

Relația dintre ML & CS

Conf.dr. Cristian KEVORCHIAN

Facultatea de Matematică și Informatică

Universitatea din București

Azure-platformă pentru cloud computing

- ▶ Dezvoltarea, implementarea și managementul aplicațiilor și serviciilor în regim "cloud computing", Azure are alocată o rețea globală de 54 de regiuni care includ peste 100 de centre de date cu peste 4 milioane de servere la nivel global.
- ▶ Oferă, platformă ca serviciu (PaaS), infrastructura ca serviciu (IaaS) și software ca serviciu SaaS. Suportă numeroase limbaje de programare, în diverse metodologii de dezvoltare și diverse medii de programare, incluzând atât sisteme software aparținând Microsoft dar și altor vendori (Oracle, IBM, Google, dar și limbaje cum ar fi C#, Java etc.)
- ▶ Utilizează un sistem de operare distribuit în care context se execută "fabric layer", un cluster de dimensiuni mari.
- ▶ Scalarea și fiabilitatea sunt asigurate de platforma de calcul distribuit Microsoft Azure Fabric, astfel încât serviciile și containerele nu eșuează dacă unul dintre servere este nefuncțional într-unul din centrele de date Microsoft.
- ▶ Se adaugă în mod constant noi servicii și funcții - actualizări frecvente la serviciile existente utilizând o abordare de tip Agile

Azure Cognitive Services

Serviciile cognitive oferă funcționalități de ML pentru rezolvarea unor probleme generale de ordin cognitiv, cum ar fi:

- ▶ Distribuția emoțiilor în analiza textului
- ▶ Recunoașterea obiectelor sau fețelor.

fără a avea nevoie de cunoștințe speciale de ML sau data science pentru a utiliza aceste servicii.

Categorii de servicii cognitive

Categoria	Scopul
Decizion	Crearea de aplicații care să ofere recomandări în ghidarea procesului de luare a deciziilor
Language	Crearea de aplicații care să proceseze limbajul natural în sensul evaluării distribuției sentimentelor și să identifice preferințele utilizatorilor pe baza acestor evaluări.
Search	Apelarea API-urilor de căutare Bing în aplicații și valorificarea capacității de a combina conținut din miliarde de pagini web, imagini, videoclipuri și știri prin intermediul unui singur punct de apelare.
Speech	Conversia vorbire-în-text și text-în-vorbire dar și traducerea dintr-o limbă în alta împreună cu verificarea și recunoașterea vorbitorilor.
Vision	Recunoașterea, identificarea, indexarea și comentarea imaginilor, videoclipurile și a conținutului digital.

Machine Learning

- ▶ Inițial disciplină științifică care are drept obiect de studiu construcția algoritmilor și modelelor matematice pe care sistemele IT le folosesc pentru a-și îmbunătăți progresiv performanțele relative la un anumit task. Algoritmii de învățare automată implementează un model matematic al datelor de antrenare, pentru a realiza predicții sau a genera decizii fără a fi programat în mod explicit pentru a realiza această sarcină. Ex. E-mail filtering, intrusion detection și computer vision, unde este foarte dificil să se dezvolte un algoritm care efectiv să implementeze task-ul.
- ▶ Teoria și practica ML este strâns legată de Statistica Computațională, care este un domeniu de frontieră între statistica și computer science.
- ▶ Data Mining este un subdomeniu al ingineriei datelor. Reprezintă procesul de identificare a tiparelor în volume mari de date care implică utilizarea de metode aflate la intersecția inteligenței artificiale, ML, Statisticii matematice și a sistemelor de baze de date.

Machine Learning(ML)

ML este un concept în care algoritmi specializați ”învăță” pe baza datelor de antrenare furnizând un model care poate fi utilizat în diferite contexte computationale. Modelul poate fi instruit și pe baza noilor date oferite de acesta(Q-learning, reinforcement learning).

Procesul de construire a unui sistem de învățare automată necesită unele cunoștințe de învățare automată sau știința datelor.

Serviciile de învățare automată sunt furnizate Azure Machine Learning (AML).

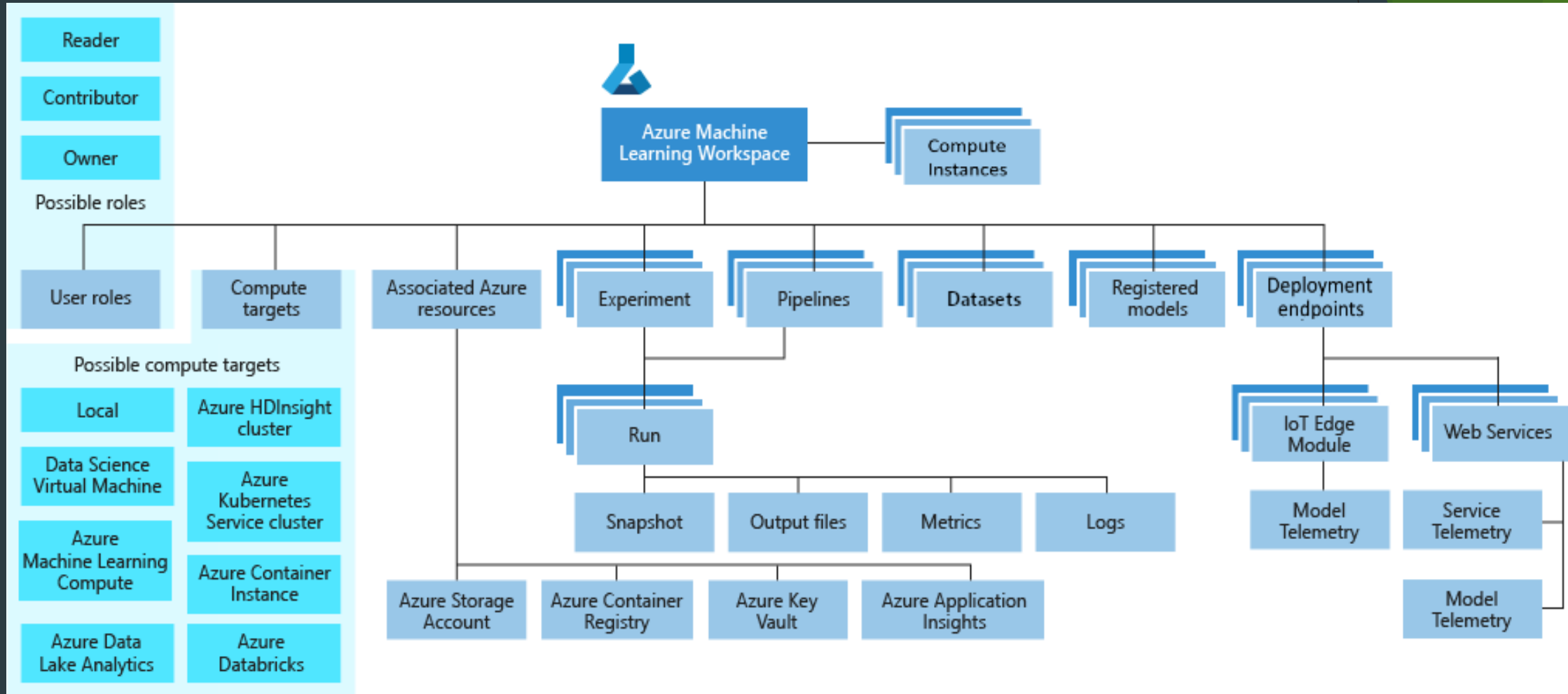
ML & Servicii Cognitive

- ▶ Un serviciu cognitiv poate include toate sau numai unele componente ale unei soluții de machine learning: date, algoritmi și modele de instruire.
- ▶ Aceste servicii sunt menite să necesite cunoștințe generale despre datele care vor genera modelul, fără a fi nevoie de experiență în ML sau data science.
- ▶ Aceste servicii oferă atât API-uri REST, cât și SDK-uri operaționale într-o serie de limbaje de programare.
- ▶ Prin urmare, trebuie să aveți cunoștințe în aceste limbaje de programare pentru a utiliza serviciile cognitive.
 - **Cognitive Services sunt destinate dezvoltatorilor fără a fi necesară experiența în machine-learning.**
 - **Azure Machine Learning este destinat celor cu experiență în data science.**

Cognitiv Services vs ML Services

- ▶ Un serviciu cognitiv oferă un model instruit. Aceasta reunește date și un algoritm, disponibile prin intermediul unui API REST sau SDK.
- ▶ Este posibilă implementarea acestui serviciu, în funcție de scenariul un scenario peopus.
- ▶ Un serviciu cognitiv oferă răspunsuri la probleme generale, precum fraze cheie din text sau identificarea articolelor din imagini.
- ▶ ML este un proces care necesită, în general, o perioadă mai lungă de timp pentru a fi implementat cu succes. Acest timp este cheltuit pentru colectarea transformarea si calibrarea datelor, selectarea algoritmilor, instruirea modelului și implementarea pentru a ajunge la același nivel de funcționalitate oferit de un serviciu cognitiv. Cu învățarea automată, este posibil să oferiți răspunsuri la probleme foarte specializate și / sau specifice. Problemele de învățare automată necesită familiarizarea cu subiectul specific și cu datele problemei luate în considerare, precum și expertiză în Data Science.

Taxonomia unui WS în Azure Machine Learning



Machine Learning vs. Cognitive Computing

- ▶ **Cognitive Computing** - simularea proceselor de gândire(umană) într-un model computerizat, implicând sisteme de auto-învățare bazate pe extragerea datelor, recunoașterea tiparelor și procesarea limbajului natural în scopul de a imita modul în care funcționează creierul uman; generează o nouă clasă de probleme calculabile, abordând situații complexe care sunt caracterizate de ambiguitate și incertitudine.
- ▶ **Învățare automată:** simularea pe baze algoritmice a capacității de a învăța fără a fi implementat un program de calcul explicit în mod explicit; o metodă de analiză a datelor care automatizează construirea modelului analitic; folosind algoritmi care învață iterativ din date pentru a identifica structuri informaționale fără a fi programat în mod explicit unde și ce să caute.

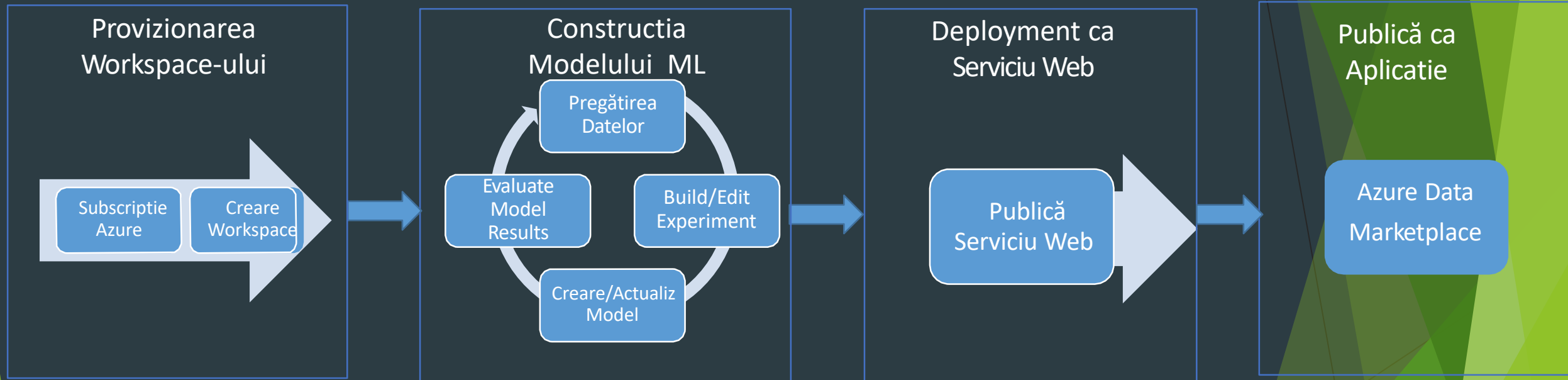
Microsoft Azure Machine Learning

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The top navigation bar includes 'Home', 'Studio', and 'Gallery'. The left sidebar contains a search bar and a list of experiment items: Data Format Conversions, Data Input and Output, Data Transformation, Feature Selection, Machine Learning (selected), Evaluate, Initialize Model, Anomaly Detection, Classification, Clustering, Regression, Score, Train, OpenCV Library Modules, Python Language Modules, R Language Modules, Statistical Functions, Test Analytics, and Web Service. The main workspace shows a 'Linear Regression' experiment in draft mode, created at 11:07:07 AM. The workflow consists of several steps: 'Enter Data Manually' (input), 'Train Model' (training), 'Score Model' (scoring), and 'Evaluate Model' (evaluation). A 'Quick Help' panel is visible at the bottom right, stating 'Enables entering and editing small datasets by typing values (more help...)'. The 'Properties' panel on the right shows the 'Enter Data Manually' settings, including 'DataFormat' set to 'CSV' and 'Has Header' checked. A table of data is displayed below these settings.

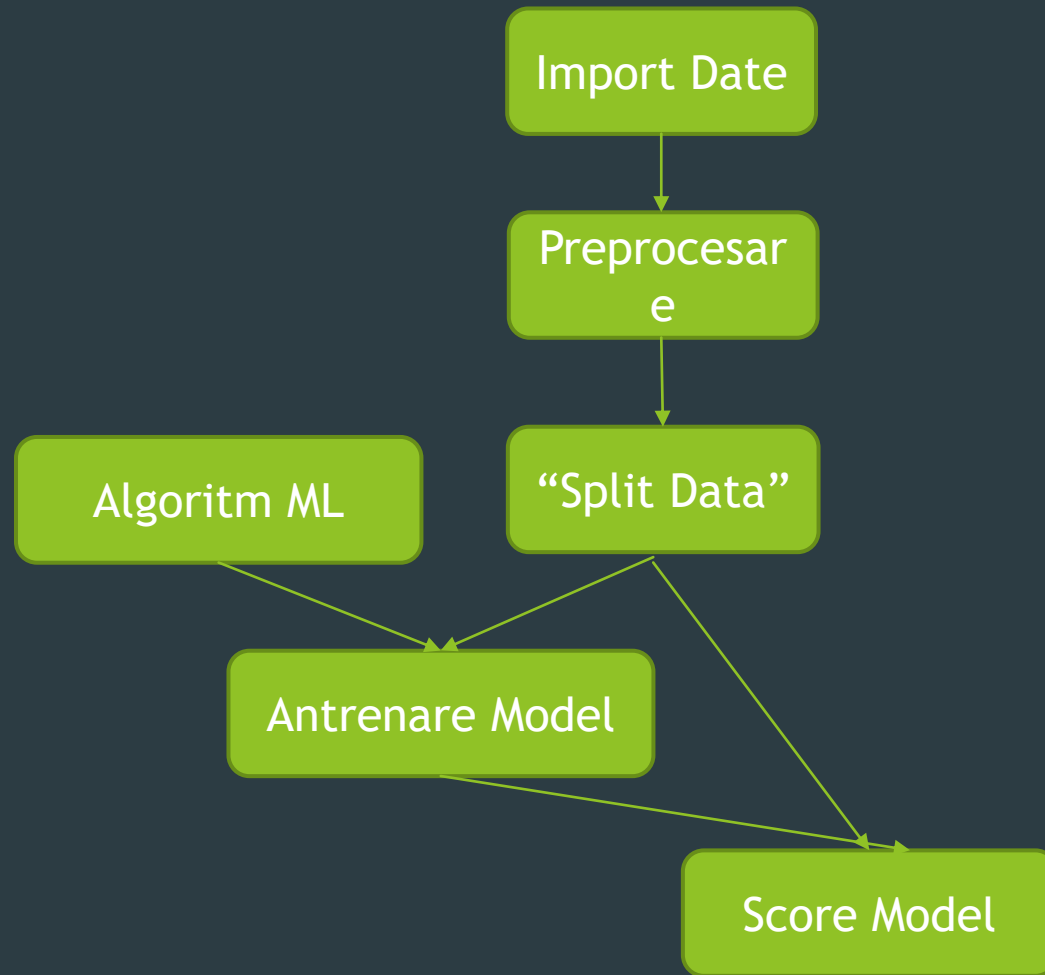
x	y	noise
1	0.251535205	
2	0.555030732	
3	2.45169005	
4	2.757234922	
5	5.536073336	

- UI accesibil din diverse browsere
- Partajază | colaborează cu alte workspace-uri ML
- Design visual Drag & Drop | Dezvoltare
- O gamă largă de algoritmi destinați ML
- Extins cu scripturi OSS R | Python.
- Partajază | Documente cu IPython | Jupyter
- Deployment | Publicare | Scalare rapida (API-uri)

Ecosistemul Azure Machine Learning

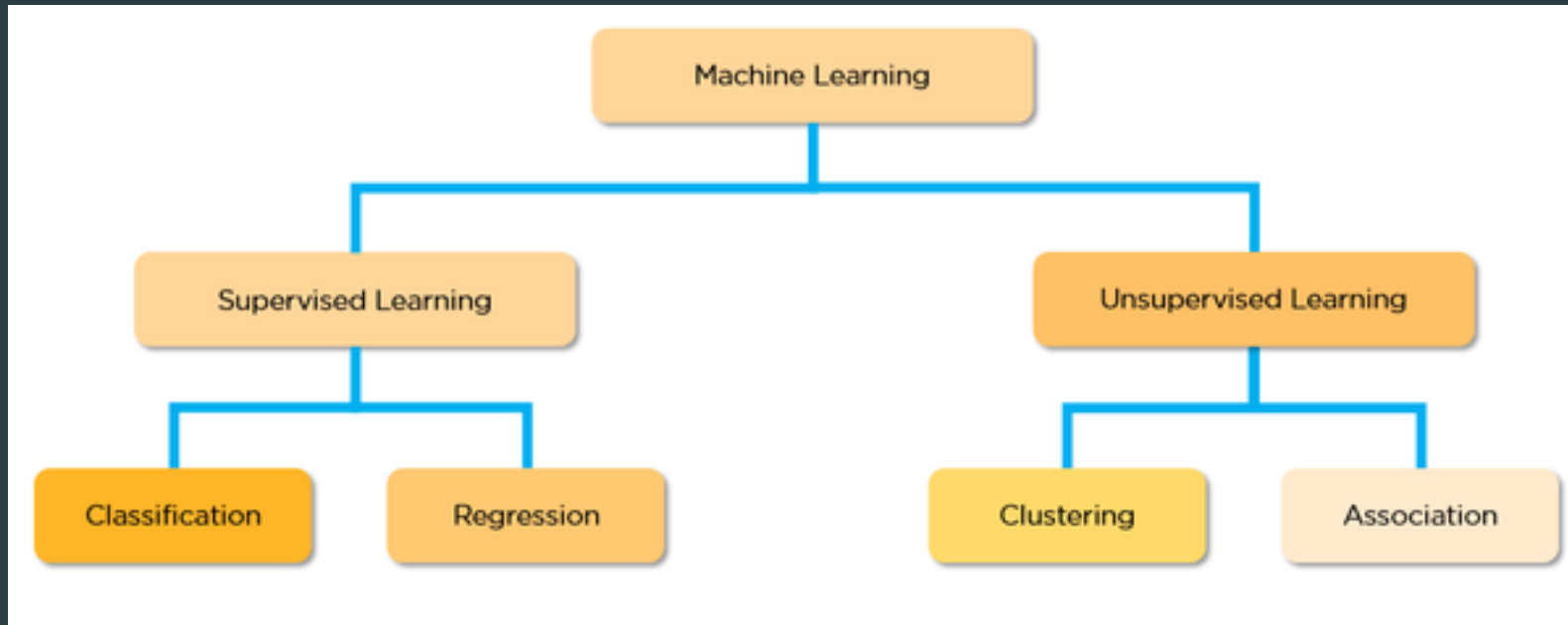


Fluxul ML



ML-Tipuri de învățare

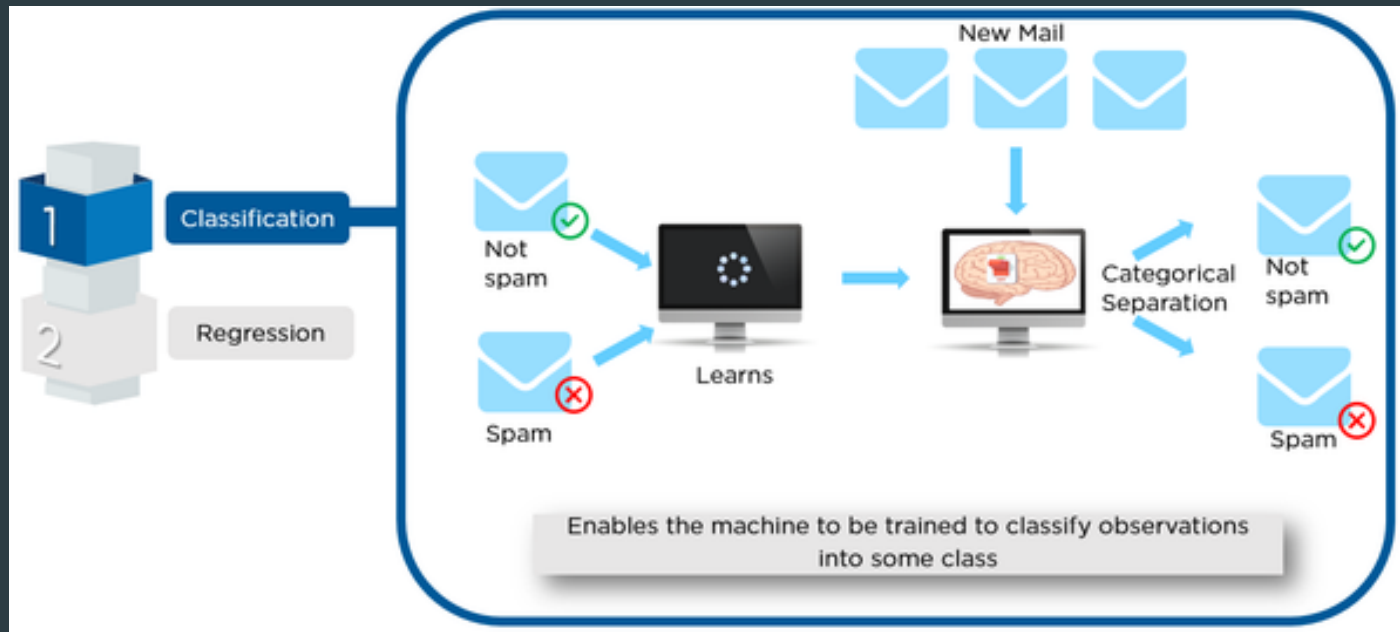
- **Învățarea supervizată:** Sistemul este instruit folosind datele anterioare (care includ intrările și ieșirile) fiind recomandat a fi aplicat la probleme de decizie sau pentru realizarea de previziuni atunci când se întâlnesc date noi.
- **Învățare nesupervizată:** Sistemul este capabil să recunoască regularități (tiparele), luând în considerare numai datele de intrare.
- **Învățare consolidată:** deciziile sunt luate de sistem pe baza recompensei / pedepsei aplicate pentru ultima acțiune efectuată.



Clasificarea

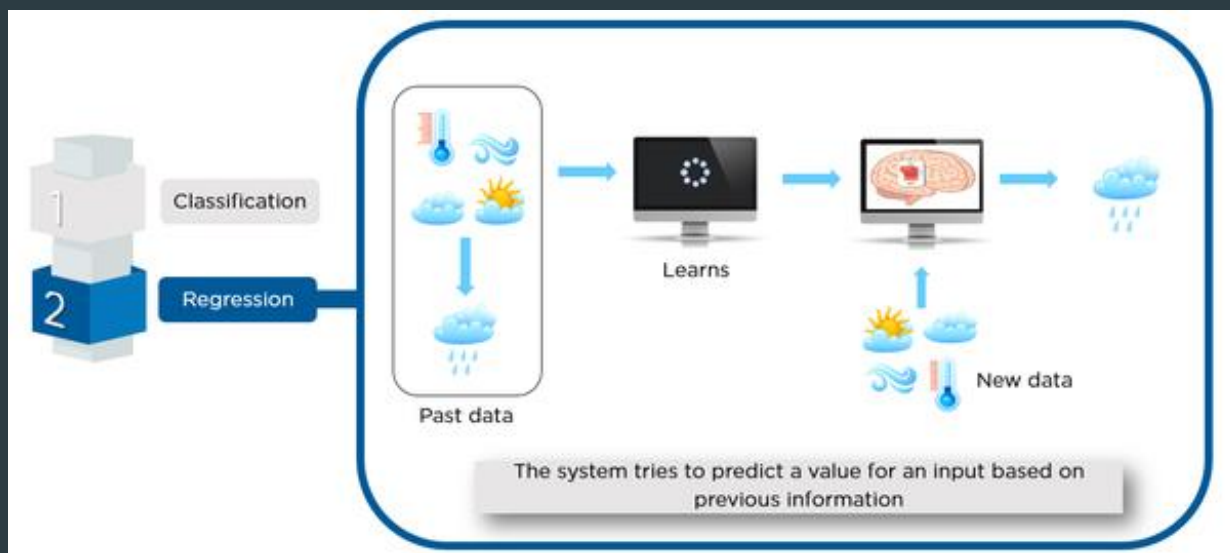
Datele trebuie să fie împărțite într-un anumit număr de categorii bazate pe un proces de "training" folosind date istorice.

Un algoritm celebru care poate fi folosit pentru a rezolva problemele de clasificare, este teorema lui Bayes.



Regresia

Anticipăm o valoare pentru o intrare pe baza informațiilor primite anterior. Deși acest lucru pare similar clasificării, având în vedere că ambele utilizează date istorice pentru a face previziuni, similitudinea lor se rezumă la acest amănunt. În cazul regresiei, trebuie să estimăm o valoare și nu doar o clasă de observație.



Încărcarea Modelului de Date și Învățarea

			Variables			Valori dorite a face predictie	
			A	B	C		
Date de Antrenare		1.2	45	14		N	
		1.4	43	16		D	
		1.1	67	3		D	
		1.2	55	7		N	
		1.2	30	13		N	
Date de test		1.3	43	16		Ar trebui luat N	
		1.1	65	3		Ar trebui luat D	

Azure Machine Learning (AML)

ML Studio - Tabloul de bord

Microsoft Azure Machine Learning

Home

Studio

Gallery PREVIEW



EXPERIMENTS



WEB SERVICES



DATASETS



TRAINED MODELS



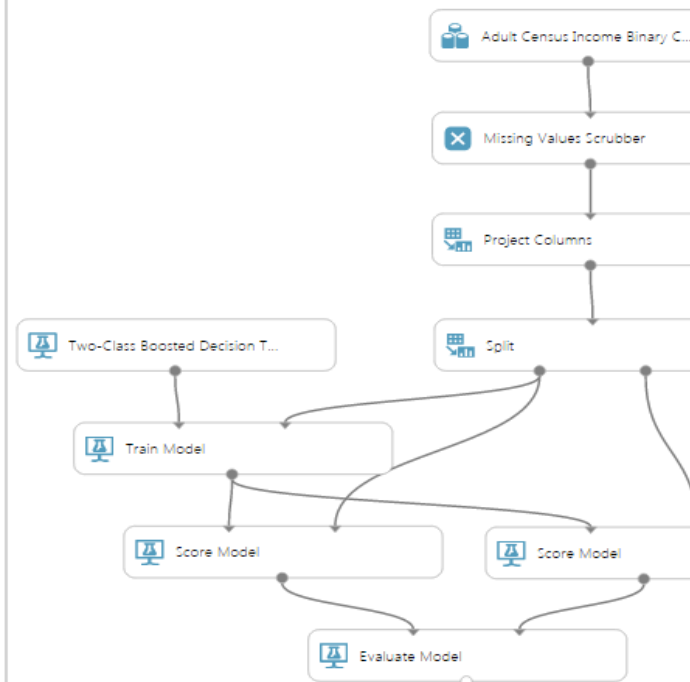
SETTINGS

experiments

MY EXPERIMENTS

SAMPLES

	NAME	AUTHOR	STATUS	LAST EDITED
<input checked="" type="checkbox"/>	Sample 5: ScoreExperiment	alannoel	Draft	2/7/2015 6:40:25 PM
<input type="checkbox"/>	Chapter 02 - Hello ML AI Noel	alannoel	Finished	12/3/2014 2:44:05 PM
<input type="checkbox"/>	Chapter 02 - Hello ML AI Noel	alannoel	Draft	12/3/2014 2:32:14 PM
<input type="checkbox"/>	Sample Network Intrusion Detect...	alannoel	Finished	9/6/2014 10:37:59 AM
<input type="checkbox"/>	Sample Experiment - Demand Fo...	alannoel	Draft	7/25/2014 4:42:53 PM
<input type="checkbox"/>	German Credit Prediction	alannoel	Draft	7/25/2014 4:18:41 PM
<input type="checkbox"/>	Sample 5: Train, Change to Boost...	alannoel	Finished	7/24/2014 6:53:54 PM
<input type="checkbox"/>	Japanese Trained Model	alannoel	Draft	7/24/2014 6:51:53 PM
<input type="checkbox"/>	Untitled	alannoel	Draft	7/24/2014 6:33:09 PM
<input type="checkbox"/>	Untitled	alannoel	Finished	7/24/2014 6:13:58 PM
<input type="checkbox"/>	Untitled	alannoel	Finished	7/24/2014 6:07:17 PM
<input type="checkbox"/>	Untitled	alannoel	Draft	7/24/2014 6:02:34 PM
<input type="checkbox"/>	Sample 5: TrainedModel	alannoel	Draft	7/24/2014 4:23:23 PM
<input type="checkbox"/>	Sample 5: TrainCopy	alannoel	Finished	7/24/2014 4:21:47 PM
<input type="checkbox"/>	Income Level Prediction	alannoel	Finished	7/24/2014 1:12:12 PM
<input type="checkbox"/>	Flight Delay Prediction	alannoel	Draft	7/23/2014 8:32:21 PM
<input type="checkbox"/>	Untitled	alannoel	Finished	7/23/2014 6:25:35 PM
<input type="checkbox"/>	Writer01	alannoel	Failed	7/23/2014 6:22:41 PM
<input type="checkbox"/>	Untitled	alannoel	Draft	7/23/2014 6:05:33 PM
<input type="checkbox"/>	Demo01	alannoel	Draft	7/23/2014 6:00:41 PM



Portalul pentru management

Microsoft Azure | | alannoel@msn.com

STORAGE 5

HDINSIGHT 0

MEDIA SERVICES 0

SERVICE BUS 0

VISUAL STUDIO ONLINE 1

CACHE 0

BIZTALK SERVICES 0

RECOVERY SERVICES 0

CDN 0

AUTOMATION 0

SCHEDULER 0

API MANAGEMENT 0

MACHINE LEARNING 1

NETWORKS 1

TRAFFIC MANAGER 0

REMOTEAPP

all items

NAME	TYPE	STATUS	SUBSCRIPTION	LOCATION	
alnoel01 →	Storage Account	✓ Online	Pay-As-You-Go	East US	
alnoelws01	Storage Account	✓ Online	Pay-As-You-Go	South Central US	
anoelw2k12	Storage Account	✓ Online	Pay-As-You-Go	East US	
portalvhds4vxb6nzk1fsc	Storage Account	✓ Online	Pay-As-You-Go	North Central US	
portalvhds20ws6qp6033k	Storage Account	✓ Online	Pay-As-You-Go	East US	
alnoel01	Cloud service	✓ Running	Pay-As-You-Go	North Central US	
anoelW2K12	Cloud service	✓ Running	Pay-As-You-Go	East US	
anoelW2K12	Virtual machine	✓ Running	Pay-As-You-Go	East US	
wolfie	Website	■ Stopped	Pay-As-You-Go	East US	
alnoel12feb	Website	■ Stopped	Pay-As-You-Go	East US	
Group Group anoelW2K12	Virtual Network	✓ Created	Pay-As-You-Go	East US	
Default Directory	Directory	✓ Active	Shared by all Default Directory subscriptions	United States	
alnoel	Visual Studio Online	✓ Active	Pay-As-You-Go	North Central US	
alnoelws01	ML Workspace	✓ Online	Pay-As-You-Go	-	

- ▶ <https://docs.microsoft.com/en-us/azure/cognitive-services/cognitive-services-and-machine-learning>

Mulțumesc pentru
atenție!