

<https://colab.research.google.com/drive/1mfjOKgtl9ocR0yyDlInnDO5qtu6a0kn4?usp=sharing>

Introduction To Machine Learning Project

Mobile price prediction and explainability

PRANEETH RAM K (B22MT031)

ANSH BANSAL(B22ME012)

OJASVI PANDEY(B22ME043)

SURALKAR PRANAV NANASAHEB (B22ME066)

MOHIT PAREEK(B22ME040)

OBJECTIVE :

To build a model that accurately predicts the prices of mobile phones based on various features and attributes. Additionally, we aim to create an understandable and interpretable model, providing insights into why the model made certain predictions.

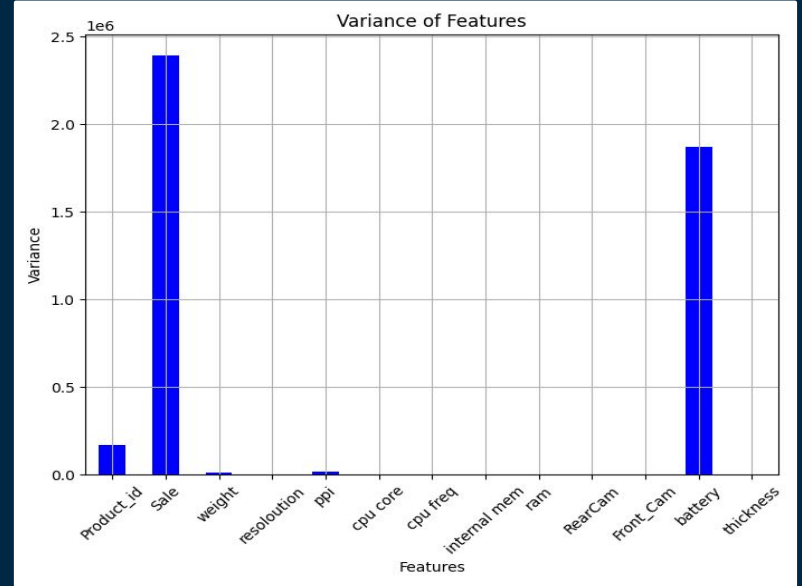


Data familiarization

Data set used :

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

	Product_id	Price	Sale	weight	resolution	ppi	cpu core	cpu freq	internal mem	ram	RearCam	Front_Cam	battery	thickness	
	150	826	614	2171	69.8	1.4	129	0	0.000	0.000	0.004	0.0	0.0	800	14.1
	149	826	614	2159	69.8	1.4	129	0	0.000	0.000	0.004	0.0	0.0	800	14.1
	81	1026	791	106	66.0	1.5	121	1	0.208	0.000	0.004	0.0	0.0	800	13.2
	86	1026	791	114	66.0	1.5	121	1	0.208	0.000	0.004	0.0	0.0	800	13.2
	113	64	754	308	77.9	2.4	167	0	0.000	0.004	0.004	0.0	0.0	850	12.4
	110	64	754	299	77.9	2.4	167	0	0.000	0.004	0.004	0.0	0.0	850	12.4
	125	1198	705	423	110.0	2.2	128	0	0.000	0.128	0.032	2.0	0.0	900	15.6
	126	1198	705	427	110.0	2.2	128	0	0.000	0.128	0.032	2.0	0.0	900	15.6
	141	701	628	1274	102.9	2.2	128	0	0.000	0.256	0.128	1.3	0.0	950	18.5
	140	701	628	1224	102.9	2.2	128	0	0.000	0.256	0.128	1.3	0.0	950	18.5



Insights from 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

Keypoints :

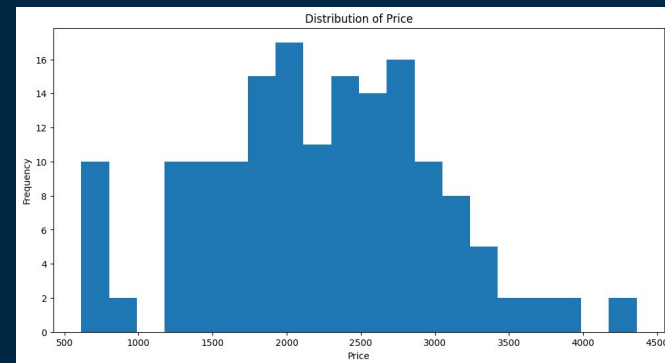
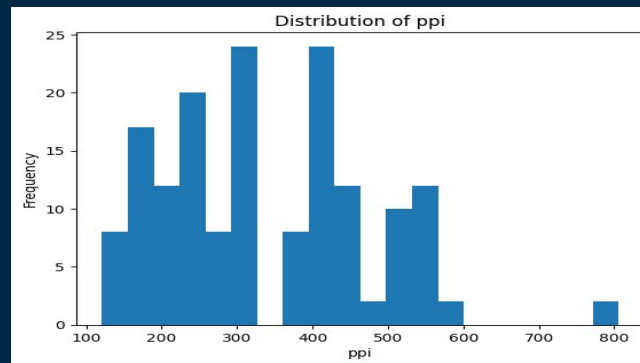
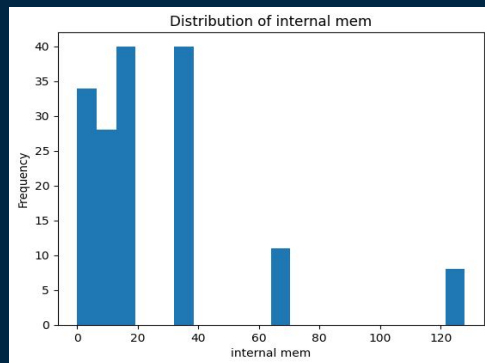
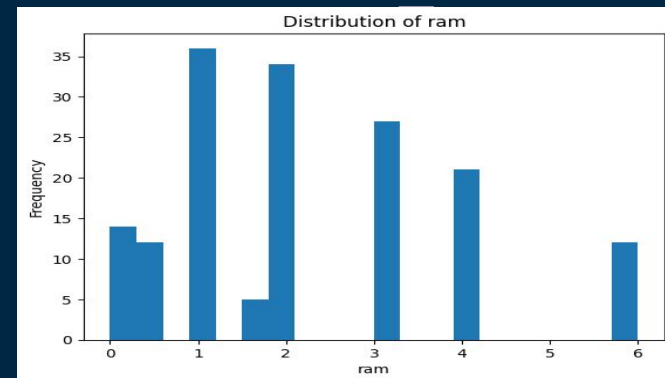
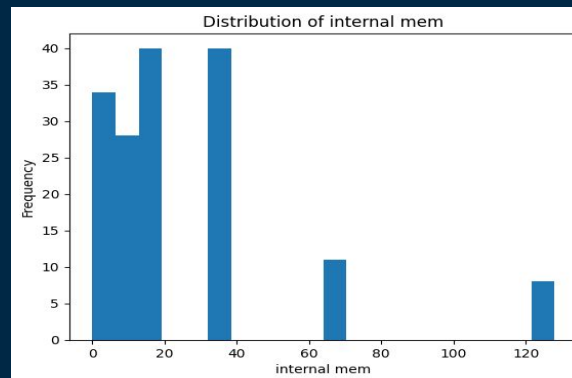
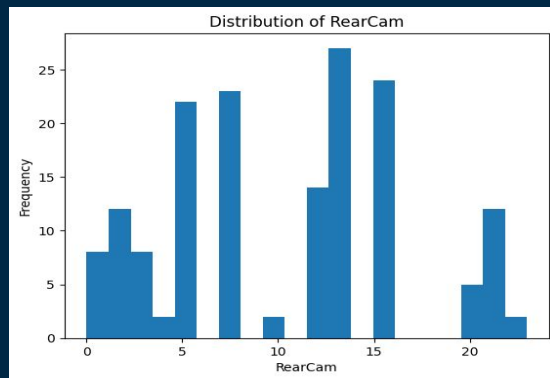
Data preprocessing involved:

1. Handling null values in the dataset and found that there is no null value 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.
- Plotted a histogram showcasing the frequency distribution of RAM.
- Illustrated a graph depicting a strong positive correlation between mobile price and RAM, indicating a general trend of price increase with higher RAM

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



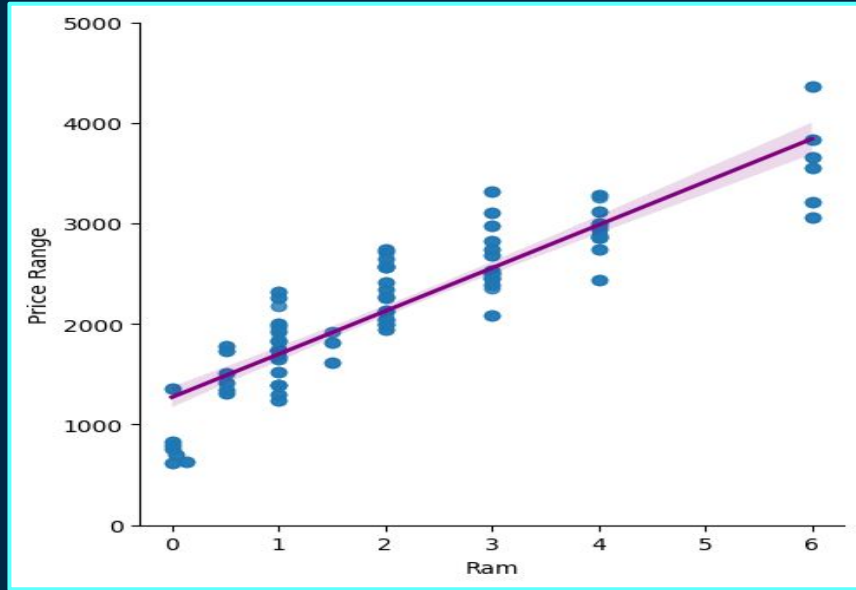
- Utilized a correlation matrix to generate a heatmap, identifying features correlated with mobile price.



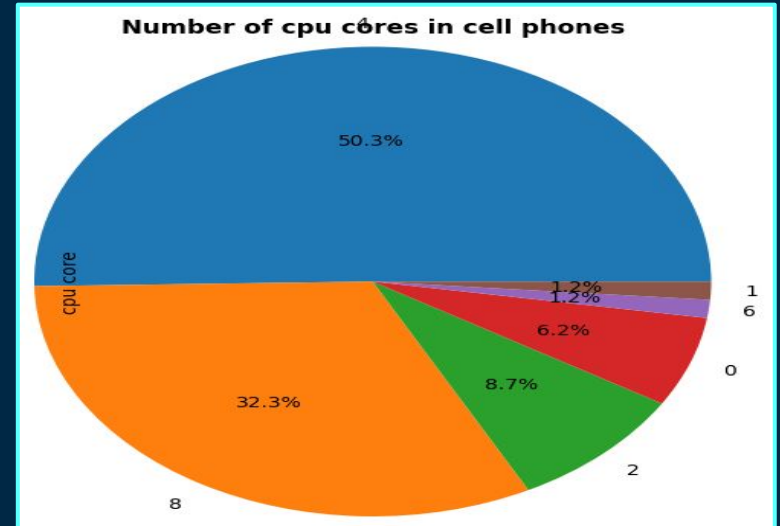
```
Price 1.000000
ram 0.896915
ppi 0.817614
internal mem 0.776738
RearCam 0.739538
cpu freq 0.727383
thickness 0.716773
cpu core 0.686811
Front_Cam 0.675286
battery 0.559946
resolution 0.404010
Sale 0.273263
Product_id 0.165136
weight 0.144555
```

