M. Tech. Programme: PEOs

The Program Educational Objectives of M. Tech. programs are (PEOs):

1. To prepare graduates who will be successful professionals in industry, government, academia, research, entrepreneurial pursuit and consulting firms.

2. To prepare graduates who will contribute to society as broadly educated, expressive, ethical and responsible citizens with proven expertise.

3. To prepare graduates who will achieve peer-recognition, as an individual or in a team, through demonstration of good analytical, research, design and implementation skills.

4. To prepare graduates who will thrive to pursue life-long reflective learning to fulfill their goals.

M. Tech. Programme Objectives

Program Outcomes of M. Tech. Program are:

Engineering programs of NU have been designed to prepare graduates for attaining the following program outcomes:

- An ability to apply knowledge of mathematics, statistics, and engineering in practice.
- An ability to identify, critically analyze, formulate and solve engineering problems with comprehensive knowledge in the area of specialization.
- An ability to select modern engineering tools and techniques, and use them with appropriate skills.
- An ability to design a system and process to meet desired needs within realistic constraints such as health, safety, security and manufacturability.
- An ability to contribute by research and innovation to solve engineering problems.

M. Tech. Programme Objectives

Program Outcomes of M. Tech. Program are … :

- An ability to understand the impact of engineering solutions in a contemporary, global, economical, environmental, and societal context for sustainable development.
- An ability to function professionally with ethical responsibility as an individual as well as in multidisciplinary teams with positive attitude.
- An ability to communicate effectively.
- An ability to appreciate the importance of goal setting and to recognize the need for life-long reflective learning.

Applied Machine Learning (AML)

Applied Machine Learning

AML Course Code: 6CS203

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to:

- Comprehend statistical methods as basis of machine learning domain.
- Apply and evaluate variety of machine learning algorithms.
- Implement machine learning techniques to solve complex problems in interdisciplinary domains.

Applied Machine Learning

Introduction to AML Course:

1.1 Importance of the course

Machine learning is the science of getting computers to act without being explicitly programmed. In the past decades, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome.

1.2 Objective of the Course

The Objective of the course is to help students learn the statistical and mathematical interpretation of the data, Basic Machine Learning Algorithm and apply it effectively for the application domain.

1.3 Pre-requisite:

Basic knowledge of statistics and mathematics concepts.

Applied Machine Learning

Syllabus of AML Course:

Unit-I: Introduction: Motivation and Applications, Basics of Supervised and Unsupervised Learning (3 Hours)

Unit-II: Regression Techniques: Basic Concepts and applications of Regression, Simple Linear Regression – Gradient Descent and Normal Equation Method, Multiple Linear Regression, Non-Linear Regression, Linear Regression with Regularization, Hyper-parameters tuning, Loss Functions, Decision Tree Regression, Evaluation Measures for Regression Techniques (13 Hours)

Unit-III: Classification Techniques: Naïve Bayes Classification: Fitting Multivariate Bernoulli Distribution, Gaussian Distribution and Multinomial Distribution, K-Nearest Neighbors, Classification Trees, Linear Discriminant Analysis, Support Vector Machines: Hard Margin and Soft Margin, Kernels and Kernel Trick, Evaluation Measures for Classification Techniques (10 Hours)

Applied Machine Learning

Syllabus of AML … :

Unit-IV: Artificial Neural Networks: Biological Neurons and Biological Neural Networks, Perceptron Learning, Activation Functions, Multilayer Perceptrons, Back-propagation Neural Networks, Learning with Momentum, Winner-take-all Learning, Competitive Neural Networks, Adaptive ANN

(9 Hours)

Unit-V: Clustering: Hierarchical Agglomerative Clustering, k-means Clustering Algorithm, Self-Organizing Maps (4 Hours)

Unit-VI: Advances in Machine Learning: Basics of Semi-Supervised and Reinforcement Learning, Introduction to Deep Learning, Best Practices for Machine Learning, Case Studies in interdisciplinary domain (6 Hours)

Applied Machine Learning

Suggested Readings:

1. C. Bishop, Pattern Recognition and Machine Learning, Springer
2. R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification and Scene Analysis, Wiley
3. KishanMehrotra, Chilukuri Mohan and Sanjay Ranka, Elements of Artificial Neural Networks, Penram International
4. Tom Mitchell, Machine Learning, TMH
5. RajjanShinghal, Pattern Recognition, Techniques and Applications, OXFORD
6. AthemEalpaydin, Introduction to Machine Learning, PHI
7. Andries P. Engelbrecht, Computational Intelligence - An Introduction, Wiley Publication
8. Andrew Kelleher, Adam Kelleher, Applied Machine Learning for Data Scientist and Software engineers, Addison-Wesley Professional

AML: Practical Sessions

		Marks
1.	Overview of Python	-
2.	Unsupervised Learning- K means	10
3.	Unsupervised Learning- Fuzzy C Means	
4.	Statistical approach of Linear Regression for univariate	10
5.	Gradient descent approach of Linear Regression univariate and Multivariate	
6.	Gradient Descent for classification	10
7.	Naïve Bayes Classification for Categorical data	10
8.	Naïve Bayes Classification Continuous data	
9.	Decision Tree-ID3	10
10.	Decision Tree-C4.5 and CART	
11.	Support Vector Machines	10
12.	Kernel Tricks	
13.	Image Classification	40
14.	Image Classification and Analysis	

AML: Assessment Policy

Assessment Policy:

Assessment of "Continuous Evaluation (CE)" comprises of three components:

- One class Test will be conducted as per academic calendar. It will be conducted for the duration of 1 hour and 15 min and will be of 35 marks.
- One Sessional Exam will be conducted as per academic calendar. It will be conducted for the duration of 1 hour and 15 minutes and will be of 35 marks. (Comprehensive Assessment)
- Assignments comprising of all the topics covered in the class (30 marks).

Assessment Policy for Semester End Examination (SEE) :

- A written examination of 3 hour duration will be conducted for the course as per academic calendar.
- It will carry 100 marks and the marks obtained out of 100 will be converted as per weightage assigned.

UNIT-I: Motivation and Applications

What is Data Science:

- It is said that Data is not just Data, but it has more than it (More information. It has something to tell.
- It has some Known and Unknown information hidden in it.
- Known info is obtained from Statistical Analysis of parameters such as: Mean, Standard Deviation, Mode, Median, Minimum, Maximum, Quartile Values, interquartile ranges and Outliers etc.
- And also some patterns in the data are known a priori. Such patterns are retrieved by querying the data. For querying, we need to know what to ask or query.

