

Modulo IA

Analisi di Big Data basata su
"Intelligenza Artificiale"

Docente

- **Attualmente**

- Libero professionista e consulente in ambito ICT e Data Science/Analytics
- Docente, formatore, educatore
- Ricercatore indipendente, divulgatore, comunicatore ed attivista

- **25+ anni di carriera in ICT**

- Accademia, R&D, industria, agenzie multilaterali, non-profit
- PM, coordinatore di progetti e team, ricercatore, docente
- Italia ed estero (EU, UK, US, Sud America, UN)

- **Laurea specialistica (v.o.) e Master di II livello**

- Ingegneria Informatica
- Ingegneria e progettazione per la cooperazione e lo sviluppo

- **Specializzazione e certificazioni**

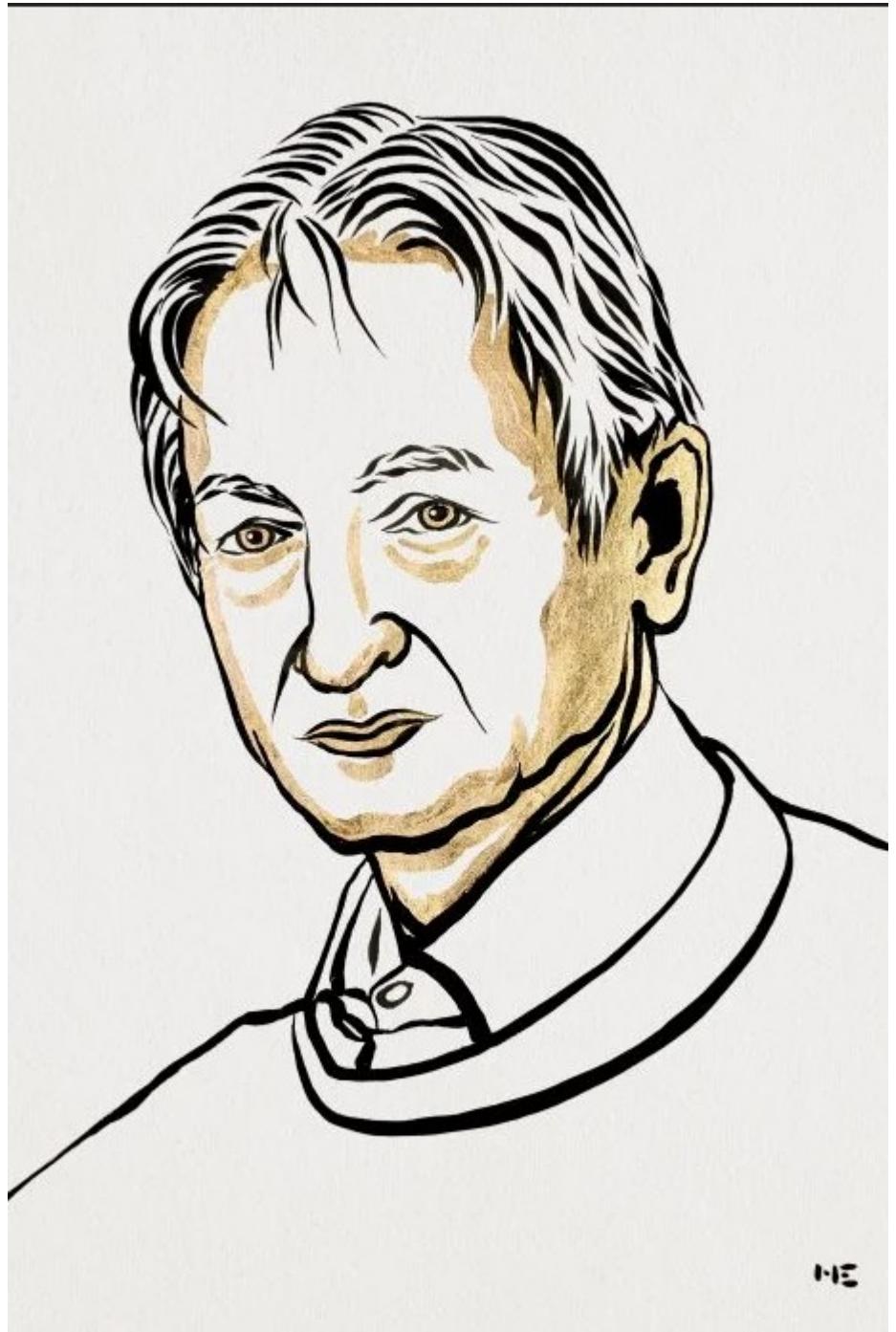
- Data Science, Humane Technology
- Piattaforme di analytics (KNIME)

Studenti

- Percorso scolastico e formativo
- Perché un ITS?
- Sogni e ambizioni

Storiella “leggera”...

... sul “peso” odierno di Analisi Big Data ed Intelligenza Artificiale



Geoffrey Hinton Facts



Ill. Niklas Elmehed © Nobel
Prize Outreach

Geoffrey Hinton
The Nobel Prize in Physics 2024

Born: 6 December 1947, London, United Kingdom

Affiliation at the time of the award: University of Toronto,
Toronto, Canada

Prize motivation: “for foundational discoveries and
inventions that enable machine learning with artificial
neural networks”

Prize share: 1/2

Work

When we talk about artificial intelligence, we often mean machine learning using artificial neural networks. This technology was originally inspired by the structure of the brain. In an artificial neural network, the brain's neurons are represented by nodes that have different values. In 1983–1985, Geoffrey Hinton used tools from statistical physics to create the Boltzmann machine, which can learn to recognise characteristic elements in a set of data. The invention became significant, for example, for classifying and creating images.

Geoffrey Hinton Facts



Ill. Niklas Elmehed © Nobel
Prize Outreach

Geoffrey Hinton
The Nobel Prize in Physics 2024

Born: 6 December 1947, London, United Kingdom

Affiliation at the time of the award: University of Toronto,
Toronto, Canada

Prize motivation: “for foundational discoveries and
inventions that enable machine learning with artificial
neural networks”

Prize share: 1/2

Work

When we talk about artificial intelligence, we often mean machine learning using artificial neural networks. This technology was originally inspired by the structure of the brain. In an artificial neural network, the brain's neurons are represented by nodes that have different values. In 1983–1985, Geoffrey Hinton used tools from statistical physics to create the Boltzmann machine, which can learn to recognise characteristic elements in a set of data. The invention became significant, for example, for classifying and creating images.

Demis Hassabis Facts



Demis Hassabis
The Nobel Prize in Chemistry 2024

Born: 27 July 1976, London, United Kingdom

Affiliation at the time of the award: Google DeepMind,
London, United Kingdom

Prize motivation: “for protein structure prediction”

Prize share: 1/4

Ill. Niklas Elmehed © Nobel
Prize Outreach

Work

Proteins control and drive all the chemical reactions that together are the basis of life. Proteins generally consist of 20 different amino acids. These are linked together in long strings that fold up to make a three-dimensional structure. In 2020, Demis Hassabis and John Jumper presented an AI model called AlphaFold2. With its help, they have been able to predict the structure of virtually all known proteins. AlphaFold2 has been widely used in many areas, including research into pharmaceuticals and environmental technology.

Demis Hassabis Facts



Demis Hassabis
The Nobel Prize in Chemistry 2024

Born: 27 July 1976, London, United Kingdom

Affiliation at the time of the award: Google DeepMind,
London, United Kingdom

Prize motivation: "for protein structure prediction"

Prize share: 1/4

Ill. Niklas Elmehed © Nobel
Prize Outreach

Work

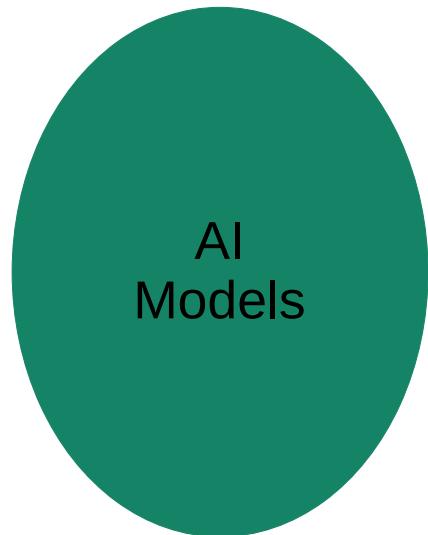
Proteins control and drive all the chemical reactions that together are the basis of life. Proteins generally consist of 20 different amino acids. These are linked together in long strings that fold up to make a three-dimensional structure. In 2020, Demis Hassabis and John Jumper presented an **AI model** called AlphaFold2. With its help, they have been able to predict the structure of virtually all known proteins. AlphaFold2 has been widely used in many areas, including research into pharmaceuticals and environmental technology.

Tre parole per due Nobel

- Machine Learning
- Neural Networks
- AI Model

Tre parole per due Nobel

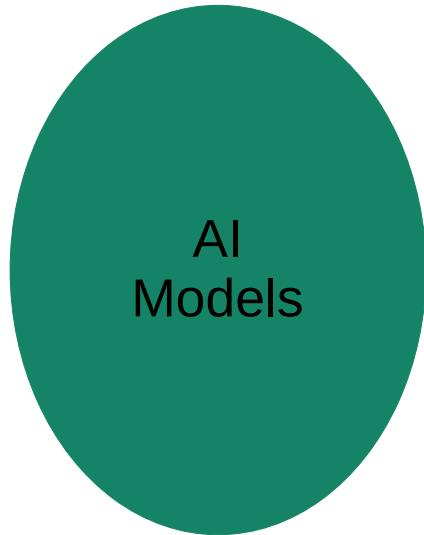
- Machine Learning
- Neural Networks
- AI Model



AlphaFold2

Tre parole per due Nobel

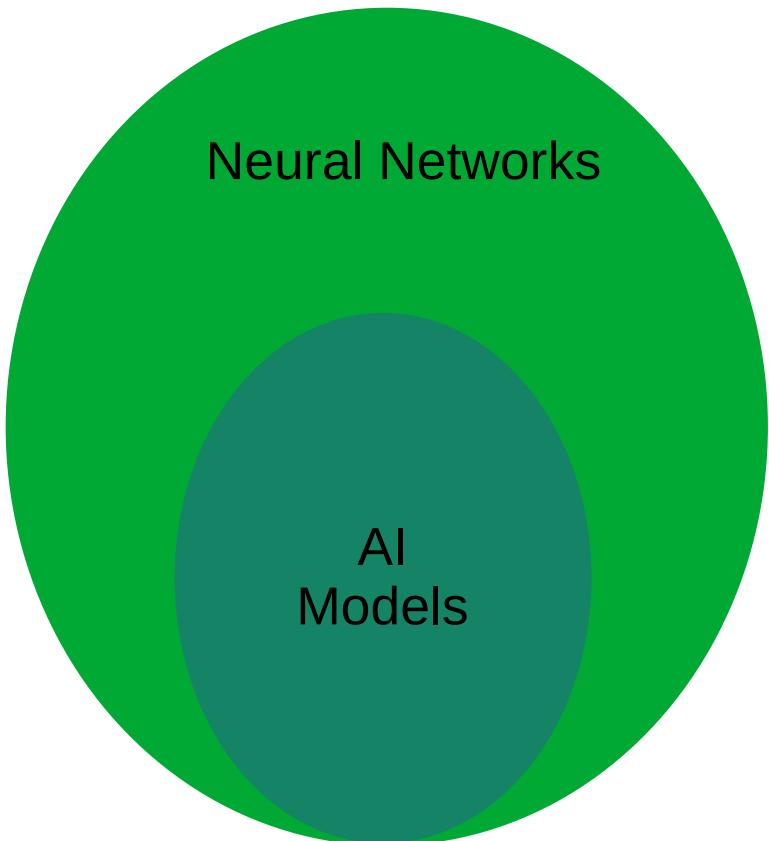
- Machine Learning
- Neural Networks
- AI Model



ChatGPT, GPT4, Gemini, Claude, Llama, Grok, DeepSeek, AlphaGo, AlphaZero, AlphaFold(2), DALL-E 3, MidJourney, Stable Diffusion, SORA...

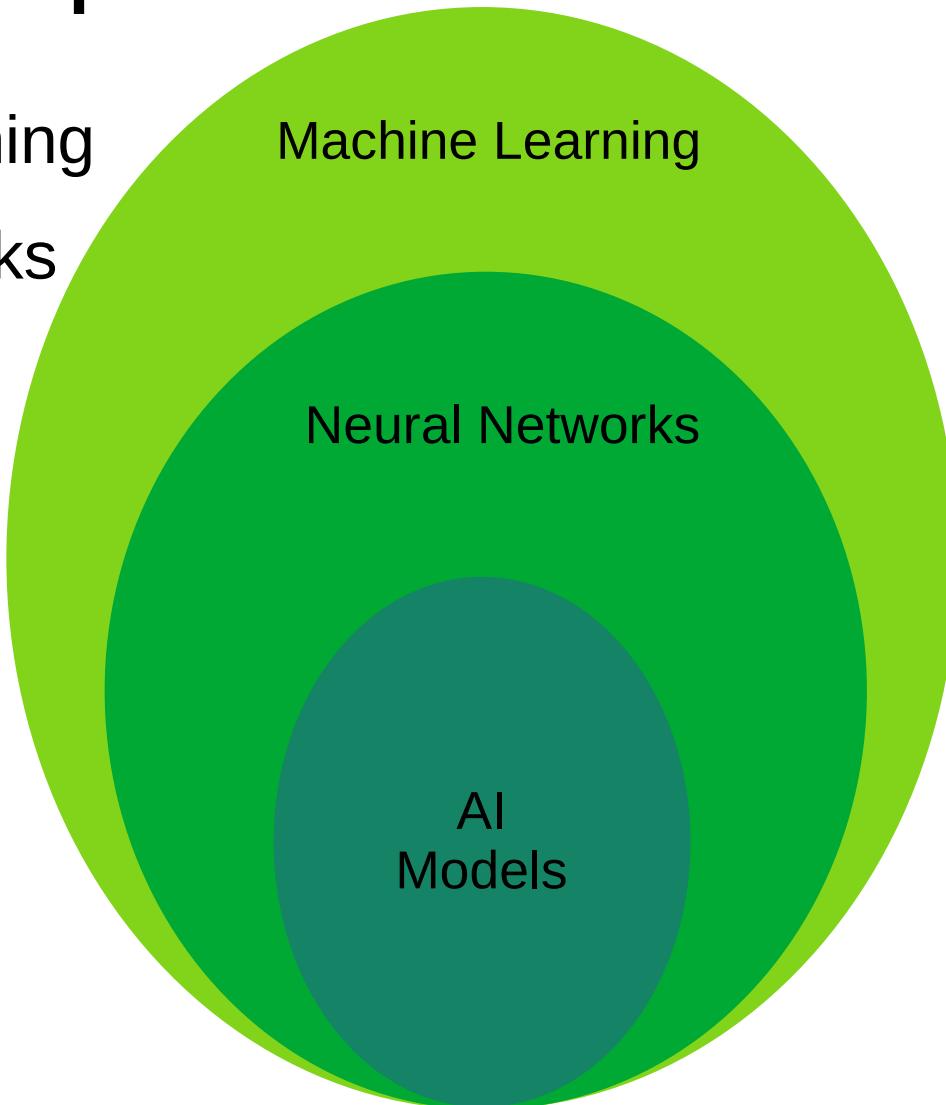
Tre parole per due Nobel

- Machine Learning
- Neural Networks
- AI Model



Tre parole per due Nobel

- Machine Learning
- Neural Networks
- AI Model







Social Media



Social Media ↔ Economia Digitale



Social Media ↔ Economia Digitale ↔ Big Tech



Machine Learning





Capitalizzazione (totale):
ca. **\$17000B**, 2nd trimestre 2025

Fatturato annuale (individuale):
centinaia di miliardi di dollari

Profitto annuale (individuale):
decine di miliardi di dollari



Pioniere

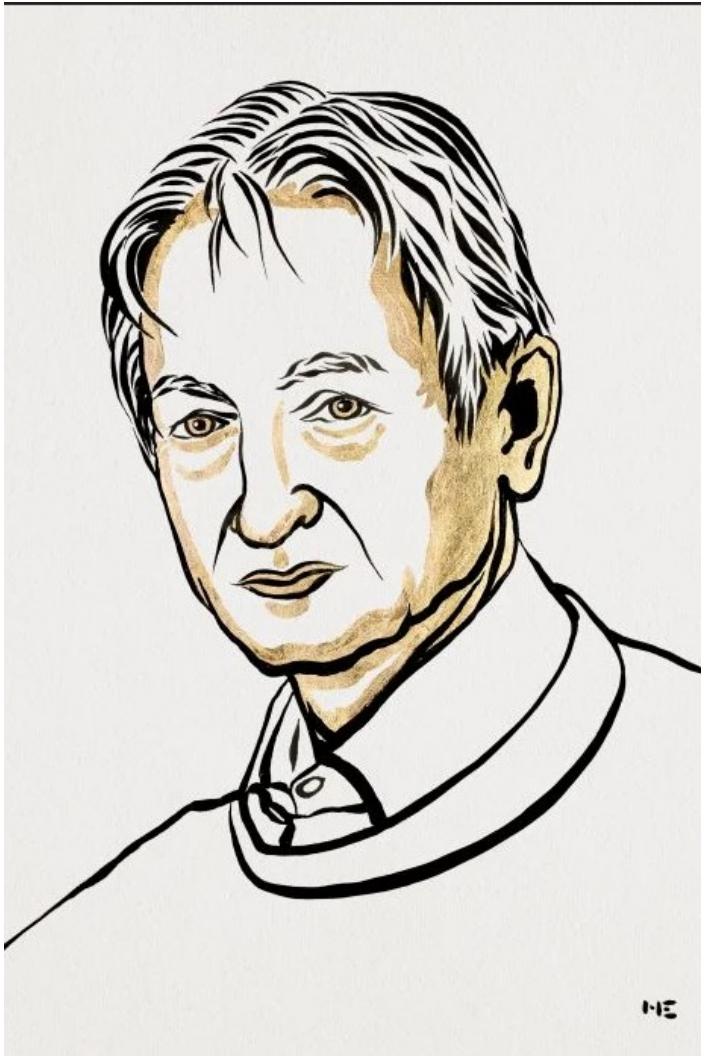


Indovinate un po' chi ci (ha) lavora(to)?

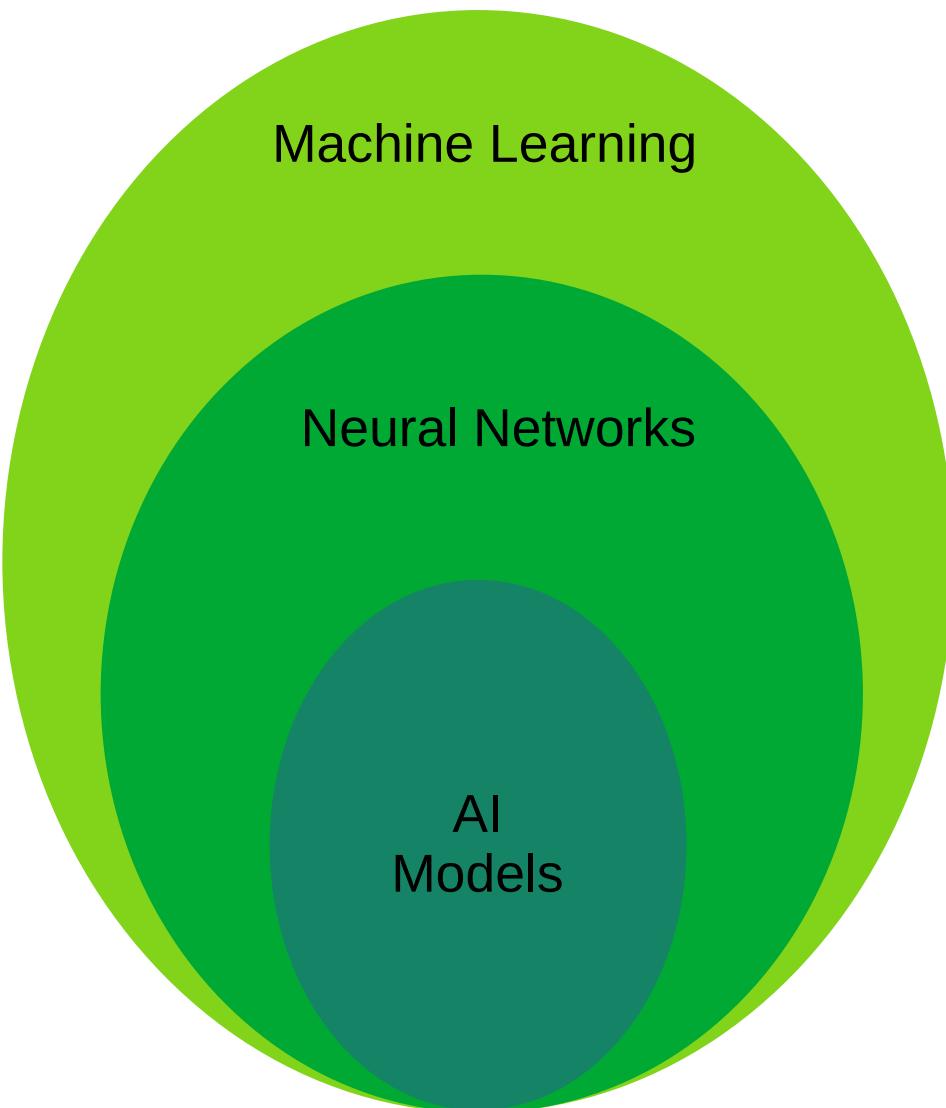
Google

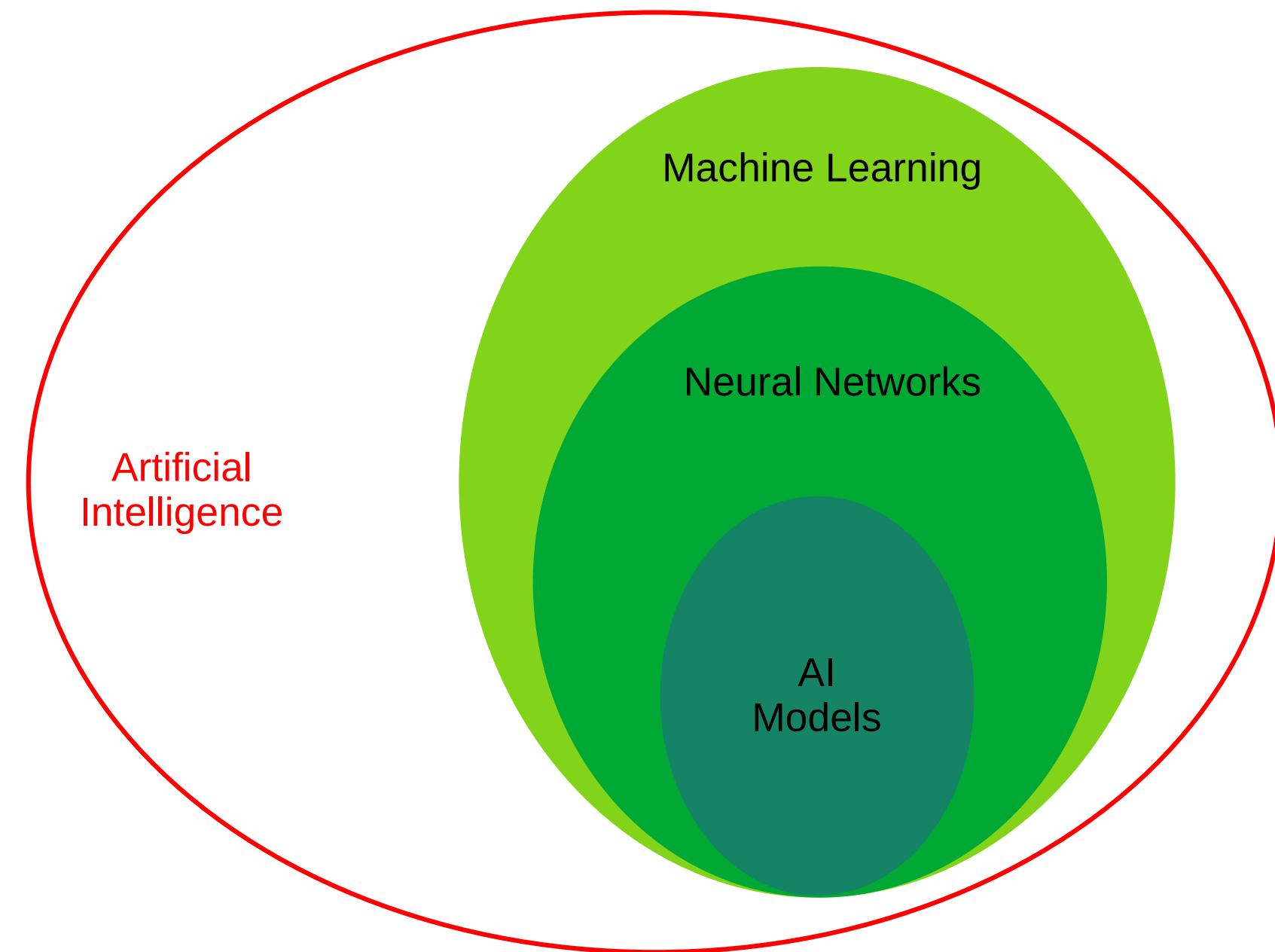


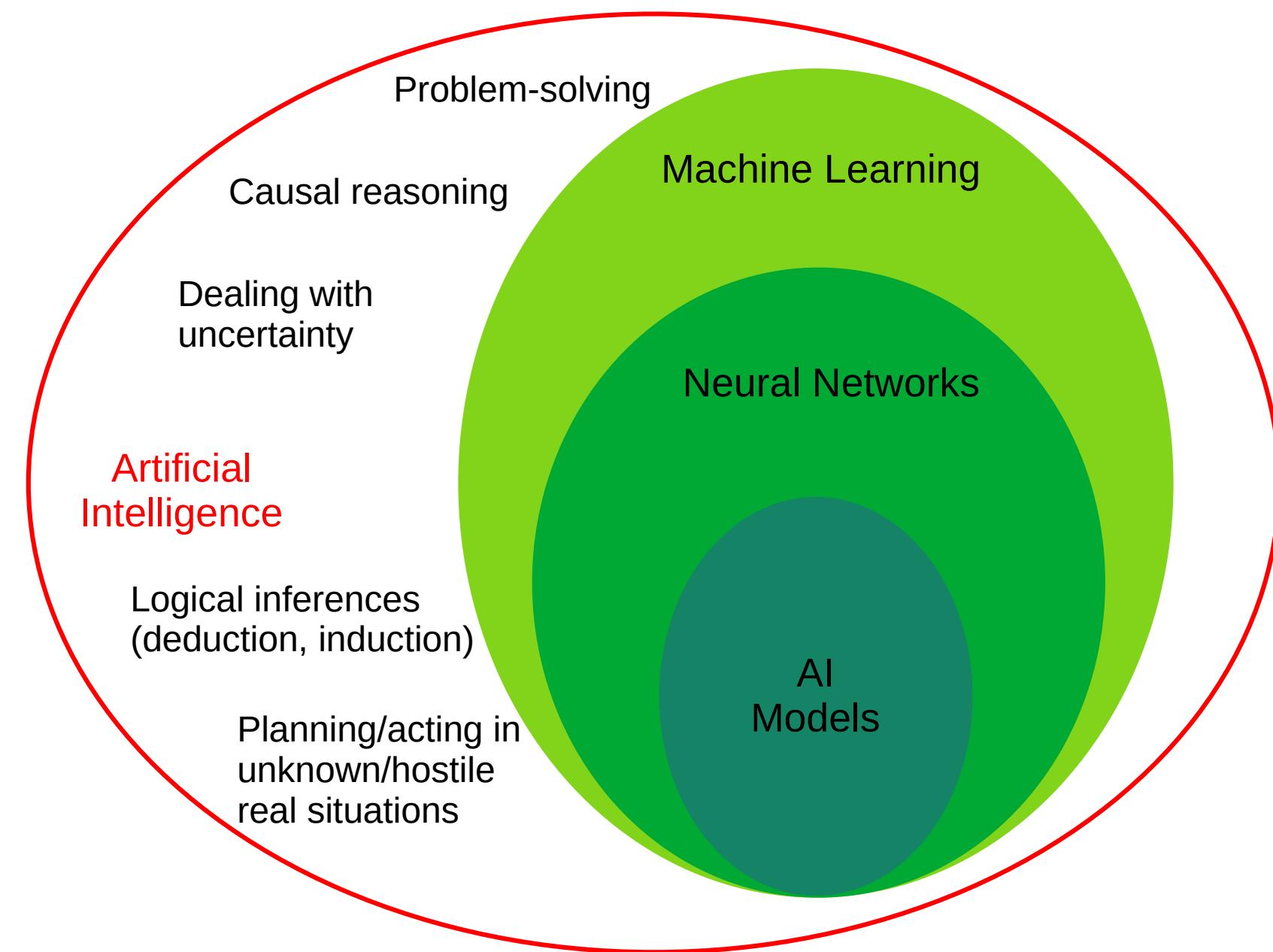
Google

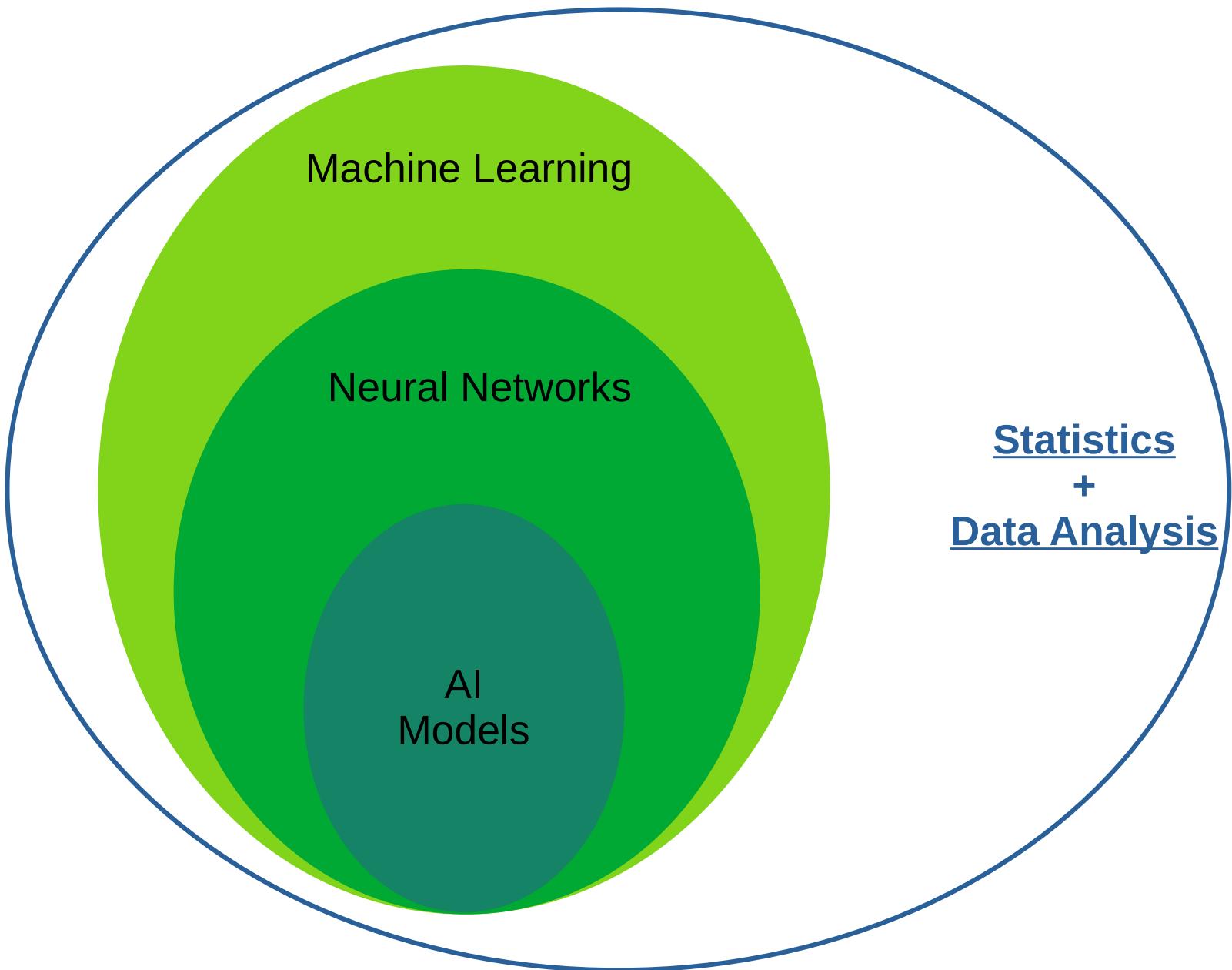


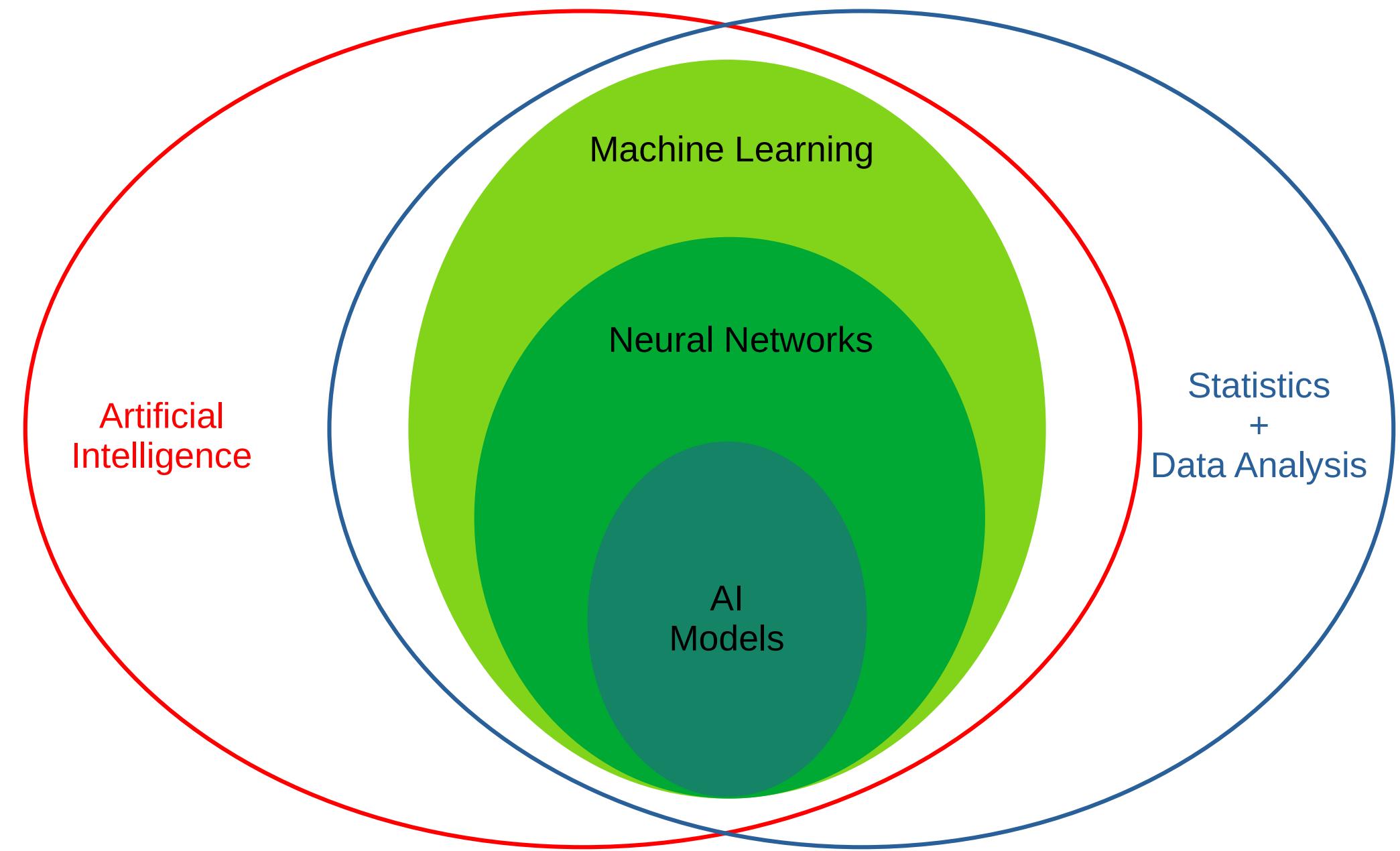
Machine Learning



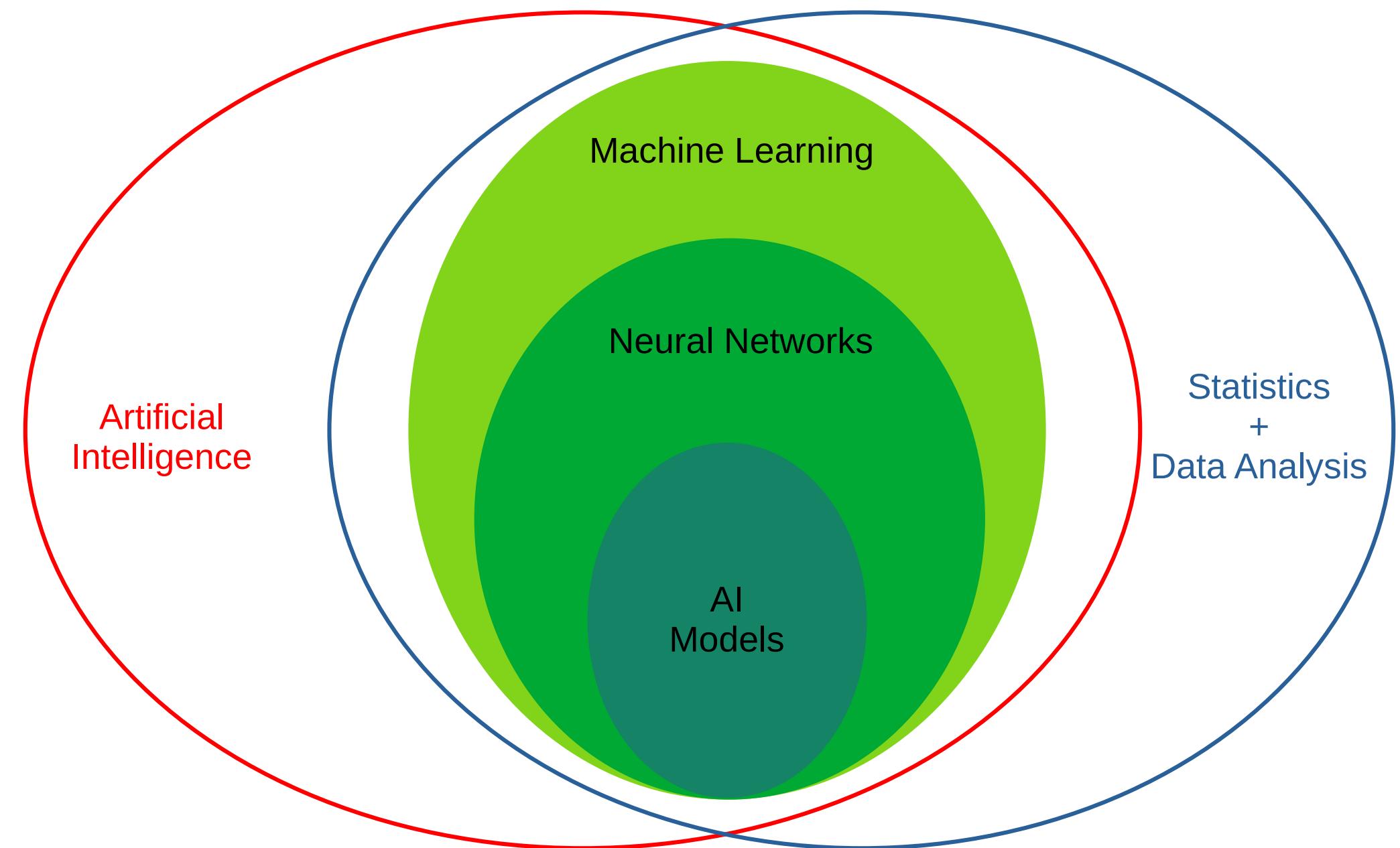




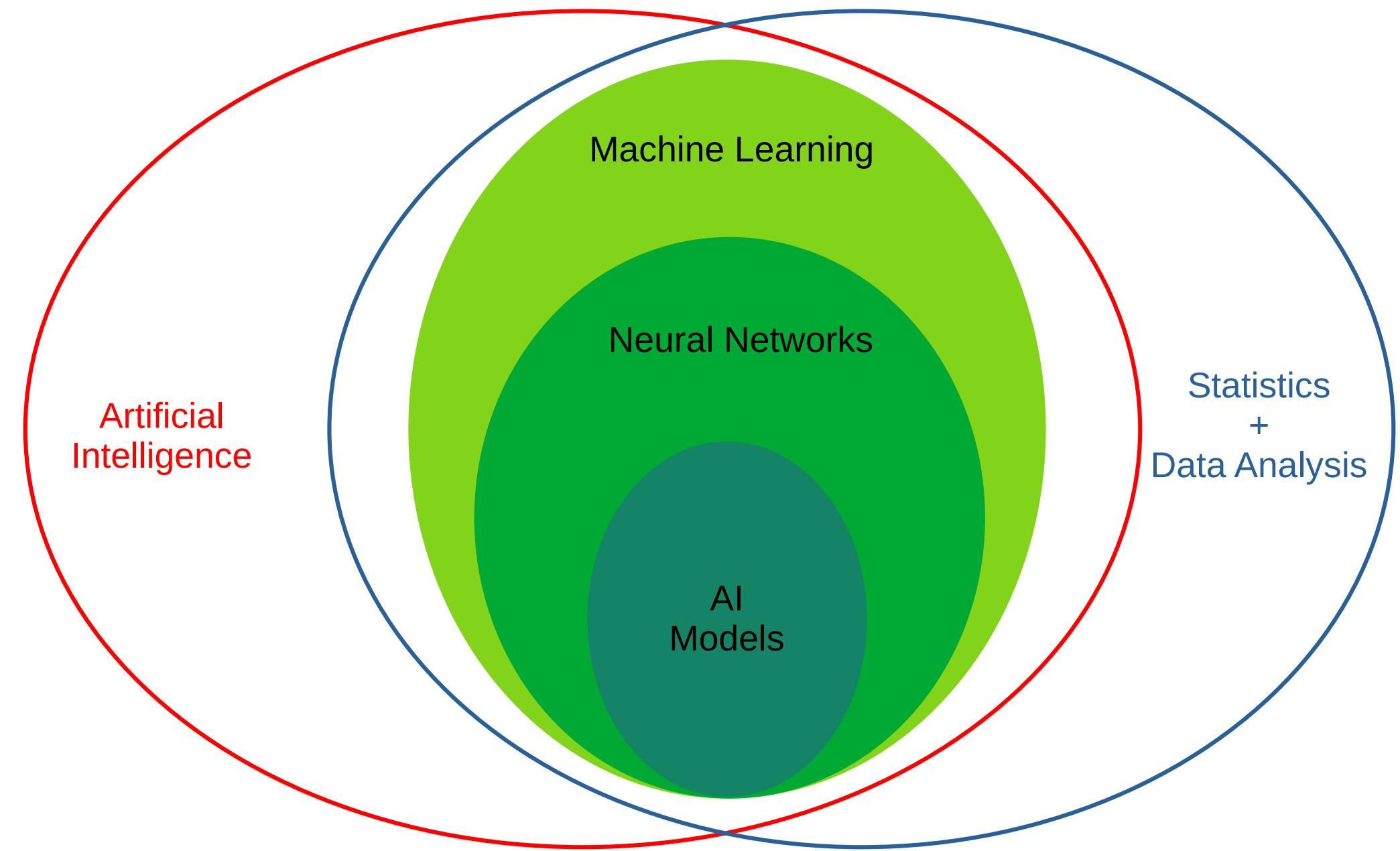




Potenti strumenti statistici, **NON** agenti pensanti



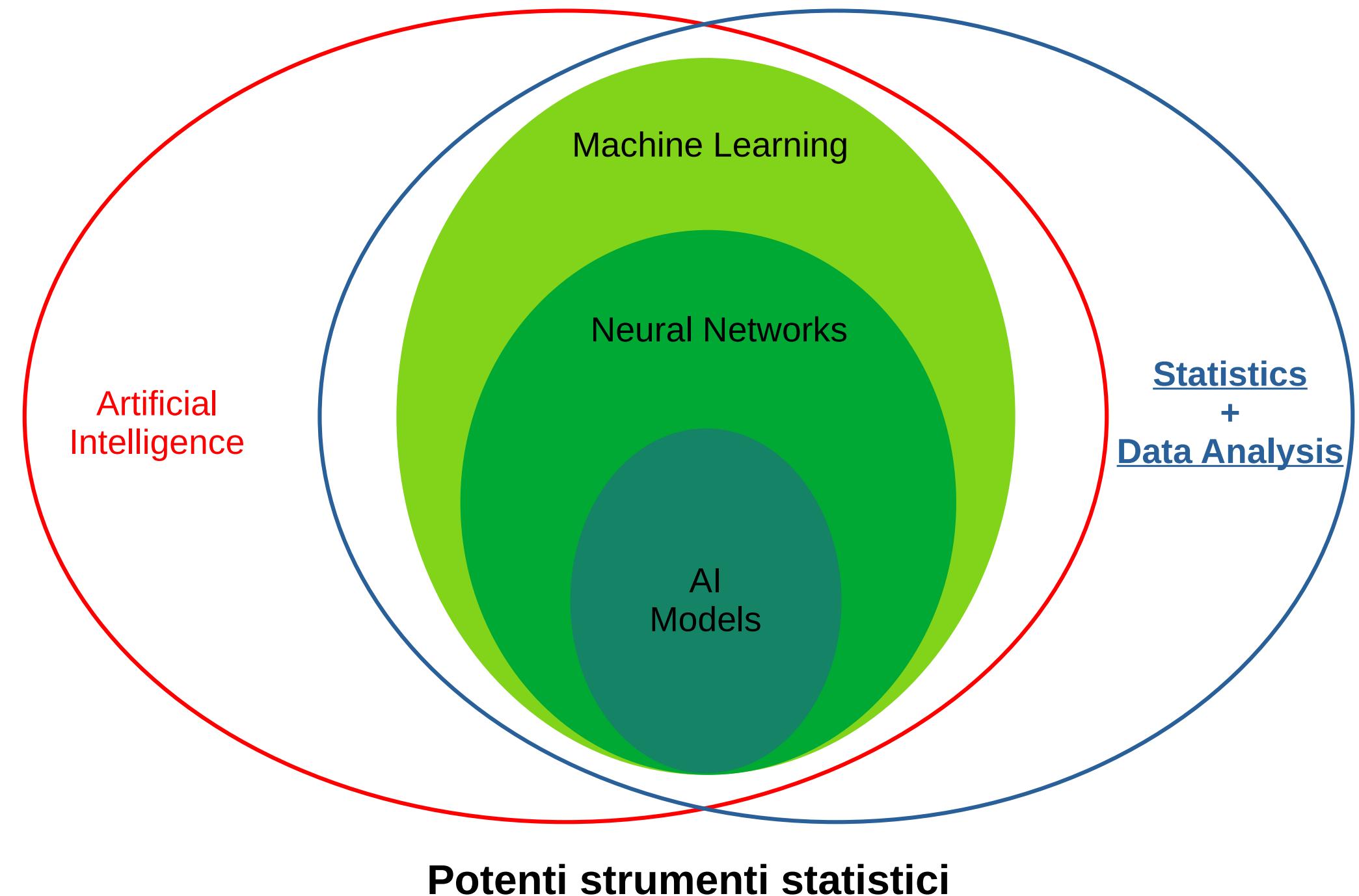
Alcune competenze intelligenti simulate (es. per inferenze, previsioni)...



...ma **NESSUNA** intelligenza di tipo umano

Un questionario per pensare

1. A partire dalle tue esperienze d'uso personale (se ve ne sono), e anche in base a quanto iniziato a vedere oggi, rifletti un momento sugli strumenti attuali di cosiddetta "Intelligenza Artificiale", disponibili al grande pubblico e che conosci. Descrivi brevemente cosa ti sembra convincente di questi strumenti e cosa invece ti lascia maggiori perplessità.
2. Secondo te, oggi, chi trae maggior beneficio dalla narrazione che racconta l'esistenza di strumenti "davvero intelligenti" e perché/come? Ti sembra convincente questa narrazione o solleva qualche dubbio?
3. Ragiona un po' sui "costi" e sugli impatti di ciò che oggi viene chiamato "Intelligenza Artificiale". Scrivi qualche riga su eventuali rischi, criticità, insostenibilità che ti sembra di poter individuare.
4. Cosa ti piacerebbe che ti venisse comunicato quando si parla della cosiddetta "Intelligenza Aritificiale" e che è invece mancante o carente?



Cosa imparo nell'unità IA.1?

- **Primi strumenti moderna Data Analysis/Analytics**
 - Tipologie di analisi
 - Descrittiva (→ Analysis)
 - Predittiva/Prescrittiva (→ Analytics)
 - Preparazione dei dati (“Data Wrangling”)
 - Visualizzazione dei dati per primi rapporti statistici
 - “Affaccio” sul Machine Learning
- **L'uso a livello iniziale di alcuni strumenti di lavoro**
 - Python-based (pandas, Matplotlib, seaborn, Plotly, scikit-learn)

Cosa imparo nell'unità IA.1?

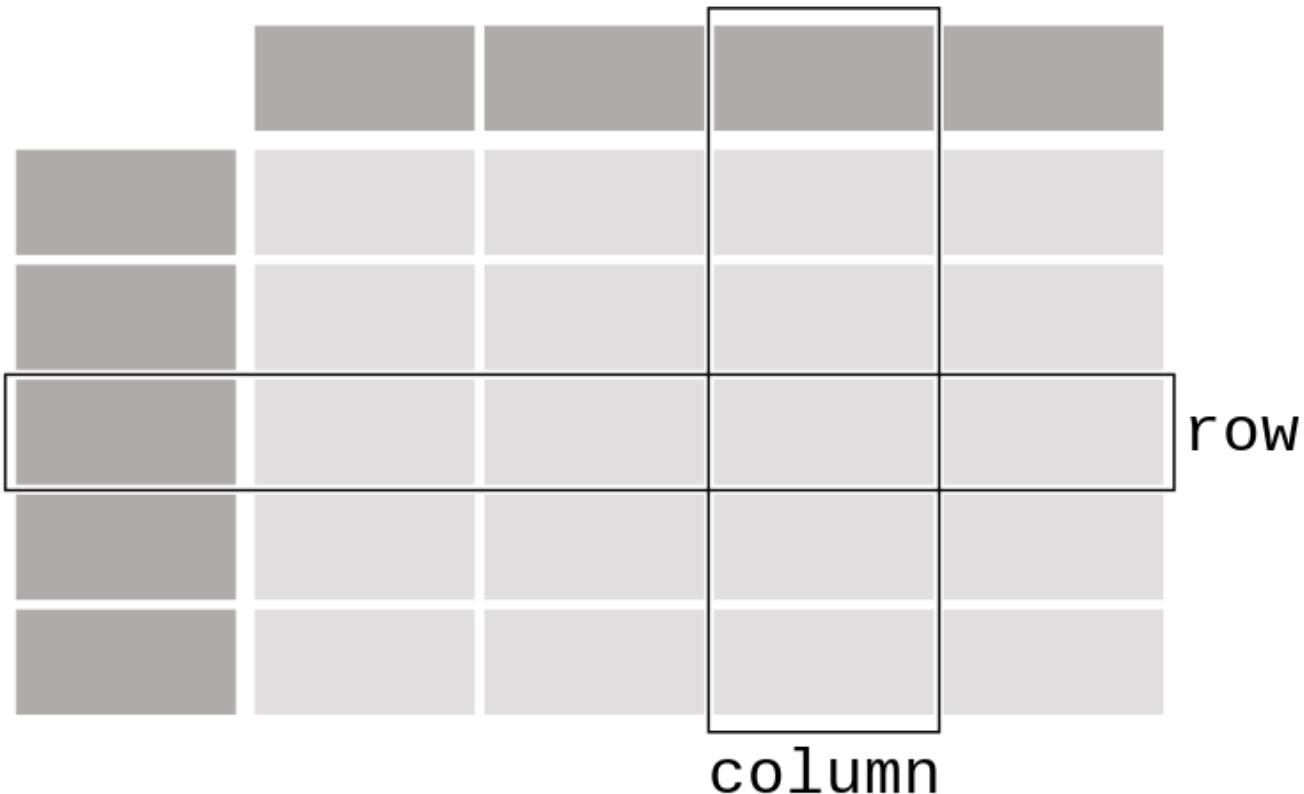
- **Primi strumenti moderna Data Analysis/Analytics**
 - Tipologie di analisi
 - Descrittiva (→ Analysis)
 - Predittiva/Prescrittiva (→ Analytics)
 - Preparazione dei dati (“Data Wrangling”)
 - Visualizzazione dei dati per primi rapporti statistici
 - “Affaccio” sul Machine Learning
- **L'uso a livello iniziale di alcuni strumenti di lavoro**
 - Python-based (pandas, Matplotlib, seaborn, Plotly, scikit-learn)
 - KNIME Analytics Platform (?)

Dettagli

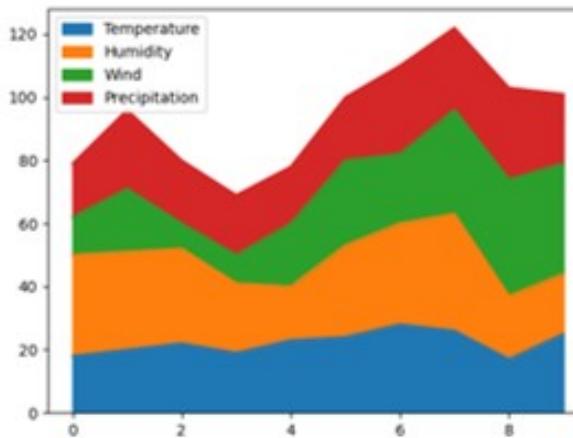
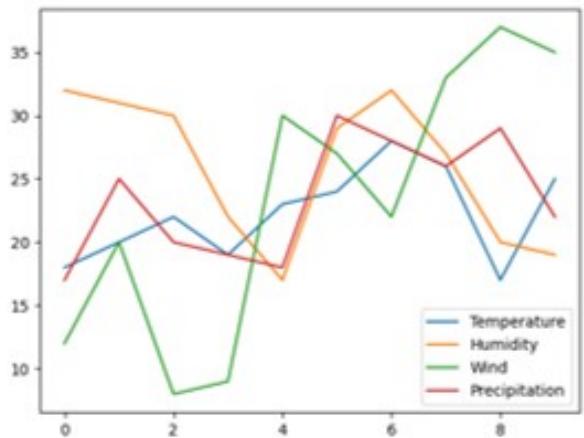
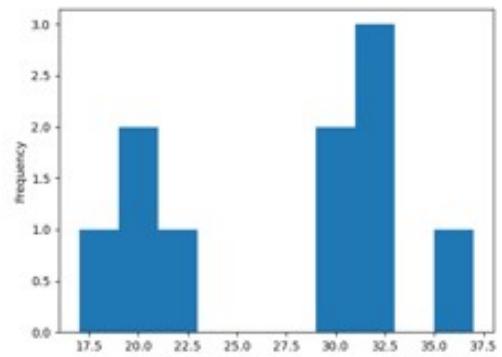
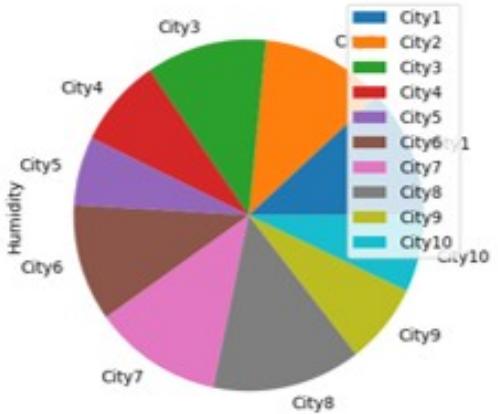
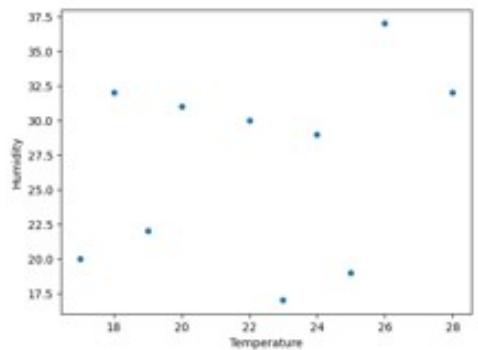
- **Durata**
 - 40 ore
- **Prerequisiti**
 - (Obbligatori) Curiosità, apertura mentale, attitudine alla novità
 - (Preferibile) Familiarità di base con:
 - Linguaggi di programmazione (es. Python, R, Java)
 - Database relazionali
 - Installazione applicazioni (es. IDE, Docker) su comuni sistemi operativi (Windows, macOS, Linux)
- **Modalità**
 - (Un po' di) Teoria
 - Esempi ed esercizi (in aula e a casa)
 - Test finale (quiz da 10 domande + mini lab)

Domande?

DataFrame

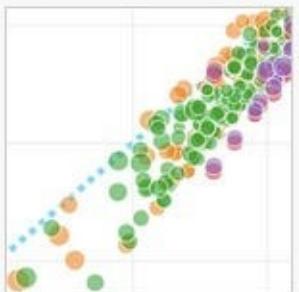


Python-based (pandas)
 pandas

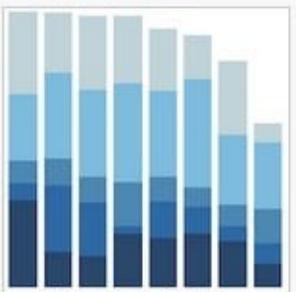


Python-based (Matplotlib)
matplotlib

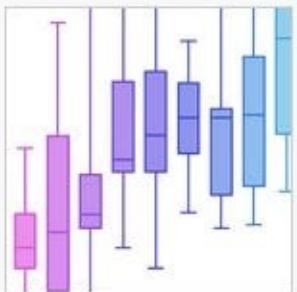
Line and Scatter Plots



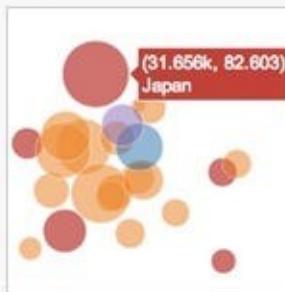
Bar Charts



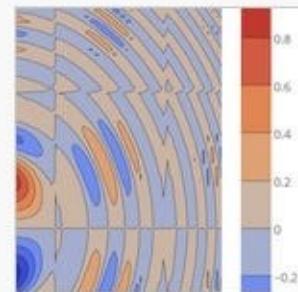
Box Plots



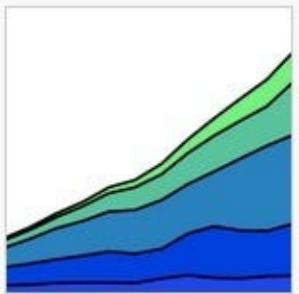
Bubble Charts



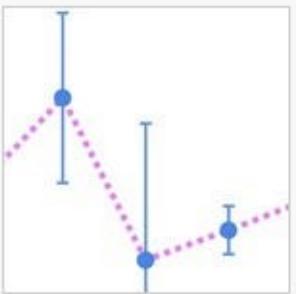
Contour Plots



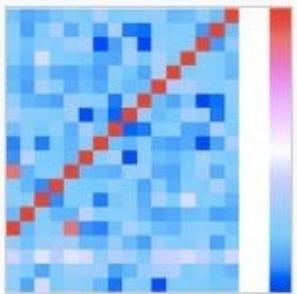
Filled Area Plots



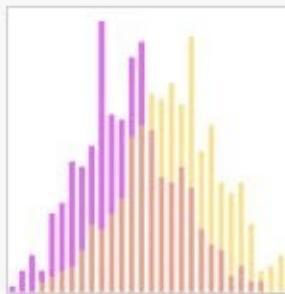
Error Bars



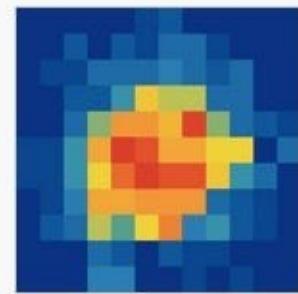
Heatmaps



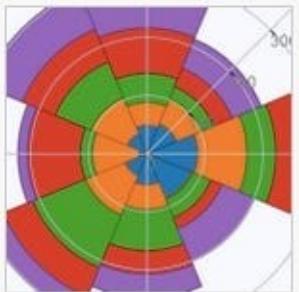
Histograms



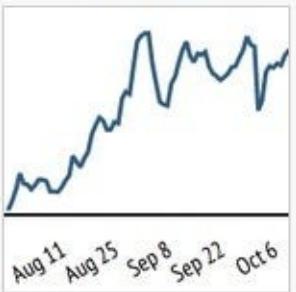
2D Histograms



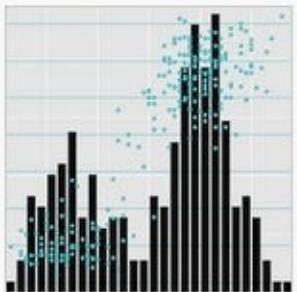
Polar Charts



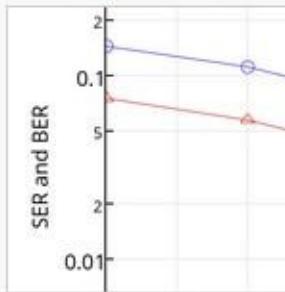
Time Series



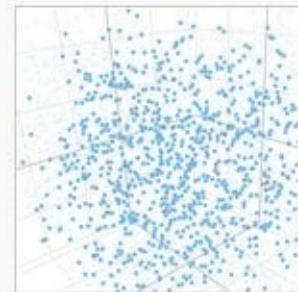
Multiple Chart Types



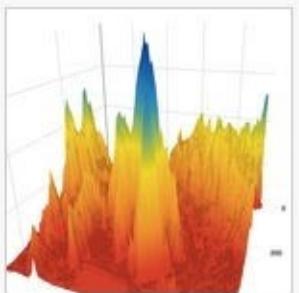
Log Plots



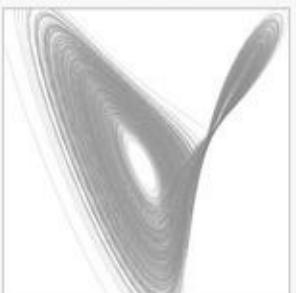
3D Scatter Plots



3D Surface Plots



3D Line Plots

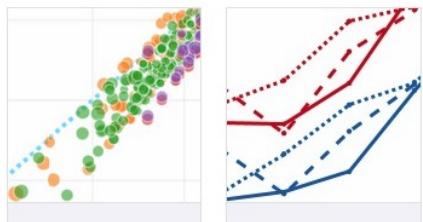


Python-based (seaborn)

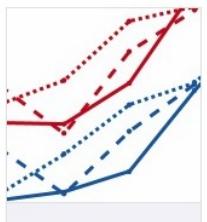


seaborn

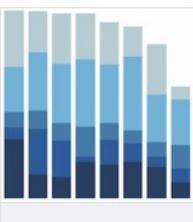
Basic Charts



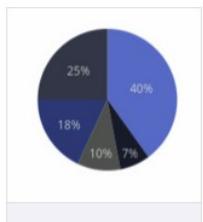
Scatter Plots



Line Charts



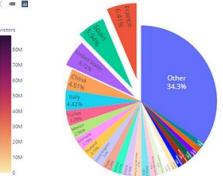
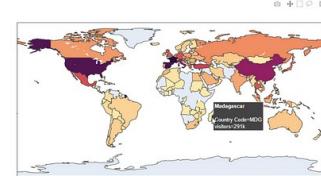
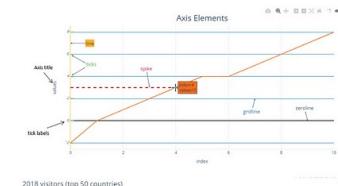
Bar Charts



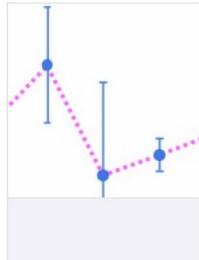
Pie Charts



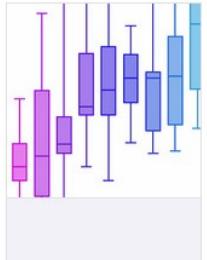
More Basic Charts



Statistical Charts



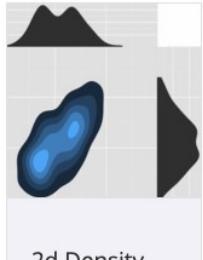
Error Bars



Box Plots



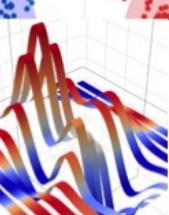
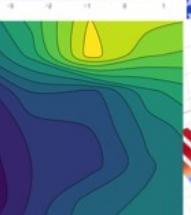
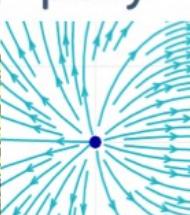
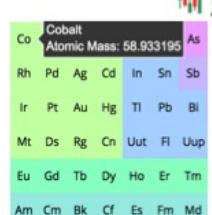
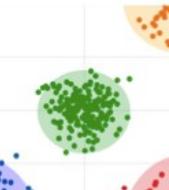
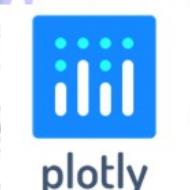
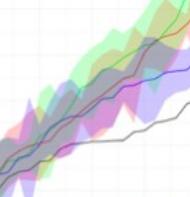
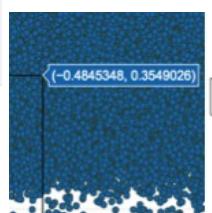
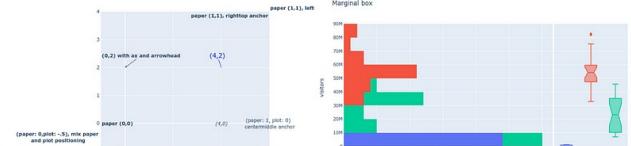
Histograms



2d Density Plots



More Statistical Charts



Python-based (Plotly)



Python-based (scikit-learn)

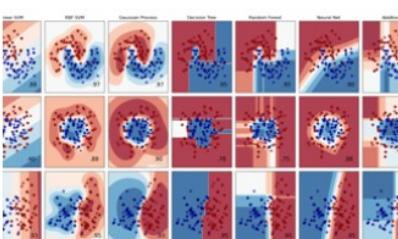


Classification

Identifying which category an object belongs to.

Applications: Spam detection, image recognition.

Algorithms: [Gradient boosting](#), [nearest neighbors](#), [random forest](#), [logistic regression](#), and more...



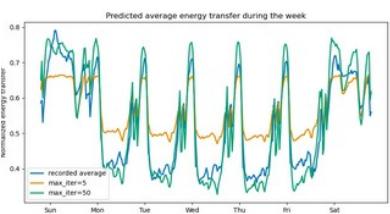
Examples

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, stock prices.

Algorithms: [Gradient boosting](#), [nearest neighbors](#), [random forest](#), [ridge](#), and more...



Examples

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, grouping experiment outcomes.

Algorithms: [K-Means](#), [HDBSCAN](#), [hierarchical clustering](#), and more...



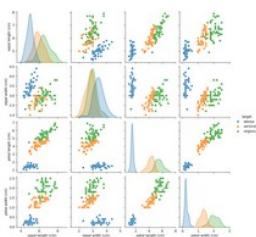
Examples

Dimensionality reduction

Reducing the number of random variables to consider.

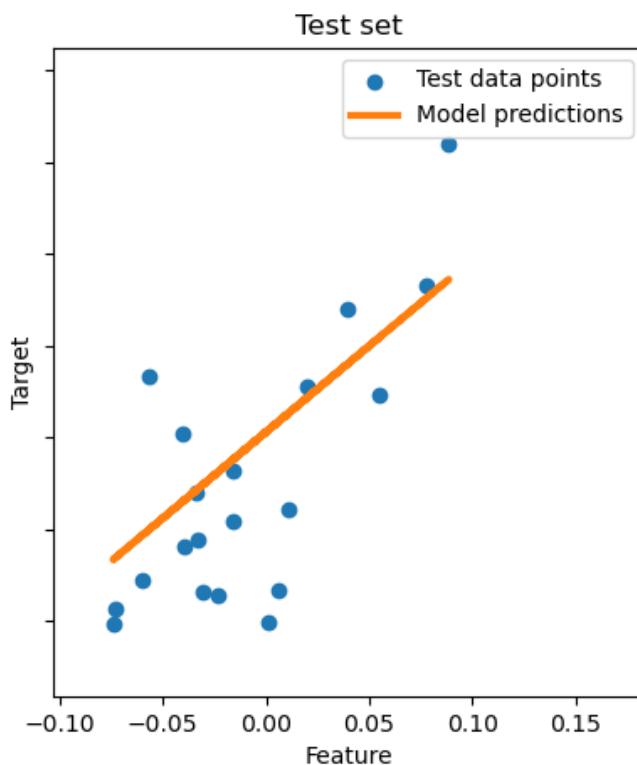
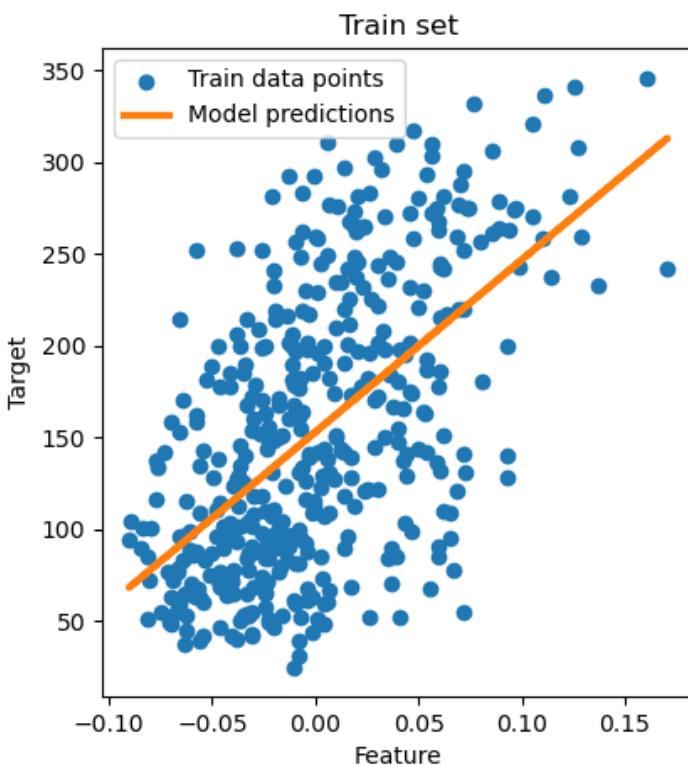
Applications: Visualization, increased efficiency.

Algorithms: [PCA](#), [feature selection](#), [non-negative matrix factorization](#), and more...



Examples

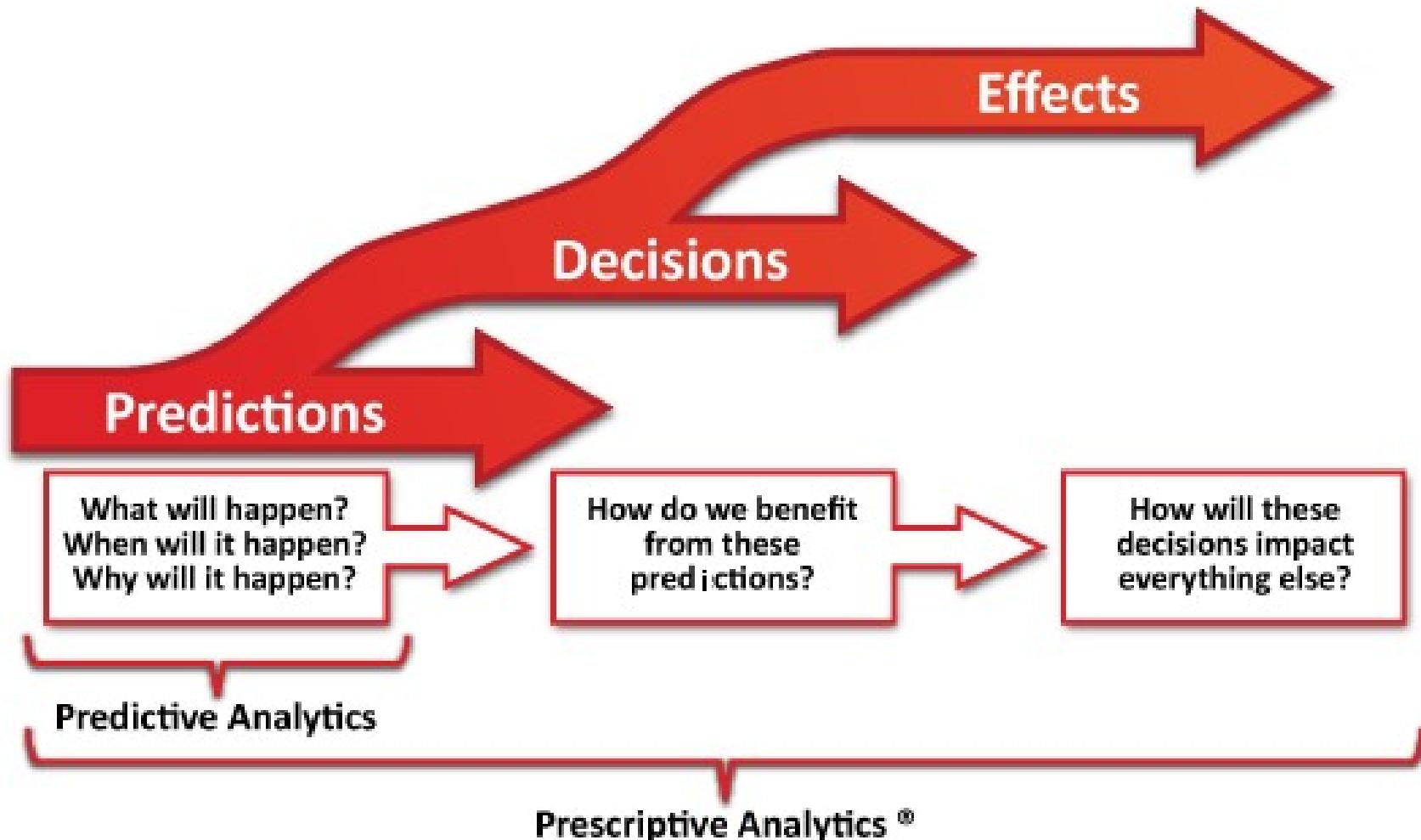
Linear Regression



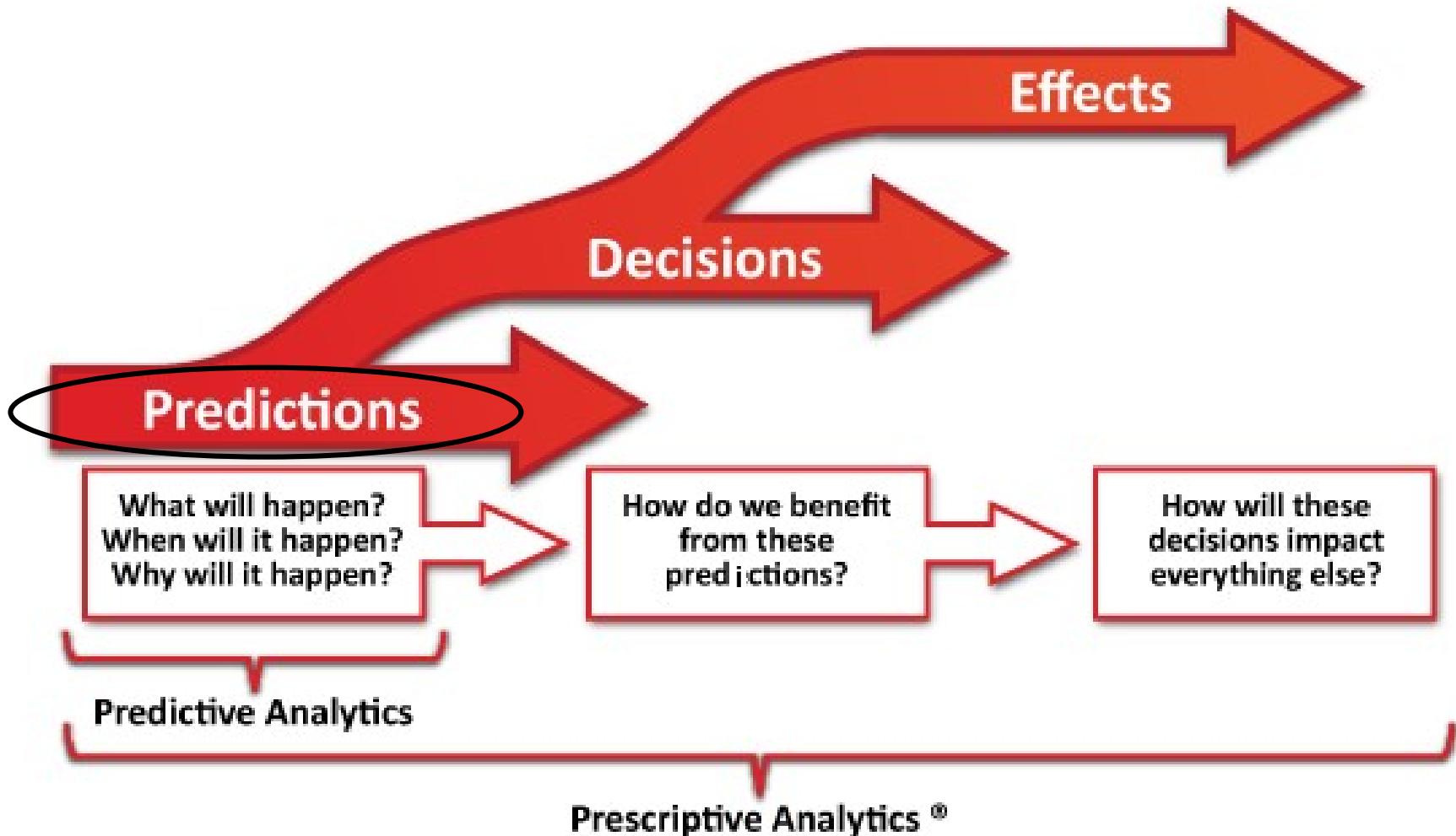
Data analysis vs. Data analytics

- Data analysis
 - Dati strutturati (modello relazionale)
 - Tecnologie SQL (RDBMS, BI, DWH)
 - Focus (principalmente) sul passato
 - “Human-driven evidence” (es. insiemi di regole)
- Data analytics
 - Dati strutturati e non strutturati (es. data lakes)
 - Tecnologie SQL e NoSQL
 - Focus (principalmente) sul futuro
 - “Machine-driven evidence” (es. modelli intelligenti automatici)

Data analytics...e il futuro!



Data analytics...e il futuro!



Data analytics...e il futuro!



III. Niklas Elmehed © Nobel
Prize Outreach

Demis Hassabis

The Nobel Prize in Chemistry 2024

Born: 27 July 1976, London, United Kingdom

Affiliation at the time of the award: Google DeepMind,
London, United Kingdom

Prize motivation: "for protein structure prediction"

Prize share: 1/4



Data analysis vs. Data analytics

- Data analysis (passato)

- Descrittiva
 - Visualizzazioni
 - Report
 - Dashboard interattive

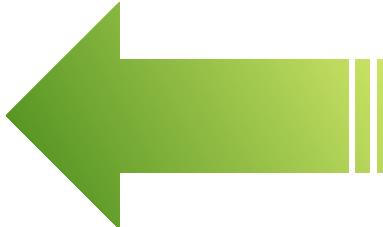
- Data analytics (futuro)

- Predittiva
 - Diagnostica
 - Segmentazioni
 - Previsioni
 - Prescrittiva
 - Scelte/azioni ottimizzate
 - Modelli di propensione
 - Simulatori
 - Agenti generativi

Data analysis vs. Data analytics

- Data analysis (passato)

- Descrittiva
 - Visualizzazioni
 - Report
 - Dashboard interattive



Unità IA.1

- Data analytics (futuro)

- Predittiva
 - Diagnostica
 - Segmentazioni
 - Previsioni
 - Prescrittiva
 - Scelte/azioni ottimizzate
 - Modelli di propensione
 - Simulatori
 - Agenti generativi



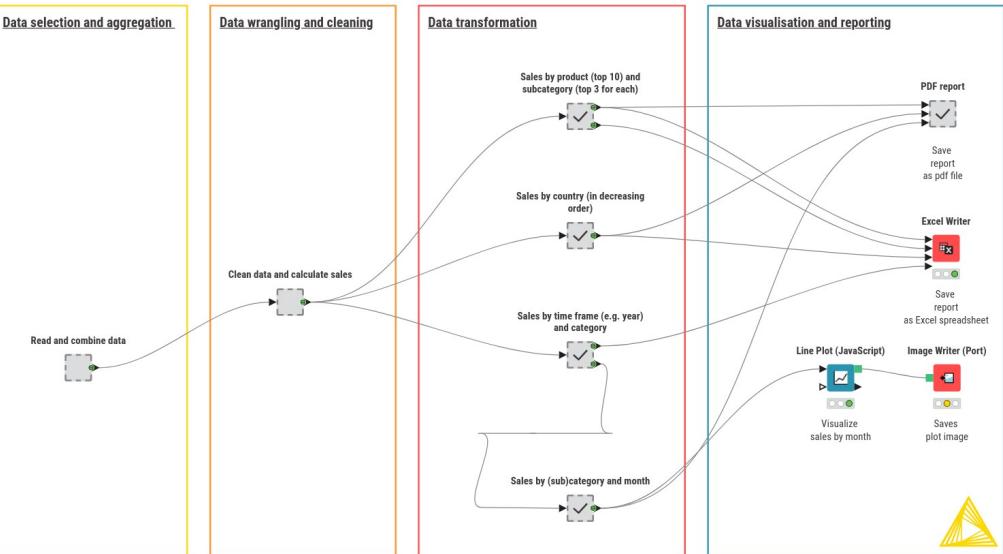
Esempi di analisi

Provate ad indovinare il tipo di analisi in ogni esempio

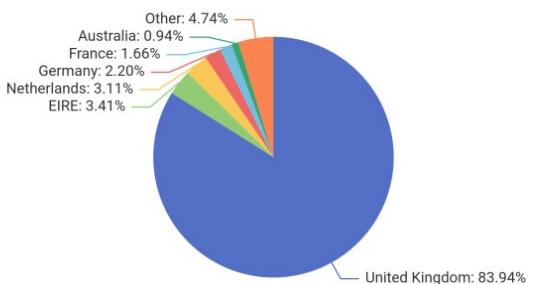
KNIME AP (Manipolazione)

Rows: 675802 | Columns: 10

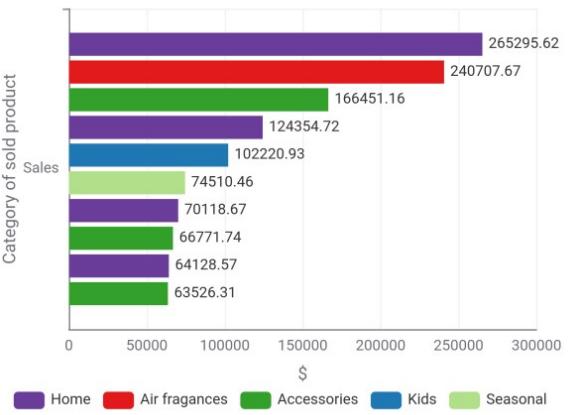
#	RowID	Invoice	StockCode	Quantity	Price	Customer_ID	Invoice_time	Country	Description	Category	Subcategory
4/4101	Row47	554794	ZZ491	90	1.65	16966	D26/5/19T12:07:00	United Kingdom	PVA UP 12 COLOURED P	Stationery	Desktop Stationery
474151	Row47	554789	22561	48	0.42	16966	D26/5/19T12:07:00	United Kingdom	WOODEN SCHOOL COOL	Stationery	Desktop Stationery
474151	Row47	554789	21703	120	0.42	16966	D26/5/19T12:07:00	United Kingdom	BAG 125g SWIRLY MARBL	Accessories	Bags
474151	Row47	554789	22560	48	1.06	16966	D26/5/19T12:07:00	United Kingdom	TRADITIONAL MODELLINI	Hobbies	Crafts
474151	Row47	C554790	22961	-4	1.45	14911	D26/5/19T12:10:00	EIRE	JAM MAKING SET PRINTI	Home	Kitchen
474151	Row47	554791	22961	4	1.45	14911	D26/5/19T12:12:00	EIRE	JAM MAKING SET PRINTI	Home	Kitchen
474151	Row47	554792	21080	24	0.85	12547	D26/5/19T12:38:00	Spain	SET/20 RED RETROSPOT	Home	Paper
474161	Row47	554792	POST	1	28	12547	D26/5/19T12:38:00	Spain	POSTAGE	Others	Code
474161	Row47	C554793	84077	-96	0.11	12901	D26/5/19T12:41:00	United Kingdom	WORLD WAR 2 GLIDERS A	Kids	Toys
474161	Row47	554794	23209	10	1.65	12720	D26/5/19T12:41:00	Germany	LUNCH BAG DOLEY PATT	Accessories	Bags
474161	Row47	554794	23203	10	2.08	12720	D26/5/19T12:41:00	Germany	JUMBO BAG VINTAGE DO	Accessories	Bags
474161	Row47	554794	22963	12	0.85	12720	D26/5/19T12:41:00	Germany	JAM JAR WITH GREEN LIL	Home	Kitchen
474161	Row47	554794	22962	12	0.85	12720	D26/5/19T12:41:00	Germany	JAM JAR WITH PINK LID	Home	Kitchen
474161	Row47	554794	22961	12	1.45	12720	D26/5/19T12:41:00	Germany	JAM MAKING SET PRINTI	Home	Kitchen
474161	Row47	554794	23111	2	12.5	12720	D26/5/19T12:41:00	Germany	PARISIENNE SEWING BOX	Hobbies	Crafts
474161	Row47	554794	23112	2	7.5	12720	D26/5/19T12:41:00	Germany	PARISIENNE CURIO CABIN	Home	Furniture



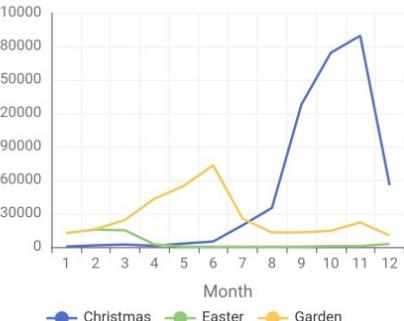
Top sales by country (%)



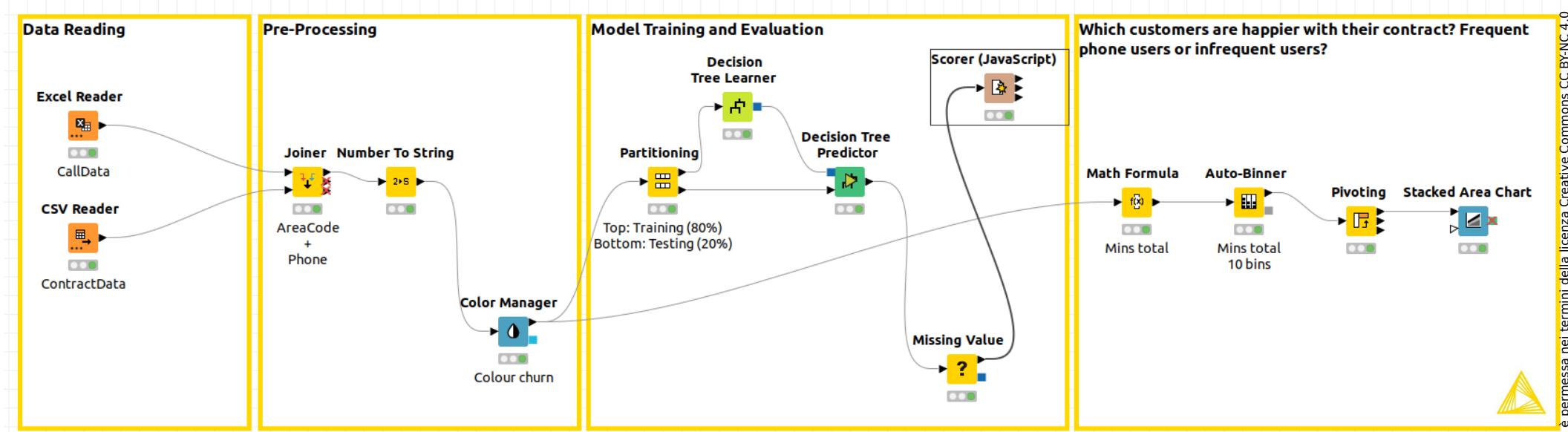
Top 10 sales by product (in \$)



Sales of some items across the year (in \$)

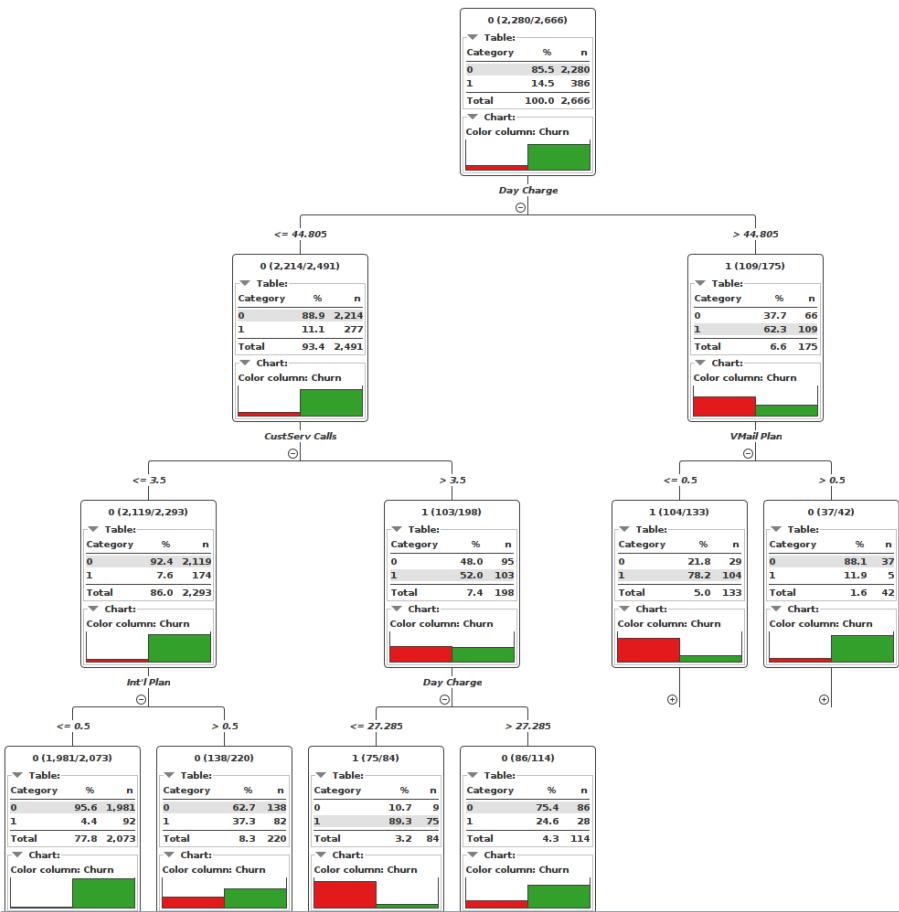


Apriamo un report pdf...



La distribuzione di questa presentazione è permessa nei termini della licenza Creative Commons CC BY-NC 4.0

KNIME AP (Classificazione)



