Day:3

1st NOVEMBER 2023

1. **Semi-supervised Machine Learning:**

Semi-supervised Machine Learning is the combination of both Supervised and Unsupervised Machine Learning, since they use both labeled data(contains both input and output columns) and unlabeled data(contains only input data). for training—typically a small amount of labeled data and a large amount of unlabeled data. This is a rarely used technique.

The system that use this method can considerably improve learning accuracy.

Example:

Suppose you are doing Sentiment Analysis and you have around 1000s of data of students about how they are feeling (Happy, sad, etc.) . Lets assume that you have only data of 1 to 100 student’s (which is called labeled data) and we do not have the information about the rest of the data(Which is also called unlabeled data). In this case, Supervised Machine Learning is used for the 1-100 students data and Un-supervised Machine Learning is used for the rest of the data.

This case uses both Supervised and Unsupervised ML. Hence, it is called Semi-supervised Machine Learning. Unsupervised learning is not used for predictions.

1. **Reinforcement Machine Learning/ Reward based Machine Learning:**

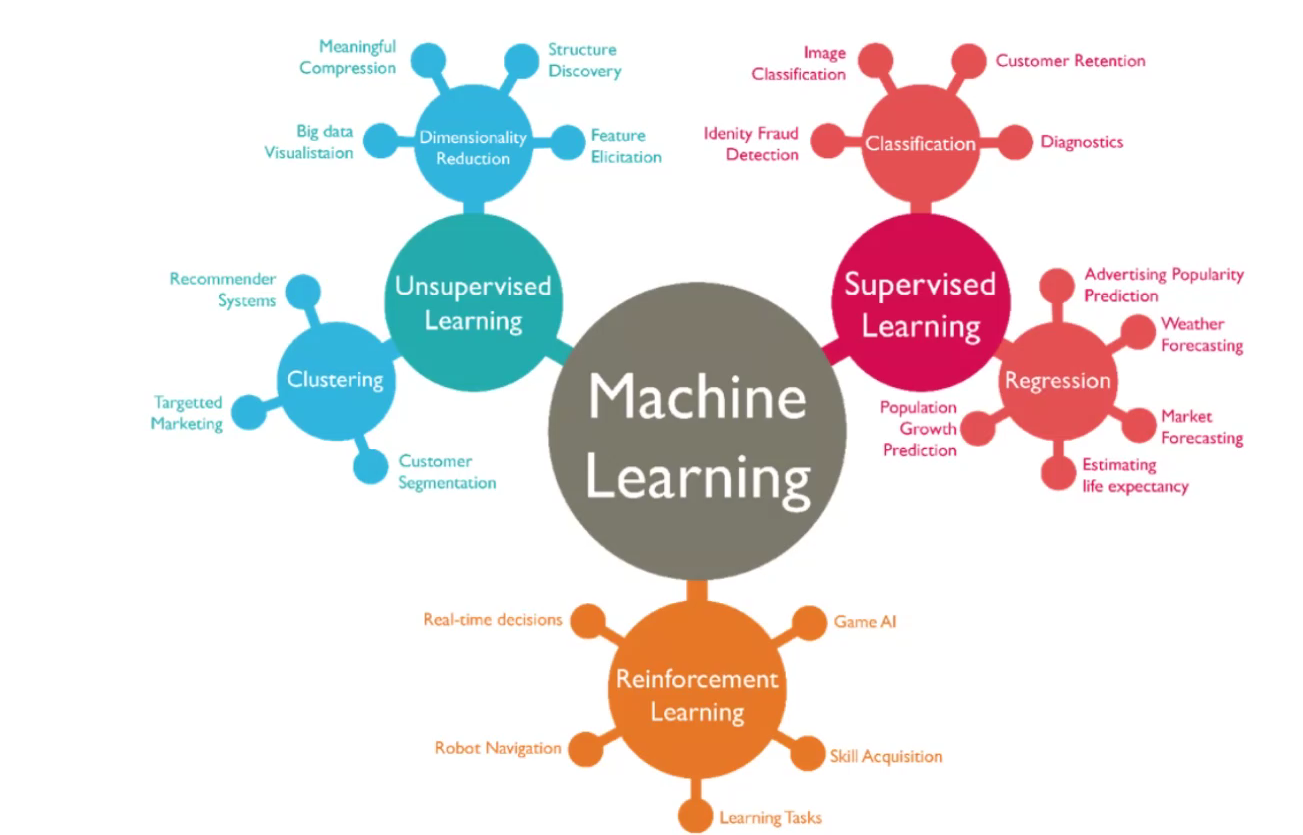
Reinforcement Learning is a learning by interacting with it’s environment by producing actions and discovers errors or rewards. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize it’s performance.

Simple reward feedback is required for agent to learn which action is best; this is known as the reinforcement signal.