UNIT-I: Motivation and Applications

What is Data Science … :

- The SQL query language deals mainly with extraction of known information and patterns.
- Such a pattern tells more about the properties of sample or population.
- But, the Data Science deals with extraction of information of unknown patterns.
- Such extracted information can then be used for prediction (Not necessarily in time domain. Could be spatial prediction or in some abstract feature space).
- Prediction is an important aspect of any science and technology.

UNIT-I: Motivation and Applications

What is Data Science … :

- Also it is said that “the future belongs to those companies and people who convert data in to products”.
- This comes under the domain of Data Science.
- There are many faces and sides of Data Science:

  (i) Technology, (ii) The companies and (iii) The unique skill sets.
- The web is full of “data Science Apps”, most common being e-commerce applications.
- For every web-app, there is database in the back end. There is a middleware that queries the data base, extracts unknown patterns and then delivers the data services.
- A data application acquires its values from data itself and creates more data as a result in the process of running applications.

Applied Machine Learning

Unit-1: Motivation and Applications

UNIT-I: Motivation and Applications

What is Data Science … :

- The Data Science enables the creation of data products.
- Examples: The well known CDDB data base (Compact Disk-Data Base).
- This is based on unique track length of data on a CD and coupled to a data base of album metadata. (Metadata is the data about the data.)
- Just based on track length mostly unique to a CD, the meta data, all info of the CD could be retrieved.
- It is said the Google is a master at creating data products.
- For example: Google Page Rank Algorithm, Spell check, Speed recognition, year 2009 Flu epidemic predictions based on data of searches of flu related query topics.

UNIT-I: Motivation and Applications

What is Data Science … :

- The Data collected from users provides added value. The users are in a feedback loop in which they contribute to products they use.
- There has been a Big-Bang (large scale) explosion in the amount of data generated in last a few decades.
- Main Sources of the these data are: (i) Web server logs, (ii) Tweet streams, (iii) Online transaction records, (iv) Citizen Science, (v) Sensors data (IoT instruments), (vi) Govt. data etc. etc.
- The present day problem is not about finding data, but what to do with the data, how to learn from the data and how to use them more effectively.
- Using data effectively requires something different form traditional statistics.
- What differentiates data science from statistics is that the Data Science needs a holistic approach and based on advanced statistical thinking.
- We need the data to tell story and present the story to others.
- Any story must have logical beginning, logical flow and logical end.

UNIT-I: More about Data Science

More about the Data Science … :

- The Data Science concept is @50 yrs. old.
- Any Data has a life cycle:
  (i) Where it comes from, (ii) How it is used and (iii)Where it goes.
- The Data Science is the Science of understanding data using processes, skills and techniques that aid/help in decision making.
- It involves techniques for identifying, collecting and exploring data using plots (especially colourful plots) and graphs.
- The Data Mining and Machine Learning are processes of probing more deeply to understand the data better, bring out hidden and unknown patters and use them effectively.

UNIT-I: More about Data Science

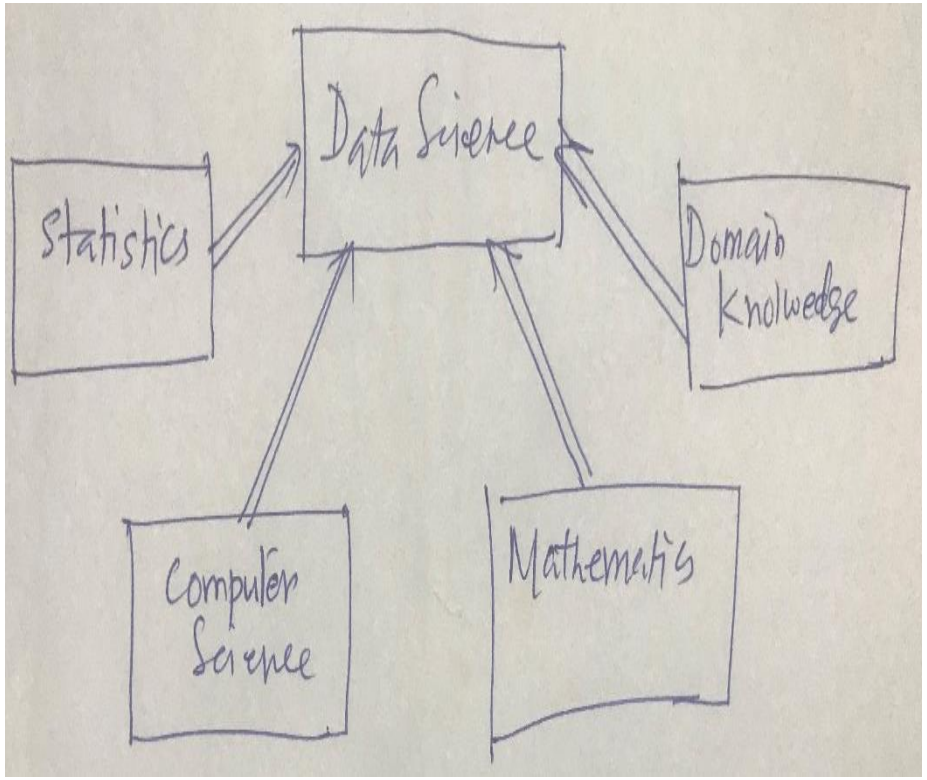
Increased attention to Data Science is due to:

- Availability and increased use of the internet.
- Growing use of smart phones, tablets and digital devices capturing large volume of data.
- Increased use of social media leaving digital foot print behind (also known as Data Exhaust).
- Increased computational capability in both hardware and software.
- Development in programming languages (available through Open Source) to work with such data.
- Easy and speedy access to such data
- Increased storage capacity at cheaper rates.
- Most important: Development of variety of complex algorithms.

