

# Breast Cancer (Binary Classification)

The goal of this study is to train a model in order to predict whether the cancer is benign (B) malignant (M). The dataset used in this case studv is found or in https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data\_and has 32 features and 569 labelled samples. The features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.

# Step 1: Import data from file

Right click on the input spreadsheet and choose the option "Import from file". Then navigate through your files to load the one with the breast cancer data.



1

# Step 2: Manipulate data

In our Dataset there are not empty values, and the only categorical feature is the label ("Diagnosis") which has two categories and the number of samples in each category are:

- Benign (B): 357
- Malignant (M): 212

In order to use the data for training we have to exclude any columns that do not contain features, like the "id" column. We follow these steps to execute this:

- On the menu click on "Data Transformation"  $\rightarrow$  "Data Manipulation"  $\rightarrow$  "Select Column(s)"
- Select all columns except the one that corresponds to the id.

<u> </u>		_								
File Edit	Data Transform	tion Anal	ytics Statistics	Plot	Help					
	Normalizers	•								
IMPORT	Data Manipulati	on 🔸 Remo	ve Column(s)							
	Split	Select	t Column(s)							
	Variable Selectio	n 🔸 Matri	x Transpose		_					
	Col1	C Sort b	y Column		(D)		Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
User Header	r User Row ID	id Fill M	issing Column(s)	Values	mean	texture_mean	perimeter_me	area_mean	smoothness_	compactnes
1		842302	м	17.99		10.38	122.8	1001	0.1184	0.2776
2		842517	м	20.57		17.77	132.9	1326	0.08474	0.07864
3		84300903	м	19.69		21.25	130	1203	0.1096	0.1599
4		84348301	м	11.42		20.38	77.58	386.1	0.1425	0.2839
		84358402	м	20.29		14.34	135.1	1297	0.1003	0.1328
6		843786	м	12.45		15.7	82.57	477.1	0.1278	0.17
7		844359	м	18.25		19.98	119.6	1040	0.09463	0.109
8		84458202	м	13.71		20.83	90.2	577.9	0.1189	0.1645
9		844981	м	13		21.82	87.5	519.8	0.1273	0.1932
10		84501001	м	12.46		24.04	83.97	475.9	0.1186	0.2396
11		845636	м	16.02		23.24	102.7	797.8	0.08206	0.06669
12		84610002	м	15.78		17.89	103.6	781	0.0971	0.1292
13		846226	м	19.17		24.8	132.4	1123	0.0974	0.2458
14		846381	M	15.85		23.95	103.7	782.7	0.08401	0.1002
15		84667401	м	13.73		22.61	93.6	578.3	0.1131	0.2293
16		84799002	M	14.54		27.54	96.73	658.8	0.1139	0.1595
17		848406	M	14.68		20.13	94.74	684.5	0.09867	0.072
18		84862001	M	16.13	-	20.68	108.1	798.8	0.117	0.2022
19		849014	м	19.81		22.15	130	1260	0.09831	0.1027
20		8510426	В	13.54		14.36	87.46	566.3	0.09779	0.08129
21		8510653	B	13.08		15.71	85.63	520	0.1075	0.127
22		8510824	B	9.504		12.44	60.34	273.9	0.1024	0.06492
23		8511133	м	15.34	_	14.26	102.5	704.4	0.1073	0.2135
	< (									>

The data without the "id" column will appear in the output spreadsheet.

# 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 to create 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".

TRAIN_TEST_SPLIT						
	Col1	Col2	Col3	Col4	Col5	Col6
User Header	User Row ID					
1						
2						
			Import	from SpreadSh	eet	
4			Import	from file		
5			import			
6			Export	Spread Sheet D	ata	
7			Clear S	preadSheet		
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

Split the dataset by choosing: "Data Transformation"  $\rightarrow$  "Split"  $\rightarrow$  "Random Partitioning". Then choose the "Training set percentage" and the column for the sampling as shown below:

Random Partitioning	×
Training set percentage	75
Usage of random generator seed	6615372866300
Stratified sampling	Col2 diagnosis
Execute	Cancel

The results will appear on the output spreadsheet.

														- 0	$\times$
File Edit D	ata Transforma	tion Analytic	ts Statistics	Plot Help											
	TRAIN, TEST, SPUT	NORMA ISE, TR	NN, SET												
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	с				Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6
User Header	User Row ID	diagnosis	radius_mean	texture_mean	perimeter_me	ar	Random Parti	tioning	User Header	User Row ID	diagnosis	radius_mean	texture_mean	perimeter_me	e area_r
1		м	17.99	10.38	122.8	10	Training set percentage	75	1		м	17.99	10.38	122.8	1001
2		м	20.57	17.77	132.9	13			2		M	20.57	17.77	132.9	1326
3		м	19.69	21.25	130	12	<ul> <li>Usage of random generator seed</li> </ul>	6615372866300	3		м	19.69	21.25	130	1203
4		м	11.42	20.38	77.58	38	/ Stratified compling		4		м	11.42	20.38	77.58	386.1
5		м	20.29	14.34	135.1	12	Stratiled sampling		5		м	18.25	19.98	119.6	1040
6		м	12.45	15.7	82.57	47			6		м	13.71	20.83	90.2	577.9
7		м	18.25	19.98	119.6	10			7		м	13	21.82	87.5	519.8
8		м	13.71	20.83	90.2	57	Reconfigure		8		м	16.02	23.24	102.7	797.8
9		м	13	21.82	87.5	51			9		м	19.17	24.8	132.4	1123
10		м	12.46	24.04	83.97	47			10		м	15.85	23.95	103.7	782.7
11		м	16.02	23.24	102.7	79			11		м	13.73	22.61	93.6	578.3
12		м	15.78	17.89	103.6	78			12		м	14.68	20.13	94.74	684.5
13		м	19.17	24.8	132.4	11			13		м	16.13	20.68	108.1	798.8
14		м	15.85	23.95	103.7	78			14		м	19.81	22.15	130	1260
15		м	13.73	22.61	93.6	57			15		В	13.54	14.36	87.46	566.3
16		м	14.54	27.54	96.73	65			16		В	13.08	15.71	85.63	520
17		м	14.68	20.13	94.74	68			17		В	9.504	12.44	60.34	273.9
18		м	16.13	20.68	108.1	75			18		м	15.34	14.26	102.5	704.4
	<(					>	< (	) >							>

# Step 4: Normalize the training set

Create a new tab by pressing the "+" button on the bottom of the page with the name "NORMALISE\_TRAIN\_SET".

Import data into the input spreadsheet of the "NORMALISE\_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"

IMPORT	INVERCIES LISPLE	eckwarte j	wingen				
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Co
User Header	User Row ID	diagnosis	radius_mean	texture_mean	perimeter_me	area_mean	sm î
1		м	17.99	10.38	122.8	1001	0.1
2		м	20.57	17.77	132.9	1326	0.0
3		м	19.69	21.25	130	1203	0.1
4		м	11.42	20.38	77.58	386.1	0.1
5		м	20.29	14.34	135.1	1297	0.1
6		м	12.45	15.7	82.57	477.1	0.1
7		м	13	21.82	87.5	519.8	0.1
8		м	16.02	23.24	102.7	797.8	0.0
9		м	15.78	17.89	103.6	781	0.0
10		м	19.17	24.8	132.4	1123	0.0
11		м	15.85	23.95	103.7	782.7	0.0
12		м	14.54	27.54	96.73	658.8	0.1
13		м	14.68	20.13	94.74	684.5	0.0
14		м	19.81	22.15	130	1260	0.0
15		в	13.54	14.36	87.46	566.3	0.0
16		в	13.08	15.71	85.63	520	0.1
17		м	15.34	14.26	102.5	704.4	0.1
18		м	17.14	16.4	116	912.7	0.1
19		м	14.58	21.53	97.41	644.8	0.1
			1757	15.05	115	955.1	0.0
20		M	11.31	10.00	1112	a second s	

Normalize the data using Z-score by browsing: "Data Transformation"  $\rightarrow$  "Normalizers"  $\rightarrow$  "Z-Score". Then select all columns and click "Execute".

Col6 (D) Cc			
me 🗧 🔳 ZScore Normalizer			× ead
Excluded Columns	·	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 ~	
e e	Execute	Cancel	
			45

The results will appear on the output spreadsheet.

IMPORT	TRAIN_TEST_SPLIT	NDRWALSE, TR	NOL SET											
	Col1	Col2 (S)	Col3 (D)	Col4 (I	7900	ve Norm	olizor		Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)
User Header	User Row ID	diagnosis	radius_mean	texture	230	ore North	alizei	User Header	User Row ID	diagnosis	radius_mean	texture_mean	perimeter_me	area_mean
1		M	17.99	10.38	Excluded Columns	1 1	Included Columns	1		м	1.1414062322 03168	-2.103062476 3416603	1.3105995227 949832	1.0289232935 50912
2		M	20.37	01.07				2		м	1.8911172683	-0.331678704	1.7343491516	1.9749845091 -
4		M	11.42	20.38			Col4 texture_mean	3		м	841118 1.6354018761 08264	28159096 0.5024776592 966151	506528 1.6126784661 276299	1.6169367260 0
5		M	20.29	14.34			Col5 perimeter_mean Col6 area_mean	4		м	-0.767741639 001792	0.2939385683 945634	-0.586624063	-0.761024526 3 4177266 3
7		M	12.45	21.02			Col7 smoothness_mean	5		м	1.8097532799	-1.153850062	1.8266510510	1.8905667391
,		M	16.02	22.34			Col8 compactness_mean				613734 -0.468438395	6261157	053526	692518 8
0			16.70	17.00			Col9 concavity_mean	6		м	8752916	5419894	1321035	0418181
,		M	13.70	17.07			Col10 concave points_me	7		м	-0.308616275	0.6391067188	-0.170426407	-0.371829497
10		M	19.17	24.8							7591988	3658	7599795	09619967
11		м	15.85	23.95				8		м	146185	891234	634041	6426796
12		м	14.54	27.54				9		м	0.4992118950	-0.302914691	0.5050556738	0.3885126245
13		м	14.68	20.13							094144	744756	812356	1025426 8
14		м	19.81	22.15		Reconfigure		10		м	70422	67975	518573	098223
15		В	13.54	14.36				11		м	0.5195528921	1.1496679413	0.5092512147	0.3934612524
16		в	13.08	15.71							150989	653954	609951 0.2168220154	073867 3
17		м	15.34	14.26				12		м	6586013	256997	4178575	34034144 1
18		м	17.14	16.4				13		м	0.1795676547	0.2340135422	0.1333307519	0.1076052174 0
19		м	14.58	21.53							1.6702721569	0.7182077533	1.6126784661	1.7828613084 (
20		м	17.57	15.05				14		м	509653	12875	376388	669593
21		м	18.63	25.11				40	e (		-0.151700012	-1.149056060	-0.172104624	-0.236469969
MPORT TRAIN	TEST_SPLIT	NORMALISE_TR	RAIN_SET +											

# Step 5: Normalize the test set

Create a new tab by pressing the "+" button on the bottom of the page with the name "NORMALISE\_TEST\_SET".

Import data into the input spreadsheet of the "NORMALISE\_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".

