

Ratings Prediction

Submitted by:

S. Praveena

**INTRODUCTION**

I used pandas library to analyse the data.

* info(): It informs about data columns and data types.
* head(): It returns the first five data.
* tail(): It returns the last five data.
* columns: It returns data columns
* shape: It gives number of rows and columns in a table.

We can combine tables to make it easier to see data. For this, we are using ‘concat function’ that is found in pandas library.

**DATA CLEANING**

Dataset can contain missing data, numerical string value, various cues. We can make NaN category value. At that time we can get a clean data. I used shift() method that is found in pandas library. Data type of Rating is object. If we convert from string to numeric, we can make easy. Data type of Size is objet. I had to convert the column because it contains the application sizes. Finally I changed ‘Varies with device’ value with NaN. After, I dropped ‘M’ and ‘K’. I changed from ‘1000+’ to 1000. Finally, I converted float value.

EXPLORATORY DATA ANALYSIS

After, I prepared to analyse our data, somewhat let’s explore the data.

* Corr(): It returns correlation.
* Describe(): It returns number of entries, average of entries, outliers values, standard deviation, minimum and maximum entry.

CONCLUSION

This is the end of the project. I hope it benefits. You can visualize with a lot of different model. Actually, I thought that EDA study could be boring. But its important for ML models.