UNIT-I: Motivation and Applications

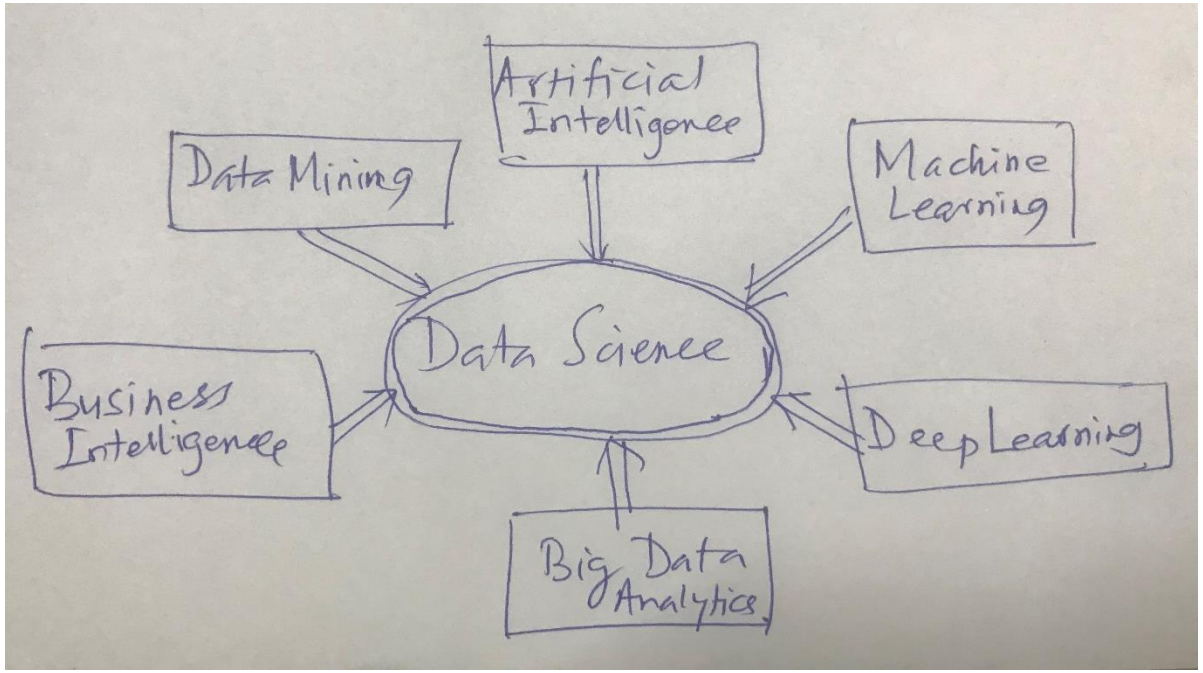
Fundamental Fields of Study related to Data Science:

- Maths and Statistics are foundations of Data Science
- Computer Science and Information Technology give Structure of programming languages.
- Domain Knowledges help in clear formulation of problems and objectives in mind and interpreting outputs of Data Science algorithms.
- Clear understanding of the related technologies help in comprehending their applications in the real world.



UNIT-I: Motivation and Applications

The technologies related to the Data Science:



- Clear understanding of the related technologies help in comprehending their applications in the real world.

UNIT-I: Motivation and Applications

Data life cycle:

(i) Where it comes from, (ii) How tit is used, and (iii)Where it goes.

Where the data Comes from?

- The data comes from Your web server (Web 2.0), Business partners, Facebook accounts, twitter etc.
- Sites like Infochimps and Factuals provide access to large volumes of data sets.
- Point-of-Sale devices, credit cards usages.
- The social data are expanding to fill the space created for storing.
- Mobile applications leave richer data trails (known as Data Exhaust).

Hardware speeds have increased and Costs have come down:

- Since early 1980s CPU processor speeds have increased from 10 MHz to 3.6GHz (increase by 360 times).
- RAM cost decreased from $1000/MB to $25/GB (by factor of 1/40000)
- Now, the 32 GB microSD cards have become cheaper and they weighs only ½ gram.

UNIT-I: Motivation and Applications

How to make use of Data:

- General Steps in Handing Web Data:
  - First step: Data Conditioning (i.e. Bring the Data in usable form).
  - Clean HTML data with “Beautiful Soup” app (It is an NLP application to parse HTML data to plain text in English and other languages).
  - Second step: Replace missing data values
  - Third Step: Try Google trends to examine Casandra DB (use the NLP library). Casandra DB is NoSQL data base.
  - Fourth Step: Use Mechanical Turk to replace AI having Human Intelligence.

UNIT-I: Motivation and Applications

What is the Big Data:

- Definition of Big Data: Big Data is when the size of the data itself becomes part of the problem of data processing.
- That is the traditional techniques for working with Big Data run into out of capacity for its processing, due to limitations of hardware as well as software.
- The Big Data is characterized by three Vs, i.e. Volume, Velocity and Variety.
  - Volume refers to large size (Terra to Peta bytes in size).
  - Velocity refers to its property of continuously being generated and streaming,
  - Variety refers to the data formats like Formatted data, unformatted data, structured and unstructured data in text, audio and video formats.

UNIT-I: Motivation and Applications

What is the Big Data:

- Traditional RDBMS stops being effective at the scale of the Big Data.
- To store such huge data sets effectively, NoSQL /Non-RDBMS are usefully applied (These are Logical descendants of Google Table/Amazon Dynamo).
- They are designed to be distributed across many compute nodes and to have flexible schema.
- Most Common NoSQL databases are: Casandra and HBase (part of Hadoop Distributed Computing System).

UNIT-I: Motivation and Applications

Big Data Sources:

- Most common sources of the big data are Telecom companies and several data centric industries.
- Size and complexity of the Big Data is time dependent. Today's Big Data may become tomorrows Medium Scale data and later Small scale data.
- It depends on the available hardware and software capabilities and their data storage and processing capabilities at any time.
- Big Data Platforms: They are akin to data warehouses, they expose rich APIs, designed to explore and understand data.

UNIT-I: Motivation and Applications

Hadoop Big Data Processing Framework:

- Hadoop framework is a “one stop solution” to Big Data processing.
- It has components like Hive app, HBase data storage, Pig/Pigeon query language, Zookeeper for application management, workflow management, etc.
- These all deal with Big Data file in the form of distributed chunks across multiple nodes stored in a fault tolerant style.
- The big data file is divided into chunks of a fixed size (typically 64GB); known as Hadoop Distributed File System (HDFS).
- The data is divided into chunks and stored in HDFS format, which is distributed file system stored in fault tolerant architecture.

UNIT-I: Motivation and Applications

Hadoop Big Data Processing Framework:

- The data chunks are processed in parallel by nodes using Mapper program.
- The results of processing by all nodes are combined by Reducer software (MapReduce programming environment).
- Hadoop has enabled an ‘agile’ data analysis system.
- The agile system is described by:

  (i) Faster product generation cycle, (ii) closer interaction between developers and consumers and (iii) Faster computations and testing to make it easy to test different assumptions, data sets and algorithms).

UNIT-I: Motivation and Applications

Who is a Data Scientist (DS):

- DS requires skills of computer science, mathematics, statistics.
- DS has to think about a big picture and big problem.
- DS has to make the data to tell a story.
- DS has to have the ability to break a large problem into smaller problems.
- DS has to combine entrepreneurship with patience, the willingness to build data products incrementally, the ability to explore and ability to iterate over a solution.
- DS has to have ability to think out-of-box solution to come up with new ways to view problem. **"Here is a lot of data, what can you make from it"**.
- DS should have the ability to take data, understand, process , extract value(s) from it, visualize to effectively communicate.

