



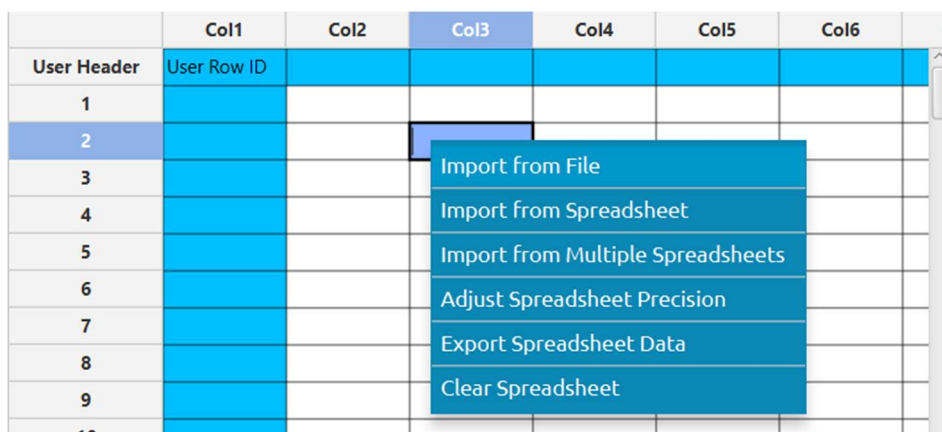
Breast Cancer Wisconsin (Diagnostic) Data set

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.

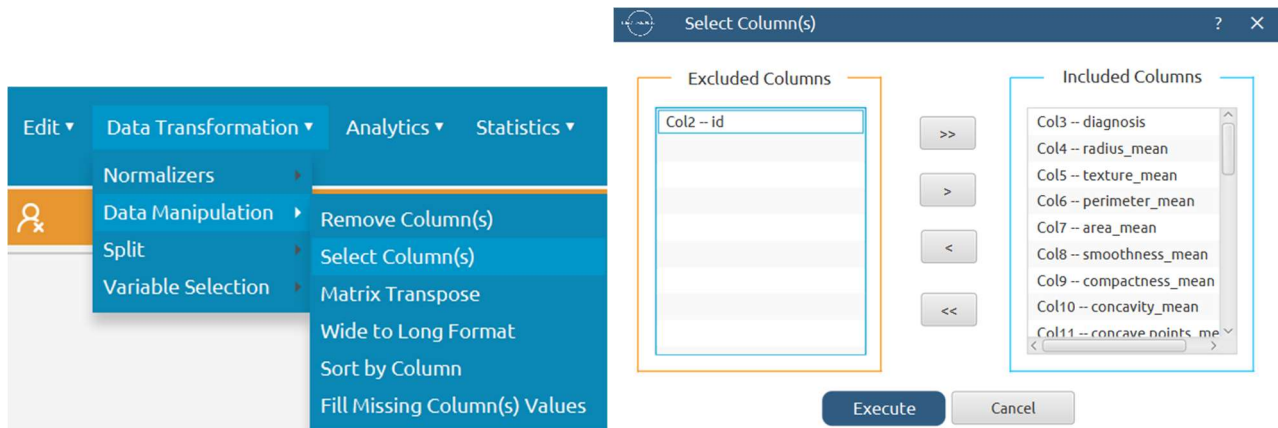


The data will appear on the left spreadsheet.

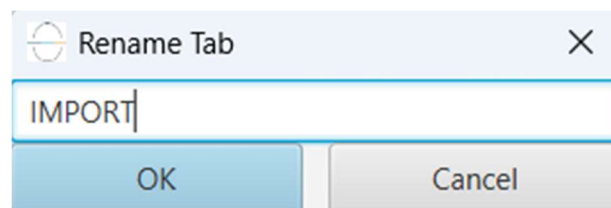
User Header	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12	Col13	Col14	Col15	Col16	Col17	Col18	Col19	Col20	Col21	Col22	Col23	Col24	Col25	Col26	Col27	Col28	Col29	Col30
	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1	Col Row 1
1	42332	M	17.99	10.38	122.8	1001	0.1184	0.2778	0.3001	0.1471	0.2419	0.0781	1.095	0.9053	0.889	153.4	0.00899	0.04904	0.05373	0.01887	0.00003	0.00893	25.38	17.33	184.5	2019	0.1622	0.8658	0.7119	
2	42317	M	20.57	17.77	152.9	1028	0.08474	0.07864	0.0899	0.07077	0.1412	0.05987	0.5453	0.7339	3.360	74.08	0.00325	0.01038	0.0148	0.0174	0.01032	0.00352	24.99	12.47	168.8	1958	0.1238	0.1685	0.2416	
3	423003	M	19.69	21.25	150	1203	0.1095	0.1599	0.1974	0.1279	0.2009	0.05999	0.7458	0.7869	4.55	94.03	0.00915	0.04008	0.03832	0.02028	0.02025	0.00471	23.57	25.53	1709	0.1468	0.4345	0.4524		
4	4234001	M	11.42	20.38	77.58	388.1	0.1425	0.0839	0.2414	0.1052	0.2597	0.09744	0.4958	1.156	3.445	27.23	0.00911	0.07408	0.05681	0.01887	0.00963	0.00028	14.91	26.5	98.07	0.2088	0.8863	0.8869		
5	4235402	M	20.29	14.34	135.1	1287	0.1009	0.1328	0.198	0.1048	0.1809	0.05883	0.7572	0.7813	5.458	94.44	0.01489	0.05481	0.05688	0.01758	0.00915	0.00915	22.54	16.07	152.2	1575	0.1374	0.2025	0.4	
6	423768	M	12.45	15.7	82.97	471.1	0.1278	0.171	0.1578	0.08089	0.2087	0.07953	0.3445	0.6902	2.217	27.59	0.00751	0.03345	0.03972	0.01137	0.02165	0.00382	15.47	23.75	103.4	741.6	0.1791	0.5249	0.5355	
7	424559	M	18.55	19.89	119.6	1500	0.09483	0.109	0.1127	0.074	0.1784	0.07162	0.4487	0.7732	3.15	53.91	0.00414	0.01382	0.02254	0.01039	0.01369	0.00179	22.88	17.66	150.2	1695	0.1442	0.2378	0.3784	
8	4248202	M	15.71	20.83	95.2	577.9	0.1189	0.1045	0.09085	0.0985	0.2195	0.07451	0.5835	1.377	3.855	50.95	0.00805	0.03029	0.02488	0.01448	0.01485	0.005412	17.06	28.14	110.5	997	0.1854	0.3582	0.2878	
9	424981	M	13	21.82	87.5	519.8	0.1279	0.1892	0.1859	0.0933	0.235	0.07389	0.3063	1.002	2.405	24.52	0.00731	0.03952	0.05853	0.01265	0.00143	0.00749	15.49	30.73	106.2	738.3	0.1703	0.5401	0.539	
10	4250101	M	12.45	24.04	83.97	475.9	0.1188	0.2395	0.2273	0.0843	0.203	0.08343	0.2978	1.599	2.039	23.84	0.00748	0.07217	0.07743	0.01432	0.01789	0.0108	15.09	45.05	97.65	711.4	0.1833	1.058	1.105	
11	425056	M	18.02	23.24	102.7	797.8	0.08208	0.06669	0.02099	0.0323	0.1038	0.05987	0.2795	1.187	2.465	40.21	0.00429	0.000268	0.0101	0.007581	0.0148	0.00342	19.19	33.88	125.8	1153	0.1181	0.1551	0.1459	
12	4251002	M	15.78	17.89	103.6	781	0.0971	0.1282	0.09954	0.09606	0.1842	0.06082	0.3058	0.9649	3.884	54.16	0.00771	0.04061	0.02791	0.01382	0.00208	0.004144	20.42	27.28	156.5	1269	0.1395	0.5509	0.3945	
13	42526	M	19.17	24.8	132.4	1123	0.0974	0.2488	0.2085	0.1118	0.2397	0.078	0.8555	3.588	11.07	116.2	0.00939	0.08297	0.0889	0.0408	0.04484	0.01284	20.95	29.84	151.7	1332	0.1037	0.3903	0.3639	
14	42531	M	15.85	23.95	103.7	782.7	0.08401	0.1002	0.09058	0.05364	0.1847	0.05338	0.4033	1.078	2.903	38.58	0.00789	0.03128	0.05051	0.01982	0.00361	0.003002	16.84	27.66	112	878.5	0.1131	0.1824	0.2322	
15	4257491	M	15.73	22.61	93.6	578.3	0.1131	0.2289	0.2128	0.08025	0.2099	0.07882	0.2121	1.189	2.061	19.21	0.006429	0.03838	0.03801	0.01828	0.01961	0.000893	15.03	32.01	158.8	897.7	0.1681	0.7725	0.6843	

