



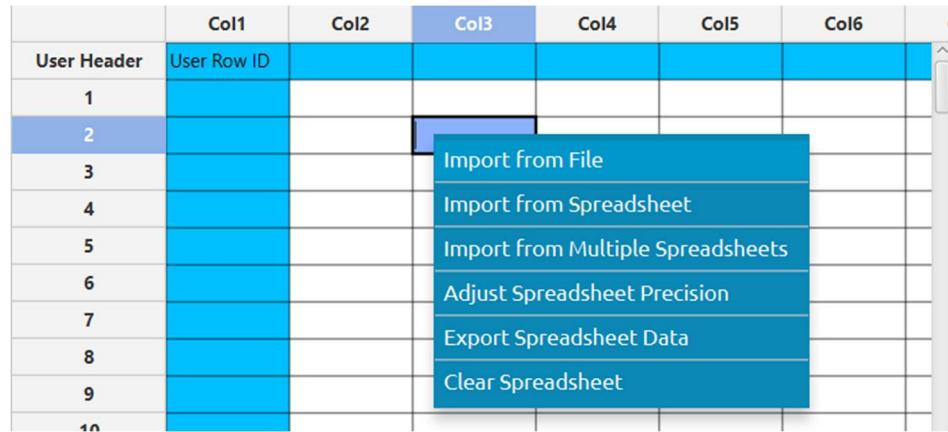
Breast Cancer Wisconsin (Diagnostic) Data set

This dataset contains, which can be found in <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data> 31 features and 569 samples and it is used to predict whether a breast cancer diagnosis is malignant (M) or benign (B). The features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass and they describe characteristics of the cell nuclei present in the image.

Isalos version used: 2.0.6

Step 1: Import data from file

Right click on the input spreadsheet (left) and choose the option “Import from File”. Then navigate through your files to load the one with the breast cancer data.

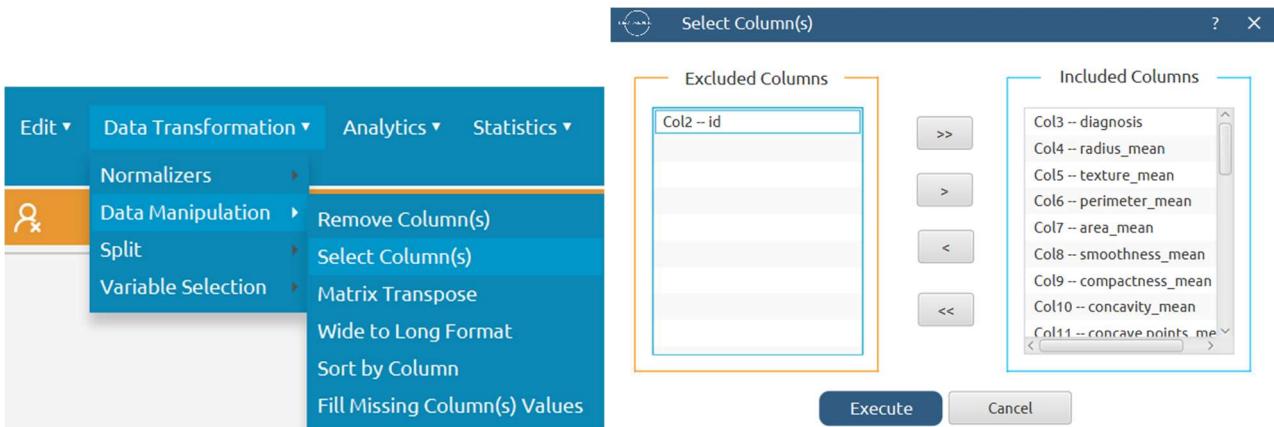


The data will appear on the left spreadsheet.

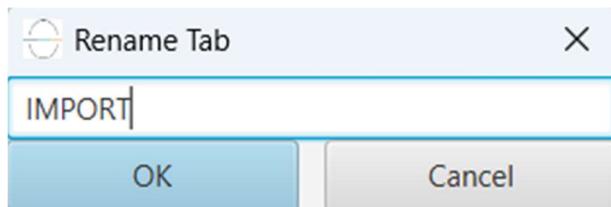
User Header	Col1	Col2 (B)	Col3 (S)	Col4 (D)	Col5 (E)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (D)	Col14 (D)	Col15 (D)	Col16 (D)	Col17 (D)	Col18 (D)	Col19 (D)	Col20 (D)	Col21 (D)	Col22 (D)	Col23 (D)	Col24 (D)	Col25 (D)	Col26 (D)	Col27 (D)	Col28 (D)	Col29 (D)	Col30		
User Row ID	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean	fractal_dimension_mean	radius_se	texture_se	perimeter_se	area_se	smoothness_se	compactness_se	concavity_se	concave points_se	symmetry_se	fractal_dimension_se	radius_worst	texture_worst	perimeter_worst	area_worst	smoothness_worst	compactness_worst	concavity_worst	concave points_worst	symmetry_worst	fractal_dimension_worst
1	642302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.3419	0.07871	1.002	0.9533	8.589	153.4	0.05699	0.04924	0.05773	0.01587	0.00033	0.005193	25.38	17.33	184.5	2019	0.1622	0.6656	0.7119			
2	642303	M	20.57	17.77	188.5	1005	0.1095	0.2926	0.3047	0.1382	0.3455	0.07871	1.002	0.9533	8.589	153.4	0.05699	0.04924	0.05773	0.01587	0.00033	0.005193	25.38	17.33	184.5	2019	0.1622	0.6656	0.7119			
3	6423003	M	18.49	21.25	150	1003	0.1096	0.1599	0.1974	0.1579	0.2049	0.05999	0.7456	0.7469	0.9391	0.0423	0.00415	0.04056	0.03332	0.02063	0.0232	0.004771	23.57	24.53	152.3	1709	0.1444	0.4149	0.4554			
4	64343101	M	11.42	20.38	77.58	1061	0.1425	0.2339	0.2414	0.1052	0.2897	0.09744	0.4956	1.156	3.445	27.23	0.00911	0.07468	0.05691	0.01867	0.000208	14.91	26.5	98.87	567.7	0.0663	0.4669	0.4669				
5	64383420	M	20.29	14.34	135.1	1297	0.1005	0.1238	0.1048	0.1059	0.09883	0.7572	0.713	5.439	94.44	0.01149	0.03451	0.05688	0.01883	0.01756	0.005115	22.54	16.67	152.2	1975	0.1374	0.2369	0.4				
6	642708	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.0809	0.2087	0.07813	0.3348	0.896	2.217	27.19	0.00751	0.03344	0.08572	0.01137	0.02165	0.05050	15.47	23.75	103.4	741.6	0.1791	0.5249	0.5355			
7	644319	M	18.25	19.98	119.6	1024	0.09483	0.109	0.1127	0.074	0.1794	0.05742	0.4487	0.7714	3.18	53.91	0.004314	0.01834	0.02354	0.01030	0.01369	0.002179	22.88	27.66	133.2	160.1	0.1442	0.3276	0.3784			
8	64458202	M	19.71	20.83	90.2	577.9	0.1189	0.1645	0.09596	0.09885	0.2196	0.07451	0.5848	1.377	3.854	50.96	0.00805	0.03029	0.02488	0.01448	0.01485	0.05012	17.06	28.14	110.6	897	0.1654	0.3882	0.2878			
9	644981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09693	0.235	0.07389	0.3046	0.1002	1.002	2.406	24.23	0.005731	0.03502	0.03933	0.01232	0.02143	0.00749	18.49	30.73	106.2	739.3	0.1709	0.5401	0.539		
10	64450091	M	12.48	24.34	89.5	475.9	0.1188	0.2239	0.2273	0.08647	0.285	0.08387	0.32779	0.0939	0.0939	0.0423	0.007148	0.01743	0.01423	0.01769	0.01543	0.01423	20.73	40.68	97.65	711.4	0.1833	1.058	1.165			
11	645124	M	16.92	22.34	103.7	1026	0.09449	0.09333	0.09333	0.09333	0.09337	0.37749	1.157	0.446	0.421	0.008267	0.019191	0.019191	0.019191	0.019191	0.019191	0.002045	16.3	22.13	110.7	241.1	0.1474	0.4457	0.4457			
12	6461002	M	19.79	17.89	103.6	781	0.0971	0.1329	0.0954	0.09605	0.13842	0.09082	0.5058	0.9849	3.554	54.16	0.007771	0.04081	0.01382	0.02008	0.004144	20.42	27.29	136.5	1299	0.1394	0.5659	0.3945				
13	6445238	M	19.17	24.8	132.4	1123	0.0974	0.2458	0.2095	0.1118	0.2397	0.0718	0.9555	3.568	11.07	116.2	0.003139	0.06237	0.0889	0.0409	0.04484	0.01284	20.95	29.94	151.7	1332	0.1037	0.3903	0.3639			
14	6463831	M	19.85	23.95	103.7	782.7	0.09401	0.1002	0.09388	0.09384	0.1847	0.05338	0.4033	1.078	2.903	36.55	0.000769	0.03156	0.05051	0.01962	0.02081	0.030302	16.84	27.66	112	876.5	0.1311	0.7124	0.2322			
15	64657401	M	17.73	22.61	93.6	578.3	0.1131	0.2239	0.2128	0.08025	0.2099	0.07882	0.2131	1.169	2.061	19.21	0.006429	0.06938	0.05501	0.01628	0.008039	15.03	32.01	108.8	697.7	0.1651	0.7725	0.6943				