- These are the important skills of a data scientist.
- Data is indeed like a new slogan "Intelligence Inside" (like Intel Inside)

Applied Machine Learning

Unit-1: Motivation and Applications

UNIT-I: Motivation and Applications

What is Machine Learning (ML):

- Machine Learning is science and art of programming computers so that computers (i.e. Machines) can learn the patterns in data, i.e. they learn from the data. It is mimicking human way learning.
- General Definition: ML is the field of study that gives computers the ability to learn without being explicitly programmed.
- Engineering-Oriented Definition: A computer is said to learn from experience E with respect to task T and some performance measure P, if its performance on task T as measured by measure P improves with experience.
- Here E is Data, task T is what information we are looking for from the data, and P is the accuracy required.

UNIT-I: Motivation and Applications

History of Machine Learning:

- It is more than 20 yrs. since the computer program DeepBlu, by IBM, defeated world class Russian chess champion Garry Kasporov.
- The computer program DeepBlue is a ML program.
- Google has become one of the front runners focussing on a lot of research on ML and AI, e.g. Googles Self Driving Cars, Google Brain.
- The foundation of ML was laid in 18^{th} and 19^{th} centuries. First related work was dated back to year 1763.
- In 1763, Thomas Bayes work on “An essay towards solving problems in Doctrine of Chances”, was published 2 yrs. after his death.

UNIT-I: Motivation and Applications

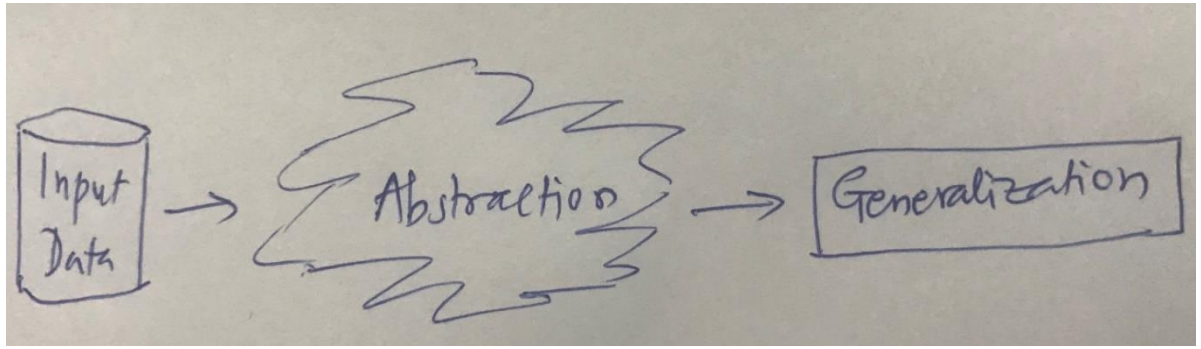
History of Machine Learning:

- In 1982, the Bayes Theorem was actually formalized by French Mathematician Pierre-Simon Laplace. The Method of Least Square was formulated in 1805.
- The real start of focussed work on ML is considered to be Alan Turing's work in 1950. He was the first person to propose that the Machines can learn and become intelligent like human beings.
- In 1952, Arthur Samuel of IBM Labs started working on ML programs and first developed programs that could play checkers game.
- In 1957, Frank Rosenblatt designed first Neural Network (NN) program simulating the human brain. Known as Rosenblatt Perceptron.
- The NN algorithm developed in 1969, the Recurrent Neural Network was in 1982, the Support Vector Machine and Random Forest methods were developed in 1995, and Google AlphaGo was developed in 2016.

UNIT-I: Motivation and Applications

How do Machines Learn:

- The ML imbibes the philosophy of human leaning, i.e. Learning from an expert guidance and learning from experience and examples.
- Thus ML has become a way of life now a days.
- The ML process can be divided into 4 parts:
  1. **Data Input,**
  2. **Abstraction and learning, and**
  3. **Generalization.**



- Between Abstraction and Generalization, there is step of Regularization, if needed.

UNIT-I: Motivation and Applications

What is Machine Learning….

- The ML has been adopted by various organizations and domains, such as Banking and Financial Services, Insurance, health care, Life Sciences etc. to solve complex problems.
- To avoid ethical issues, the critical consideration is required before applying ML and using any outcome from ML.

Typical Application Areas: Supervised Learning, Unsupervised Learning, Semi-supervised Learning, Deep Learning and Reinforced Learning.

Examples: Supervised Leaning:

- Handwriting Recognition
- Stock Market Predictions
- Epidemic (Disease) Prediction
- Fraud Detection

UNIT-I: Motivation and Applications

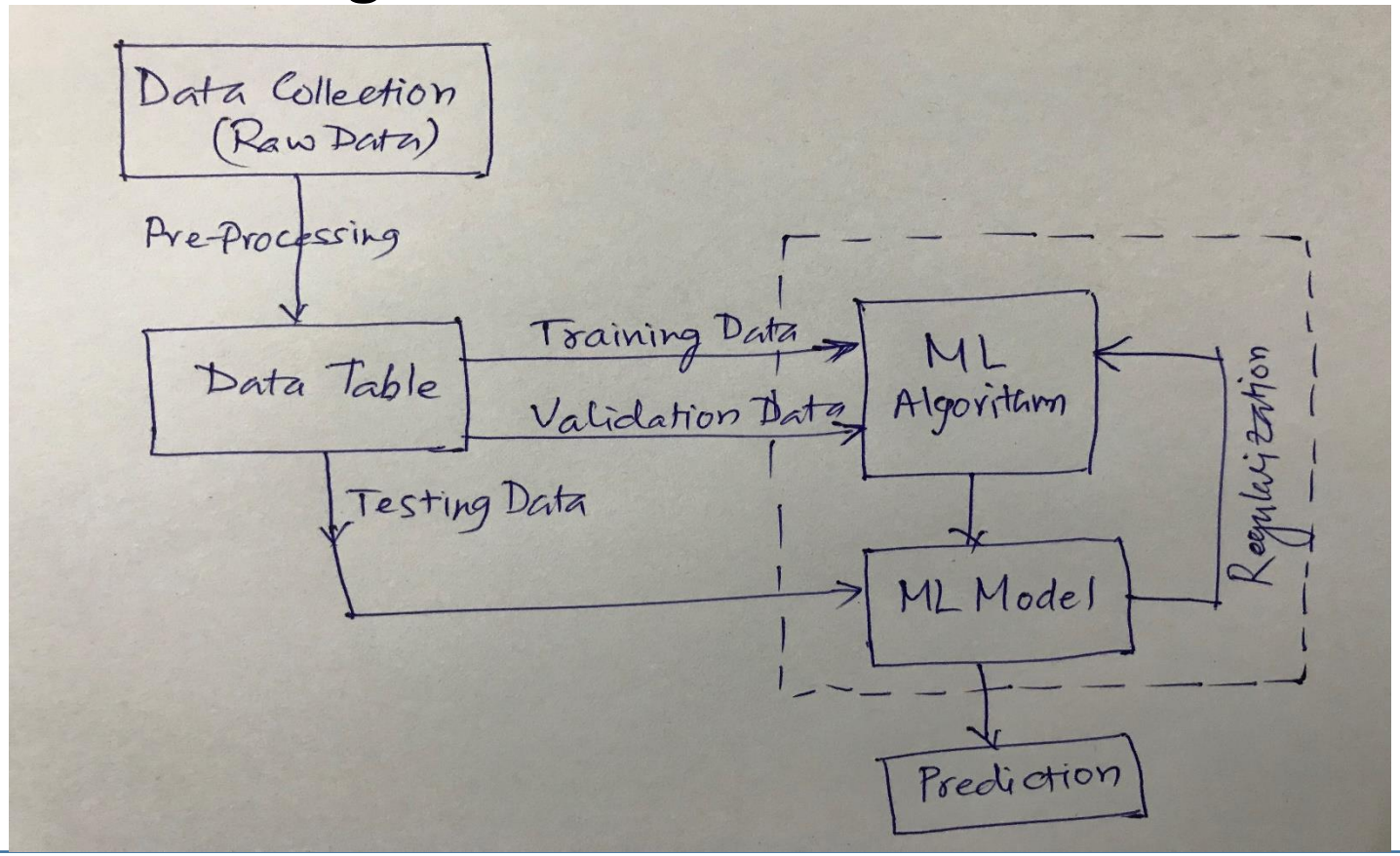
Machines Learning Work Flow:

There are five components of ML workflow:

1. Raw data collection, Data extraction, preprocessing and putting it a tabular format.
2. Randomly Splitting the data into 2 or 3 components, viz. Training Data, Testing data, and Validation data:
3. Applying ML algorithm to training data and optionally to validation data. This results into a ML model.
4. Apply the model to testing data to evaluate its performance. Regularize the model if needed to avoid both overfitting and underfitting, and obtain good generalization.
5. Use the model for forecasting.

UNIT-I: Motivation and Applications

Machines Learning Work Flow:



UNIT-I: Motivation and Applications

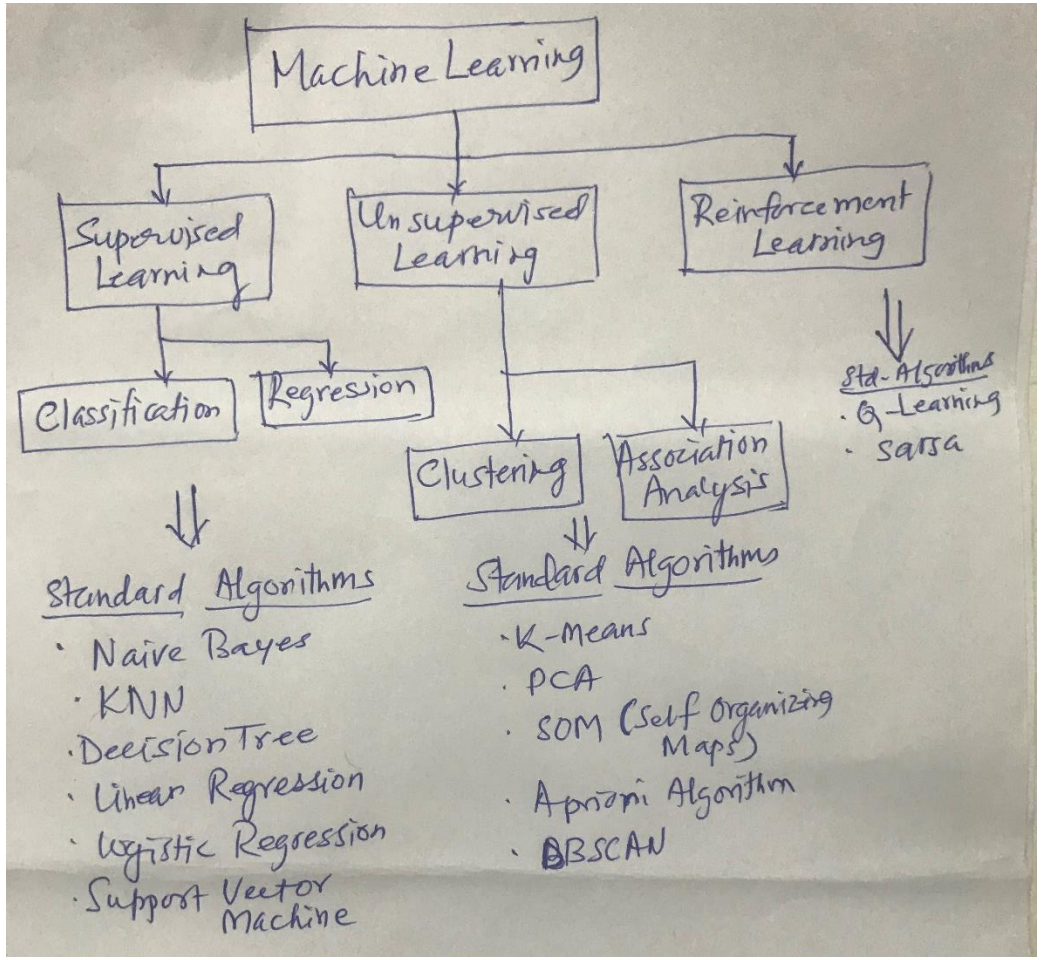
Terminologies used:

1. The object in the table: Label, Object, Sample, Instance, and Example. Denoted by variable ‘y’ generally, and it is a dependent variable.
2. The object properties in the table: Attribute, Feature. Denoted by variable X (Capital x) generally, and they are independent variables.
3. Algorithm is a mathematical and statistical data processing method.
4. Model is what results after applying algorithm to data. It depends on the quality and quantity of the data used.
5. Regularization is method used to avoid either model Overfitting or underfitting for good generalization.
6. Generalization means ensuring the model performance on testing data at par with that on training data.
7. Validation data is used during training to test the model being developed.

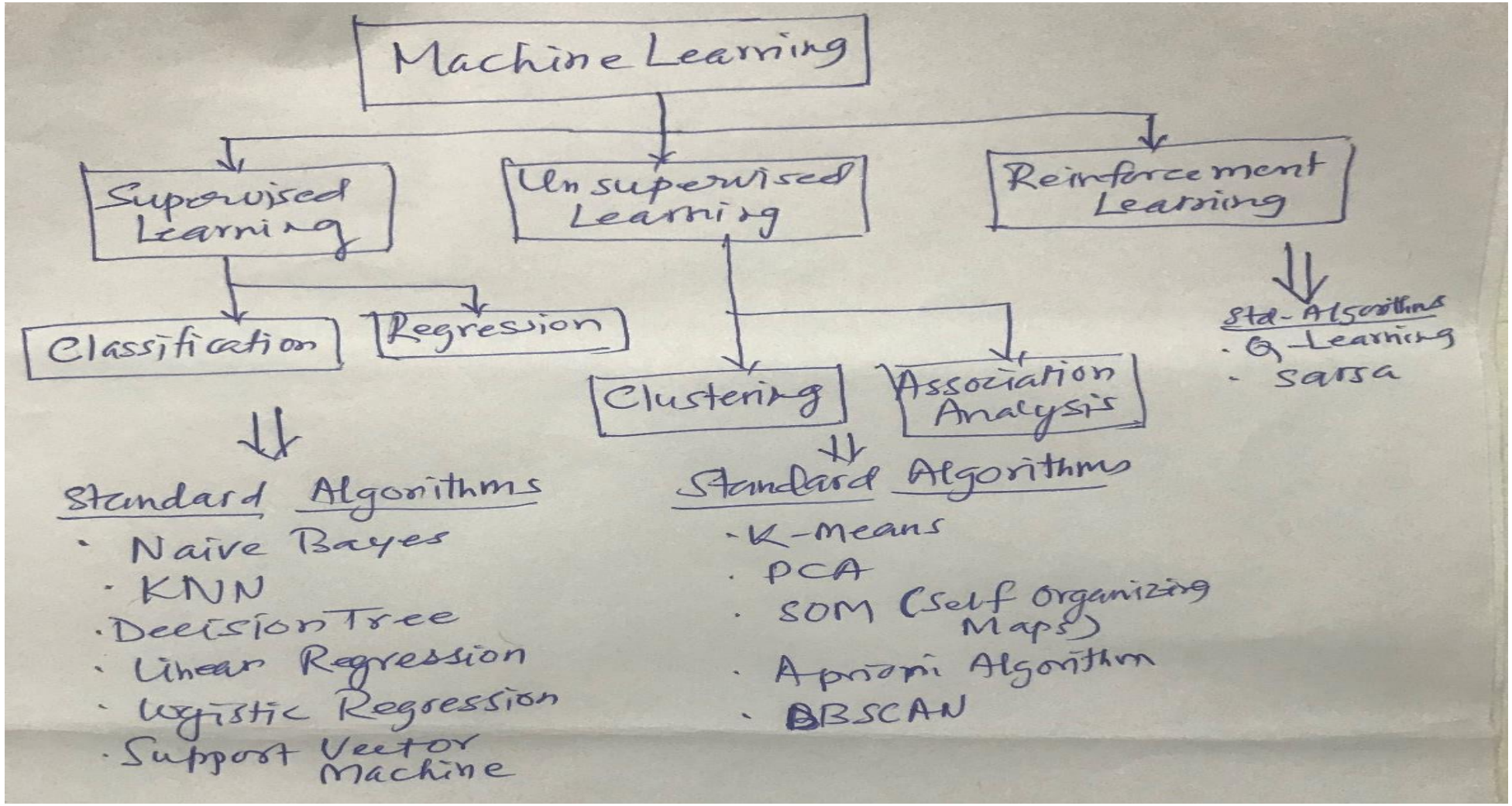
UNIT-I: Motivation and Applications

Typical Application Areas:

- Unsupervised Leaning:
  - Market Basket Analysis
  - Recommender Systems
  - Customer Segmentation
- Reinforcement Learning:
  - Self Driving Cars
  - Intelligent Robots
  - AlphaGo Zero (The latest version of DeepMinds AI System Playing Go)



UNIT-I: Motivation and Applications



UNIT-I: Features and Features Engineering

What are features and their properties:

- A feature is an attribute of a data set that is used in ML, be it supervised or unsupervised learnings.
- Feature extraction is responsible for taking raw input data and converting into well known features useful in ML model development.
- Only those attributes which are meaningful to a ML problem are called features (but may not be so always).
- A data set having n-district features is known as n-dimensional data set.
- Features are independent properties (also distinct) of a data set.

UNIT-I: Features and Features Engineering

Features Engineering:

- It is process of transforming a data set into features data set such that they are able to represent data set more effectively and hence result into better learning performance.
- Feature Extraction: It is the process of extracting or creating a new set of features from the original set of features using some functional mapping.
- Feature extraction involves getting rid of redundant features without loosing much information existing in the data.
- This is one of the time consuming process in ML.
- Selecting right features has a critical role to play in success of ML model developed.
- Features construction expands the features space, whereas features extraction and feature selection reduces the features space.

UNIT-I: Advanced Learning Algorithms

What are the advanced learnings:

- There are two more types of advanced learning, viz. (i) Semi-supervised learning, (ii) Reinforcement learning and (iii) Deep Learning.
- In Semi-Supervised learning, some classes are derived from unsupervised learning and some are from supervised learning.
- This case occurs when the full information in data are not available.
- In Reinforcement Learning, the algorithm deals with examples with no labels (as in unsupervised learning case).
- But examples with the +ve or –ve feedbacks as the solution the algorithm proposes.

UNIT-I: Advanced Learning Algorithms

What are the advanced learnings:

- It is related to applications for Deep Learning and Reinforcement Learning algorithms.
- The RL must make decisions bearing consequences.
- It (RL) is just like learning by Trial and Error. It is used extensively in Video Games.
- The RL algorithm lets one to know the outcomes of an action taken and the learning occurs by avoiding what it discovers to be dangerous.
- Example: Google's mind for playing video games.
- Deep Learning: It is learning through use of the multiple layers of neural networks.

UNIT-I: Motivation and Applications

ML can also help Humans to learn:

- ML algorithms can be inspected to see what they have learnt from data.
- For example: Once the Spam filter is trained on enough data on Spams, it can be inspected to reveal the list of words and combinations of words which form Spam.
- This can sometimes reveal unsuspected correlations or new trends, thus lead to better understanding of Spam problems.
- ML technique applied to large volume of data can help to discover patterns that were not immediately apparent (from the Technology of Data Mining and visualization)

Applied Machine Learning

Unit-1: Motivation and Applications

UNIT-I: Motivation and Applications

When to use Machine Learning (ML):

- For problem which require long and complex list of rules.
- Complex problems for which there is no good solution at all using traditional approaches.
- In Fluctuating Environment, ML system has capability to adapt to new data.
- For getting insight into complex problems and large amount of data.

More about Machine Learning:

- It is the main tool for data scientists.
- Many libraries are available for ML programming, such as
  (i) Python, (ii) Elefant, (iii) Weka, (iv) Mahout, (v) KNime etc.
- The Google prediction API for public use. It exposes their ML algorithms using an interface.
- OpenCV is a defacto standard library for computer vision.
- “Mechanical Turk”, a tool box for ML used to develop excellent training sets.

UNIT-I: Motivation and Applications

What is Supervised Learning:

- In supervised learning, an algorithm learns from the example data.
- It is similar to humans learning under the supervision of a teacher.
- Associated target responses consists of numerical values or string labels (categorical variables, such as classes or tags).
- The main purpose of Supervised Learning is to predict the correct response to a new data set or example.
- There are two types of Supervised Learnings:

  (i) Regression approach: in which target is a numerical value and the attributes are also numerical values.

  (ii) Classification approach: in which target has a qualitative value/variable (such as a class or tag or label or categorical, nominal)
- In case of modelling a numerical value the price of an item on its parameters, the regression approach determines/models the price on its attribute parameters.
- If the attributes and the dependent variable are not numerical values (may be nominal, logical etc.), then they have to be assigned numerical values.

UNIT-I: Motivation and Applications

What is unsupervised Learning:

- In unsupervised leaning, the algorithm learns from plain examples without any associated responses.
- The algorithm determines the data patterns on its own.
- The algorithm restructures the data in such a way that each structure represents a class or new classes of uncorrelated values of parameters.
- It is quite useful in providing humans with insights into the meaning of data and new useful inputs to supervised learning algorithms.
- It resembles the methods which humans use to figure out certain objects from same class, such as degree of similarity between objects.
- Example: Recommender system in form of marking automation.
- The marketing algorithm derives its suggestions from what customers have bought in the past.
- It classifies the customers based on their buying patterns.

UNIT-I: Motivation and Applications

Advanced Machine Learnings:

- There are two types of advanced learning techniques, viz. (i) Semi-supervised Learning and (ii) Reinforcement Leaning
- In semi-supervised learning some classes are derived from unsupervised learning and some are derived from supervised learning.
- This case occurs when the full information in data is not available.
- In Reinforcement learning, the algorithm deals with examples with no labels (as in unsupervised learning).
- But, examples with + ve or – feedback as the solution the algorithm proposes.
- It is related to applications for the RL algorithm must make decisions bearing consequences.
- It is just like learning by trial and error. Used extensively in video games.
- The algorithm lets to know the outcome of an action taken and learning occurs by avoiding what it discovers to be dangerous.

UNIT-I: Motivation and Applications

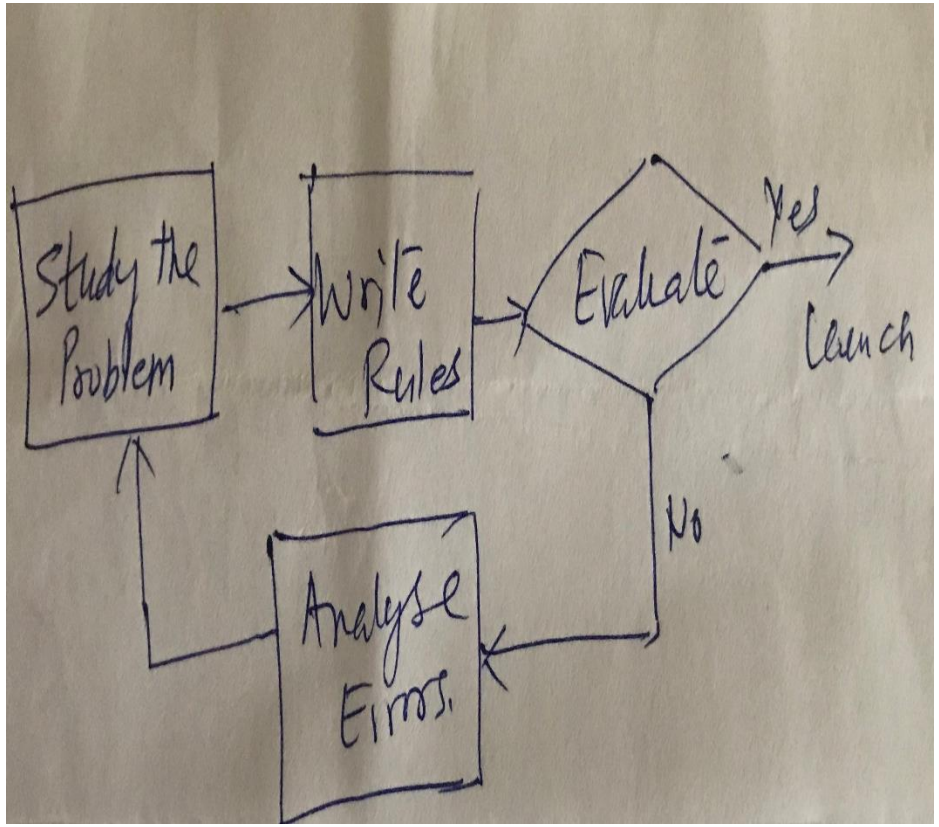
Example of ML:

- Well known example is the “Spam Filter used in eMail”.
- The task of Spam filter is check every incoming email for its nature as per the criteria set and if found to be spam, then send to a Spam folder.
- The Spam Filter is Machine Learning program.
- In this the given examples are of Spam emails and regular emails (non-Spam). These act as training data sets.
- Here,
  - T ➔ To flag (or label) Spam to new emails being received
  - E ➔ Training data
  - P ➔ Performance measure (Could be ratio of correctly classified Spam emails to total emails received).