File Edit	Data Transforma	tion Analytic	s Statistics	Plot Help													-	
IMPORT	TRAIN_TEST_SPL	NORMALISE TR NORMALISE TR	NN, 5(1 151, 921															
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)			Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
User Header	User Row ID	diagnosis	radius_mean	texture_mean	perimeter_me an	area_mean	smoothness_ mean	compactness mean	â	User Header	User Row ID							
1		м	20.29	14.34	135.1	1297	0.1003	0.1328	U	1								
2		м	12.45	15.7	82.57	477.1	0.1278	0.17										
3		м	12.46	24.04	83.97	475.9	0.1186	0.2396		2								
4		м	15.78	17.89	103.6	781	0.0971	0.1292		3								
5		м	14.54	27.54	96.73	658.8	0.1139	0.1595		4								
6		м	21.16	23.04	137.2	1404	0.09428	0.1022		•								
7		м	18.61	20.25	122.1	1094	0.0944	0.1066		5								
8		м	19.27	26.47	127.9	1162	0.09401	0.1719		6								
9		м	16.13	17.88	107	807.2	0.104	0.1559		_								
10		м	14.25	21.72	93.63	633	0.09823	0.1098		1								
11		В	13.03	18.42	82.61	523.8	0.08983	0.03766		8								
12		м	14.99	25.2	95.54	698.8	0.09387	0.05131		9								
13		м	13.44	21.58	86.18	563	0.08162	0.06031										
14		м	13.17	21.81	85.42	531.5	0.09714	0.1047		10								
15		В	13.49	22.3	86.91	561	0.08752	0.07698		11								
16		В	11.76	21.6	74.72	427.9	0.08637	0.04966		42								
17		В	13.64	16.34	87.21	571.8	0.07685	0.06059		12								
18		В	11.94	18.24	75.71	437.6	0.08261	0.04751	~	13								

Normalize the test set using the existing normalizer of the training set: "Analytics"  $\rightarrow$  "Existing Model Utilization"  $\rightarrow$  "Model (from Tab:) NORMALISE\_TRAIN\_SET".

File Edit	Data Transforma	tion Analyti	cs Statistics	Plot Help				
IMPORT _	TRAIN_TEST_SPUT	Regress Classific Clusteri Anomal Existing	sion cation ng ly Detection Model Utilizati	) ) on				
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	diagnosis	radius_mean	texture_mean	perimeter_me an	area_mean	smoothness_ mean	compactness, 2 mean
1		м	20.29	14.34	135.1	1297	0.1003	0.1328
2		м	12.45	15.7	82.57	477.1	0.1278	0.17
3		м	12.46	24.04	83.97	475.9	0.1186	0.2396
4		м	15.78	17.89	103.6	781	0.0971	0.1292
5		м	14.54	27.54	96.73	658.8	0.1139	0.1595
6		м	21.16	23.04	137.2	1404	0.09428	0.1022
7		м	18.61	20.25	122.1	1094	0.0944	0.1066
8		м	19.27	26.47	127.9	1162	0.09401	0.1719
9		м	16.13	17.88	107	807.2	0.104	0.1559
10		м	14.25	21.72	93.63	633	0.09823	0.1098
11		В	13.03	18.42	82.61	523.8	0.08983	0.03766
12		м	14.99	25.2	95.54	698.8	0.09387	0.05131
13		м	13.44	21.58	86.18	563	0.08162	0.06031
14		м	13.17	21.81	85.42	531.5	0.09714	0.1047
15		В	13.49	22.3	86.91	561	0.08752	0.07698
16		В	11.76	21.6	74.72	427.9	0.08637	0.04966
17		В	13.64	16.34	87.21	571.8	0.07685	0.06059
18		В	11.94	18.24	75.71	437.6	0.08261	0.04751
18	<.	В	11.94	18.24	75.71	437.6	0.08261	0.04751

Existing Model Executi	on	>
Mode	el (from Tab: )NORMALISE_TR 👻	
Ту	pe Z Score Normalizer Model	
Description		
Vodel Input		
Header -> Datatype		
radius_mean -> Dou	ble	
perimeter mean -> Dou	Double	
area_mean -> Dou	ble	
smoothness_mean ->	Double	
compactness_mean ->	Double	
concevity mean	Double	
concavity_mean ->	- 2 1 1/11/11/4	
concave points_mean	Double	
concave points_mean Transfer Column(s)	Double to Output	
Concave points_mean	to Output	

The results will appear on the output spreadsheet.

																	-	0	×
File Edit (	ata Transforma	tion Analytic	s Statistics	Plot Help															
IMPORT	TRAIN, TEST, SPUT	NORMALISE, TR	NN 587 51,987																
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)			Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	
User Header	User Row ID	diagnosis	radius_mean	texture_mean	perimeter_me	area_mean	smoothness_ mean	compactness mean	â	User Header	User Row ID	diagnosis	radius_mean	texture_mean	perimeter_me	area_mean	smoothness_	compactness_	cor
1		м	20.29	14.34	135.1	1297	0.1003	0.1328	111				1.6544727897	-1.159126809	1.6746400856	1.7050626433	0.2624790087	0.5063876786	1.3
2		м	12.45	15.7	82.57	477.1	0.1278	0.17	-111			m	423337	2511253	237293	426805	0224555	651897	134
1		м	12.46	24.04	92.07	475.9	0.1196	0.2296	- 11	2		м	-0.488001579	-0.840771103	-0.401173408	-0.514131135	2.1695790672	1.1951997060	0.8
		M	15 70	17.00	102.6	701	0.0071	0.1202	-				-0.485268831	1.1114984449	-0.345849995	-0.517379132	1.5315674113	2.4839447895	1.6
-		M	13.70	17.07	105.0	/01	0.0971	0.1252	- 11	3		M	94744735	532763	3294507	4931154	25884	738505	540
5		м	14.54	27.54	96.73	658.8	0.1139	0.1595	-	4		м	0.4220034776	-0.328124783	0.4298632901	0.3084240410	0.0405619109	0.4397284502	0.1
6		м	21.16	23.04	137.2	1404	0.09428	0.1022	- 11				0.0831427355	1.9307962170	0.1583833986	-0.022330294	1.2056266740	1.0007769563	0.9
7		м	18.61	20.25	122.1	1094	0.0944	0.1066		5		м	0015698	88885	2639889	592290908	400317	92355	657
8		м	19.27	26.47	127.9	1162	0.09401	0.1719		6		м	1.8922218588	0.8774133672	1.7576252053	1.9946756868	-0.155002531	-0.060215763	0.2
9		м	16.13	17.88	107	807.2	0.104	0.1559					189/63	022451	186572	385154	39580372	224370306	32
10		м	14.25	21.72	93.63	633	0.09823	0.1098		7		м	11576	6928855	884596	879654	2310589	12690638	461
11		В	13.03	18.42	82.61	523.8	0.08983	0.03766		8		м	1.3757325018	1.6803251838	1.3901196752	1.3396630090	-0.173726786	1.2303809655	0.9
12		м	14.99	25.2	95.54	698.8	0.09387	0.05131					593738	931417	41119	255055	5164808	24072	808
13		м	13,44	21.58	86.18	563	0.08162	0.06031		9		м	471241	0053954	582618	6081137	485547	347598	425
14		м	13.17	21.81	85.42	531.5	0.09714	0 1047	-	10		м	0.0038930458	0.5684210645	0.0358815552	-0.092162224	0.1189263861	0.0805092746	0.5
15		D	12.40	22.2	96.01	561	0.09752	0.07699	-				0794308	66244	67218826	70623979	1039118	3055303	524
15		0	11.76	21.6	74.72	407.0	0.00637	0.07090	-	11		В	4834407	01875798	51283275	90946595	42177117	8502594	451
16		5	11.76	21.0	14.12	427.9	0.08637	0.04966	-	12		м	0.2061163919	1.3830371351	0.1113584974	0.0859362637	-0.183435659	-1.002518023	-0.8
17		в	13.64	16.34	87.21	571.8	0.07685	0.06059	-	12			1911024	467923	6593966	2390911	5420169	2818771	543
18		В	11.94	18.24	75.71	437.6	0.08261	0.04751	~	13		м	-0.217459535	0.5356491536	-0.258518036	-0.281628701	-1.032962049	-0.835869952	-0.1
											12								>

IMPORT TRAIN\_TEST\_SPLIT NORMALISE\_TRAIN\_SET NORMALISE\_TEST\_SET +

# Step 6: Feature selection

Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE\_SELECTION\_BEST\_FIRST".

Import data into the input spreadsheet of the "FEATURE\_SELECTION\_BEST\_FIRST" tab from the output of the "NORMALISE\_TRAIN\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

																	-	0	×
File Edit D	ata Transforma	tion Analyti	cs Statistics	Plot Help															
	TRAIN_TEST_SPUT	NORMALISE, TR	RAN LEET	COULELA															
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	c		Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col	18
User Header	User Row ID	diagnosis	radius_mean	texture_mean	perimeter_me an	area_mean	smoothness_ mean	compactness_ mean	.α n	User Header	User Row ID								î
1		м	1.0259407680	-2.086103717 1531285	1.1885843845 534358	0.9038901118	1.5176975927	3.1875699788 484675	2. 76	1									
2		м	1.7309897315	-0.356214992	1.5877032935	1.7835558981	-0.816592878	-0.496463380	-0	2									
3		м	1.4905079145	0.4584010780	1.4731047949	1.4506362312	0.9074255739	1.0081835373	1.	3									
4		м	-0.769474615	0.2547470603	-0.598361859	-0.760437555	3.1890107349	3.3042236286	1.	4									
5		м	1.0969922139	0.1611130292	1.0621308688	1.0094500061	-0.130730348	49259	0.	5									-
6		м	764139 -0.143675341	7597024 0.3600853453	278308 -0.099660806	610954 -0.241299408	8319634 1.5523721392	510875 1.0933592181	24 0.	6									
7		м	89480025 -0.337700444	660462 0.5918295723	90116359 -0.206355960	78680536 -0.398556584	377474 2.1349045207	390153 1.6247814005	20	7									-
		M	244704 0.4875894277	415474 0.9242303827	7946428 0.3942982389	74108596 0.3538959955	699087 -1.002448448	648444 -0.717734986	0€ -C										
-		m 	224913 1.3484050226	508512 1,2894031040	019804	5650414 1.2341031146	339021 0.0613666388	1491507 2.5987467941	7:	8									-
9		м	551627 0.4411327130	455802	302507 0.4338149625	672103 0.3130253697	8757059	39709	9(	9									
10		м	753313	555035	662319	921389	9119114	9230344	12	10									-
11	< (	м	-0.138209846	0.7767567837	0.0346960535	-0.240216743	1.1501473996	2.2932253303	<mark>1.</mark> ∼ >	**	< (								> ×

Choose the most important features for the classification using the Best First Function by browsing: "Data Transformation"  $\rightarrow$  "Variable Selection"  $\rightarrow$  "Best First". Then choose the "diagnosis" column as the target variable and the direction as forward.

Best First	×
Target Column	Col2 diagnosis 🔹
Select Direction	forward •
#Nodes before termination	Integer (0,+∞), Default: 2
Execute	Cancel

The results will appear on the output spreadsheet.

k.