Step 2: Manipulate data

In our dataset there are not any empty values, so we can select all the columns to be used. However, since the column “id” does not offer any significant information about the breast cancer diagnosis we will exclude it. On the menu click on *Data Transformation → Data Manipulation → Select Column(s)* and select all columns except “id”.



All of the data will appear in the output (right) spreadsheet. This tab can be renamed “IMPORT” by right-clicking on it and choosing the “Rename” option.



Step 3: Split data

Create a new tab by pressing the “+” button on the bottom of the page with the name “TRAIN_TEST_SPLIT” which we will use for splitting the train and test set.

Import data into the input spreadsheet of the “TRAIN_TEST_SPLIT” tab from the output of the “IMPORT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

	Col1	Col2	Col3	Col4	Col5	Col6
User Header	User Row ID					
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

Split the dataset by choosing Data Transformation → Split → Random Partitioning. Then choose the “Training set percentage” and the column for the sampling as shown below:



The results will be two separate spreadsheets, “TRAIN_TEST_SPLIT: Training Set” and “TRAIN_TEST_SPLIT: Test Set”, which will be available to import into the next tabs.

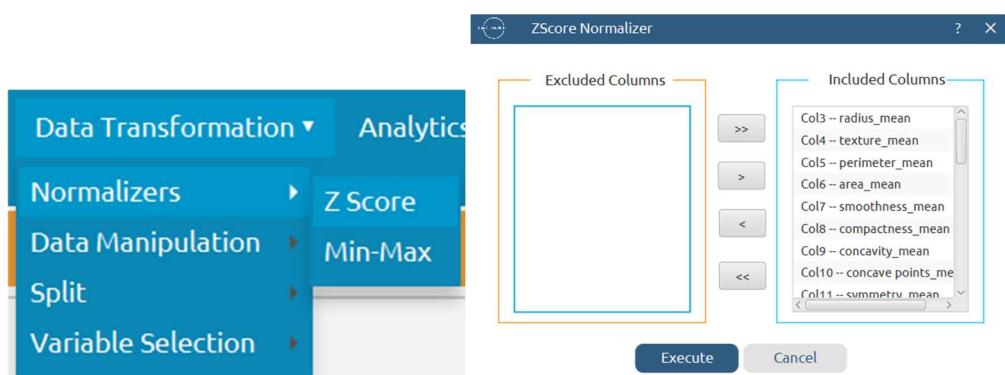
Step 4: Normalize the training set

Create a new tab by pressing the “+” button on the bottom of the page with the name “NORMALIZE_TRAIN_SET”.

Import into the input spreadsheet of the “NORMALIZE_TRAIN_SET” tab the train set from the output of the “TRAIN_TEST_SPLIT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”. From the available Select input tab options choose “TRAIN_TEST_SPLIT: Training Set”.