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”.



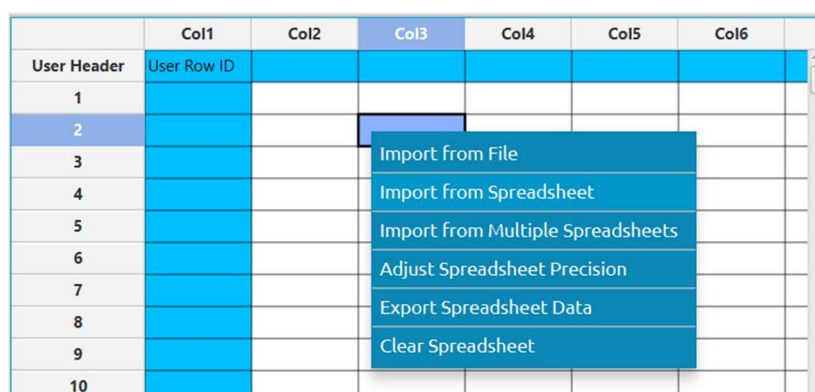
All of the data will appear in the output (right) spreadsheet. This tab can be renamed “IMPORT” by right-clicking on it and choosing the “Rename” option.



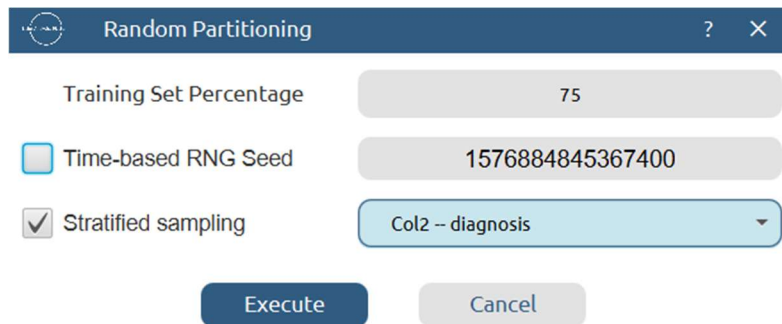
Step 3: Split data

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”.



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:



Random Partitioning

Training Set Percentage: 75

☐ Time-based RNG Seed: 157688445367400

☒ Stratified sampling: Col2 -- diagnosis

Execute Cancel

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.

