

Task 1 - Classification

Hi there!

It's taking longer than usual to solve the glitch. We will go ahead with the first classification task. Here's the link to your dataset: [Task 1](#)

The dataset talks about various features of many mobile phones. The description of all the features(column names) are quite intuitive. The rating feature provides a phone's like/dislike value.

Goal: Predict if a user will like the phone or dislike.

Since the rating feature is a numeric quantity you are free set your thresholds to make a binary prediction. Remember that in the end your model should produce a binary outcome depending on the threshold of probability.

Report requirements:

1. You will work on Jupyter Notebook setup on your local machine. As part of your submission process you will upload your work on Github and provide us a link of your repo.
2. A clear representation of your thought process via data analysis is a must. Although you know matplotlib but you're free to use any visualization library viz. Plotly, seaborn, bokeh etc.
3. Along the way you will go through an interesting data cleaning pipeline. Try writing readable and modular code for us to interpret and debug if required. Your interpretation of data will be tested here.
4. While writing explanations under various sections or subsections, remember to be succinct and logical. For instance: If you are cleaning/encoding a feature then your explanations should answer WHY, WHAT, HOW and all possible WH-questions of your code.
5. You are free to use any algorithm that will maximize a model's accuracy score without compromising the fit.
6. The report should also represent a comparative study of all the training algorithms along with their metrics as stated below.

Metrics required: A classification will consist of accuracy score, a labelled confusion matrix and feature correlation matrix.