User Header	User Row ID	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (D)	Col14 (D)	Col15 (D)	Col16 (D)	Col17 (D)	Col18 (D)	Col19 (D)	Col20 (D)	Col21 (D)	Col22 (D)	Col23 (D)	Col24 (D)	Col25 (D)	Col26 (D)	Col27 (D)	Col28 (D)	Col29 (D)	Col30
1	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07371	1.005	0.9033	0.8369	151.4	0.006399	0.0404	0.05773	0.01487	0.000303	0.006193	21.33	17.33	184.6	2019	0.1622	0.6556	0.7119	0.2054		
2	M	19.89	21.25	130	1208	0.1096	0.1599	0.1074	0.1279	0.2069	0.05999	0.7456	0.7869	0.4815	94.03	0.00615	0.04206	0.03832	0.02058	0.004771	23.57	25.53	182.5	1709	0.1444	0.2445	0.4804	0.243			
3	M	11.42	20.38	77.58	386.1	0.1425	0.2339	0.2414	0.1052	0.2997	0.09744	1.156	1.3465	2.723	0.0911	0.07488	0.05681	0.01887	0.05963	0.002028	14.91	26.5	98.07	587.7	0.2098	0.0885	0.689	0.2675			
4	M	20.29	14.34	181.1	1297	0.1003	0.1328	0.1016	0.1043	0.1809	0.05883	0.7572	0.7813	5.428	44.44	0.0149	0.02481	0.05683	0.01885	0.01719	0.001115	22.34	16.67	181.2	157	0.1312	0.207	0.445	0.1625		
5	M	12.45	10.91	62.1	477	0.1171	0.1719	0.1179	0.1049	0.174	0.07179	0.7742	0.4487	0.7732	3.19	0.5371	0.02414	0.02352	0.01939	0.02179	23.88	27.66	193.2	1697	0.1442	0.2376	0.4385	0.1747			
6	M	18.29	19.99	119.6	1640	0.0845	0.109	0.1127	0.074	0.1794	0.07482	0.7745	0.5885	0.7395	1.377	0.5646	0.00805	0.03029	0.04488	0.01448	0.01458	0.02542	17.04	28.14	110.6	997	0.1554	0.3482	0.2679	0.1556	
7	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.0946	0.0985	0.2195	0.07451	0.5835	0.5885	0.7395	0.7395	0.00639	0.02359	0.02199	0.02179	23.88	27.66	193.2	1709	0.1444	0.2445	0.4804	0.243				
8	M	13	21.82	87.5	519.8	0.1273	0.1932	0.0939	0.0935	0.238	0.07399	0.5063	1.002	2.456	24.32	0.005731	0.03032	0.03533	0.02125	0.02143	0.03749	15.49	30.73	106.2	5401	0.1559	0.2679	0.1703	0.2675		
9	M	15.85	23.95	103.7	782.7	0.0841	0.1020	0.0938	0.0934	0.1547	0.05338	0.4233	1.076	2.923	36.53	0.00769	0.03197	0.05051	0.01992	0.02681	0.03039	16.84	27.66	110.6	875.5	0.1311	0.1924	0.2322	0.1119		
10	M	13.73	22.81	93.6	578.3	0.1131	0.2293	0.2128	0.08035	0.2059	0.07882	0.2121	1.169	2.061	19.21	0.006429	0.05649	0.05650	0.01628	0.01961	0.008993	15.03	32.01	108.8	697.7	0.1551	0.6843	0.2208	0.1119		
11	M	14.54	27.54	96.73	658.1	0.1139	0.1595	0.1639	0.07364	0.2399	0.07077	0.37	1.033	2.879	32.55	0.005607	0.0424	0.04741	0.0109	0.01837	0.005466	17.45	37.13	124.1	943.2	0.1678	0.6577	0.7026	0.1112		
12	M	14.68	20.13	94.74	684.5	0.09887	0.07075	0.07395	0.05259	0.1686	0.05923	0.4727	1.24	3.95	45.4	0.005718	0.01162	0.01909	0.01417	0.020365	19.07	30.88	123.4	115	0.1454	0.1571	0.2914	0.1659			
13	M	16.13	20.63	101.1	793.8	0.117	0.2032	0.1722	0.1029	0.2164	0.07158	0.5652	1.07	3.41	5.0703	0.02151	0.01189	0.00412	20.94	31.43	135.3	1315	0.1769	0.4233	0.4784	0.2073					
14	M	13.54	14.36	87.45	586.3	0.09779	0.08129	0.09664	0.04791	0.1685	0.09766	0.2699	0.7866	2.068	0.006482	0.0148	0.02387	0.01515	0.0198	0.0323	15.11	19.26	99.7	711.2	0.144	0.1773	0.259	0.1358			
15	M	15.08	16.71	55.63	525	0.1073	0.121	0.04688	0.0211	0.1907	0.0582	0.7477	1.383	14.67	0.004097	0.01698	0.00649	0.01678	0.003425	14.5	20.49	96.09	695.5	0.1312	0.2776	0.169	0.0738				

Normalize the data using Z-score: Data Transformation → Normalizers → Z Score and select all columns except the “diagnosis” target column.



The results will appear on the output spreadsheet.

User Header	User Row ID	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (D)	Col14 (D)	Col15 (D)	Col16 (D)	Col17 (D)	Col18 (D)	Col19 (D)	Col20 (D)	Col21 (D)	Col22 (D)	Col23 (D)	Col24 (D)	Col25 (D)	Col26 (D)	Col27 (D)	Col28 (D)	Col29 (D)	Col30						
1	M	1.0293984	0.0001558	1.2424927	0.0005449	1.5020493	3.07793	2.604002	2.5227065	2.2194160	2.0496660	2.2104160	2.0431043	-0.5748493	2.8323293	2.3831581	-0.2108640	1.2442638	0.6760713	1.1842766	0.0773378	1.3116275	-0.1714614	2.393638	2.0161784	1.3440113	2.0897118	2.158272	2.088712								
2	M	1.5228309	0.3946059	1.5137327	1.4935954	0.9620020	1.0551109	1.307081	2.0297705	0.9632077	-0.3006457	1.2105058	-0.7010103	0.8484489	1.1140304	-0.2919770	0.7972708	0.2138944	1.4469146	0.2358685	0.2837770	1.4885815	-0.0509662	1.338814	1.4550028	0.3556440	1.0263814	0.678970	1.9460247								
3	M	-1.6873312	0.214055	-0.9319165	-0.7995040	1.280923	3.427987	1.691284	1.4469761	2.0083781	4.8143176	0.3211997	-0.1174073	0.2861118	-0.2795904	0.8711440	2.7087783	0.2816118	1.1345023	1.3821911	0.105412	-0.2386083	0.1058412	-0.3571870	1.3621911	0.2937477	2.1900250	1.2030023	0.3537810	1.3621911	0.2937477						
4	M	-0.6887160	-1.1502088	1.7718149	1.7628003	0.342002	0.5431760	1.3639868	1.4213869	-0.0212273	-0.5419525	1.260339	-0.0203112	1.2718975	1.1223888	0.4591781	0.0505491	0.8402107	1.1630116	-0.3421285	0.4293179	1.2862958	-0.1777141	1.3296116	1.2197448	0.2420302	-0.3132368	0.6599149	0.7114439								
5	M	-0.4785537	-0.2457164	-0.2847164	-0.2847164	1.280923	3.427987	1.691284	1.4469761	2.0083781	4.8143176	0.3211997	-0.1174073	0.2861118	-0.2795904	0.8711440	2.7087783	0.2816118	1.1345023	1.3821911	0.105412	-0.2386083	0.1058412	-0.3571870	1.3621911	0.2937477	2.1900250	1.2030023	0.3537810	1.3621911	0.2937477						
6	M	1.0293984	0.0001558	1.2424927	0.0005449	1.5020493	3.07793	2.604002	2.5227065	2.2194160	2.0496660	2.2104160	2.0431043	-0.5748493	2.8323293	2.3831581	-0.2108640	1.2442638	0.6760713	1.1842766	0.0773378	1.3116275	-0.1714614	2.393638	2.0161784	1.3440113	2.0897018	0.678970	1.9460247								
7	M	-0.1020542	0.2029114	-0.0041162	-0.3019155	1.280923	3.427987	1.691284	1.4469761	2.0083781	4.8143176	0.3211997	-0.1174073	0.2861118	-0.2795904	0.8711440	2.7087783	0.2816118	1.1345023	1.3821911	0.105412	-0.2386083	0.1058412	-0.3571870	1.3621911	0.2937477	2.1900250	1.2030023	0.3537810	1.3621911	0.2937477						
8	M	0.3265371	0.2135981	-0.1893227	-0.3052963	2.007385	1.696665	1.734854	1.1473954	1.9953220	1.547582	-0.3316053	-0.3316053	-0.3316053	-0.3316053	-0.3316053	0.5131985	0.1273096	-0.0171381	0.2320027	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042	0.0950042
9	M	0.4619311	1.0001733	0.4741957	0.3349581	-0.07247477	0.1594854	0.1523273	0.1594854	0.1523273	0.1594854	0.1523273	0.1594854	-0.0444747	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847	0.2545847					
10	M	-0.1247271	0.7070551	0.0518920	-0.2295253	1.2707055	2.3075650	1.5019489	0.9620020	0.8540252	0.9529771	1.9539695	-0.0684845	-0.0684845	-0.0684845	-0.0684845	-0.0684845	0.0503297	0.2039374	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	0.1853873	
11	M	0.0991990	1.0304150	0.1775761	-0.0707965	1.2635242	1.0574969	0.9568227	0.6718702	1.1173705	1.8218702	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167	0.2961167							
12	M	0.1379025	0.1454704	0.0761675	0.0379798	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282	0.1962282							
13	M	0.3837777	0.2685597	0.0640792	-0.3039787	1.4212932	1.8711620	1.0610966	1.3853591	1.308894	1.5707155	0.3833719	-0.2678477	-0.2678477	-0.2678477	-0.2678477	-0.2678477	0.4745778	0.2827342	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	0.0371049	
14	M	-0.1772330	-1.0453377	-0.1948288	-0.3021401	0.17117508	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351	0.1628351						
15	M	-0.3934209	-0.8434790	-0.2681043	-0.3098718	0.3107055	0.4031967	0.3444844	0.3810755	0.3107055	0.4031967	0.3810755	0.3107055	0.4031967	0.3810755	0.3107055	0.4031967	0.3810755	0.3107055	0.4031967	0.3810755	0.3107055	0.4031967	0.3810755	0.3107055	0.4031967	0.3810755	0.3107055	0.4031967	0.3810755	0.3107055	0.4031967	0.3810755				

