**WHY IS RANDOM FOREST AND XGBOOST USED AND NOT THER METHODS LIKE LINEAR REGRESSION,SVM ETC.?**

1)RANDOM FOREST AND XGBOOST BOTH ARE ENSEMBLE LEARNING MODELS WHERE MULTIPLE DECISION TREES ARE USED TO FORM A MORE ACCURATE AND EFFICIENT PREDICTIVE MODEL.

2)IN RANDM FOREST, THE TREES ARE BUILT INDEPENDENTLY AND ALL THEIR RESULTS ARE AVERAGED OUT SO THAT THE VARIANCE IS MINIMIZED.

3)IN XG BOOST,THE TREES ARE FORMED SEQUENTIALL SO THAT THE NEW TREES CAN CORRECT THE ERRORS VREATED BY PREVIOUS TREES.

4)IT IS VERY EFFICIENT IN REPLACING MISSING VALUES UN THE DATASET HENCE MAKING IT EASIER DURING PREPROCESSING.

5)THEY DO NOT REQUIRE SCALING OR NORMALIZATION LIKE SVM OR K-NEAREST NEIGHBOURS

**HOW DOES RANDOM FOREST WORK?**

**1)**FROM THE RAW DATASET, MANY RANDOM SAMPLES ARE MADE TO FORM SUBSETS FOR EACH TREE.

2)A DECISION TREE IS TRAINED ON EACH SAMPLE OF THE DATASET.

3)AT EACH SUBSET OF THE TREE, WE FIND RANDOM SET OF FEATURES IMPROVING DIVERSITY.

4)FINALLY, AVERAGE OF ALL THE TREE OUTPUTS ARE TAKEN TO FIND THE PREDICTED VALUE.

**HOW DOES XGBOOST WORK?**

**1)**HERE, WE START WITH AN INITIAL PREDICTION

2)THE AVERAGE OF ALL THE TARGETS IS TAKE AS THE INITIALLY PREDICTED VALUE.

3)CALCULATE THE DIFFERENCE BETWEEN EACH PREDICTED VALUE AND THE ACTUAL TARGET FOR THAT FEATURE.

4)A NEW TREE WILL BE TRAINED EACH ONE FOR EVERY ERRORS FOUND.

5)THE NEW TREE’S PREDICTION IS SCALED DOWN USING A PARAMETER CALLED LEARNING RATE WHICH ENSURES THAT THE MODEL TRAINS GRADUALLY AND IS NOT OVERSTUFFRD.

6)THE UPCOMING TREES LEARN AND CORRECT THE PREVIOUS ERRORS AND THE FINAL PREDICTION IS THE SUM OF ALL THE SCALED OUTPUTS.

OTHER MODELS LIKE SVM AND LINEAR REGRESSION CAN BE MOSTLY USED FOR LINEARLY RELATED DATA UNLIKE RANDOM FOREST AND XGBOOST WHICH HAS THE ABILITY TO WORK WITH NON LINEAR RELATIONSHIPS.