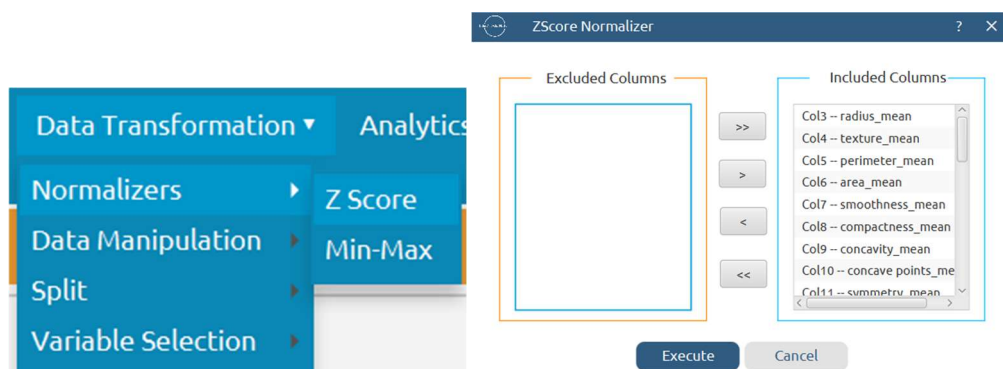
Step 4: Normalize the training set

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	Col1	Col2 (S)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)	Col9 (S)	Col10 (S)	Col11 (S)	Col12 (S)	Col13 (S)	Col14 (S)	Col15 (S)	Col16 (S)	Col17 (S)	Col18 (S)	Col19 (S)	Col20 (S)	Col21 (S)	Col22 (S)	Col23 (S)	Col24 (S)	Col25 (S)	Col26 (S)	Col27 (S)	Col28 (S)	Col29 (S)	Col30	
1	M	1739	133.8	132.8	1021	0.1154	0.2776	0.2021	0.1471	0.2473	0.2771	1.08	0.3053	4.55	153.4	0.02639	0.04024	0.2573	0.0187	0.0303	0.02618	25.38	17.33	18.6	2219	0.1822	0.8555	0.7119	0.2854		
2	M	1989	21.25	130	1203	0.1096	0.1999	0.1974	0.1279	0.2069	0.00999	0.1458	0.1969	4.55	64.03	0.00911	0.04009	0.03832	0.02058	0.0225	0.00471	23.87	26.53	152.5	1709	0.1444	0.4345	0.4304	0.243		
3	M	1142	20.88	77.58	386.1	0.1425	0.2889	0.2414	0.1052	0.2597	0.00764	0.4955	1.158	3.445	27.23	0.00911	0.07459	0.03661	0.01887	0.08963	0.009208	14.91	26.5	98.87	547.7	0.2098	0.8868	0.6889	0.2875		
4	M	2028	14.34	135.1	1297	0.1003	0.1328	0.1168	0.1043	0.1929	0.00883	0.7972	0.7913	4.438	94.44	0.01149	0.03481	0.03688	0.01885	0.01758	0.008119	22.84	16.67	182.2	1975	0.1376	0.225	0.4	0.1625		
5	M	1245	18.7	82.37	477.1	0.1278	0.17	0.1578	0.08369	0.2047	0.01913	0.3345	0.8402	2.217	27.19	0.00791	0.03145	0.03872	0.01137	0.02163	0.00392	15.47	23.75	102.4	741.6	0.1791	0.5248	0.5165	0.1741		
6	M	1825	19.96	119.6	1040	0.09483	0.108	0.1127	0.074	0.1794	0.05742	0.4487	0.7732	3.18	53.91	0.00434	0.01382	0.02354	0.01039	0.01359	0.002179	22.88	27.66	153.2	1606	0.1442	0.2576	0.3784	0.1932		
7	M	1571	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.09865	0.2196	0.07461	0.5835	1.377	3.856	50.96	0.00805	0.03029	0.02488	0.01448	0.01488	0.005412	17.08	28.14	110.6	897	0.1654	0.3582	0.2878	0.1585		
8	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1889	0.09353	0.235	0.07389	0.3063	1.002	2.406	24.32	0.00731	0.03832	0.03553	0.01226	0.02143	0.003748	15.49	30.73	106.1	739.3	0.1709	0.5421	0.539	0.206		
9	M	1585	23.95	110.7	152.7	0.06491	0.1002	0.09958	0.05354	0.1647	0.00458	0.4533	1.978	3.903	34.58	0.00799	0.03128	0.03551	0.01942	0.02041	0.003002	16.54	27.86	112	875.5	0.1131	0.1624	0.2122	0.1119		
10	M	1373	22.61	92.4	578.3	0.1131	0.2233	0.2128	0.06325	0.2059	0.07862	0.2121	1.159	2.951	19.21	0.006429	0.03598	0.03551	0.01658	0.01981	0.000593	15.53	32.51	108.8	687.7	0.1451	0.7725	0.6943	0.2238		
11	M	1454	27.84	96.73	658.8	0.1139	0.1995	0.1839	0.07364	0.2303	0.07077	0.37	1.033	2.879	32.55	0.005007	0.0424	0.04741	0.0109	0.01837	0.005486	17.48	37.13	124.1	942.2	0.1678	0.6577	0.7028	0.1712		
12	M	1468	20.13	94.74	684.5	0.09887	0.072	0.07395	0.05259	0.1558	0.05822	0.4727	1.24	3.195	45.4	0.00578	0.01162	0.01998	0.01109	0.0141	0.002385	19.07	30.83	123.4	1138	0.1454	0.1871	0.2914	0.1609		
13	M	1613	20.68	108.1	798.8	0.117	0.2022	0.1722	0.1038	0.2164	0.07558	0.5452	1.973	3.854	54.18	0.007028	0.03251	0.03168	0.01287	0.01689	0.004742	20.98	31.48	136.8	1315	0.1759	0.4233	0.4784	0.2073		
14	S	1534	14.35	87.45	595.5	0.09779	0.20129	0.09564	0.04781	0.1833	0.05766	0.2949	0.7884	2.955	23.56	0.004452	0.0148	0.03387	0.01015	0.0198	0.003	15.11	19.24	99.7	771.2	0.144	0.1773	0.239	0.1288		
15	S	1308	15.71	88.63	520	0.1075	0.127	0.04688	0.0311	0.1967	0.08811	0.1852	0.7477	1.383	14.67	0.004097	0.01898	0.01988	0.00649	0.01678	0.003425	14.5	20.49	96.09	630.5	0.1312	0.2778	0.189	0.07283		

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



ZScore Normalizer

Excluded Columns: (empty)

Included Columns: Col3 -- radius_mean, Col4 -- texture_mean, Col5 -- perimeter_mean, Col6 -- area_mean, Col7 -- smoothness_mean, Col8 -- compactness_mean, Col9 -- concavity_mean, Col10 -- concave points_me, Col11 -- symmetry_mean

Execute Cancel

The results will appear on the output spreadsheet.

User Header	Row Col	Col0 (S)	Col1 (S)	Col2 (S)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)	Col9 (S)	Col10 (S)	Col11 (S)	Col12 (S)	Col13 (S)	Col14 (S)	Col15 (S)	Col16 (S)	Col17 (S)	Col18 (S)	Col19 (S)	Col20 (S)	Col21 (S)	Col22 (S)	Col23 (S)	Col24 (S)	Col25 (S)	Col26 (S)	Col27 (S)	Col28 (S)	Col29 (S)	Col30	
1	M	1.025084	-2.030558	1.224927	0.955549	1.550469	1.397793	0.800522	2.527055	2.499980	2.218140	2.488183	-0.174549	2.832329	2.133181	-0.210589	1.264236	0.733743	0.676773	1.184276	0.877378	1.871625	-1.3714614	2.291638	2.0016784	1.344337	2.058718	2.182478	2.288572				
2	M	1.520268	0.390839	1.513769	1.493854	0.960291	1.065190	1.380781	2.030774	0.956297	0.956297	1.219088	-0.780103	0.848469	-0.281979	0.797208	0.213844	1.448148	0.288865	2.887770	1.498815	-0.050062	1.338838	1.450065	0.856415	1.062814	0.878970	1.849487					
3	M	-0.783312	0.201405	-0.591846	0.759540	0.200223	0.477872	1.009124	1.448971	2.906716	4.414073	0.325987	-0.117493	0.248116	-0.279586	0.871448	0.708783	0.809023	1.143870	4.822023	1.801291	0.246063	0.159642	-0.253964	-0.557349	0.403970	0.302191	2.037477	2.188250				
4	M	1.684780	-1.150086	1.718148	1.763003	0.304032	0.547780	1.388188	1.423484	2.960339	-0.003152	1.271875	-0.000152	1.128888	1.631761	-0.036261	0.8400107	1.168118	-0.340288	0.462975	2.068364	-1.4777341	1.326928	1.218748	0.3452010	-0.312586	0.6299168	0.715439					
5	M	-0.478587	-0.845714	-0.390951	-0.038158	2.344077	1.257785	0.878270	0.822842	1.024784	1.883247	-0.208800	-0.801881	-0.323197	-0.280423	0.183743	0.401049	0.199149	-0.05843	0.154808	0.470873	-0.170810	-0.337285	-0.121832	-0.207276	2.083478	1.738818	1.293872	0.806802				
6	M	1.124838	0.111908	1.058120	1.044104	-0.094705	0.268197	0.3058741	0.643950	-0.556845	-0.737487	0.150289	-0.815079	0.135782	0.277133	-0.655740	-0.318493	-0.219143	-0.823845	-0.591855	1.358705	0.292489	1.596744	1.273483	0.545775	0.026820	0.524967	1.196098					
7	M	-0.130252	0.320914	-0.054915	-0.230155	1.815211	1.157738	0.064357	0.328714	1.429888	1.833329	0.834808	0.283232	0.484477	0.215756	0.578189	0.255287	-0.239501	0.448481	-0.800581	0.591834	0.158824	0.368123	0.023281	1.489169	0.711951	-0.317426	0.627032					
8	M	0.3285271	0.323561	-0.193027	-0.360369	2.287785	0.696955	1.248544	1.147054	1.955328	1.567326	0.329103	-0.389154	-0.226028	-0.340337	0.519185	0.197744	0.068448	0.127596	-0.0771301	0.1607389	0.797002	-0.038547	-0.2547832	1.703192	0.820682	1.311032	1.589747					
9	M	0.485101	1.000103	0.425167	0.334912	-0.244881	-0.075487	0.109483	0.123273	0.190110	-1.297812	-0.004810	-0.236872	0.016502	-0.084872	0.884134	0.308084	0.821206	1.338744	-0.290883	0.111508	0.263499	0.134192	-0.012802	-0.031776	-0.3939564	-0.192837	-0.034291					
10	M	-0.124721	0.700561	0.051858	1.207505	0.287580	1.576262	0.8064160	0.955297	1.933956	-0.6884564	-0.068073	-0.4040742	-0.4468414	-0.2205372	1.858873	0.778628	0.743659	-0.065014	1.573953	-0.2615362	0.8931915	0.0389736	-0.3281354	-0.1472669	0.3244274	2.072007	1.618647					
11	M	0.269199	1.804150	-0.177561	-0.007895	1.283342	0.071469	0.855887	0.6587124	1.8218702	1.114708	-0.1238751	-0.3416409	0.048467	-0.188079	-0.469440	0.028468	0.520540	-0.137742	-0.2241134	0.8114071	0.232955	1.817925	0.444000	0.1047484	1.582511	2.589164	2.112878	0.883061				
12	M	0.157923	0.145474	0.707675	0.067196	0.100248	-0.008458	-0.187488	0.064799	-0.024045	-0.074428	0.240323	0.055784	0.181066	0.095549	-0.235963	-0.777882	-0.444801	-0.104744	-0.775892	-0.026068	0.571190	0.811172	0.474251	0.440214	0.644076	0.427915	0.097765	0.701321				
13	M	0.538737	0.286297	0.634792	0.376878	1.482162	0.871603	1.041066	1.385391	1.308854	1.501791	0.588319	-0.2887487	0.487473	0.262765	-0.048739	-0.058065	-0.0035971	0.203805	-0.405789	0.108742	0.906832	0.8079191	0.871840	0.760336	2.064888	1.064957	1.014071	1.405769				
14	M	-0.177250	-1.145837	-0.184288	-0.2821431	0.127708	-0.432821	-0.278099	-0.024407	0.3762483	-0.073816	-0.4877782	-0.7870124	-0.401959	-0.036713	0.881767	-0.012548	-0.273721	0.232031	-0.072811	-0.5475614	-0.248040	-1.060572	-0.231706	-0.043312	0.5377138	-0.490878	-0.118687	0.2214810				
15	B	-0.354428	-0.484790	-0.268104	-0.898178	0.810775	0.438197	-0.444884	-0.454491	0.818887	0.747429	0.7848157	-0.8815447	-0.782341	-0.841761	-0.486387	-0.3700712	-0.050043	-0.858765	-0.444875	-0.571800	-0.370077	-0.852489	-0.393919	-0.468572	-0.026838	0.1517286	-0.039638	-0.6254812				

