Intelligent System for Industry, Supply Chain and Environment

Parte I

Indice

Capitolo 1

Introduzione

1.1 What is an intelligent system (IS)

a computer-based system that aims to replicate human cognitive abilities such as learning, perception, reasoning, and decision-making.

By utilizing Machine Learning (ML), and other related technologies, these systems are capable of processing and analyzing data to perform tasks that typically require human intelligence, make predictions, or provide insights

Capitolo 2

Legislation, Artificial and human learning, Gestalt, applications and opinions about AI

2.1 Laws about AI in Europe

The AI Act is a European law on artificial intelligence (AI), the first comprehensive law on AI by a major regulator anywhere. The AIA was published in the Official Journal of the EU on 12 July 2024 and entered into force on 1 August 2024 .

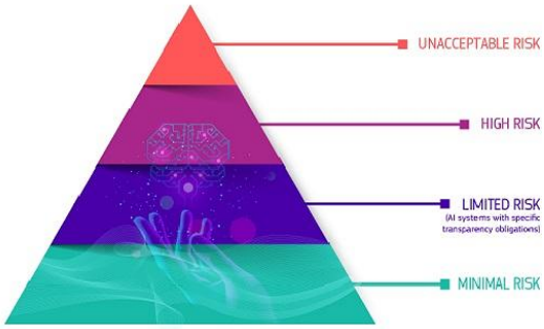
Why do we nee rules on AI?

- To avoid undesirable outcomes
- It is often not possible to find out why an AI system has made a decision or prediction and taken a particular action.
- It may become difficult to assess whether someone has been unfairly disadvantaged, such as in a hiring decision or in an application for a public benefit scheme

According to the European Union's Artificial Intelligence Act (AI Act), an AI system is defined as:

"a machine-based system designed to operate with varying levels of autonomy and that may exhibit, for explicit or implicit objectives , infers from the input it receives how to generate outputs such as predictions , content , recommendations, or decisionsthat can influence physical or virtual environments"

2.1.1 Four risk levels



- **Unacceptable risk:** All AI systems considered a clear threat to the safety, livelihoods and rights of people will be banned
- **High risk:**
  - critical infrastructures (e.g. transport), that could put the life and health of citizens at risk;
  - educational or vocational training, that may determine the access to education and professional course of someone's life (e.g. scoring of exams);
  - safety components of products (e.g. AI application in robotassisted surgery);
  - employment, management of workers and access to selfemployment (e.g. CV-sorting software for recruitment procedures);
  - essential private and public services (e.g. credit scoring denying citizens opportunity to obtain a loan);
  - law enforcement that may interfere with people's fundamental rights (e.g. evaluation of the reliability of evidence);
  - migration, asylum and border control management (e.g. verification of authenticity of travel documents);
- **Limited risk:** refers to AI systems with specific transparency obligations

Capitolo 3

lez 3

Capitolo 4

lez 4

Capitolo 5

lez 5

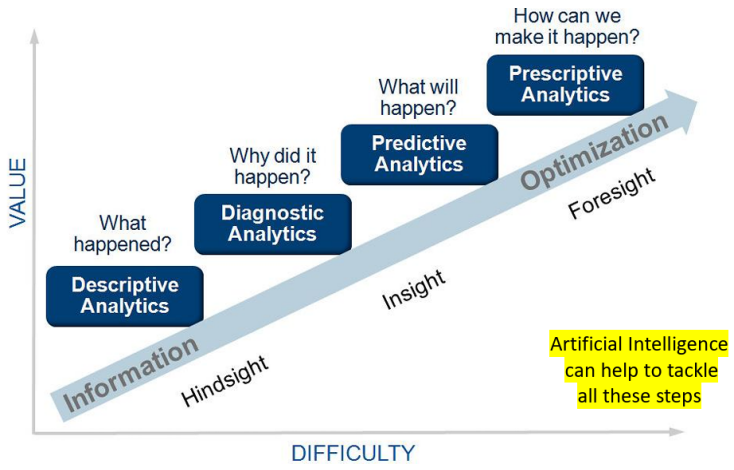
Capitolo 6

Intelligent transportations and Vehicle to everything protocol Examples of IoT, IoT security Artificial Intelligence of Things (AIoT), HW/SW environments for IoT/AIoT

Capitolo 7

Data Gathering, Data Preprocessing, Data Harmonization for intelligent system learning

