ME043), Suralkar Pranav Nanasahab(B22ME066), Praneeth R. Karuturi (B22MT031), Mohit Pareek (B22ME040)

Abstract

This report presents the findings and results of a machine learning project aimed at predicting the price of mobile phones based on various features. The objective is to develop a robust and accurate model that can assist in pricing mobile devices effectively. The project involved data collection, preprocessing, model selection, training, and evaluation. The final model demonstrates promising predictive capabilities.

Objective:

- In this project, our primary focus is on predicting mobile prices using various machine learning models.
- We aim to provide accurate predictions that can assist both consumers and industry stakeholders in understanding and estimating mobile device costs.

Datasets used:

1. <https://drive.google.com/file/d/1c-lgrUY4K7luqVlgGEzaSuWeioc3Al76/view?usp=sharing>
2. https://drive.google.com/file/d/1ziTWfH-OjU1VBzmFcImFZwFt99uTocOm/view?usp=drive_link

Data familiarization:

Significant Correlation:

"Price" and "RAM" exhibit a very high correlation, suggesting the importance of incorporating RAM in predicting mobile phone prices.

Key Correlations:

Explored correlations between "price_range" and other features in the dataset.

Identified top correlations with the target variable ("Price"):

1. RAM
2. PPI (Pixel per inch)
3. Internal Memory
4. Rear Camera

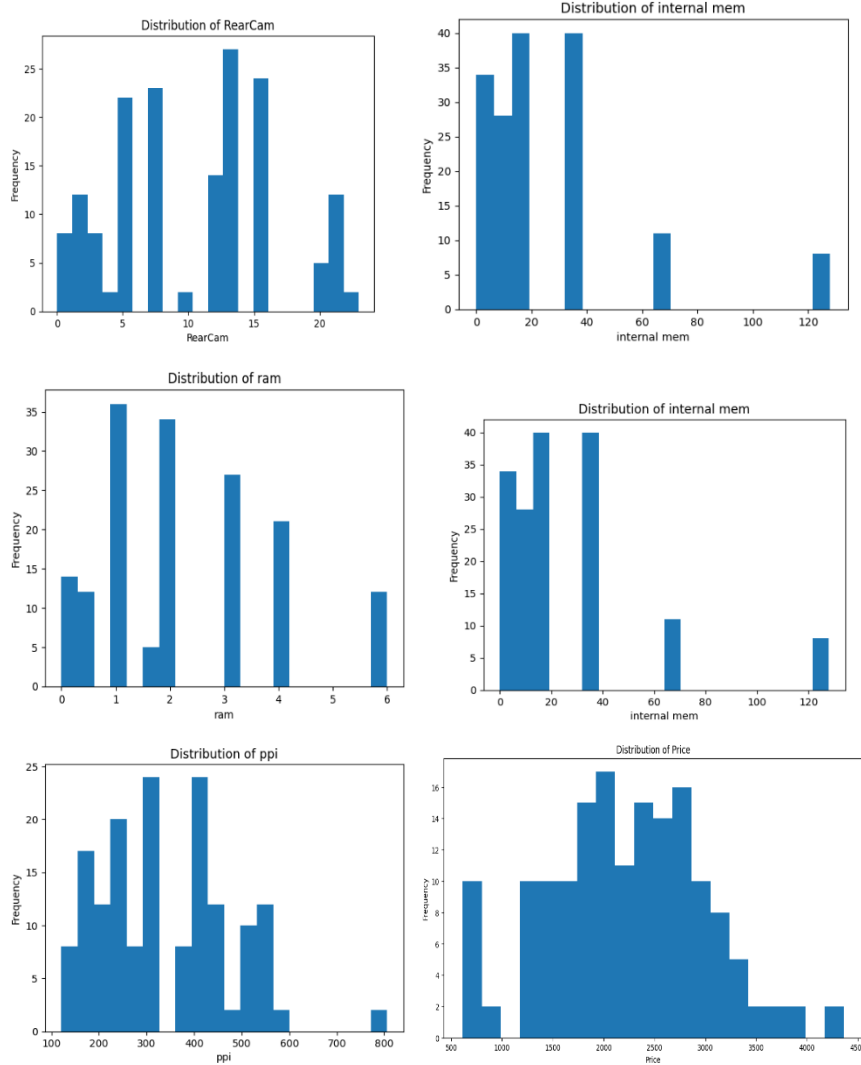
Data Preprocessing:

Data preprocessing involved:

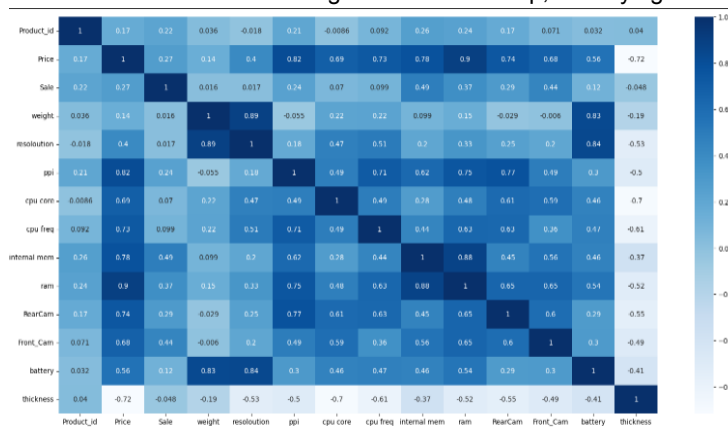
1. Handling null values in the dataset and found that there is no null value in the dataset.
2. Checking for duplicate rows in the dataset.
3. Scaling features.

EDA (Exploratory Data Analysis):

- Analyzed frequency distributions for all columns to understand the dataset's overall distribution.



- Utilized a correlation matrix to generate a heatmap, identifying features correlated with mobile price.

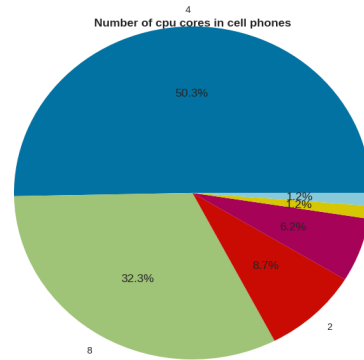
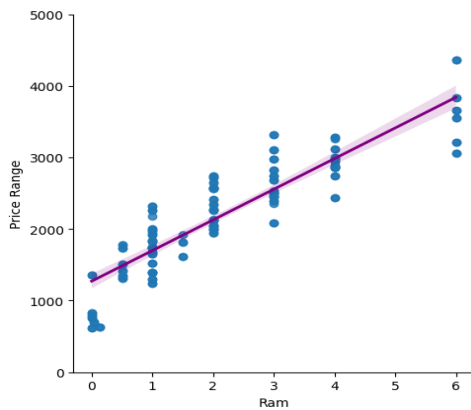


- Identified RAM as the most correlated feature with price through correlation analysis.

- Sorted feature correlations with price in descending order for a clearer understanding.

Price	1.000000
ram	0.896915
ppi	0.817614
internal mem	0.776738
RearCam	0.739538
cpu freq	0.727383
thickness	0.716773
cpu core	0.686811
Front_Cam	0.675286
battery	0.559946
resolution	0.404010
Sale	0.273263
Product_id	0.165136
weight	0.144555

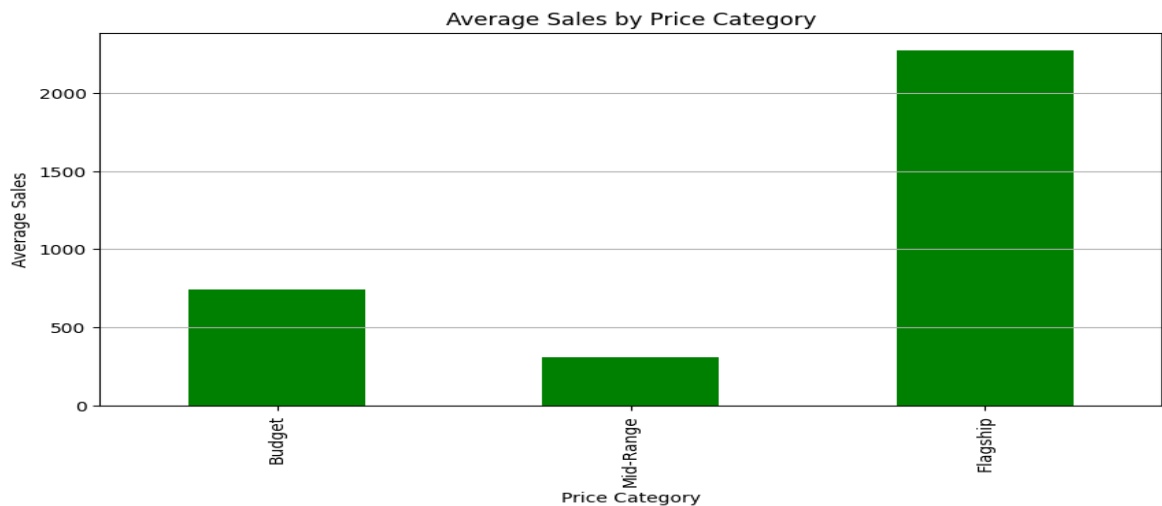
- Plotted a histogram showcasing the frequency distribution of RAM.
- Illustrated a graph depicting a strong positive correlation between mobile price and RAM, indicating.



Scatter plot of "Sales" 'vs.' ``features`` phones.helps in predicting cellphone prices and priorities of consumers when purchasing cell phones.

Plotting a pie chart for the number of cpu cores in cell

- Prices below 1000 are categorized as "Low," prices between 1000 and 3000 are categorized as "Medium," and prices above 3000 are categorized as "High"



- We Have Normalized the Features Using **StandardScaler** Function.

Setting the price column as target column and done train test split in 70 : 30 ratio by using sklearn library.

□ Applying the KNN

We have applied KNN on scaled features and got the following result

Model score = 0.0204

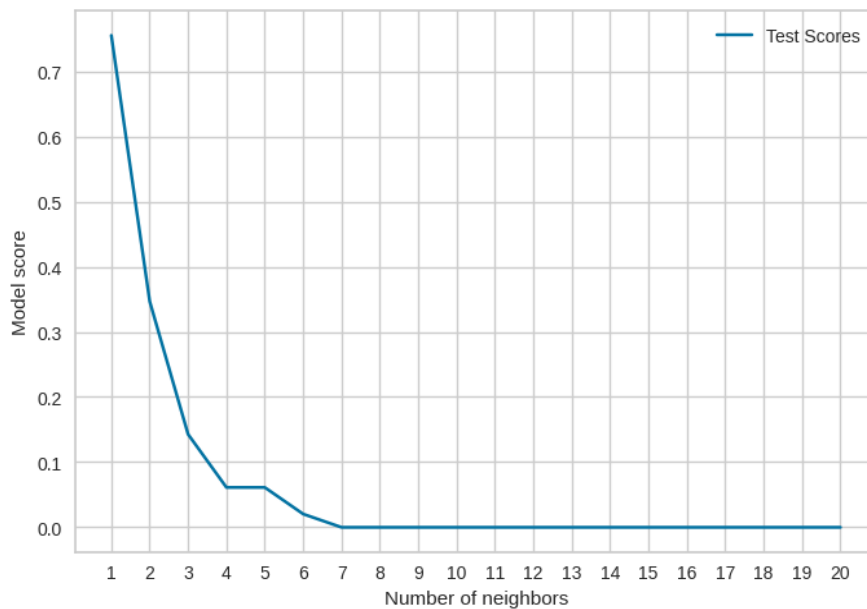