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	Col1	Col2 (S)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)	Col9 (S)	Col10 (S)	Col11 (S)	Col12 (S)	Col13 (S)	Col14 (S)	Col15 (S)	Col16 (S)	Col17 (S)	Col18 (S)	Col19 (S)	Col20 (S)	Col21 (S)	Col22 (S)	Col23 (S)	Col24 (S)	Col25 (S)	Col26 (S)	Col27 (S)	Col28 (S)	Col29 (S)	Col30	
1	M	20.57	17.77	132.9	1326	0.08474	0.07884	0.0889	0.07017	0.1812	0.06867	0.5405	0.7339	3.39	74.08	0.015225	0.01388	0.0188	0.0134	0.0138	0.003532	24.99	23.41	158.8	1956	0.1238	0.1886	0.2416	0.168		
2	M	12.48	24.04	89.97	475.9	0.1108	0.2198	0.2273	0.08343	0.203	0.08343	0.2478	1.89	2.09	23.84	0.007149	0.07217	0.07743	0.01432	0.01789	0.01008	19.59	40.88	87.61	711.4	0.1833	1.038	1.106	0.221		
3	M	16.02	23.34	102.7	787.8	0.08306	0.26689	0.01039	0.01033	0.1528	0.00887	0.3798	1.187	2.49	48.51	0.004029	0.00009	0.01031	0.007181	0.01448	0.000462	19.18	33.88	123.8	1190	0.1181	0.1551	0.1488	0.26975		
4	M	18.78	17.89	103.6	791	0.0871	0.1382	0.09654	0.06806	0.1942	0.00882	0.1058	0.0648	8.64	54.16	0.007791	0.04081	0.02791	0.01282	0.00088	0.004144	20.42	27.28	18.5	1299	0.1036	0.1808	0.3885	0.181		
5	M	18.17	24.8	132.4	1123	0.0874	0.2483	0.2085	0.1118	0.2397	0.078	0.9555	3.58	11.07	116.2	0.003139	0.06287	0.0889	0.04049	0.04464	0.01284	20.96	29.84	151.7	1332	0.1037	0.1893	0.3889	0.1787		
6	M	18.81	22.15	130	1260	0.08613	0.1027	0.1479	0.04688	0.1582	0.08393	0.7582	1.017	8.85	112.4	0.006448	0.01883	0.03391	0.01521	0.01358	0.001987	27.32	38.88	188.8	2388	0.1512	0.315	0.5372	0.2388		
7	M	21.18	20.54	197.2	1454	0.04828	0.1022	0.1087	0.08832	0.1786	0.02729	0.8917	1.137	4.93	99.99	0.004728	0.01029	0.01715	0.01038	0.01083	0.001987	28.17	38.59	188	2416	0.1401	0.28	0.3168	0.2008		
8	M	16.85	21.38	110	804.8	0.1121	0.1487	0.1535	0.0917	0.1995	0.0883	0.3088	0.9017	5.45	102.6	0.00048	0.01882	0.02741	0.0113	0.01485	0.002801	26.46	31.56	177	3215	0.1805	0.3878	0.4895	0.2095		
9	M	14.88	21.53	97.41	644.8	0.1054	0.1888	0.1425	0.08783	0.2252	0.08024	0.2545	0.8032	2.11	21.05	0.004462	0.00055	0.00881	0.01352	0.01454	0.003711	17.62	33.21	122.4	886.9	0.1535	0.6543	0.5339	0.2701		
10	M	18.81	20.25	122.1	1094	0.0844	0.1586	0.1149	0.07731	0.1997	0.08899	0.8829	1.849	6.62	93.84	0.01075	0.02722	0.05081	0.01911	0.02283	0.004217	21.31	27.28	139.9	1403	0.1338	0.2177	0.3448	0.148		
11	M	18.81	25.11	124.8	1588	0.1084	0.1887	0.2319	0.1044	0.2181	0.06197	0.8807	1.488	5.74	195	0.004048	0.00374	0.01186	0.02007	0.00489	0.01158	34.01	160.3	1870	0.1491	0.4287	0.8139	0.1648			
12	M	14.88	25.2	95.54	688.8	0.08887	0.26131	0.01388	0.02899	0.1585	0.08804	1.214	2.18	10.77	106	0.008883	0.01084	0.01818	0.01917	0.007882	0.001754	14.89	25.2	95.54	688.8	0.08887	0.01311	0.02388	0.02319		
13	M	13.28	20.28	87.32	545.2	0.1041	0.1438	0.09847	0.06158	0.1974	0.08782	0.3704	0.8349	2.427	31.33	0.000372	0.02147	0.02185	0.00856	0.01718	0.003317	17.38	28	113.1	887.2	0.153	0.3724	0.3884	0.1482		
14	M	15.17	21.81	85.42	531.5	0.08714	0.1547	0.08239	0.05252	0.1748	0.06177	0.1938	0.8123	1.354	14.49	0.000833	0.01384	0.01452	0.00883	0.01113	0.01072	16.23	28.89	105.5	140.7	0.1583	0.3904	0.3728	0.1607		
15	M	15.17	18.86	85.94	534.6	0.1158	0.1231	0.1228	0.0734	0.2128	0.08777	0.2871	0.8897	1.897	24.25	0.005632	0.02336	0.02005	0.01215	0.01743	0.003548	15.67	27.95	102.8	789.4	0.1788	0.4786	0.3006	0.2088		

Normalize the test set using the existing normalizer of the training set: *Analytics* → *Existing Model Utilization* → *Model (from Tab:) NORMALIZE_TRAIN_SET*

Existing Model Execution

Model

(from Tab:)NORMALIZE...

Type

Z Score Normalizer Model

Description

Model In...

Header -> Datatype
radius_mean -> Double
texture_mean -> Double
perimeter_mean -> Double
area_mean -> Double
smoothness_mean -> Double
compactness_mean -> Double
concavity_mean -> Double
concave points_mean -> Double
symmetry_mean -> Double

Transfer Column(s) to Output

Execute

Cancel

Data Transformation

Analytics

Statistics

Regression

Classification

Clustering

Anomaly Detection

Existing Model Utilization

The results will appear on the output spreadsheet.

