BLUEPRINT

For predicting Billboard Hits using Spotify Data

* We have to create a target variable which contains two classes “Hit” and “Non hit”. The songs which are present in billboard dataset should be “Hit” i.e. 1 and those songs in Million dataset whose Spotify ID is not present in Billboard dataset should be “Non hit” i.e. 0.
* Handle Missing values
* Drop duplicate values
* Remove outliers (those points which are outside the whiskers in box plot are outliers)
* Convert categorical values into numerical by using encoding techniques.
* Perform EDA for each columns and decide which feature is most relevant for prediction.
* Standardize the overall dataset.
* Divide the dataset into Training and Test sets.
* As we mentioned above the target of the data contains only two class, so it is a classification problem. We simply try each and every classification model into the resultant dataset and will find the best accuracy score. To simply this process we will use Lazy Predict Classifier which will help us to predict the accuracy scores of all classification models at a time.
* The model which will give us the highest accuracy in the first trial will be the desired model.
* After that we have to push the accuracy score of that model into to top notch by the help of hyperparameter tuning.
* Now, our model is ready to predict and also ready to deploy.