In reinforcement Machine Learning, no data and training are provided. This ML first observes the environment and based on the knowledge it has got from the environment, it will perform certain task.

For example: Automated cars.

If this model performs the tasks, we can give positive feedback. If it mess up the task, we can also give negative feedback. Based on these feedbacks, the machine would learn. Example: TAY (Thinking About You) chatbot (This AI got shut down after it caused subsequent controversy)

**Fig: Overview of Machine Learning Categories and Their Applications**

**Machine Learning Algorithms:**

**Unsupervised Learning:**

Continuous

* Clustering and Dimensionality Reduction:
* SVD
* PCA
* K-means

Categorical

* Association Analysis
* Apriori
* FP-Growth
* Hidden Markov Model

**Supervised Learning:**

Continuous

* Regression
* Linear
* Polynomial
* Decision Trees
* Random Forests

Categorical

* Classification
* KNN
* Trees
* Logistic Regression
* Naive-Bayes

Categorical means “Text Data”.

Continuous means “Numerical data”- data which have decimal values.

**Fig: Types of Numerical Datatype**

**Fig: Types of Categorical Datatype**

**Data Science Industry Applications, Workflow and Job Roles:**

**Health Care:**

(AI Predict Heart Attack- Better than Doctors)

University of Nottingham researchers created an AI system that scanned routine medical data to predict which patients would have strokes or heart attacks within 10 years.

<https://futurism.com/neoscope/confirmed-ai-can-predict-heart-attacks-and-strokes-more-accurately-than-doctors>

**Retail:**

(Machine Learning in Retail)

What to stock, how much to buy, what products to suggest to repeat customers. But doing more with that data using machine learning is just what retailers need to really succeed in the current market.

<http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world/>

**Banking:**

(AI in Banking)

* Fraud Transactions Detection
* Predicting risk of asset class based on future context models.
* More personalized and faster customer experiences.

**Financial Services-** Experian:

* With approximately 3.6 petabytes of data (and growing) about individuals around the world, credit reference agency, Experian gets its extraordinary amount of data from marketing databases, transactional records and public information records.
* They are actively embedding machine learning into their products to allow for quicker and more effective decision-making. Over time, the machines can learn to distinguish what data points are important from those that aren’t. Insight extracted from the machines will allow Experian to optimize its process.

**Financial Services-** American Express:

* American Express process $1 trillion in transaction and has 110 million AmEx cards in operation.
* They rely heavily on data analytics and machine learning algorithms to help detect fraud in near real time, therefore saving millions in loses.
* Additionally, AmEx is leveraging its data flows to develop apps that can connect a cardholder with products or services and special offers.
* They are also giving merchants online business trend analysis and industry peer benchmarking.

**Healthcare-** Infervision:

* AI and deep learning is being put to use to save lives by Infervision.
* In China, where there aren’t enough radiologists to keep up with the demand of reviewing 1.4 billion CT scans each year to look for every signs of lung cancer.
* Radiologists need to review hundreds of scans each day which is not only tedious, but human fatigue can lead to errors.
* Infervision trained and taught algorithms to augment the work of radiologists to allow them to diagnose cancer more accurately and efficiently.

**Manufacturing-** Volvo:

* Cars are increasingly connected and generate data that can be used in a number of ways.
* Volvo uses data to help predict when parts would fail or when vehicles need servicing, uphold its impressive safety record by monitoring vehicle performance during hazardous situations and to improve driver and passenger convenience.
* Volvo is also conducting its own research and development on autonomous vehicles.

**Manufacturing-** BMW:

* BMW has big data-related technology at the heart of its business model and data guides decisions throughout the business from design and engineering to sales and aftercare.
* The company is also a leader in driverless technology and plans for its cars to deliver Level 5 autonomy- the vehicle can drive itself without any human intervention- by 2021.

**Media-** BBC:

* The BBC project, Talking with Machines is an audio drama that shows listeners to join in and have a two-way conversation via their smart speaker.
* Listeners get to be a part of the story as it prompts them to answer questions and insert their own lines into the story.
* Created specifically for smart speakers Amazon Echo and Google Home, the BBC experts to expand to other voice-activated devices in the future.

**Media-** Netflix:

* Big data analytics is helping Netflix predict what its customers will enjoy watching.
* They are also increasingly a content creator, not just a distributor, and use data to drive what content it will invest in creating.
* Due to the confidence they have in the data findings, they are willing to buck convention and commission multiple seasons of a new show rather than just a pilot episode.

**Retail-** Burberry:

* When you first think of Burberry, you likely consider its luxury fashion and not first consider them a digital business.
* However, they have been busy reinventing themselves and use big data and AI to combat counterfeit products and improve sales and customer relationships.
* The company’s strategy for increasing sales is to nurture deep, personal connections with its customers.
* As part of that, they have reward and loyalty programs that create data to help them personalize the shopping experience for each customer.

