SNS COLLEGE OF ENGINEERING
Kurumbapalayam(Po), Coimbatore – 641 107
(An autonomous institution)
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Mrs. N.Padmashri/Assistant Professor

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATASCIENCE

Basic Concepts in Machine Learning

- Machine Learning is defined as a technology that is used to train machines to perform various actions such as predictions, recommendations, estimations, etc., based on historical data or past experience.
- Machine Learning enables computers to behave like human beings by training them with the help of past experience and predicted data.
- There are three key aspects of Machine Learning, which are as follows:

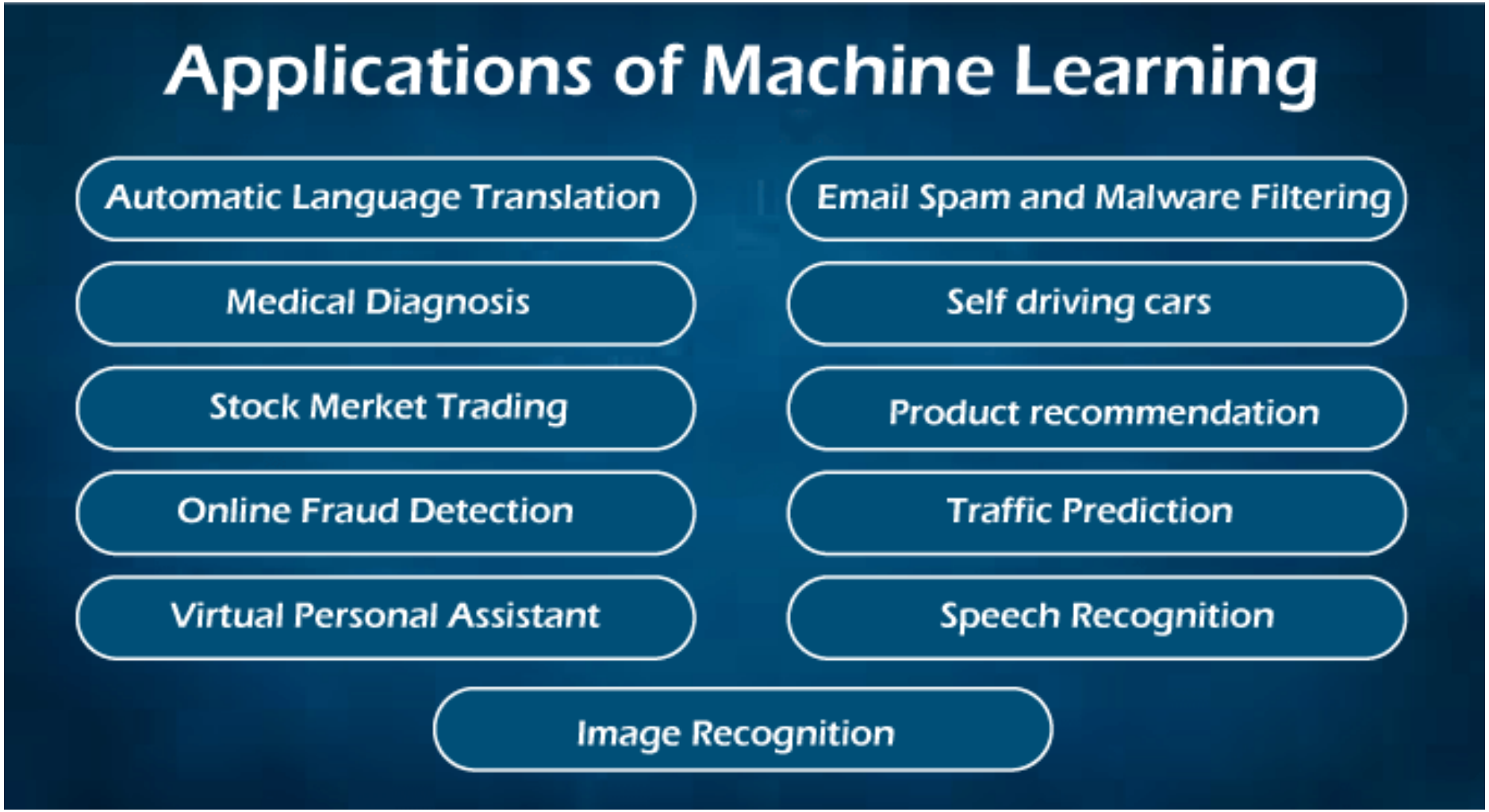
There are three key aspects of Machine Learning

- Task: A task is defined as the main problem in which we are interested. This task/problem can be related to the predictions and recommendations and estimations, etc.
- Experience: It is defined as learning from historical or past data and used to estimate and resolve future tasks.
- Performance: It is defined as the capacity of any machine to resolve any machine learning task or problem and provide the best outcome for the same. However, performance is dependent on the type of machine learning problems.

Techniques in Machine Learning

- ✓ Supervised Learning
- ✓ Unsupervised Learning
- ✓ Reinforcement Learning
- ✓ Semi-supervised Learning(is an intermediate technique of both supervised and unsupervised learning)

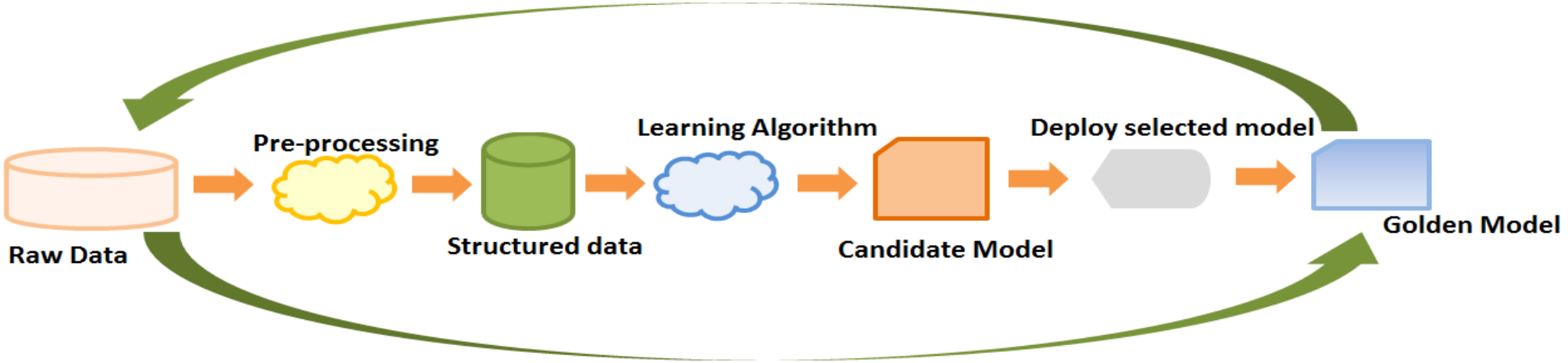
Applications of Machine Learning



Machine Learning Process

- Machine learning is the process of making systems that learn and improve by themselves, by being specifically programmed.

- The ultimate goal of machine learning is to design algorithms that automatically help a system gather data and use that data to learn more.



Steps In Developing A ML Application.

1. Gathering data
2. Preparing that data
3. Choosing a model
4. Training
5. Evaluation
6. Hyper parameter tuning
7. Prediction

Case Study :

1. Price Comparison Tool: The tool will allow you to compare similar products offered online by different brands and retailers by web scraping. You can compare the prices of your desired products and pay the lowest price to buy them. The tool will also send a notification about the company with the lowest price, along with a link that can be used to buy the product.
2. Crime Analysis: Our Crime Analysis project creates a website showcasing crime data insights. The website will serve as a valuable resource for law enforcement agencies, policymakers, and the general public, aiding in evidence-based decision-making and the development of targeted crime prevention strategies.
3. Disease Detection
4. Disease Prevention

Steps In Developing A ML Application.

Objective/Define Problem/ Hypothesis

1. Gathering data
2. Preparing that data
3. Choosing a model
4. Training
5. Evaluation
6. Hyper parameter tuning
7. Prediction

Define the Problem:

1. Clearly define the problem you want to solve using machine learning.
2. Identify the goals, requirements, and desired outcomes of the application.
3. This step involves understanding the problem domain, gathering relevant data, and defining the evaluation metrics for measuring success.

Price Comparison Tool	Crime Analysis
Clearly define the problem and the objectives of the Price Comparison Tool.	Clearly define the problem and the objectives of the Crime Analysis project.
Develop a tool that allows users to compare prices of similar products offered by different brands and retailers, enabling them to find the lowest price and make informed purchasing decisions.	Develop a website that showcases crime data insights to support law enforcement agencies, policymakers, and the general public. The project aims to provide evidence-based decision-making and aid in the development of targeted crime prevention strategies.

Steps In Developing a ML Application.

1. Gathering data

This step is very important because the quality and quantity of data that you gather will directly determine how good your predictive model can be.

- Investigate and obtain data that you will use to feed your machine.
- The quality and quantity of data
- Existing dataset
- Develop your own datasets (survey forms/experiments)
- Web scraping technique to automatically collect information from various sources such as APIs.

Price Comparison Tool	Crime Analysis
Collecting relevant data is crucial for the tool's functionality. In this case, web scraping can be used to gather product information, including prices, from various e-commerce websites. Identify the target websites and determine the data attributes to be extracted, such as product name, brand, retailer, and price.	Collecting relevant crime data is crucial for the project's functionality. Identify reliable sources for crime data, such as law enforcement agencies, government databases, or publicly available crime data repositories. Gather information on various crime-related variables, such as location, date/time, type of crime, and other relevant attributes.

Steps In Developing A ML Application.

2. Preparing that data

i. Not all the collected data is useful for a machine learning application.

ii. We may need to clean the irrelevant data, which may affect the accuracy of prediction or may take additional computation without aiding in the result.

iii. Once the data ready for the machine learning algorithm, we need to transform the data in the form that the ML system can understand.

iv. Machines cannot understand an image or text. We need to convert it into numbers.

v. It also requires building data pipeline depending on the machine learning application needs.

Price Comparison Tool	Crime Analysis
Preprocess the collected data to ensure its quality and compatibility with the machine learning model. This involves cleaning the data, handling missing values, standardizing formats, and resolving any inconsistencies. Additionally, perform any necessary data transformations or feature engineering to extract relevant features for the model.	Preprocess the collected crime data to ensure its quality and compatibility with the machine learning model. This involves cleaning the data, handling missing values, standardizing formats, and resolving any inconsistencies. Additionally, perform any necessary data transformations or feature engineering to extract meaningful features for the model.

Steps In Developing A ML Application.

3. Choosing Model

- Choose an appropriate machine learning algorithm to build the price comparison model. This could involve techniques such as natural language processing (NLP) for text analysis, clustering or similarity algorithms for product matching, or regression models for price prediction. The choice of model depends on the specific requirements and characteristics of the data.

Model	Applications
Logistic Regression	Price prediction
Fully connected networks	Classification
Convolutional Neural Networks	Image processing
Recurrent Neural Networks	Voice recognition
Random Forest	Fraud Detection
Reinforcement Learning	Learning by trial and error
Generative Models	Image creation
K-means	Segmentation
k-Nearest Neighbors	Recommendation systems
Bayesian Classifiers	Spam and noise filtering

Price Comparison Tool	Crime Analysis
Choose an appropriate machine learning algorithm to build the price comparison model. This could involve techniques such as natural language processing (NLP) for text analysis, clustering or similarity algorithms for product matching, or regression models for price prediction. The choice of model depends on the specific requirements and characteristics of the data.	Choose appropriate machine learning algorithms or techniques to build models for crime analysis. This can involve techniques such as clustering to identify crime hotspots, classification for predicting crime types or severity, or anomaly detection for identifying unusual patterns. The choice of models depends on the specific requirements and objectives of the project.

Steps In Developing A ML Application.

4. Training

- The goal of training is to answer a question or make a prediction correctly as often as possible.
- Linear regression example: algorithm would need to learn values for *m* (or *W*) and *b* (*x* is input, *y* is output)
- Each iteration of process is a training step

- Split the preprocessed data into training and validation sets. Use the training set to train the machine learning model on historical data, including product information and prices. Validate the model's performance using the validation set to assess its accuracy and generalization capabilities

Price Comparison Tool	Crime Analysis
Split the preprocessed data into training and validation sets. Use the training set to train the machine learning model on historical data, including product information and prices. Validate the model's performance using the validation set to assess its accuracy and generalization capabilities. Iteratively refine the model by adjusting hyperparameters and evaluating its performance.	Split the preprocessed data into training and validation sets. Use the training set to train the machine learning models on historical crime data. Validate the models' performance using the validation set to assess their accuracy, precision, recall, or other relevant metrics. Iteratively refine the models by adjusting hyperparameters and evaluating their performance.

Steps In Developing A ML Application.

5. Evaluation

- Uses some metric or combination of metrics to "measure" objective performance of model.
- Test the model against previously unseen data.
- This unseen data is meant to be somewhat representative of model performance in the real world, but still helps tune the model (as opposed to test data, which does not).
- Good train/eval split? 80/20, 70/30, or similar, depending on domain, data availability, dataset particulars, etc.

Steps In Developing A ML Application.

6. Hyper parameter Tuning

- Once the evaluation is over, any further improvement in your training can be possible by tuning the parameters.
- There were a few parameters that were implicitly assumed when the training was done.
- Another parameter included is the learning rate that defines how far the line is shifted during each step, based on the information from the previous training step.
- These values all play a role in the accuracy of the training model, and how long the training will take.
- Tune model parameters for improved performance
- Simple model hyper parameters may include: number of training steps, learning rate, initialization values and distribution, etc.

Steps In Developing A ML Application.

7. Prediction

Using further (test set) data which have, until this point, been withheld from the model (and for which class labels are known), are used to test the model; a better approximation of how the model will perform in the real world.