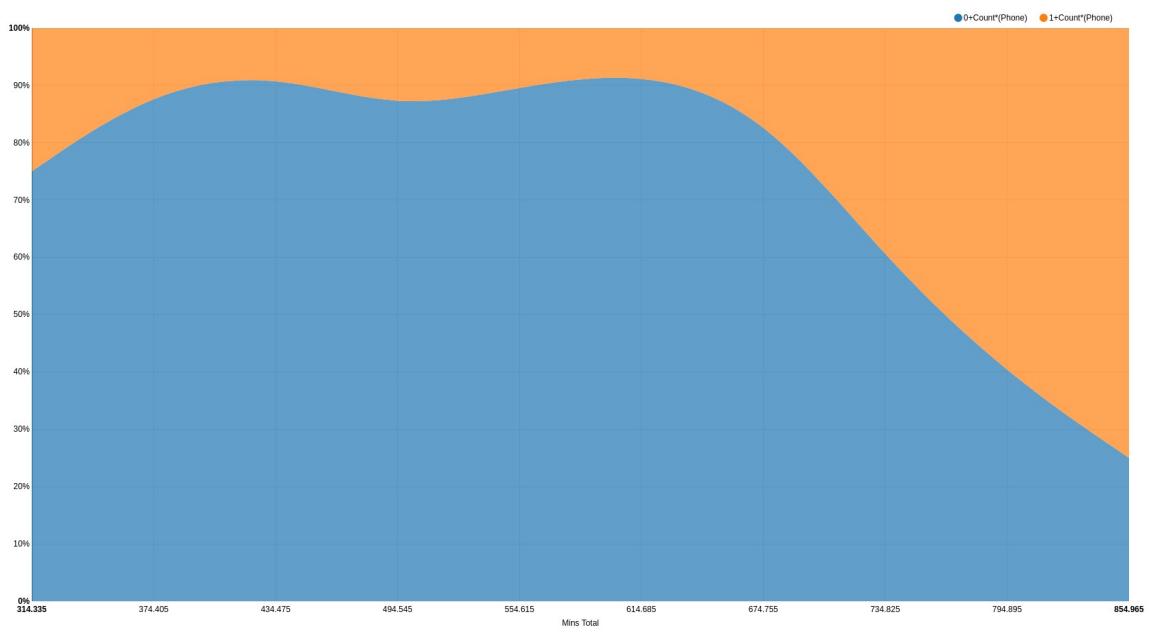
Scorer View

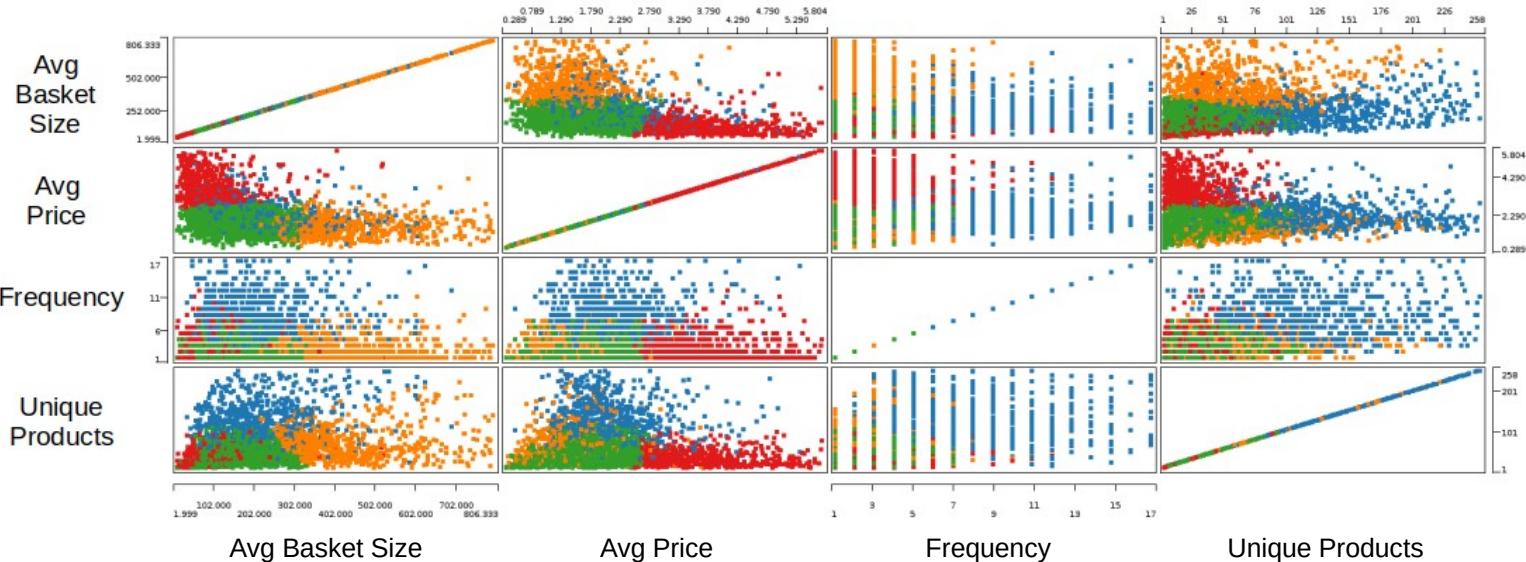
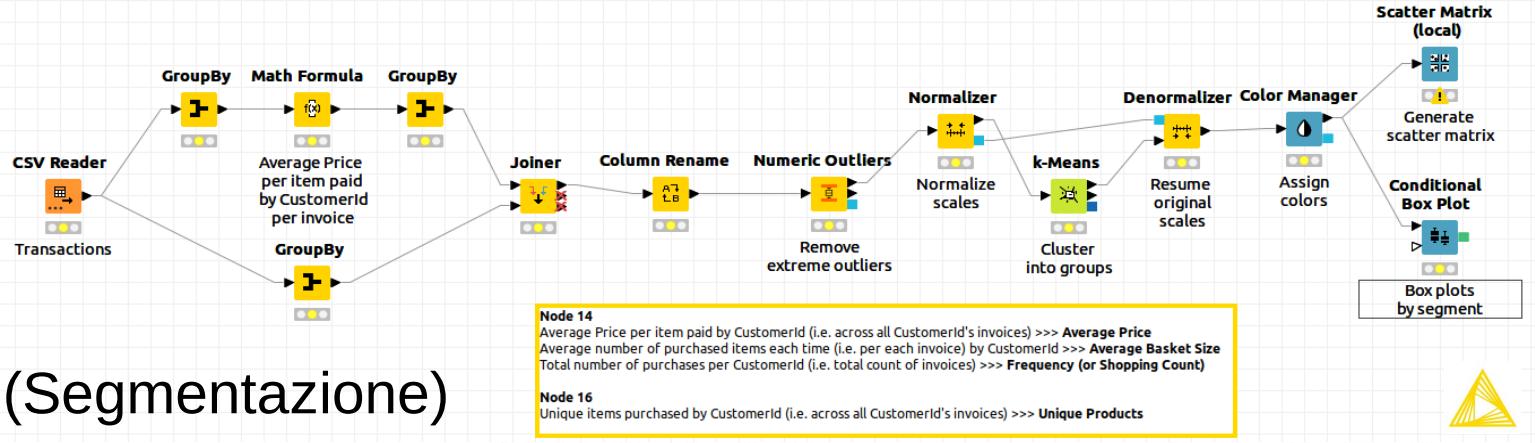
Confusion Matrix

		0 (Predicted)	1 (Predicted)	
		0 (Actual)	521	95.07%
		1 (Actual)	46	50.54%
		91.89%		63.51%

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
88.61%	11.39%	0.498	568	73





- **Segmento/cluster blu:** ~700 clienti, alta “Frequenza acquisto”, alta “Varietà prodotti” (ABITUALI, CURIOSI, APERTI)*
- **Segmento/cluster arancione:** ~600 clienti, alta “Dimensione carrello”, bassa “Frequenza acquisto”, basso “Prezzo medio”, bassa “Varietà prodotti (ABITUALI, RISPARMIATORI, RIPETITIVI)”
- **Segmento/cluster verde:** ~1900 clienti, bassa “Frequenza acquisto”, basso “tutto” (OCCASIONALI)^
- **Segmento/cluster rosso:** ~600 customers, bassa “Dimensione carrello”, bassa “Varietà prodotti”, bassa “Frequenza acquisto”, alto “Prezzo medio” (TOP, PREMIUM, LUSSO)#

* Es. Invio frequente di comunicazioni, buoni sconto, campioni gratuiti

° Es. Segnalare novità di assortimento prodotti d'interesse, offerte speciali per grosse spese

^ Es. Sconti su prodotti già acquistati, informazione su simili/altre categorie merceologiche

Es. Enfasi su qualità e valore, segnalazione assortimenti di “alta gamma”

Demo

Una piccola dashboard con Python e Docker per cominciare a esplorare...

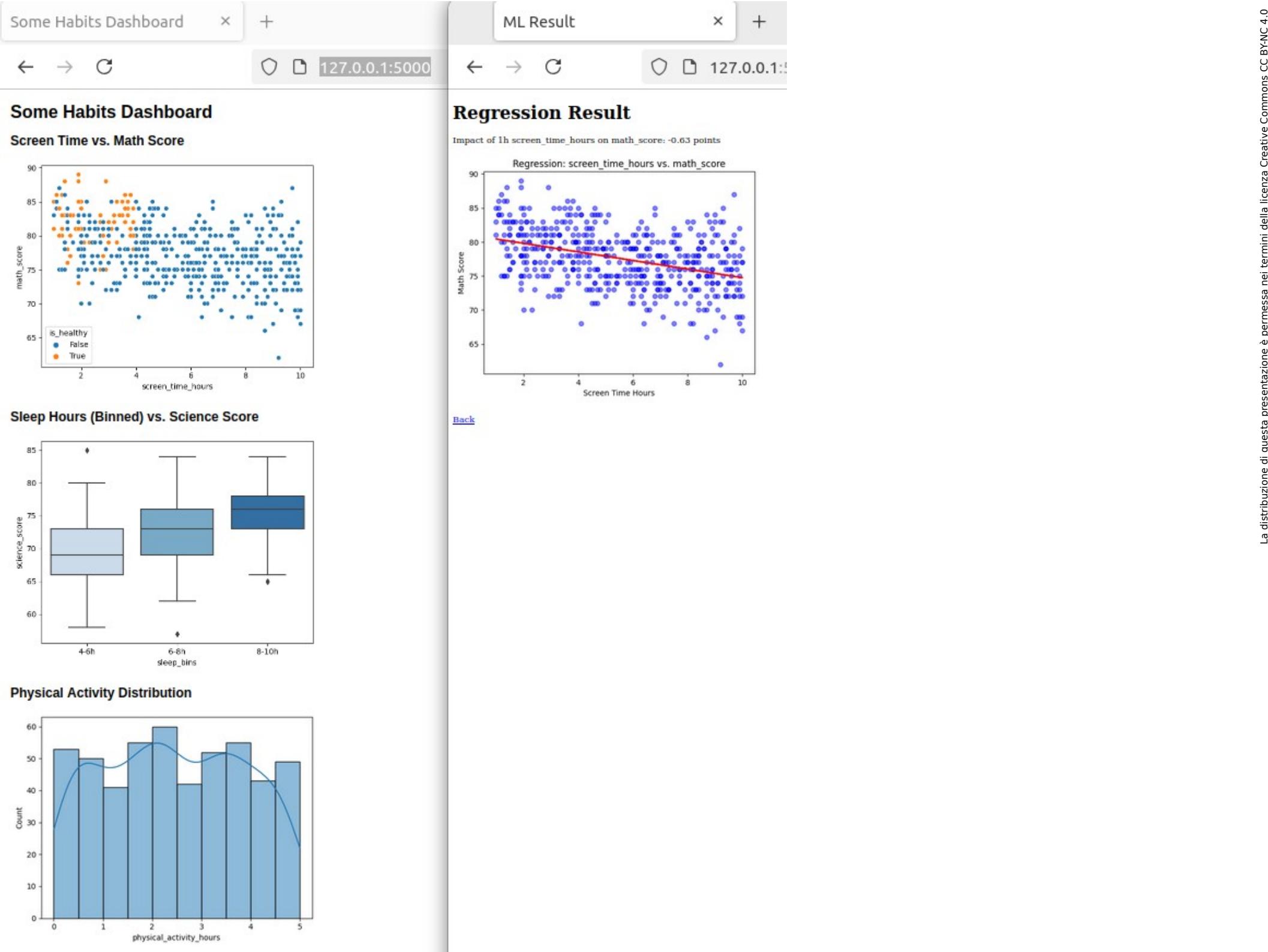
Activities Terminal dom 30 mar 15:09:14 ~/Projects/Python/Fun/Docker/docker-api-edu-demo - Spyder (Python 3.9) stefano@stefano-ThinkPad-T450s: ~/Pro... it 98 %

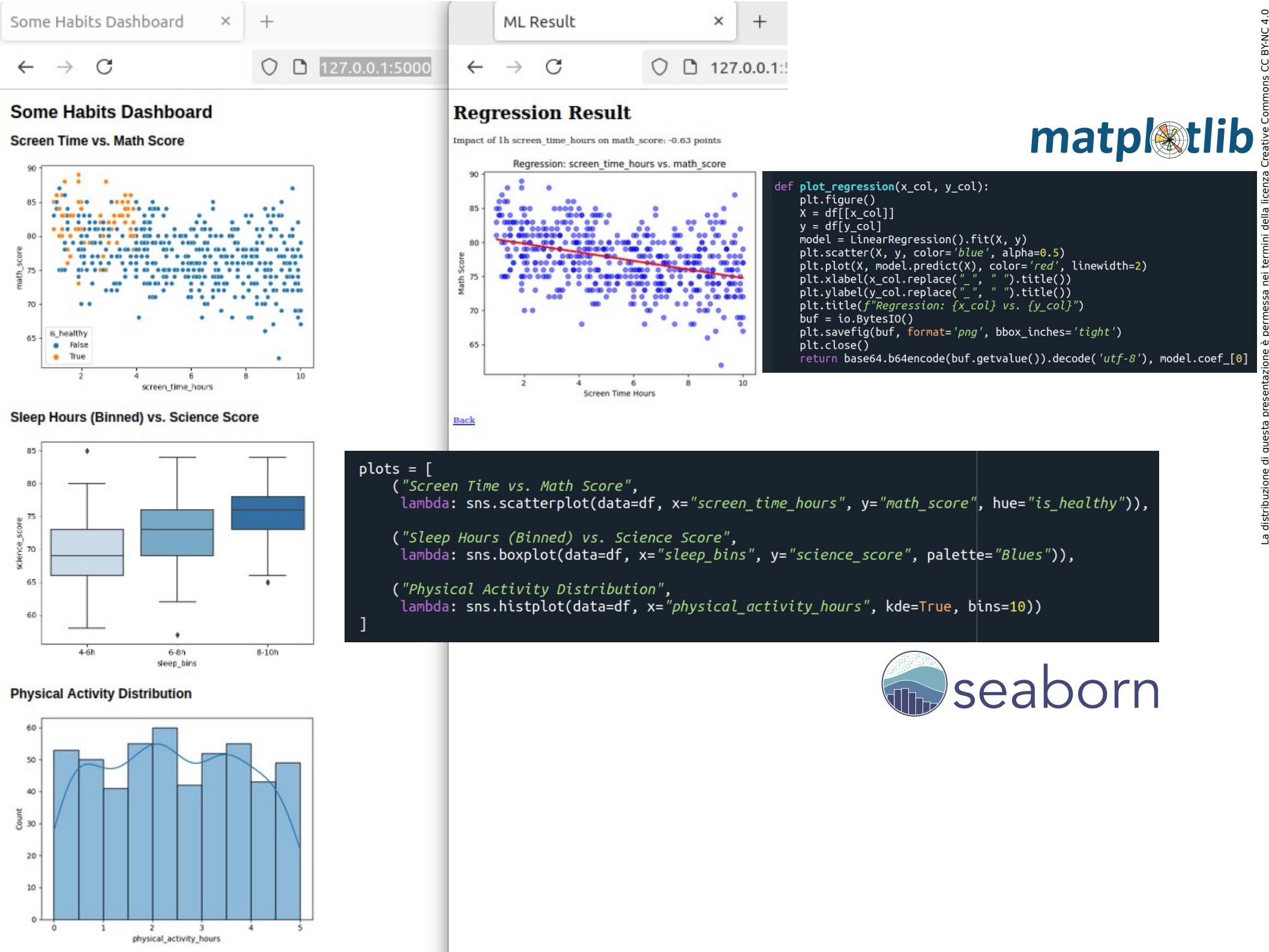
The screenshot displays a desktop environment with several windows open:

- Terminal:** Shows the command `sudo docker compose up` being run, followed by a list of created containers and their details. It also shows logs for each container starting up.
- File Browser:** Shows a project structure for "docker-api-edu-demo". The `docker-compose.yml` file is open in the code editor, defining services for data-generation, data-preprocess, data-visualize, and ml-basics.
- Code Editor:** Shows the `docker-compose.yml` file with code like:


```

1 services:
2   data-generation:
3     build: ./data-generation
4     volumes:
5       - ./data:/data
6
7   data-preprocess:
8     build: ./data-preprocess
9     volumes:
10    - ./data:/data
11    depends_on:
12      - data-generation
13
14   data-visualize:
15     build: ./data-visualize
16     volumes:
17       - ./data:/data
18     ports:
19       - "5000:5000"
20     depends_on:
21       - data-preprocess
22
23   ml-basics:
24     build: ./ml-basics
25     volumes:
26       - ./data:/data
27     ports:
28       - "5001:5001"
29     depends_on:
30       - data-preprocess
      
```
- Terminal (Bottom):** Shows the command `sudo docker container ls` being run, listing the four containers created by the compose up command.





Prima lezione

- Introduzione alla libreria **pandas**
- Concetto di DataFrame (“contenitore” di dati)
- Propedeutico ad analisi e visualizzazioni
- Primi esempi ed esercizi

Prima lezione

- Introduzione alla libreria **pandas**
- Concetto di DataFrame (“contenitore” di dati)
- Propedeutico ad analisi e visualizzazioni
- Primi esempi ed esercizi



[Apriamo il tutorial ufficiale...](#)