Introduction to Machine Learning

12 March 2023 10:48

1. Python Basics

2. EDA
   - → matplotlib → simply plot w lists, multiple dfs etc
   - → Seaborn → specific dataset (data=df)

3. NULL Values
   - → Most entries in column are NULL → drop the column
   - → Numeric + some values are null (30–50%) → Mean
     - → Categorical → Replace w/ most occuring
       - Student / Subject → NM, M

Types of Variables / Column / Attributes

1) String / Text; → Eg: Names, Movie Description, Movie-Reviews, Actors (Cast)

2) Numeric: → Id, Age, No of matches played, Ratings, marks
   - Id: "AB101", Ratings: (1hr, 20mins, 40secs)

3) Categorical: Reviews / Rating
   - 4
   - 2
   - 3
   - 1
   - → categorical [ ] [male, female]

→ Very few rows are NULL (<5%) → drop the rows (subset)

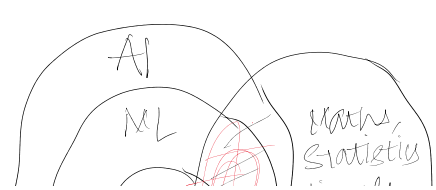
→ For all other categories
- → Actor → NULL ("Unknown cast")

Where does ML Come from?

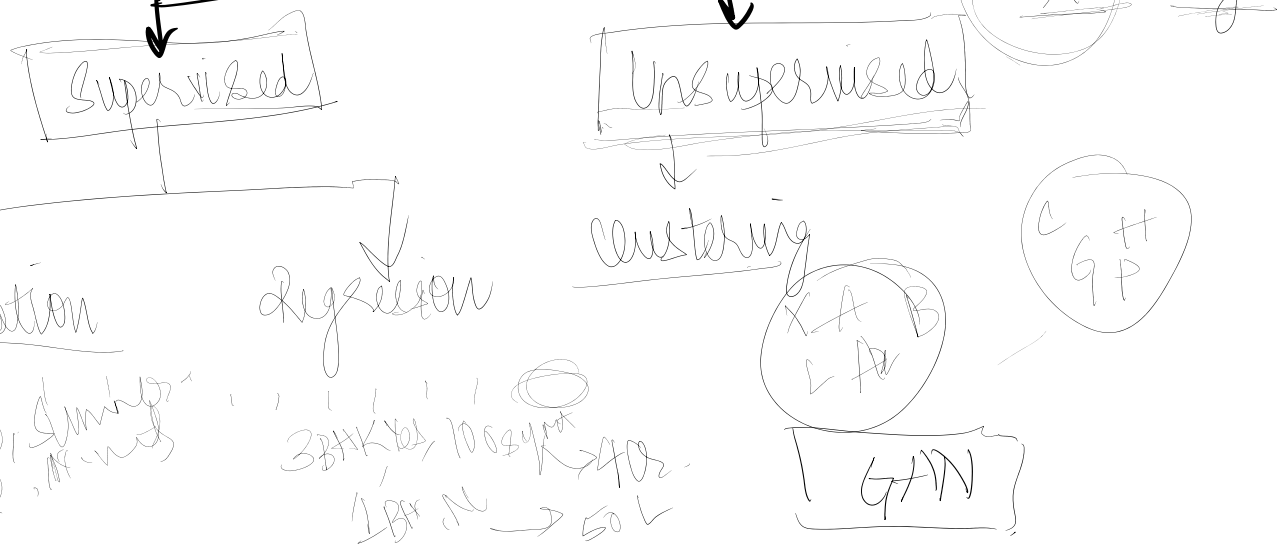
1. What is ML?
2. Where was it originated
3. Why is it used?
4. Current Scenario

Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed.

From <https://www.javatpoint.com/machine-learning>



Machine learning



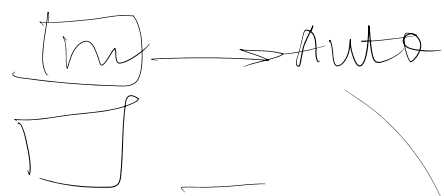
* Supervised V/S Unsupervised

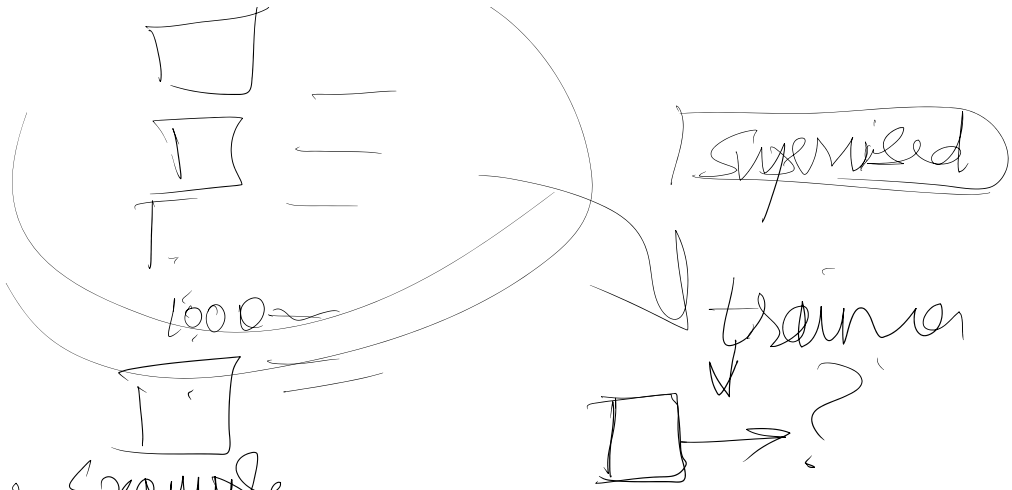
Dataset, 1000 images (Diff Vehicles)

→ Cars → 250
→ Scooters → 100
→ Planes → 300
→ Autos → 350

* CLUSTERING

Dataset, 1000 images





Pizza - Clustering Example

Social Media