Azure Machine Learning Studio

Partea a II-a

Machine Learning Studio (clasic)

Machine Learning Studio (clasic) este un instrument drag-and-drop care poate fi utilizat pentru a construi, testa și implementa soluții de analiză predictivă.

https://studio.azureml.net/

CONTINUAREA PARTII I (LAB. 1)

Data Input and Output

Include urmatoarele module:

- Enter Data Manually: Permite crearea de seturi mici de date tastând valori.
- Export Data: Scrie setul de date la anumite adrese (URL-uri) web sau sub diferite forme de stocare în cloud în Azure, ar fi tabele, blobs sau o bază de date SQL.
-
- Import Data: Încarcă date din surse externe pe web și din diverse forme de stocare în cloud în Azure sau dintr-o bază de date SQL Server locală
-
- Load Trained Model: Permite incarcarea unui model antrenat deja.
-
- Unpack Zipped Datasets: Dezarhiveaza un set de date care a fost stocat în format zip si apoi adaugă setul de date la spațiul de lucru.

Baze de date

Baza de date poate sa aiba urmatoarele tipuri de formate

- fisier text care contine pe prima linie numele atributelor bazei de date, separate de tab, iar pe urmatoarele linii valorile atributelor respective, câte o linie pentru fiecare înregistrare.
  - *Fisierul text poate fi creat cu orice editor de texte*

Baze de date

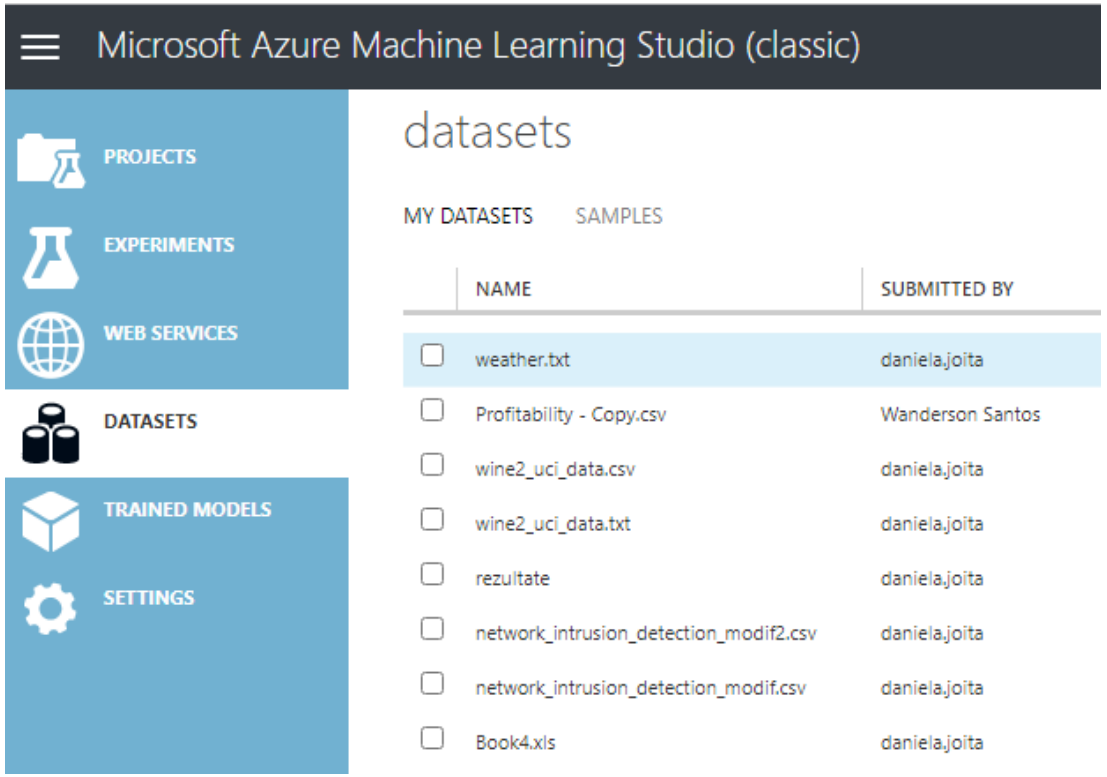
Fisier .arff (format Weka)

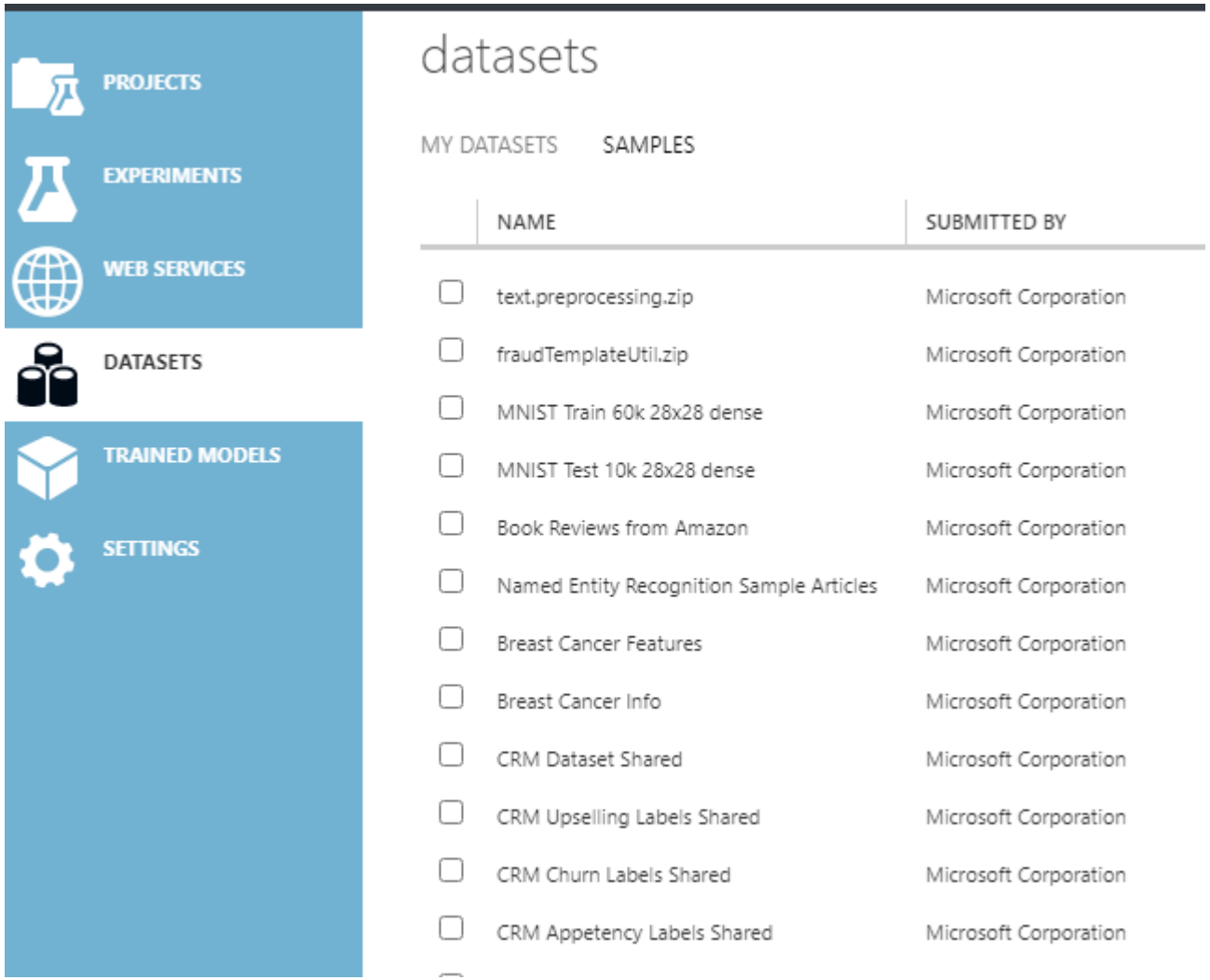
Weka – un alt open source software pentru data mining http://www.cs.waikato.ac.nz/ml/weka/

Creat de un grup de cercetatori de la University of Waikato, Noua Zeelanda – Eibe, Witten

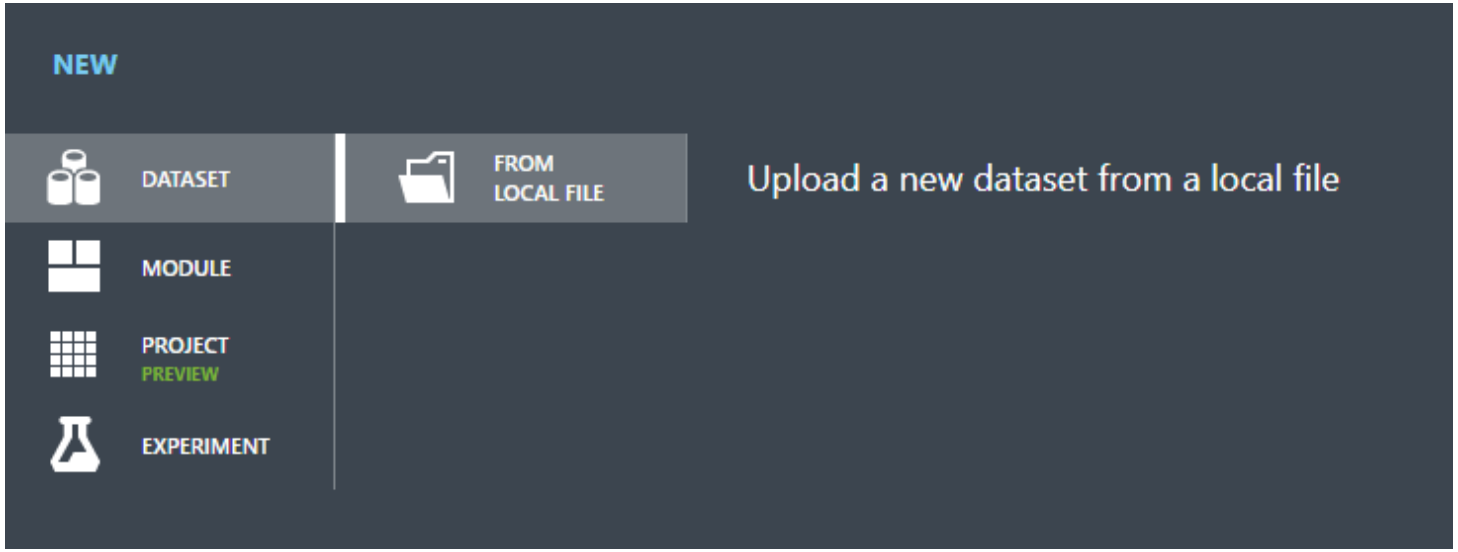
O colectie de algoritmi de machine learning pentru data mining

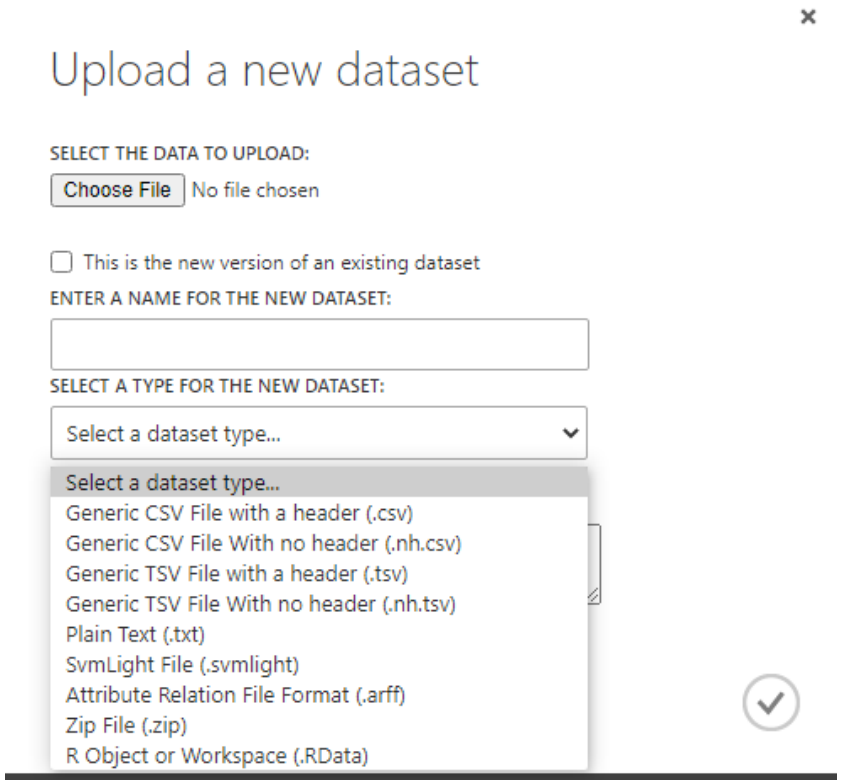
Datasets


Microsoft Azure Machine Learning Studio (classic)
PROJECTS
EXPERIMENTS
WEB SERVICES
DATASETS
TRAINED MODELS
SETTINGS
datasets
MY DATASETS
SAMPLES
NAME
SUBMITTED BY
weather.txt
daniela.joita
Profitability - Copy.csv
Wanderson Santos
wine2_uci_data.csv
daniela.joita
wine2_uci_data.txt
daniela.joita
rezultate
daniela.joita
network_intrusion_detection_modif2.csv
daniela.joita
network_intrusion_detection_modif.csv
daniela.joita
Book4.xls
daniela.joita


PROJECTS
EXPERIMENTS
WEB SERVICES
DATASETS
TRAINED MODELS
SETTINGS
datasets
MY DATASETS
SAMPLES
NAME
SUBMITTED BY
text.preprocessing.zip
Microsoft Corporation
fraudTemplateUtil.zip
Microsoft Corporation
MNIST Train 60k 28x28 dense
Microsoft Corporation
MNIST Test 10k 28x28 dense
Microsoft Corporation
Book Reviews from Amazon
Microsoft Corporation
Named Entity Recognition Sample Articles
Microsoft Corporation
Breast Cancer Features
Microsoft Corporation
Breast Cancer Info
Microsoft Corporation
CRM Dataset Shared
Microsoft Corporation
CRM Upselling Labels Shared
Microsoft Corporation
CRM Churn Labels Shared
Microsoft Corporation
CRM Appetency Labels Shared
Microsoft Corporation

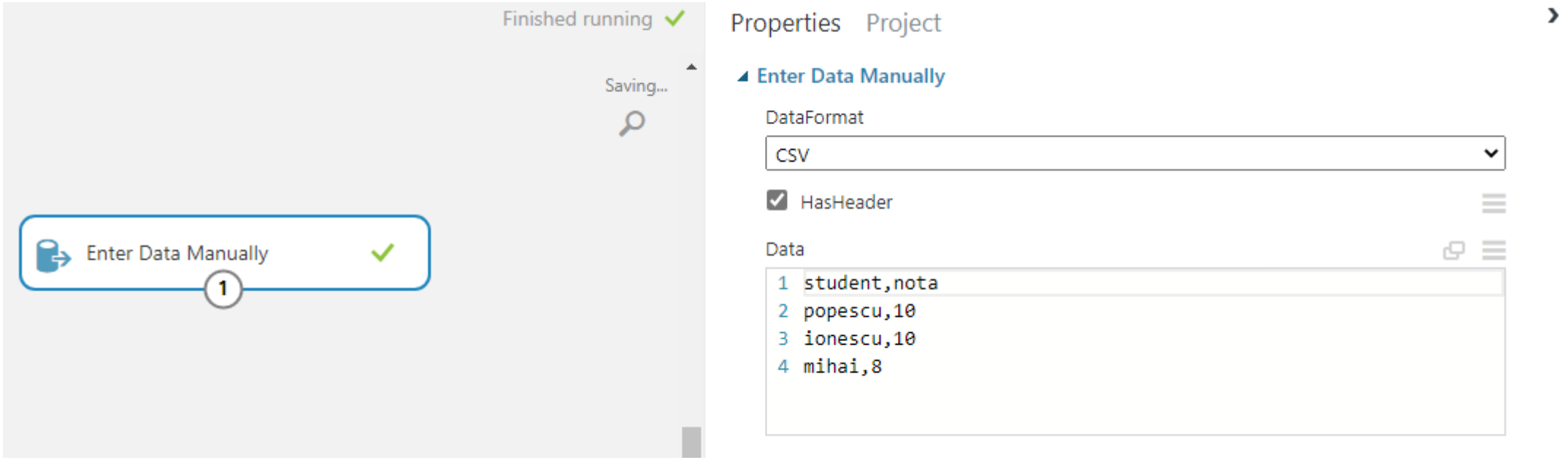
Datasets

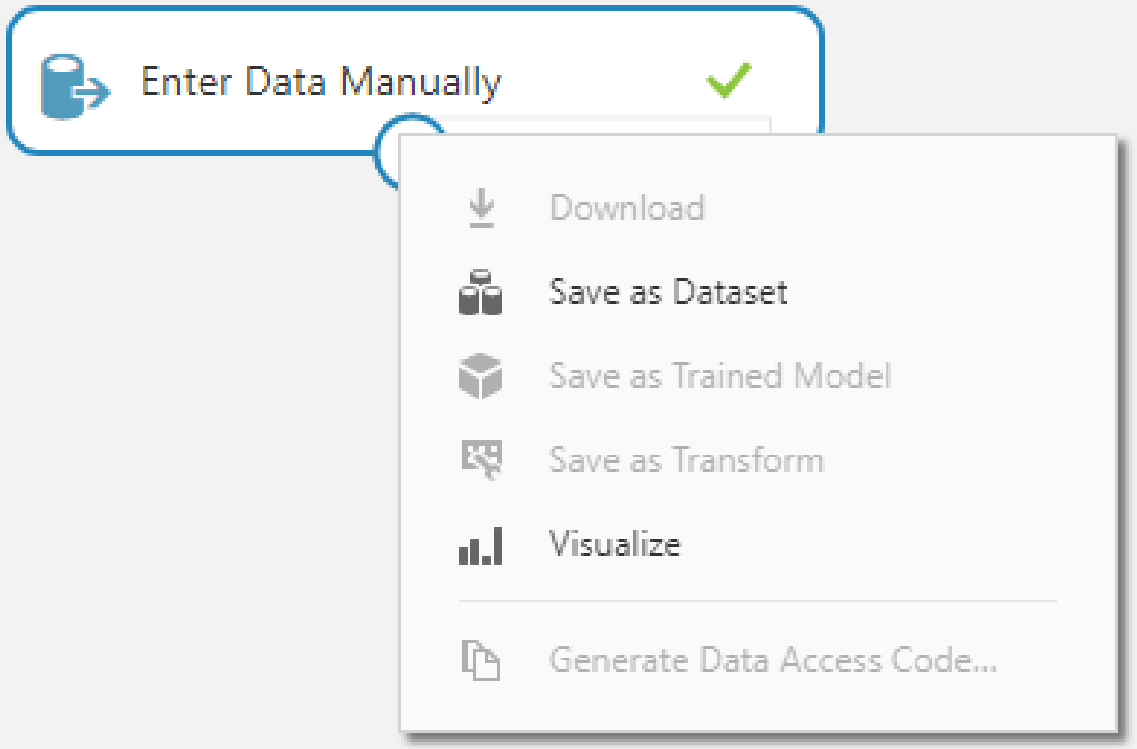

NEW
DATASET
FROM LOCAL FILE
MODULE
PROJECT
PREVIEW
EXPERIMENT
Upload a new dataset from a local file


Upload a new dataset
SELECT THE DATA TO UPLOAD:
Choose File
No file chosen
This is the new version of an existing dataset
ENTER A NAME FOR THE NEW DATASET:
SELECT A TYPE FOR THE NEW DATASET:
Select a dataset type...
Select a dataset type...
Generic CSV File with a header (.csv)
Generic CSV File With no header (.nh.csv)
Generic TSV File with a header (.tsv)
Generic TSV File With no header (.nh.tsv)
Plain Text (.txt)
SvmLight File (.svmlight)
Attribute Relation File Format (.arff)
Zip File (.zip)
R Object or Workspace (.RData)

Enter Data Manually

Permite crearea de seturi mici de date tastând valori.





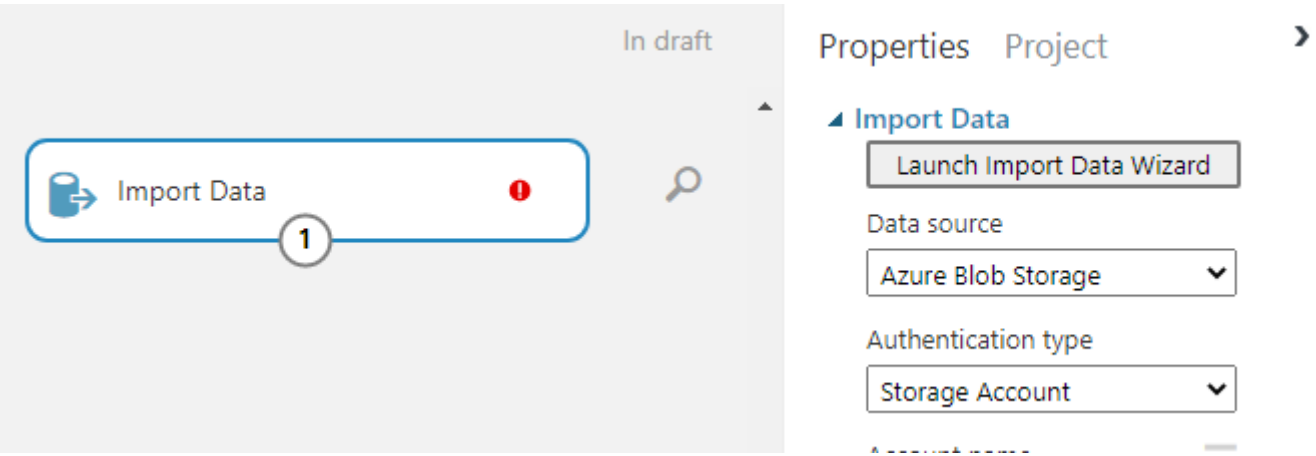
rows	columns
3	2

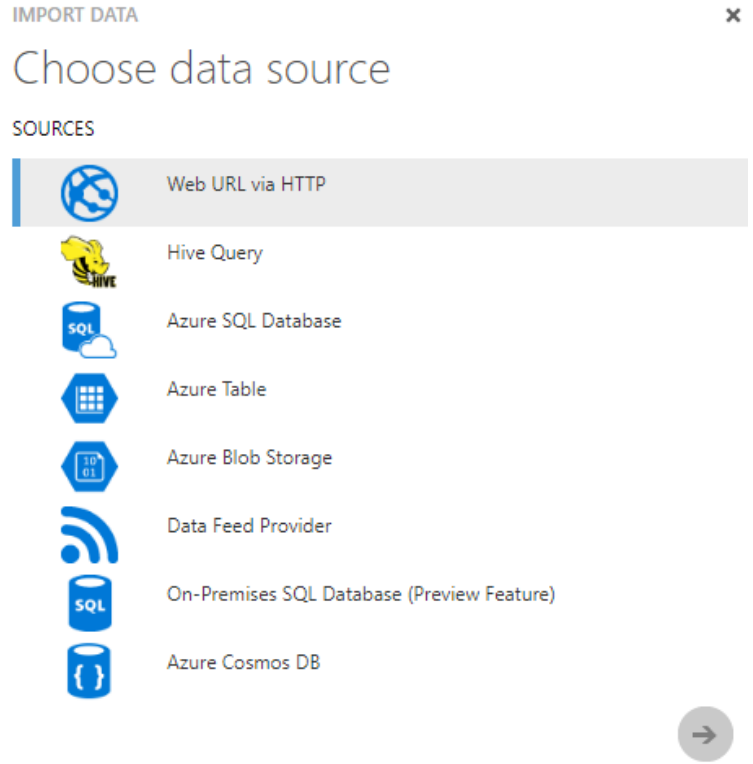
view as

student	nota
popescu	10
ionescu	10
mihai	8

Import data

Via http – se furnizeaza adresa web URL





Web URL – fisier CSV

archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame

UCI Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Tic-Tac-Toe Endgame Data Set

Download: Data Folder, Data Set Description

Abstract: Binary classification task on possible configurations of tic-tac-toe game

Jocul x si 0

Index of /ml/machine-learning-databases/tic-tac-toe

- Parent Directory
- Index
- tic-tac-toe.da
- tic-tac-toe.na

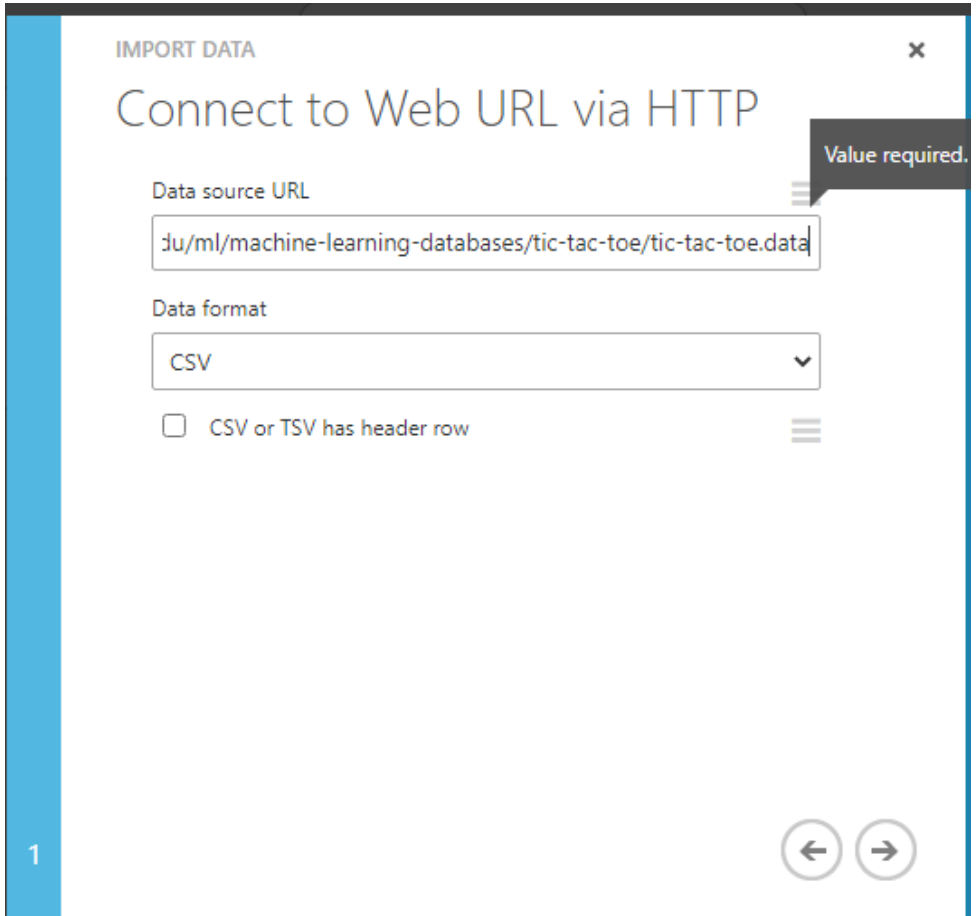
Apache/2.4.6 (Cent ... Passenger/4.0.53 mod_perl/2.0.11 Perl/v5.16.3

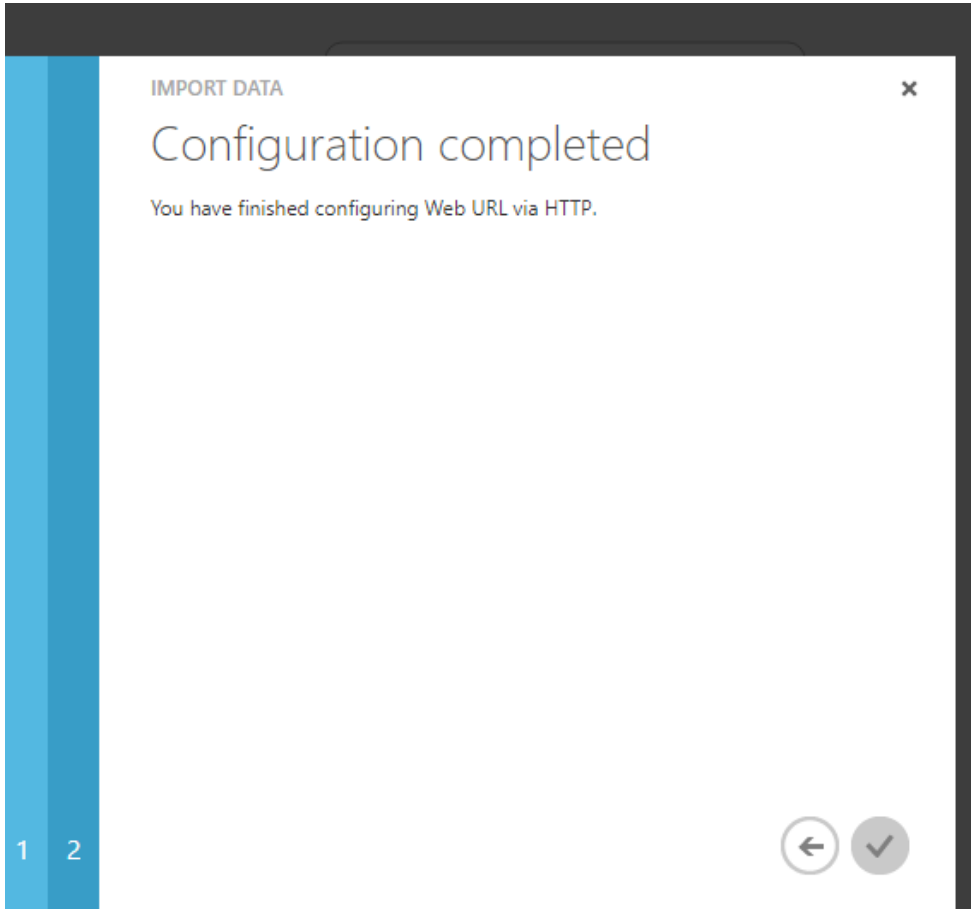
Open link in new tab
Open link in new window
Open link in incognito window
Save link as...
Copy link address
Inspect Ctrl+Shift+I

https://archive.ics.uci.edu/ml/index.php

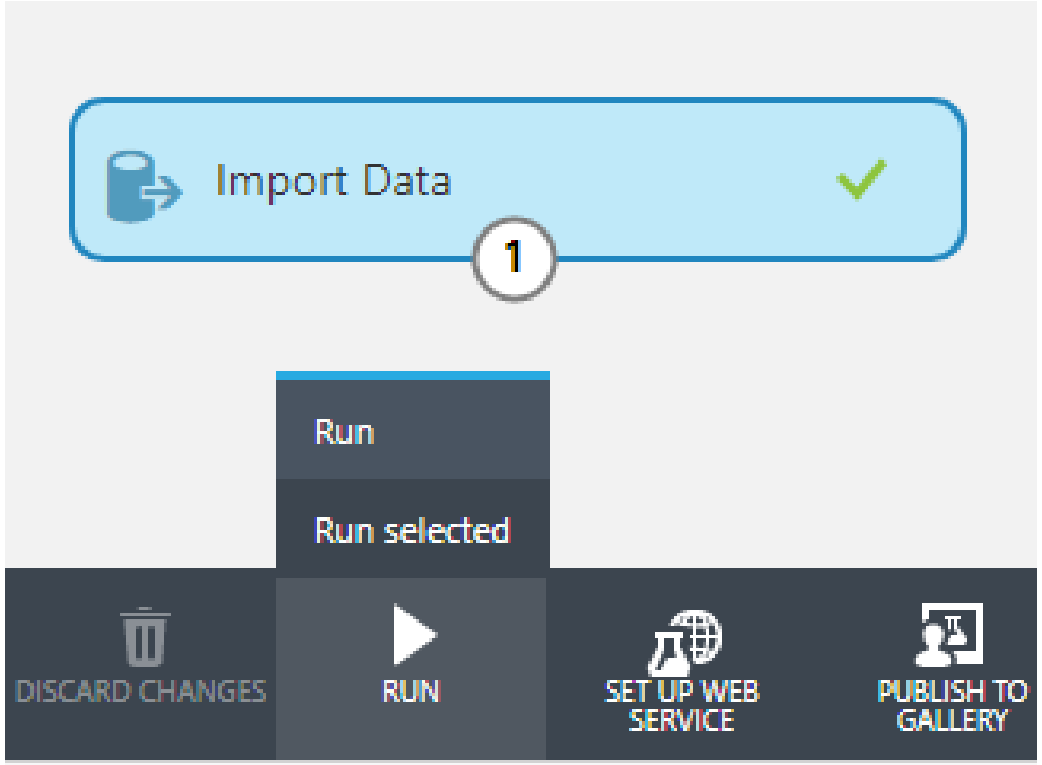
UCI machine learning repository

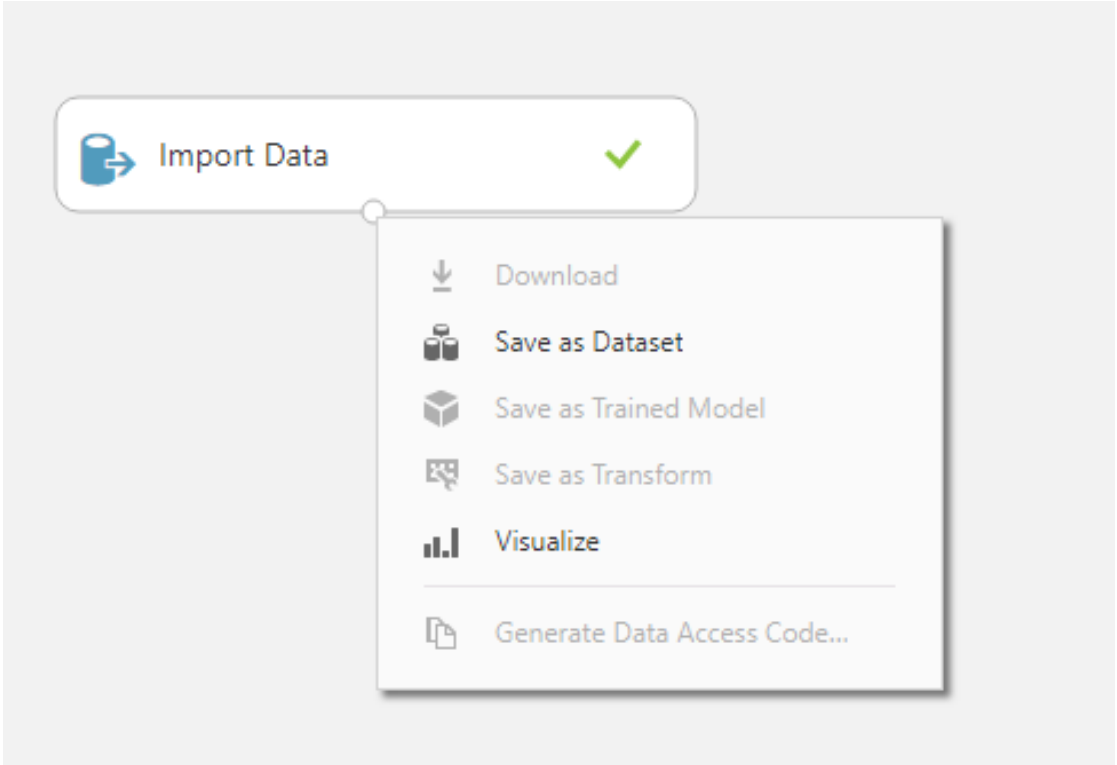
Adresa URL a fisierului este https://archive.ics.uci.edu/ml/machine-learning-databases/tic-tac-toe/tic-tac-toe.data





https://archive.ics.uci.edu/ml/machine-learning-databases/autos/imports-85.data


Import Data
1
Run
Run selected
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Import Data
Download
Save as Dataset
Save as Trained Model
Save as Transform
Visualize
Generate Data Access Code...

Experiment created on 11/12/2020 › Import Data › Results dataset

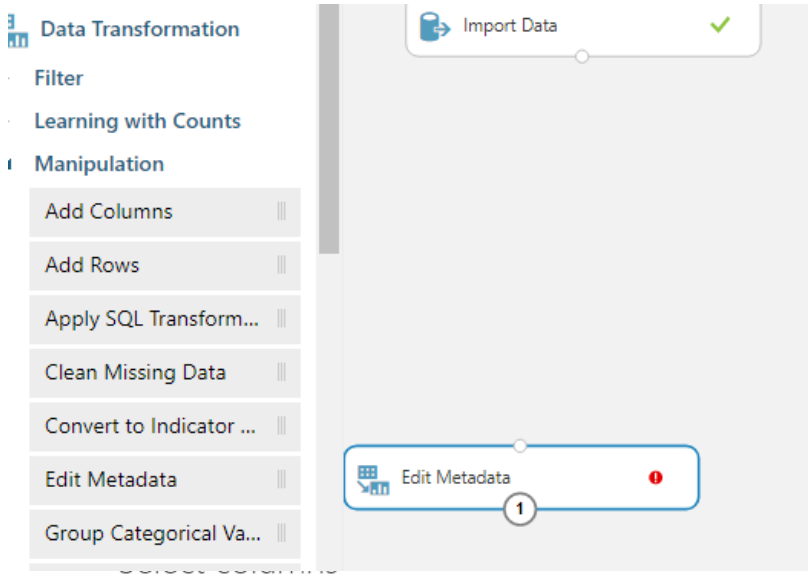
rows 958 columns 10

view as

Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
x	x	x	x	o	o	x	o	o	positive
x	x	x	x	o	o	o	x	o	positive
x	x	x	x	o	o	o	o	x	positive
x	x	x	x	o	o	o	b	b	positive
x	x	x	x	o	o	b	o	b	positive
x	x	x	x	o	o	b	b	o	positive

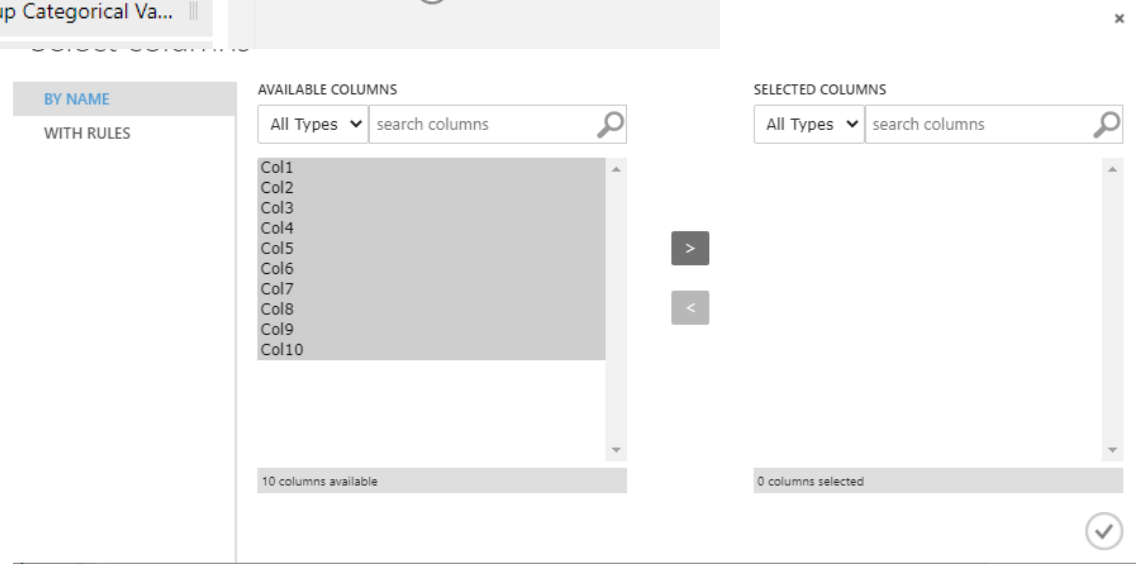
Import data

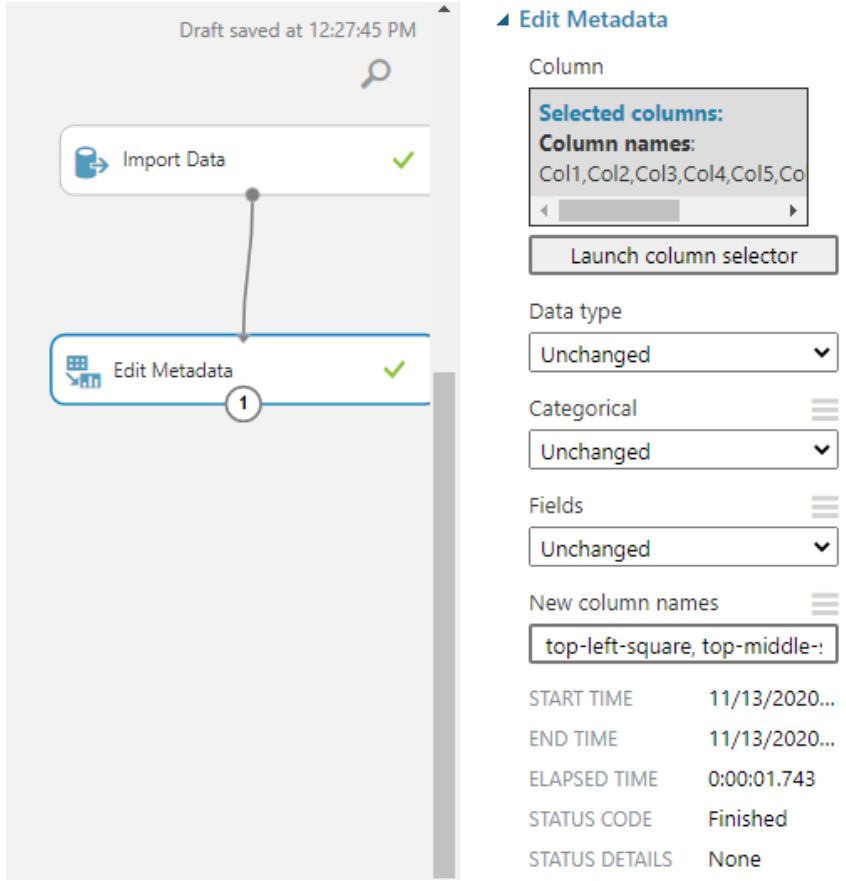
Adaugarea numelor atributelor



Attribute Information:

1. top-left-square: {x,o,b}
2. top-middle-square: {x,o,b}
3. top-right-square: {x,o,b}
4. middle-left-square: {x,o,b}
5. middle-middle-square: {x,o,b}
6. middle-right-square: {x,o,b}
7. bottom-left-square: {x,o,b}
8. bottom-middle-square: {x,o,b}
9. bottom-right-square: {x,o,b}
10. Class: {positive,negative}





https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame

Adaugarea numelor atributelor

https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame

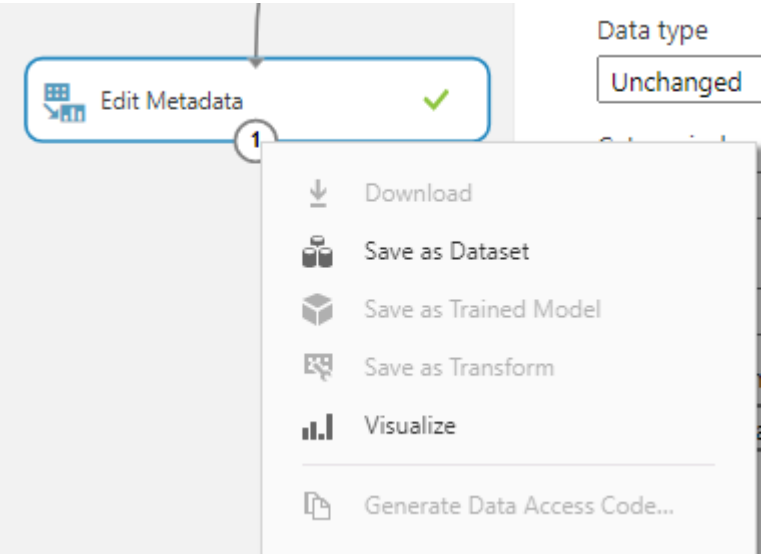
Attribute Information:

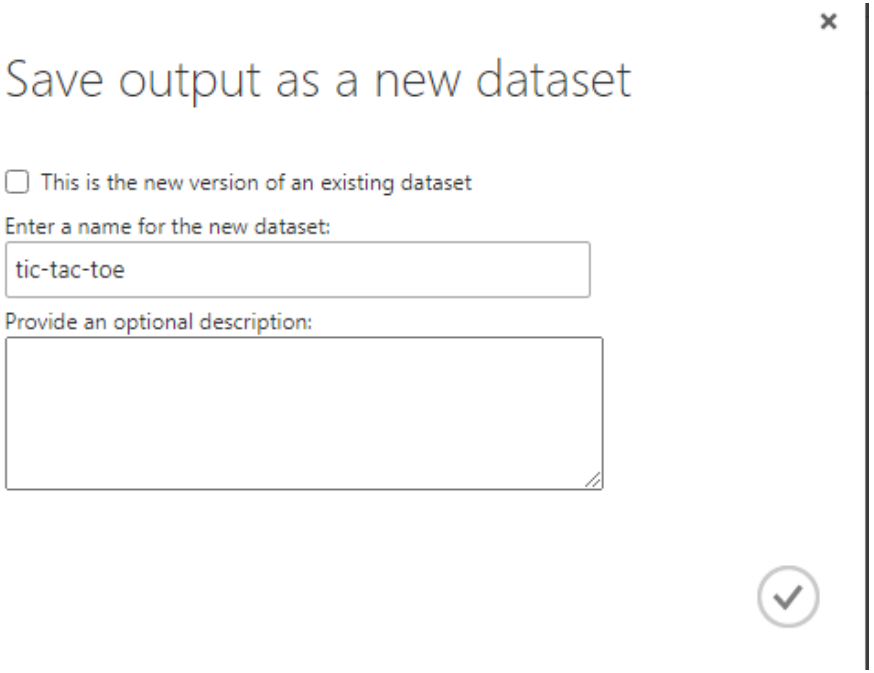
1. top-left-square: {x,o,b}
2. top-middle-square: {x,o,b}
3. top-right-square: {x,o,b}
4. middle-left-square: {x,o,b}
5. middle-middle-square: {x,o,b}
6. middle-right-square: {x,o,b}
7. bottom-left-square: {x,o,b}
8. bottom-middle-square: {x,o,b}
9. bottom-right-square: {x,o,b}
10. Class: {positive,negative}

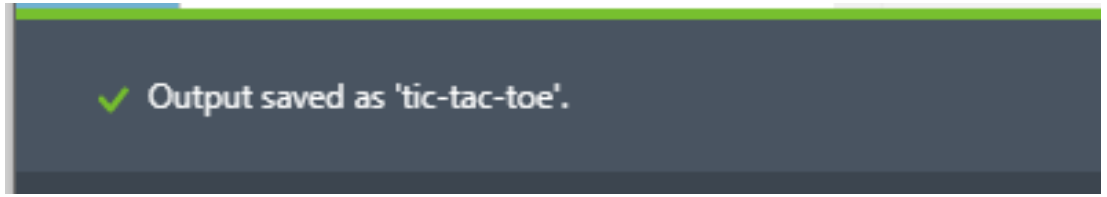
view as

top-left-square	top-middle-square	top-right-square	middle-left-square	middle-middle-square	middle-right-square	left-square
x	x	x	x	o	o	x
x	x	x	x	o	o	o
x	x	x	x	o	o	o
x	x	x	x	o	o	o
x	x	x	x	o	o	b
x	x	x	x	o	o	b
x	x	x	x	o	b	o
x	x	x	x	o	b	o
x	x	x	x	o	b	b
x	x	x	x	b	o	o

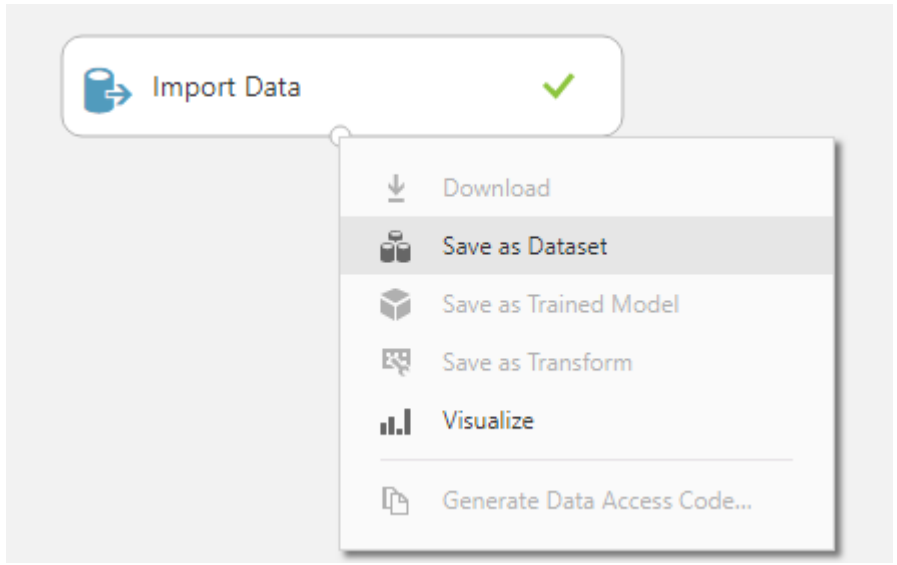
Save data - ca o noua baza de date

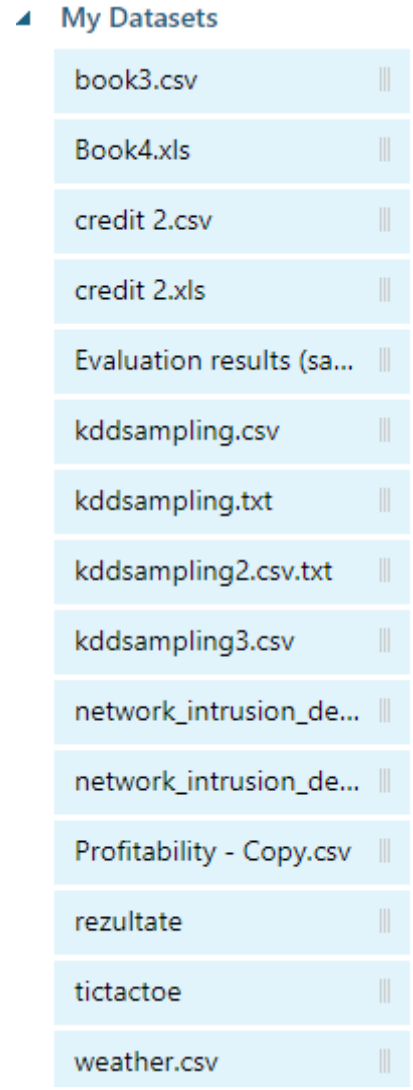


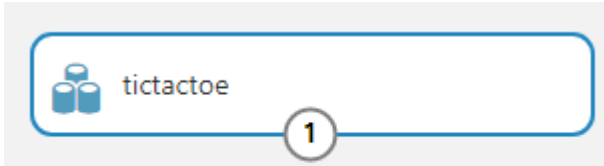




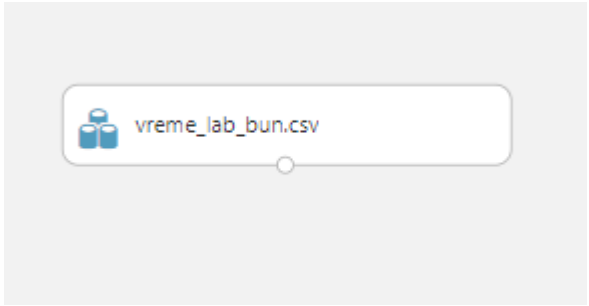
Se regaseste in Saved Datasets - > My Datasets


Import Data
Download
Save as Dataset
Save as Trained Model
Save as Transform
Visualize
Generate Data Access Code...


My Datasets
book3.csv
Book4.xls
credit 2.csv
credit 2.xls
Evaluation results (sa...
kddsampling.csv
kddsampling.txt
kddsampling2.csv.txt
kddsampling3.csv
network_intrusion_de...
network_intrusion_de...
Profitability - Copy.csv
rezultate
tictactoe
weather.csv


tictactoe
1

Incarcare baza de date

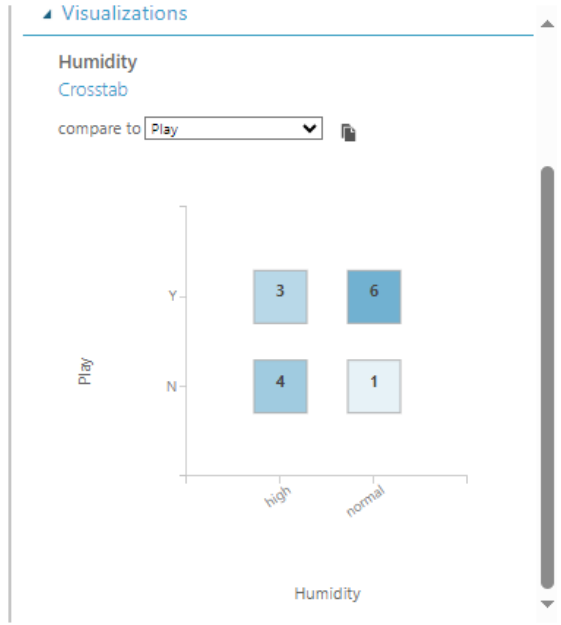


Experiment created on 19/02/2024 › vreme_lab_bun.csv › dataset

rows 14 columns 5

view as

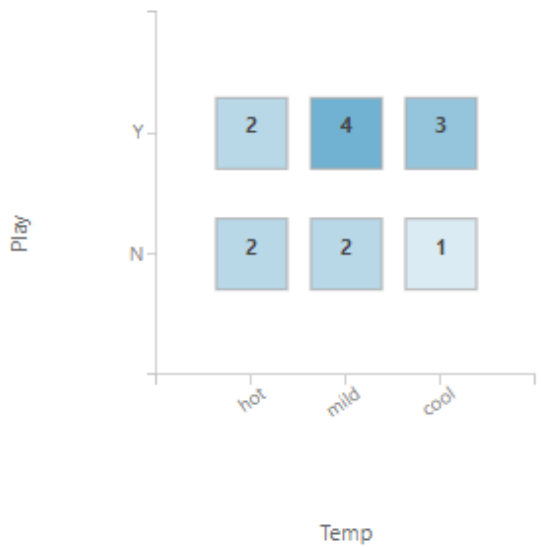
Outlook	Temp	Humidity	Windy	Play
sunny	hot	high	F	N
sunny	hot	high	T	N
overcast	hot	high	F	Y
rain	mild	high	F	Y
rain	cool	normal	F	Y
rain	cool	normal	T	N
overcast	cool	normal	T	Y
sunny	mild	high	F	N
sunny	cool	normal	F	Y
rain	mild	normal	F	Y
sunny	mild	normal	T	Y
overcast	mild	high	T	Y
overcast	hot	normal	F	Y
rain	mild	high	T	N



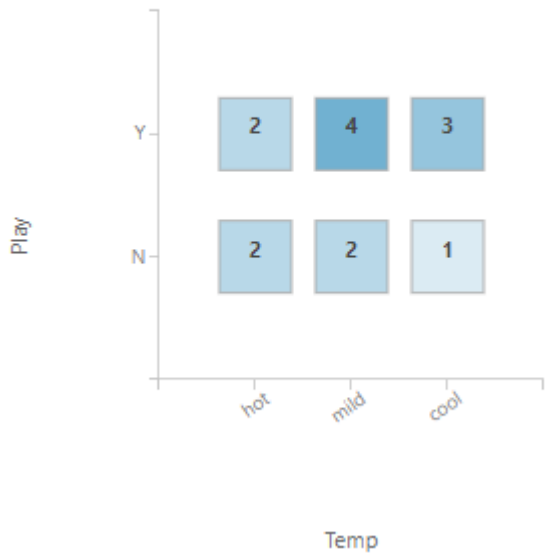
Temp

Crosstab

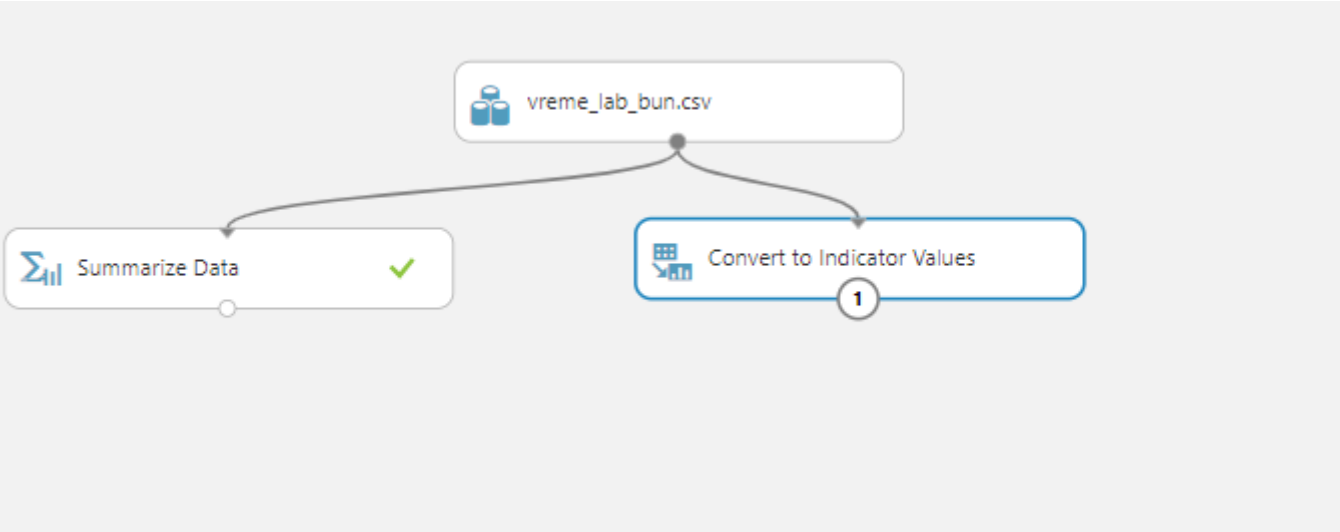
compare to Play



Temp



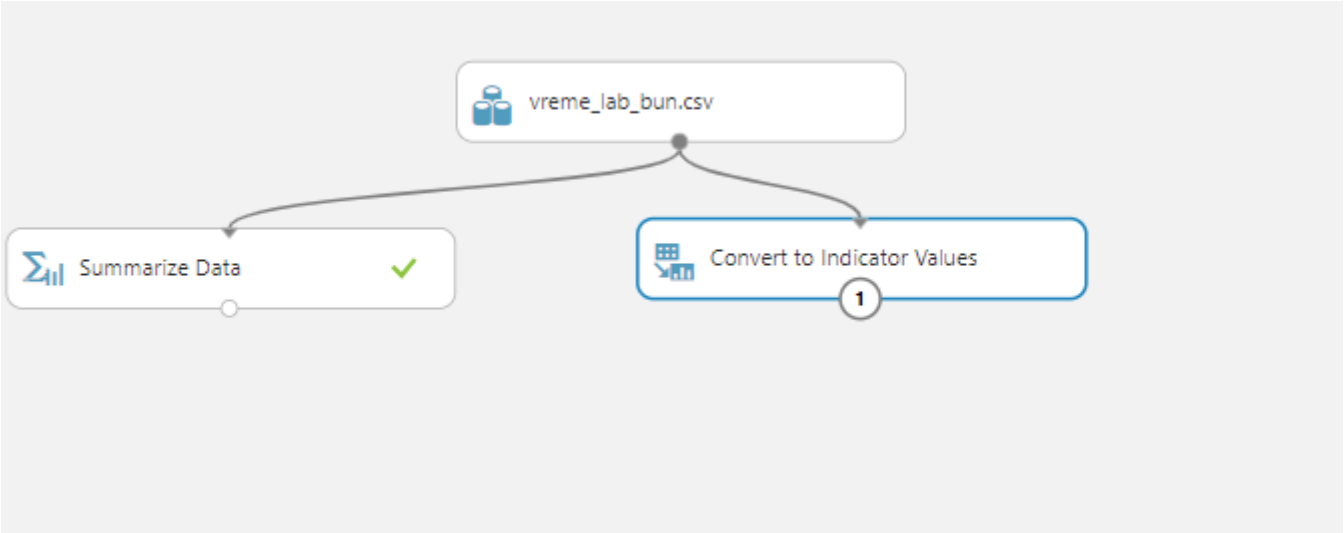
Humidity

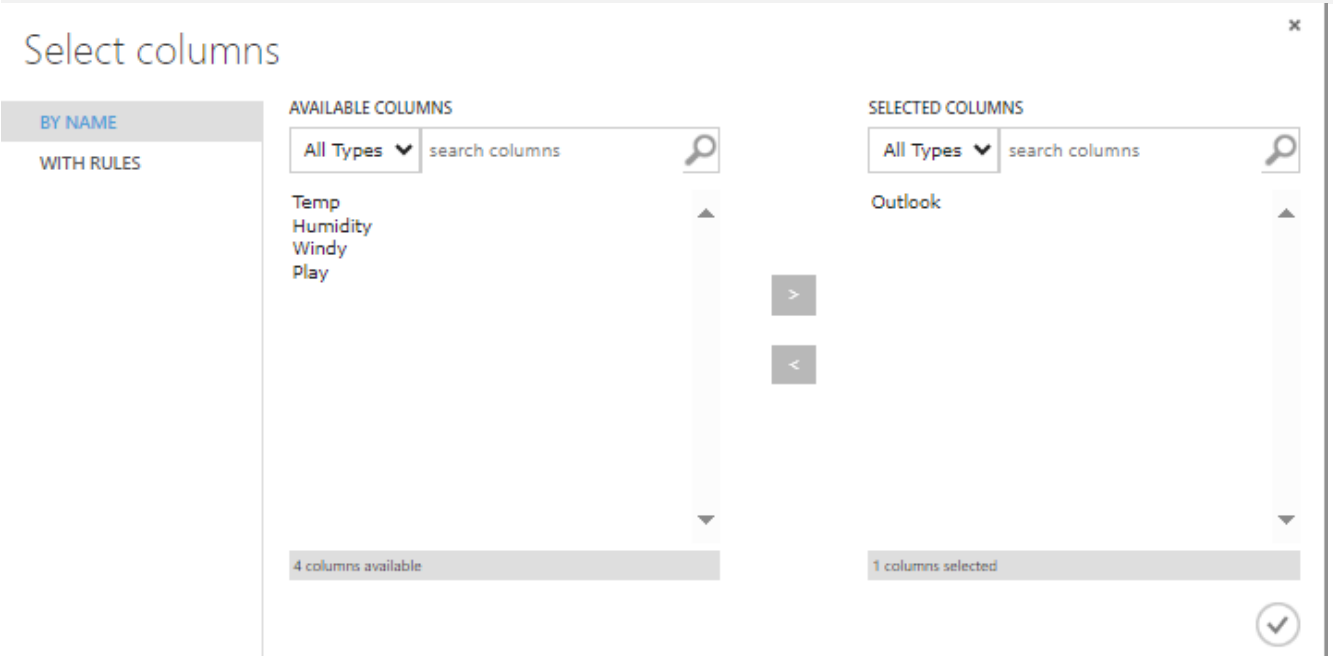


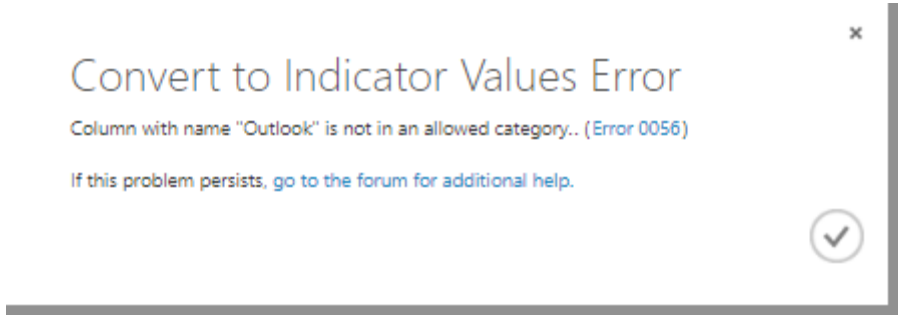
Experiment created on 19/02/2024 › Summarize Data › Results dataset

rows 5 columns 23

Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile	Median	3rd Quartil
Outlook	14	3	0							
Temp	14	3	0							
Humidity	14	2	0							
Windy	14	2	0							
Play	14	2	0							


vreme_lab_bun.csv
Summarize Data
Convert to Indicator Values
1


Select columns
BY NAME
WITH RULES
AVAILABLE COLUMNS
All Types
search columns
Temp
Humidity
Windy
Play
4 columns available
SELECTED COLUMNS
All Types
search columns
Outlook
1 columns selected


Convert to Indicator Values Error
Column with name "Outlook" is not in an allowed category.. (Error 0056)
If this problem persists, go to the forum for additional help.

Outlook	Temp	Humidity	Windy	Play
sunny	hot	high	F	N
sunny	hot	high	T	N
overcast	hot	high	F	Y
rain	mild	high	F	Y
rain	cool	normal	F	Y
rain	cool	normal	T	N
overcast	cool	normal	T	Y
sunny	mild	high	F	N
sunny	cool	normal	F	Y
rain	mild	normal	F	Y
sunny	mild	normal	T	Y
overcast	mild	high	T	Y

Statistics

Unique Values	3
Missing Values	0
Feature Type	String Feature

Visualizations

Outlook

Histogram

compare to None

Properties Project

Convert to Indicator Values

Categorical columns to con...

Selected columns:
Column names:
Outlook

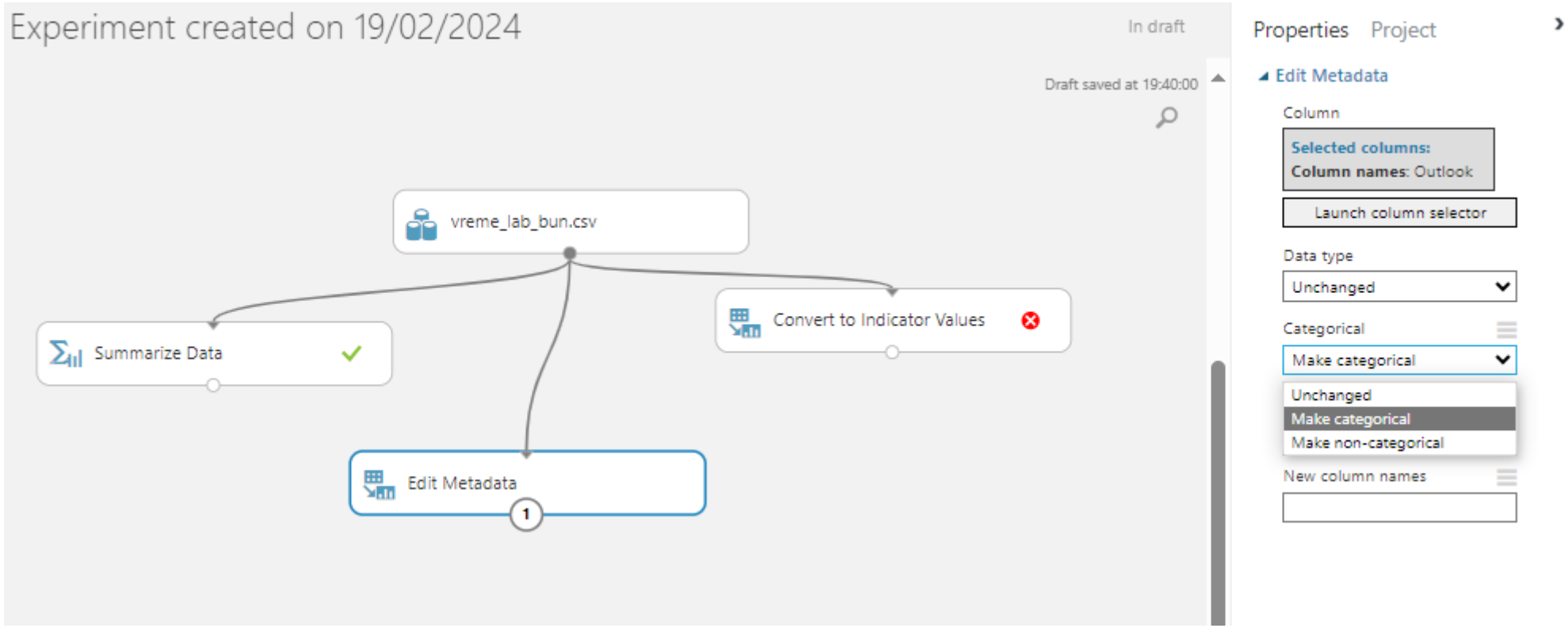
Launch column selector

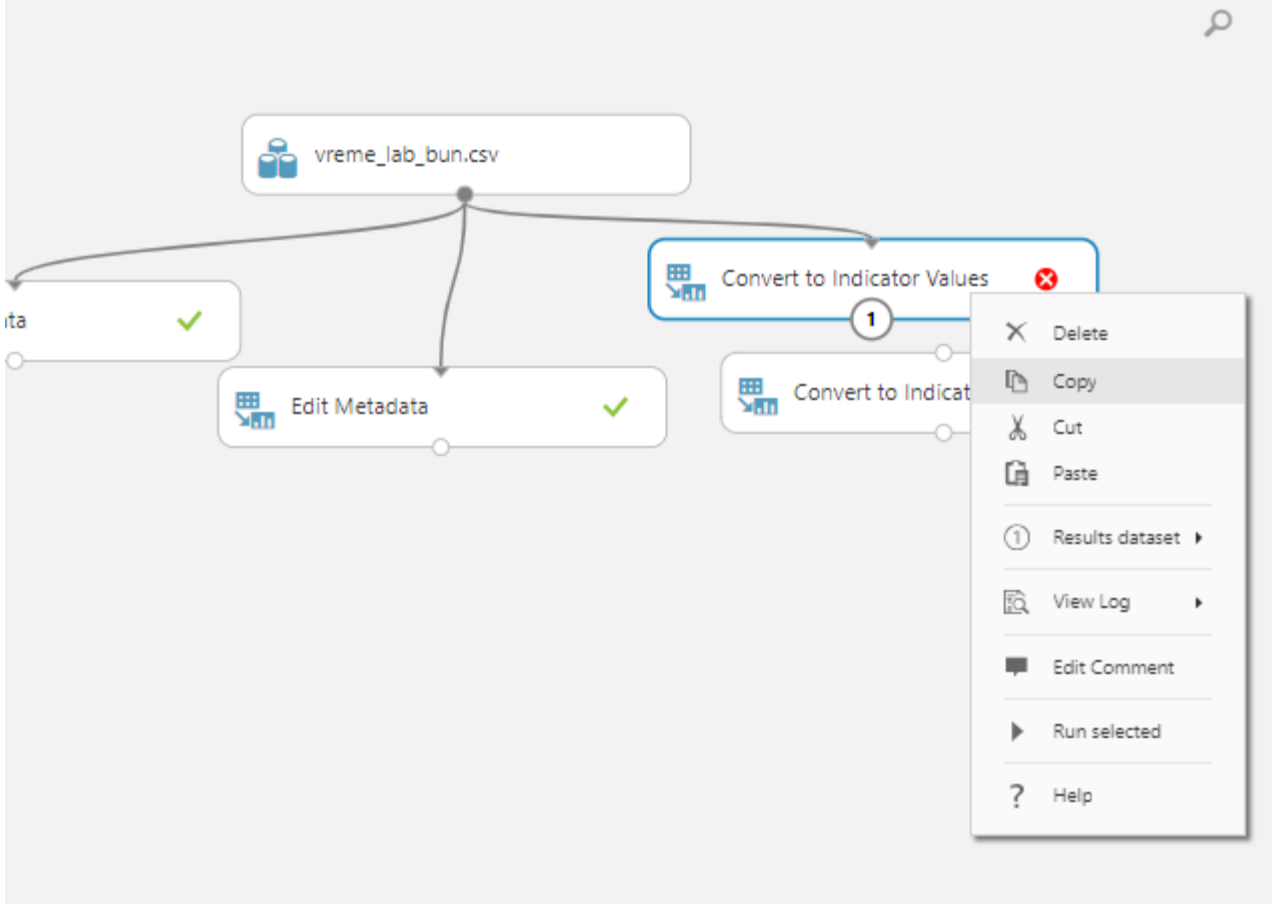
Overwrite categori...

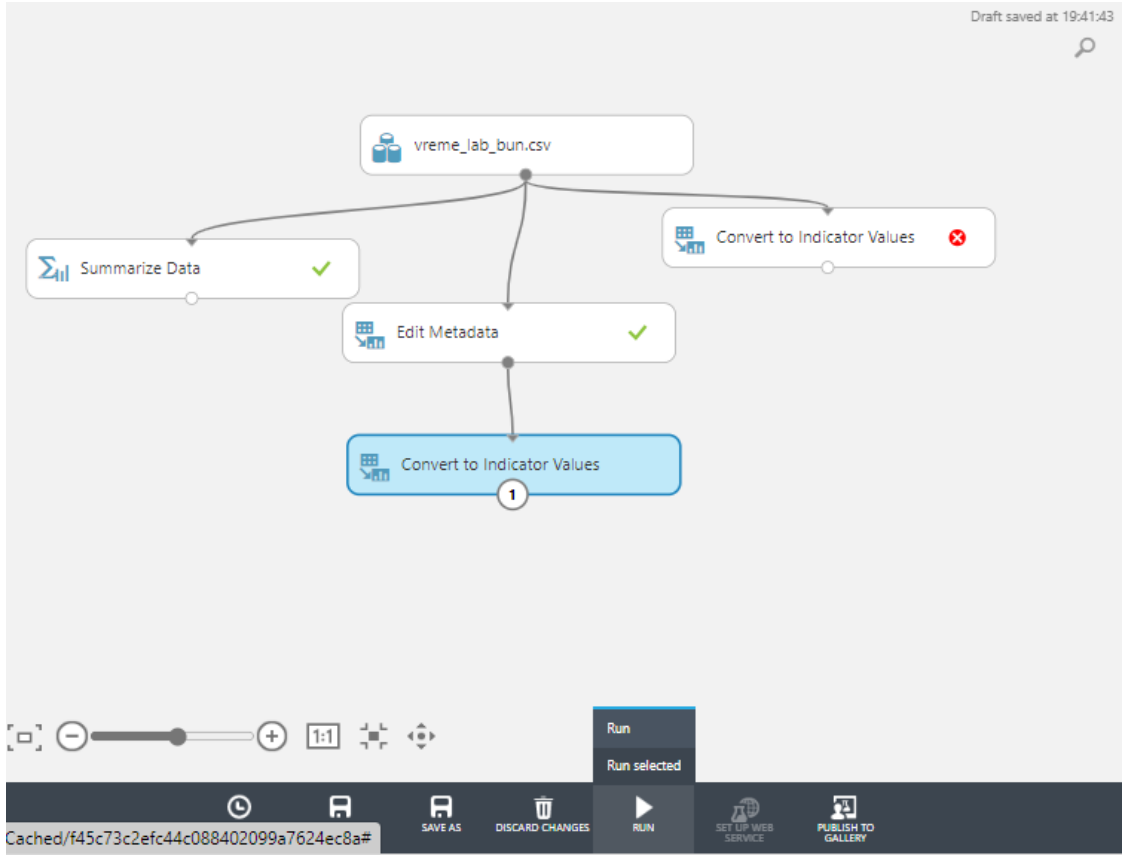
START TIME	2/19/20...
END TIME	2/19/20...
ELAPSED TIME	0:00:03....
STATUS CODE	Failed
STATUS DETAILS	requestId = dd4fb9a... errorCo... taskStat... {"Excepti... {"ErrorId... 0056: Column with name \"Outlo...

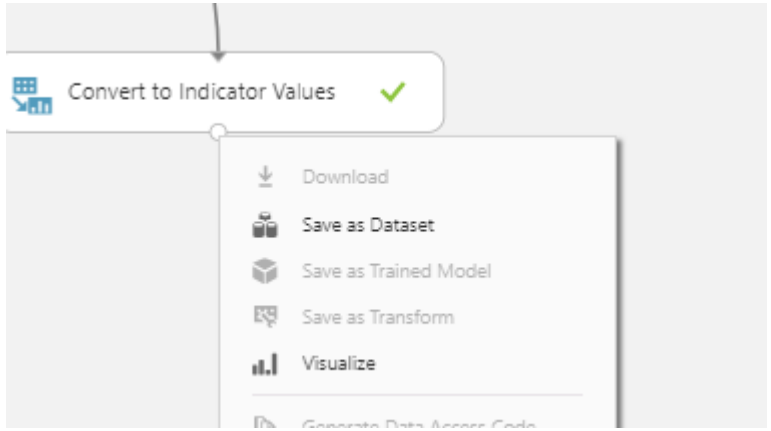
Quick Help

Converts categorical values in columns to indicator values
(more help...)


Experiment created on 19/02/2024
In draft
Draft saved at 19:40:00
vreme_lab_bun.csv
Summarize Data
Convert to Indicator Values
Edit Metadata
1
Properties
Project
Edit Metadata
Column
Selected columns:
Column names: Outlook
Launch column selector
Data type
Unchanged
Categorical
Make categorical
Unchanged
Make categorical
Make non-categorical
New column names


vreme_lab_bun.csv
Convert to Indicator Values
1
Edit Metadata
Convert to Indicat
Delete
Copy
Cut
Paste
Results dataset
View Log
Edit Comment
Run selected
Help


Draft saved at 19:41:43
vreme_lab_bun.csv
Summarize Data
Edit Metadata
Convert to Indicator Values
Convert to Indicator Values
1
1:1
Run
Run selected
Cached/f45c73c2efc44c088402099a7624ec8a#
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY



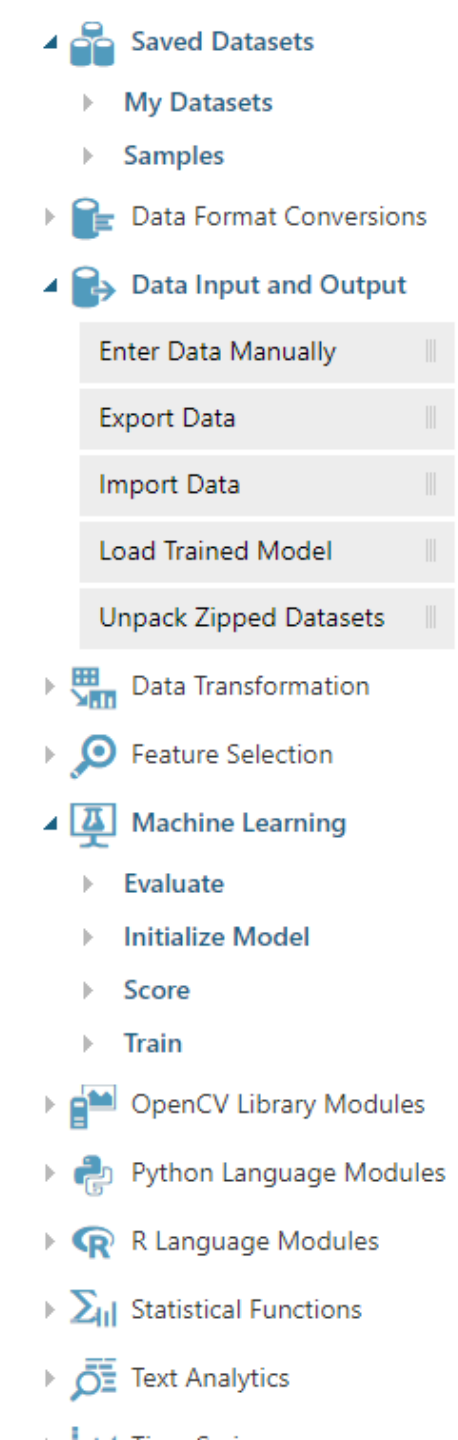
Experiment created on 19/02/2024 › Convert to Indicator Values › Results dataset

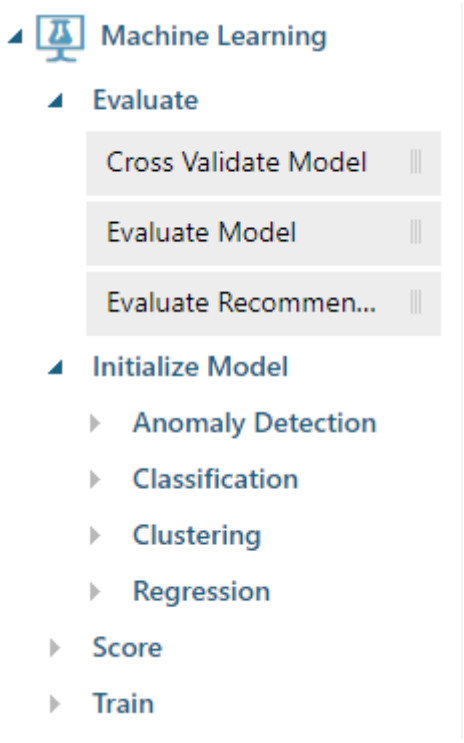
rows 14 columns 8

view as

Outlook	Temp	Humidity	Windy	Play	Outlook-overcast	Outlook-rain	Outlook-sunny
sunny	hot	high	F	N	0	0	1
sunny	hot	high	T	N	0	0	1
overcast	hot	high	F	Y	1	0	0
rain	mild	high	F	Y	0	1	0
rain	cool	normal	F	Y	0	1	0
rain	cool	normal	T	N	0	1	0
overcast	cool	normal	T	Y	1	0	0
sunny	mild	high	F	N	0	0	1
sunny	cool	normal	F	Y	0	0	1
rain	mild	normal	F	Y	0	1	0
sunny	mild	normal	T	Y	0	0	1
overcast	mild	high	T	Y	1	0	0
overcast	hot	normal	F	Y	1	0	0
rain	mild	high	T	N	0	1	0

Componentele unui experiment ML Studio (clasic)





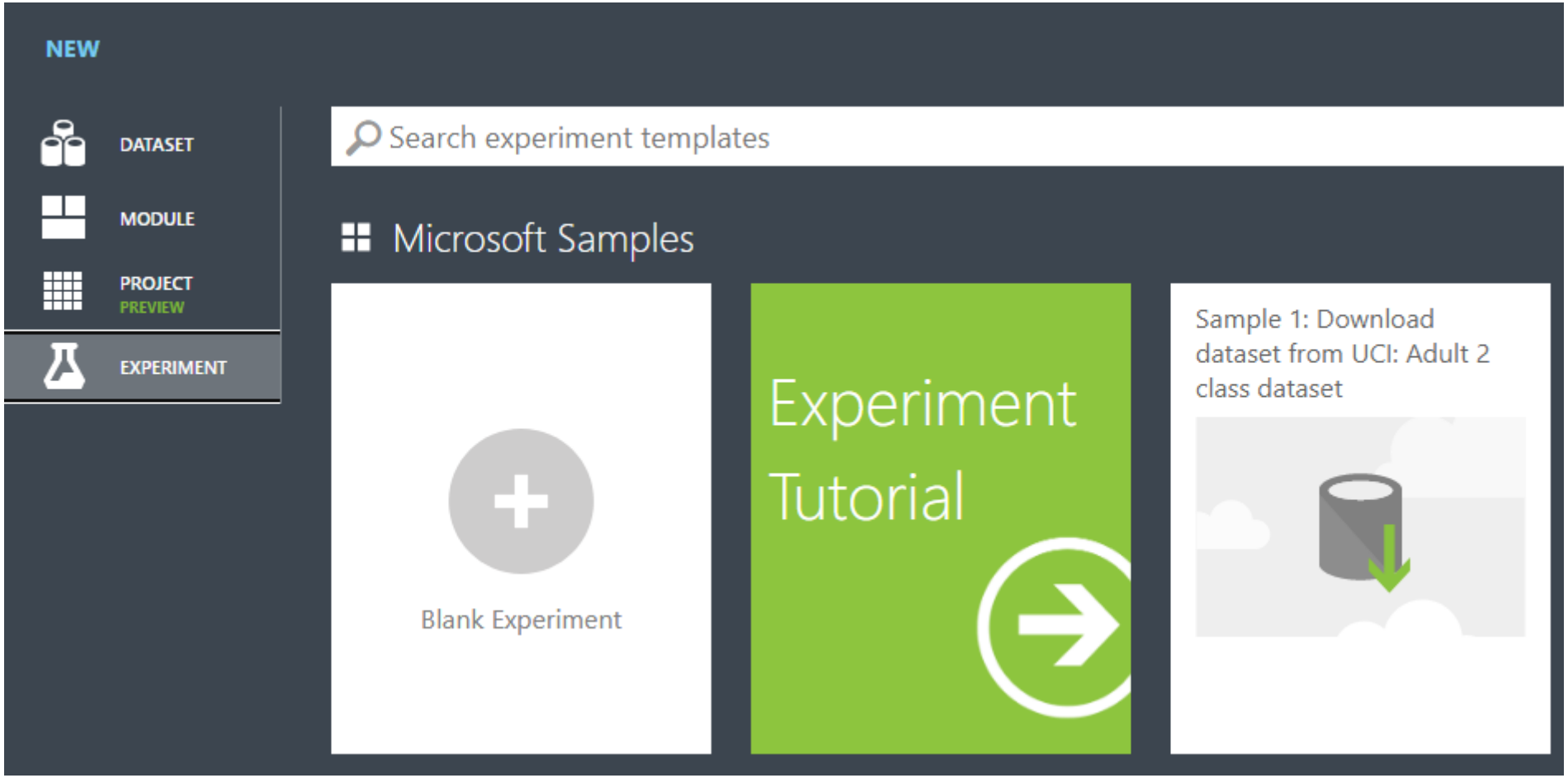
Modulele (Modules)

Un modul este un algoritm care poate fi aplicat unor date.

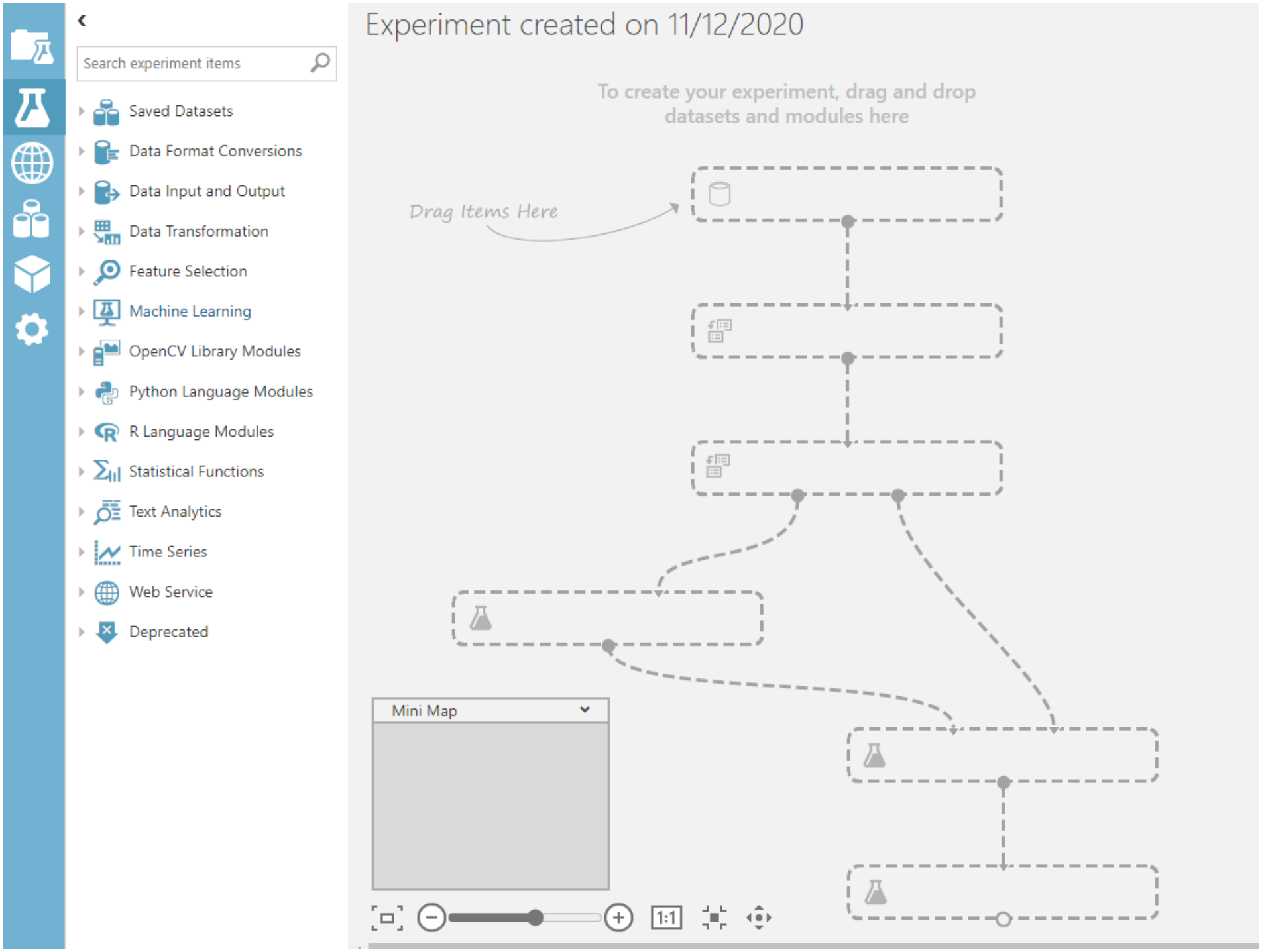
Machine Learning Studio (clasic) ofera o serie de module de tip built-in, de la module de pregatire a datelor, de implementare a diverselor operatii specifice (clasificare, clusterizare, etc.) la cele de validare a rezultatelor.

Lista de module disponibile se gaseste in partea stanga a suprafetei interactive si poate fi accesata f usor.

Experiment - new



Blank Experiment



Experiment – tutorial

+ New -> Experiment->Experiment tutorial

Urmariti experimental.

Unpack Zipped Datasets

Bibliografie

https://docs.microsoft.com/en-us/azure/machine-learning/classic/studio-classic-overview

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/