**Service-** Microsoft:

* Central to everything Microsoft does is leveraging smart machines.
* Microsoft has Cortana, a virtual assistant; chatbots that run Skype and answer customer service queries or deliver info such as weather or travel updates and the company has rolled out intelligent features within its Office enterprise.
* Other companies can use the Microsoft AI Platform to create their own intelligent tools.
* In the future, Microsoft wants to see intelligent machines with generalized AI capabilities that allow them to complete any task

**Service-** Disney:

* Always at the top of delivery extraordinary service, Disney is getting even better thanks to big data.
* Every visitor gets their own Magic Band wristband that serves as ID, hotel room key, tickets, Fast Passes and payment systems.
* While guest enough the convenience, Disney gets a lot of data that helps them anticipate guests needs and deliver an amazing, personalized experience.
* They can resolve traffic jams, gives extra services to guests who may have been inconvenienced by a closed attraction and data even allows the company to schedule staff more effectively.

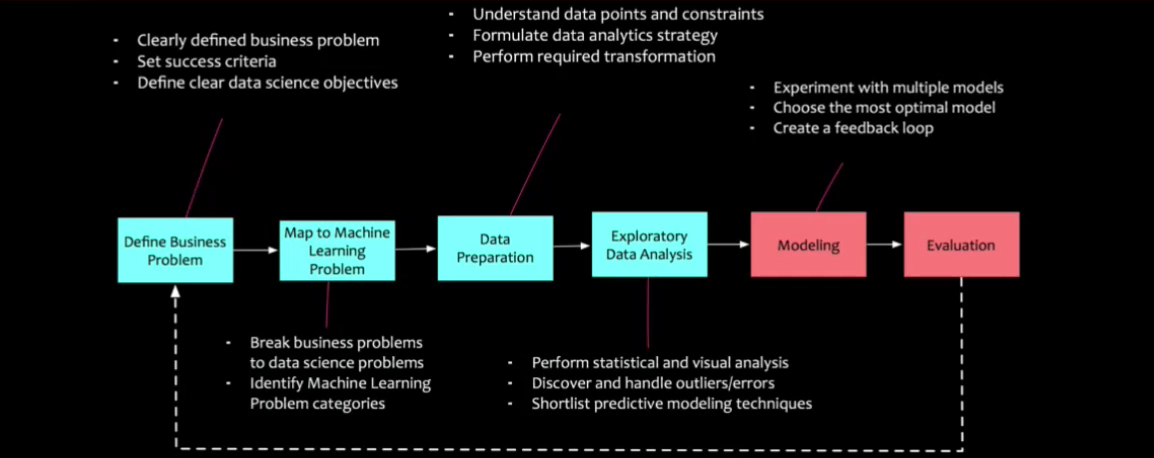
**Social Media-** Facebook:

* Deep learning is helping Facebook draw value from a larger portion of its unstructured datasets created by almost 2 billion people updating their statuses 293,000 times per minute.
* Most of its deep learning technology is built on the Torch platform that focuses on deep learning technologies and neural networks.

**Social Media-** Twitter:

* From what tweets to recommend to fighting inappropriate or racist content and enhancing the user experience, Twitter has begun to use artificial intelligence behind the scenes to enhance their product.
* They process lots of data through deep neural networks to learn over time what user preferences are.

**Data Science Workflow:**

****

**Fig: Data Science Workflow**

1. **Define Business Problem:**

Identifying the business problem is the first step. To keep the project focused on important business outcomes, success criteria and data science goals are established.

1. **Map to Machine Learning Problem:**

This stage breaks the business problem into specific data science tasks. Only the forward process is carried out after it has determined the kind of machine learning problem (classification, regression, etc.) that needs to be addressed. Data science is used to translate the business problem into a technical one.

1. **Data Preparation:**

Following problem mapping, pertinent data is gathered, cleansed, and prepared. Understanding data points, dealing with missing values, and carrying out the required transformations to prepare the data for analysis are all part of this.

1. **Exploratory Data Analysis:**

The fourth step is exploratory data analysis, which looks for patterns, errors, or outliers in the data using both statistical and visual analysis. Choosing the best predictive modeling methods to use later is made easier with this step.

1. **Modeling:**

A variety of machine learning models are created and evaluated. To find the best model for resolving the issue, it is necessary to test out various algorithms and parameters. The best model is selected for implementation.

1. **Evaluation:**

The last phase is to assess how well the model performs using test data. During this stage, the model's robustness and compliance with the established success criteria are confirmed. Feedback loops are made if needed to improve the model.