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

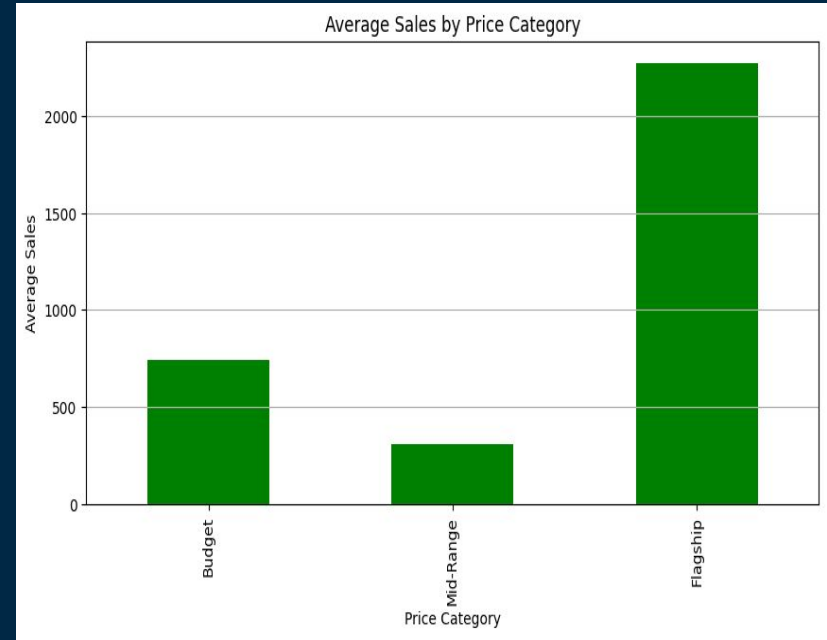
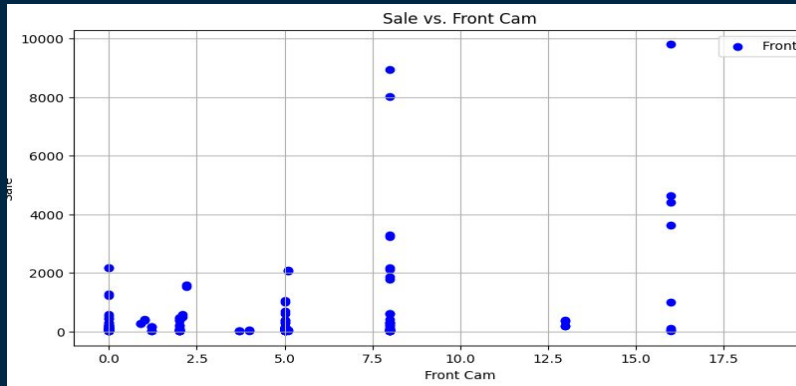
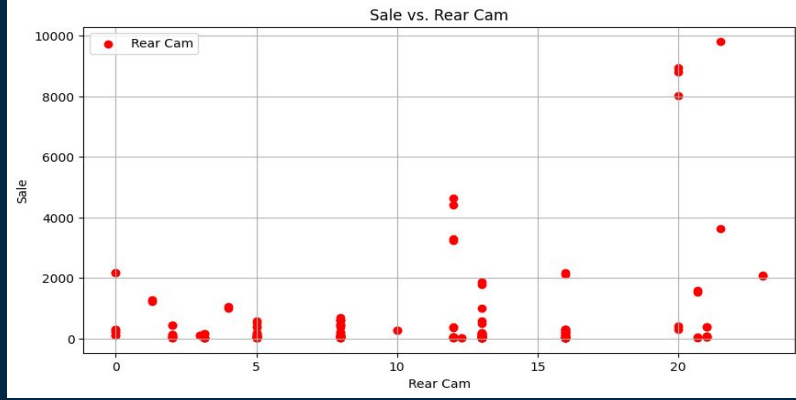
- Illustrated a graph depicting a strong positive correlation between mobile price and RAM, indicating a general trend of price increase with higher RAM .



- Plotting a pie chart for number of cpu cores in cell phones.



Scatter plot of "Sales" vs. "features" helps in predicting cellphone prices and priorities of consumers when purchasing cellphones.



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

Normalizing the dataset

Normalization by use of standard scaler function we have normalized the all features

	cpu_freq	ram	RearCam	Front_Cam
0	-2.513446	-1.371486	-1.684138	-1.040327
1	-2.513446	-1.371486	-1.684138	-1.040327
2	-2.165572	-1.371486	-1.684138	-1.040327
3	-2.165572	-1.371486	-1.684138	-1.040327
4	-2.513446	-1.371486	-1.684138	-1.040327
...
156	0.747876	2.364749	0.263169	0.807868
157	1.266342	-0.127736	-0.872760	-0.578278
158	1.266342	-0.127736	-0.872760	-0.578278
159	0.162510	0.495385	-0.385934	-0.578278
160	0.162510	0.495385	-0.385934	-0.578278

Model training and testing on scaled data

- Setting the price column as target column as we are predicting the prices.
- We have applied the same 5 models on scaled data.
- We have 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

