

AI/ML

Bibek Thapa

INTRODUCTION TO AI, ML, DL

AI vs ML vs DL

- **Artificial Intelligence (AI):** Developing machines to mimic human intelligence and behaviour.
- **Machine Learning (ML):** Algorithms that learn from structured data to predict outputs and discover patterns in that data.
- **Deep Learning (DL):** Algorithms based on highly complex neural networks that mimic the way a human brain works to detect patterns in large unstructured data sets.

A Venn diagram consisting of three concentric circles. The outermost circle is light blue and labeled 'AI'. Inside it is a medium blue circle labeled 'ML'. Inside the 'ML' circle is a dark blue circle labeled 'DL'. Each circle contains a descriptive text about its respective field.

AI

Ability to learn and reason like humans

ML

**Algorithms with ability to learn
without being explicitly
programmed**

DL

**ML in which artificial neural
networks adapt and learn
from vast amount of data.**

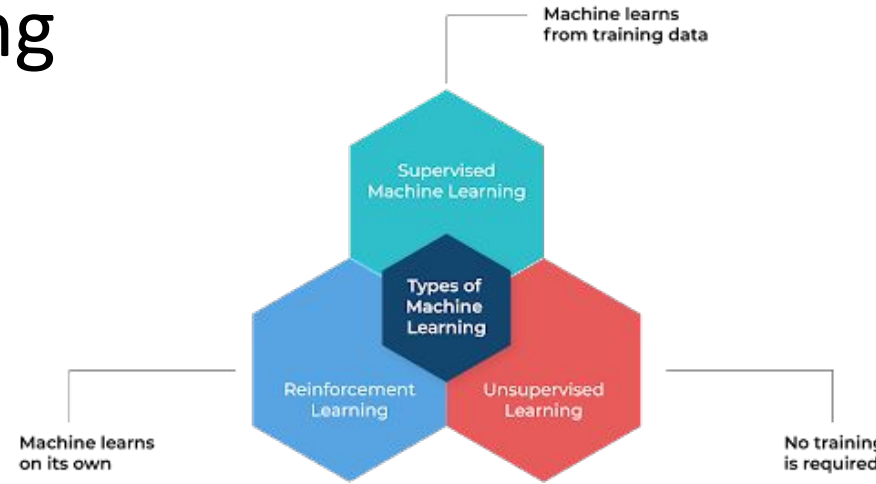
Machine learning

- How does machine learning differ from other traditional programming ideas?
- What makes machine learning so special?

We will discuss it now.



Types of Machine Learning



- A very short intro for the types could be:
 - **Supervised Learning:** "Guide me through training!"
 - **Unsupervised Learning:** "I learn independently!"
 - **Reinforcement Learning:** "I make my own choices!"

Machine learning Lifecycle

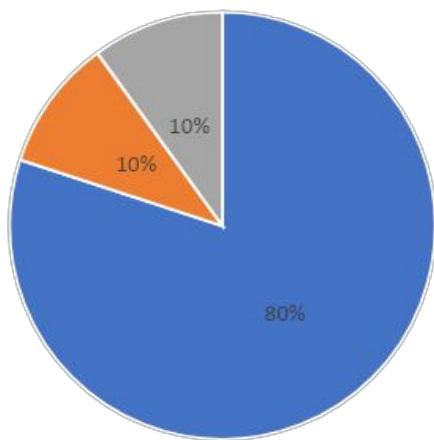
- Problem Definition
- Data Collection
- Data Preprocessing
- Feature Engineering
- Model Selection and Training
- Model Evaluation and Fine Tuning
- Model Deployment
- Monitoring and Maintenance

Python and Data Structures

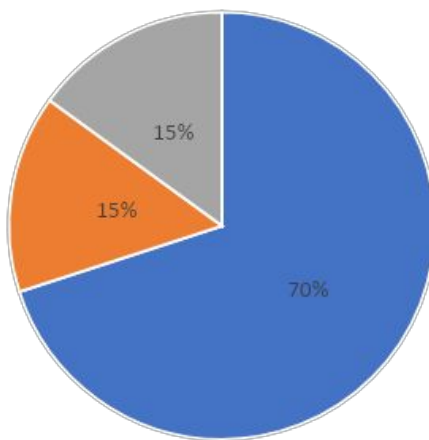
- Demo

Data For Training

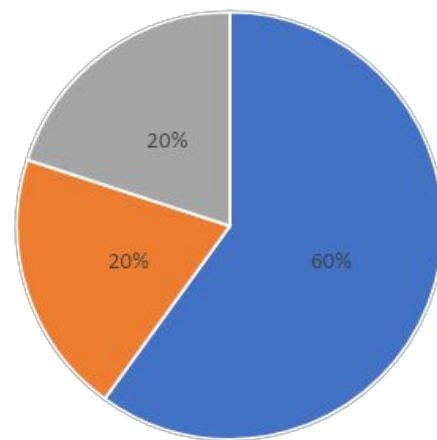
- Training Data
- Validation Data
- Test Data



■ Train ■ Validation ■ Test

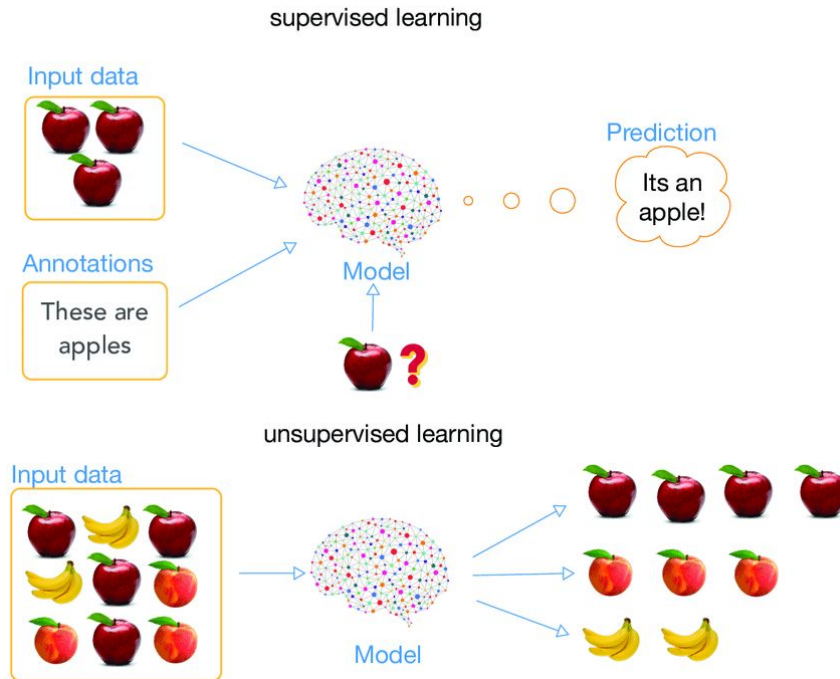


■ Train ■ Validation ■ Test



■ Train ■ Validation ■ Test

Supervised Learning and Unsupervised Learning



Training Data

- **Training data** is the data we use to train a machine learning algorithm.
- Training data in **supervised learning** consists of generally input data and their correct labels.
- Training data in **unsupervised learning** consists of no labels and machine learning model is task to find patterns in the data.

Validation Data

- Validation Data is a test data that our model isn't train on but our model can predict on.
- It's unseen data **during** training. It helps to further tune our model.
- The validation data must be labeled for supervised learning model.

Test Data

- Validation Data is a test data that our model isn't train on but our model can predict on.
- It's unseen data **after** training. It helps to evaluate our model.
- The test data can be labeled or not labeled for supervised learning model.

Regression Model

- Works on continuous data
- Given a value of X determines the Value of Y where Y is dependent on X
- Example: House price prediction, sales prediction

YearsExperience	Field	Salary
1.1	Technical	39343
1.3	Non-Technical	26205
1.5	Technical	37731
2	Non-Technical	28525
2.2	Technical	39891
2.9	Non-Technical	36642
3	Technical	60150
3.2	Non-Technical	34445
3.2	Technical	64445
3.7	Technical	57189
3.9	Non-Technical	43218
4	Technical	55794
4	Technical	56957

Classification Model (Numeric)

- Task of identifying the label or class of the data
- Works on the discrete data
- Example 0 or 1 . Here 1 means Death and 0 means Alive

age	anaemia	creatinine	diabetes	ejection_f	high_blood	platelets	serum_cre	serum_soc	sex	smoking	time	DEATH_EVENT
75	0	582	0	20	1	265000	1.9	130	1	0	4	1
55	0	7861	0	38	0	263358	1.1	136	1	0	6	1
65	0	146	0	20	0	162000	1.3	129	1	1	7	1
50	1	111	0	20	0	210000	1.9	137	1	0	7	1
65	1	160	1	20	0	327000	2.7	116	0	0	8	0
90	1	47	0	40	1	204000	2.1	132	1	1	8	1
75	1	246	0	15	0	127000	1.2	137	1	0	10	0
60	1	315	1	60	0	454000	1.1	131	1	1	10	1
65	0	157	0	65	0	263358	1.5	138	0	0	10	1
80	1	123	0	35	1	388000	9.4	133	1	1	10	1
75	1	81	0	38	1	368000	4	131	1	1	10	0
62	0	231	0	25	1	253000	0.9	140	1	1	10	1
45	1	981	0	30	0	136000	1.1	137	1	0	11	1
50	1	168	0	38	1	276000	1.1	137	1	0	11	1
49	1	80	0	30	1	427000	1	138	0	0	12	1
82	1	379	0	50	0	47000	1.3	136	1	0	13	1

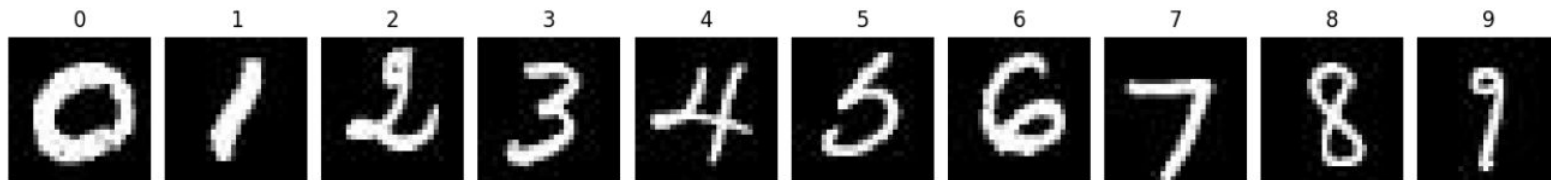
Classification Model (Text)

- The input data are text and labels are also text.
- But while training on the data, the data must be encoded (converted to numeric data.)

Text	Category
worldcom ex-boss launches defence lawyers defending former worldcom chief bernie ebbers against a battery of fraud charges have called a company whistleblower as their first witness. cynthia co	business
german business confidence slides german business confidence fell in february knocking hopes of a speedy recovery in europe s largest economy. munich-based research institute ifo said that its confi	business
bbc poll indicates economic gloom citizens in a majority of nations surveyed in a bbc world service poll believe the world economy is worsening. most respondents also said their national economy wa	business
lifestyle governs mobile choice faster better or funkier hardware alone is not going to help phone firms sell more handsets research suggests. instead phone firms keen to get more out of their cust	tech
enron bosses in \$168m payout eighteen former enron directors have agreed a \$168m (Â£89m) settlement deal in a shareholder lawsuit over the collapse of the energy firm. leading plaintiff the univer	business
howard truanted to play snooker conservative leader michael howard has admitted he used to play truant to spend time with his school friends at a snooker hall. mr howard said his time at jack s	politics
wales silent on grand slam talk rhys williams says wales are still not thinking of winning the grand slam despite a third six nations win. that s the last thing on our minds at the moment said williams a	sport
french honour for director parker british film director sir alan parker has been made an officer in the order of arts and letters one of france s highest cultural honours. sir alan received his decoration	entertainment
car giant hit by mercedes slump a slump in profitability at luxury car maker mercedes has prompted a big drop in profits at parent daimlerchrysler. the german-us carmaker saw fourth quarter operatin	business
fockers fuel festive film chart comedy meet the fockers topped the festive box office in north america setting a new record for christmas day. the sequel took \$44.7m (Â£23.2m) between 24 and 26 d	entertainment
blair rejects iraq advice calls tony blair has rejected calls for the publication of advice on the legality of the iraq war amid growing calls for an investigation. the prime minister told his monthly press co	politics
housewives lift channel 4 ratings the debut of us television hit desperate housewives has helped lift channel 4 s january audience share by 12% compared to last year. other successes such as celebrity	entertainment
uk coal plunges into deeper loss shares in uk coal have fallen after the mining group reported losses had deepened to Â£51.6m in 2004 from Â£1.2m. the uk s biggest coal producer blamed geological	business
bp surges ahead on high oil price oil giant bp has announced a 26% rise in annual profits to \$16.2bn (Â£8.7bn) on the back of record oil prices. last week rival shell reported an annual profit of \$17.5br	business
ireland 21-19 argentina an injury-time dropped goal by ronan o gara stole victory for ireland from underneath the noses of argentina at lansdowne road on saturday. o gara kicked all of ireland s point	sport
wenger signs new deal arsenal manager arse wenger has signed a new contract to stay at the club until may 2008. wenger has ended speculation about his future by agreeing a long-term contract th	sport
u2 s desire to be number one u2 who have won three prestigious grammy awards for their hit vertigo are stubbornly clinging to their status as one of the biggest bands in the world. the most popular	entertainment
hantuchova in dubai last eight daniela hantuchova moved into the quarter-finals of the dubai open after beating elene likhotseva of russia 7-5 6-4 and now faces serena williams. australian open cha	sport
melzer shocks agassi in san jose second seed andre agassi suffered a comprehensive defeat by jurgen melzer in the quarter-finals of the sap open. agassi was often bamboozled by the austrian s drop	sport
moving mobile improves golf swing a mobile phone that recognises and responds to movements has been launched in japan. the motion-sensitive phone - officially titled the v603sh - was developed b	tech

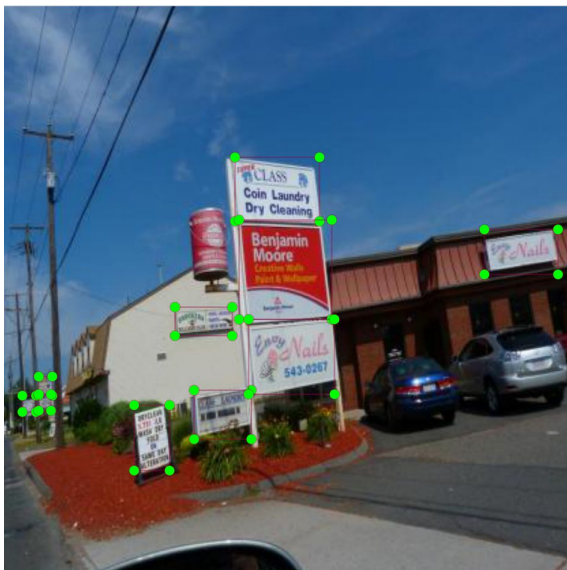
Classification Model (Image)

- The input data are image and labels are text.
- But while training on the data, the data must be encoded (converted to numeric data.)



Object Detection

- It will be a combination of classification and regression.
- The bounding box is a regression problem and the area inside the bounding box is a classification problem.



000f269d7fece350.jpg.rf.b8fe9106ca7426a036d41c004df8dbe1 - Notepad

File Edit Format View Help

```
0 0.057692307692307696 0.4519230769230769 0.11538461538461539 0.06009615384615385
0 0.1466346153846154 0.5913461538461539 0.052884615384615384 0.028846153846153848
0 0.33413461538461536 0.5985576923076923 0.09014423076923077 0.03245192307692308
0 0.5300480769230769 0.6021634615384616 0.12740384615384615 0.06009615384615385
0 0.8100961538461539 0.5084134615384616 0.3269230769230769 0.27163461538461536
```

Data Quality

I tried to explain the concept of 'clean data' to my LLM. It responded with, 'I prefer my data like my room: messy and full of surprises.'

-ChatGPT

Data Quality

- Satisfy the requirements of its intended use

Why Data Quality matters in Data Science, AI and ML

- Foundation for Model Training
- Ensures Model Performance
- Reduces Bias and Variance
- Enables Trust and Interpretability
- Supports Data-driven Insights
- Ensures Ethical and Responsible AI

Key Aspects of Data Quality

- Accuracy: Ensuring the data is correct and free from errors.
- Completeness: All required data is present.
- Consistency: Data is consistent across different datasets and sources.
- Timeliness: Data is up-to-date and available when needed.
- Believability: Data is credible and trustworthy.
- Interpretability: Data is easily understood and usable.