Step 5: Normalize the test set

Create a new tab by pressing the “+” button on the bottom of the page with the name “NORMALIZE_TEST_SET”.

Import into the input spreadsheet of the “NORMALIZE_TEST_SET” tab the test set from the output of the “TRAIN_TEST_SPLIT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”. From the available Select input tab options choose “TRAIN_TEST_SPLIT: Test Set”.

User Header	User Row ID	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (D)	Col14 (D)	Col15 (D)	Col16 (D)	Col17 (D)	Col18 (D)	Col19 (D)	Col20 (D)	Col21 (D)	Col22 (D)	Col23 (D)	Col24 (D)	Col25 (D)	Col26 (D)	Col27 (D)	Col28 (D)	Col29 (D)	Col30		
1	M	20.57	17.77	132.9	132	0.00474	0.07764	0.0089	0.07717	0.132	0.05667	0.5435	0.7339	3.936	74.03	0.005225	0.01028	0.01688	0.01314	0.01389	0.0030312	24.99	13.41	1958	0.1338	0.1366	0.2416	0.1816	0.2416	0.1816	0.2416	0.1816	0.2416
2	M	12.48	24.04	83.97	475.9	0.118	0.2396	0.2273	0.05843	0.2032	0.05669	0.20329	0.05669	0.20329	0.05669	0.20329	0.05669	0.20329	0.05669	0.20329	0.05669	0.20329	0.05669	0.20329	0.05669	0.20329	0.05669	0.20329	0.05669	0.20329			
3	M	16.02	23.24	102.7	102.7	79.78	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001		
4	M	19.78	17.89	103.6	781	0.0071	0.0192	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	
5	M	19.81	24.25	120	112.3	0.074	0.147	0.0119	0.02497	0.074	0.0555	2.005	11.7	1.07	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
6	M	19.81	21.25	130	100	0.0081	0.1027	0.1479	0.04049	0.182	0.00395	0.7382	1.017	5.865	11.2	0.00404	0.01889	0.0121	0.0121	0.0121	0.00197	0.00197	0.00197	0.00197	0.00197	0.00197	0.00197	0.00197	0.00197	0.00197	0.00197		
7	M	21.15	20.04	128.2	140.4	0.05442	0.1022	0.0319	0.1244	0.2183	0.06197	0.1807	1.4656	5.7474	105	0.002648	0.0374	0.01048	0.01048	0.01048	0.00207	0.00207	0.00207	0.00207	0.00207	0.00207	0.00207	0.00207	0.00207	0.00207	0.00207		
8	M	16.65	21.38	110	90.6	0.017	0.1537	0.125	0.0033	0.0917	0.0033	0.0																					

The results will appear on the output spreadsheet.