User Header	Var Col 1	Col2 (D)	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	
		Segment	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max	metric_min	metric_max
1	M	1.7602045	-0.2823559	1.6326135	1.8227993	-0.7937371	-0.4832219	0.5479243	0.2098442	-0.8119119	0.4954651	-0.8809525	0.2819385	0.8909132	-0.9391412	-0.6967173	-0.4513540	0.2729008	-0.7902874	-0.0965987	0.7912474	-0.2820959	1.5262493	1.8905822	-0.9375519	-0.4311109	-0.1482237	1.0870561			
2	M	-0.4738211	1.0203103	-0.3347543	-0.9114628	1.5905048	2.8838835	1.7902348	0.9304050	0.8143890	2.7821729	-0.3827385	0.8897841	-0.4115951	-0.3482402	0.9354388	2.5733271	1.5332590	0.4323629	-0.9378828	2.3004006	-0.2491700	3.3937442	-0.2926889	-0.3039785	2.3681522	5.1499778	4.5844392	1.6188912		
3	M	0.5888480	0.8413139	0.4172881	0.2762022	-0.8824182	0.7710338	-0.7091457	-0.2076420	1.0325254	-0.7899958	0.0089984	-0.8810583	-0.1999994	-0.0324811	-0.3871232	0.2697372	-0.7074505	0.4789962	-0.7120111	-0.2789188	0.5998813	1.2844109	0.4831583	0.4669184	-0.8197193	0.6323529	-0.8151542	-0.2181181		
4	M	0.8419895	-0.5357166	0.4548404	0.3290854	0.0794760	0.6057188	0.1338787	0.4421554	0.1025038	-0.2655503	0.3919582	-0.4320253	0.3484612	0.3823490	-0.4161929	0.8277385	-0.1372859	0.1788880	-0.0385402	0.1274664	0.8483277	0.2312859	0.8623243	0.7312121	0.3427952	1.5661956	0.8157368	1.0118339		
5	M	1.3791717	1.1903859	1.6097383	1.2728862	0.0996403	0.2701769	1.4862806	1.6164229	2.1887754	2.1176588	1.8965429	4.2779028	4.0897929	1.8791777	-1.2788158	3.1733854	1.8201961	4.7882323	3.0540886	3.3707888	0.8906832	0.8597851	3.1510262	0.7003900	-1.2481882	0.3733396	0.4330487	0.8483246		
6	M	1.5581010	0.8974887	1.5133769	1.6507709	0.1638308	-0.0548482	0.7526488	1.1545904	-0.8388964	-1.2184824	1.2641093	-0.3707978	1.6833470	1.4876343	-0.1778239	-0.3727799	0.0651244	0.5687811	-0.8388231	-0.6584797	2.2714889	0.8111712	2.3591237	2.8699993	0.8889180	0.3912635	1.3022172	1.8880485		
7	M	1.6261112	0.7946681	1.6024811	-0.1232439	-0.0344789	0.2879052	0.5652356	-0.1483485	-1.3807915	1.035428	-0.1703442	0.7098116	1.1134888	-0.7530301	-0.7123838	-0.8002494	-0.2207788	-1.1733298	-0.6821433	2.6527370	1.5868775	2.3448183	3.0523497	0.3648952	0.0390203	0.2158896	1.3128293			
8	M	1.6822127	0.4451927	0.7105653	0.6707166	1.1065551	0.7940255	0.8110210	1.1002005	0.8652121	0.0781153	0.4278755	-0.0090996	1.2002050	1.3931109	-0.2524073	0.3788710	-0.1541512	-0.0735860	-0.7821784	-0.2641657	0.8007755	2.0971177	2.2474201	0.1533599	0.8652859	0.9754070	1.4242862			
9	M	0.1102571	0.4687123	0.2048665	-0.0456793	0.6039479	1.5777090	0.8841348	1.0010230	1.6335557	-0.6328917	-0.4323917	-0.5388379	-0.4085467	-0.8044875	0.2706950	-0.1743821	0.2925170	-0.7193849	-0.0310487	0.2727200	1.1846873	0.0331088	0.9444458	0.2438456	1.3844473	2.3998936				
10	M	1.2248386	0.1723197	1.1981873	1.1930172	-0.0446888	0.7688074	0.7039351	-0.4146811	-0.7947821	1.6027801	1.1453855	1.3679504	1.0407388	1.2156454	0.0860898	0.6325410	1.2081158	0.3117139	0.1541693	1.0237892	0.2280620	0.8640591	0.9155017	0.0854880	-0.2705462	0.8347846	0.5271561			
11	M	1.2288887	1.2597186	1.3368593	1.1764719	0.7344853	1.6139142	1.8185770	1.9389122	1.3730125	-0.1058886	5.2323384	0.4474174	1.3382378	1.3423109	-0.2386889	0.4473077	0.4740359	-0.0248147	-0.0397891	0.2797350	1.4120183	1.3153513	1.5781847	1.3862659	0.7657438	1.1002871	1.6751075	1.0688972		
12	M	0.2284054	1.2796038	0.1209179	0.1002258	-0.1493187	1.0841558	-0.2194862	-0.0960267	-0.9507548	-1.0872906	0.2691902	1.7812121	0.3797891	1.0648718	-0.0517414	0.3910174	-0.4855228	1.2491229	-1.0378275	-0.7152423	0.2497803	-0.1037486	-0.3304519	-0.3291948	-1.6880515	-0.2479480	-1.2105648	-1.2880970		
13	M	-0.2481305	0.1790320	-0.2020485	-0.3320275	0.5722480	0.7545182	0.1524505	0.3270869	0.6077020	0.7059149	-0.1224449	-0.7208827	-0.2192333	-0.1940572	-0.6433881	-0.2323302	-0.2417462	-0.3348339	-0.3937101	-0.1732711	0.2228353	0.3472411	0.1668793	0.0412755	0.9389341	0.7388950	0.4852567	0.5303826		
14	M	-0.2785402	0.5213807	-0.2763559	-0.3381059	0.0813009	0.0132816	-0.0781104	0.0648827	-0.2337276	-0.1373240	-0.7538869	-1.1002849	-0.7620499	-0.6454321	-1.2086918	-0.6548331	-0.5889917	-0.7973370	-1.1384808	-0.7588803	-0.0142128	0.6517021	-0.0591668	-0.2523146	0.8196209	0.8742301	0.4866967	0.7042058		
15	M	-0.2785402	-0.1834336	-0.2344516	-0.3485575	1.3975471	0.3638807	0.4818013	0.6305507	1.1700381	0.6882791	-0.4202787	-0.9564881	-0.4481805	-0.3417733	-0.1889798	-0.1274735	-0.0882083	0.0885552	-0.3542150	-0.0559333	-0.1206304	0.3392071	-0.1384995	-0.2183414	0.2171878	1.0419840	1.1282771	1.4202755		

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 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*

Data Transformation ▾

Analytics ▾

Statistics ▾

DOE

Normalizers

Data Manipulation

Split

Variable Selection ▾

Best First

Stepwise

Regression Analysis

Genetic Algorithm

Boruta

Generalized Simulated Annealing

Particle Swarm Optimization

Recursive Feature Elimination

Successive Projections Algorithm

Lasso

MIVIF

Best First

?

×

Target Column

Col2 – diagnosis

Select Direction

forward

#Nodes before termination

2

Execute

Cancel

The results will appear on the output spreadsheet.

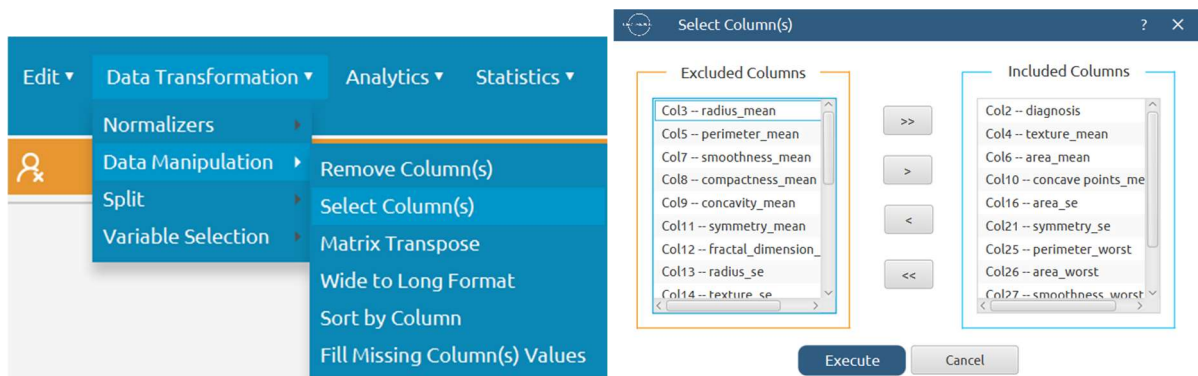
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (S)
User Header	User Row ID	texture_mean	area_mean	concave points_mean	area_se	symmetry_se	perimeter_worst	area_worst	smoothness_worst	concavity_worst	concave points_worst	diagnosis
1		-2.0360356	0.9365649	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	0.2586885	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	-0.8238466	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	-0.8317766	-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		0.2685297	0.3789878	1.3853591	0.2827663	-0.4305789	0.8718480	0.7603336	2.0844838	1.0140971	1.4093769	M
14		-1.1455337	-0.2621431	-0.0264407	-0.3561713	-0.0729511	-0.2317106	-0.3043312	0.5377135	-0.1589637	0.2214810	B
15		-0.8434790	-0.3898178	-0.4554491	-0.5416761	-0.4440975	-0.3390919	-0.4466274	-0.0295834	-0.4039638	-0.6254812	B

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.

