

Data mining, analysis and visualization Using Python-Orange

diagraff pains

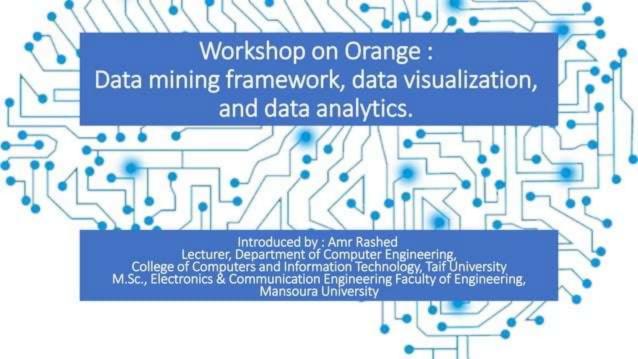
Eng. Amr Rashid

🖥 تاريخ الورشة

27/3/2021

8:30 - 10:30 PM

Jing!





- ➤Topics:
- >Introduction & overview
- ➤ Application 1:

Fault Detection for Attaining Service Continuity of Photovoltaic Power System

➤ Application 2:

Data Mining for Diagnosis of Breast Cancer in Medical Ultrasonic Images

Project 1

Seport Strapes Project 2 Confusion Metro **Distinct Booking** Rendon Feeth Heard Saturals

Applied Machine Learning Process





Step 1: What is the problem?

Define the Problem

?

Step 2: Why does the problem need to be solved?



Step 3: How would I solve the problem?

Data Preparation Techniques

1.Common Data Preparation Tasks

1.Data Cleaning

1.Feature Selection

1.Data Transforms 1.Feature Engineering 1.Dimensionality Reduction

Common Data Preparation Tasks

•Step 1: Define Problem.

•Step 2: Prepare Data.

•Step 3: Evaluate Models. •Step 4: Finalize Model.

Data Cleaning



Data cleaning involves fixing systematic problems or errors in "messy" data.



Using statistics to define normal data and identify outliers.



Identifying columns that have the same value or no variance and removing them

Identifying duplicate rows of data and removing them.

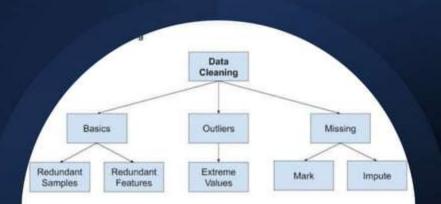


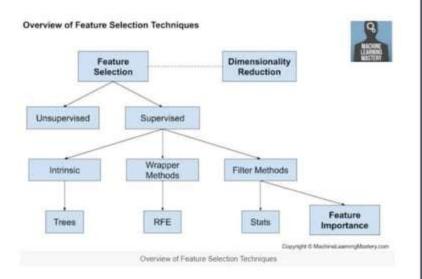
Marking empty values as missing.



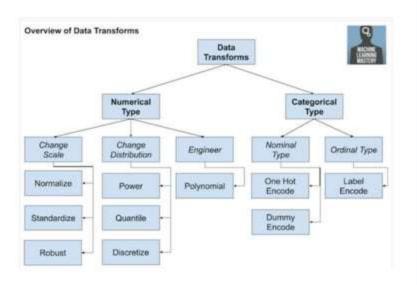
Imputing missing values using statistics or a learned model.

Overview of data cleaning





Overview of feature selection techniques



Overview of Data Transforms

Feature Engineering

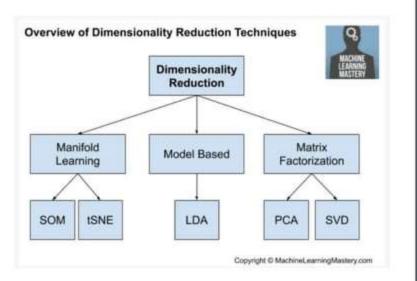
Feature engineering is the process of creating new input variables from the available data.

Adding a Boolean flag variable for some state.

Adding a group or global summary statistic, such as a mean.

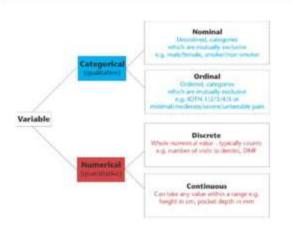
Adding new variables for each component of a compound variable, such as a date-time.

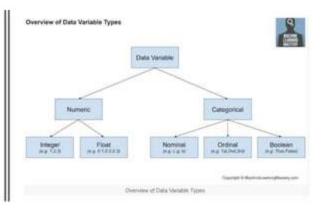
Polynomial Transform: Create copies of numerical input variables that are raised to a power(raising them to a power or multiplied with other input variables).



Overview of Dimensionality Reduction Techniques

Overview of data types





Orange installation



Screenshots

Worldlows

Download

Blog

Docs

Workshops



Data Mining Fruitful and Fun

Open source machine learning and data visualization.

Build data analysis workflows visually, with a large, diverse toolbox.

Download Orange



Cont.



Screenshots.

Workflows

Download

Workshops









Download the latest version for Windows

Cont.



Change 3-3.28 (0 Ministrata widt, 6-1 mm (0-4 tist) Care be used without administrative priviledges.

Portable Orange

Orange 2-325 Gales

Ariaconda

If you are using python provided by Ansaumba distribution, you are simpet ready to go. Add conductor farge to the fist of channels you can install packages from

runts cortig - and charmals comis-forge

and run

cook install owners

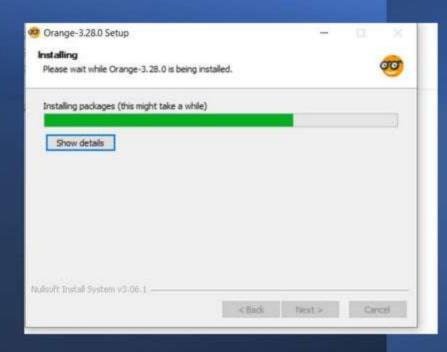
Pip

Drange can also be installed from the Python Package Index.

pip install image?

Cont.

Today Orange3-3.28.0-Miniconda-x86_64.exe https://download.biolab.si/download/files/Orange3-3.28.0-Miniconda-x86_64.exe 376 KB/s - 29.6 MB of 414 MB, 17 mins left. Cancel Pause





ANACONDA NAVIGATOR



At Connecty

Applications on the root



Comment Address of

the party speciments with party sorted. producing to from temperary a modest



Strategy Selfa Kindyser Tool water servers builting abstraction by particular, facilitations group Replace, following to the child and District Service and General Service



destinate description and solutions date for DRAMARIES PAIN SALE TO CHARLES AND THE inschine learning restals. Present day, with Sufficient No. 10 years again, according State attended to the state of the state of the



the beautiful final provide paying its executive exposured by transfer. with the restorate series of the beautiful the Junglet Assistant profit acrossment



make based, trade to be accommon to contribute descriptions, \$40 and but Appropriate Associated and a security of the Territor and delivery



0.60

But a biomerical internal solit plancolonial and constraint from the patrol Allower





PpChare-Coremety: SW1.1.2

an 200 by and some his paid former. development, Supports obtained announced String and Military



Qt-Corrocia

665 Nath Cornel Assessed minerapore. problem multi-free withing-wint option Equitopiting generical settings, and entire



Supports A Print Speciments Studiosewic householdprint 68 with exhaltited editing, how become steeling. belongs of the second state from



100 from the property of the countries of th the burney waterprop with end Arrivago relational decisions:



1.04.0 Comparison Classed Date Strong Pages and St. Detection better and easy engage. For receive and works. Promitting world have with a wage technic



.

Politicano Professionali

Child Redgell (M. Ny Jedharia Novemb branching and later, further development. Ingenior WHO, JL and Kill.



Test programming to

Distance in Free

KARLOTTERSTEEL COLLARSPANCE WAS you be more productive over 4, recludes 4 married and resident













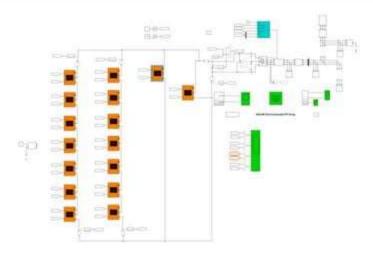


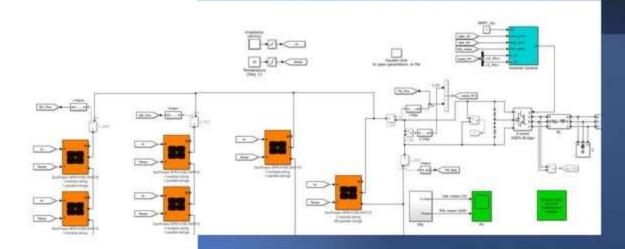


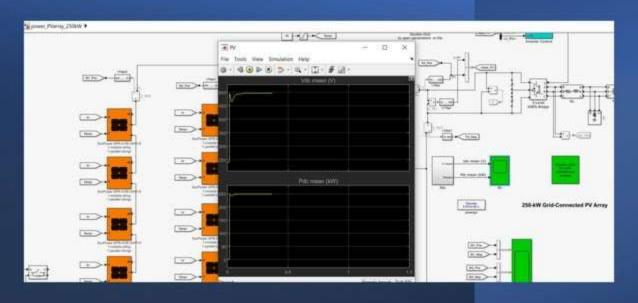




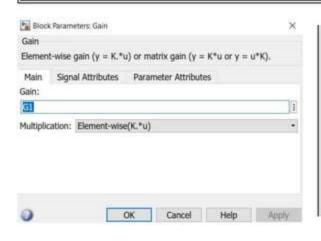
Create Dataset for the First Project (250kw grid connected PV array)

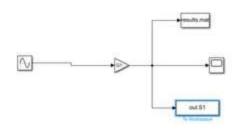


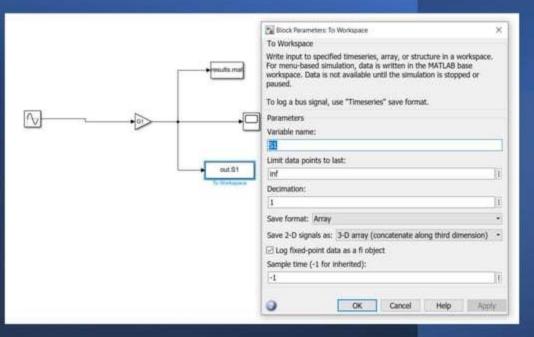




Example

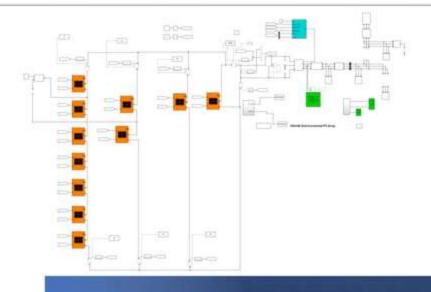


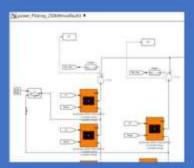


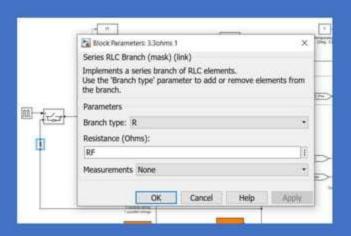


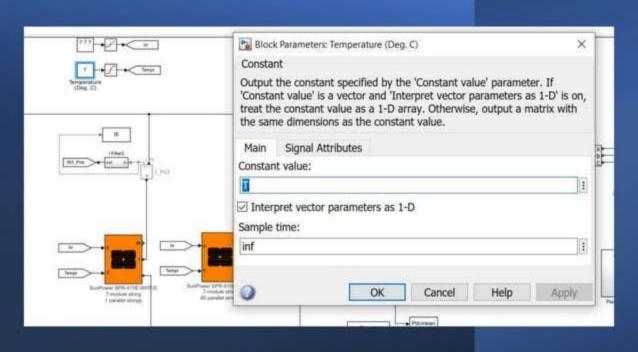
MATLAB Script

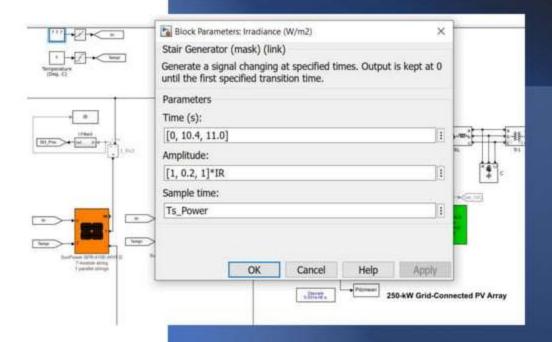
```
clc;clear;close all
                                                                                            61 +
     | for i=1:5
                                                                                              5.3417
       G1=randi([1,10],1);
                                                                                            mil in
       %simulation
                                                                                              K-8718
5 -
6
7 -
        a=sim('E:\FE project\all codes\new1', 'SimulationMode', 'normal');
        textract data generated by simulink model
                                                                                            60 +
                                                                                              1,7627
        S1 = a.get('S1');
8 -
       ml=mean(S1)
                                                                                            mi.-
9 -
       end
                                                                                               1.1141
10
                                                                                            46. -
                                                                                              Links
```

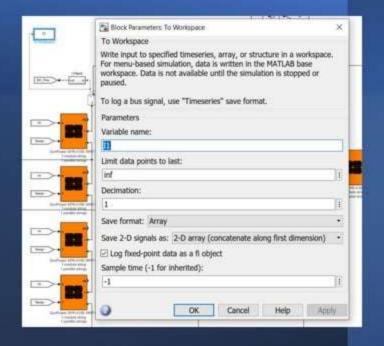


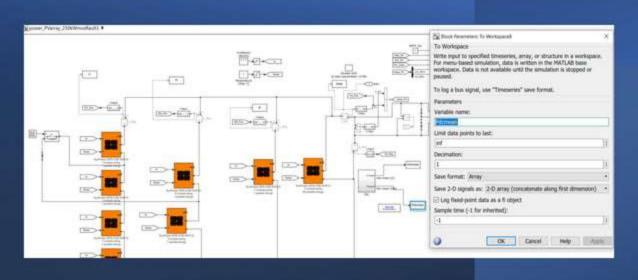












```
solar3.m × sa.m × 4
       clerclear; close all ;
       format short
       tuiopen('E)\FE project\all codes\matlab simulink\power_FVarray 250kMmod.mlx',1)
       Arin.
       Adedine model constant
       mimouti=zeros(1,30);
     | for i=1:2810:50
         disp(1)
       %Ir-0:1000
10
11
       %rf=0.001:20000hm
12 -
       T=randi((10,351,1);%1-var
12 -
       IR=randi [[100, 1000], 11/42-way
14 -
       LOC=26;33-VAE
15
       %500-0:%normal condition
16-
       Rf-randi([1,2000],1):54-var
17
       %EF=100007%normal condition
111
       Strample-D.0005;
19-
       tfinal=0.4/%sim stop time %5-var
10
       %mamplet=Autor%mample time 2+0.005
11 -
       timef=0.2:%fault time
12
       bulmulation.
13 -
       awsin('E)\FE project\all codes\nutlab simulink\power FVarray 250Weodfault3', 'Simulation'
14
       textract data generated by simulink model
       Il = a.get('Il');
15 -
16 -
       12 = a.get('12');
17 -
      13 - a.get('13'):
      I4 = a.get("14");
18 -
     15 = a.det('15');
```

```
market to saw to wi
                      THE ROBERT STORY
                           District v. a. art ("hornout") I
 61 -
                            Witness - a. prill'However'll.
                            Philal - 4.381 ("Distail")
 10 -
                          includes a second section of
 60 -
                           to-disses (n/21)
 64 =
                            quedimos (196/4)
                          9-Seogra (Vicenaria)
 66.0
                           Sent Liver to City
 16-
                           quedhair (Trackly)
 40
                            Different (2) (income 1) 40-max
                            EDD-main (ED that open in this year)
                            Clearman ICL (Re-set L/RF-less)
                            Elaboration (23.00x-1901) a 64-160
 44.-
                            CINAD-MARKET DIA - GREEK AND - MARK
                            CONNECTED AND STREET, SAN DESCRIPTION OF STREET,
67 -
                            Distance CD theory 1772 -- ex-
44 -
                            40-
                           Different Gliffer; and Life in con-
                            Likeway (1) the capt (1) (1) year.
61 -
                            Characterist (Chirage) ) / Crimena
48 6
                            Challe and a LEY Maniput 1 ( 1) Tower
44 -
                            Elisarysas IES (temperatural)
                            DESCRIPTION OF STREET
 68 -
                            Distriction of Children and Colorest
 to a
                            Children and Children and Links over
                            Distriction (U.S. Charles S. C. C. Charles
```

```
meries - David St. +1
                                        ASSESSMENT OF THE PROPERTY OF THE PARTY OF T
                                        PERSONAL PRINTED PROPERTY AND ADDRESS OF THE PERSONS ASSESSED.
16 -
                                        Stomest Jensel / Richest things ) / 126-rell
40 -
                                        Printed Lines (Vilones Oter pril 1 i 127-41)
44 --
                                        William Land read to Philippens and Their special of A VIV a read
44.-
                                        Directal Limited (Chotal (Springs)) (LDT-room
44.-
                                        25 ( Calibrati-Stat ( Divided ( December 2 ) / Villemont
44 -
                                        ProteinghiseastProteinmought (1975-ma)
ea.
                                        Shadle vigors
146
                                        discounted of
                                        Arrivant Smill. S.
                                        Assertagence atting family 5
                                        ALC: DE-COST
                                                               Black-Blateness condition of the fault objects
                                       brite
                                                       & whose-behavioral conditions
                                        epicort (1, 1)-11(1, 1); Thor. Their, Civir, Cher. Cher. Their, Their, 131, 144, Thur. Thirt, Their, Their, Their,
 11.
                                        Athinas
16 -
                                        444
                                        hard's C. Lill Posse
                                        by a collabor or highligh headighteen
                                       Spin-1992, Company of the Company of
                                        Tilleman 'ni bendamp benasebeni alaa'y
45 -
                                       allertywithlesses, title, "mostly, 'mp's
14 -
                                       singsite Hilmann, simusi, "Short", "At"
                                       Schools (ECComme) 2
```



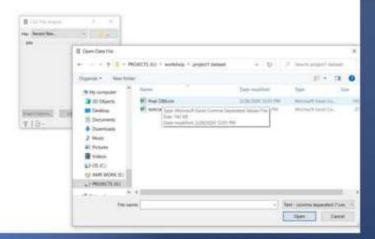
U.	Vinne	W	N. A.	- Y	1	AA.	All	A.	'AD
Vdcmean1	Vdcmas1	Vdcmin1	Pdcmean1	R T		LOC	RF	simstop fir d	lass
483.2946	483.6525	482,3762	58.59916	242	35	26	1942	0.4	/3
494.6765	494,9698	494,4559	133,0941	537	- 34	26	1601	0.4	- 3







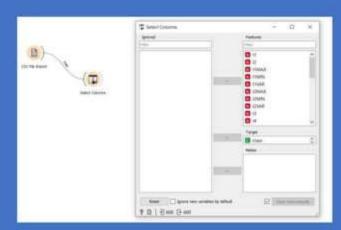










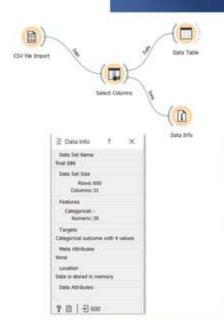




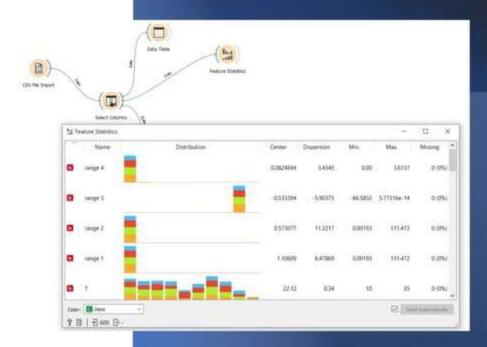
Vacc the detect in a special dust.

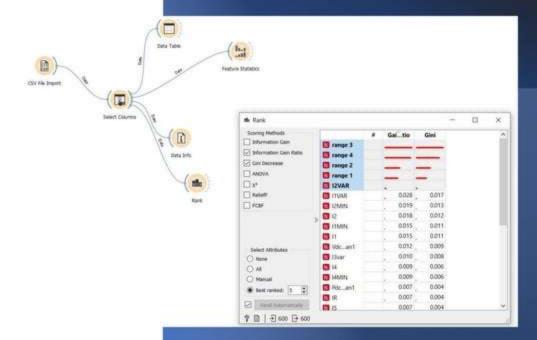


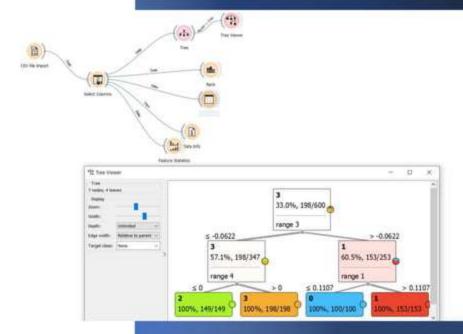
**			(ine		36	(imp)	- /4
SEE negacion (no coming twee). 26 Septions:		. 1		540413	34603	9.7	ne8
Target with 4 values to make attributes		- 8		23407	226477	24	HE.
Territor.				3.87836	3,87800	4.1	gt.
There is national to be the Companion of the control of the contro		- 1		2.61629	241679	3.2	at-
Trauma famore values	1	- 8		12388	1,2166	5.4	ista.
Cities by Indiana (fema)		- 1		496736	456750	5.1	14
	. ,	- 8		3.15704	3,15754	4.7	79
Trained full cours	5 4	- 8		2,2575	22575	23	910
C INTELLIGENCE		- 8		3.08912	1,000	43	NEM .
		- 1		290525	19250	6.4	air.
	- 1	- 1		134945	4,34943	4.	56.
	1	: 1		440027	4.60037	53	24
	1	- 8		1.62548	882946	1.3	Mb-
		- 1		281784	2,81794	9,1	231
Retting Drighed Diller	1	- 8		14621	5/625	5.0	ar.
(S) the second (à	J.	-	4.44707	6.68767	40	vist."

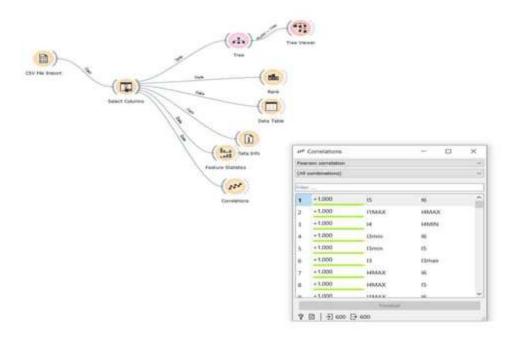


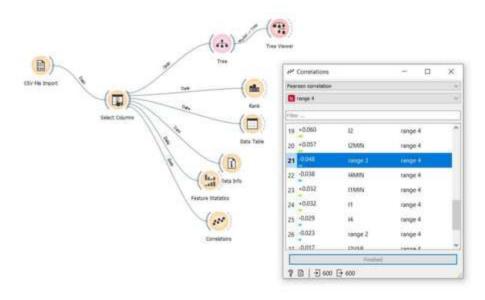


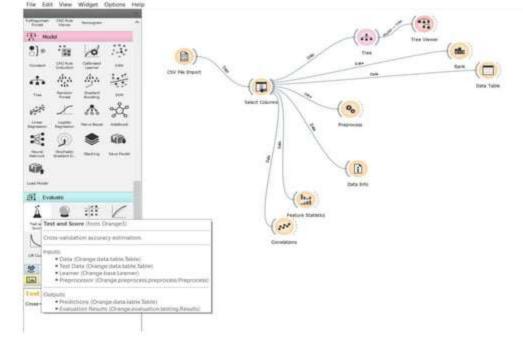


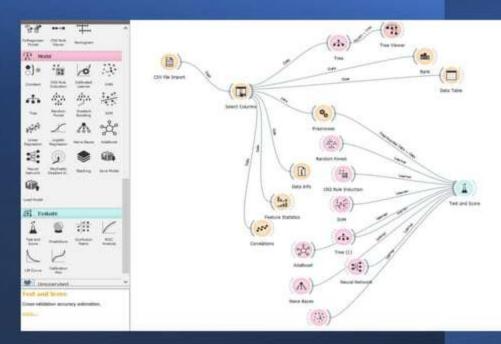














• Commonwell III

Associated Associated

Citmetted

take on or

Aspect Variable: 36

Separate and Mile.

Andre State State

| Moder | MC | Ch | 11 | Parcero Reset | MN | SAD | SA

Two 0.995 0.996 0.996 0.996 0.996 0.996 0.995 0.997 0.527 0.476 0.498 0.523 0.498 0.996 0.

Name Sapes APSC 6500 DADS AREA ASSE CNC not reduce 1991 1992 1999 1999 1999

Append 5,000 1,000 1,000 1,000 1,000

O Test on test date
O Test on test date
Target Class
Orderings over classes

Mark Comparison Annual ASS 2019

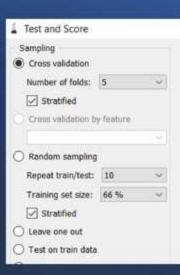
Treatment SE sere

Make (properties by REC.

Two Sangton France ** saw Since Yeners Save Save Old role retigen Salahora: 8000 Sidel 2.860 history 0.000 9000 1400 1.000 0.000 51279 2.757 3.000 5817 9,071 **Date** 100 School Printers 100 DVM 1.704 SART 2000 SHIP Salaton Passet 1,000 8.729 0.044 21244 0.000 3.667 2300 Newson's Restractor's Alter 9.258 SHOW E 136 67% 5457 20104 Same Same 1,000 2.170 0.000 nors. 2.264 2544 stins. 1000 3160 0.136 0.256 8119 2762 con before: 0.000 5.145 Addison 1000 8 579 1.000 1100 11864 2.361 1000

- 75

No de la completa del completa de la completa de la completa del completa de la completa del la completa del la completa de la completa del la completa de la completa de la completa del la com

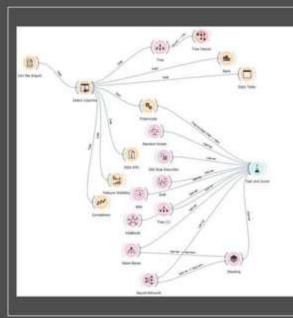


Evaluation Results

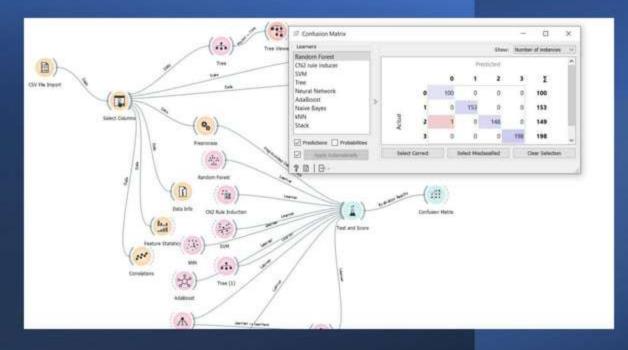
Model	AUC.	CA	F1	Precision	Recall	Train time (¢1
kNN	0.685	0.420	0.435	0.471	0.420	Test time (s	
Tree	0.999	0.998	0.998	0.998	0.998	✓ AUC	
SVM	0.801	0.522	0.470	0.490	0.527	✓ CA	
Random Forest	1.000	0.998	0.998	0.998	0.998	✓ F1 ✓ Precision	
Neural Network	0.996	0.963	0.964	0.964	0.96		
Naive Bayes	0.992	0.920	0.920	0.931	0.920	LogLoss	
CN2 rule inducer	0.991	0.958	0.958	0.959	0.958	Specificity	
AdaBoost	1.000	1.000	1.000	1.000	1.000		

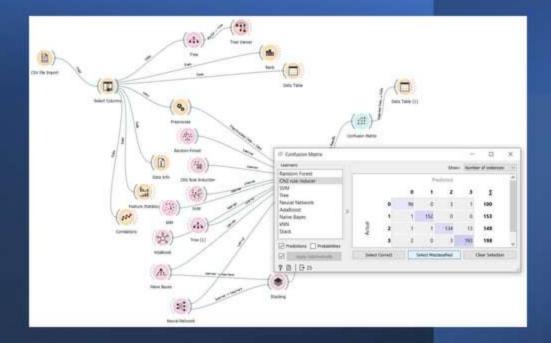
Evaluation Results

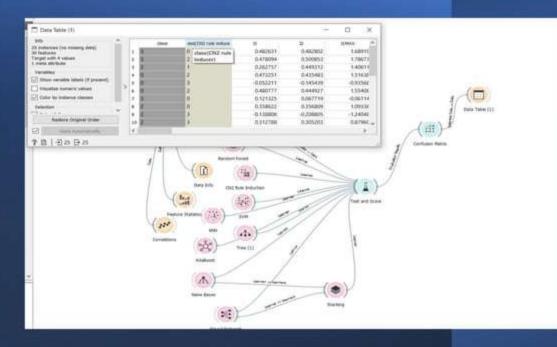
Model	Train time [s]	Test time [s]	AUC	ČA	F1	Precision	Recall	LogLoss	Specificity
AdaBoost	0.084	0.020	1.000	1.000	1.000	1.000	1.000	0.000	1.000
Random Forest	0.111	0.029	1.000	0.998	0.998	0.998	0.998	0.073	1.000
Tree	0.097	0.001	0.999	0.998	0.998	0.998	0.998	0.058	0.999
Neural Network	3.269	0.014	0.996	0.963	0.964	0.964	0.963	0.293	0.991
CN2 rule inducer	13.204	0.037	0.991	0.958	0.958	0.959	0.958	0.213	0.983
Naive Bayes	0.154	0.023	0.992	0.920	0.920	0.931	0.920	0.249	0.979
SVM	0.457	0.054	0.801	0.522	0.470	0.490	0.522	1.057	0.798
kNN	0.075	0.138	0.685	0.420	0.435	0.471	0.420	5.483	0.820

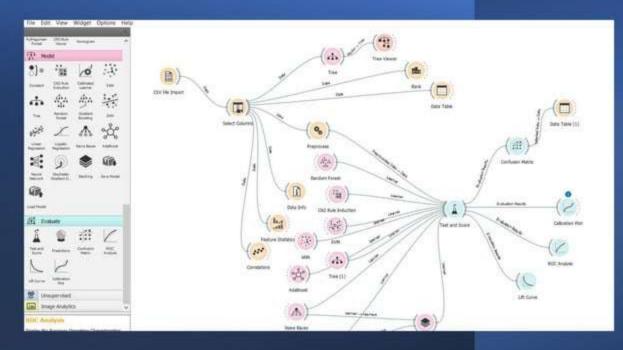


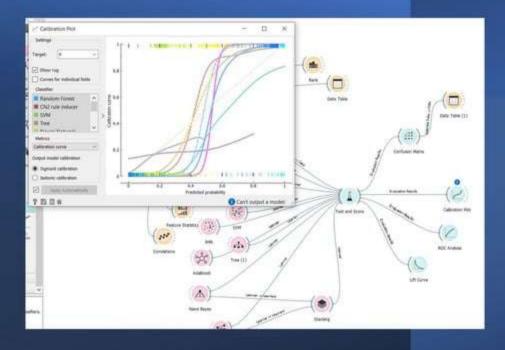
Sangering:	Delivative function									
* 100 0000	Medic	See tow 14	feet time (a)	AUC	-CA	.81	Person	food	Ingine	Secretary
Actor of the T	Abdust	5.004	0.000	tak	180	1.69	1000	1.69	Naki	5 (6)
[[mens	Distance Committee	8111	\$10%	1,000	2.000	210	- tree	2,560	Battle	1000
THE REST OF SHARE	Terr	0.007	- ban	Ame	0.000	phe) THE	0.69	0.26	0.004
Chian make	Set	2.76	- 1	100	181	-	- 100	-		- 100
Stanforming, 16	State Street	336	STR	100	090	5.004	5364	0.00	670	3.00
Santan Mile	No se sere	16396	9407	desi	684	194	0.004	ANN	820	0.00
Chinate	Name State	309	910	TMG	141	1400	0.091	non	534	1479
Total street	(NW)	0.407	596	8401	101	6479	540	KNI	100	479
The other has	Toron .	9875	\$100	1485	340	100	3471	140	140	1004

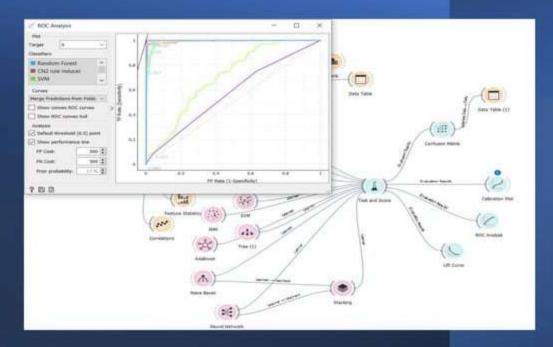


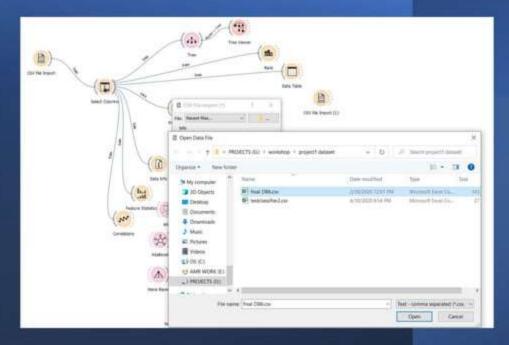


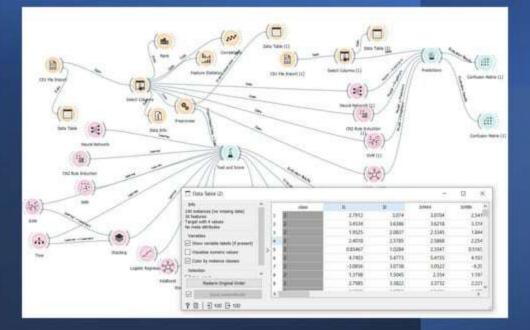


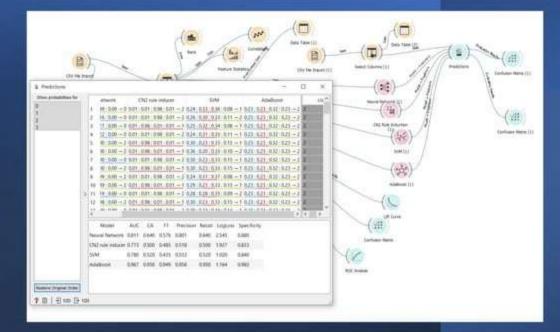










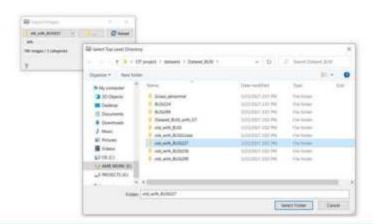


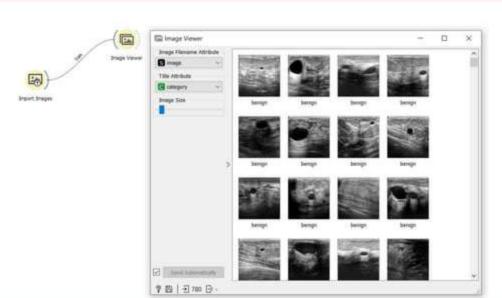
Project 2

 Data Mining for Diagnosis of Breast Cancer in Medical Ultrasonic Images









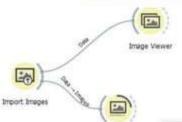


Image Embedding 21%

