RANDOM FOREST

CANCER DISEASE PREDICTION

MACHINE LEARNING

STEPS

1. PROBLEM ASSESMENT
2. DATA-PREPROCESSING
3. FEATURE SELECTION
4. MODEL TRAINING
5. MODEL TESTING
6. MODEL PERFORMANCE VISUALISATION
7. MODEL PERFORMANCE ENHANCEMENT

MODEL TRAINING AND TESTING STATS

TRAINING SATS

	precision	recall	f1-score	support
0	0.96	0.98	0.97	248
1	0.98	0.96	0.97	251
avg / total	0.97	0.97	0.97	499

ROC AUC 96.99%

		PREDICTED	
		0	1
ACTUAL	0	243	5
	1	10	241

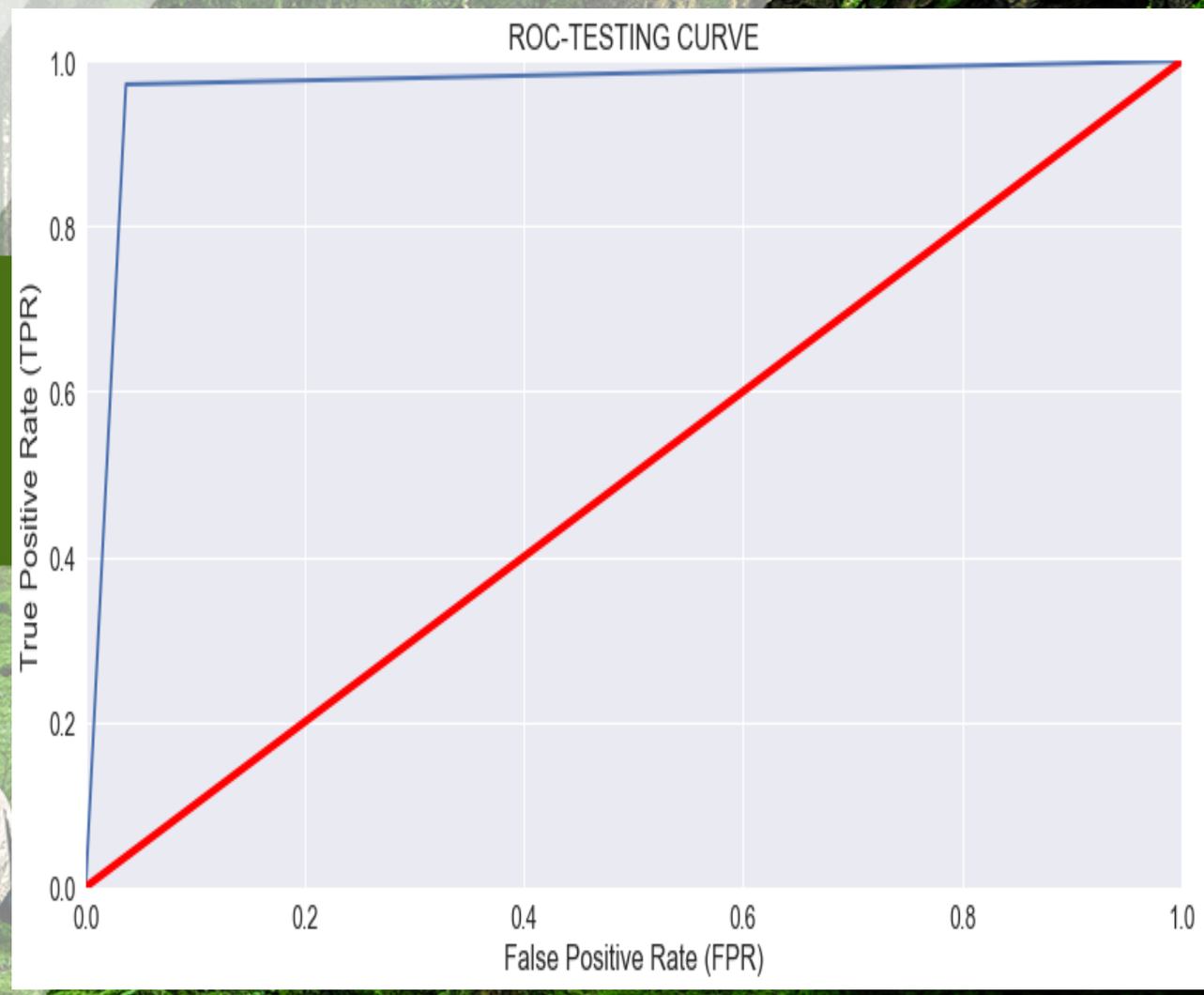
TESTING STATS

	precision	recall	f1-score	support
0	0.97	0.96	0.97	109
1	0.96	0.97	0.97	106
avg / total	0.97	0.97	0.97	215

ROC AUC 96.75%

		PREDICTED	
		0	1
ACTUAL	0	105	4
	1	3	103

ROC CURVE COMPARISON



OBSERVATION

Original dataset had 30 independent features, but for creation of model without degrading model performance ,7 features were found to be sufficient which could predict the event with same accuracy.
Features with importance is mentioned here.

FEATURE	IMPORTANCE
radius_mean	46.20%
texture_mean	18.50%
smoothness_mean	14.70%
fractal_dimension_mean	8.90%
texture_se	4.70%
smoothness_se	4.00%
symmetry_se	3.00%
total_importance	100.00%

FINAL VERDICT

Based upon comparison of training and testing evaluation parameter ,it can be said that model so formed has low bias and low variance.