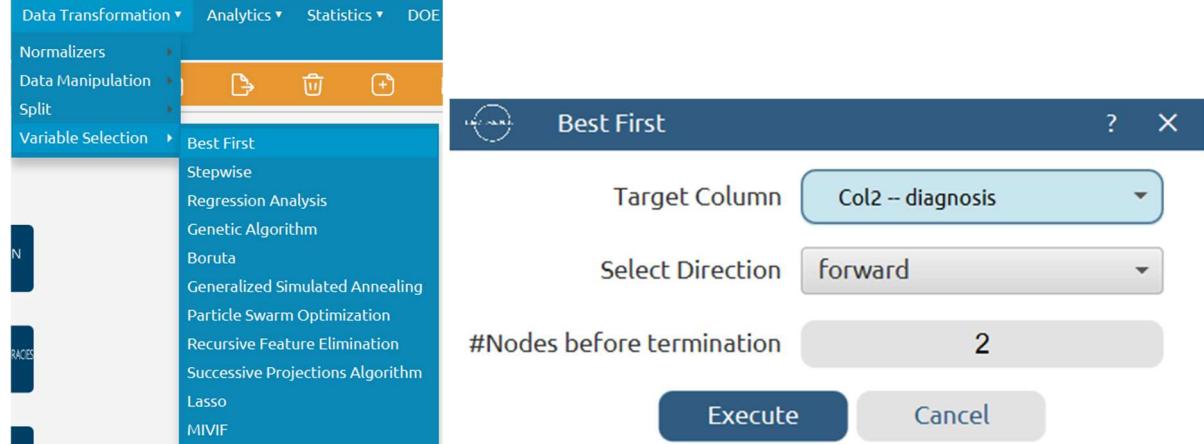
User Header	User Row ID	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (D)	Col14 (D)	Col15 (D)	Col16 (D)	Col17 (D)	Col18 (D)	Col19 (D)	Col20 (D)	Col21 (D)	Col22 (D)	Col23 (D)	Col24 (D)	Col25 (D)	Col26 (D)	Col27 (D)	Col28 (D)	Col29 (D)	Col30				
		diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	concave_points_mean	area_se	symmetry_se	perimeter_worst	area_worst	smoothness_worst	concavity_worst	concave_points_worst	radius_worst	texture_worst	perimeter_worst_sq	area_worst_sq	smoothness_worst_sq	concavity_worst_sq	concave_points_worst_sq	radius_worst_sq	texture_worst_sq	perimeter_worst_sq_sq	area_worst_sq_sq	smoothness_worst_sq_sq	concavity_worst_sq_sq	concave_points_worst_sq_sq	radius_worst_sq_sq	texture_worst_sq_sq	perimeter_worst_sq_sq_sq	area_worst_sq_sq_sq	smoothness_worst_sq_sq_sq	concavity_worst_sq_sq_sq	concave_points_worst_sq_sq_sq
1	M	-0.3823559	1.6291156	1.8227160	-0.7937751	-0.4833219	0.5476245	0.098842	-0.8411952	0.6901032	0.5914713	0.6917173	0.4813540	0.2733056	0.7932674	-2.095587	1.7912474	-2.3920958	1.5262403	1.9050522	-0.357319	-0.4311039	-0.1452237	1.9870561											
2	M	-0.4785211	0.0203103	-0.3347143	-0.3114258	1.5950545	2.5838330	1.7602048	0.8914690	0.814690	2.7312729	-0.3827376	0.6877841	-0.411691	-0.3402420	0.0954884	2.8753271	1.5333990	0.4323929	-0.3076838	2.3004143	2.3970700	-0.2497100	-0.2626889	-0.3097885	2.3481324	5.1499778	0.4034938	1.6169912						
3	M	0.008380	0.4191549	0.4172661	0.3742362	-0.9834180	-0.7110338	-0.7055107	-0.4076460	-0.7996568	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908	-0.089908			
4	M	0.4149995	-0.3357166	0.6340480	0.0784793	0.4011656	0.1389767	0.4612054	0.1250558	-0.2635053	0.3816692	-0.4292939	0.3484832	0.2833490	-0.4161920	0.8277059	-0.1372389	0.1786880	-0.0384602	0.1747468	0.4893377	0.2131080	0.8630240	0.731212	0.3427076	1.9619816	4.7682323	0.9597181	0.4519583	0.4693934	0.1915140	0.2181151			
5	M	1.3971717	1.9193559	1.6097383	1.2729861	0.0964608	2.7019762	1.4962606	1.8145222	1.2167554	2.1176554	1.9694249	2.2719038	1.4097929	1.7879177	1.2718184	1.921961	4.7682323	0.9597181	0.4519583	0.4693934	0.1915140	0.2181151	0.7309999	0.8735998	0.4304687	0.3427076	1.9619816	4.7682323	0.9597181	0.4519583	0.4693934	0.1915140	0.2181151	
6	M	1.8201112	0.8955561	0.1313774	0.8814244	0.0784793	0.2410549	-0.1204389	-0.0474769	0.2879052	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156				
7	M	0.8031327	0.7955561	0.1313774	0.8814244	0.0784793	0.2410549	-0.1204389	-0.0474769	0.2879052	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156					
8	M	0.105371	0.4581107	0.7105453	0.7346235	0.3110210	1.5930855	0.6533203	0.7075153	1.4273754	-0.5520956	1.9352055	1.3931208	2.2334073	0.2737010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010			
9	M	0.105371	0.4581107	0.7105453	0.7346235	0.3110210	1.5930855	0.6533203	0.7075153	1.4273754	-0.5520956	1.9352055	1.3931208	2.2334073	0.2737010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010	0.2377010				
10	M	1.2434558	0.1723197	1.5961873	1.1930172	-0.119743	0.0549466	0.7666074	0.7093551	-0.1459511	-0.0749761	1.6037031	1.1453065	1.1679564	1.1040788	1.3205170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170	0.2035170		
11	M	1.2288777	0.2597166	1.3045659	1.1764719	0.7344863	1.6139142	1.8818570	1.9399112	1.3790125	-0.1595082	1.5233284	0.44747474	1.3382376	1.3452109	1.2556384	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474	0.44747474			
12	M	0.2236024	1.2798504	0.1397879	0.1032282	0.5723409	0.7451612	0.1520408	0.3270869	0.0677020	-0.0707008	1.7611211	1.7617261	1.3877681	1.3840778	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220	0.2161220			
13	M	-0.2491305	0.1793055	-0.2020498	-0.2020498	0.0784793	0.2410549	-0.1204389	-0.0474769	0.2879052	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156	0.9623156						
14	M	-0.2795402	0.8119007	-0.2785359	-0.3381059	0.0781008	0.3091326	-0.0781104	0.0648827	-0.2237276	-0.1373740	-0.7358869	-0.1020849	-0.7040499	-0.6686918	0.5658351	-0.5388997	-0.7977370	-0.1384370	-0.1384370	0.1223102	0.1223102	0.1223102	0.1223102	0.1223102	0.1223102	0.1223102	0.1223102	0.1223102	0.1223102	0.1223102				
15	M	-0.2795402	-0.1034036	-0.2340516	-0.3495795	0.3975471	0.3638807	0.5416913	0.0650507	0.1170381	0.6968791	-0.4020767	-0.5954881	-0.4171905	0.3417733	-0.1667799	-0.1247479	-0.0998265	0.0589352	-0.0593933	-0.1247479	0.1167799	-0.1247479	-0.0998265	-0.0589352	-0.0593933	-0.1247479	-0.1247479	-0.1247479	-0.1247479	-0.1247479				

Step 6: Best First Algorithm

We want to choose the features that will be the most useful for predicting the diabetes outcome. Create a new tab by pressing the “+” button on the bottom of the page with the name “BEST_FIRST”.

Import data into the input spreadsheet of the “BEST_FIRST” tab from the output of the “NORMALIZE_TRAIN_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Use the best first algorithm by choosing: *Data Transformation → Variable Selection → Best First*.



The results will appear on the output spreadsheet.

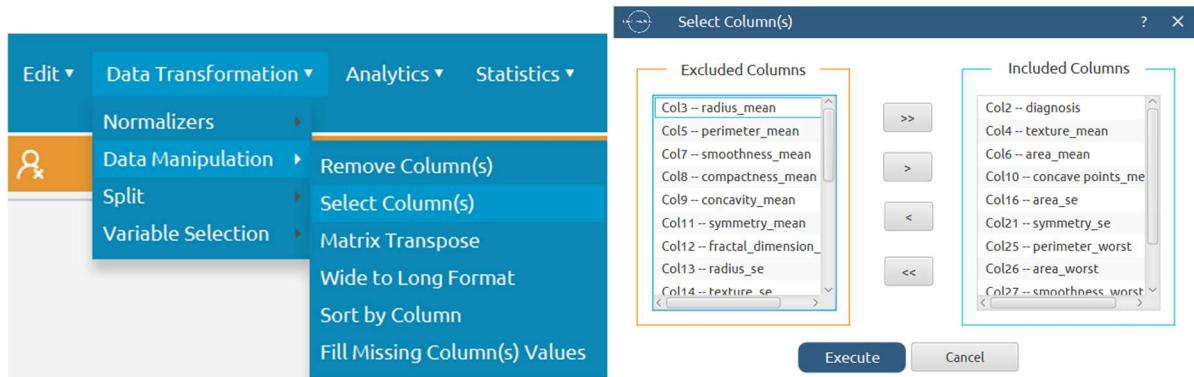
User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (S)
1		-2.0360356	0.9365649	2.5227065	2.3531581	1.1842766	2.2936836	2.0016784	1.3443387	2.1582478	2.2885712	M	
2		0.3960639	1.4935904	2.0297704	1.1143034	0.2588685	1.3388526	1.4550635	0.5554415	0.8768970	1.9496047	M	
3		0.2014065	-0.7590540	1.4469761	-0.2795906	4.8220028	-0.2563994	-0.5573610	3.4539739	2.0357477	2.1690250	M	
4		-1.1500086	1.7528003	1.4238698	1.1228588	-0.3482385	1.3299289	1.2187848	0.2452010	0.6299368	0.7314439	M	
5		-0.8457164	-0.5081168	0.8228472	-0.2804253	0.1544068	-0.1216522	-0.2507276	2.0933478	1.2938872	0.9069802	M	
6		0.1119088	1.0441094	0.6459550	0.2771323	1.28238466	1.3596744	1.2734463	0.5465775	0.5240967	1.1960096	M	
7		0.3020914	-0.2301555	0.2826714	0.2155756	-0.6800581	0.0925155	0.0232851	1.4861629	-0.0178436	0.6270302	M	
8		0.5235981	-0.3903693	1.1473634	-0.3403126	0.1273696	-0.0383647	-0.2547832	1.7033312	1.3110372	1.3897047	M	
9		1.0001733	0.3345912	0.1232373	-0.0844872	1.1572394	0.1341592	-0.0128620	-0.817766	-0.1922837	-0.0342571	M	
10		0.7003561	-0.2290525	0.8064160	-0.4469414	-0.0963014	0.0389736	-0.3281354	1.4728669	2.0720077	1.6136647	M	
11		1.8034150	-0.0070695	0.6367124	-0.1685799	-0.2241134	0.4940800	0.1047484	1.5925311	2.1126778	0.8630961	M	
12		0.1454704	0.0637996	0.0962799	0.0995569	-0.7734592	0.4732581	0.4482341	0.6440816	0.0977965	0.7072321	M	
13		0.2665297	0.3789878	1.3853591	0.2827663	-0.4305789	0.8718480	0.7603336	2.0844838	1.0140971	1.4093769	M	
14		-1.1455337	-0.2621431										

Step 7: Feature Selection: Test set

We need to select the features of the test set that the best first algorithm indicated. Create a new tab by pressing the “+” button on the bottom of the page with the name “FEATURE_SELECTION”.