🕘 bc.ekk																- 0	) ×
File Edit (	Data Transforma	tion Analytic	s Statistics	Plot Help													_
IMPORT	TRAIN_TEST_SPLIT	NCRIVALISE, TRA NCRIVALISE, TRA	NN 527 - PRIVE, SL 51, 527	LCCN.JET.JRC													
	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11			Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
User Header	texture_mean	perimeter_me an	area_mean	smoothness_ mean	compactness_ mean	concavity_me	a concave	symm ^		User Header	User Row ID	texture_mean	concave	fractal_diment	area_se	radius_worst	perimeter.
1	-2.086103717 1531285	1.1885843845 534358	0.9038901118 028003	1.5176975927 179759	3.1875699788 484675	2.5892476816 768712	2.3793395976 20931	2.2256 89504	Target (	1		-2.086103717 1531285	2.3793395976 20931	2.3607370906 60363	2.2227153166 031446	1.7948318609 964493	2.1783380 462767
2	-0.356214992 5582287	1.5877032935 623765	1.7835558981 219255	-0.816592878 9930453	-0.496463380 57463285	-0.036997120 22566861	0.4936137407 6952486	-0.009 73900	Select Di	2		-0.356214992 5582287	0.4936137407 6952486	-0.899012720 6795266	0.6342849827 997221	1.7167157861 821813	1.4404918 822503
3	0.4584010780	1.4731047949	1.4506362312	0.9074255739	1.0081835373	1.3241663441	1.9087047808	0.9367	#Nodes before term	3		0.4584010780	1.9087047808	-0.407979627	1.0337956378	1.4322931548	1.2603200
4	0.2547470603	-0.598361859	-0.760437555	3.1890107349	3.3042236286	1.8661680856	1.3522771590	2.8811		4		0.2547470603	1.3522771590	5.1309328241	-0.303914224	-0.302284301	-0.273427
	7718237	544018 1.0621308688	9129843	709633	49259	558627	134446	66397		-		7718237	134446	62909	6976347	3250492	79837576
5	7597024	278308	610954	8319634	510875	24451	164621	60970		5		7597024	164621	1121528	556092	742195	43546
6	0.3600853453	-0.099660806	-0.241299408	1.5523721392	1.0933592181	0.0462740564	0.2406475267	1.4044		6		0.3600853453	0.2406475267	1.7395506474	0.1712931860	0.1283555982	0.0620350 6340836
7	0.5918295723	-0.206355960	-0.398556584	2.1349045207	1.6247814005	1.1825067981	1.0662194344	1.9715		7		0.5918295723	1.0662194344	1.6478516963	-0.362188711	-0.186111677	-0.063799
	415474	7946428	74108596	699087	648444	061863	657272	51831				415474	657272	473742	2252187	24229186	99557276
8	508512	019804	5650414	339021	1491507	7536514	8476715	81000		8		508512	8476715	4525769	92570521	546085	4035256
9	1.2894031040	1.5679449317	1.2341031146	0.0613666388	2.5987467941	1.4362621588	1.5140578772	2.1446		9		1.2894031040	1.5140578772	2.2557270014	1.4777631176	0.9095163464	1.2374411
10	1.0904307879	0.4338149625	0.3130253697	-0.867217716	-0.097248667	0.1167342828	0.0884265782	0.1192		10		1.0904307879	0.0884265782	-1.385608767	-0.116674895	0.0842901201	0.1020732
10	555035	662319	921389	9119114	9230344	1252875	3535445	87352		10		555035	3535445	8350728	13306085	9171221	7308441
11	664419	5729124	20364345	083975	757304	933265	562423	2595		11		0.7767567837 664419	562423	30581	-0.464519510	99758234	93825156
12	0.1962257909	0.0797451185	0.0472309691	0.1494399870	-0.619412624	-0.196518087	0.0626887366	-0.841		12		0.1962257909	0.0626887366	-0.521863808	0.0599508681	0.5309538300	0.4280982
	389246	3453799	0.3566026595	4778985	1/419/6	3205931	9507133	1.2866				389246	9507133	7504553	56729696	271414	804454
13	0309184	889382	144091	626156	088324	63293	21369	54629		13		0309184	21369	977301	767253	347472	873434
14	0.6690776480	1.4731047949 360467	1.6049160769 001953	0.1244743135	-0.050957537 04970428	0.7144143850	1.1017621680 21356	-0.856		14		0.6690776480 000473	1.1017621680	-1.301304893 4038687	1.4016658500	2.1834092587 905	2.2412551 75768
	< (							>			<	)					>

#### Step 7: Feature selection: test set

Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE\_SELECTION\_TEST\_SET".

Import data into the input spreadsheet of the "FEATURE\_SELECTION\_TEST\_SET" tab from the output of the "NORMALISE\_TEST\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Manipulate the data by choosing the columns that correspond to the significant features (from the previous step): "Data Transformation"  $\rightarrow$  "Data Manipulation"  $\rightarrow$  "Select Column(s)".

ilo Edit		_					
- Lun	Data Transforma	tion Analyti	ics Statistics	Plot Help			
	Normalizers						
IMPORT		n 🕨 Remove	e Column(s)				
	Split	Select C	Column(s)				
	Variable Selection	Matrix 1	Transpose				
		Sort by	Column				
		Fill Mice	sing Column(s) V	aluer			
	Coll	Col2 (5)	Col3 (D)	Cold (D)	Col5 (D)	Col6 (D)	C.
	con	COI2 (3)	COIS (D)	C014 (D)	COIS (D)	C010 (D)	C.C.
					perimeter me		sm ^
Jser Header	User Row ID	diagnosis	radius_mean	texture_mean	perimeter_me an	area_mean	sm 🏫 me
User Header 1	User Row ID	diagnosis M	radius_mean 1.6544727897 423337	texture_mean	perimeter_me an 1.6746400856 237293	area_mean 1.7050626433 426805	sm me 0.2
Jser Header 1 2	User Row ID	diagnosis M	radius_mean 1.6544727897 423337 -0.488001579	texture_mean -1.159126809 2511253 -0.840771103	perimeter_me an 1.6746400856 237293 -0.401173408	area_mean 1.7050626433 426805 -0.514131135	sm me 0.2 02. 2.1
Jser Header 1 2	User Row ID	diagnosis M M	radius_mean 1.6544727897 423337 -0.488001579 86786896 0.465268931	texture_mean -1.159126809 2511253 -0.840771103 5070031	perimeter_me an 1.6746400856 237293 -0.401173408 4594031 -0.245840005	area_mean 1.7050626433 426805 -0.514131135 7436293 -0.517370132	sm me 0.2 02. 2.1 899
User Header 1 2 3	User Row ID	diagnosis M M M	radius_mean 1.6544727897 423337 -0.488001579 86786896 -0.485268831 94744735	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763	erimeter_me an 1.6746400856 237293 -0.401173408 4594031 -0.345849995 3294507	area_mean 1.7050626433 426805 -0.514131135 7436293 -0.517379132 4931154	sm () () () () () () () () () ()
Jser Header 1 2 3 4	User Row ID	diagnosis M M M M	radius_mean 1.6544727897 423337 -0.488001579 86786896 -0.485268831 94744735 0.4220034776 2222	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783	perimeter_me           an           1.6746400856           237293           -0.401173408           4594031           -0.345849995           3294507           0.4298632901	area_mean 1.7050626433 426805 -0.514131135 7436293 -0.517379132 4931154 0.3084240410 c32005	sm ^ 0.2 02. 2.1 899 1.5 25: 0.0
Jser Header 1 2 3 4	User Row ID	diagnosis M M M M M	radius_mean 1.6544727897 423337 -0.488001579 86786896 -0.485268831 94744735 0.4220034776 32383 0.0831427355	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.9307962170	perimeter_me an 1.6746400856 237293 -0.401173408 4594031 -0.345849995 3294507 0.4298632901 9980646 0.1583833986	area_mean 1.7050626433 426805 -0.514131135 7436293 -0.517379132 4931154 0.3084240410 6370025 -0.022330294	sm 0.2 0.2 2.1 899 1.5 25; 0.0 75 1.2
User Header 1 2 3 4 5	User Row ID	diagnosis M M M M M M	radius_mean 1.6544727897 423337 -0.488001579 86786896 -0.485268831 94744735 0.4220034776 32383 0.0831427355 0015698	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.9307962170 88885	perimeter_me an           1.6746400856           237293           -0.401173408           4594031           0.345849995           3294507           0.4298632901           9980646           0.1583833986           2639889	area_mean 1.7050626433 426805 0.514131135 7436293 -0.517379132 4931154 0.3084240410 6370025 -0.022330294 592290908	sm 0.2 0.2 0.2 2.1 89( 1.5 25; 0.0 75 1.2 40(
Jser Header 1 2 3 4 5 6	User Row ID	diagnosis M M M M M M M	radius_mean 1.6544727897 423337 -0.488001579 86786896 -0.485268831 94744735 0.4220034776 32383 0.0831427355 0015698 1.8922218588	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.9307962170 88885 0.8774133672 0.02451	perimeter_me an 1.6746400856 237293 -0.401173408 4594031 -0.345849995 3294507 0.4298632901 9980646 0.1583833986 2639889 1.7576252053	area_mean 1.7050626433 426805 -0.514131135 7436293 -0.517379132 4931154 0.3084240410 6370025 -0.022330294 592290908 1.994675688 1.994675688	sm 0.2 02. 2.1 899 1.5 25: 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.
Jser Header 1 2 3 4 5 6 7	User Row ID	diagnosis M M M M M M M	radius_mean 1.6544727897 423337 -0.488001579 86786896 -0.485268831 94744735 -0.425268831 94744735 -0.4220034776 32383 -0.0831427355 -0.015698 1.892218588 1.89763 1.1953711391	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.9307962170 88885 0.8774133672 002451 0.22443160002	perimeter_me an 1.6746400856 237293 -0.401173408 4594031 -0.345849995 3294507 0.429863200 0.429863200 0.1583833986 2639889 1.7576252053 186572 1.1609226779	area_mean 1.7050626433 426805 -0.514131135 7436293 -0.517379132 4931154 0.3084240410 6370025 -0.022330294 592290908 1.9946756668 385154 1.1556098598	sm ( 0.2 02. 2.1 899 1.5 25: 0.0 75 1.2 400 -0, 399 -0,
Jser Header 1 2 3 4 5 6 7	User Row ID	diagnosis M M M M M M M M	radius_mean 1.6544727897 423337 -0.488001579 86786896 -0.485268831 94744735 0.4220024776 22383 0.0831427355 0015698 1.8922218588 1.8952218588 1.8952711391 1.1953711391	texture_mean -1.159126809 2511253 -0.440771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.9307962170 88885 0.8774133672 0.2243160002 6928855	perimeter_me an 1.6746400856 237293 -0.401173408 4594031 -0.345849995 3294507 0.4298632901 9980646 0.158383986 1.7576252053 186572 1.1609226779 884596	area_mean 1.7050626433 426805 -0.514131135 7436293 -0.517379132 4931154 0.3084240410 6370025 -0.022330294 592290908 1.9946756868 385154 1.1556098598 879654	sm me 0.2 0.2 2.1 1.5 25; 0.0 75; 1.2 400 -0, 39; -0, 23 0
Jser Header 1 2 3 4 5 6 7 8	User Row ID	diagnosis M M M M M M M M M M M	radius_mean 1.6544727897 423337 -0.488001579 86786896 -0.485268831 94744735 0.4220034776 22383 0.0831427355 0015698 1.8922218588 1.89763 1.1953711391 11576 1.3757325018 593738	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.9307962170 88885 0.8774133672 002451 0.2243160002 6928855 1.6803251838 931417	perimeter_me an 1.6746400856 237293 -0.401173408 4594031 0.4298632901 9980646 0.1583833986 2639889 1.7576252053 186572 1.1609226779 884596 1.3901196752	area_mean 1.7050626433 426805 -0.514131135 7436293 -0.517379132 4931154 0.3084240410 6370025 -0.022330294 592290908 1.9946756868 385154 1.1356098598 879654 1.3396630090 555055	970 0.2 0.2 2.1 889 2.5 1.5 2.5 2.5 2.5 1.5 2.5 1.5 2.5 1.5 2.5 3.0 0.0 7.5 7.5 1.2 400 0.0 2.3 3.9 9.0 0.2 2.3 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0
User Header 1 2 3 4 5 6 7 8 8	User Row ID	diagnosis M M M M M M M M M M M M	radius_mean 1.6544727897 423337 -0.488001579 86766896 -0.485268831 94744735 -0.4220034776 32383 0.0831427355 0015698 1.892218588 189763 1.953711391 11576 1.3757325018 593753 0.5176496548	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.930762170 88885 0.8774133672 002451 0.0243160002 6928855 1.6803251838 931417 -0.330465634	perimeter_me an 1.6746400856 237293 -0.401173408 4594031 -0.345849995 3294507 0.4298632901 0.4298632901 0.1533833986 2639889 1.7576252053 186572 1.1609226779 884596 1.3901196752 41119 0.5642201506	area_mean 1.7050626433 426805 -0.514131135 7436293 -0.517379132 4931154 -0.3084240410 6370025 -0.022330294 592290908 385154 1.356008598 879654 1.3396630090 255055 0.3793386637	am (1) 102 022 022 2.1 2.5 2.5 2.5 2.5 0.0 7.5 1.2 0.0 0.0 7.5 1.2 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0
User Header 1 2 3 4 5 6 7 8 8 9	User Row ID	diagnosis M M M M M M M M M M M	radius_mean 1.6544727897 423337 -0.488001579 86766896 0.485268831 94744735 0.422034776 32383 0.0831427355 0015698 1.9922218588 1.9922318588 1.993711391 1.576 1.3757325018 593738 0.5176496548 471241 0.033924459	texture_mean -1.159126809 2511233 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 -0.328124783 2278647 -0.328124783 2278647 -0.33046555 1.6803251838 931417 -0.330465551838 931417	perimeter_me an 1.6746400856 237293 -0.401173404 4594031 -0.345849995 3294507 0.429863201 9980646 2639869 2639869 1.7576252053 1.66572 1.1609226779 884596 1.3901196752 41119 0.564220150 582618	area_mean 1.7050626433 426805 -0.514131135 7436293 -0.517379132 4931154 0.3084240410 6370025 -0.022330294 592290908 1.9946756868 395154 1.1556096598 879654 1.3396630090 255055 0.3793386367 6081137 -0.027333024 -0.0739386367 6081137 -0.00255734 -0.0025574 -0.0025734 -0.0025734 -0.0025734 -0.0025734 -0.0025734 -0.002574 -0.005574 -0.002574 -0.002574 -0.002574 -0.0057474 -0.005574 -0.0057474 -0.005574 -0.00	sm 6 mi 2.1 2.1 2.2 2.1 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5

The results will appear on the output spreadsheet.

File Edit I	Data Transformat	tion Analytic	s Statistics	Plot Help														
IMPORT	TRAIN_TEST_SPUT	NCRIMALISE, TRA	NIN SET - BRINE SEL ST. SET - BRINE SE	500/19/19														
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col		Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	diagnosis	radius_mean	texture_mean	perimeter_me an	area_mean	smoothness_ mean	compactness_ mean	con 🏫	User Header	User Row ID	diagnosis	texture_mean	concave points mean	fractal_dimens	area_se	radius_worst	perimeter_w
1		м	1.6544727897	-1.159126809	1.6746400856	1.7050626433	0.2624790087	0.5063876786	1.33	1		м	-1.159126809	1.3302161519	-0.579545407	1.0420061325	1.2259865982	1.251740490
2		м	-0.488001579	-0.840771103	-0.401173408	-0.514131135	2.1695790672	1.1951997060	0.83	2		м	-0.840771103	0.7563848467	1.9791511327	-0.304715248	-0.190117629	-0.14387551
3		м	-0.485268831	1.1114984449	-0.345849995	-0.517379132	1.5315674113	2.4839447895	1.69	3		м	1.1114984449	0.8676703711	2.9109307974	-0.369798437	-0.266230728	-0.30831797
4		м	94744735 0.4220034776	-0.328124783	3294507 0.4298632901	4931154 0.3084240410	25884 0.0405619109	0.4397284502	0.11	4		м	532763 -0.328124783	549723 0.3928684753	74537 -0.285221354	98827083 0.2353750956	3338487 0.8013556274	9510933 0.802741079
			32383 0.0831427355	2278647	9980646 0.1583833986	6370025 -0.022330294	7570802	0759466	360	-		IVI	2278647	118454	2067252	4182776	611457 0.2084746493	924124
5		м	0015698	88885	2639889	592290908	400317	92355	657	5		м	88885	026507	870997	01490397	8362542	35283
6		м	189763	002451	186572	385154	39580372	224370306	325	6		м	002451	891167	2889713	901676	2.5559598700 8895	554897
7		м	1.1953711391 11576	0.2243160002 6928855	1.1609226779 884596	1.1556098598 879654	-0.146680640 2310589	0.0212566271 12690638	0.72 461	7		м	0.2243160002 6928855	0.6686310632 434499	-0.851684229 7707804	1.0239830954 549745	0.9796205161 39859	0.899976620
8		м	1.3757325018	1.6803251838	1.3901196752	1.3396630090	-0.173726786	1.2303809655	0.93	8		м	1.6803251838	0.6348041857	-0.020477608	0.5159337060	1.5484657788	1.514848426
9		м	0.5176496548	-0.330465634	0.5642201506	0.3793386367	0.5190706529	0.9341177279	0.56	9		м	-0.330465634	0.6737786315	0.3551922884	-0.147714570	0.7592931256	0.694066062
10		м	471241 0.0038930458	0053954 0.5684210645	582618 0.0358815552	6081137 -0.092162224	485547	347598 0.0805092746	429	10		м	0053954 0.5684210645	515065 0.1457851965	0224587 -0.221623694	08761582 -0.350373609	380783 -0.105992626	232919 0.222187699
		-	0794308	66244 -0.204059692	67218826 -0.399592739	70623979	1039118	3055303	524 -0.7	10		IVI	66244	2512817	5480976 -0.609125713	1457429	150735	0211217
11		в	4834407	01875798	51283275	90946595	42177117	8502594	451	11		В	01875798	0011308	863456	1531406	9685651	24154
12	< (	м	101103919	467000	6503066	2200011	C 4204 CO	2010771	-0.8 	12	<	М	1.50503/1351	-0.515/99892	-1.140092221	1.2755020308	-0.200260491	-0.50006120

#### 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 "FEATURE\_SELECTION\_BEST\_FIRST" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

ekk							
Edit [	Data Transforma	tion Analytic	s Statistics	Plot Help			
PORT	TRAN_TEST_SPU	NCRIVALISE, TRA	IN SET	2000,087,987	RAIN, MODELL <sup>4</sup> H		
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Cc
Header	User Row ID	texture_mean	concave	fractal_diment	s area_se	radius_worst	pe ^
		-2.086103717	2 3703305076	2 3607370906	2 2227153166	1 794831860	9 2 1
1		1531285	20931	60363	031446	964493	46.
2		-0.356214992	0.4936137407	-0.899012720	0.6342849827	1.716715786	1 1.4
-		5582287	6952486	6795266	997221	821813	82.
		0.4584010780	1.9087047808	97962750128	1.0337956378	1.432293154	8 1.2
		2231966	843263		599626	071551	88
4		0.2547470603	1.3522771590	5.1309328241	-0.303914224	-0.30228430	1 -0.
	-	7/18237	134440	02909	09/034/	3250492	6 1 2
5		7597024	164621	1121528	556092	742195	43
		0.3600853453	0.2406475267	1.7395506474	0.1712931860	0.128355598	2 0.0
0		660462	736	83071	5823035	9206822	63-
7		0.5918295723	1.0662194344	1.6478516963	-0.362188711	-0.186111677	7 -0.
		415474	657272	473742	2252187	24229186	99.
		0.9242303827	-0.411868036	-0.854642260	-0.037974299	0.554989545	3 0.4
8		1 2004021040	1.5140578772	2,2557270014	1.4777631176	0.909516346	4 1.2
8				1212001210014	1.4111001110	0.303310340	4 1.6
8 9		455802	666524	565827	93824	347472	68
8		455802 1.0904307879	666524 0.0884265782	565827 -1.385608767	93824 -0.116674895	347472 0.084290120	68 1 0.1
8 9 10		455802 1.0904307879 555035	666524 0.0884265782 3535445	565827 -1.385608767 8350728	93824 -0.116674895 13306085	347472 0.084290120 9171221	68: 1 0.1 73
8 9 10 11		455802 1.0904307879 555035 0.7767567837	666524 0.0884265782 3535445 0.7406970195	565827 -1.385608767 8350728 2.0812031912	93824 -0.116674895 13306085 2 -0.464519510	347472 0.084290120 9171221 -0.278248585	68 1 0.1 73 5 0.0
8 9 10 11		455802 1.0904307879 555035 0.7767567837 664419	666524 0.0884265782 3535445 0.7406970195 562423	565827 -1.385608767 8350728 2.0812031912 30581	93824 -0.116674895 13306085 2 -0.464519510 59152585	347472 0.084290120 9171221 -0.278248585 99758234	68 1 0.1 73 5 0.0 93

Use the Random Forest Method to train and fit the model by browsing: "Analytics"  $\rightarrow$  "Classification"  $\rightarrow$  "Random Forest" and adjust the model parameters based on training set performance.

🕀 bc.ekk									
File Edit	Data Transforma	tion Analyti	s Statistics	Plot Help					
		Regress	ion	•					
IMPORT	TRAIN_TEST_SPLIT	Classific	ation		at Mariahha an G	LAIND			
		Chustani		k Neare	est iveignbors (	KININ)			
		Cluster	ng	* Multiple	e Layer Percept	tron (MLP)			
		Anomal	y Detection	Radial E	Basis Function (	(RBF)			
		\ Existing	Model Utilizati	on XGBoos	st			Random Forest Classification N	lodel
				J48					•
				Randon	n Forest				
								Eastures fraction	0.0
								realures fraction	0.9
								Min impurity decrease	0.1
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Cc	with imparity deeredee	0.1
User Header	r User Row ID	texture_mean	concave points_mean	fractal_dimens	area_se	radius_worst	pest		Time-based RNG Seed
1		-2.086103717	2.3793395976	2.3607370906	2.2227153166	5 1.7948318609	2.1		
		-0.356214992	0.4936137407	-0.899012720	0.6342849827	7 1.7167157861	1.4	Seed	1234
2		5582287	6952486	6795266	997221	821813	82		
		0.4584010780	1.9087047808	979627501285	1.0337956378	3 1.4322931548	1.2		
		2231966	843263	5 1309328241	599626 -0 303914224	0/1551	-0	Number of ensembles	610
4		7718237	134446	62909	6976347	3250492	79:		
5		0.1611130292	0.5874955818	-0.788086570	0.2303686964	1.2940877916	1.2	Target column	Col12 diagnosis
		7597024	164621	1121528	556092	742195	43	rarget oblamm	Corre allagriosis
6		660462	736	83071	5823035	9206822	63.		
7		0.5918295723	1.0662194344	1.6478516963	-0.362188711	-0.186111677	-0.	<b>E</b> ve evite	Canaal
		415474	657272	473742	2252187	24229186	99	Execute	Cancel
8		0.9242303827	8476715	4525769	92570521	546085	40		
9		1.2894031040	1.5140578772	2.2557270014	1.4777631176	5 0.9095163464	1.2		
9		455802	666524	565827	93824	347472	68		
10		1.0904307879	0.0884265782	-1.385608767 8350728	-0.116674895	0.0842901201	0.1		
		0.7767567837	0.7406970195	2.0812031912	-0.464519510	-0.278248585	0.0		
11		664419	562423	30581	59152585	99758234	93		
12		0.1962257909	0.0626887366	-0.521863808	0.0599508681	0.5309538300	0.4		
	< (						> <		

IMPORT TRAIN\_TEST\_SPLIT NORMALISE\_TRAIN\_SET NORMALISE\_TEST\_SET FEATURE\_SELECTION\_BEST\_FIRST

The predictions will appear on the output spreadsheet.

File Edit D	ata Transforma	tion Analytic	s Statistics	Plot Help														
IMPORT	TRAIN_TEST_SPUT	normalise, tra	IN SET BUTHE SET	ECCA, BC, HS TA	IAIN, MODELL #G													
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	c		Col1	Col2 (S)	Col3 (S)	Col4	Col5	Col6	Col7
User Header	User Row ID	texture_mean	concave	fractal_dimens	area_se	radius_worst	perimeter_wor	smoothness_v	concavity_wo	r o î	User Header	User Row ID	diagnosis	Prediction				
1		-2.086103717	2.3793395976	2.3607370906	2.2227153166	1.7948318609	2.1783380031	1.2814547279	2.131001227	3 2	1		м	м				
		1531285	20931	60363	031446	964493	462767	261252	75943	9	2		м	м				
2		-0.356214992 5582287	0.4936137407	-0.899012720	0.6342849827	1.7167157861	1.4404918359	-0.373256661	-0.147856471	1	3		м	м				
		0.4584010780	1.9087047808	-0.407979627	1.0337956378	1.4322931548	1.2603200974	0.5144270526	0.863892386	0 1	4		м	В				
3		2231966	843263	501285	599626	071551	887083	071184	800449	3	5		м	м				
4		0.2547470603	1.3522771590	5.1309328241	-0.303914224	-0.302284301	-0.273427574	3.3326073877	2.009862715	2 2	6		м	м				_
		7718237	134446	62909	6976347	3250492	79837576	679616	759523	5	7		м	в				
5		7597024	164621	1121528	556092	742195	43546	203878	92073	3			M	M				
6		0.3600853453	0.2406475267	1.7395506474	0.1712931860	0.1283555982	0.0620350430	1.4193473437	-0.020903309	0	0			M				
		660462	736	83071	5823035	9206822	6340836	13811	95028407	8	9		M	m				
7		415474	657272	473742	2252187	24229186	99557276	387067	344106	7	10		M	8				
8		0.9242303827	-0.411868036	-0.854642260	-0.037974299	0.5549895453	0.4395377332	-0.618877883	-0.611574697	-	11		м	В				
		508512	8476715	4525769	92570521	546085	4035256	3979307	3878364	9	12		м	м				
9		455802	666524	565827	93824	347472	688933	-1.239394034	8407855	8	13		м	м				
10		1.0904307879	0.0884265782	-1.385608767	-0.116674895	0.0842901201	0.1020732071	-0.834335095	-0.193404551	-	14		м	м				
10		555035	3535445	8350728	13306085	9171221	7308441	5661905	89267017	8	15		В	В				
11		0.7767567837	0.7406970195	2.0812031912	-0.464519510	-0.278248585	0.0105574034	1.4064199109	2.045719715	1	16		В	в				
		0.1962257909	0.0626887366	-0.521863808	0.0599508681	0.5309538300	0.4280982577	0.6006099374	0.093451445	9 0	17		в	в				
12	<									>	<	< (						>
IMPORT TRAIN	I_TEST_SPLIT	NORMALISE_TR	AIN_SET NOF	RMALISE_TEST_	SET FEATURE,	SELECTION_BE	ST_FIRST FEA	TURE_SELECTIC	N_TEST_SET	TRAIN	MODEL(.fit) +							

## 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\_TEST\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



To validate the model browse: "Analytics"  $\rightarrow$  "Existing Model Utilization". Then choose Model "(from Tab:) TRAIN\_MODEL (.fit)".

🕀 bc.ekk								Existing Model Execution	$\times$
bc.ekk File Edit D IMPORT	TRAIN_TEST_SPUT	tion Analytik Regress Classific Clusterii Anomal Existing	cs Statistics ion aation ng y Detection Model Utilizatio	Plot Help	RAIN, MODELL M) LDATE, MOBELIPHICK			E Existing Model Execution  Model (from Tab:)TRAIN_MODEL(   Type  Description  It is a Random forest model  Model Input Header -> Datatype teader -> Datatype teader -> Datatype	×
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Cc	concave points_mean -> Double fractal_dimension_mean -> Double area_se -> Double radius_worst -> Double	
User Header	User Row ID	diagnosis	texture_mean	concave	fractal_dimens	area_se	rac	perimeter_worst -> Double smoothness_worst -> Double	
1 2 3 4 5 6 7 8 9 10 11		M M M M M M M M M M M M M M M M M M M	-1.159126809 2511253 -0.840771103 5070031 -1.1114994449 532763 -0.328124783 -2.278647 -0.328124783 -0.328124784 -0.328124784 -0.328124784 -0.328124784 -0.328124784 -0.32844784 -0.3284478447844 -0.328447844 -0.3284	1.3302/c16/S19 728/164 0.756340467 807959 0.8676703711 549723 0.392664757 0.392664757 0.85786711790 0.26507 0.26507 0.6864106255 881167 0.6686310632 434499 0.6346041857 905062 0.6737766315 51065 2512817 - 6.50991657 0.011308	-0.579545407 04549 19791511327 085978 23109307974 74537 0-0.28221354 2067252 2067252 11.663980909 870997 1.474349688 2889713 -0.851684229 7707804 -0.20477600 155926533 0.3515922884 0.224582 1.6592152284 0.224582 5480976 -0.609125771 853455	1.0420061325 233611 -0.304715248 56742965 -0.369798437 98827083 -0.3597998437 98827083 -0.3593798437 -0.197378050 -0.197378050 -0.197378050 -0.197378050 -0.197378050 -0.19737805 -0.19757805 -0.19757	1.2 46. -0. 799 -0. 33. 0.8 61 0.2 883 2.5 883 2.5 883 0.9 399 0.7 380 -0. 151 -0. 96i -0.	concavity_worst     >> Double       Image: Transfer Column(s) to Output       Excluded Columns       Col3 texture_mean       Col4 concave points_mea       Col6 ractal_dimension_m       Col6 ractal_dimension_m       Col6 ractal_dimension_m       Col6 ractal_dimension_m       Col6 ractal_dimension_m       Col7 radius_worst       Col9 smoothness_worst       Col10 concavity_worst	
IMPORT TRAIN	TEST_SPLIT	NORMALISE_TF	RAIN_SET NOF	RMALISE_TEST_	SET FEATURE	_SELECTION_BE	ST_FIRST	Execute Cancel	

The predictions will appear on the output spreadsheet.

IMPORT	TRAIN_TEST_SPUT	NORMALISE, TR	NAN, ST. PRILES. 151, ST. PRIMES	12100, 1927, 1927 TR 122100, 1927, 1927 TR	RAIN_MODEL(#R) LOATEMODEL(predict												
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)		Col1	Col2 (S)	Col3 (S)	Col4	Col5	Col6	Col7	Col8
Jser Header	User Row ID	diagnosis	texture_mean	concave	fractal_dimens	area_se	radius_worst	perimeter_wc	User Header	User Row ID	Prediction	diagnosis					
1		м	-1.159126809	1.3302161519	-0.579545407	1.0420061325	1.2259865982	1.251740490	1		м	м					
		141	2511253	789164	04549	253611	463962	937775	2		В	м					
2		м	-0.840771103	0.7563848467	085978	-0.304715248	7968696	21492445	3		В	м					
		м	1.1114984449	0.8676703711	2.9109307974	-0.369798437	-0.266230728	-0.308317974	4		м	м					
1		IVI	532763	549723	74537	98827083	3338487	9510933	5		м	м					
4		м	-0.328124783	0.3928684753	-0.285221354	0.2353750956	0.8013556274	0.802741079	6		м	м					
			1.9307962170	0.5786711790	1.1863989099	-0.197378050	0.2084746493	0.448117339	7		м	м					
5		м	88885	026507	870997	01490397	8362542	35283	8		м	м					
6		м	0.8774133672	0.8894862558	-1.474349688	1.0329946139	2.5539598700	2.275573544	9		м	м					
-			0.2243160002	0.6686310632	-0.851684229	1.0239830954	0.9796205161	0.899976620	10		м	м					
7		м	6928855	434499	7707804	549745	39859	016255	11		p	p					-
8		м	1.6803251838	0.6348041857	-0.020477608	0.5159337060	1.5484657788	1.514848426			0						
			-0.330465634	905062	0.3551922884	-0.147714570	0.7592931256	0.694066062	12		5	M					
9		м	0053954	515065	0224587	08761582	380783	232919	13		в	м					
10		м	0.5684210645	0.1457851965	-0.221623694	-0.350373609	-0.105992626	0.222187699	14		В	м					
			66244	2512817	5480976	1457429	150735	0211217	15		В	В					
11		В	01875798	0011308	863456	1531406	9685651	24154	16		В	В					
12		м	1.3830371351	-0.515799892	-1.140092221	1.2735020308	-0.286260491	-0.368661208	17		в	В					
			467923	2103384	245952	96107	10673785	0021048	18		В	В					
13		м	0.5356491536	-0.728566048	-1.01/333947	-0.438085722	-0.09/980721	-0.169614334	19		м	м					-
			0.5894887215	0.0609728805	-0.144714896	-0.559040327	-0.037891432	-0.083818265	20		R	R					-
14		IVI	640160	0000570	0242052	2272224	70004477	05044070			~						1

# 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".

KNOCT       TAAL IES SHIT       CORE ERRE       DALEEREE	File Edit D	ata Transforma	tion Analytic	s Statistics	Plot Help										
Colt         Col2 (S)         Col3 (S)         Col4         Col5         Col6         Col5         Col7         Col2         Col3         Col4         Col5           User Reader         User Roution         Mayoritic         Mayoritic <td< th=""><th>IMPORT</th><th>TRAIN_TEST_SPUT</th><th>NORMALISE TR NORMALISE TR</th><th>NILSET BILLES ST.SET BILLES</th><th>ettolijet, jezt</th><th>RAIN_MODELLIN) ALOA'R MORELped (</th><th>STATSTICS ACCURACES</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></td<>	IMPORT	TRAIN_TEST_SPUT	NORMALISE TR NORMALISE TR	NILSET BILLES ST.SET BILLES	ettolijet, jezt	RAIN_MODELLIN) ALOA'R MORELped (	STATSTICS ACCURACES								
User Roade         User Roade         Sequence		Col1	Col2 (S)	Col3 (S)	Col4	Col5	Col6	â		Col1	Col2	Col3	Col4	Col5	
1       M	User Header	User Row ID	Prediction	diagnosis					User Header	User Row ID					
2     8     M <td>1</td> <td></td> <td>м</td> <td>м</td> <td></td> <td></td> <td></td> <td></td> <td>1</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	1		м	м					1						
3       8       M	2		В	м					2						
4       M	3		В	м					3						
5       M	4		м	м					4						
6         M         M         A	5		м	м					5						
7     M     M     C     C       8     M     M     C     C       9     M     M     C     C       10     M     M     C     C	6		м	м					6						
8         M	7		м	м					7						
9 M M	8		м	м					8						
10 M M	9		м	м					9						
	10		м	м					10						
11 <u>8 B</u> <u>11</u>	11		В	В					11						
12 B M I I I I I I I I I I I I I I I I I I	12		В	м				U	12						
13 B M 13 13	13		В	м					13						
14 B M 14 14	14		В	м					14						
15 B B I I I I I I I I I I I I I I I I I	15		В	В					15						
	16		В	В					16						
17 8 8 1 I I I I I I I I I I I I I I I I I	17		В	В					17						
<b>18</b> 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	18		В	В					18						
19 M M J J J J J J J J J J J J J J J J J	19		м	м				~	19						

Calculate the statistical metrics for the classification by browsing: "Statistics"  $\rightarrow$  "Model Metrics"  $\rightarrow$  "Classification Metrics".

File Edit Dat	ta Transformat	NORMAUSE_TEXT_	Statistics F Domain - APE Model Metric Probability Di Descriptive St Confidence In	Plot Help D cs istribution Fun itatistics ntervals	Rep actions	gression Metrics				
IMPORT	TRAIN_TEST_SPUT	NORMALISE, TRAIN, S	Domain - AP( Model Metric Probability Di Descriptive St Confidence Ir	D cs istribution Fun itatistics ntervals	Report	gression Metrics issification Metrics				
IMPORT	TRAIN_TEST_SPLIT	NORMALISE (TRAIN)	Model Metric Probability Di Descriptive St Confidence In St	cs listribution Fun itatistics ntervals	Cla	gression Metrics ssification Metrics				
		NORMAUSE_TEST_3	Probability Di Descriptive St Confidence In	istribution Fun itatistics ntervals	ictions Cla	ssification Metrics				
		NORMAUSE_TEST_S	Descriptive St Confidence In	tatistics ntervals						
		NORMAUSE_TEST_S	Confidence Ir	ntervals						
			1		•	STATISTICS ACTURACIES				
			Hypothesis Te	esting				Classification Castistics Materia		
			Weight Cases	s				Classification Statistics Metrics		
			Random Num	nber Generato	r					
	Col1	Col2 (S)	Design of Exp	periments	•	Col6				
User Header	Iser Row ID	Prediction d	iagnosis			colo	<u>^</u>			
1		M N	1					Actual Value Column	Col3 diagnosis	•
2		B N	4							
3		B N	4							
4		M N	4					Prediction Value Column	Col2 Prediction	-
5		M N	4							
6		M N	4							
7		M N	4					beta of F Score	2	
8		M N	4						<u> </u>	
9		M N	4							
10		M N	4							
11		B B						Evecute	Cancel	
12		B N	4					LACCULE	Cancer	
13		B N	4							
14		B N	4					L.		
15		B B								
10		D D								
18		R R								
19		M N	4							
C TRAIN				AAUGE TEET C	CT CEATURE	SELECTION REST				

The results will appear on the output spreadsheet.

Accuracy: 0.923

F1-Score = 0.914

IMPORT	TRAIN_TEST_SPUT	NORMALISE TR	AN ST		IAIN_MODEL(#c)												
	Col1	Col2 (S)	Col3 (S)	Col4	Col5					Col1 (S)	Col2 (5)	Col3 (5)	Col4 (5)	Col5	Colfi	Col7	Col8
User Header	User Row ID	Prediction	diagnosis				â	Classificatior	User Header	User Row ID	con (o)	0015 (0)	0011 (0)				
1		м	M						1			Predicted	Predicted				
2		в	м					Actual Value Colu				Class	Class				
		в	м					Prediction Value Colu	2	Astual Class		M	10				
4		м	м						3	Actual Class	P	40	00				
5		м	м					beta of F Scc	5	Pictour cross	0		~				-
6		м	м						6								
7		м	м						7	Classification	0.9225352112						
8		м	м						'	Accuracy	676056						
9		м	м				- 11	Rit	8			0.0772727272	0.9070501926				
10		м	м				- 11		9	Precision		727273	734694				
11		B	B				- 11		10								
12		в	M				- 11		11	Recall/Sensitiv	1	0.8113207547	0.9887640449				
13		8	M				- 11		12	ny		109012	430202				
14		D	0				- 11		13	Specificity		0.9887640449	0.8113207547				
16		R	R				-					438202	169812				
17		B	B				-		14			0.8865979381	0.9411764705				
18		В	в				-		15	F1 Score		4433	882354				
19		м	м				-		16	_			0.00040200055				
20		В	В						17	F (beta=2)		0.83984375	947136				
21		в	в						18								
22		в	в						19	мсс	0.8368152413						
**	< (	n	0				> <			< (	323554	)			-		>
IMPORT TRAIN	_TEST_SPLIT	NORMALISE_TR	RAIN_SET NOF	RMALISE_TEST_	SET FEATURE	SELECTION_BEST	F_FIRST F	EATURE_SELECTION_TEST_SET	TRAIN_MODEL(	fit) VALIDATE	_MODEL(.predi	t) STATISTICS	S_ACCURACIES	+			

# Step 11: Reliability check of 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 "FEATURE\_SELECTION\_BEST\_FIRST" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Manipulate the data to exclude the column that corresponds to the diagnosis by browsing: "Data Transformation"  $\rightarrow$  "Data Manipulation"  $\rightarrow$  "Select Columns". Then select all the columns except the "diagnosis".



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 of applicability by browsing: "Statistics"  $\rightarrow$  "Domain APD".

File Edit I	Data Transforma	tion Analytic	cs Statistics	Plot Help			
	TRAIN_TEST_SPUT	NORMALISE, IR	Model Me Probability Descriptive Confidence	APD trics / Distribution Fi e Statistics e Intervals s Tection	unctions	DOMAIN	
		NORMAUSE, IS	Weight Ca Random N Design of	ises Jumber Genera Experiments	tor	STAT STICS ACCURACE	
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
User Header	User Row ID	texture_mean	concave	fractal_dimen	<sup>6</sup> area_se	radius_worst	perimeter_wor
1		-2.086103717	2.3793395976	2.3607370906	5 2.2227153166	1.7948318609	2.1783380031
2		-0.356214992 5582287	20931 0.4936137407 6952486	-0.899012720 6795266	031446 0.6342849827 997221	964493 1.7167157861 821813	462/6/ 1.4404918359 822503
3		0.4584010780	1.9087047808 843263	0.407979627 501285	1.0337956378	1.4322931548 071551	1.2603200974 887083
4		0.2547470603 7718237	1.3522771590	5.1309328241 62909	-0.303914224 6976347	-0.302284301 3250492	+0.273427574 79837576
5		0.1611130292	0.5874955818	-0.788086570 1121528	0.2303686964	1.2940877916	1.2803391795 43546
6		0.3600853453	0.2406475267	1.7395506474	0.1712931860	0.1283555982	0.0620350430
7		0.5918295723	1.0662194344	1.6478516963	-0.362188711 2252187	-0.186111677	+0.063799186
8		0.9242303827	-0.411868036 8476715	-0.854642260	-0.037974299	0.5549895453	0.4395377332 4035256
9		1.2894031040	1.5140578772	2.2557270014	1.4777631176	0.9095163464	1.2374411465
10		1.0904307879 555035	0.0884265782	-1.385608767 8350728	-0.116674895	0.0842901201 9171221	0.1020732071 7308441
11		0.7767567837	0.7406970195	2.0812031912	-0.464519510	-0.278248585	0.0105574034

IMPORT TRAIN\_TEST\_SPLIT NORMALISE\_TRAIN\_SET NORMALISE\_TEST\_SET FEATURE\_SELECTION\_BEST\_FIRST FEATURE

The results will appear on the output spreadsheet.

IMPORT	TRAIN_TEST_SPUT	INT 32 INVRCM	n, st - fitte, se 51, st - fitte, se	ELLON FR. THE MAN	IAN, MODEL, RY CLUDE, D'AGNOSS	DOMAIN SWISHS & CORICE											
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)		Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5	Col6	Col7	Col8
er Header	User Row ID	texture_mean	concave	fractal_dimens	area_se	radius_worst	perimeter_wor	smoothness_w ( )	User Header	User Row ID	Domain	APD	Prediction				
1		-2.086103717	2.3793395976	2.3607370906	2.2227153166	1.7948318609	2.1783380031	1.2814547279	1		0.0	3.0635684293	reliable				
		1531285	20931	60363	031446	964493	462767	261252	2		0.0	3.0635684293	reliable				
2		5582287	6952486	6795266	997221	821813	822503	52611414			0.0	235745	renouve				
3		0.4584010780	1.9087047808	-0.407979627	1.0337956378	1.4322931548	1.2603200974	0.5144270526 (	3		0.0	235745	reliable				
		2231966	843263	501285	599626	071551	887083	071184			0.0	3.0635684293	cellable				
4		0.254/4/0603	1.3522771590	5.1309328241	-0.303914224	-0.302284301	-0.273427574	5.33200738777	•		0.0	235745	reliable				
		0.1611130292	0.5874955818	-0.788086570	0.2303686964	1.2940877916	1.2803391795	0.5058087641	5		0.0	3.0635684293	reliable				
2		7597024	164621	1121528	556092	742195	43546	203878 9				3.0635684293					
6		0.3600853453	0.2406475267	1.7395506474	0.1712931860	0.1283555982	0.0620350430	1.4193473437	6		0.0	235745	reliable				
		0.5918295723	1.0662194344	1.6478516963	-0.362188711	-0.186111677	-0.063799186	1.6304954116	7		0.0	3.0635684293	reliable				
7		415474	657272	473742	2252187	24229186	99557276	387067		_		235745					+
8		0.9242303827	-0.411868036	-0.854642260	-0.037974299	0.5549895453	0.4395377332	-0.618877883	8		0.0	235745	reliable				
		1 2894021040	8476715	4525769	92570521	546085	4035256	3979307	9		0.0	3.0635684293	reliable				
9		455802	666524	565827	93824	347472	688933	44252				235745					
40		1.0904307879	0.0884265782	-1.385608767	-0.116674895	0.0842901201	0.1020732071	-0.834335095	10		0.0	3.0635684293	reliable				
		555035	3535445	8350728	13306085	9171221	7308441	5661905 8				3.0635684293					1
10		0.7767567837	0.7406970195	2.0812031912	-0.464519510	-0.278248585	0.0105574034	1.4064199109	11		0.0	235745	reliable				
10		664410	1663433	1 1 / 1 / 1 / 1 / 1		COMPANY AND A STATE	110020100	1037137				3 0635684203					-
11		664419 0.1962257909	562423	-0.521863808	0.0599508681	0.5309538300	0.4280982577	0.6006099374 (	12		0.0	5.0055004255	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\_TEST\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Filter the data to exclude the column that corresponds to the diagnosis by browsing: "Data Transformation"  $\rightarrow$  "Data Manipulation"  $\rightarrow$  "Select Columns". Then select all the columns except diagnosis.

File Edit C	Data Transforma	ation Analyt	ics Statistics	Plot Help							
	lormalizers										
IMPORT	ata Manipulatio	n 🕨 Remov	e Column(s)	I	AIN_MODEL(.fr)						
s s	plit	> Select (	Column(s)								
v	ariable Selectio		e oranni (a)					Colort Column (a)			
	undere bereede	Mathx	Transpose					Select Column(s)			
		V Sort by	Column	( )	CLUDE, D'AGNOSIS	DOMAIN					
		Fill Mis	sing Column(s) V	alues				<b>-</b>			Included Colum
								Excluded Colun	mns —		
					.DATE.NCCEI/prodct	STATISTICS ACCURACE		Col2 diagnosis			Col3 texture mean
								doite allagriosis		>>	colo texture_mean
											Col4 concave point
											ColFfractal_dimon
				10	SAL CLOWNER, SAL						COD Tractal dimen
										>	
										>	Col6 area se
										>	Col6 area_se
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Cc			>	Col6 area_se Col7 radius_worst
User Header	Col1 User Row ID	Col2 (S) diagnosis	Col3 (D)	Col4 (D) concave	Col5 (D)	Col6 (D)	Cc îac			>	Col6 area_se Col7 radius_worst
User Header	Col1 User Row ID	Col2 (S) diagnosis	Col3 (D) texture_mean -1.159126809	Col4 (D) concave points_mean 1.3302161519	Col5 (D) fractal_dimens ion_mean -0.579545407	Col6 (D)				<	Col6 area_se Col7 radius_worst Col8 perimeter_wo
User Header 1	Col1 User Row ID	Col2 (S) diagnosis M	Col3 (D) texture_mean -1.159126809 2511253	Col4 (D) concave points_mean 1.3302161519 789164	Col5 (D) fractal_dimens ion_mean -0.579545407 04549	Col6 (D) area_se 1.0420061325 253611	<b>C</b> c 1.2 16			<	Col6 area_se Col7 radius_worst Col8 perimeter_wo Col9 smoothness
User Header 1 2	Col1 User Row ID	Col2 (S) diagnosis M	Col3 (D) texture_mean -1.159126809 2511253 -0.840771103 r070e01	Col4 (D) concave points_mean 1.3302161519 789164 0.7563848467 02356	Col5 (D) fractal_dimens ion_mean -0.579545407 04549 1.9791511327	Col6 (D) area_se 1.0420061325 253611 -0.304715248	<b>C</b> 1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2			<	Col6 area_se Col7 radius_worst Col8 perimeter_wo Col9 smoothness_
User Header 1 2	Col1 User Row ID	Cot2 (S) diagnosis M M	Col3 (D) texture_mean -1.159126809 2511253 -0.840771103 5070031 11114984449	Col4 (D) concave points_mean 1.3302161519 789164 0.7563848467 807959	Col5 (D) fractal_dimens ion_mean -0.579545407 04549 1.9791511327 085978 2 9109307974	Col6 (D) area_se 1.0420061325 253611 -0.304715248 56742965 -0.3697084 <sup>32</sup> 7	<b>C</b> 1.2 1.6 0. 79 1.0			> < <<	Col6 area_se Col7 radius_worst Col8 perimeter_wo Col9 smoothness_ Col10 concavity_w
User Header 1 2 3	Col1 User Row ID	Cot2 (S) diagnosis M M	Col3 (D) texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984499 532763	Col4 (D) concave points_mean 1.3302161519 789164 0.7563848467 807959 703711549728	Col5 (D) fractal_dimens ion_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537	Col6 (D) area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083	Cc 1.2 1.6 0. 79( 0. 3.3			<	Col6 area_se Col7 radius_worst Col8 perimeter_wo Col9 smoothness_ Col10 concavity_w
User Header 1 2 3	Col1 User Row ID	Col2 (S) diagnosis M M M	Col3 (D) texture_mean -1.159126809 2511253 -0.840771103 5070031 1.111498449 1321763 -0.328124783	Col4 (D) concave points_mean 1.3302161519 789164 0.7563848467 807959 703711549723 0.3928684753	Col5 (D) fractal_dimens ion_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537 -0.285221354	Col6 (D) area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956	<b>c</b> c 1.2 1.6 0. 790 0. 333 3.8			<	Col6 area_se Col7 radius_worst Col8 perimeter_wo Col9 smoothness_t Col10 concavity_w
User Header 1 2 3 4	Col1 User Row ID	Col2 (S) diagnosis M M M M	Col3 (D) texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647	Col4 (D) concave points_mean 1.3302161519 789164 0.7563848467 807959 703711549723 0.3928684753 118454	Col5 (D) fractal_dimens ion_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537 -0.285221354 2067252	Col6 (D) area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956 4182776	Cr ar 1.2 1.2 1.3 3.3 3.3 3.3 51			<	Col6 area_se Col7 radius_worst Col8 perimeter_wo Col9 smoothness_ Col10 concavity_w
User Header 1 2 3 4 5	Col1 User Row ID	Col2 (S) diagnosis M M M M M M M	Col3 (D) texture_mean -1.159126809 2511253 -0.840771103 5070031 -1.1114984449 532763 -0.328124783 -0.328124783 -0.328124783 -0.328124783	Col4 (D) Concave points_mean 1.3302161519 789164 0.7563848467 807959 703711549722 0.3928684753 118454 0.5786711790 025607	Co15 (D) fractal_dimens on_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537 -0.285221354 2067252 1.18633980099 970007	Col6 (D) area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956 4182776 -0.197378050 01409107	Cc 11.2 12.2 16. 0. 79.9 0. 0. 13.3 13.3 15.2 12.2 12.2 13.3 15.2 15.			<	Col6 area_se Col7 radius_worst Col8 perimeter_wo Col9 smoothness_v Col10 concavity_wo
User Header 1 2 3 4 5	Col1 User Row ID	Cot2 (5) diagnosis M M M M M M M	Col3 (D) texture_mean -1.159126809 2511253 -0.840771103 5070031 -0.328124783 2278447 1.9307962170 8885 0.8774133672	Col4 (D) concave points_mean 1.3302161519 789164 0.7563848467 807959 703711549722 0.3928684753 118454 0.5786711790 026507 0.8894862558	Col5 (D) fractal_dimens ion_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537 -0.285221354 2067252 1.1863989099 870997 -1.474349688	Col6 (D) area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956 4182776 -0.197378050 01490397 1.0329946139	Cc 27 12 12 12 16 0. 79 0. 0. 13 13 13 13 15 15 15 15 15 15 15 15 15 15		Exect	< <li>&lt;</li> <li>&lt;</li> <li>ute</li>	Col6 area_se Col7 radius_worst Col8 perimeter_wor Col9 smoothness_1 Col10 concavity_work
User Header 1 2 3 4 5 6	Col1 User Row ID	Col2 (5) diagnosis M M M M M M M M	Col3 (D) texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.9307962170 88895 0.8774133672 002451	Col4 (D) concave points_mean 1.3302161519 789164 0.7563848467 807959 703711549723 0.3928684753 118454 0.3786711790 0.26507 0.8894862558 891167	Col5 (D) fractal_dimens cn_mean -0.57954407 04549 1.9791511327 045578 2.9109307974 74537 -0.285221354 2067252 1.18639809099 870997 -1.474349668 2869713	Col6 (D) area_se 1.0420061325 253611 -0.304715248 56742965 -0.36979843 0.2535750956 -0.197378050 01490397 1.022946139 901676	Cc 1.2 1.2 1.3 1.2 1.3 1.2 1.3 1.3 1.3 1.3 1.3 1.3 1.3 1.3		Exect	< <ul> <li></li> <li>&lt;</li> <li>ute</li> </ul>	Col6 area_se Col7 radius_worst Col8 perimeter_wo Col9 smoothness_ Col10 concavity_w
User Header 1 2 3 4 5 6 7	Col1 User Row ID	Col2 (5) diagnoss M M M M M M M M M M M M	Cot3 (D) texture_mean -1.159126809 2511253 -0.840771103 5070031 -0.328124783 -0.32812478 -0.328178 -0.3281478 -0.39	Cold (D) Concave points, mean 1.3302161519 789164 0.7563484467 0.3928684753 118454 0.5786711790 0.28507 0.8894862558 911167 0.6868310632	Col5 (D) fractal_dimens ion_mean -0.57954540 1.9791511327 065578 2.9109307974 74537 745777 745777 745777 745777 745777 7457777 7457777 74577777 7457777777 7457777777777	Col6 (D) area_te 1.042061325 253611 -0.304715248 56742965 -0.369798437 98827008 -0.2533750954 4182776 -0.197378050 0.1490397 1.0329346139 901676 1.0239346139 901676	<b>Cr</b> <b>11</b> <b>12</b> <b>12</b> <b>12</b> <b>12</b> <b>12</b> <b>12</b> <b>12</b> <b>12</b> <b>12</b> <b>13</b> <b>13</b> <b>13</b> <b>13</b> <b>13</b> <b>13</b> <b>13</b> <b>13</b> <b>13</b> <b>13</b> <b>13</b> <b>15</b> <b>15</b> <b>15</b> <b>16</b> <b>16</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b> <b>17</b>		Exect	< <ul> <li></li> <li>&lt;</li> <li>&lt;</li> <li>ute</li> </ul>	Col6 area_se Col7 radius_worst Col8 perimeter_wo Col9 smoothness_ Col10 concavity_w
User Header 1 2 3 4 5 6 7	Col1 User Row ID	Col2 (5) diagnosis M M M M M M M M M M	Col3 (D) texture, mean -1.159126809 2511253 -0.240771103 5070031 1.1114884449 532763 -0.322184783 2278647 1.9307962170 88885 0.02741133672 0.022431 0.02	Cold (D) concave points, mean 1.3302161519 703711549722 0.33286473 118454 0.57865711790 0.63894862558 891167 0.66686310632 434499	Co15 (D) fractal_climens co.mmean co.mmean co.579545407 04549 13791511327 085978 29109307974 74537 -0.285271354 2067252 2067252 2067252 2067252 2067252 1.18633980099 870997 -1.474349688 2889713 -0.85166422 -0.201477684	Col6 (D) area_rea 1.0420061325 253611 -0.30471524 56742965 -0.369798437 9827083 0.2353750956 4182776 -0.197378050 01490397 -0.197378050 01490397 -0.1579337065 549745 549745 -0.5159337061 -0.515937061 -0.515937060	Cr. 1.2 1.2 1.2 1.3 1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2		Exect	< <ul> <li></li> <li>&lt;</li> <li>&lt;</li> <li>ute</li> </ul>	Col6 area_se Col7 radius_worst Col8 perimeter_wo Col9 smoothness_ Col10 concavity_w
User Header 1 2 3 4 5 6 7 8	Col1 User Row ID	Cot2 (5) diagnosis M M M M M M M M M M M M M M M M M M	Col3 (D) texture_mean -1.159126809 2511253 -0.040771103 5070031 -1.1114984445 -0.28214784 -0.28214784 -0.28214784 -0.28214784 -0.28214100002 -0.2451 -0.02251 -0.02451 -0.02251 -0.02451 -0.02251 -0.02451 -0.02251 -0.0251 -	Col4 (0) Concave points, mean 1:302-161518 789164 0:7563249467 807959 703741549722 0:3728684753 118454 0:3786711790 0:36846753 881167 0.6686510632 884499 0:6384041857 0:636461857	Col5 (D) fractal_dmens lot_mean -0.57954540 -0.57954540 -0.57954540 -0.57954540 -0.57954578 -0.285221354 -0.285221354 -0.285221354 -0.285221354 -0.2854713 -0.851684229 7107804 -0.20477608	Col6 (D) #rea_se 1.0420061325 233611 1.0420061325 233611 1.030471524 56742965 -0.367799437 98827003 0.2335750956 4182776 0.197378050 01490397 1.0239830954 549745 0.5159337060 375155	Cr 1.2 1.2 1.6 1.0 1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2		Exect	< <ul> <li></li> <li>&lt;</li> <li>&lt;</li> <li>ute</li> </ul>	Col6 area_se Col7 radius_worst Col8 perimeter_wo Col9 smoothness_1 Col10 concavity_w

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 of the test set predictions by browsing: "Analytics"  $\rightarrow$  "Existing Model Utilization". Then select as Model "(from Tab:) DOMAIN".

IMPORT	TRAIN_TEST_SPUT	Classific Clusterin Anomal	ation ng		RAIN_MODEL(.M)		Existing Model Execution
		Existing	Model Utilizati		KLUDE, DIAGNOSS LOATE, KOSE, gewärt	DOMAIN SATSIC_KOMOS	Model (from Tab: )DOMAIN - Type APD Model Description
							Model Input
	Cell	Cel2 (D)	Cel2 (D)	Cell (D)	Col5 (D)	Calf (D)	Header -> Datatype radius_mean -> Double
	2011	COIL (D)	COIS (D)	C014 (D)	cois (o)	C010 (D)	
User Header	User Row ID	texture_mean	concave	fractal_dimen	area_se	radius worst pe	perimeter_mean -> Double
ser Header 1	User Row ID	texture_mean -1.159126809	concave points_mean 1.3302161519	fractal_dimen ion_mean -0.579545407	area_se 1.0420061325	radius_worst pr st 5 1.2259865982 1.	area_mean -> Double
ser Header	User Row ID	texture_mean -1.159126809 2511253 -0.840771103	concave points_mean 1.3302161519 789164 0.7563848467	fractal_dimen ion_mean -0.579545407 04549 1.9791511327	area_se 1.0420061325 253611 -0.304715248	radius_worst         pt st           1.2259865982         1           463962         93           -0.190117629         -0	area_mean -> Double smoothness_mean -> Double
lser Header 1 2	User Row ID	texture_mean -1.159126809 2511253 -0.840771103 5070031	concave points_mean 1.3302161519 789164 0.7563848467 807959	fractal_dimen ion_mean -0.579545407 04549 1.9791511327 085978	area_se 1.0420061325 253611 -0.304715248 56742965 0.020700427	radius_worst 81 5 1.2259865982 1 463962 93 5 -0.190117629 -0 7968696 21 0 026630320 0	perimeter_mean -> Double area_mean -> Double smoothness_mean -> Double compactness_mean -> Double
Jser Header 1 2 3	User Row ID	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763	concave points_mean 1.3302161519 789164 0.7563848467 807959 0.8676703711 549723	fractal_dimen ion_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537	area_se 1.0420061325 253611 - 0.304715248 56742965 - 0.369798437 98827083	radius_worst st 1.2259865982 1 463962 93 -0.190117629 -0 7968696 21 -0.266230728 -0 3338487 95	area_mean -> Double smoothness_mean -> Double compactness_mean -> Double concavity_mean -> Double
lser Header 1 2 3 4	User Row ID	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783	concave points_mean 1.3302161519 789164 0.7563848467 807959 0.8676703711 549723 0.3928684753	fractal_dimen ion_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537 -0.285221354	area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956	radius_worst 01 5 1.2259065982 14 463962 93 1 -0.190117629 -0 7968696 21 -0.266230728 -0 3338487 95 5 0.8013556274 0.3	perimeter_mean -> Double area_mean -> Double smoothness_mean -> Double compactness_mean -> Double concavity_mean -> Double concave points_mean -> Double
Jser Header 1 2 3 4	User Row ID	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647	concave points_mean 1.3302161519 789164 0.7563848467 807959 0.8676703711 549723 0.3928684753 118454	fractal_dimen ion_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537 -0.285221354 2067252	area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956 4182776 -0.303720000	radius_worst pr 463962 93 1-0.190117629 -0 7968696 21 -0.266230728 -0 3338487 95 5 0.8013556274 0. 611457 92 0 200474600 0	perimeter_mean -> Double area_mean -> Double smoothness_mean -> Double compactness_mean -> Double concavity_mean -> Double concave points_mean -> Double
Jser Header 1 2 3 4 5	User Row ID	10xture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.9307962170 88885	concave points_mean 1.3302161519 789164 0.7563848467 807959 0.8676703711 549723 0.3928684753 118454 0.5786711790 026507	fractal_dimen ion_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537 -0.285221354 2067252 1.1863989095	area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956 4182776 -0.197378050 01490397	radius_worst         Pt           463962         93           4-0.190117629         -0           7968696         21           -0.266230728         -0           3338487         95           50.8013556274         0.3           611457         92           8362542         4	perimeter_mean -> Double area_mean -> Double smoothness_mean -> Double compactness_mean -> Double concavity_mean -> Double concave points_mean -> Double Transfer Column(s) to Output
Jser Header 1 2 3 4 5 6	User Row ID	10xture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.9307962170 88885 0.8774133672	concave points_mean 1.3302161519 789164 0.7563848467 807959 0.8676703711 549723 0.3928684753 118454 0.5786711790 026507 0.8894862558	fractal_dimen ton_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537 -0.285221354 2067252 1.1863989099 -1.474349688	area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956 4182776 -0.197378050 01490397 1.0329946139	radius_worst         P%           1        2259865982         1           463962         9            1         -0.190117629         9           7968696         21            1         -0.266230728         -0           3334487         92         0.08013556274         0           611457         92         8         8           8362542         35         3         8           2.25539598700         2         3         3	perimeter_mean -> Double area_mean -> Double smoothness_mean -> Double compactness_mean -> Double concavity_mean -> Double concave points_mean -> Double Transfer Column(s) to Output
User Header 1 2 3 4 5 6	User Row ID	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.9307962170 8885 0.8774133672 002451	concave points_mean 1.3302161519 789164 0.7563848467 807959 0.8676703711 549723 0.3928684753 118454 0.5786711790 026507 0.8894862558 891167	fractal_dimen ton_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537 -0.285221354 2067252 1.1863989099 -1.474349688 2889713	area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956 4182776 -0.197378050 01490397 1.0329946139 901676	radius_wonst         Product           1.2259805982         1.           463962         9.           -0.190117629         0.           -0.266230728         -0.           333847         92           0.0013556274         0.           0.144539         2.           0.2084746493         0.           2.5539593702         2.           8895         5	perimeter_mean -> Double area_mean -> Double smoothness_mean -> Double compactness_mean -> Double concavity_mean -> Double concave points_mean -> Double Transfer Column(s) to Output
Jser Header 1 2 3 4 5 6 7	User Row ID	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984449 532763 -0.328124783 2278647 1.9307962170 8885 0.8774133672 0.22451 0.22451 0.22451 0.22451 0.22455	concave points_mean 1.3302161519 789164 0.7563848467 807959 0.8676703711 549723 0.3928684753 118454 0.5786711790 026507 0.8894862558 891167 0.6688310632 434499	fractal_dimen ton_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 74537 -0.285221354 2067252 1.1863988095 870997 -1.474349688 2889713 -0.851684229 7707804	area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956 4182776 -0.190337 01490397 1.0329946135 901676 330954549743	radius_worst         P           1.2259865982         1.           463962         9.           -0.190117629         -0.79686966           2         -0.266230728           0.2084746493         0.           838642         9.           2.25539598700         2.           8395         55           9.98560         1.5	perimeter_mean     ->     Double       area_mean     ->     Double       smoothness_mean     ->     Double       concavity_mean     ->     Double       concave points_mean     ->     Double       Transfer Column(s) to Output     Execute     Cancel
Jser Header 1 2 3 4 5 6 7 2	User Row ID	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114994449 532763 -0.328124783 2278647 1.9307962170 8885 0.8774133672 002451 0.2243160002 6928855 1.6803251838	concave points_mean 1.3302161519 789164 0.7563848467 807959 0.8676703711 549723 0.3928684753 118454 0.5786711790 026507 0.8894862558 891167 0.6886310632 434499 0.6348041857	fractal_dimen ton_mean -0.579545407 04549 1.9791511327 085978 2.9109307974 2.9109307974 2.9109307974 -0.285221354 2067252 1.1863989095 870997 -1.474349688 2889713 -0.851664229 7707804 -0.020477608	area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956 4182776 -0.197378050 01490397 1.0329946133 901676 330954549743 0.5159337060	radius_worst         Product           1         2.259865982         1.           463962         9.         -           -0.90117629         -         7968696         21           -0.266230728         -         0.66230728         0.           610457         2         0.2084744493         0.           0.2084744493         0.2084744493         0.         2.553958700           2.553959700         2.         39859         10           39859         1.54846577788         1.         1.5484657788	permeter_mean     ->     Double       area_mean     ->     Double       smoothness_mean     ->     Double       concave points_mean     ->     Double       concave points_mean     ->     Double       Transfer Column(s) to Output     Execute     Cancel
User Header 1 2 3 4 5 6 7 8	User Row ID	texture_mean -1.159126809 2511253 -0.840771103 5070031 1.1114984407 532763 -0.328124783 2278647 1.9307962170 88885 0.8774133672 002451 0.2243160002 6928855 1.8603251838 931417	concave points_mean 1.3302161519 789164 0.7563848467 807959 0.8676703711 549723 0.3928684753 118454 0.5786711790 0.5786711790 0.5586711790 0.65894862558 891167 0.6686310632 434499 0.6348041857 905062	fractal_dimen ion_mean ion_mean ion_mean ion_mean ion_nean ion_mean i	area_se 1.0420061325 253611 -0.304715248 56742965 -0.369798437 98827083 0.2353750956 4182776 -0.197378050 01490397 1.0329946135 901676 30954547474 0.5159337060 375155	radius_worst         Project           5         2.25965982         1.25965982           -0.190117629         -0.7968696         2           -0.266230726         2         -333847           9         0.801355627         0.           0.208474693         0         8362542           9         2.5539598700         2           9.38895         0         9.9796205161           9.939590         1         5.48457788           1         1.5484657788         1	perimeter_mean -> Double area_mean -> Double smoothness_mean -> Double compactness_mean -> Double concavtp_mean -> Double concave points_mean -> Double Transfer Column(s) to Output  Execute Cancel

The results will appear on the output spreadsheet. There is one unreliable sample in the test set.

File Edit Data Transformation Analytics Statistics Plot Help	
Cott Cot2 (D) Cot4 (D) Cot4 (D) Cot6 (D) Cot6 (D) Cot7 (D) Cot8 (D) 🙃 Cott Cot2 (D) Cot4 (S) Cot5 Cot5 Cot6 Cot7 Cot8	
User Header User Row ID texture mean concave fractal_dimensiances are and use works perimeter_worksmoothness.w 2 User Header User Row ID Domain APD Prediction	Â
Opinis, mean lion, mean lion, mean lion, mean lit lion lion lion lion lion lion lion lion	٦
2511253 789164 04549 253611 469962 937775 7155378 392725 392725 715378 39275 715378 39275 715775 71578	
2 35745 10581 2357	-11
3 1.1114964449 (0.8676/03711 29/030794 (-).269796437 (-).26630728 (-).30381794 (-).269796437 (-).26630748 (-).2676670481 (-).267645 (-).26765 (-).26765 (-	
4 0.328124783 (0.32864473) -0.25873795 (0.0135524 (0.02714079) (0.30758128) 4 1136009599 (0.0350424) eliable	
227647 118454 207252 4182776 611457 924124 1253866 5105075 5 10363564293 eliable	1
5 1300 Sec 10 Garbon 112 11000 Sec 2010 1130 1000 Control 1130 100	-
6 0.877413872 0.885465259 1.474349688 10329945139 12553958970 1225573545 0.3231338501 6 120000105 0.30000000000000000000000000000000000	
0.2243160002 [0.6668310632 ] -0.851684229 [0.97956205161 [0.8999766205 [0.0576577628] 7 [0.904254822] [0.90425482] [0.90425482] [0.90425482] [0.90425482] [0.90425482] [0.90425482] [0.90425482] [0.90425482] [0.90425482] [0.904254822] [0.904254222] [0.904254222] [0.904254222] [0.904254222] [0.904254222] [0.9042	
622855 43499 7707804 2005452202 3999 016255 1040675 8 095442935 3.063564293 reliable	
8 1.0005210591005401610/100204761708103793100113499377861.51499454400124545694 883412 235745 00000000000000000000000000000000000	-
e -0.330465634 0.6737786315 0.3551922884 -0.147714570 0.7592931256 0.6940660622 0.5230453410 9 168394 235745 reliable	
0053954 515055 0224587 08761582 380783 232919 93849 Y Y	~~~

# Final Isalos Workflow

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