Accuracy achieved by Model = 2.04 %

R-squared Score = 0.7529

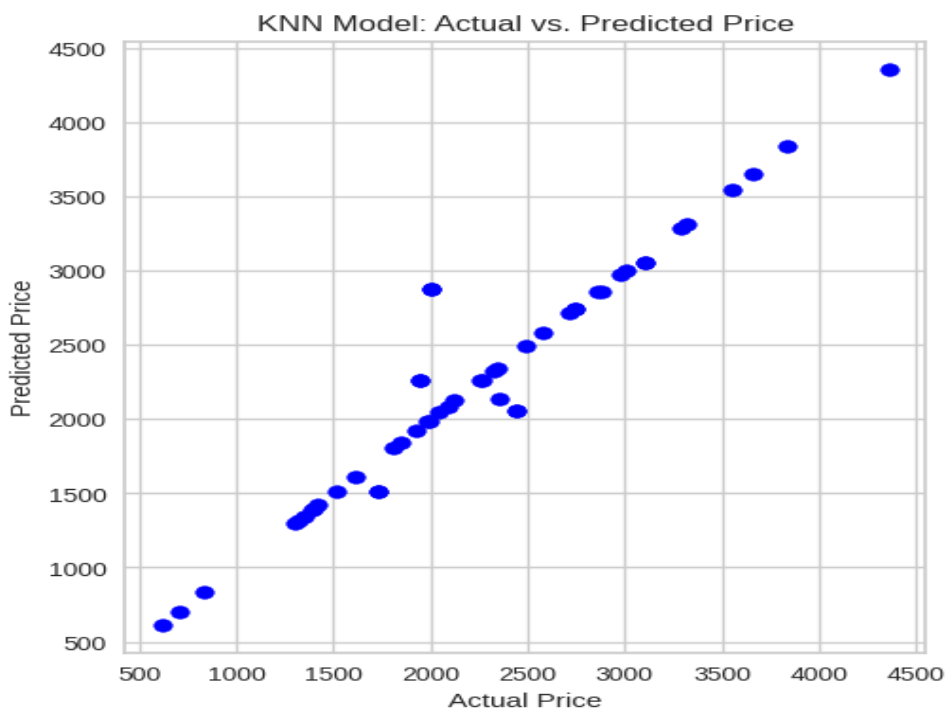
After doing **hyperparameter tuning** ,we got better model score :

For **K=1** => Model score = 75.51 % R2 Score = 0.9299 RMSE score = 212.0606

Graph between Model score and different k values



Graph between predicted and actual price:



❖ Applying Regression tree

We have applied Decision tree regression on scaled features and got the following result and graph:

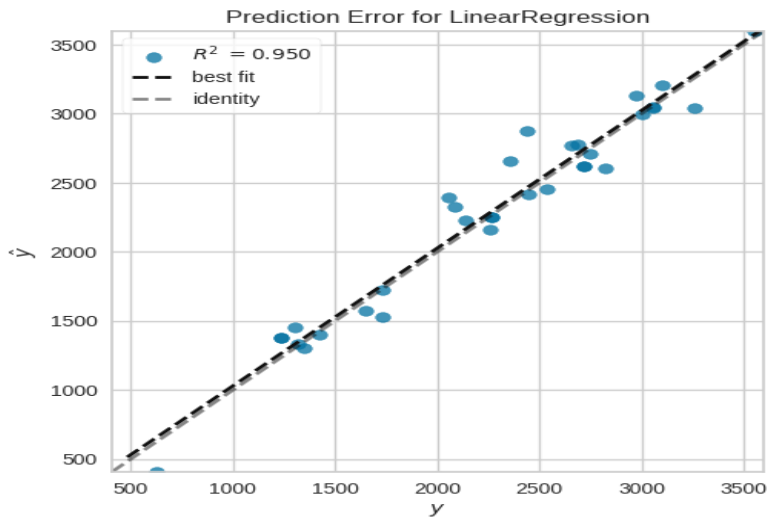


R-squared = **0.8758**

Mean Squared Error = **61266.8182**

❖ Applying Linear Regression:

We have applied Linear regression on scaled features and got the following result and graph:



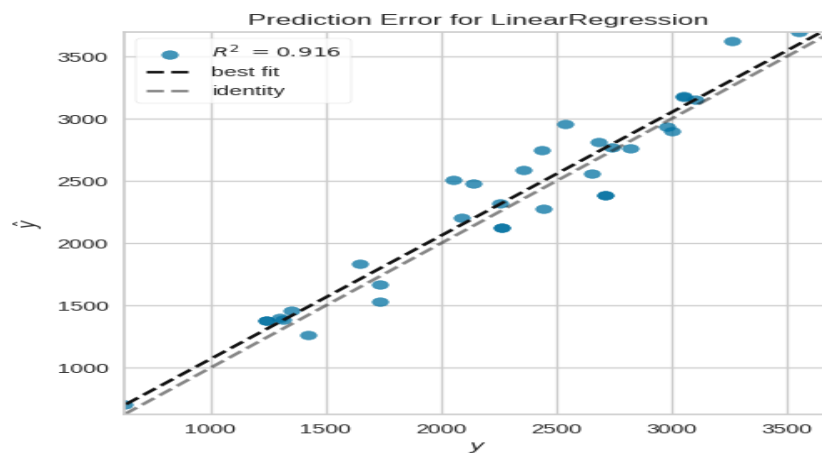
R-squared Score = **0.9499**

Mean Squared Error = **24694.8175**

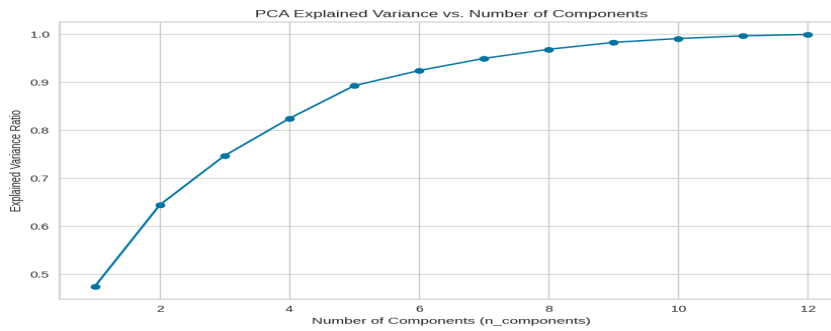
Applying Linear Reg with PCA

R-squared: **0.9161**

Mean Squared Error: **41386.5338**

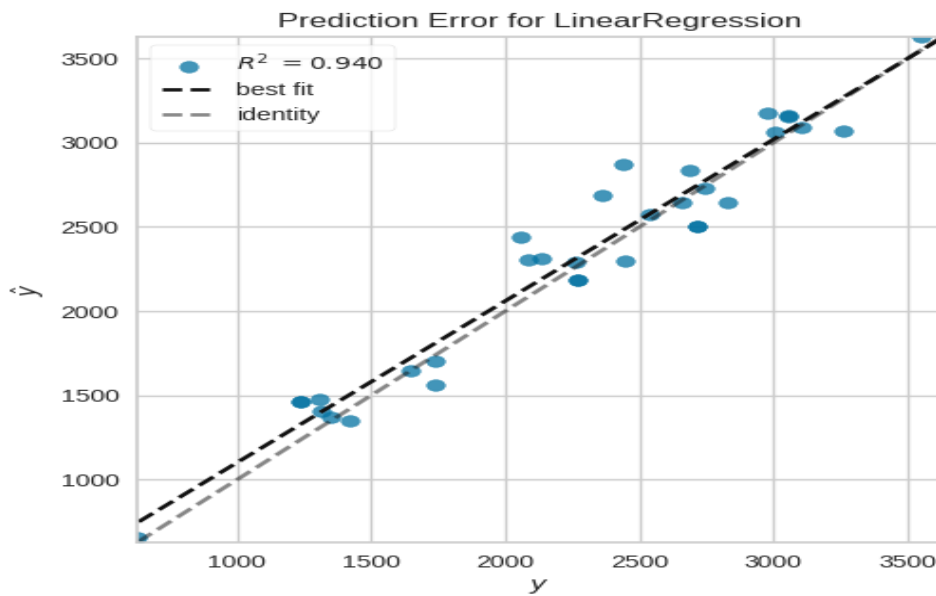


To get better result ,we've checked **best value of number of component in PCA** using Explained variance ratio



Best value of **Number of component** is **8**

Graph between predicted price vs actual price for number of component 8:



R-squared = **0.9396**

MSE = **41386.5338**

Applying Neural Network

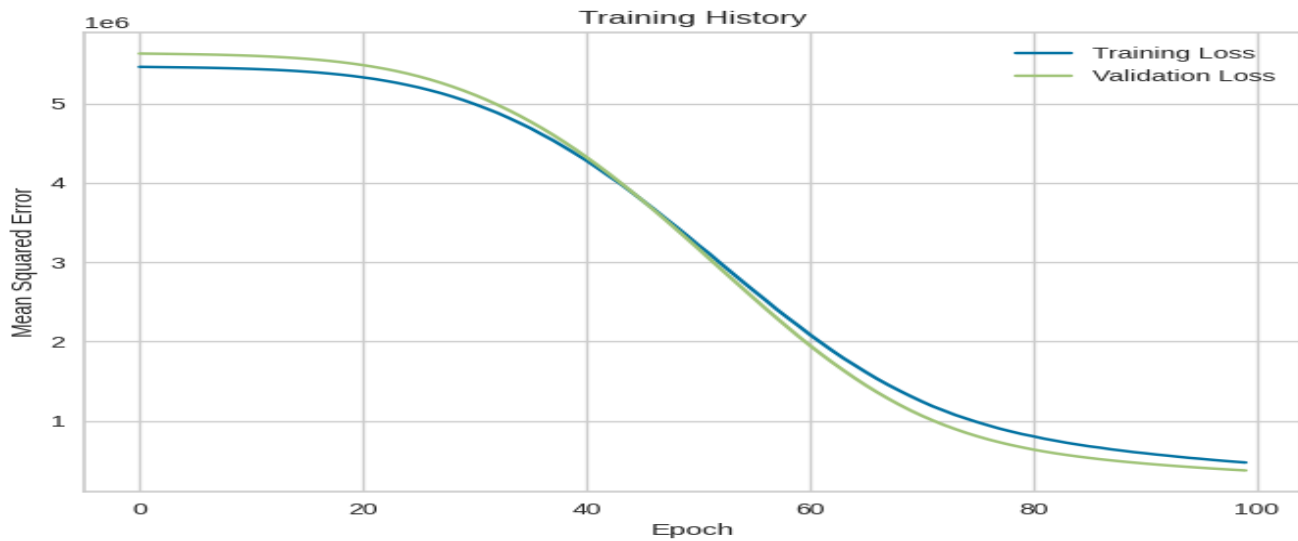
We have applied Neural network on scaled feature and got following results:

Optimizer = **Adam**

Loss function = **Mean squared error**

Batch size = **32**

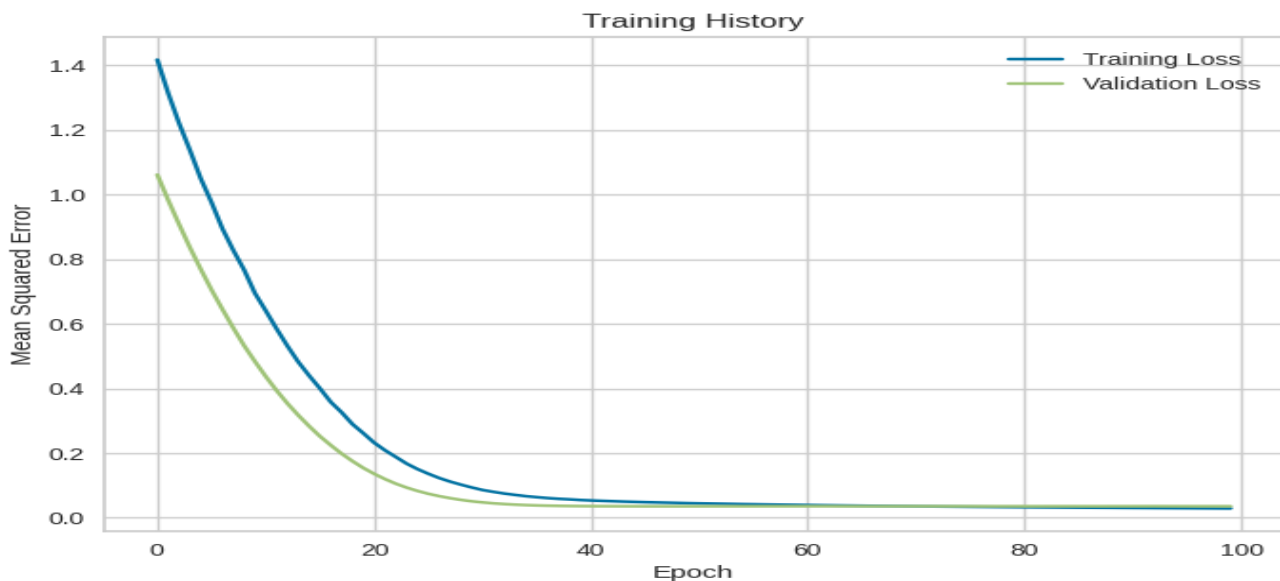
Epoch = **100**



MSE = 376785.6068

R-squared = 0.23615

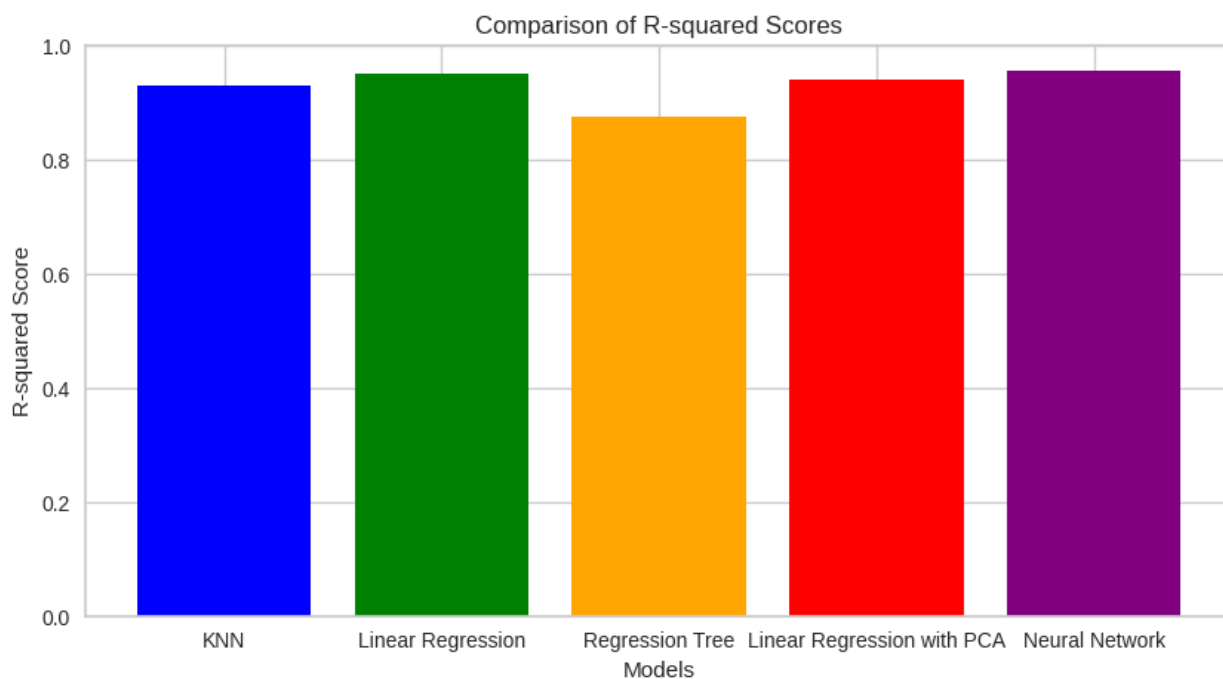
Then we've applied NN on scaled feature and scaled target variable and got the following results and graph:



MSE: 21667.6830

R-squared: 0.9560

- comparison graph of R2 Scores between different Models:



Neural Network has **best R2 score** , which is **0.9560**

Prediction on Custom values

We've predicted the price of cell phone by using Neural Network on custom value:

Actual price = 2357

Predicted price = 2668.0131

❖ Applying NN Model on Custom Dataset :