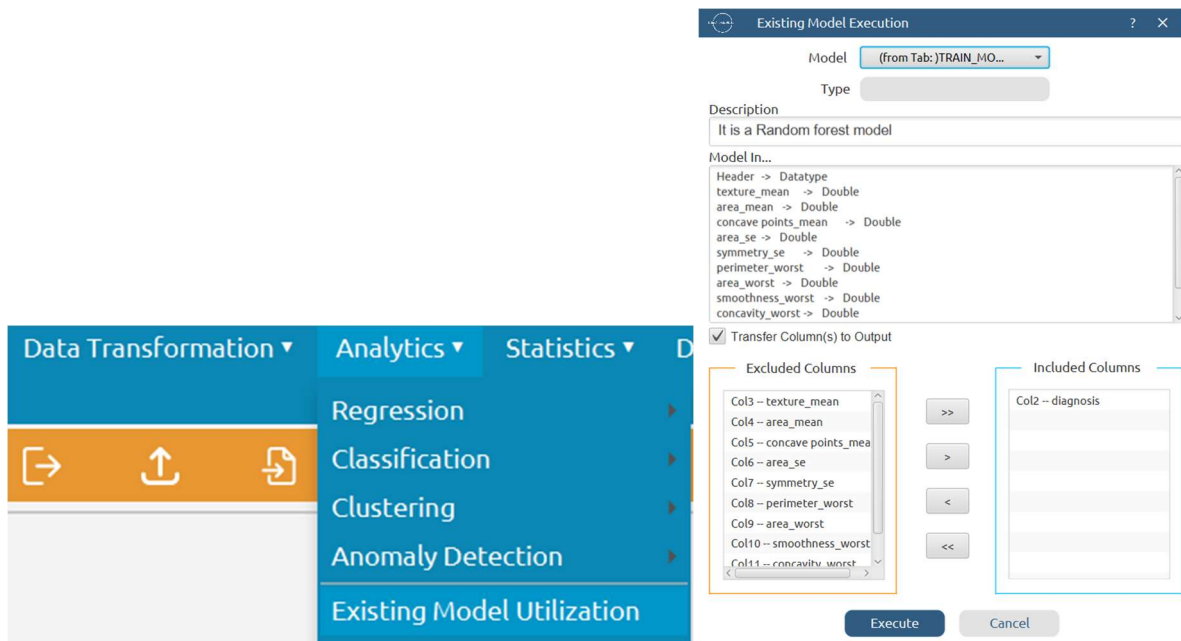
	Col1	Col2 (S)	Col3 (S)
User Header	User Row ID	diagnosis	Prediction
1		M	M
2		M	M
3		M	B
4		M	M
5		M	B
6		M	M
7		M	M
8		M	B
9		M	M
10		M	B
11		M	M
12		M	M
13		M	M
14		B	B
15		B	B

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.

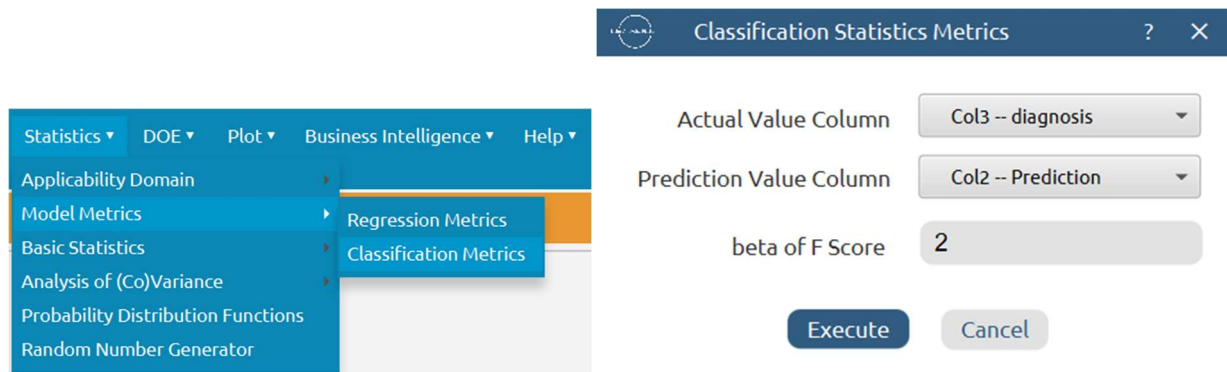
	Col1	Col2 (S)	Col3 (S)
User Header	User Row ID	Prediction	diagnosis
1		M	M
2		B	M
3		M	M
4		M	M
5		M	M
6		M	M
7		M	M
8		M	M
9		M	M
10		M	M
11		M	M
12		B	M
13		M	M
14		B	M
15		B	M

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 image shows the 'Classification Statistics Metrics' dialog box and the 'Statistics' menu. The dialog box has the following fields:

- Actual Value Column: Col3 -- diagnosis
- Prediction Value Column: Col2 -- Prediction
- beta of F Score: 2
- Buttons: Execute, Cancel

The 'Statistics' menu is open, showing the following options:

- Statistics ▾
- DOE ▾
- Plot ▾
- Business Intelligence ▾
- Help ▾
- Applicability Domain
- Model Metrics ▸
 - Regression Metrics
 - Classification Metrics
- Basic Statistics
- Analysis of (Co)Variance
- Probability Distribution Functions
- Random Number Generator

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		

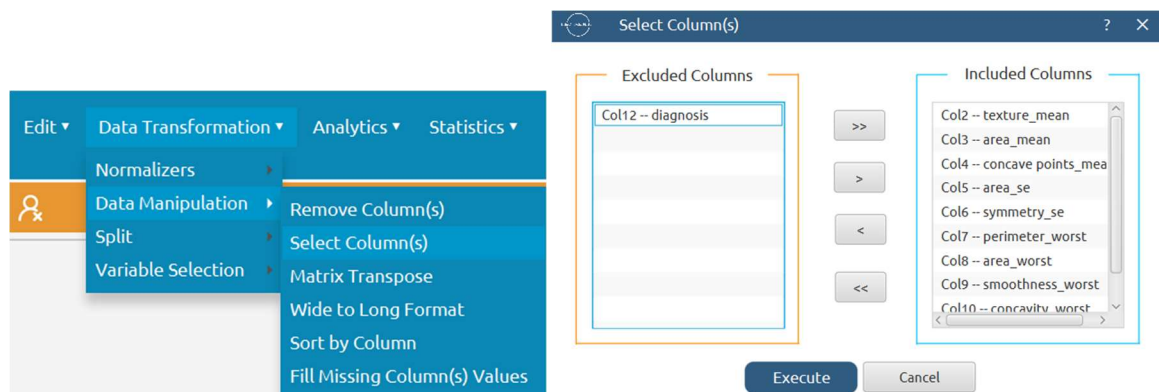
Step 11: Reliability check for each record of the test set

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)*

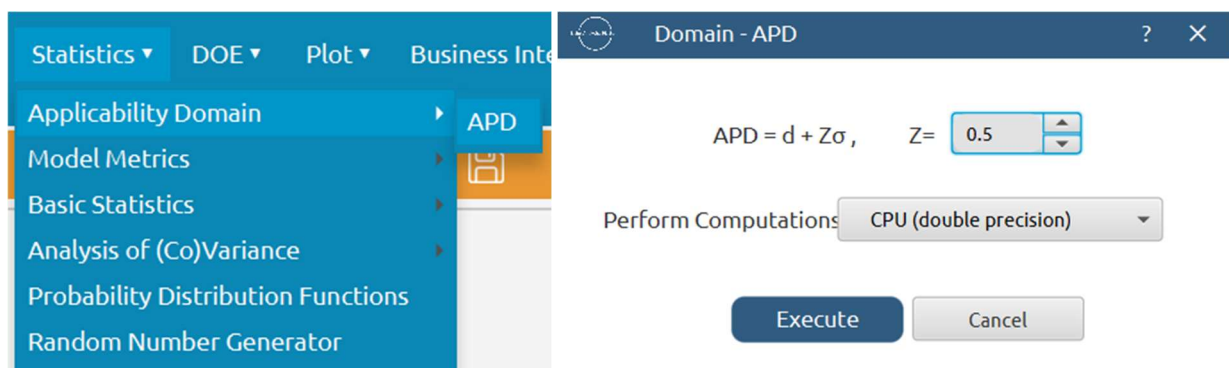


The results will appear on the output spreadsheet.

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.

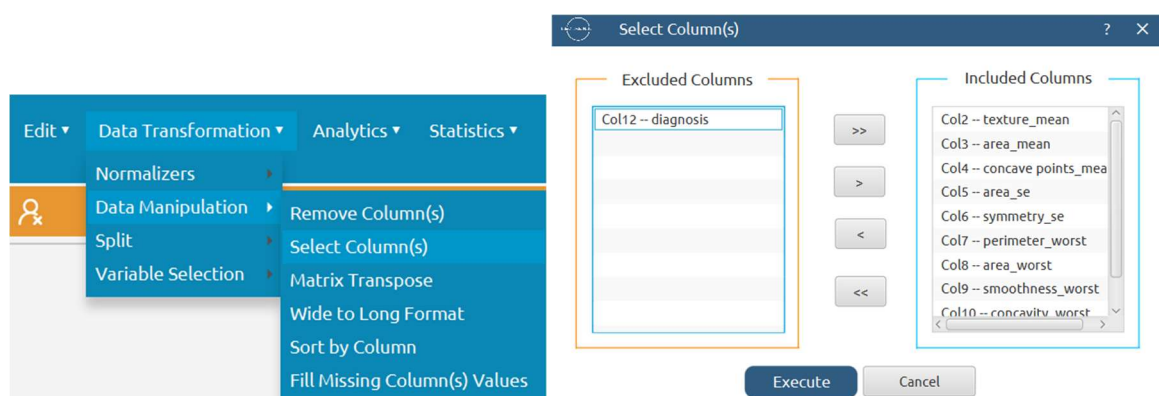
	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		0.0	2.9702124	reliable
2		0.0	2.9702124	reliable
3		0.0	2.9702124	reliable
4		0.0	2.9702124	reliable
5		0.0	2.9702124	reliable
6		0.0	2.9702124	reliable
7		0.0	2.9702124	reliable
8		0.0	2.9702124	reliable
9		0.0	2.9702124	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)*

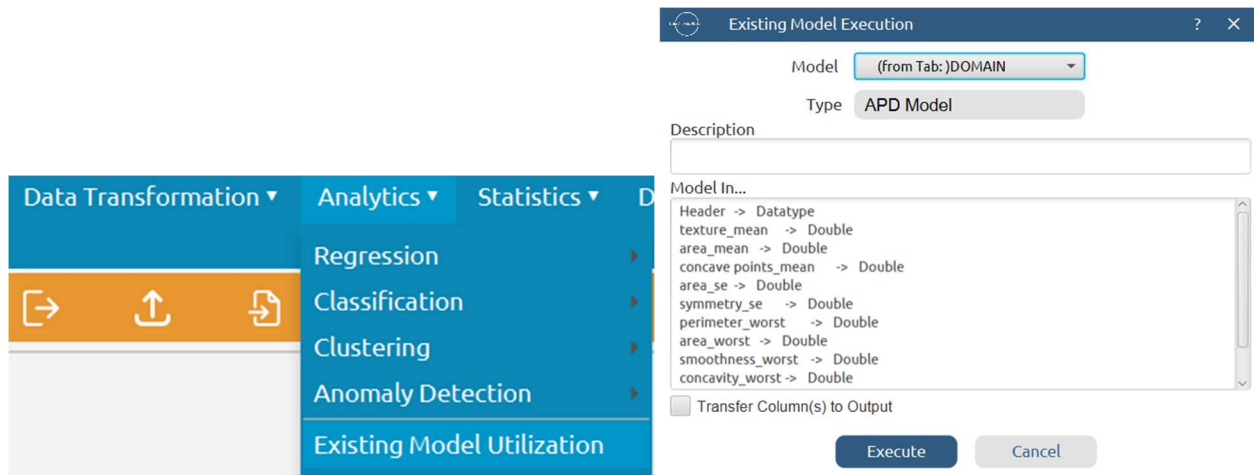


The results will appear on the output spreadsheet.

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:

