Azure Machine Learning Studio

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/

(classic)

Welcome to Azure Machine Learning Studio (classic)

Try it for free

No Azure subscription? No credit card? No problem! Choose anonymous Guest Access, or sign in with your work or school account, or a Microsoft account.

Sign In

Not an Azure ML Studio (classic) user?
Sign up here

Pricing & FAQ

By using this free version, you agree to be bound by the Microsoft Azure Website Terms of Use.

Trebuie sa aveti un cont Microsoft. Puteti folosi contul dvs de Office 365 cont sau orice alt cont (hotmail). Daca nu aveti unul, va puteti crea unul la https://signup.live.com/

Sign Up pentru un cont Free Azure ML Workspace

https://studio.azureml.net/

Click Get Started Now. Urmati instructiunile pentru inregistrarea unui free Azure ML workspace.

Spațiu de lucru (Workspace) interactiv studio (clasic)

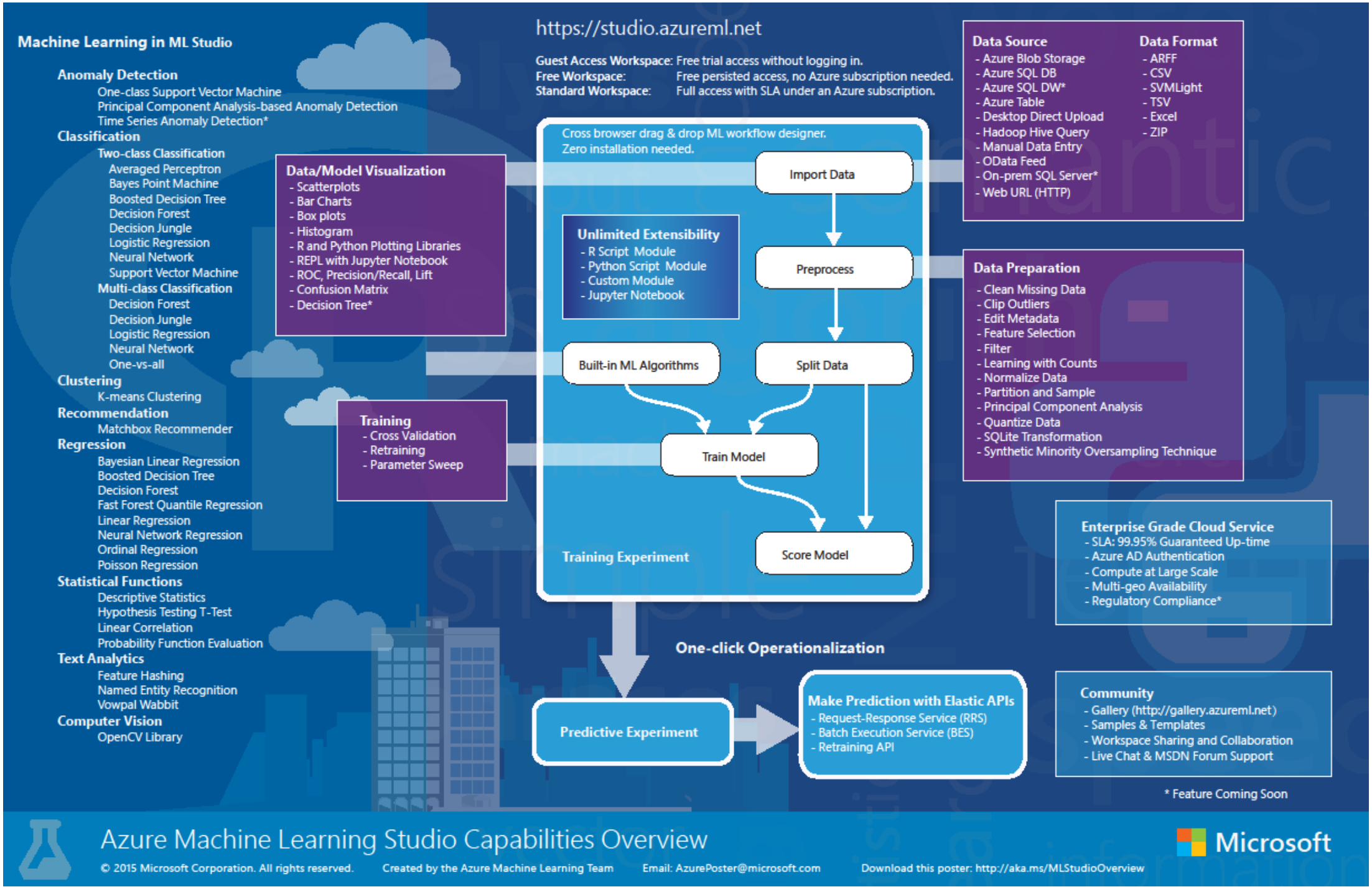
Machine Learning Studio (clasic) oferă un spațiu de lucru interactiv, vizual pentru a construi, testa și itera cu ușurință pe un model de analiză predictivă.

Seturile de date si modulele se pot incarca prin drag-and-drop pe o suprafata interactiva , conectandu-se pentru a forma un **experiment**, care apoi se poate rula. In functie de rezultatele obtinute se pot aduce imbunatatiri la experiment si rula pana cand acesta ajunge la forma finala.

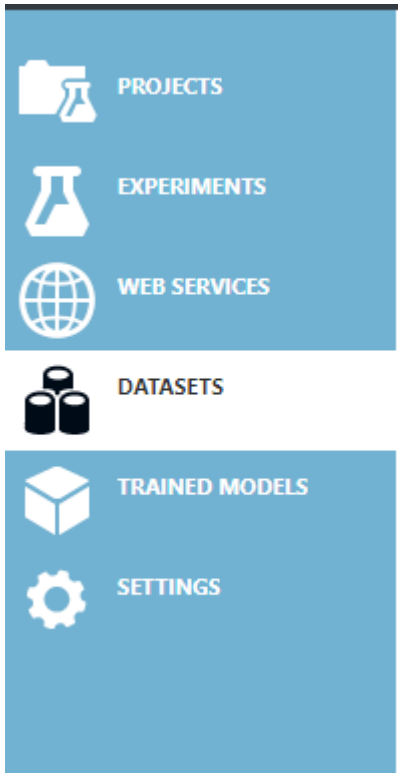
Cand modelul de analiza predictiva a fost finalizat, experimentul de instruire se poate converti într-un experiment predictiv si se poate publica pe web (ca web service), astfel încât modelul să poată fi accesat de alte persoane.

Seturile de date si modulele se conecteaza visual, fara a fi necesara scrierea de cod.

Microsoft ML Studio (classic) Capabilities Overview diagram



Componentele unui experiment ML Studio (clasic)

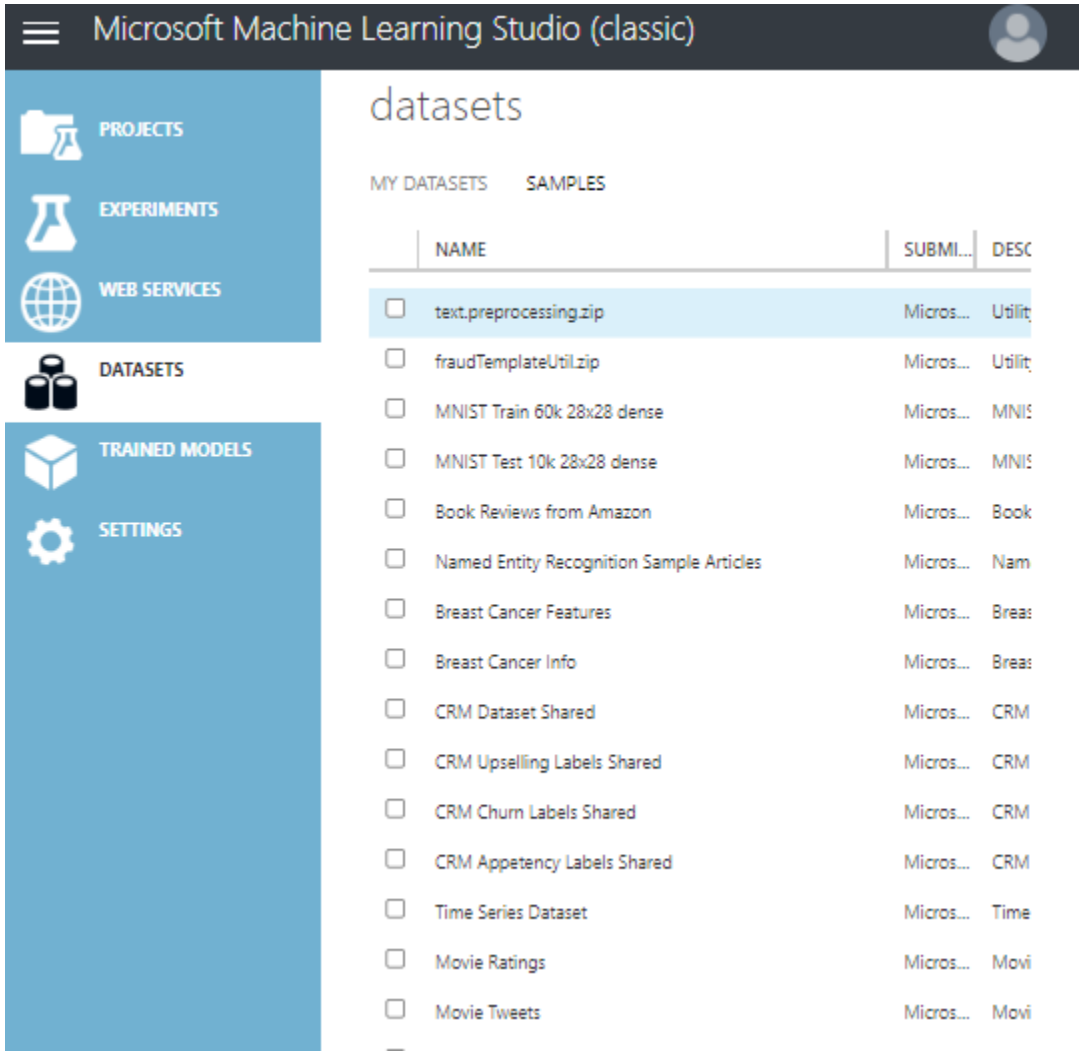


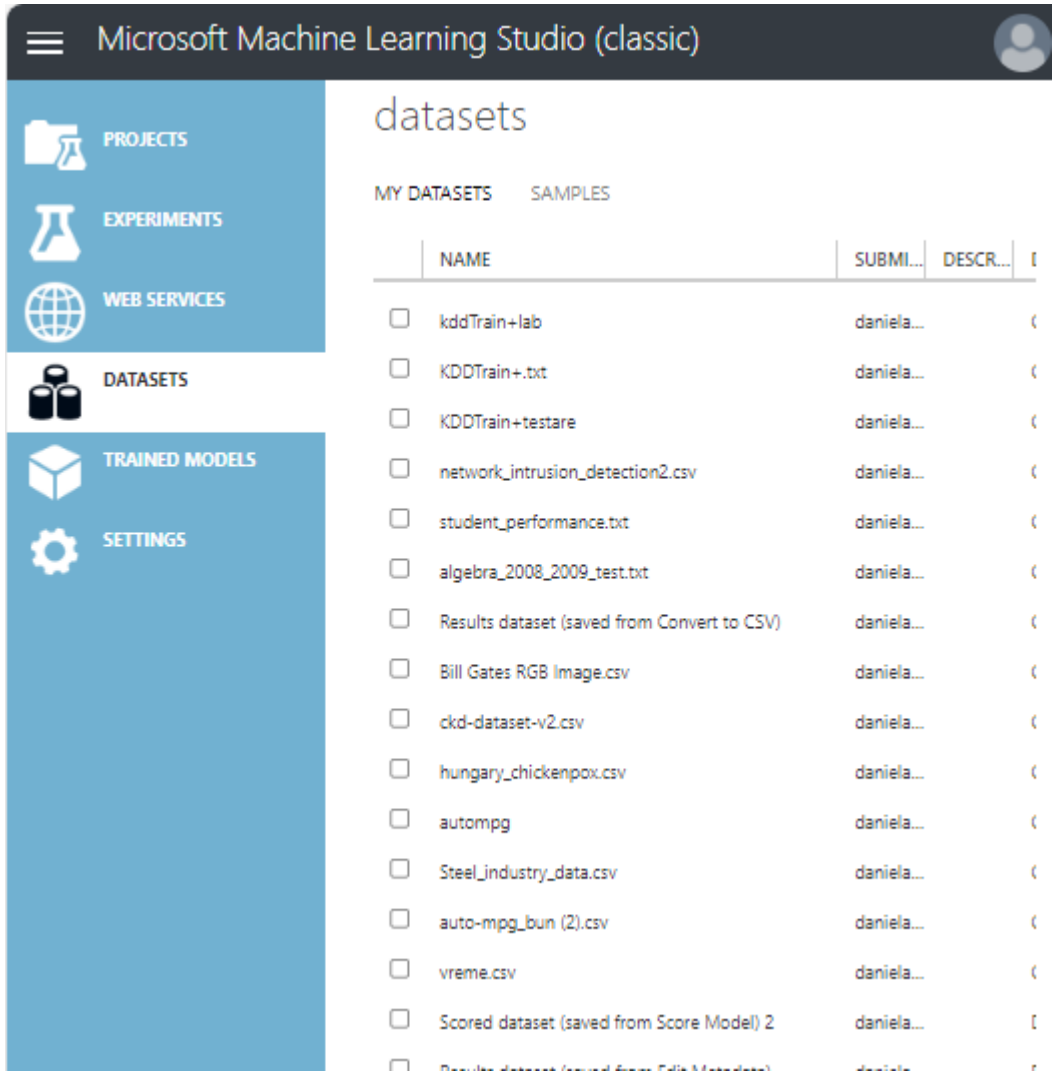
Seturile de date (Datasets)

Un dataset este un set de date care a fost încărcat în Machine Learning Studio (clasic), astfel încât să poată fi utilizat în procesul de modelare.

Un număr de seturi de date eșantion sunt incluse în Machine Learning Studio (clasic) pentru a experimenta dar se pot încărca si alte seturi de date, atunci cand este nevoie.

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

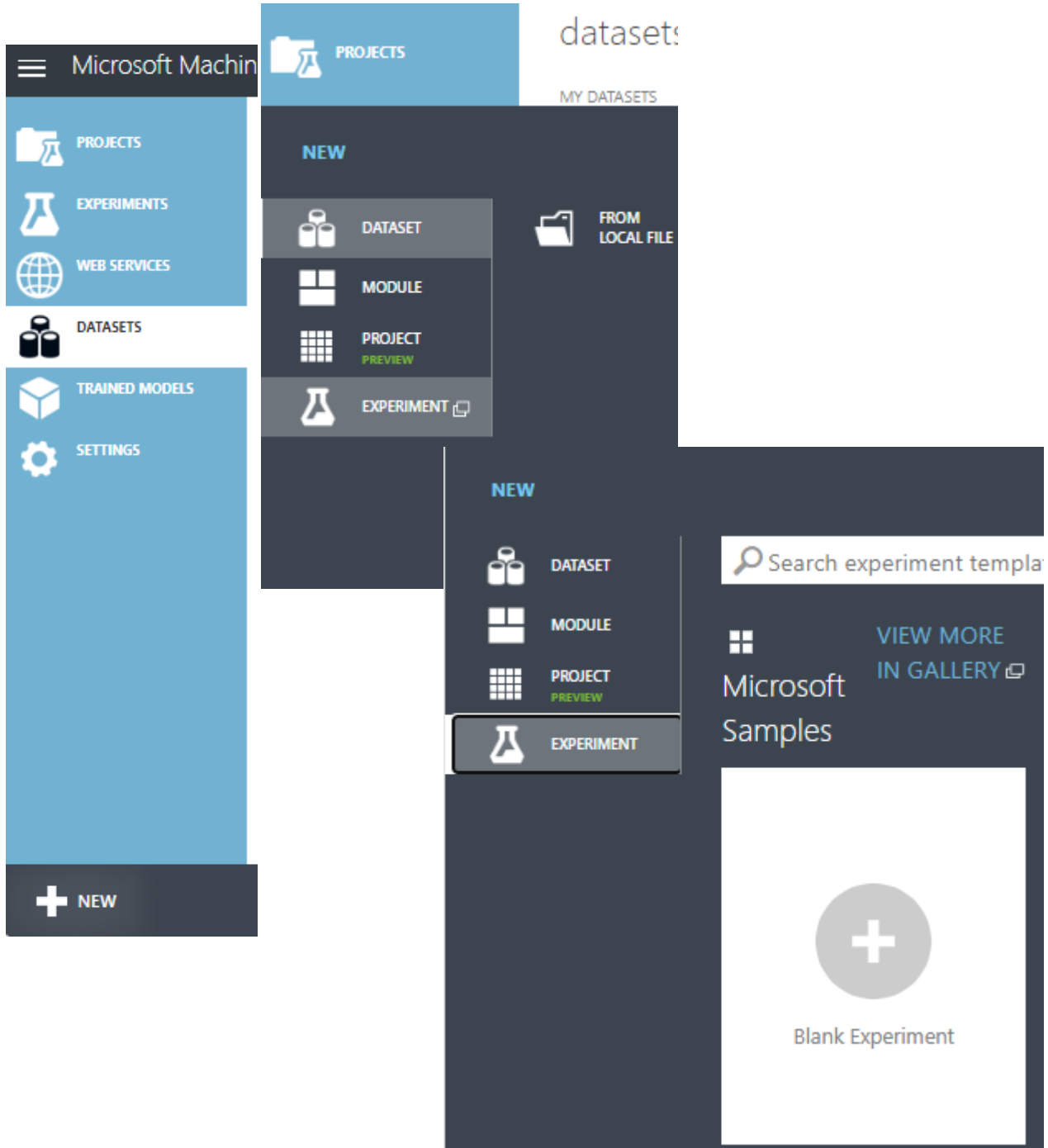

Microsoft Machine Learning Studio (classic)
PROJECTS
EXPERIMENTS
WEB SERVICES
DATASETS
TRAINED MODELS
SETTINGS
datasets
MY DATASETS
SAMPLES
NAME
SUBMI...
DESC
text.preprocessing.zip
Micros...
Utilit
fraudTemplateUtil.zip
Micros...
Utilit
MNIST Train 60k 28x28 dense
Micros...
MNIS
MNIST Test 10k 28x28 dense
Micros...
MNIS
Book Reviews from Amazon
Micros...
Book
Named Entity Recognition Sample Articles
Micros...
Nam
Breast Cancer Features
Micros...
Breas
Breast Cancer Info
Micros...
Breas
CRM Dataset Shared
Micros...
CRM
CRM Upselling Labels Shared
Micros...
CRM
CRM Churn Labels Shared
Micros...
CRM
CRM Appetency Labels Shared
Micros...
CRM
Time Series Dataset
Micros...
Time
Movie Ratings
Micros...
Movi
Movie Tweets
Micros...
Movi

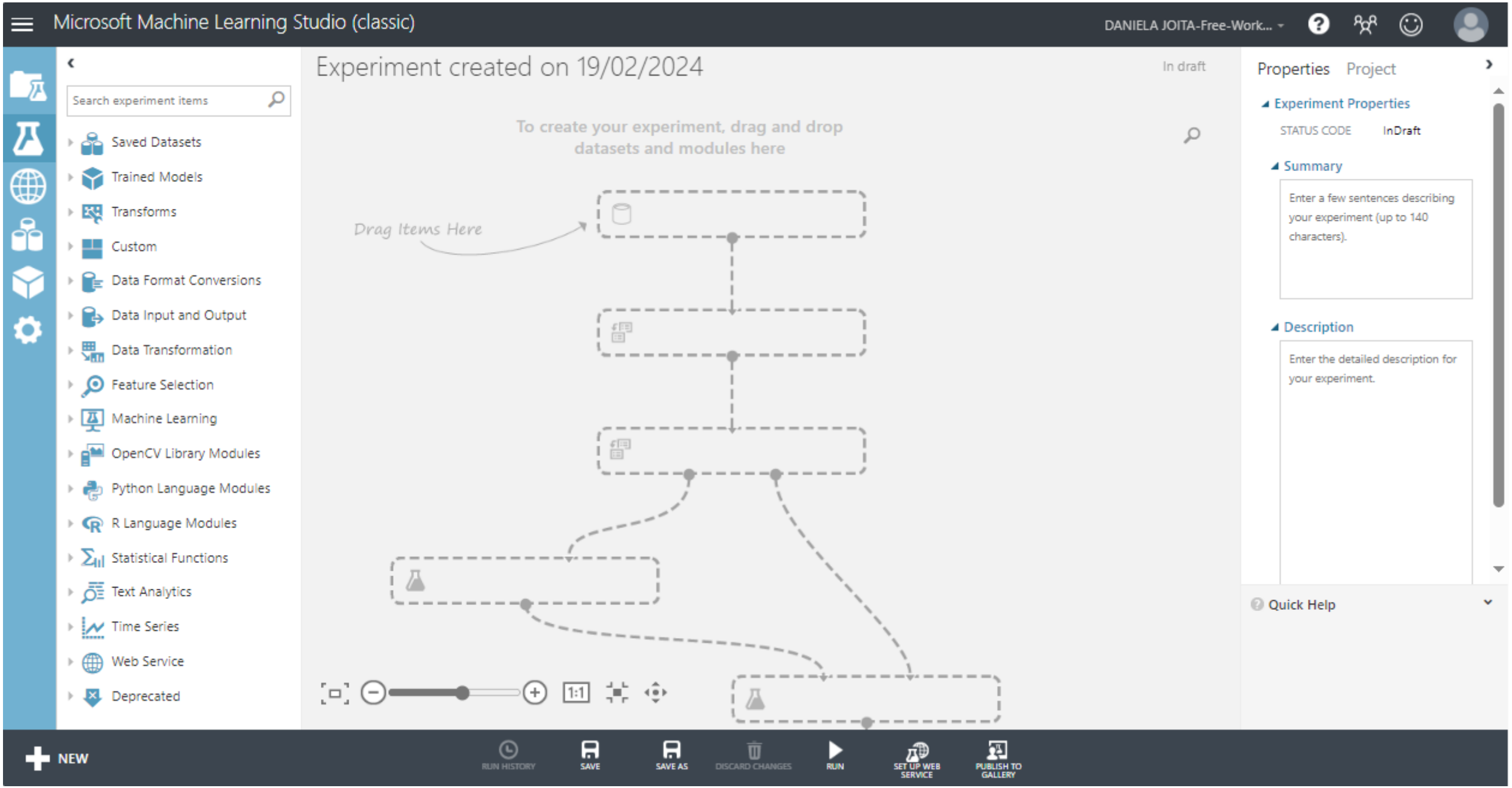

Microsoft Machine Learning Studio (classic)
PROJECTS
EXPERIMENTS
WEB SERVICES
DATASETS
TRAINED MODELS
SETTINGS
datasets
MY DATASETS
SAMPLES
NAME
SUBMI...
DESCR...
kddTrain+lab
daniela...
KDDTrain+.txt
daniela...
KDDTrain+testare
daniela...
network_intrusion_detection2.csv
daniela...
student_performance.txt
daniela...
algebra_2008_2009_test.txt
daniela...
Results dataset (saved from Convert to CSV)
daniela...
Bill Gates RGB Image.csv
daniela...
ckd-dataset-v2.csv
daniela...
hungary_chickenpox.csv
daniela...
autompg
daniela...
Steel_industry_data.csv
daniela...
auto-mpg_bun (2).csv
daniela...
vreme.csv
daniela...
Scored dataset (saved from Score Model) 2
daniela...

Crearea unui model

Crearea unui nou experiment facand click pe **+NEW** din partea de jos stanga a ferestrei Machine Learning Studio, selectarea **EXPERIMENT** si apoi selectarea **Blank Experiment**.

Acesta primeste un nume din oficiu. Poate fi redenumit.