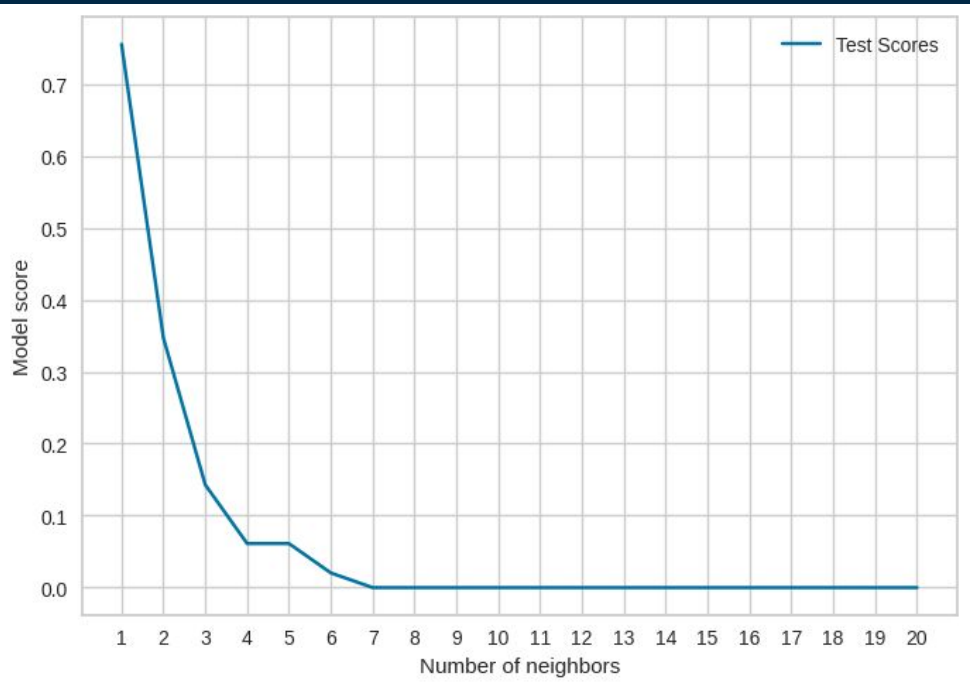
- After doing hyperparameter tuning, we got better model score