7.1 Value of Data Analytic

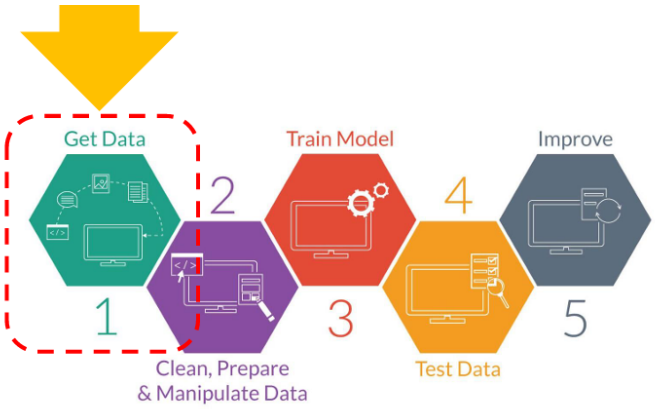


Four possibility increasing the value and the difficulty:

- Descriptive Analytics
- Diagnostic Analystics

- Predictive Analytics
- Prespective Analaytics

7.2 Data gathering - Step 1



We have so many powerful sources capable to generate data.
Data collection is the process of gathering and measuring information on targeted variables in an established system

7.2.1 IoT started to generate data...

The amount of data generated by connected internet of things (IoT) devices, forecast to grow to by 2025

- 41.6 billion connected devices
- 79.4 zettabytes (ZB) of data/year.

Data sources like phones, smart watches, ecc..

Example:
Basic GPS coordinates from smartphones @ITA

DataFromPositions/y = ItalianPopulation x CellPhoneRatio x 365 days x 24h/d x 60 min/d x 2coordinates/min = 65x10^6 x 0,83 x 365 x 24 x 60 x 2 x **8byte** = 697996800 byte ≈ 0,67GB

There are pubblic data centers, like Amazon's, Google's and Governative's ones

7.2.2 Data heterogeneity and synchronization

Heterogeneity in statistics means that your populations, samples or results are different. It is the opposite of homogeneity which means that the, population/data/results are the same. Qua fa un tot di esempi su come sia importante avere tutti i dati nello stesso formato (esempio di Marte e della NASA e della wind station)

7.2.3 Data Synchronization

The way a device adjusts its internal clock in order to align with the clocks of other devices in a network

Network Time Synchronization

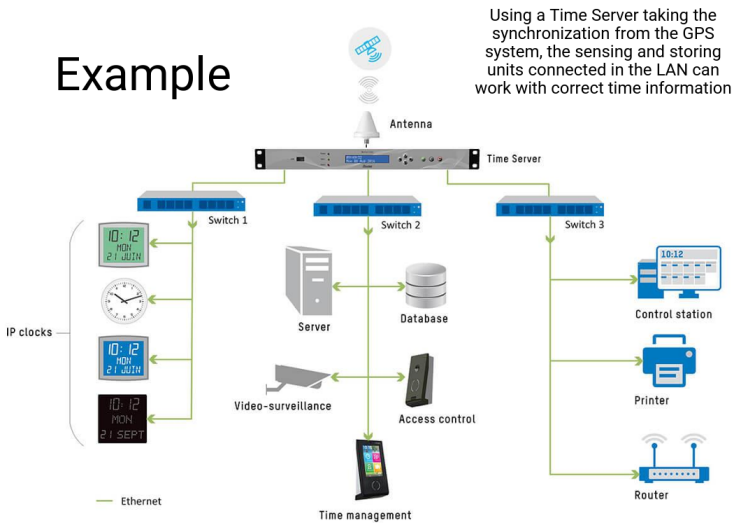
Computer clocks in servers, workstations and network devices are inherently not enough accurate Two problems:

- Clocks are set by hand to within a minute or two of actual time and are rarely checked after that
- Clocks are maintained by a battery-backed device that may drift as much as a second per day

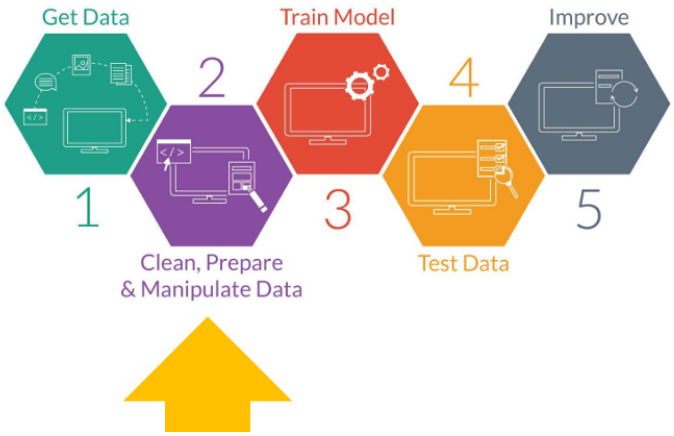
It's impossible to have accurate time synchronization without a proper method

Solutions

- **Network Time Protocol (NTP):** is a protocol for clock synchronization between computer systems over packetswitched, variable-latency data networks designed to mitigate local network latency
- **Time Server:** Dedicated network Time Server behind your firewall (devices synchronized to within 1/2 to 2 ms)



7.3 Data Preparation - Step 2

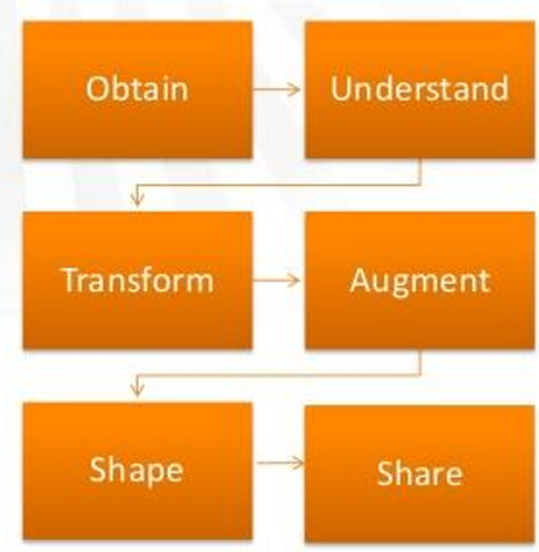


Data preparation includes two concepts such as Data Cleaning and Feature Engineering
The data wrangling problem is growing as different types of unstructured data or data in varying formats are pouring in from sensors, online and from traditional databases. All these data must be **cleaned up and organized** before data analytics/classifiers/regressors models can be applied.

7.3.1 Data wrangling

Data wrangling steps:

- Iterative process
- Understand
- Explore
- Transform
- Augment
- Visualize



Tasks of Data Wrangling:

- **Discovering:** Firstly, data should be understood thoroughly and examine which approach will best suit.
- **Structuring:** As the data is gathered from different sources, the data will be present in various shapes and sizes. Therefore, there is a need for structuring the data in proper format.
- **Cleaning:** Cleaning or removing of data should be performed that can degrade the performance of analysis.
- **Enrichment:** Extract new features or data from the given data set to optimize the performance of the applied model.

- ***Validating:*** This approach is used for improving the quality of data and consistency rules so that transformations that are applied to the data could be verified.

Data pre-processing: "is a technique that is used to convert the raw data into a clean data set" Pre-processing includes • Data cleaning • Data integration • Data transformation • Data reduction
Why is Data Preprocessing is so important? Three answers:

- Inaccurate data (missing data)
- The presence of noisy data/erroneous data/outliers
- Inconsistent data

7.3.2 Missing Data

What do we do when we have missing data?

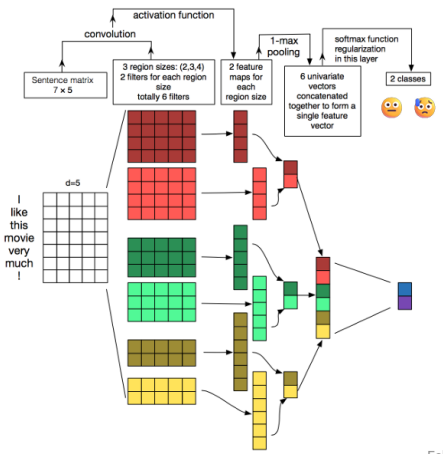
- **Ignoring the missing record:** is the simplest and efficient method for handling the missing data (not the best method when the number of missing values are immense or when the missing data problem and can solved (debugging/re-designredoing the experiment) and not just ignoring the problem causing the missing data.).
- **Filling the missing values manually:** one of the best-chosen methods, But there is one limitation that when there are large data set, and missing values are significant then, this approach is not efficient as it becomes a timeconsuming task.
- **Filling using computed values:** The missing values can also be occupied by computing mean, mode or median of the observed given values (ex: you can copy from the most similar column or generate values by using any ML or Deep Learning algorithm but it can generate bias within the data).

7.3.3 Structured and unstructured Data

Structured data usually resides in relational databases, This format is eminently searchable both with human generated queries and via algorithms using type of data and field names, such as alphabetical or numeric, currency or date. **Unstructured data** is essentially everything else. Unstructured data has internal structure but is not structured via pre-defined data models or schema (ex: sensor data, text files, emails, etc..).

Neural networks and unstructured data

It is not strictly compulsory to have structured data to use ML



structured data to use M

Some examples of text classification are:

- Understanding audience sentiment ( ) from social media
- Detection of spam & non-spam emails
- Auto tagging of customer queries
- Categorization of news articles into predefined topics

Capitolo 8

Managing a small dataset in Python Degrees of freedom/parameters Data Leakage

Sta lez probabilmente fa lab però non ci dovrebbe essere lab preciso all'esame
Lo fa sul link delle slide