Capstone Project-3

Mobile Price Range Prediction

Member

Sourabh Pramanik

Contents

Problem Statement

Here in this project we have one dataset-

1. Data Mobile Price Range

In today's world there are lots of features present in mobile phones and those features decide the price range of mobile.
In this project we will try find some important features of phones and apply different algorithms to predict the price.

Data Summary

We will complete this project by using following steps-

- After reading the data we will perform Exploratory Data analysis.
- We will check the Null values and Outliers present in our Dataset.
- We will do some statistical analysis of the data.
- We will check the distribution of all the numerical columns and correlation between the variables.
- After then we will apply different Machine Learning models and will check the performance of the Models by using some performance metrics.

Data Summary

Outcomes of this Project -

- Cost prediction of different types of mobiles
- Selection of appropriate features to decide the cost.
- Providing the best product to customer according to economic range .

Independent Variables

Some important Independent variables -

Battery power, Bluetooth, Dual sim, Front camera, 4G or 3G, RAM, Touch Screen, WiFi.

Dependent Variables

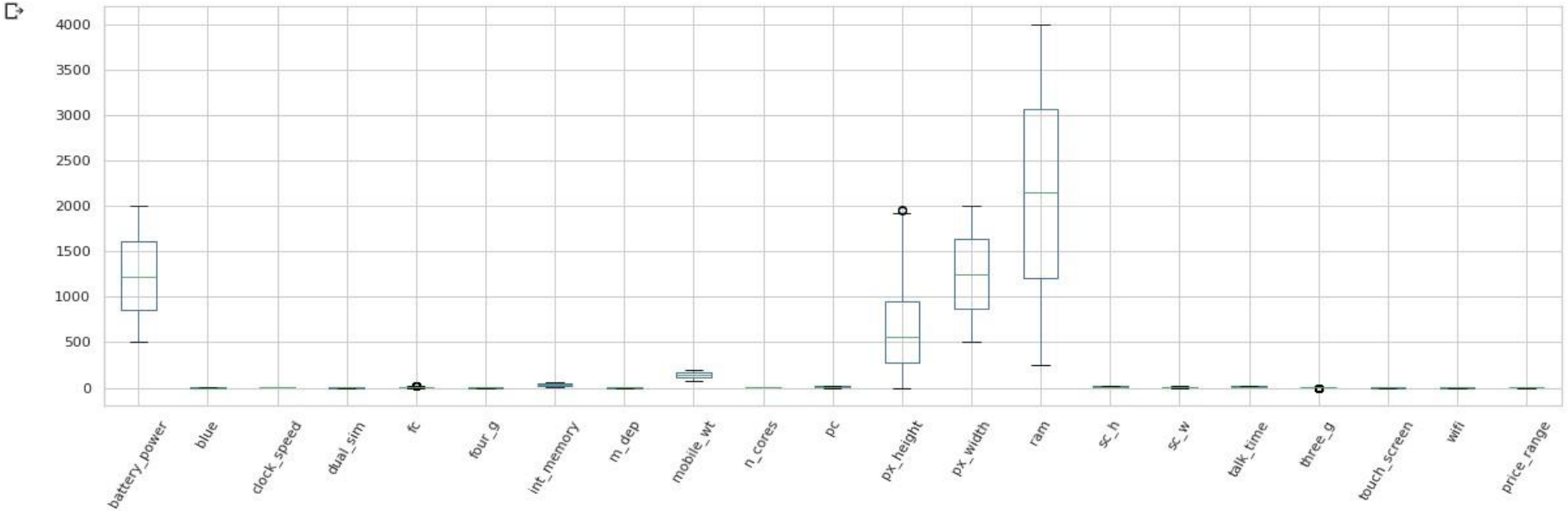
Dependent variable-

Price Ranges-
4 Types - 0(Low)
1(Medium)
2(High Cost)
3(Very High Cost)