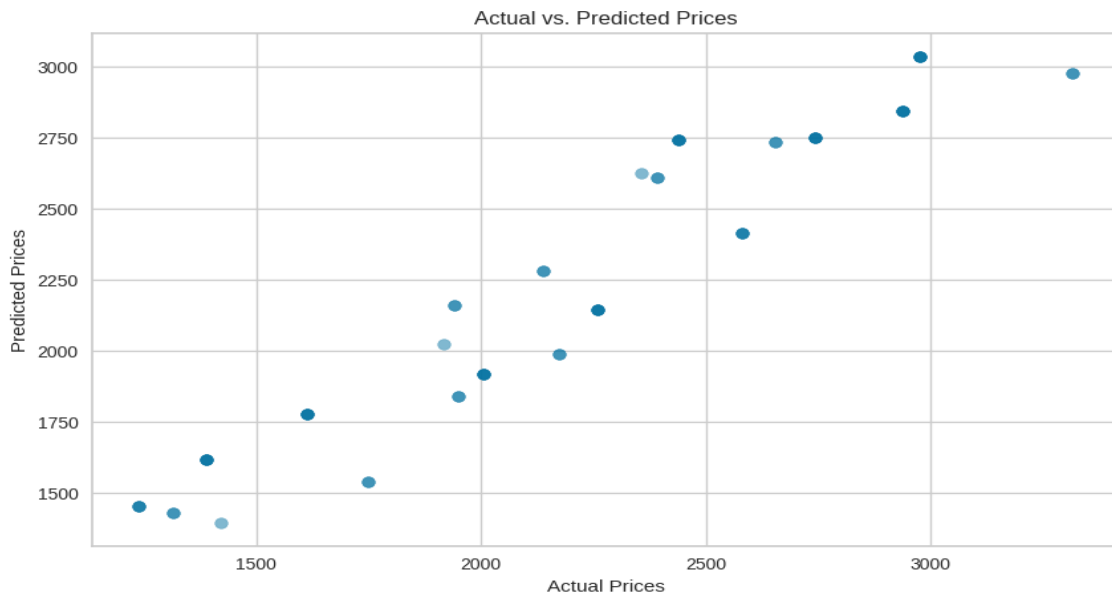
	Product_id	Price	Sale	weight	resolution	ppi	cpu core	cpu freq	internal mem	ram	RearCam	Front_Cam	battery	thickness
0	289	2438	16	162.0	5.3	277	8	1.5	32	4.0	13.0	8.0	4000	7.7
1	605	2006	16	161.0	5.5	200	8	1.4	4	1.0	5.0	0.0	2500	8.9
2	622	2174	16	140.0	5.0	294	4	1.3	16	1.0	13.0	5.0	2000	8.2
3	1058	2744	16	174.0	5.6	524	4	2.7	32	3.0	16.0	3.7	3000	8.3
4	1103	2580	16	145.0	5.1	432	4	2.5	16	2.0	16.0	2.0	2800	8.1

Mean Squared Error = 33466.3318

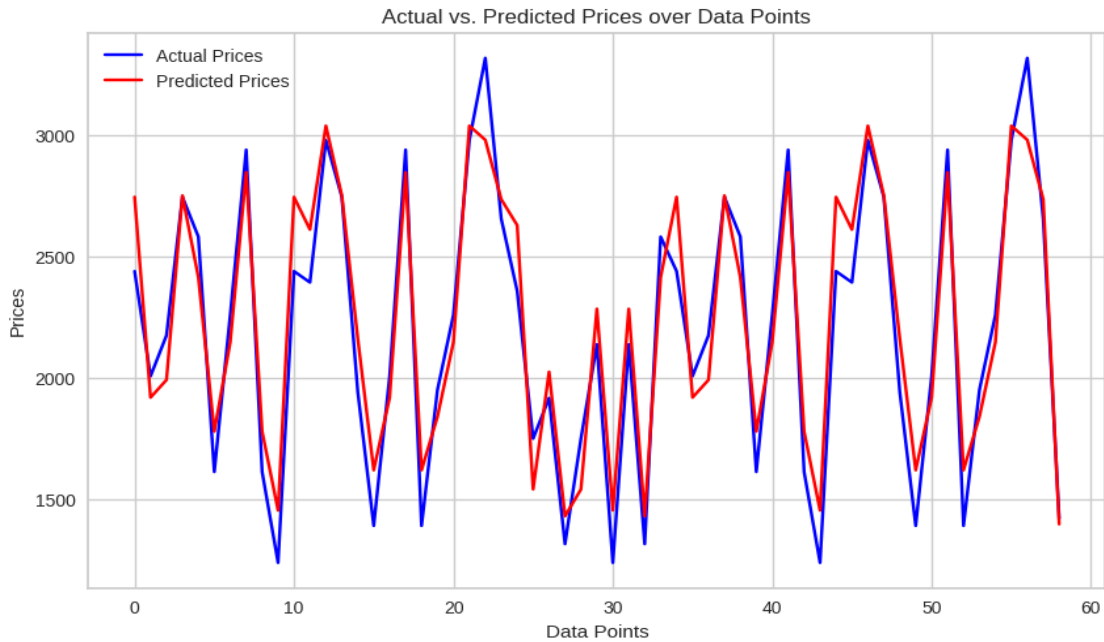
R-squared = 0.8977

Accuracy On Custom Dataset = 88.14%

- **Scatter plot between Actual and Predicted price:**



- **Graph between Actual and Predicted price :**



----- END -----