WELD-RIGHT

# PROJECT AIM : Our project's aim is to do advanced analytics on welder’s details provided in the dataset and to build a model to predict defects before it occurs as per the input parameters.

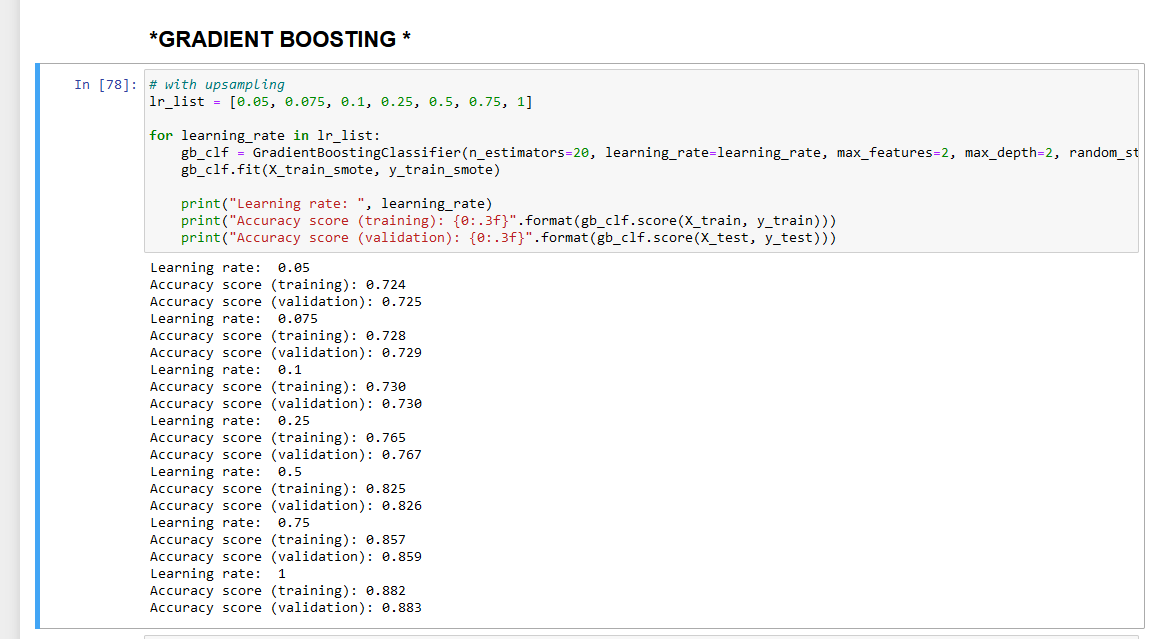
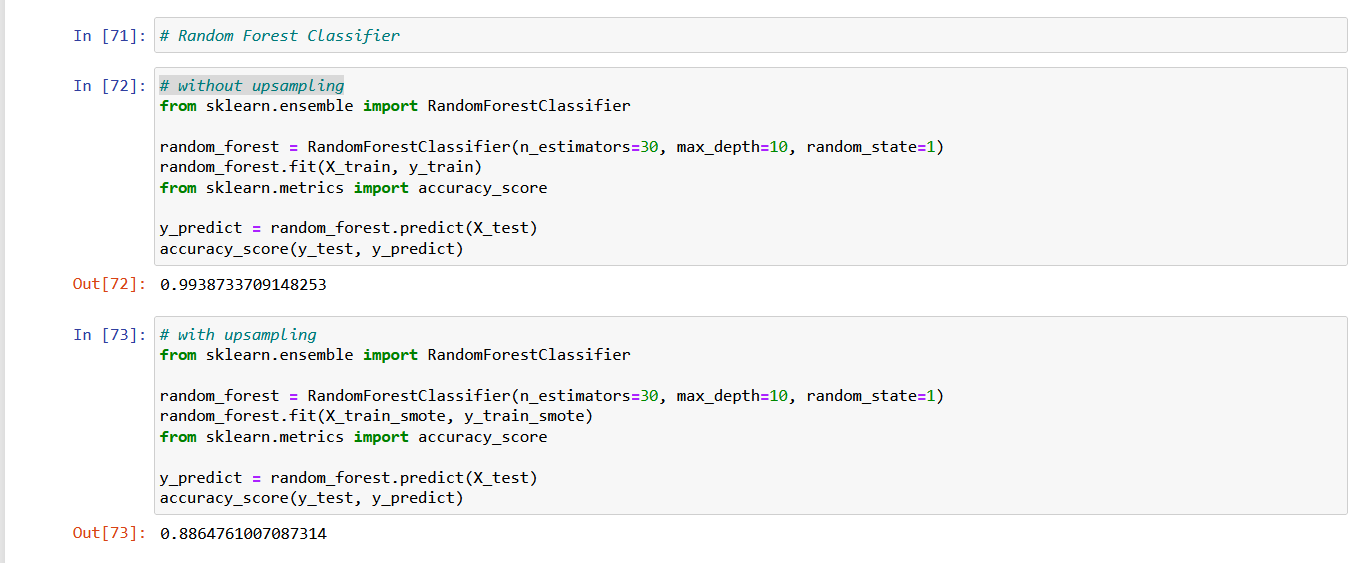
# Brief Walkthrough of work done till now :

## Cleaning and Preprocessing the data using necessary techniques .

* Getting insight from data by doing necessary EDA .
* Methods to determine welder’s performances:
  + Visualisation Techniques – plots like histplot , countplot ,box-plots ,sub-plots ,pair-plots, frequency distribution etc are used .
  + Statistical Techniques : pearson correlation , kendall correlation , p-values and heatmap for statistical visualisation like redundancy , confusion matrix .
* SWEET SPOT ANALYSIS :
  + Feature visualisation using various methods like U-Map and t-SNE for insight into how whole data is distributed over features .
  + We have also thought of an idea , sorting and dividing each feature column into 10-20 equal subsets and doing separate analysis to find correlations and zooming/analysing over subsets having higher proportion of defect and non defect in them and going iteratively to find the sweet spot of that feature .
* Using different technique to prepare data for training
  + Splitting data for training and testing .
  + Feature scaling (Standardisation like standard , quantile , robust scaler)the data to enable machine learning models to converge faster and also fit to distance based algorithms like svms , knns .
  + Dealing with an imbalanced dataset to prevent overfitting using techniques like gaussian noise upsampling , SMOTE .
* Training dataset and evaluating accuracy :
  + Fitting both original data and SMOTE generated data on various ML models like Decision Trees , Random Forest , Logistic Regression ,XGBoost , Gradient Boosting and SVMs also using cross validation techniques.
  + Obtaining Classification report from both data to get insight of overfitting .

## Overfitting can be seen in case of without upsampled data in below fig.

CLASSIFICATION REPORT OF FEW ML MODELS .



SOME ANALYSIS EXAMPLES