UNIT-I: Motivation and Applications

Traditional programming technique of SPAM filtering:

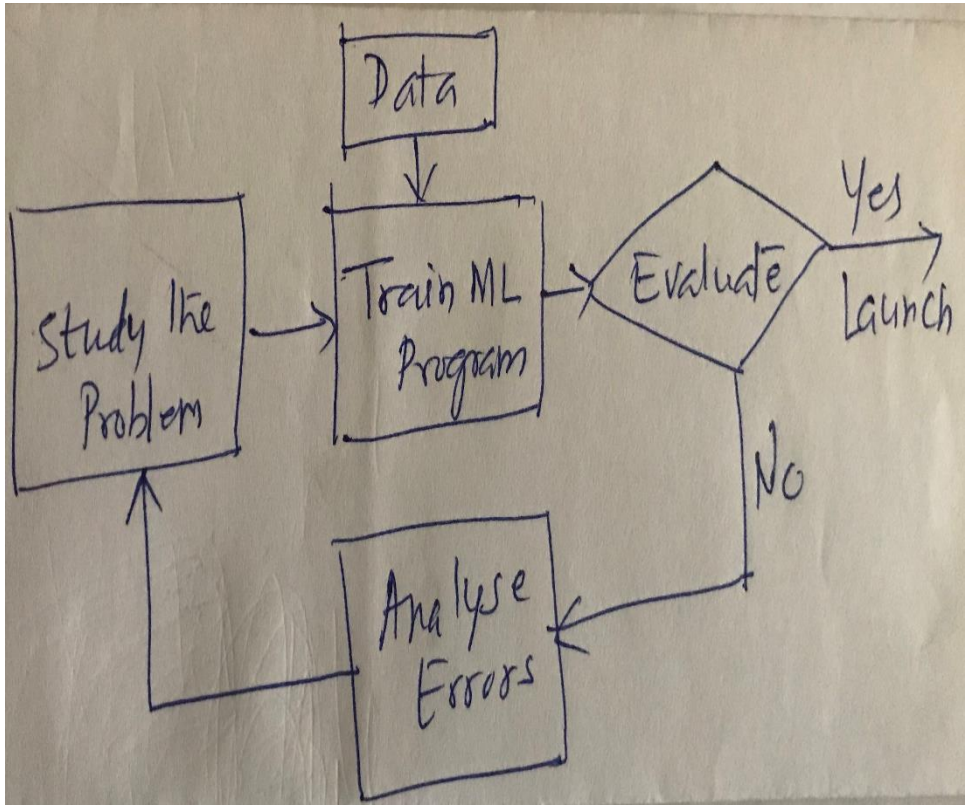
- Study the problem to find out what is (are) given, what is expected as output, and what processing need to be done.
- Write rules and develop algorithm. Evaluate the outcome /results and
- Take a decision to accept the algorithm.
- If acceptable, then launch the algorithm for wider use.
- This will become an endless process. Not a practical approach.



UNIT-I: Motivation and Applications

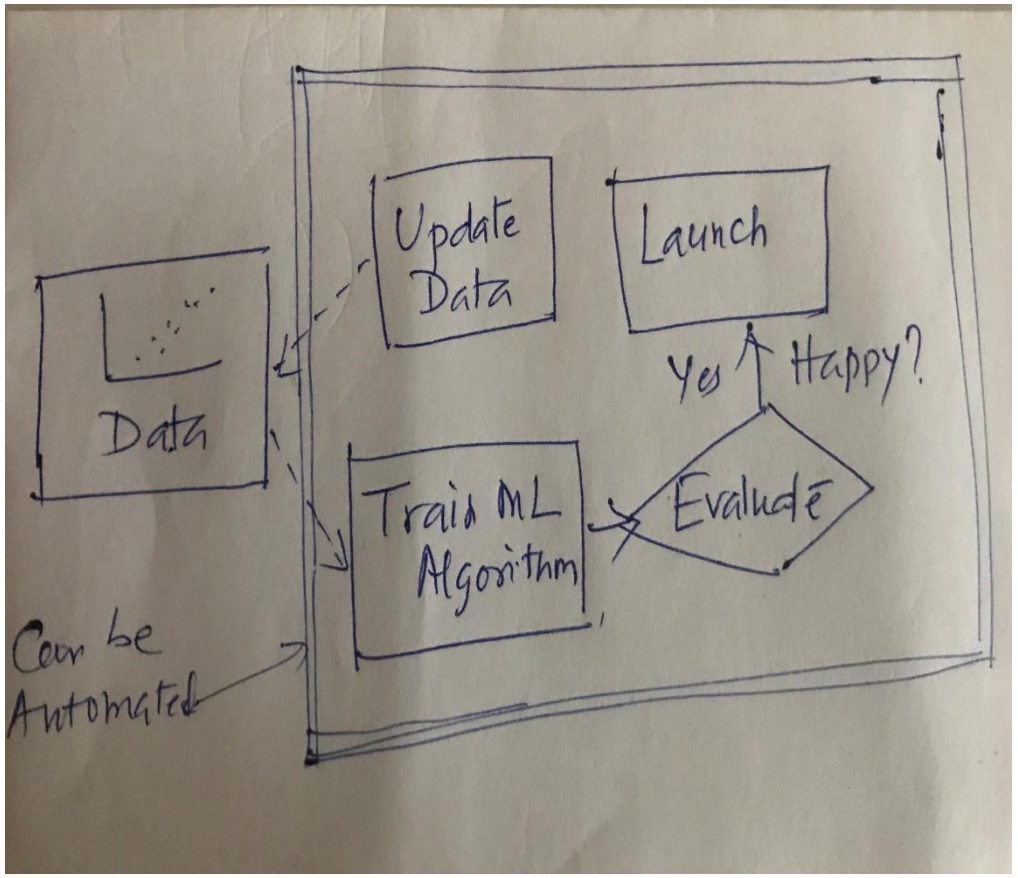
Machine Learning Approach

- Study the problem to find out what is given and what is expected from the given data
- Split the data into train and training data sets.
- Train the ML algorithm on the test data and development the model
- Apply the model on Test Data and evaluate the model performance
- If acceptable, Launch the algorithm.
- If not, go back to problem study and find better ML model.
- Since writing rules is not trivial, the program becomes a long list of complex rules



UNIT-I: Motivation and Applications

- If the above ML algorithm does not stand the test of time, then the data needs to be updated continuously.
- The model also need to be improved upon and tested.
- The updating the data base is very crucial as the model learns from the new cases in the data bases and learns from them.



UNIT-I: Motivation and Applications

Examples: **Typical Application Areas:**

- Supervised Leaning:
  - Handwriting Recognition
  - Stock Market Predictions
  - Epidemic (Disease) Prediction
  - Fraud Detection
- Unsupervised Leaning:
  - Market Basket Analysis
  - Recommender Systems
  - Customer Segmentation
- Reinforcement Learning:
  - Self Driving Cars
  - Intelligent Robots
  - AlphaGo Zero (The latest version of DeepMinds AI System Playing Go)