Import data into the input spreadsheet of the “FEATURE_SELECTION” tab from the output of the “NORMALIZE_TEST_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Select the columns that correspond to the important features: Data Transformation → Data Manipulation → Select Column(s)



The results will appear on the output spreadsheet.

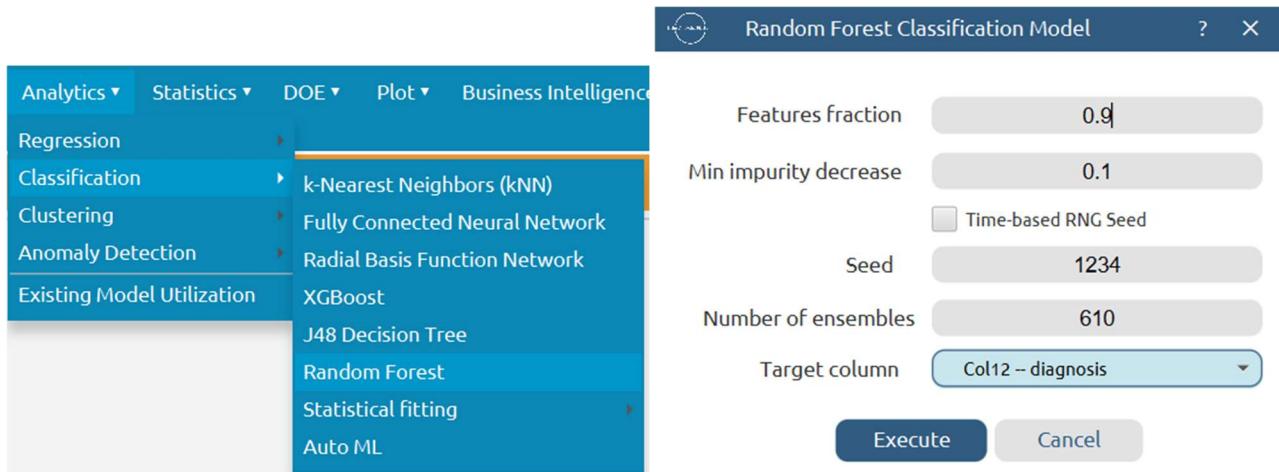
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)
User Header	User Row ID	diagnosis	texture_mean	area_mean	concave_points_mean	area_se	symmetry_se	perimeter_worst	area_worst	smoothness_worst	concavity_worst	concave_points_worst
1		M	-0.3825659	1.8327693	0.5476245	0.6980132	-0.7992674	1.5262493	1.8905922	-0.3575519	-0.1462237	1.0870561
2		M	1.0203103	-0.5114258	0.9394060	-0.3482420	-0.3076828	-0.2926889	-0.3039785	2.3681323	4.0844390	1.6166912
3		M	0.8413149	0.3762302	-0.4007640	-0.0024811	-0.7120111	0.4851563	0.4693934	-0.6101763	-0.6151540	-0.2181161
4		M	-0.3557166	0.3299034	0.4421054	0.2823490	-0.0385402	0.8629243	0.7321212	0.3427052	0.6127868	1.0113939
5		M	1.1903559	1.2729862	1.6164229	1.5769177	3.0043686	1.3150562	0.7903093	-1.2483852	0.4530467	0.9463245
6		M	0.5974337	1.6507709	1.1845904	1.4976243	-0.8398231	2.3591237	2.6699593	0.8568180	1.3022172	1.8860485
7		M	0.7965661	2.0478584	0.9622556	1.1134688	-1.1753296	2.3948183	3.0525897	0.3648652	0.2158866	1.3125293
8		M	0.4251507	0.6707368	1.1003805	1.2931309	-0.7021794	2.0676177	2.3472801	2.1553959	0.9704870	1.4426682
9		M	0.4587123	-0.0456753	1.0010230	-0.4085467	-0.7193849	0.4435126	0.0231088	0.9144340	1.3840473	2.3596936
10		M	0.1723197	1.1930172	0.7309351	1.1040788	0.3117139	0.9640591	0.9155017	0.0856488	0.3584766	0.5271561
11		M	1.2597166	1.1764719	1.9399122	1.3432109	-0.0397691	1.5768167	1.3862959	0.7637458	1.6751075	1.0688972
12		M	1.2798536	0.1032326	-0.5096207	1.3640776	-1.5376275	-0.3554519	-0.3261958	-1.6840515	-1.2125624	-1.2888870
13		M	0.1790320	-0.3203275	0.3270869	-0.1940372	-0.3937101	0.1668793	0.0412705	0.9365941	0.4652967	0.5301826
14		M	0.5213607	-0.3581059	0.0944827	-0.5454321	-1.1384608	-0.0591866	-0.2523146	0.8169299	0.4966567	0.7042056
15		M	-0.1834336	-0.3495575	0.6305507	-0.3417733	-0.3642150	-0.1394995	-0.2193414	2.0711878	1.1228771	1.4320755

Step 8: Train the model

Create a new tab by pressing the “+” button on the bottom of the page with the name “TRAIN_MODEL(.fit)”.

Import data into the input spreadsheet of the “TRAIN_MODEL(.fit)” tab from the output of the “BEST_FIRST” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Use the Random Forest method to train and fit the model: *Analytics → Classification → Random Forest*



The predictions will appear on the output spreadsheet.

