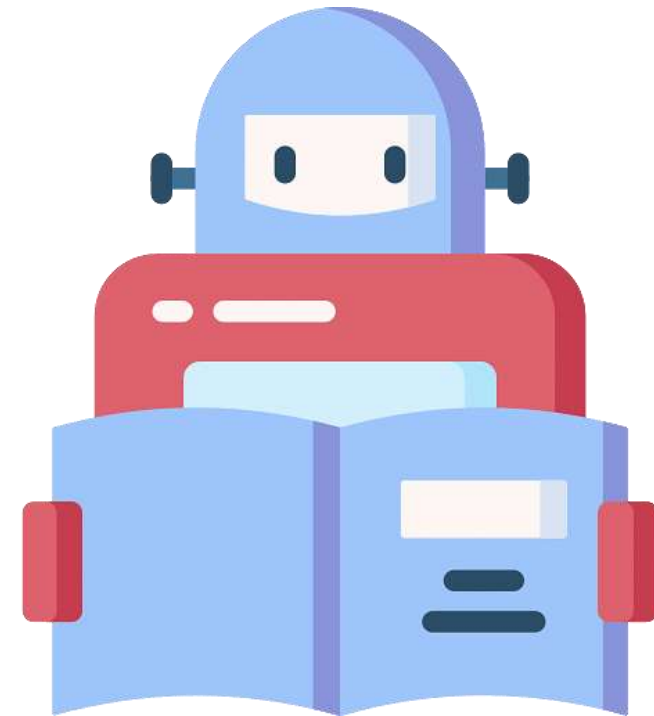
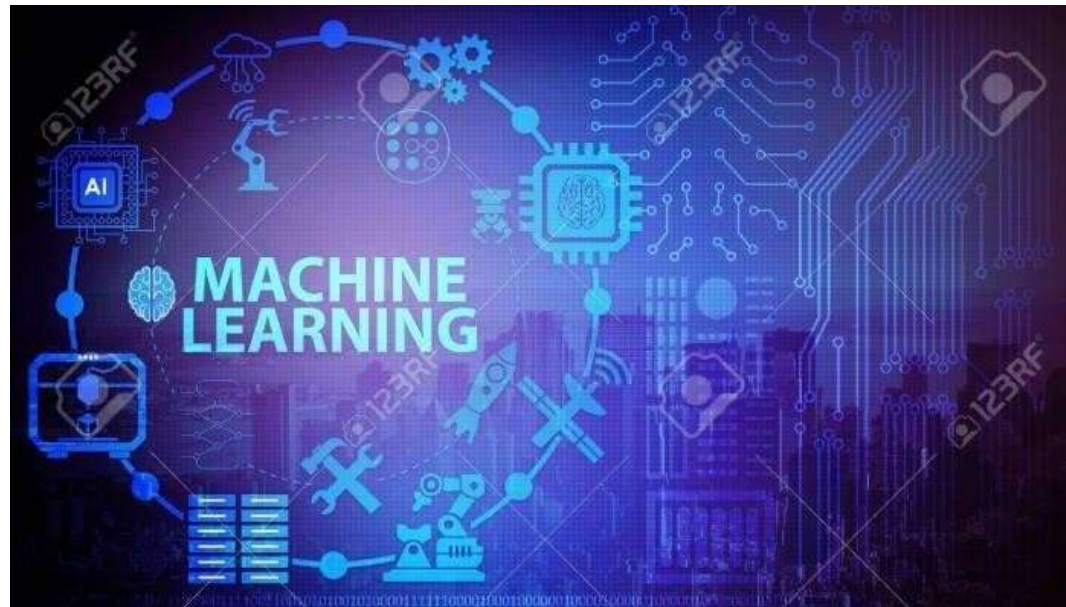




MACHINE LEARNING PROCESS



L.Poornima Devi AP/CSE

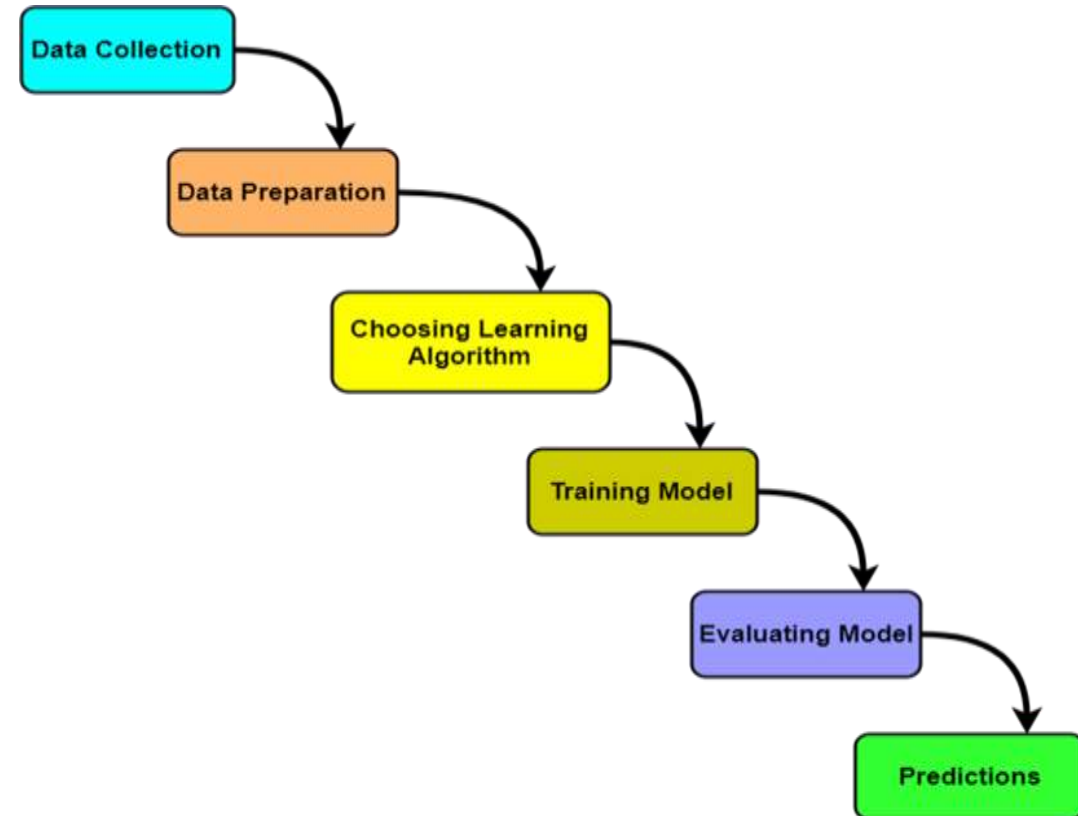
19CST301 & INTRODUCTION TO MACHINE LEARNING

19CST301 & INTRODUCTION TO MACHINE LEARNING
L.Poornima Devi Ap/CSE



STEPS IN MACHINE LEARNING PROCESS

- ❑ Data Collection
- ❑ Data Preparation
- ❑ Choosing Learning Algorithm
- ❑ Training Model
- ❑ Evaluating Model
- ❑ Predictions



Machine Learning Workflow



DATA COLLECTION

In this stage,

- Data is collected from different sources.
- The type of data collected depends upon the type of desired project.
- Data may be collected from various sources such as files, databases etc.
- The quality and quantity of gathered data directly affects the accuracy of the desired system.





DATA PREPARATION

In this stage,

- Data preparation is done to clean the raw data.
- Data collected from the real world is transformed to a clean dataset.
- Raw data may contain missing values, inconsistent values, duplicate instances etc.
- So, raw data cannot be directly used for building a model.

This is the most time consuming stage in machine learning workflow.



Different methods of cleaning the dataset are

- Ignoring the missing values
- Removing instances having missing values from



the dataset.

- Estimating the missing values of instances using mean, median or mode.
- Removing duplicate instances from the dataset.
- Normalizing the data in the dataset.



CHOOSING LEARNING ALGORITHM

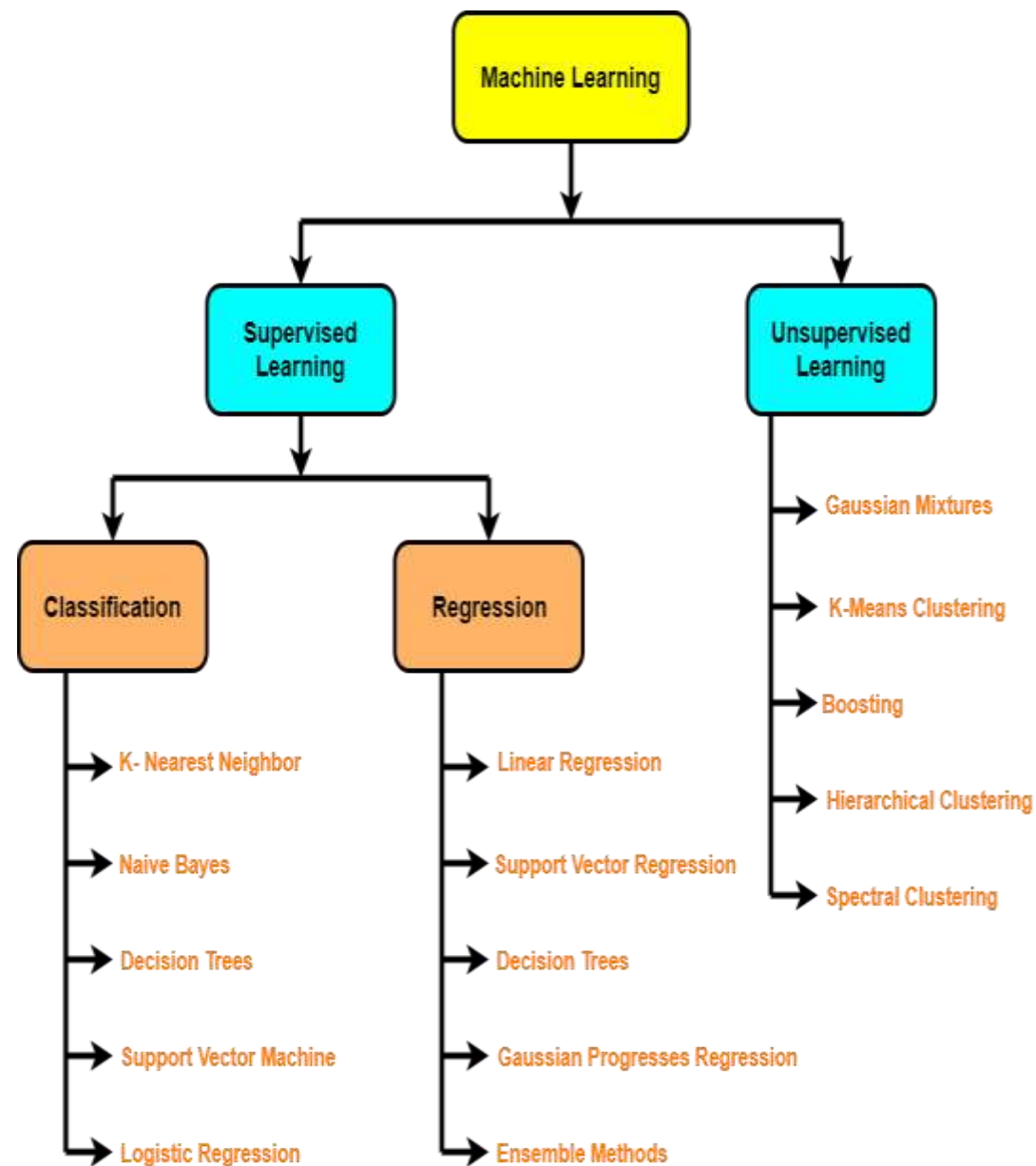
In this stage,

- The best performing learning algorithm is researched.
- It depends upon the type of problem that needs to be solved and the type of data we have.
- If the problem is to classify and the data is labeled, classification algorithms are used.
- If the problem is to perform a regression task and the data is labeled, regression algorithms are used.
- If the problem is to create clusters and the data is unlabeled, clustering algorithms are used.





- **The following chart provides the overview of learning algorithms-**

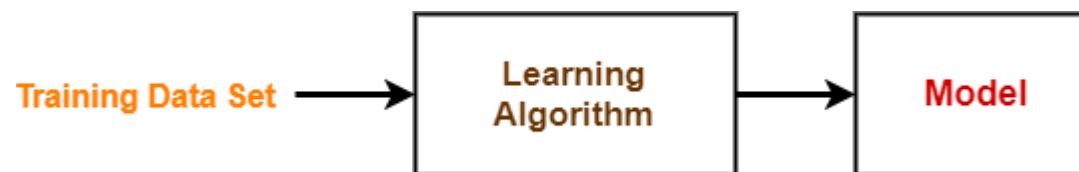




TRAINING MODEL

In this stage,

- The model is trained to improve its ability.
- The dataset is divided into training dataset and testing dataset.
- The training and testing split is order of 80/20 or 70/30.
- It also depends upon the size of the dataset.
- Training dataset is used for training purpose.
- Testing dataset is used for the testing purpose.
- Training dataset is fed to the learning algorithm.
- The learning algorithm finds a



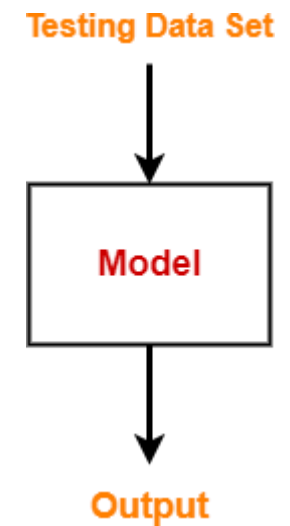


mapping between the input and the output and generates the model.

EVALUATING MODEL

In this stage,

- The model is evaluated to test if the model is any good.
- The model is evaluated using the kept-aside testing dataset.
- It allows to test the model against data that has never been used before for training.
- Metrics such as accuracy, precision, recall etc are used to test the performance.
- If the model does not perform well, the model is re-built using different hyper parameters.
- The accuracy may be further improved by tuning the hyper parameters.

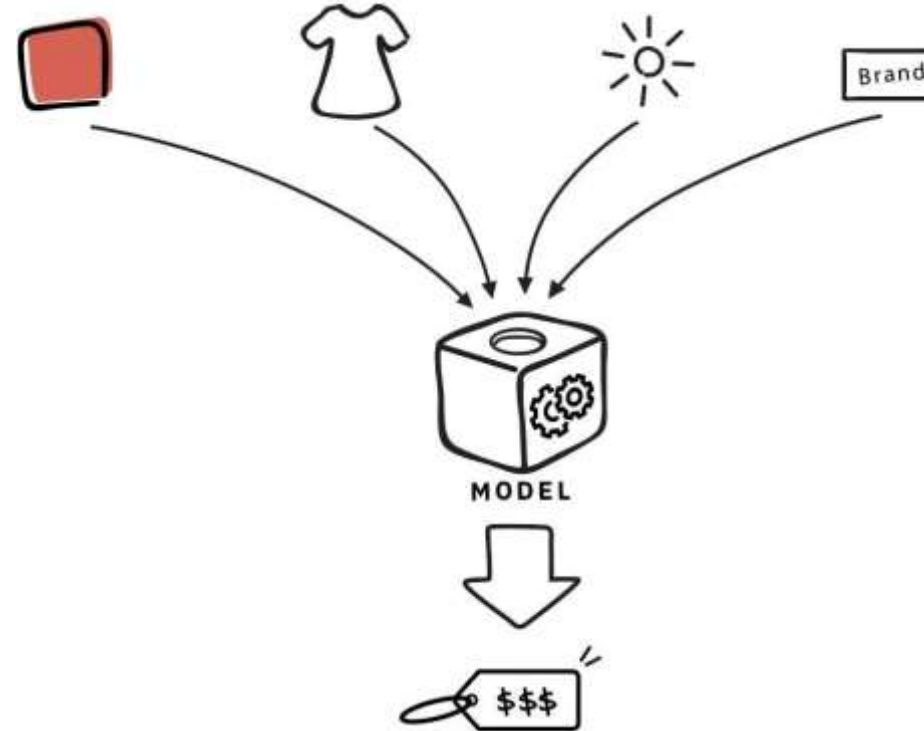




PREDICTIONS

In this stage,

- The built system is finally used to do something useful in the real world.
- Here, the true value of machine learning is realized.





SNS COLLEGE OF TECHNOLOGY, COIMBATORE –35
(An Autonomous Institution)
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



THANK
YOU
ANY QUESTIONS?