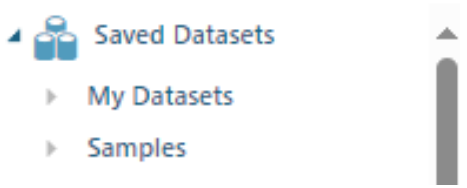



Microsoft Machine Learning Studio (classic)
DANIELA JOITA-Free-Work...
Experiment created on 19/02/2024
In draft
Search experiment items
Saved Datasets
Trained Models
Transforms
Custom
Data Format Conversions
Data Input and Output
Data Transformation
Feature Selection
Machine Learning
OpenCV Library Modules
Python Language Modules
R Language Modules
Statistical Functions
Text Analytics
Time Series
Web Service
Deprecated
To create your experiment, drag and drop datasets and modules here
Drag Items Here
Properties
Project
Experiment Properties
STATUS CODE
InDraft
Summary
Enter a few sentences describing your experiment (up to 140 characters).
Description
Enter the detailed description for your experiment.
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

Alegerea bazei de date

Pot fi alese bd deja incarcate in Azure ML sau bd de pe calculatorul local.

- Saved datasets
  - *My datasets*
  - *Samples*
- Local File
  - *Click pe **+NEW** din partea de jos stanga a ferestrei Machine Learning Studio, selecteaza Datasets->from local file. Se selecteaza baza de date dorita.*



In stanga interfetei noului experiment este o paleta de date si module.

Alegerea bazei de date

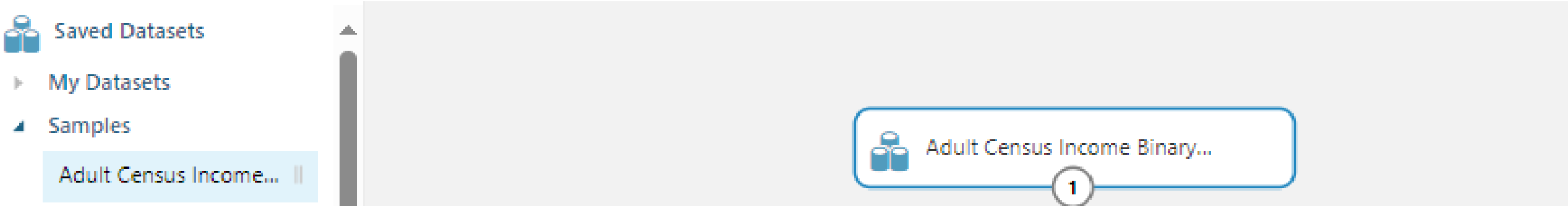
Din Saved datasets -> my datasets, selectam

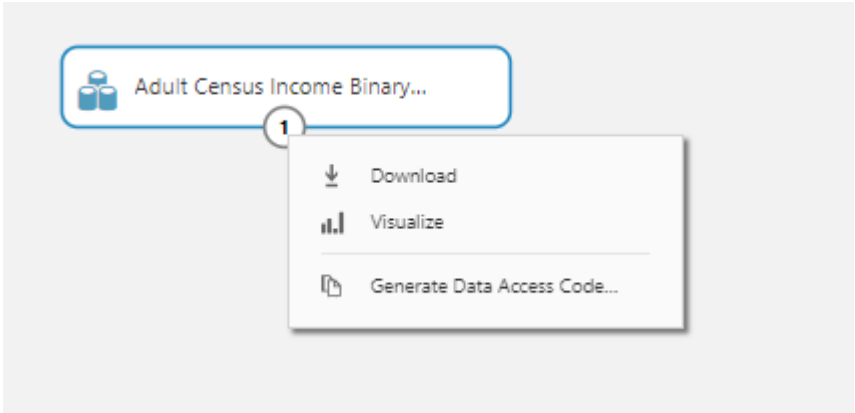
baza de date aleasa. Drag and drop baza de date.

Bazele de date si modulele au I/O porturi reprezentate prin cerculete, de intrare in partea de sus si de iesire in partea de jos.

Pentru crearea unei flow chart vom conecta un port de iesire cu unul de intrare.

Prin click la un port de iesire se poate vizualiza baza de date asa cum arata ea in acel moment.





Experiment created on 19/02/2024 › Adult Census Income Binary Classification dataset › dataset

rows 32561 columns 15

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race
view as									
	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White
	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White
	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White
	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black
	28	Private	338409	Bachelors	13	Married-civ-	Prof-	Wife	Black

Statistics

Visualizations

To view, select a column in the table.

Baza de date

Adult Census Income Binary Classification

[Home - UCI Machine Learning Repository]

[Census Income - UCI Machine Learning Repository]

Census Income

Donated on 4/30/1996

Predict whether income exceeds $50K/yr based on census data. Also known as Adult dataset.

Dataset Characteristics	Subject Area	Associated Tasks
Multivariate	Social Science	Classification
Feature Type	**# Instances**	**# Features**
Categorical, Integer	48842	14

Dataset Information

Additional Information

Extraction was done by Barry Becker from the 1994 Census database. A set of reasonably clean records was extracted using the following conditions: ((AAGE>16) && (AGI>100) && (AFNLWGT>1)&& (HRSWK>0))

Prediction task is to determine whether a person makes over 50K a year.

Atribute

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
age	Feature	Integer	Age	N/A		no
workclass	Feature	Categorical	Income	Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.		yes
fnlwgt	Feature	Integer				no
education	Feature	Categorical	Education Level	Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.		no
education-num	Feature	Integer	Education Level			no

Etc.

Experiment created on 19/02/2024 › Adult Census Income Binary Classification dataset › dataset

rows 32561 columns 15

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race
view as									
	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White
	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White
	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White
	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black
	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black
	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White
	49	Private	160187	9th	5	Married-spouse-	Other-	Not-in-	Black

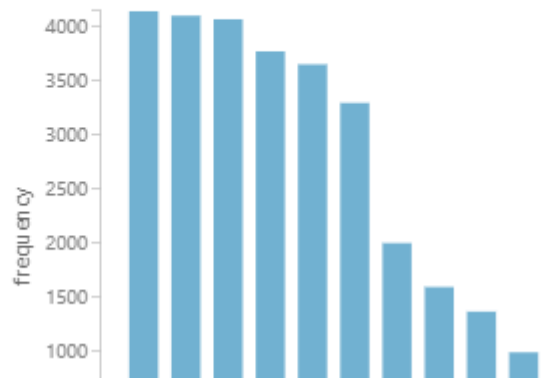
Statistics

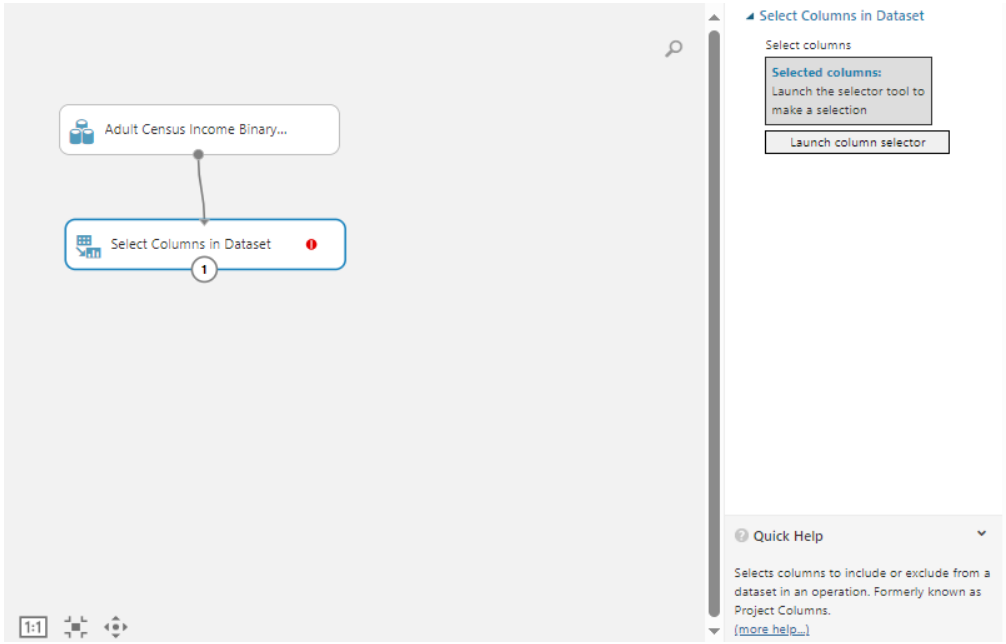
Unique Values	14
Missing Values	1843
Feature Type	String Feature