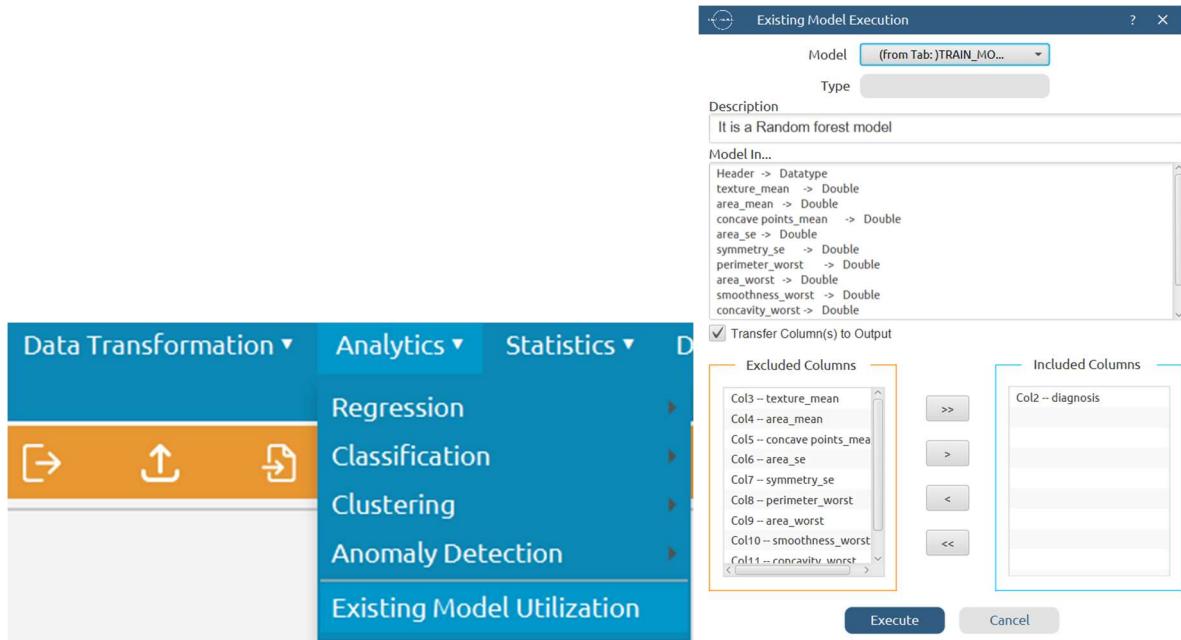
	Col1	Col2 (S)	Col3 (S)
User Header	User Row ID	diagnosis	Prediction
1		M	M
2		M	M
3		M	B
4		M	M
5		M	B
6		M	M
7		M	M
8		M	B
9		M	M
10		M	B
11		M	M
12		M	M
13		M	M
14		B	B
15		B	B

Step 9: Validate the model

Create a new tab by pressing the “+” button on the bottom of the page with the name “VALIDATE_MODEL(.predict)”.

Import data into the input spreadsheet of the “VALIDATE_MODEL(.predict)” tab from the output of the “FEATURE_SELECTION” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

To validate the model: *Analytics → Existing Model Utilization → Model (from Tab:) TRAIN_MODEL(.fit)*. Choose the column “diagnosis” to be transferred to the output spreadsheet.



The predictions will appear on the output spreadsheet.

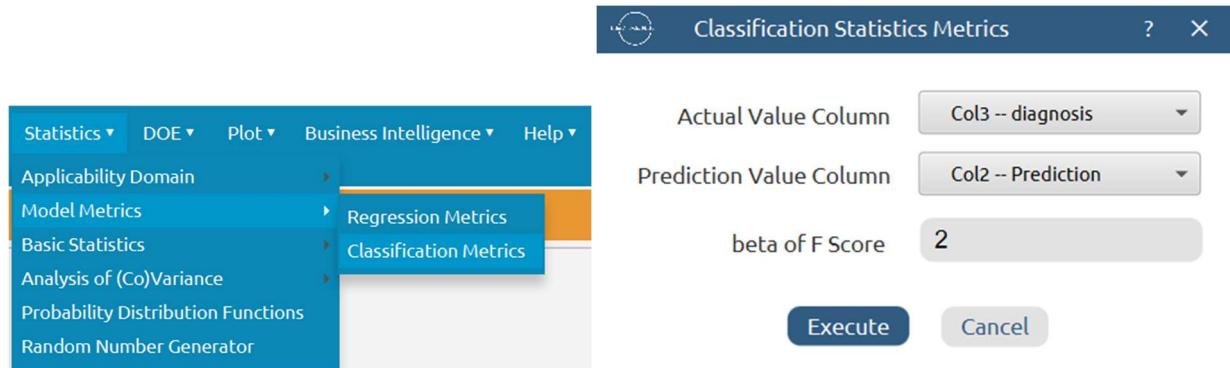
	Col1	Col2 (S)	Col3 (S)
User Header	User Row ID	Prediction	diagnosis
1		M	M
2		B	M
3		M	M
4		M	M
5		M	M
6		M	M
7		M	M
8		M	M
9		M	M
10		M	M
11		M	M
12		B	M
13		M	M
14		B	M
15		B	M

Step 10: Statistics calculation

Create a new tab by pressing the “+” button on the bottom of the page with the name “STATISTICS_ACCURACIES”.

Import data into the input spreadsheet of the “STATISTICS_ACCURACIES” tab from the output of the “VALIDATE_MODEL(.predict)” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Calculate the statistical metrics for the classification: Statistics → Model Metrics → Classification Metrics



The results will appear on the output spreadsheet.

	Col1 (S)	Col2 (S)	Col3 (S)	Col4 (S)
User Header	User Row ID			
1			Predicted Class	Predicted Class
2			M	B
3	Actual Class	M	42	11
4	Actual Class	B	4	85
5				
6				
7	Classification Accuracy	0.8943662		
8				
9	Precision		0.9130435	0.8854167
10				
11	Recall/Sensitivity		0.7924528	0.9550562
12				
13	Specificity		0.9550562	0.7924528
14				
15	F1 Score		0.8484848	0.9189189
16				
17	F (beta=2)		0.8139535	0.9402655
18				
19	MCC	0.7725647		

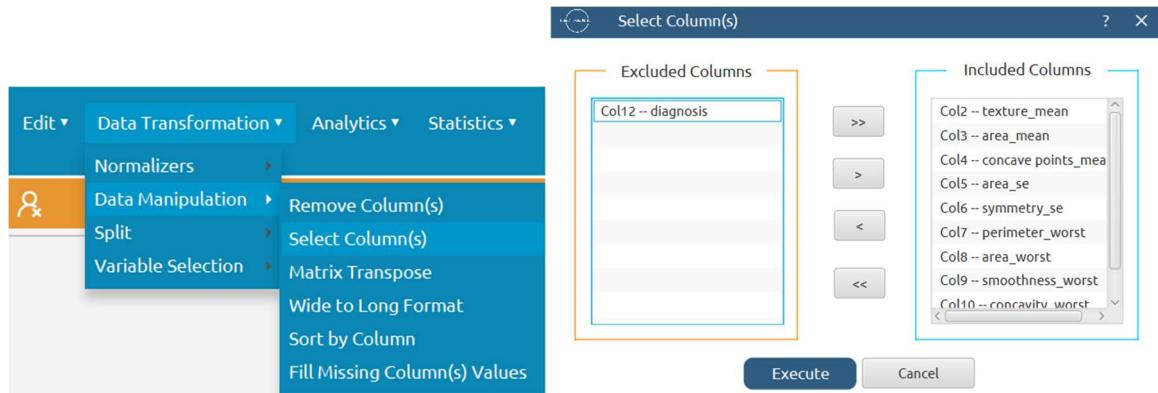
Step 11: Reliability check for each record of the test set

Step 11.a: Create the domain

Create a new tab by pressing the “+” button on the bottom of the page with the name “EXCLUDE_DIAGNOSIS”.

