

## Assignment-1 Semi-Supervised Learning

1) write a short note on semi-supervised

Ans) Semi supervised learning is a machine learning approach that lies between supervised & unsupervised learning. it leverages both labeled and unlabeled data during training, making it particularly useful when acquiring labeled data is expensive or time-consuming, but unlabeled data is abundant.

In semi supervised learning, the model learns from a small amount of labeled data to improve its performance. The key assumption is that the structure of the data distribution (from unlabeled data) aligns well with the class labels (from labeled data).

Types of Semi supervised Learning Algorithms:

- 1) Self Training: The model is initially trained on labeled data, then predicts labels for unlabeled data, which are added to the training set iteratively.
- 2) Co-training: Two models are trained on different views of the data. Each model labels the unlabeled data for the other, improving performance iteratively.
- 3) Graph based methods: These algorithms build a graph where nodes represent data points, and edges represent similarities. The labels are propagated



Through graph based on these connections.

i) Generative models: These models learn the joint probability distribution of the data and the labels, allowing them to infer labels for the unlabeled data.

Advantages: \* reduces the cost of obtaining labeled data.  
\* can improve model performance by utilizing unlabeled data.

Challenges: Requires careful balancing of labeled and unlabeled data to avoid propagating incorrect labels. \* Assumptions about data distribution may not always hold.

Application: text classification, Image recognition and speech processing.

3) Give examples of semi-supervised learning algorithms.

Ans) i) Self-training: Text classification where a small set of labeled documents is supplemented with pseudo labels generated from a large pool of unlabeled documents.

ii) Co-training: Image classification where one model uses color features and another uses texture features, each model helps labels the unlabeled images for the other.

iii) Graph Based method: social network analysis

where nodes represent users, and labels (eg user interests) are propagated through the network.

iv) Label propagation: Handwriting recognition where character samples with known labels propagate information to other samples.

v) Generative models: Generative adversarial networks (GANs) where the discriminator can be used to label unlabeled data based on its learned distribution.

vi) Consistency Regularization: Image classification where augmentations of images (eg rotations, translations) are used to ensure that predictions remain consistent.

vii) Pseudo learning: Speech recognition where a model trained on a small set of transcribed audio can generate pseudo labels for a larger set of untranscribed audio data.

3) write a few applications of semi-supervised learning algorithms.

Ans) Semi-supervised learning algorithms are applied in various domains where labeled data is limited but unlabeled data is abundant.

\* Text classification: sentiment analysis, topic categorization, and spam detection.

\* Image classification: object recognition, facial recognition, and medical imaging.

\* Speech Recognition: Transcription of spoken language & voice commands.

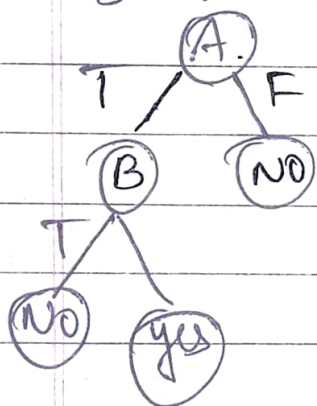


- \* web content classification: Categorizing web pages or filtering web content.
- \* Social Network Analysis: friend recommendation, community detection, and influence propagation.
- \* Anomaly detection: fraud detection, network intrusion and fault detection.
- \* Video Analysis: Action recognition, Object tracking and event detection.
- \* Natural Language Processing (NLP): Named entity recognition, machine translation, and question answering.

### Assignment - 2 Decision tree.

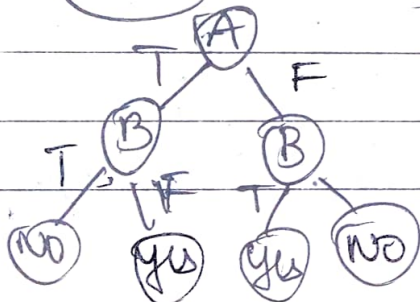
1) Develop the decision tree for the following.

a)  $A \wedge \neg B$



A	B	$\neg B$	$(A \wedge \neg B)$
T	T	F	F
T	F	T	T
F	T	F	F
F	F	T	F

2)  $A \oplus B$  ( $A \oplus B$ )



A	B	$A \oplus B$
T	T	F
T	F	T
F	T	T
F	F	F