

Source Code on App Rating Prediction

The source code starts by importing the required libraries such as pandas, numpy, matplotlib, seaborn, and scikit-learn. These libraries are used for data handling, visualization, and building the machine learning model.

First, the dataset **googleplaystore.csv** is loaded using pandas. The shape of the dataset is printed to understand the initial size. Then, null values are checked and removed to ensure clean data for analysis.

Next, data cleaning is performed. The Size column is converted into numeric format by extracting values and converting MB into KB. The Reviews column is converted from string to numeric. In the Installs column, commas and plus signs are removed before converting the values into integers. The dollar symbol is removed from the Price column, and it is converted into numeric format.

Sanity checks are applied by keeping ratings between 1 and 5, removing rows where reviews exceed installs, and filtering free apps that have a price greater than zero.

Univariate analysis is done using boxplots and histograms. After identifying extreme values, outlier treatment is performed by removing apps with price above \$200, reviews above 2 million, and installs above the 99th percentile.

For preprocessing, log transformation is applied to Reviews and Installs. Unnecessary columns are dropped, and categorical variables are converted using dummy encoding.

Finally, the dataset is split into training and testing sets. A Linear Regression model is trained, and R² scores are calculated for both training and testing data.

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns
```

```
from sklearn.model_selection import train_test_split  
  
from sklearn.linear_model import LinearRegression  
  
from sklearn.metrics import r2_score
```

```
All features are numeric. Training model...
```

```
----- MODEL PERFORMANCE -----
```

```
Train R2 Score: 0.16126843090860743
```

```
Test R2 Score : 0.11681895751120863
```

```
[Done] exited with code=0 in 1.459 seconds
```

```
[Running] python -u
```

```
"/Users/rsh/Documents/AppRatingPrediction/app_rating_prediction.py"
```

```
Initial Shape: (10841, 13)
```

```
Null Values:
```

```
App 0
```

```
Category 0
```

```
Rating 1474
```

```
Reviews 0
```

```
Size 0
```

```
Installs          0
Type             1
Price            0
Content Rating   1
Genres           0
Last Updated     0
Current Ver      8
Android Ver      3
dtype: int64
After Dropping Nulls: (9360, 13)
After Cleaning: (7723, 13)
After Sanity Checks: (7717, 13)
```