For K=1 => **Model score = 75.51 %**

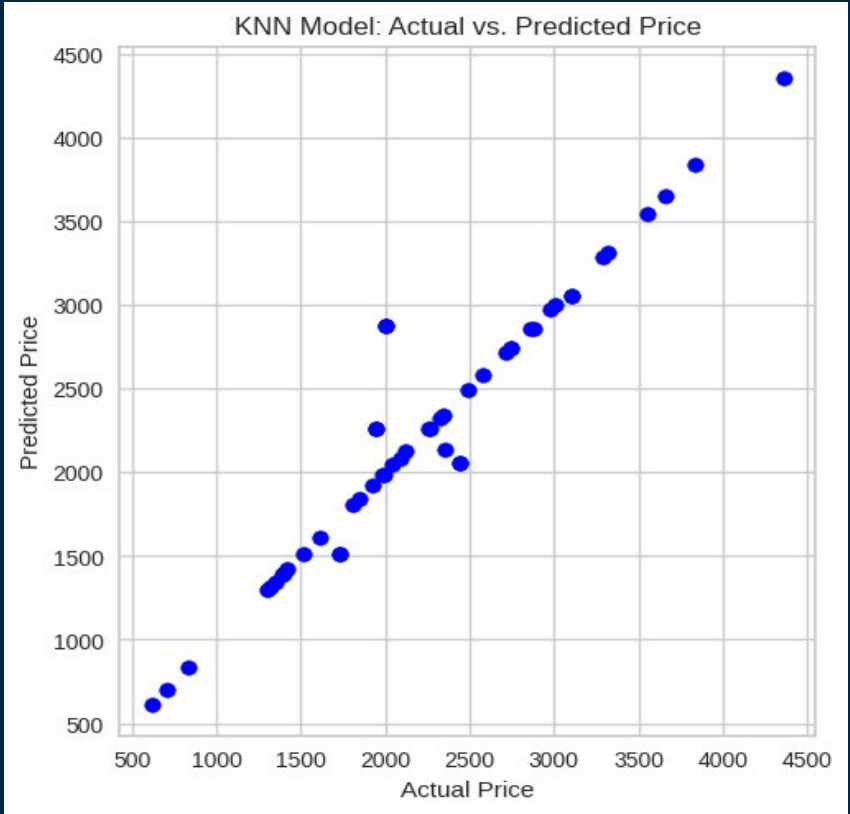
R2 Score = 0.9299

RMSE score = 212.0606

Graph between Model score and different k values

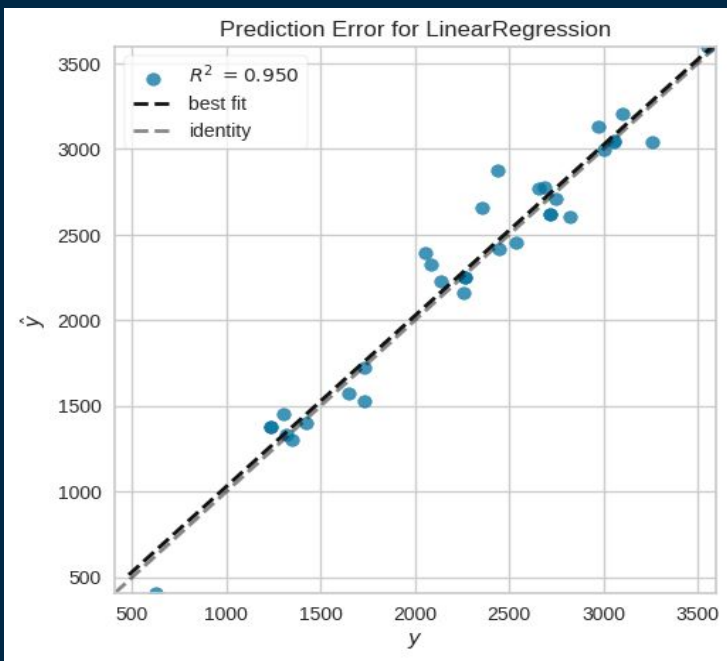


Graph between predicted and actual price



Applying Linear Regression

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

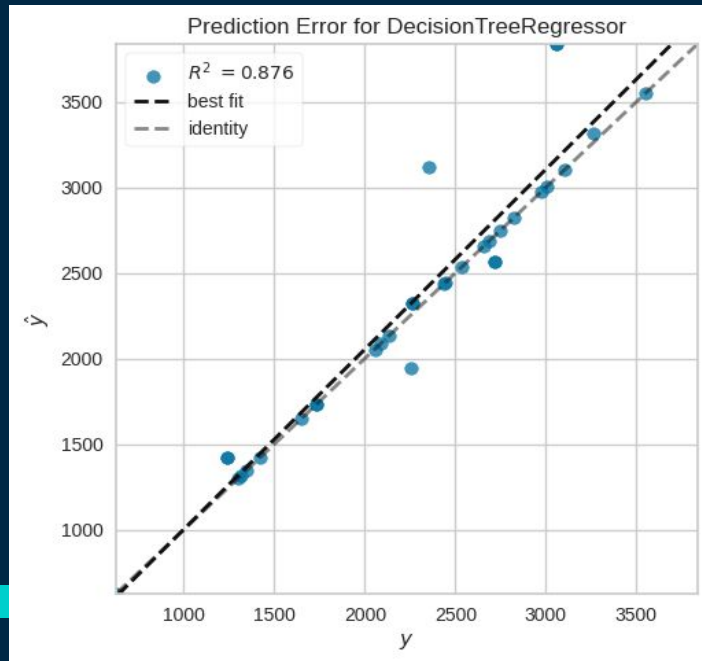


R-squared Score = 0.9499

Mean Squared Error = 24694.8175

Applying Regression tree

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

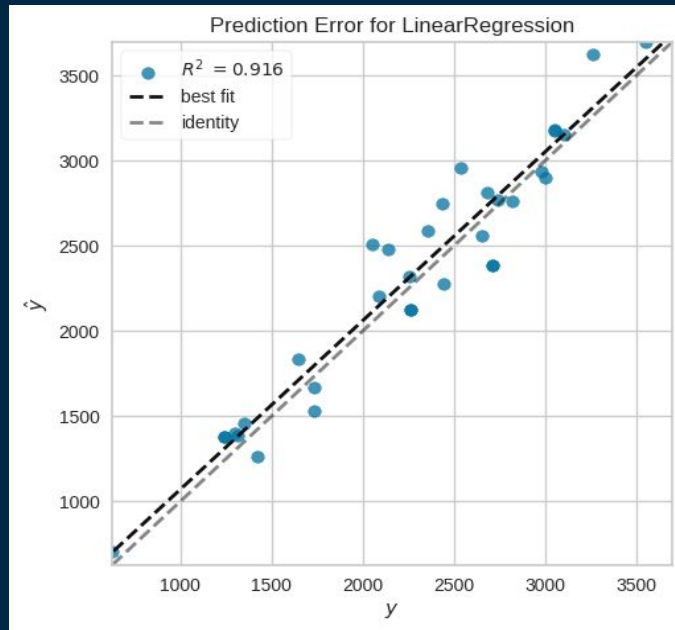


R-squared = 0.8758

Mean Squared Error = 61266.8182

Applying Linear Reg with PCA

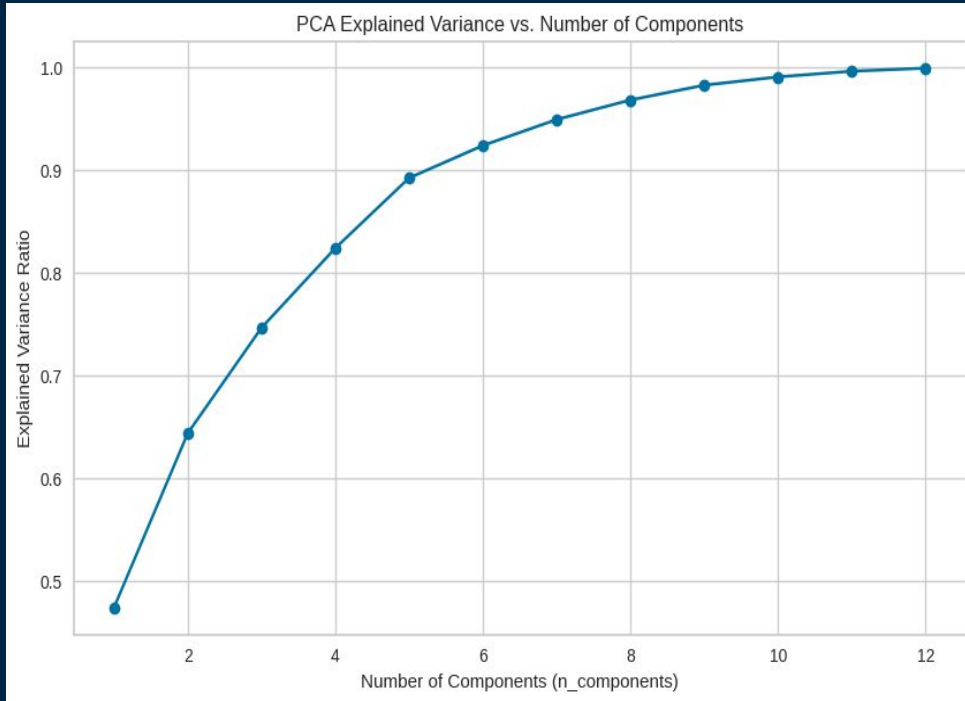
- We have applied Linear Regression with PCA on scaled features and got the following result and graph



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



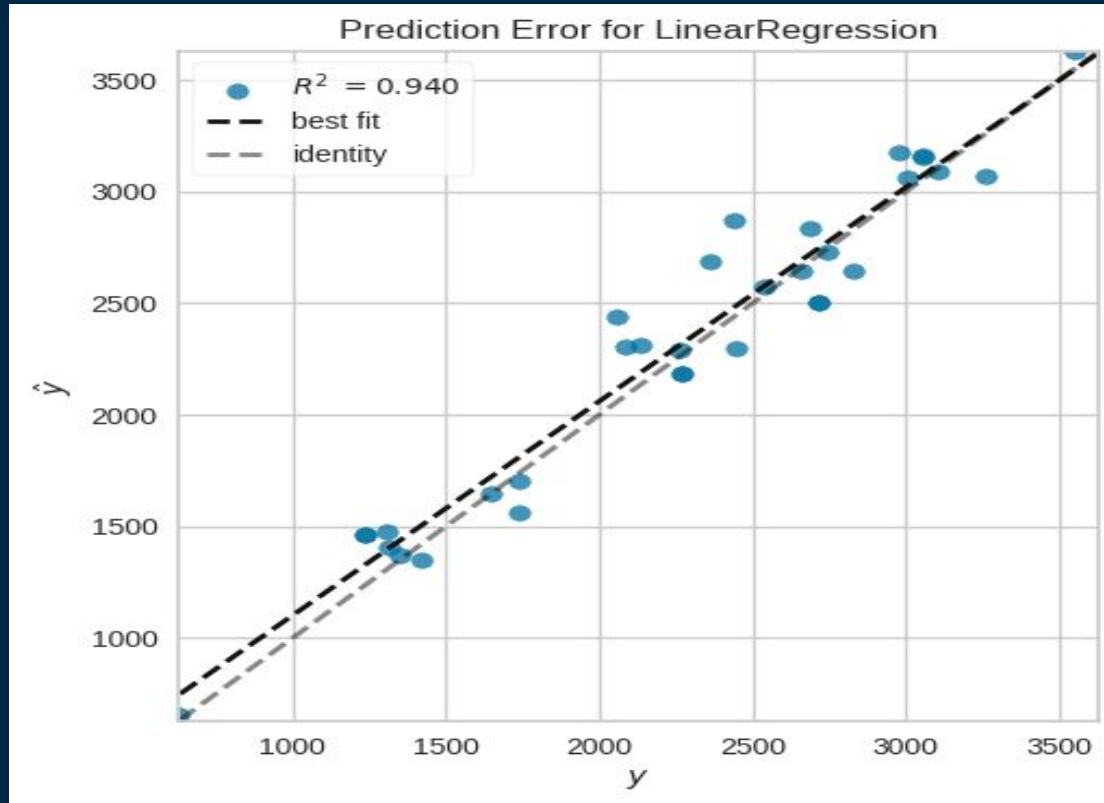
Best value of Number of component is **8**