Checking Missing values and Outliers

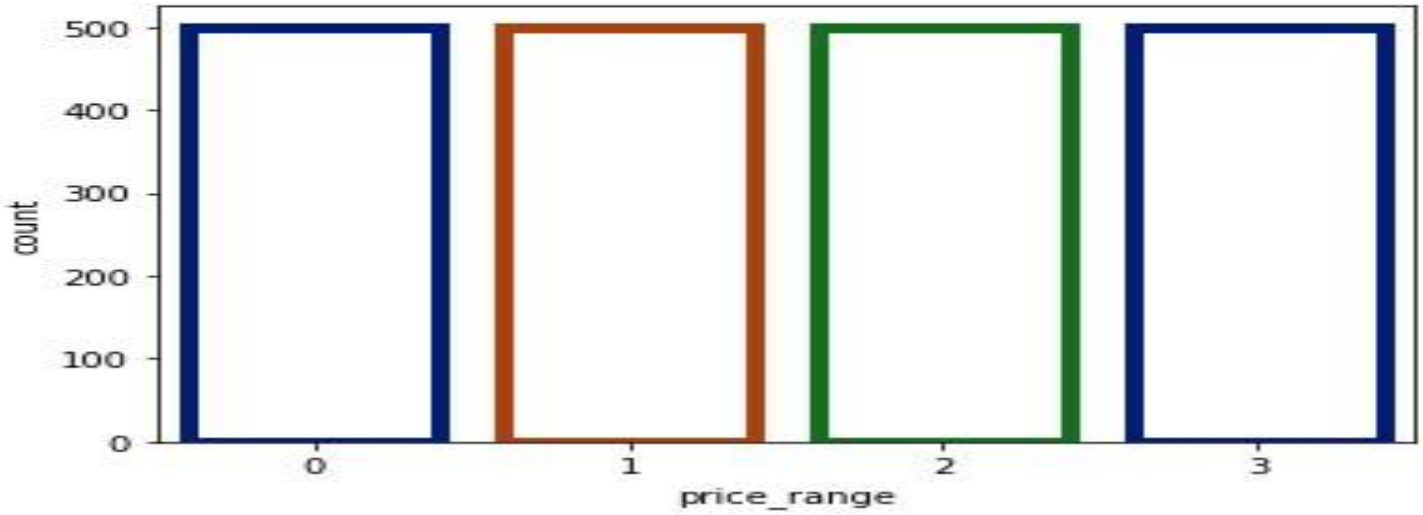
```
sns.set_theme(style="whitegrid")
df.plot(kind='box',figsize=(20,6))
plt.xticks(rotation=60)
plt.show()
```

