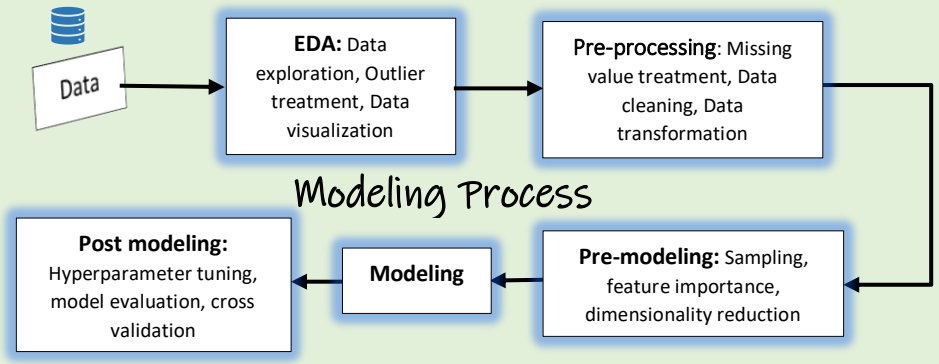
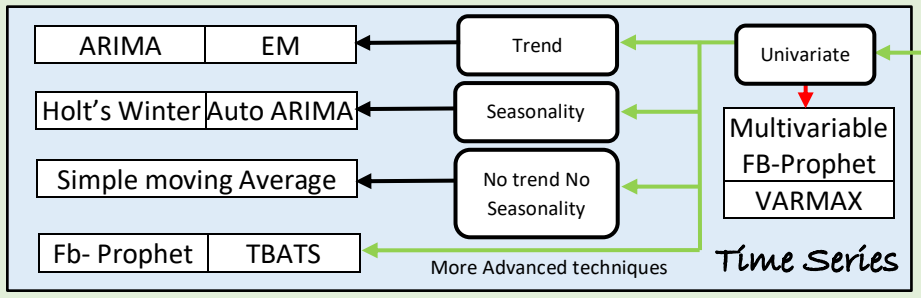
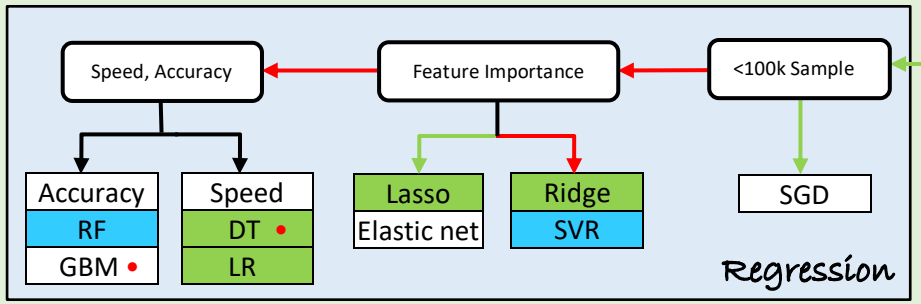
Algorithm Selection



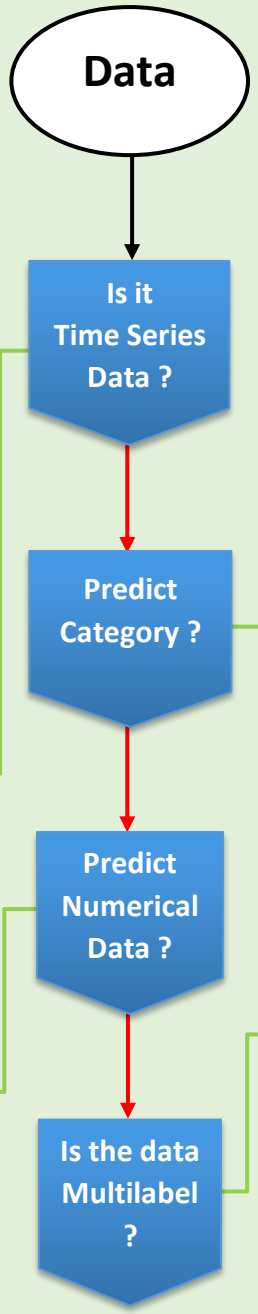
- ✓ Algorithm selection can be done as per modelling requirement from the flow chart (only for labelled data)
- ✓ For the common algorithms their characteristics mention under one type of problem will hold good for other problems as well
- ✓ If you are dealing with Text Classification, please refer to multiclass and binary classification as the algorithms used are same
- ✓ If (no. of features > no. of samples) please get more data before proceeding