For no of component = 8

R-squared = 0.9396

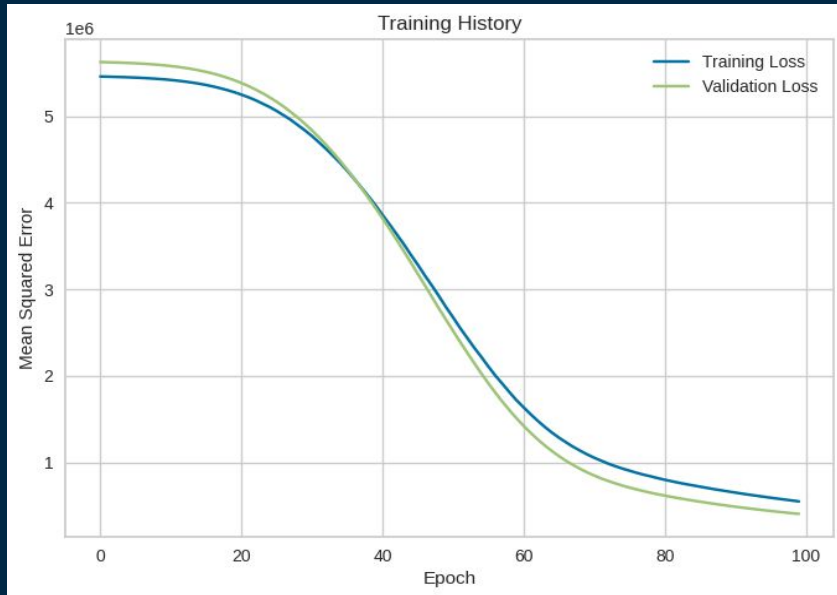
MSE = 41386.5338

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



Applying Neural Network

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



Optimizer = Adam

Loss function = Mean squared error

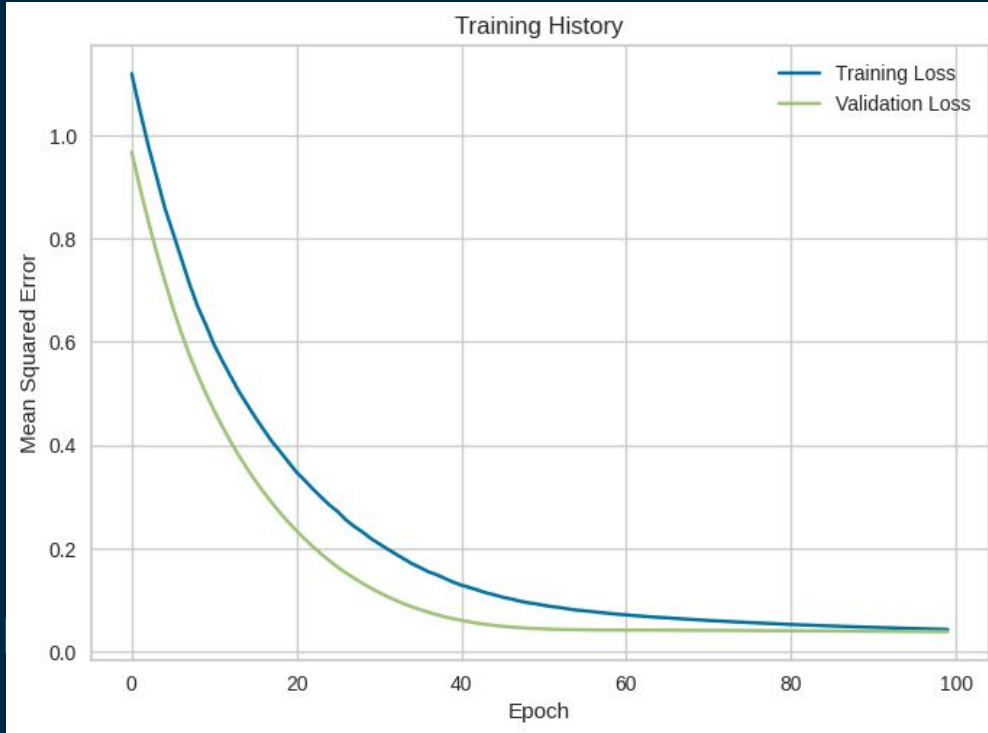
Batch size = 32

Epoch = 100

MSE = 376785.6068

R-squared = 0.2362

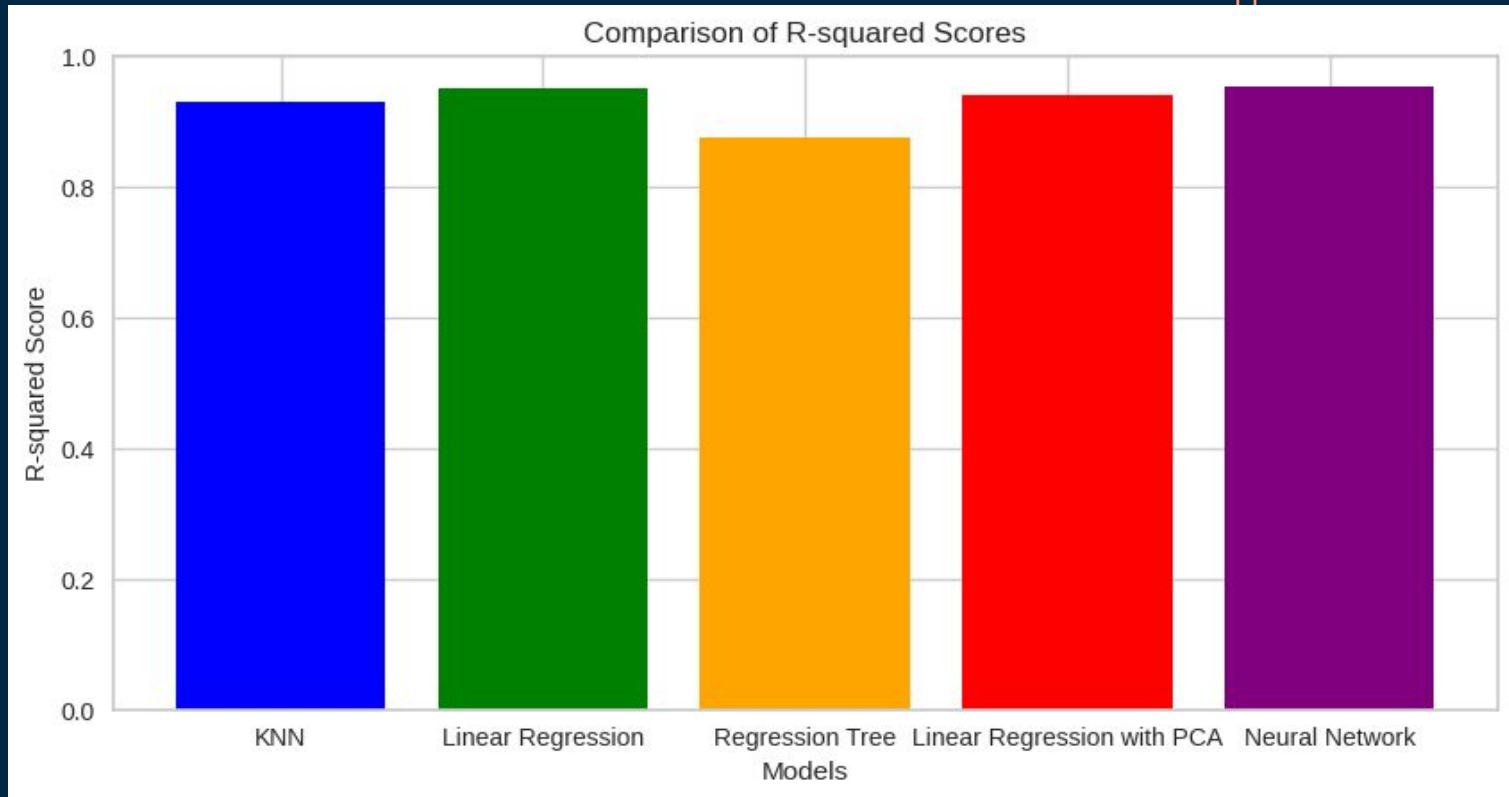
- Then we've applied NN on scaled feature and scaled target variable and got the following results and graph



MSE = 21667.6831

R-squared = 0.9561

Comparison graph of R2 Scores between different Models



- Neural Network has best R2 score ,which is 0.9561

Prediction on Custom values

- We've predicted the price of cell phone by using Neural Network on custom values