OUTLIERS

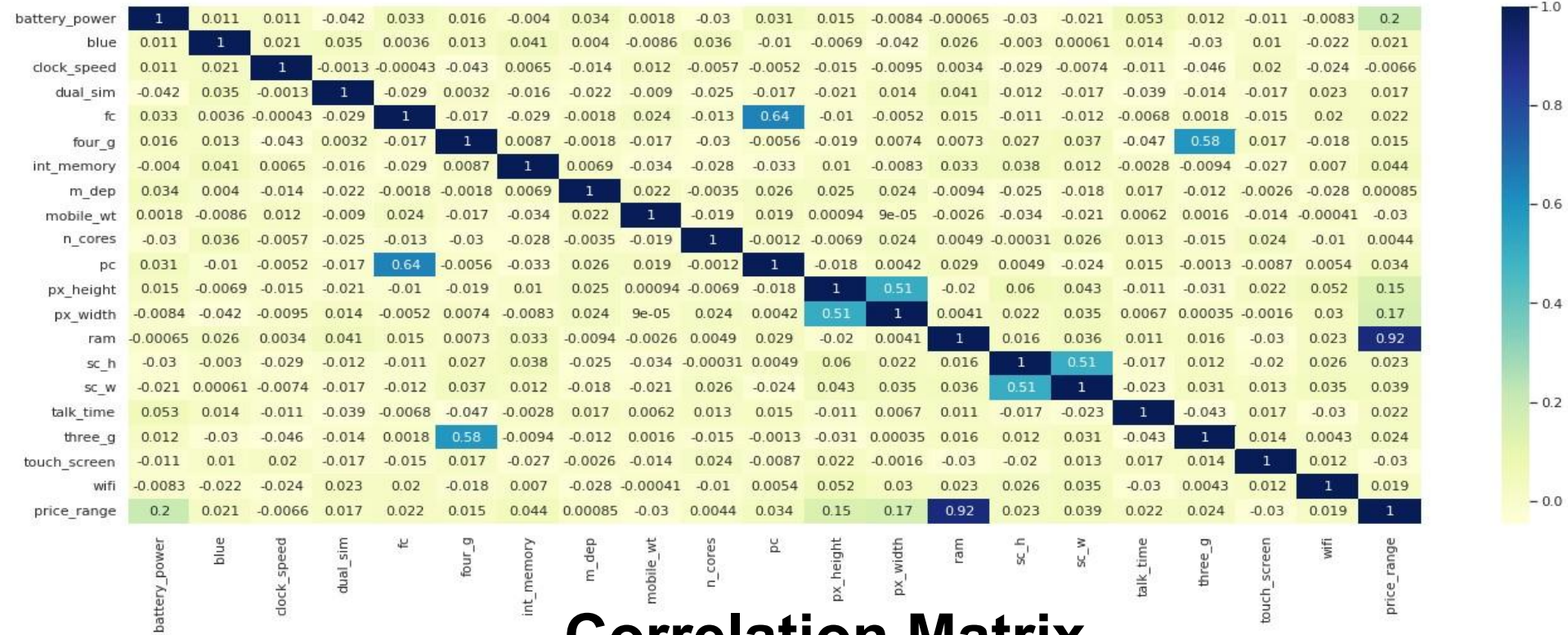


EDA and Visualization

```
[6]  ax = sns.countplot(x="price_range", data=df,
                        facecolor=(0, 0, 0, 0),
                        linewidth=5,
                        edgecolor=sns.color_palette("dark", 3))
```

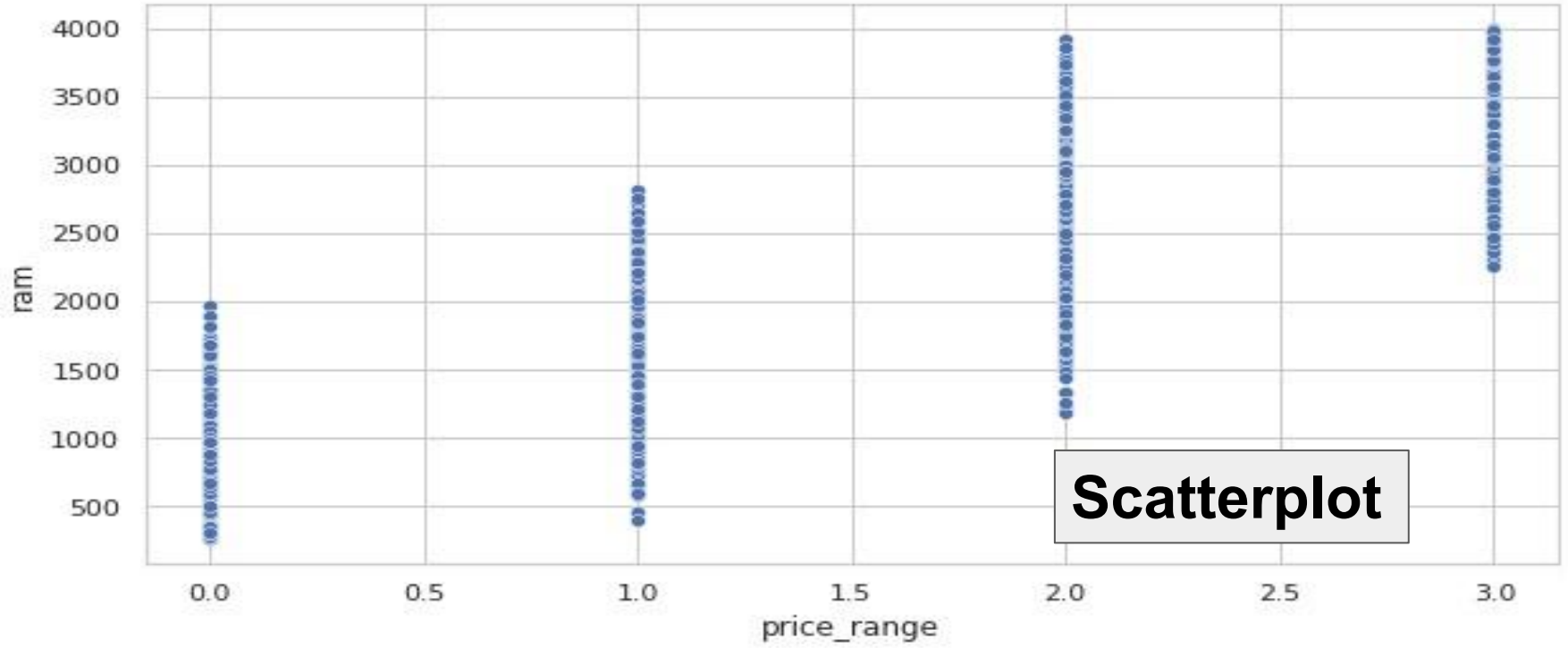


EDA and Visualization



Correlation Matrix

EDA and Visualization



Test and Train Split

```
[ ] from sklearn.preprocessing import StandardScaler
    scale = StandardScaler()

    X_train = scale.fit_transform(X_train)
    X_test = scale.fit_transform(X_test)
```

Applying Machine Learning Models and Hypertuning

The Models we will use-

1. Logistic Regression.
2. Random Forest
3. XGBoost
4. KNN
5. SVM

We will Hypertune our Models by using GridSearchCV.

Performance metrics and Classification Report

```
print(classification_report(y_train,knn_y_train_pred))
print("\n")
print(classification_report(y_test,knn_y_test_pred))
```

```
              precision    recall  f1-score   support

           0       0.76      0.86      0.81       361
           1       0.55      0.47      0.51       346
           2       0.55      0.64      0.59       360
           3       0.81      0.65      0.72       333

    accuracy                           0.66      1400
   macro avg       0.67      0.66      0.66      1400
weighted avg       0.66      0.66      0.66      1400
```

KNN

```
              precision    recall  f1-score   support

           0       0.69      0.94      0.79       139
           1       0.59      0.45      0.51       154
           2       0.53      0.66      0.58       140
           3       0.87      0.61      0.72       167

    accuracy                           0.66       600
   macro avg       0.67      0.66      0.65       600
weighted avg       0.68      0.66      0.65       600
```

Comparison

1. **AUC-ROC Score For Logistic Regression -** 0.9959478127910372
2. **AUC-ROC Score For Random Forest -** 0.9599435350025458
3. **AUC-ROC Score For XGBoost -** 0.9869522570860334
4. **AUC-ROC Score For KNN -** 0.8500126005892444
5. **AUC-ROC Score For SVM -** 0.9953712718512516

Conclusion

Except K Nearest Neighbour all the other models performing so well with respect to AUC_ROC Score and Classification Report . So if we want to do some prediction with this dataset then we can use these Models.

This work can be concluded with the comparable results of both Feature selection algorithms and classifier because we selected some relevant and important features to do the calculation.

Outcome

- Cost prediction is the very important factor of marketing and business. To predict the cost same procedure can be performed for all types of products for example Cars, Foods, Medicine, Laptops etc.
- Best marketing strategy is to find optimal product (with minimum cost and maximum specifications). So products can be compared in terms of their specifications, cost, manufacturing company etc.
- By specifying economic range a good product can be suggested to a customer.

Challenges

Finding the best K value in K Nearest Neighbour.

Elimination of features.

Finding best parameters for the model.

References

https://www.analyticsvidhya.com/

https://towardsdatascience.com/

https://stackoverflow.com/

Q & A