***Note: Domain analysis is very important.***

**Data Science Roles:**

1. Infra Engineers
2. Data Science Developers
3. Big Data Specialists
4. Data Scientist
5. Researchers
6. Analysis
7. Business Person
8. **Data Science Infrastructure Engineers:**

* Data Science Engineers are hard-core techies who deal with Data Science Infrastructure, ie; hardware, software applications and other aspects to get the back end of Data Science up and running.
* They set up entire IT infrastructure from servers, networks to processes, also manage it such as infrastructure monitoring, application management, database administration, etc.

1. **Data Science Developers:**

* Data Science Developers are the ones, who code the models and applications through programming in R Language, Python, etc. They are versatile developers, who have good knowledge on math and statistics, machine learning algorithms and related concepts.
* As this domain of data science development is evolving rapidly, these developers are expected to keep themselves updates with all latest technological advances from development perspective, so that they can use the right platform to achieve their goal inn effective manner.

1. **Machine Learning Specialist:**

* These are the professional with deep knowledge in computer science and mathematics. They engage in machine learning and deep learning heavily.
* They engage in machine learning and deep learning heavily.
* They create predictive and prescriptive models based on the machine learning algorithms, such as random forest, Artificial Neural Network, K-NN etc.
* They are masters in all kinds of data mining techniques pruning, regularization, etc, helping to create a robust data science models, which can be used in creating great business insights.

1. **Big Data Specialist:**

* Big Data Specialists focus on designing model for of Big Data Processing as one of the building block of Data Science workflow.
* Their work involve architecture of Data gathering including streaming and snapshots, storing and processing in effective and efficiently manner using Big Data Technologies.

1. **Researchers:**

* Researchers are domain experts. These professionals have more expertise in Data Analysis, Data Science related statistics along with significant expertise in specific domain such as HR, Marketing, Fraud Analytics, Health Care, Finance, etc.
* They have less emphasis and knowledge in backend part such as IT Infrastructure, coding, computer science, etc.

1. **Data Science Analyst:**

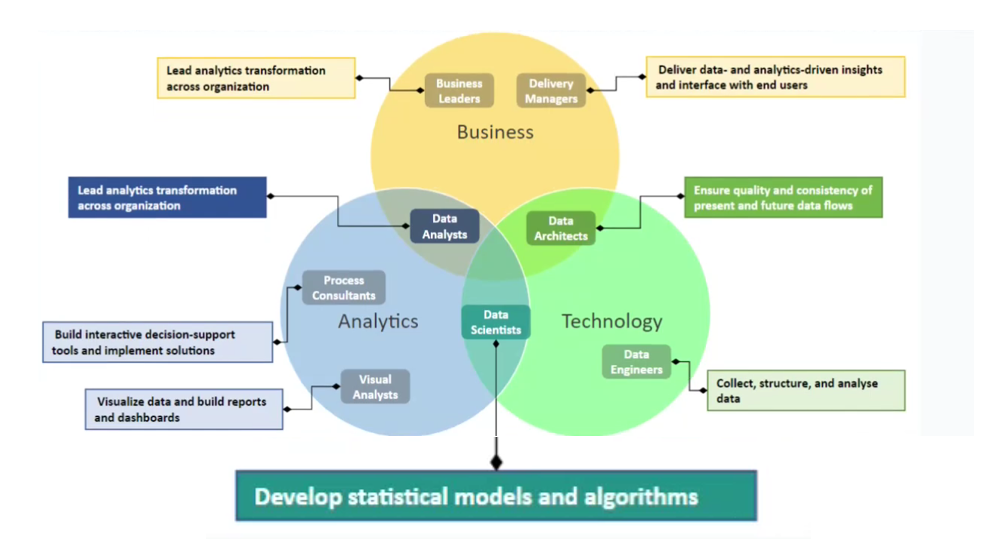
* These professionals are focused on day-to-day analysis of data, including website analytics, retrieving data from various data sources and creating data visualizations.
* They work closely with business person. Their role is to provide the reports from data analysis with appropriate visualizations in an easy to understand format. There by enabling the decision makers to gain valuable business insights.

1. **Business Person:**

* This role is the one, who manages the entire Data Science project. We can call him as sponsor. This role is predominantly business focused and, formulates the problem statements for which, Data science project needs to find the answers. This role also helps in understanding, interpreting the intermediate results of data science.
* Though this role is primarily business focused, he/she must also speak Data in order to able to perform the job well.

1. **Data Scientist:**

* These are the professional who is able to perform every aspect of Data Science sometime called as full-stack Data science professionals. Well, these people are rare as mastering difficult, if not impossible.
* If a company manages to hire full-stack Data Scientist, there would be tremendous progress in transforming the business in gain significant competitive advantage by finding solutions to important business questions.

****

**Fig: Data Science Collaboration Mode**