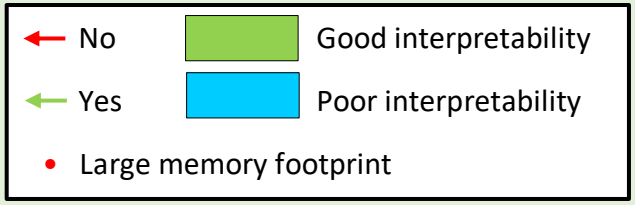
*For data too large, the limit beyond which it can be said that the data at hand can be considered large or not completely depends on the problem at hand, type of dataset and other such specific factors. However, to give an example, a dataset containing 100k datapoints can be considered large.

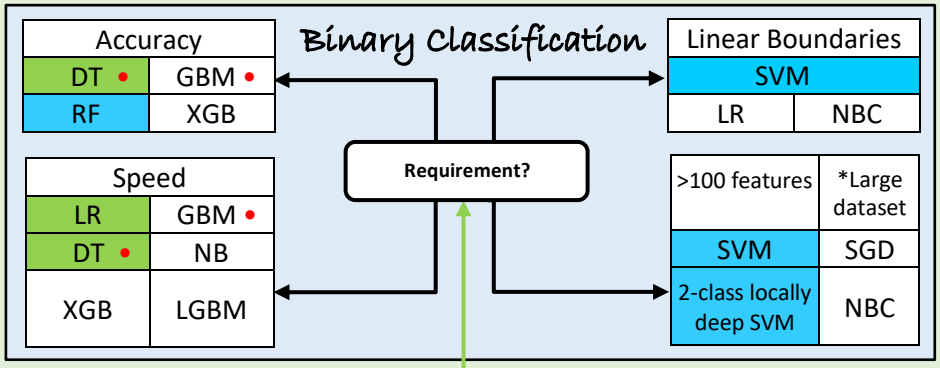


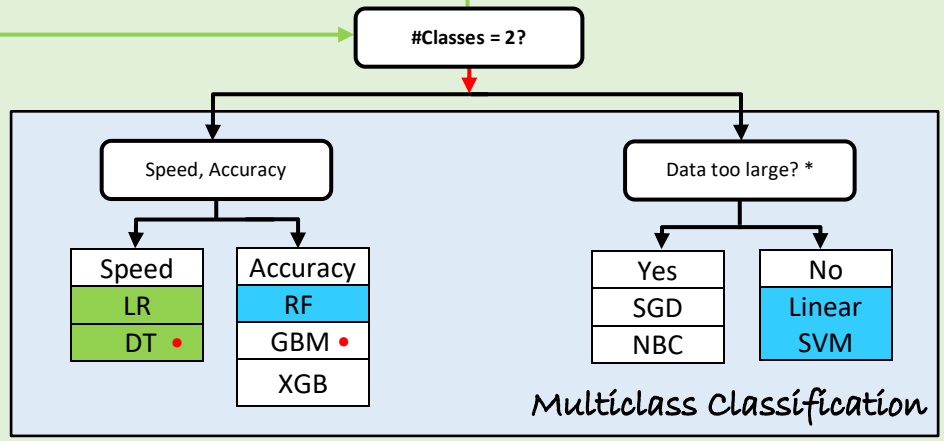


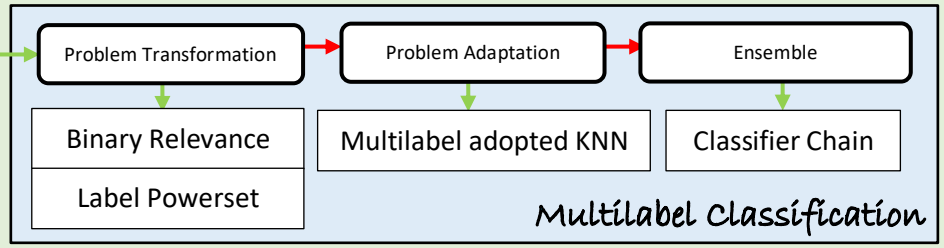
References: 1. https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/ 2. https://scikit-learn.org/stable/tutorial/machine_learning_map/ 3. https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet











LR- Linear/Logistic Regression**, SVM-** Support vector, **DT-** Decision tree, **RF-** Random Forest, **GBM-** Gradient boosting machine, **SGD-** Stochastic gradient descent, **XGB-** Extreme Gradient Boosting, **LGBM-** Light GBM, **NB-** Naïve Bayes , **EM-**Exponential Smoothening