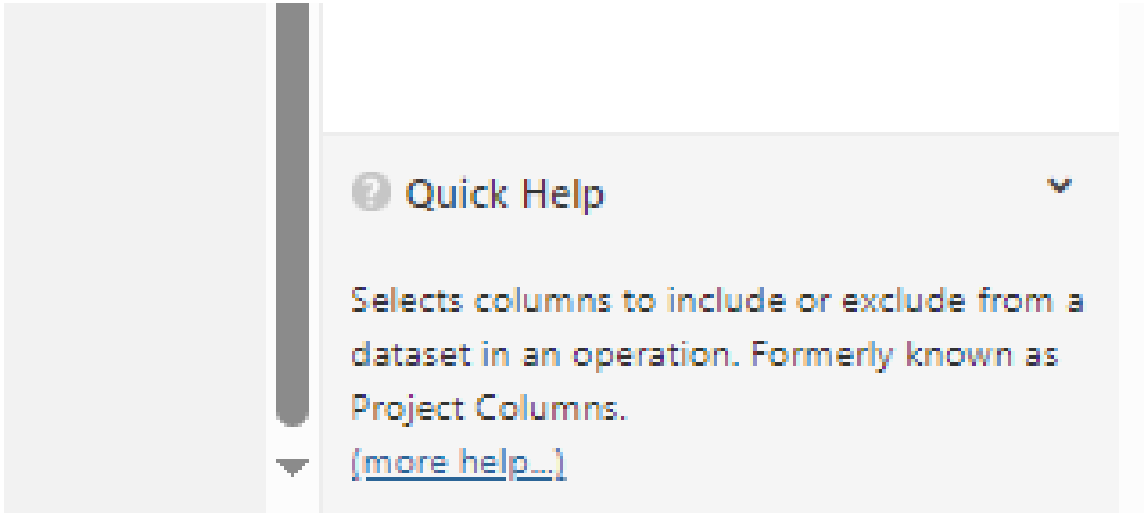
Visualizations

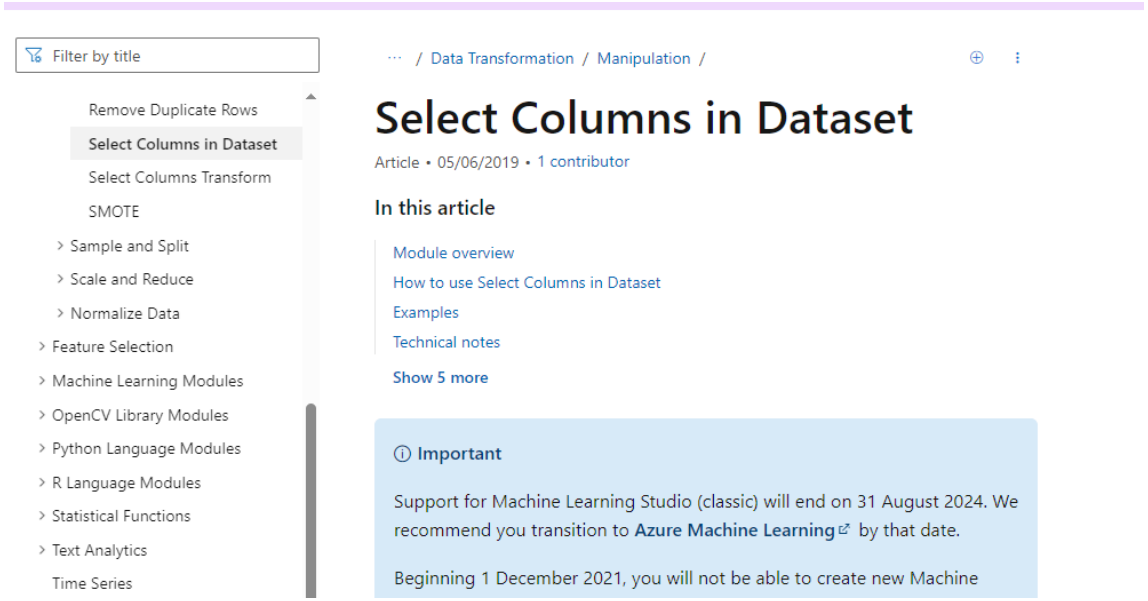
occupation

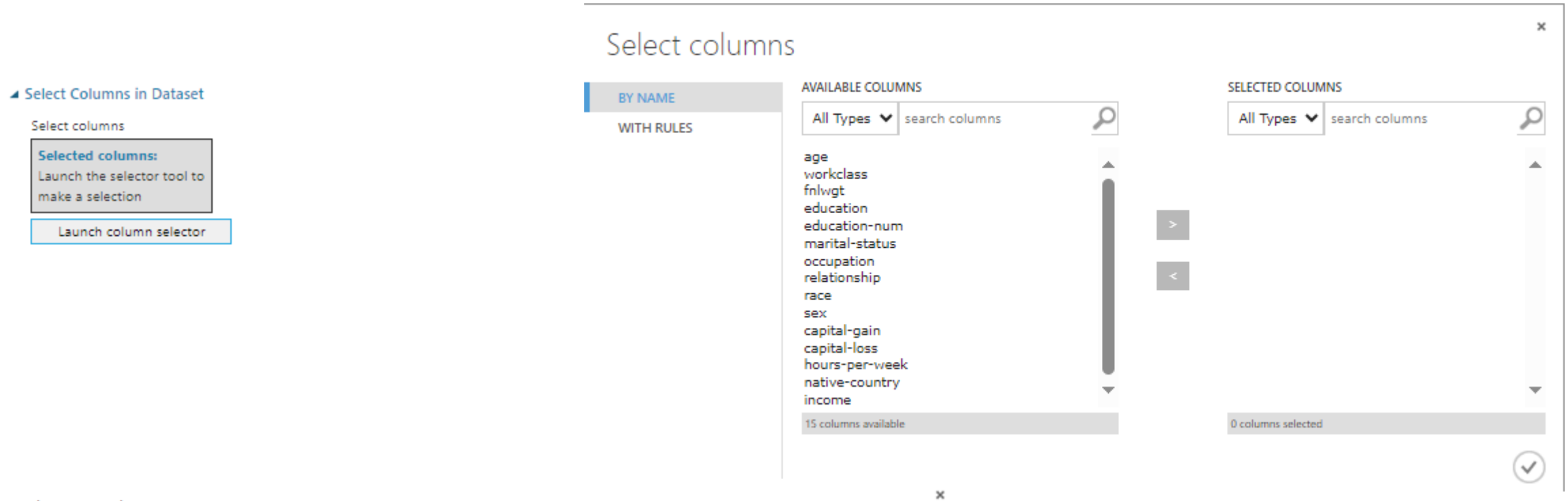
Histogram

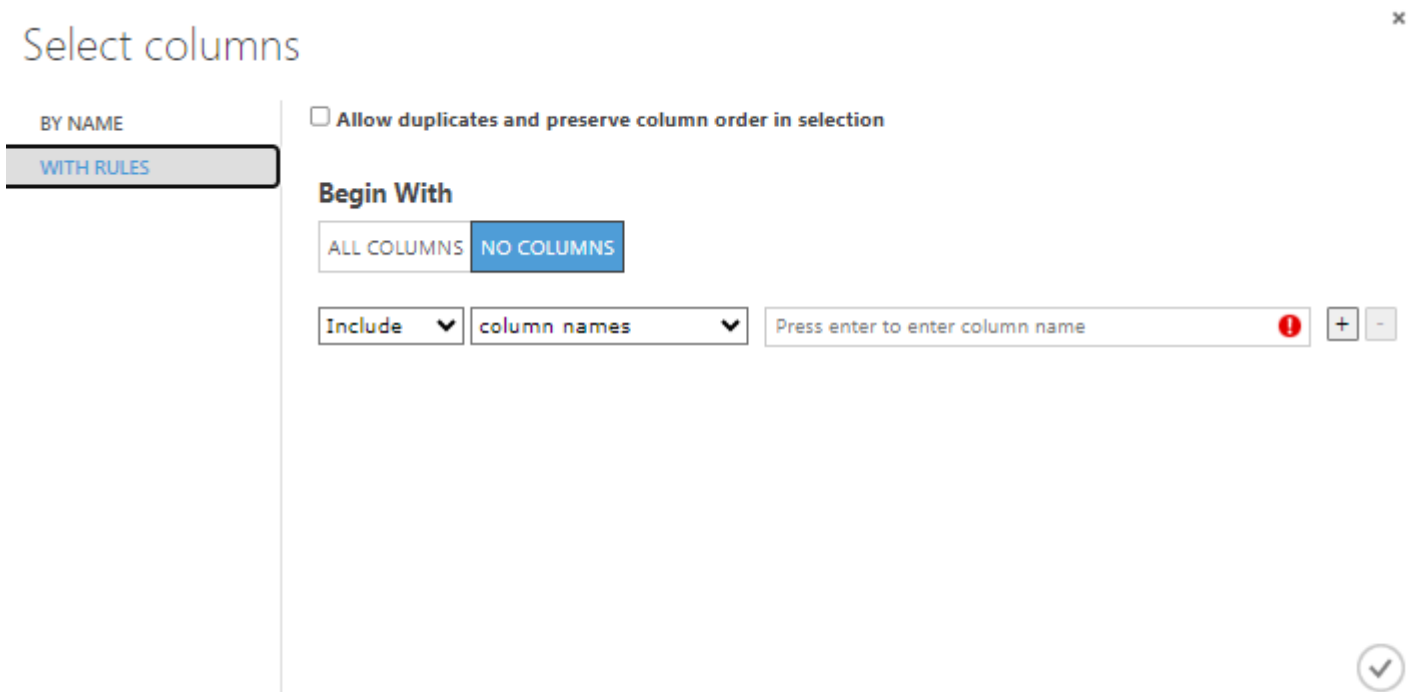


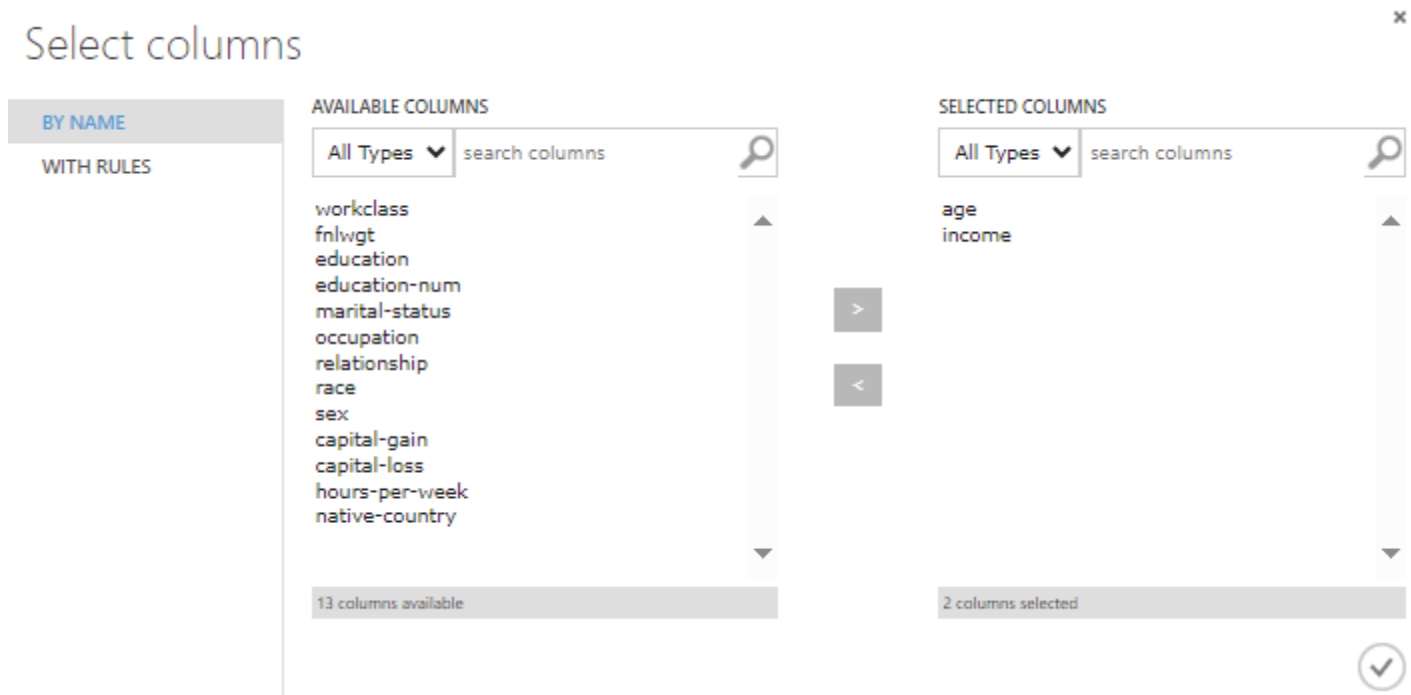

Adult Census Income Binary...
Select Columns in Dataset
1
Select Columns in Dataset
Select columns
Selected columns:
Launch the selector tool to make a selection
Launch column selector
Quick Help
Selects columns to include or exclude from a dataset in an operation. Formerly known as Project Columns.
(more help...)

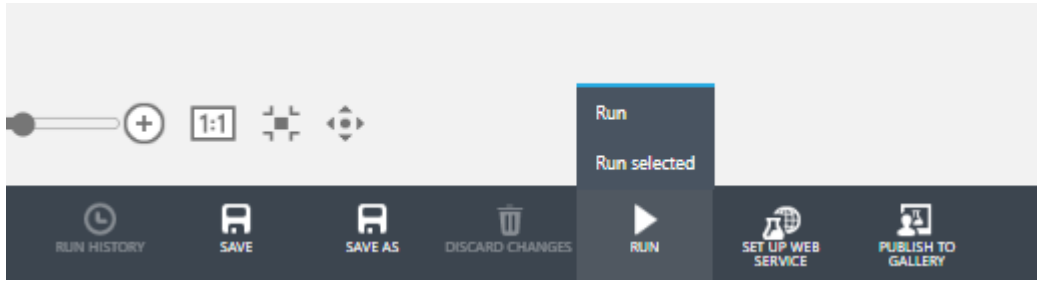

Quick Help
Selects columns to include or exclude from a dataset in an operation. Formerly known as Project Columns.
(more help...)

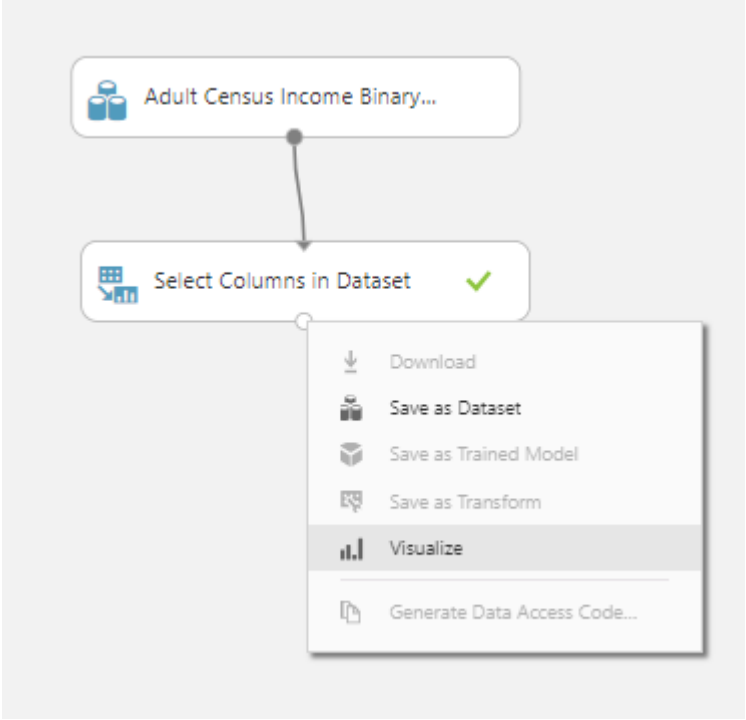

Filter by title
Remove Duplicate Rows
Select Columns in Dataset
Select Columns Transform
SMOTE
Sample and Split
Scale and Reduce
Normalize Data
Feature Selection
Machine Learning Modules
OpenCV Library Modules
Python Language Modules
R Language Modules
Statistical Functions
Text Analytics
Time Series
Data Transformation / Manipulation /
Select Columns in Dataset
Article • 05/06/2019 • 1 contributor
In this article
Module overview
How to use Select Columns in Dataset
Examples
Technical notes
Show 5 more
Important
Support for Machine Learning Studio (classic) will end on 31 August 2024. We recommend you transition to Azure Machine Learning by that date.
Beginning 1 December 2021, you will not be able to create new Machine


Select Columns in Dataset
Select columns
Selected columns:
Launch the selector tool to make a selection
Launch column selector
Select columns
BY NAME
WITH RULES
AVAILABLE COLUMNS
All Types
search columns
age
workclass
fnlwgt
education
education-num
marital-status
occupation
relationship
race
sex
capital-gain
capital-loss
hours-per-week
native-country
income
15 columns available
SELECTED COLUMNS
All Types
search columns
0 columns selected


Select columns
BY NAME
WITH RULES
Allow duplicates and preserve column order in selection
Begin With
ALL COLUMNS
NO COLUMNS
Include
column names
Press enter to enter column name


Select columns
BY NAME
WITH RULES
AVAILABLE COLUMNS
All Types
search columns
workclass
fnlwgt
education
education-num
marital-status
occupation
relationship
race
sex
capital-gain
capital-loss
hours-per-week
native-country
13 columns available
>
<
SELECTED COLUMNS
All Types
search columns
age
income
2 columns selected





Experiment created on 19/02/2024 › Select Columns in Dataset › Results dataset

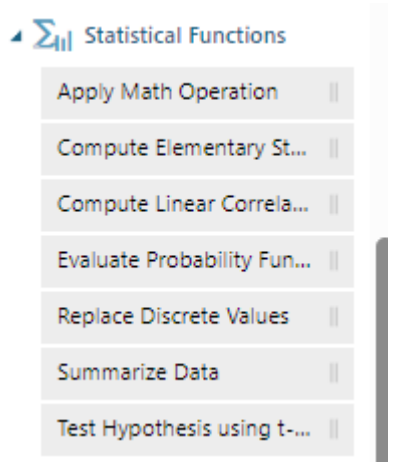
rows	columns
32561	2

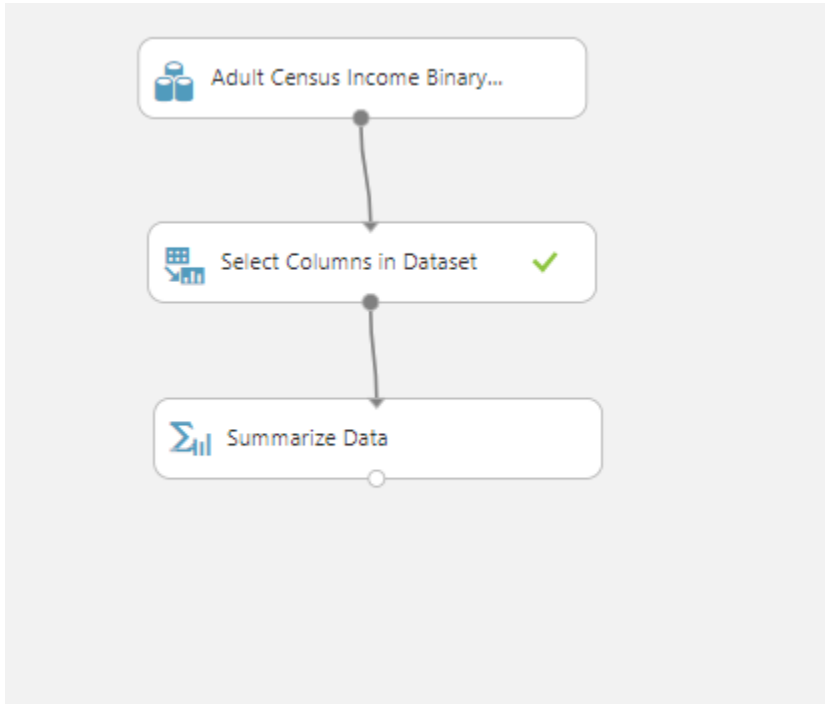
	age	income
view as		
	39	<=50K
	50	<=50K
	38	<=50K
	53	<=50K
	28	<=50K
	37	<=50K
	49	<=50K
	52	>50K
	31	>50K
	42	>50K
	37	>50K
	30	>50K

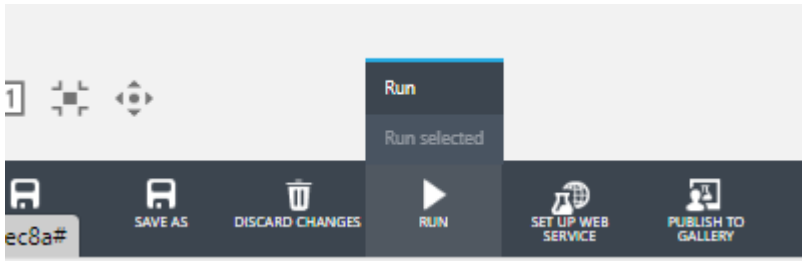
Statistics

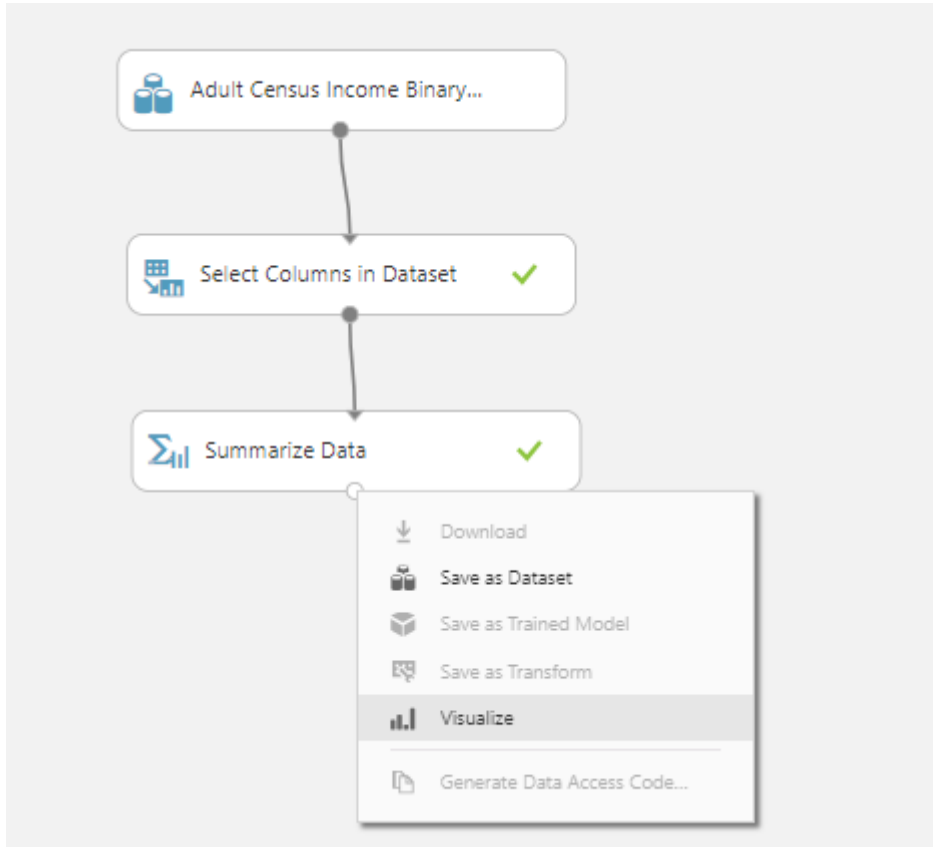
Visualizations

To view, select a column in the table.


Statistical Functions
Apply Math Operation
Compute Elementary St...
Compute Linear Correla...
Evaluate Probability Fun...
Replace Discrete Values
Summarize Data
Test Hypothesis using t-...


Adult Census Income Binary...
Select Columns in Dataset
Summarize Data


Run
Run selected
ec8a#
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Adult Census Income Binary...
Select Columns in Dataset
Summarize Data
Download
Save as Dataset
Save as Trained Model
Save as Transform
Visualize
Generate Data Access Code...

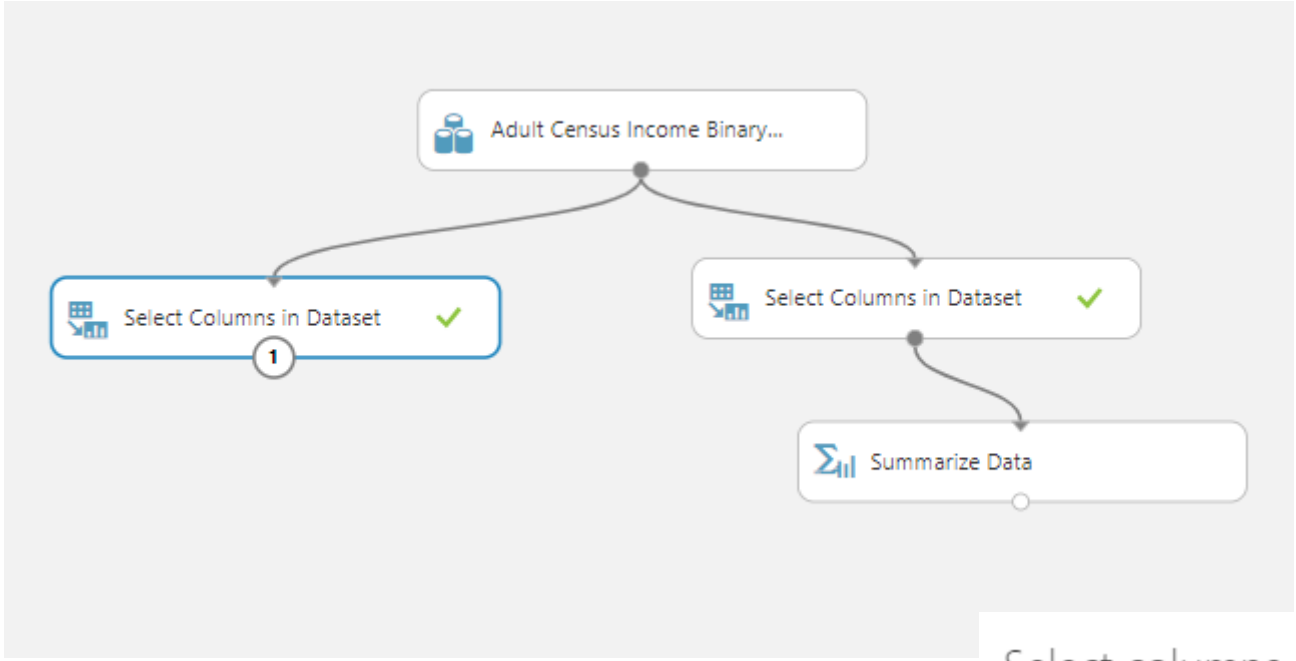
Experiment created on 19/02/2024 › Summarize Data › Results dataset

rows
2

columns
23

view as

Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile	Median	3rd Quart
age	32561	73	0	17	90	38.581647	11.189182	28	37	48
income	32561	2	0							



Selectam toate atributele numerice.

Select columns

BY NAME

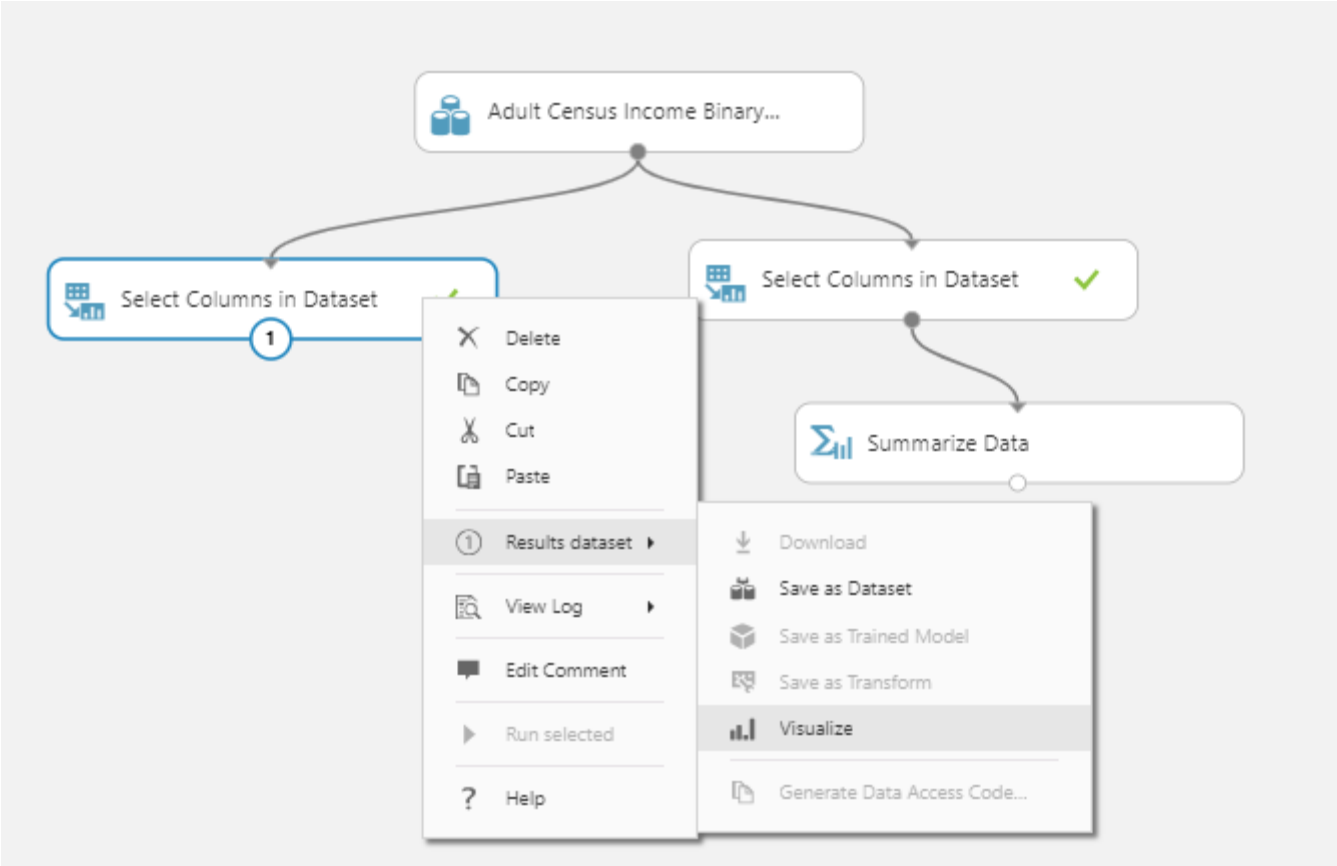
WITH RULES

Allow duplicates and preserve column order in selection

Begin With

ALL COLUMNS NO COLUMNS

Include column type Numeric + -



Cu click dreapta pe modul

Experiment created on 19/02/2024 › Select Columns in Dataset › Results dataset

rows 32561 columns 6

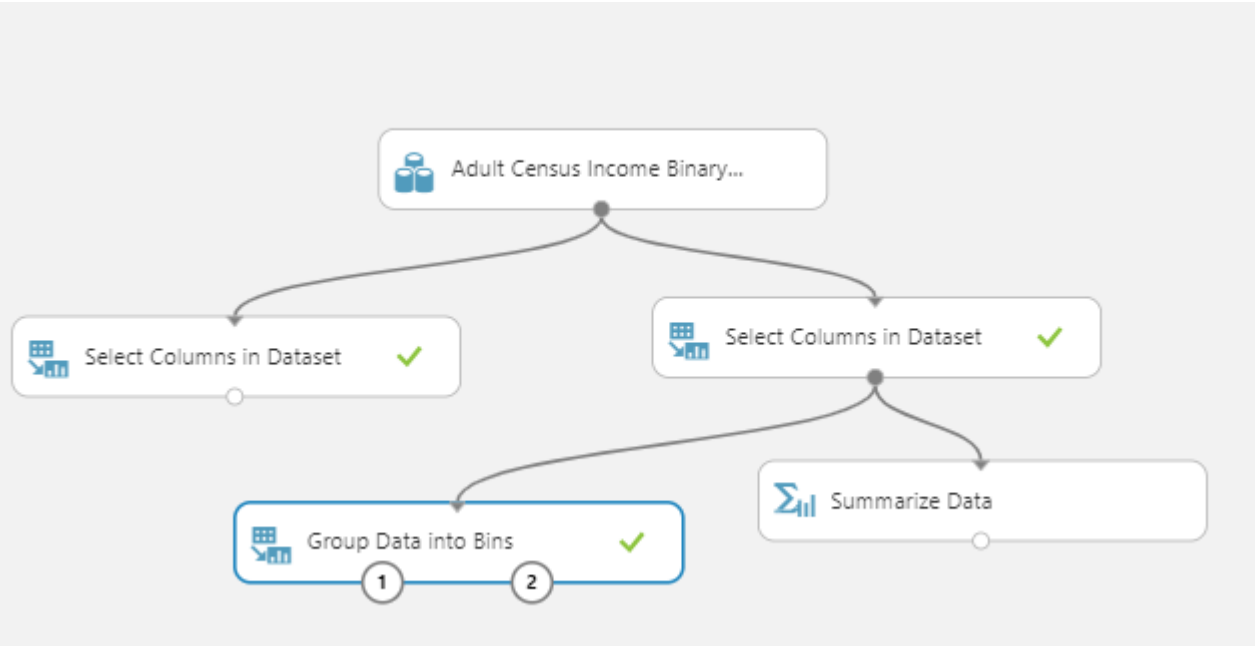
view as

age	fnlwgt	education-num	capital-gain	capital-loss	hours-per-week
39	77516	13	2174	0	40
50	83311	13	0	0	13
38	215646	9	0	0	40
53	234721	7	0	0	40
28	338409	13	0	0	40
37	284582	14	0	0	40
49	160187	5	0	0	16
52	209642	9	0	0	45
31	45781	14	14084	0	50
42	159449	13	5178	0	40
37	280464	10	0	0	80
30	141297	13	0	0	40
23	122272	13	0	0	30
32	205019	12	0	0	50

› Statistics

Visualizations

To view, select a column in the table.



Group Data into Bins

Binning mode

Custom Edges

Comma-separated list ...

20,35,60

Columns to bin

Selected columns:
Column type: Numeric
All

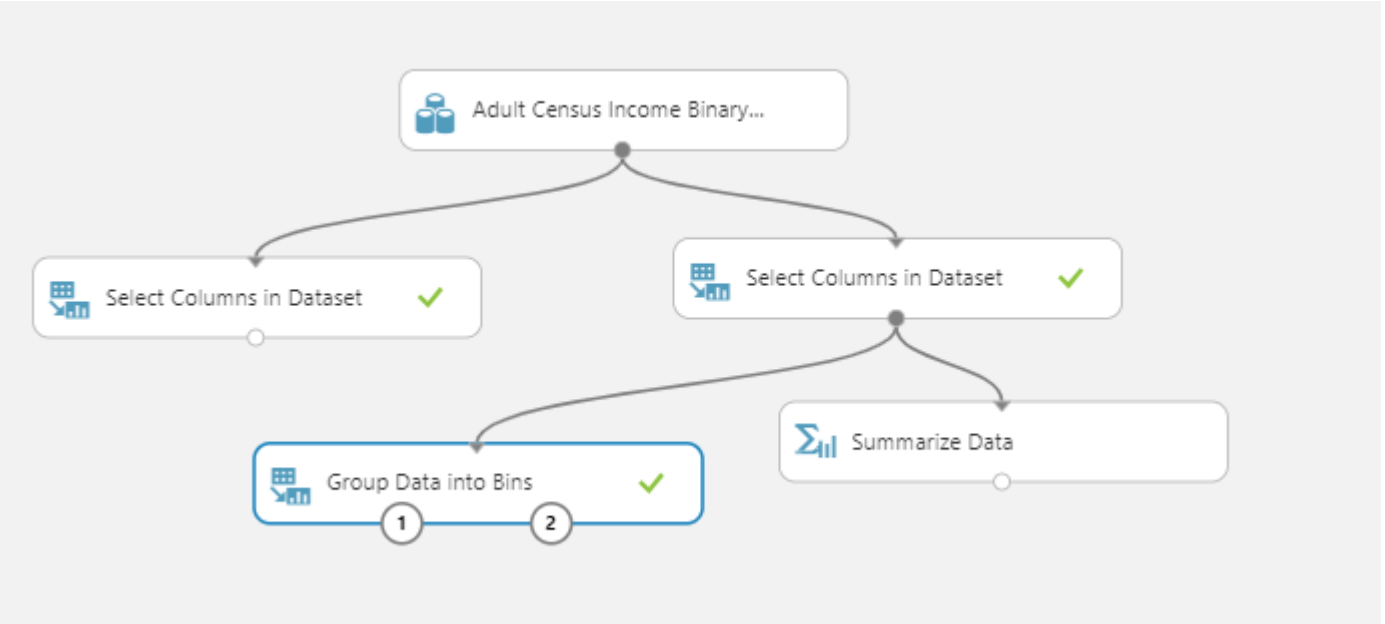
Launch column selector

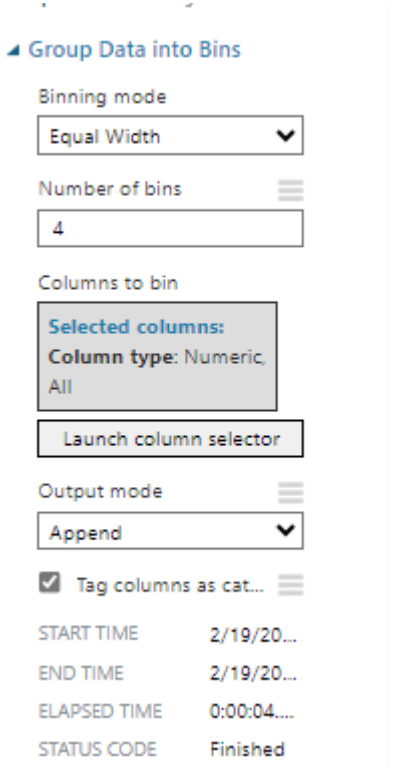
Output mode

Append

Tag columns as cat...

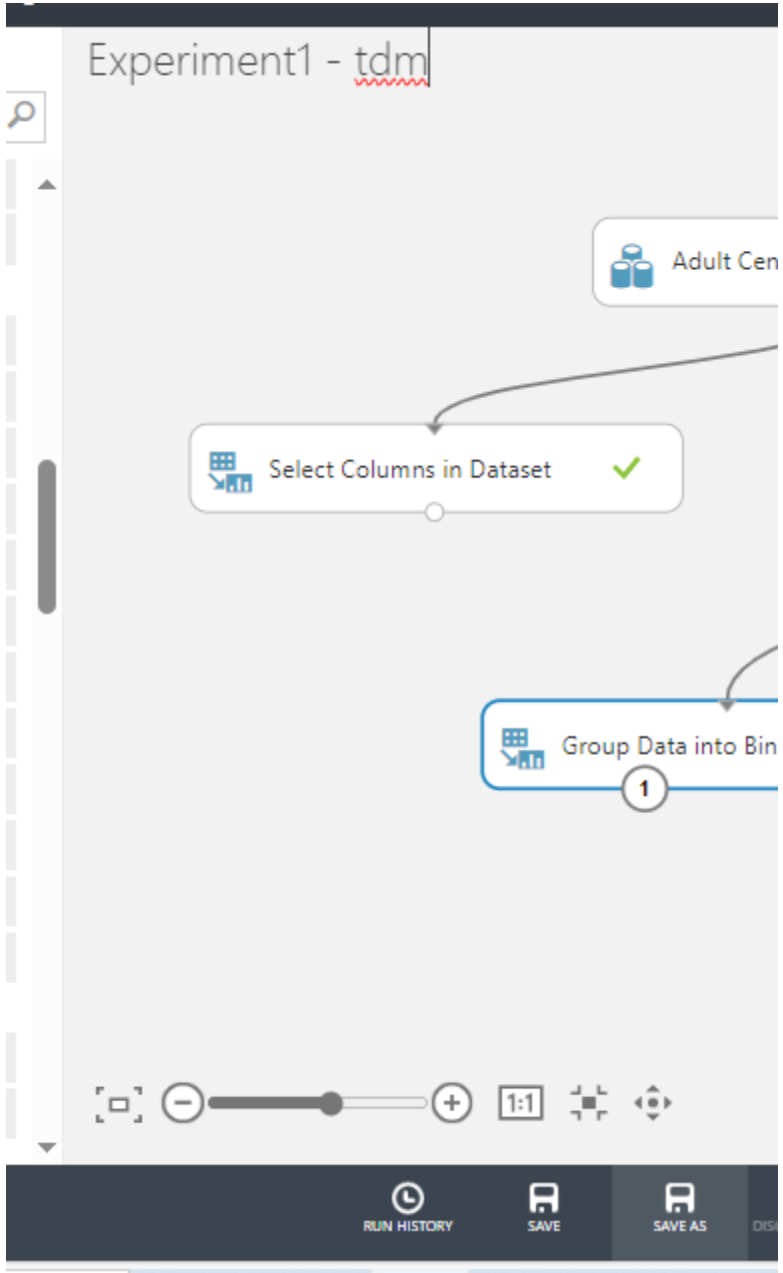
START TIME	2/19/20...
END TIME	2/19/20...
ELAPSED TIME	0:00:04....



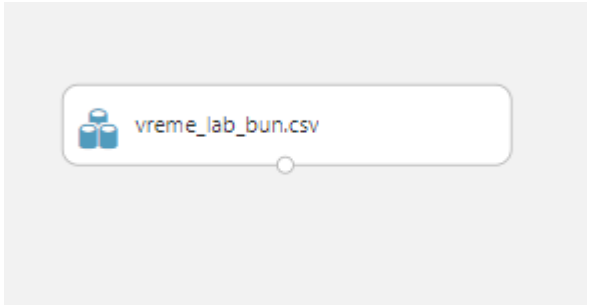


Experimentati si celelalte tipuri de discretizari(group data in bins)

Salvare experiment



Incarcare baza de date

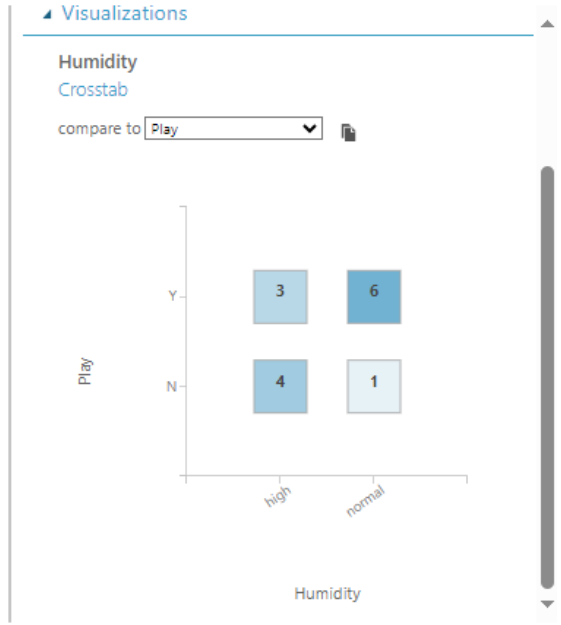


Experiment created on 19/02/2024 › vreme_lab_bun.csv › dataset

rows 14 columns 5

view as

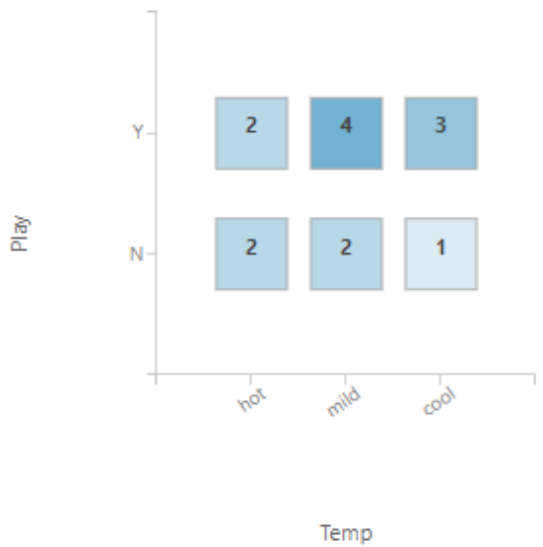
Outlook	Temp	Humidity	Windy	Play
sunny	hot	high	F	N
sunny	hot	high	T	N
overcast	hot	high	F	Y
rain	mild	high	F	Y
rain	cool	normal	F	Y
rain	cool	normal	T	N
overcast	cool	normal	T	Y
sunny	mild	high	F	N
sunny	cool	normal	F	Y
rain	mild	normal	F	Y
sunny	mild	normal	T	Y
overcast	mild	high	T	Y
overcast	hot	normal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



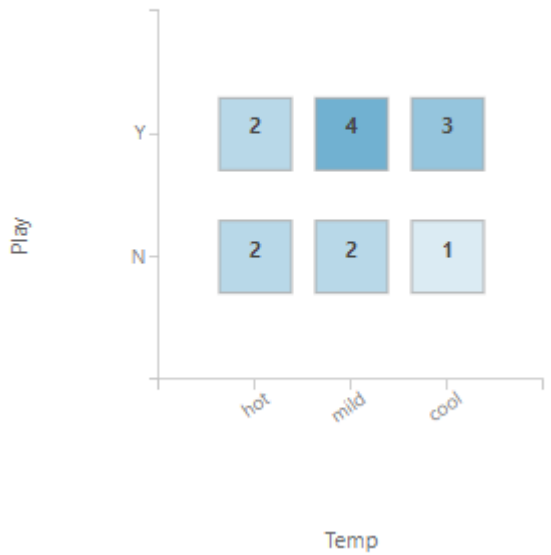
Temp

Crosstab

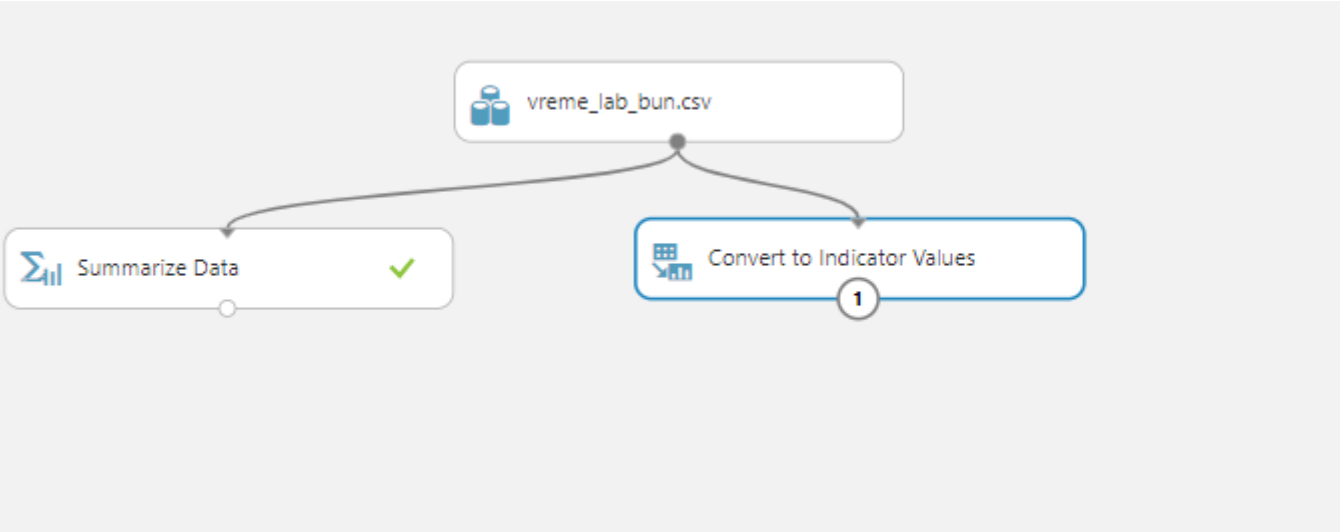
compare to Play



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							

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...)

Baze de date

Baza de date poate sa aiba urmatoarele 3 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

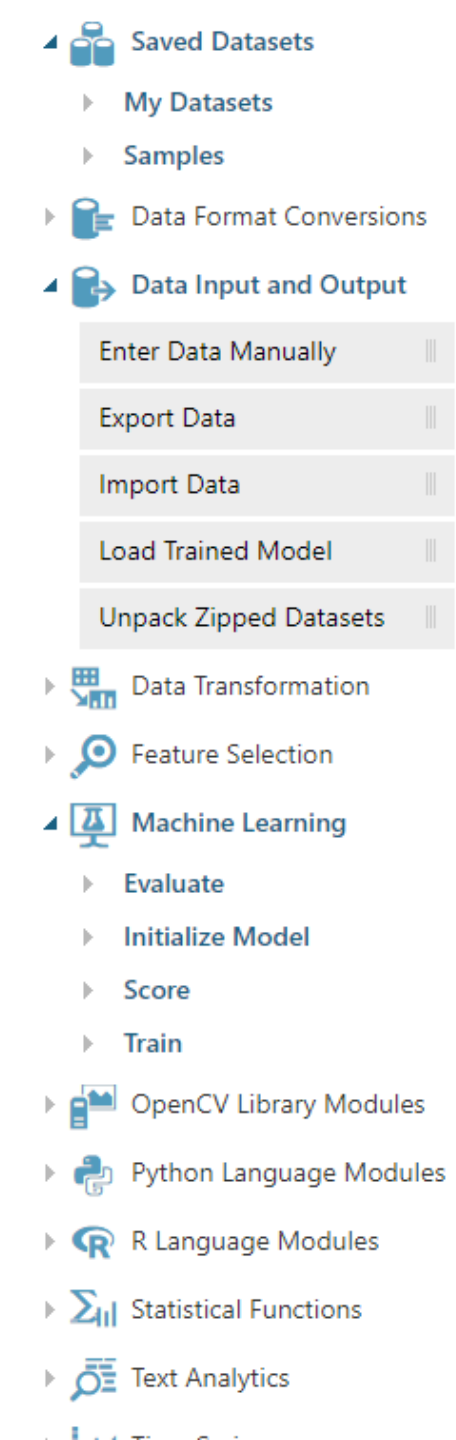
Fisier .arff (format Weka)

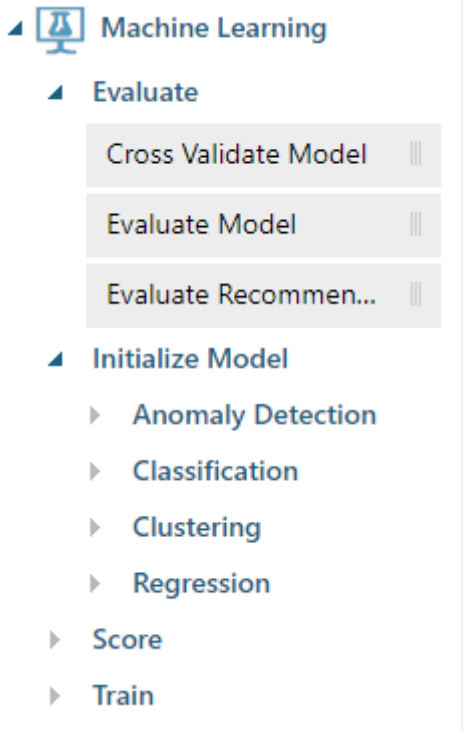
Weka – un alt open source software pentru data mining http://www.cs.waikato.ac.nz/ml/weka/

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

Componentele unui experiment ML Studio (clasic)





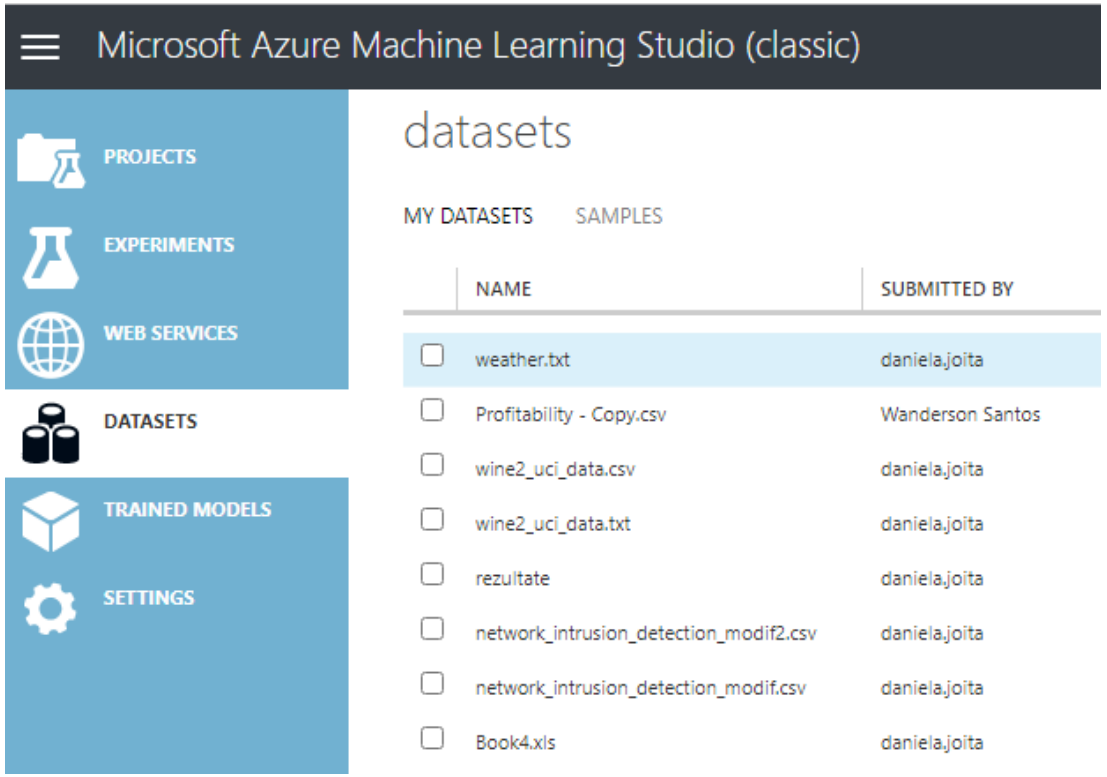
Modulele (Modules)

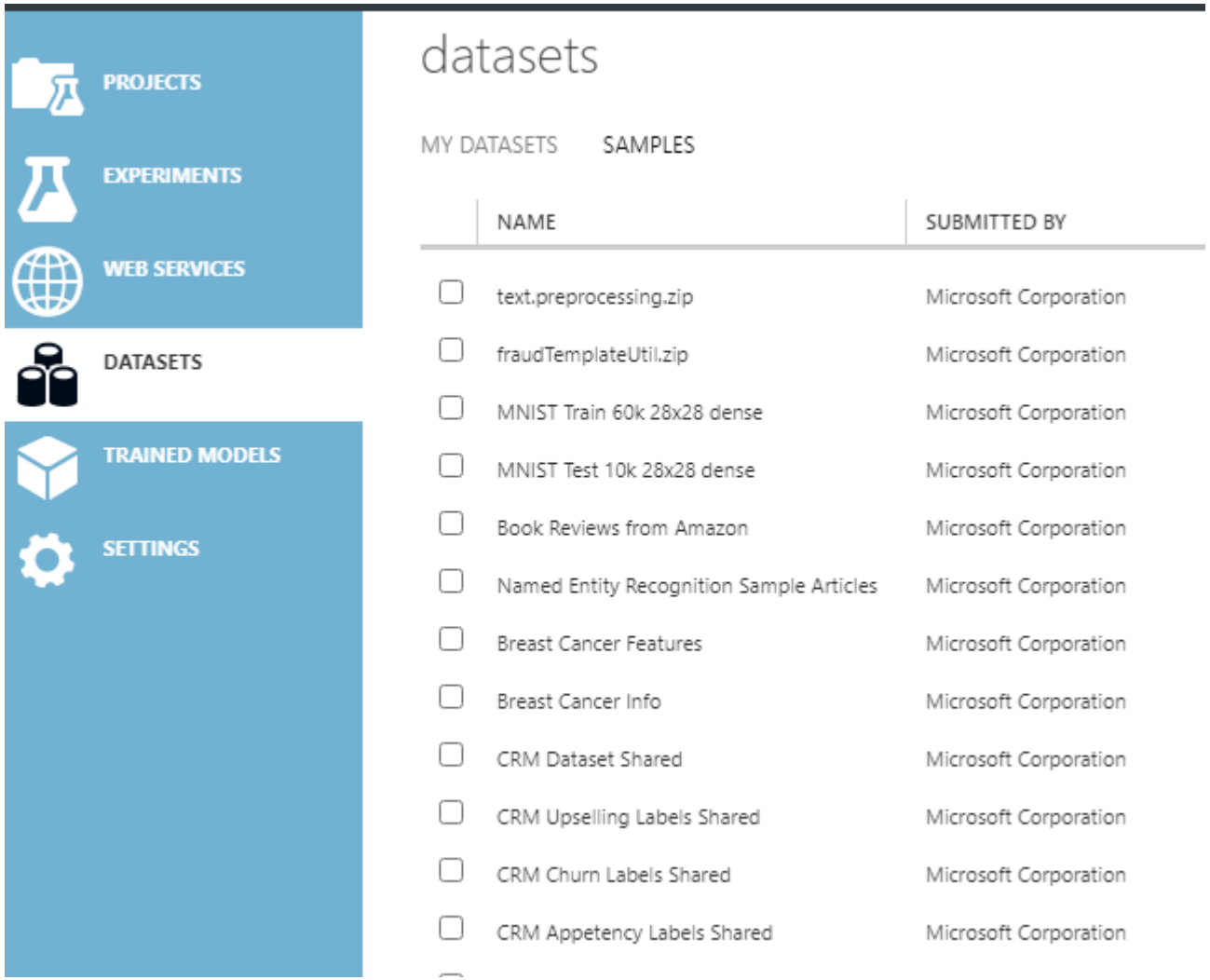
Un modul este un algoritm care poate fi aplicat unor date.

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.

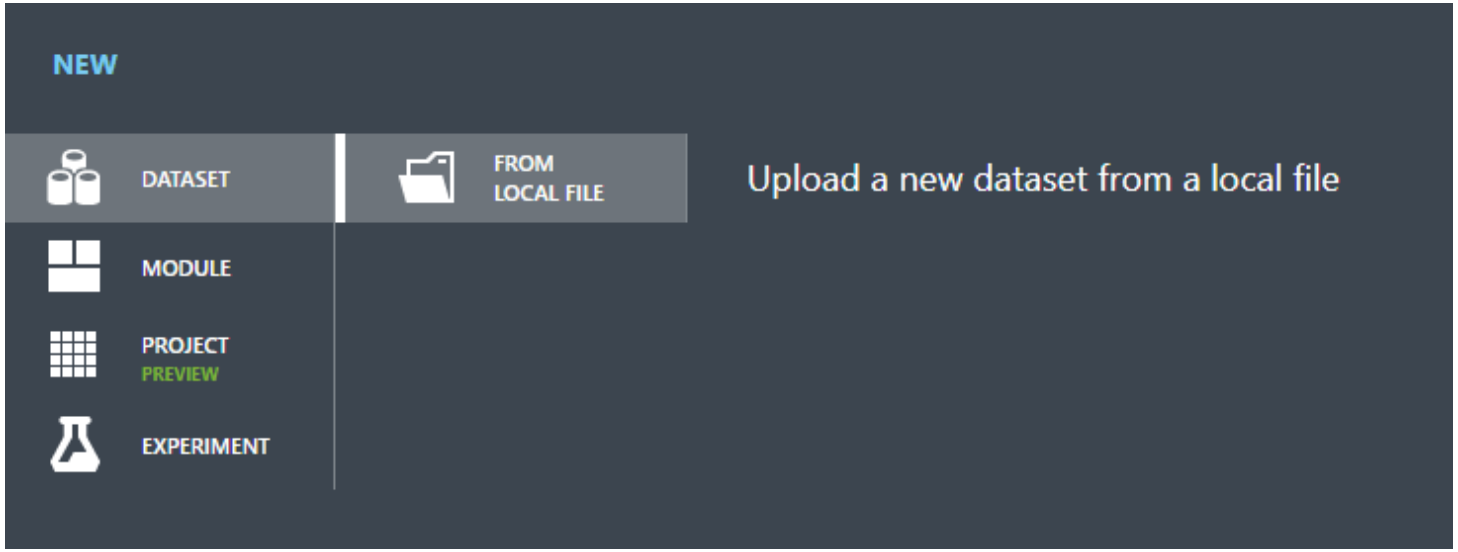
Lista de module disponibile se gaseste in partea stanga a suprafetei interactive si poate fi accesata f usor.

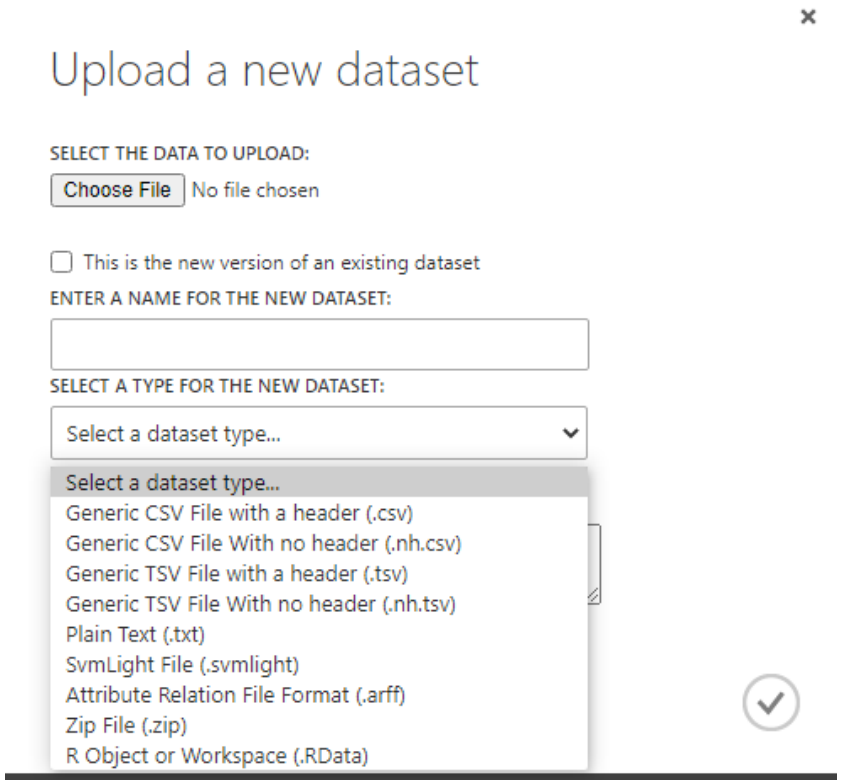
Datasets


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)