

Topic: AdaBoost- Extreme Gradient Boosting

Perform AdaBoost and Extreme Gradient Boosting for the following dataset

1.) Diabetes_RF.csv

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
1	6	148	72	35	0	33.6	0.627	50	1
2	1	85	66	29	0	26.6	0.351	31	0
3	8	183	64	0	0	23.3	0.672	32	1
4	1	89	66	23	94	28.1	0.167	21	0
5	0	137	40	35	168	43.1	2.288	33	1
6	5	116	74	0	0	25.6	0.201	30	0
7	3	78	50	32	88	31.0	0.248	26	1
8	10	115	0	0	0	35.3	0.134	29	0
9	2	197	70	45	543	30.5	0.158	53	1
10	8	125	96	0	0	0.0	0.232	54	1
11	4	110	92	0	0	37.6	0.191	30	0
12	10	168	74	0	0	38.0	0.537	34	1
13	10	139	80	0	0	27.1	1.441	57	0
14	1	189	60	23	846	30.1	0.398	59	1
15	5	166	72	19	175	25.8	0.587	51	1
16	7	100	0	0	0	30.0	0.484	32	1
17	0	110	84	47	220	45.0	0.551	31	1

2.) wbcd.csv

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity
1	87139402	B	12.320	12.39	78.85	464.1	0.10280	0.06981	0.039870
2	8910251	B	10.600	18.95	69.28	346.4	0.09688	0.11470	0.063870
3	905520	B	11.040	16.83	70.92	373.2	0.10770	0.07804	0.030460
4	868871	B	11.280	13.39	73.00	384.8	0.11640	0.11360	0.046350
5	9012568	B	15.190	13.21	97.65	711.8	0.07963	0.06934	0.033930
6	906539	B	11.570	19.04	74.20	409.7	0.08546	0.07722	0.054850
7	925291	B	11.510	23.93	74.52	403.5	0.09261	0.10210	0.111200
8	87880	M	13.810	23.75	91.56	597.8	0.13230	0.17680	0.155800
9	862989	B	10.490	19.29	67.41	336.1	0.09989	0.08578	0.029950
10	89827	B	11.060	14.96	71.49	373.9	0.10330	0.09097	0.053970
11	91485	M	20.590	21.24	137.80	1320.0	0.10850	0.16440	0.218800
12	8711003	B	12.250	17.94	78.27	460.3	0.08654	0.06679	0.038850
13	9113455	B	13.140	20.74	85.98	536.9	0.08675	0.10890	0.108500
14	857810	B	13.050	19.31	82.61	527.2	0.08060	0.03789	0.000692
15	9111805	M	19.590	25.00	127.70	1191.0	0.10320	0.09871	0.165500
16	925277	B	14.590	22.68	96.39	657.1	0.08473	0.13300	0.102900

Hints:

1. Business Problem
 - 1.1. Objective
 - 1.2. Constraints (if any)
2. Data Pre-processing
 - 2.1 Data cleaning, Feature Engineering, EDA etc.
3. Model Building
 - 3.1 Partition the dataset
 - 3.2 Model(s) - Reasons to choose any algorithm
 - 3.3 Model(s) Improvement steps
 - 3.4 Model Evaluation
 - 3.5 Python and R codes
4. Deployment
 - 4.1 Deploy solutions using R shiny and Python Flask.
5. Result Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.

Note:

1. For each assignment the solution should be submitted in the format
2. Research and Perform all possible steps for improving the model(s) accuracy
Ex: Feature Engineering, Hyper Parameter tuning, etc.
3. All the codes (executable programs) are running without errors
4. Documentation of the module should be submitted along with R & Python codes, elaborating on every step mentioned here