Import data into the input spreadsheet of the “EXCLUDE_DIAGNOSIS” tab from the output of the “BEST_FIRST” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Manipulate the data to exclude the target column “diagnosis”: Data Transformation → Data Manipulation → Select Column(s)

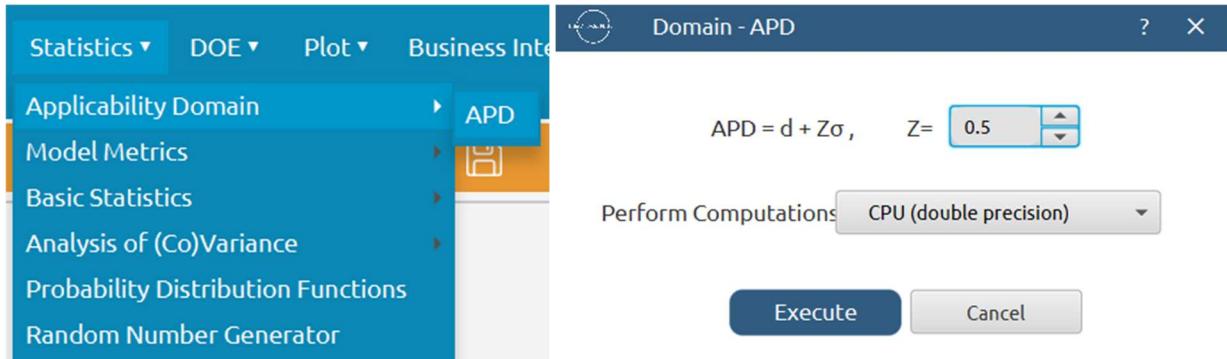


The results will appear on the output spreadsheet.

Create a new tab by pressing the “+” button on the bottom of the page with the name “DOMAIN”.

Import data into the input spreadsheet of the “DOMAIN” tab from the output of the “EXCLUDE_DIAGNOSIS” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Create the domain: Statistics → Applicability Domain → APD



The results will appear on the output spreadsheet.

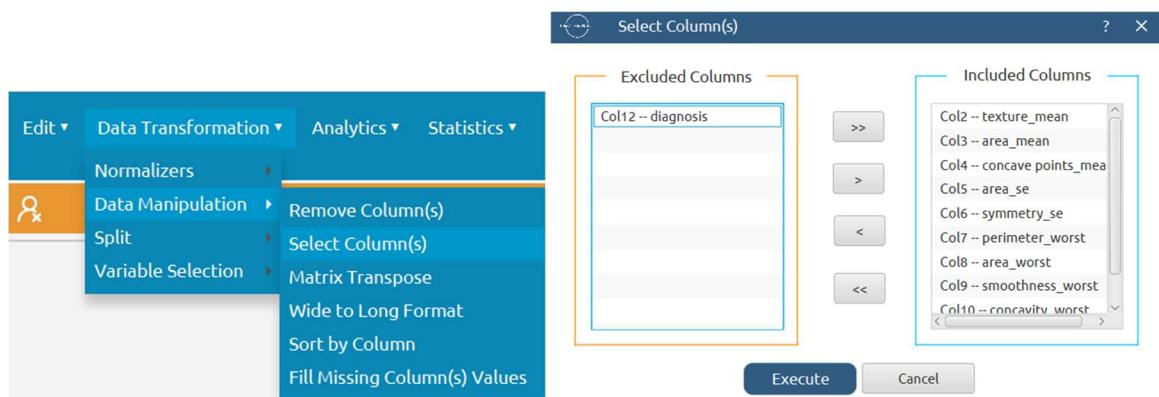
	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		0.0	2.9702124	reliable
2		0.0	2.9702124	reliable
3		0.0	2.9702124	reliable
4		0.0	2.9702124	reliable
5		0.0	2.9702124	reliable
6		0.0	2.9702124	reliable
7		0.0	2.9702124	reliable
8		0.0	2.9702124	reliable
9		0.0	2.9702124	reliable
10		0.0	2.9702124	reliable
11		0.0	2.9702124	reliable
12		0.0	2.9702124	reliable
13		0.0	2.9702124	reliable
14		0.0	2.9702124	reliable
15		0.0	2.9702124	reliable

Step 11.b: Check the test set reliability

Create a new tab by pressing the “+” button on the bottom of the page with the name “EXCLUDE_DIAGNOSIS_TEST_SET”.

Import data into the input spreadsheet of the “EXCLUDE_DIAGNOSIS_TEST_SET” tab from the output of the “FEATURE_SELECTION” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Manipulate the data to exclude the target column “diagnosis”: [Data Transformation → Data Manipulation → Select Column\(s\)](#)

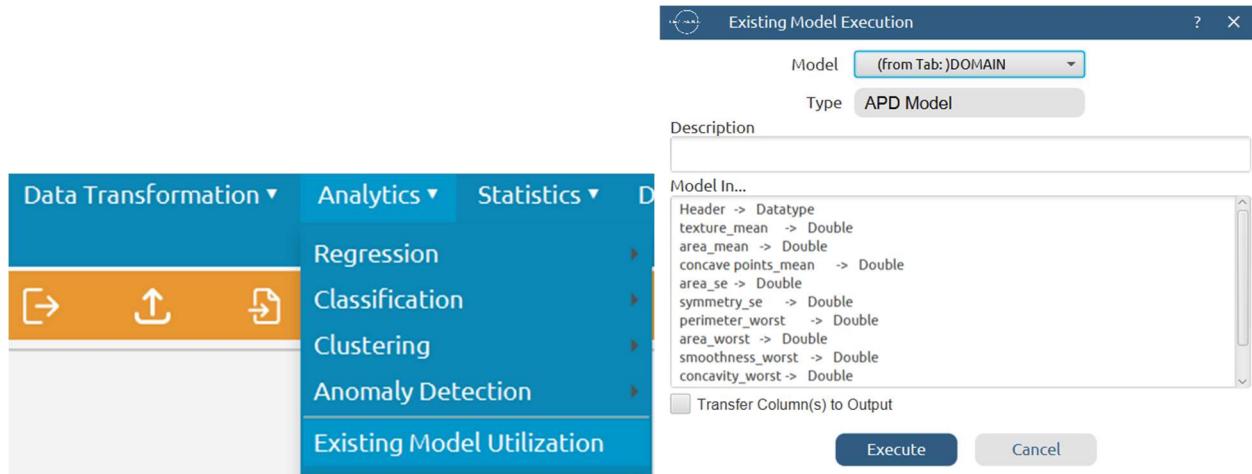


The results will appear on the output spreadsheet.

Create a new tab by pressing the “+” button on the bottom of the page with the name “RELIABILITY”.

Import data into the input spreadsheet of the “RELIABILITY” tab from the output of the “EXCLUDE_DIAGNOSIS_TEST_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Check the Reliability: [Analytics → Existing Model Utilization → Model \(from Tab:\) DOMAIN](#)



The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		1.0748767	2.9702124	reliable
2		2.2838975	2.9702124	reliable
3		0.5921315	2.9702124	reliable
4		0.9078477	2.9702124	reliable
5		1.8209589	2.9702124	reliable
6		1.5301520	2.9702124	reliable
7		1.7180436	2.9702124	reliable
8		1.3950809	2.9702124	reliable
9		1.1623175	2.9702124	reliable
10		0.9187315	2.9702124	reliable
11		1.4637847	2.9702124	reliable
12		2.1716145	2.9702124	reliable
13		0.6998135	2.9702124	reliable
14		0.8731378	2.9702124	reliable
15		1.0393431	2.9702124	reliable

Final Isalos Workflow

Following the above-described steps, the final workflow on Isalos will look like this:

