Introduction to Data Science and ML

19 May 2023 06:56

Types of data

2011 → 80%

- Structured data
  - Tabular form of Data.

20%

- Unstructured data
  - Images, videos, audio, json, xml, signal.
  - 85%+
  - 2 years
  - No Data science

* Components of Data Science Project →

1. Understanding the problem statement.
2. Collect the data. → SQL
3. Processing & Cleaning the data.
   - → Numpy
   - → Pandas
   - → SQL

1980s → Hardware limitations.

4. Explore & Visualize the data. → Statistics & Matplotlib & Seaborn
5. Apply the Machine Learning techniques.
6. Test the performance of the model.
7. Tuning the ML model.
8. Present the result.

Data Analysis + Machine Learning ↓ Data Science

Data Science

Example: Heart Disease Prediction

Age	Chest Pain	Sugar level	Blood Pressure	Cholestrol level	Output
32	5	110	L	95	Yes
35	1	78	H	65	No
27	3	130	L	73	No
45	2	66	L	43	Yes
38	4	98	H	59	Yes

Future Patient

33 | 4 | 125 | H | 72 → Yes / No

100

80 → Predicted Correctly

20 → Predicted incorrectly

Performance of our trained ML model → $\frac{80}{100}$ → 80%.

HDP → HP

HP → HDP

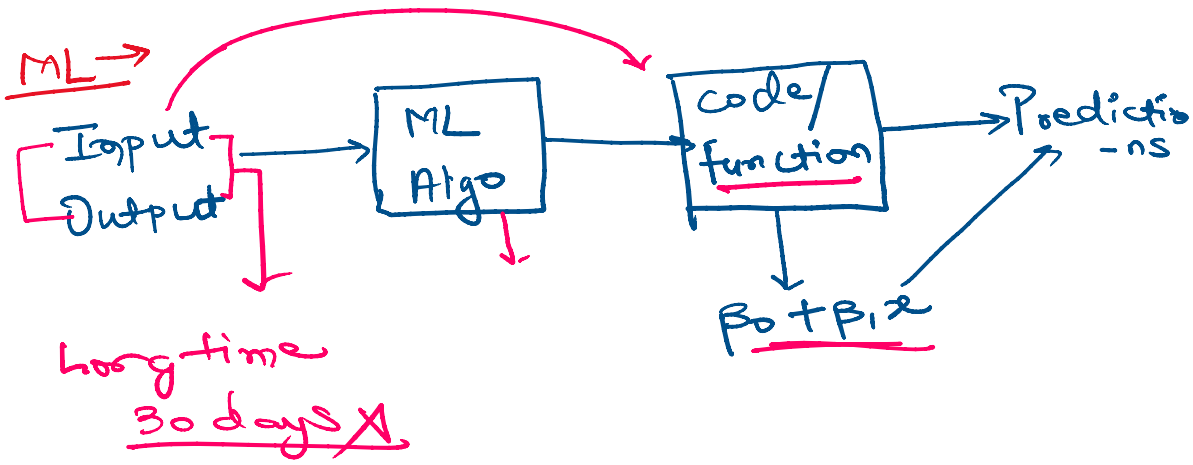
(7)

Child → Machine Learning Algorithm + Dataset → Machine Learning Model (Trained)

30+

SDEs

S/D → Input Data + code → Output



Types of Machine Learning →

1. Supervised Learning → Input & output both.
2. Unsupervised → only inputs available.
3. Reinforcement Learning.
4. Transfer Learning.

(1) Supervised Learning →

Input & Output → Yes / No

(1.1) Classification Task → categorical output

Yes / No Fraud / Not Fraud

→ Heart Disease Prediction
→ Loan Approval
→ Credit card Fraud Detection.

Classification Algorithms →

Classification Algorithms →

1. Logistic Regression
2. Decision Tree.
3. Random Forest.
4. K-Nearest Neighbour.
5. Support Vector Machines (SVM)
6. Boosting Algorithms (XG Boost, Ada Boost)

② Regression Task → Continous Output

→ House Price Prediction.
→ Sales Prediction.
→ Stock Prediction.

24.7
13.2
4.9

Algorithms for Regression Task →

1. Linear Regression
2. Polynomial Regression
3. DT Regressor.
4. RF Regressor.
5. SVM Regressor.

Input → Independant variables, Features.

Output → Dependant variable, Target.