Actual price = 2357

Predicted price = 2668.0132

Applying NN Model on Custom Dataset

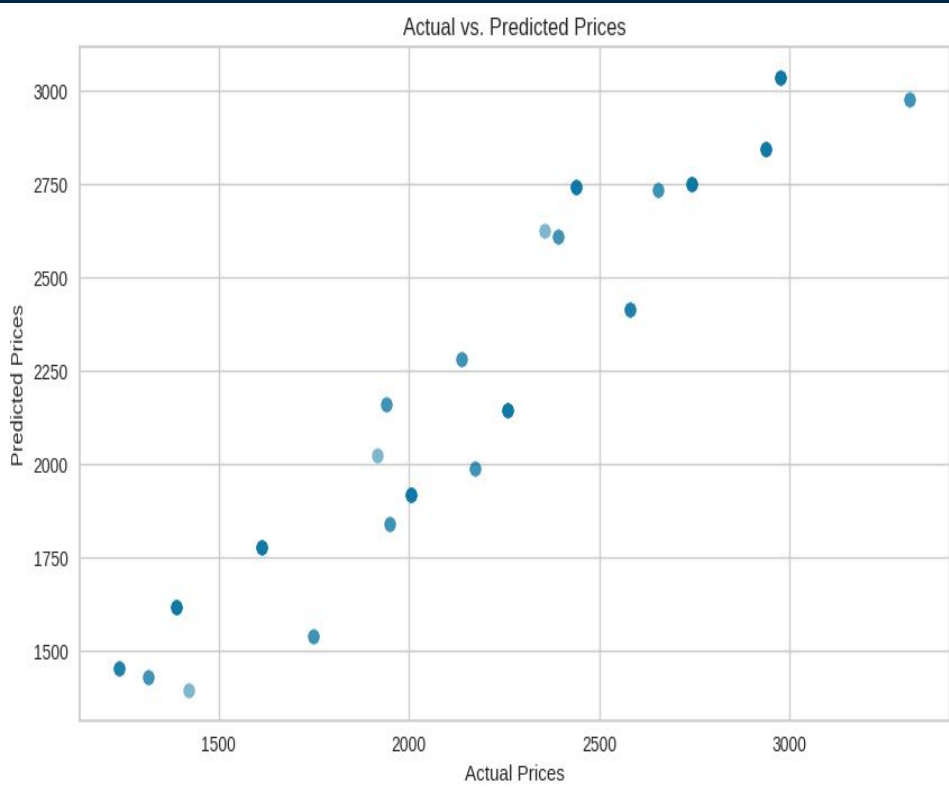
	Product_id	Price	Sale	weight	resoloution	ppi	cpu core	cpu freq	internal mem	ram	RearCam	Front_Cam	battery	thickness
0	289	2438	16	162.0	5.3	277	8	1.5	32	4.0	13.0	8.0	4000	7.7
1	605	2006	16	161.0	5.5	200	8	1.4	4	1.0	5.0	0.0	2500	8.9
2	622	2174	16	140.0	5.0	294	4	1.3	16	1.0	13.0	5.0	2000	8.2
3	1058	2744	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

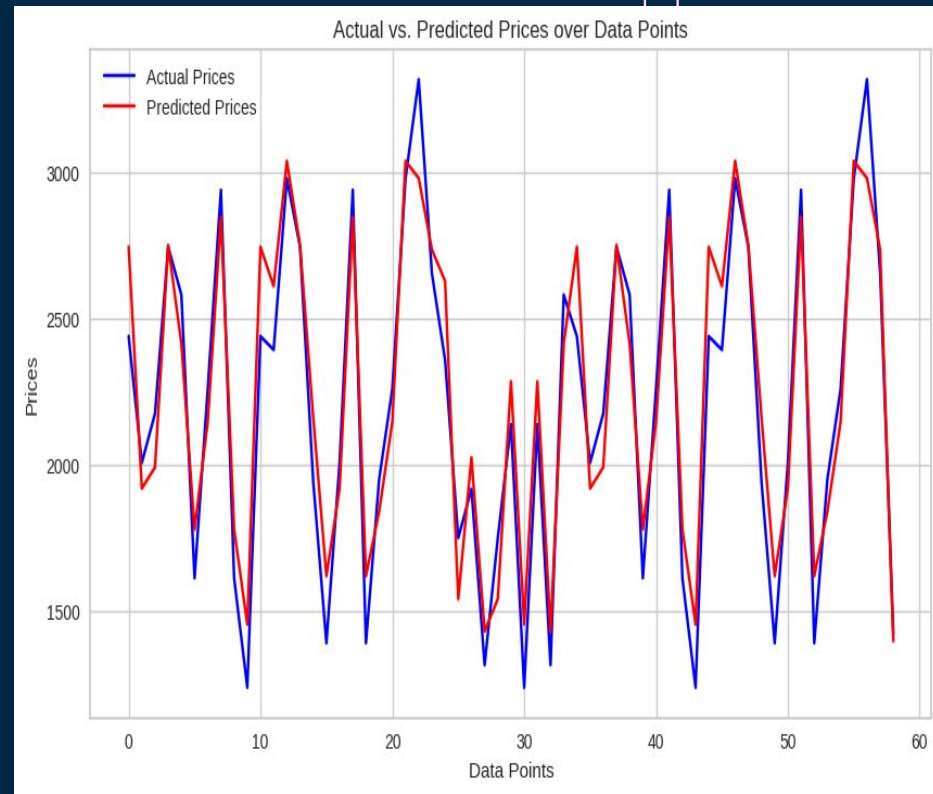
R-squared = 0.8977

Accuracy On Custom Dataset = 88.14%

Scatter plot between Actual and Predicted price



Graph between Actual and Predicted price



THANKS!!