**Numerical dataset :91\_mobiles\_uncleaned**

**Description**

This dataset contains information on over 13,000 different mobile phones scraped from 91 Mobiles, the largest gadget discovery site in India.

However, the dataset is uncleaned and contains a number of data quality issues, particularly in the last two columns (Per\_Dis\_Cam\_Bat and Other\_Features).

After cleaning and tidying the data, it can be used for exploratory data analysis (EDA), market analysis, and other research purposes related to the Indian mobile phone market.

The dataset includes the following columns:

Name: The name of the phone, including the brand and model.

Spec Score: A rating given by 91 Mobiles based on various specifications and features.missing[6729]

Operating System: The operating system the phone comes with out of the box. Features missing [4161]

Price: The cost of the phone in Indian rupees.

Per\_Dis\_Cam\_Bat: This column includes data on the phone's processor, display, battery, and camera.

Other Features: Additional details like the phone's storage capacity, SIM card type, fingerprint sensor, etc.

**Algorithm on it**

1)Feature transformation on Spec Score, Price

2)Feature Extraction on Per\_Dis\_Cam\_Bat, Other Features

3)Remove Duplicates (175)

4)impute nans by mean for numeric columns and mode for categorial columns

5)Label encoding on categorical columns

6)split data to feature and target

7)split data to train and test

8)Standard scaling to data

9)train model on data 80% of dataset

10)test Model 20% of dataset

11)evaluation of model

**Comparison on knn and linear regression**

|  |  |
| --- | --- |
| KNN | Linear Regression |
| MAE: 4728.219535135762 | MAE: 8100.659182698539 |
| MSE: 326322087.3789371 | MSE: 399542315.91445374 |
| R2\_score: 0.47241581832306767 | R2\_score: 0.354036352610561 |
| RMSE: 18064.387268294962 | RMSE: 19988.554622944946 |