Machine Learning

Unit 1

Prof. Priti Srinivas Sajja

P G Department of Computer Science,

Sardar Patel University, Vallabh Vidyanagar, Gujarat, India

About the Speaker

- **Name: Dr. Priti Srinivas Sajja**
- **Email : priti@pritisajja.info**
- **Mobile : +91 98249 26020**
- **URL: http://pritisajja.info**
- **Academic qualifications : Ph. D in Computer Science**
- **Thesis Title: Knowledge-Based Systems for Socio-Economic Rural Development**
- **Subject area of specialization : Artificial Intelligence**
- **Publications : 219 in Books, Book Chapters, Journals, Proceedings, etc.**

Unit 1 : Introduction to Machine Learning

- Types of learning: Human and machine learning
- Types of machine learning
- Applications of machine learning
- Tools for machine learning

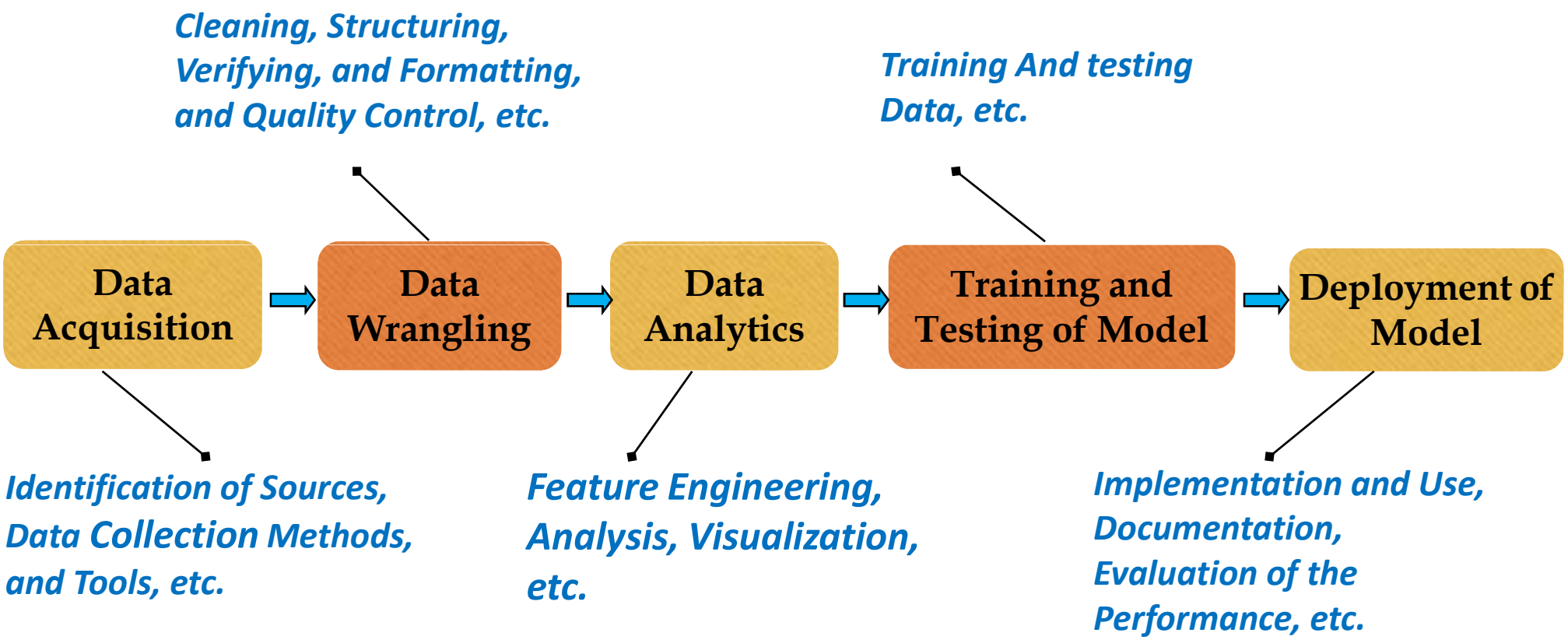
Human and Machine Learning

- Responds to situations **flexibly**.
- **Makes sense** of ambiguous or erroneous messages.
- Assigns **relative importance** to elements of a situation.
- **Finds similarities** even though the situations might be different.
- **Draws distinctions** between situations even though there may be many similarities between them.

Limitations of Symbolic Representations

- Nature of knowledge
  - Hard to characterize
  - Voluminous
  - Dynamic
- Knowledge acquisition
  - Fact finding methods support only
  - Tacit and higher level knowledge
  - Multiple experts
- Knowledge representation
  - Limited knowledge structures support
- KBS development models
  - Only SAD/SE guidelines and a few quality metrics
- Large size of knowledge base

Machine Learning Life Cycle



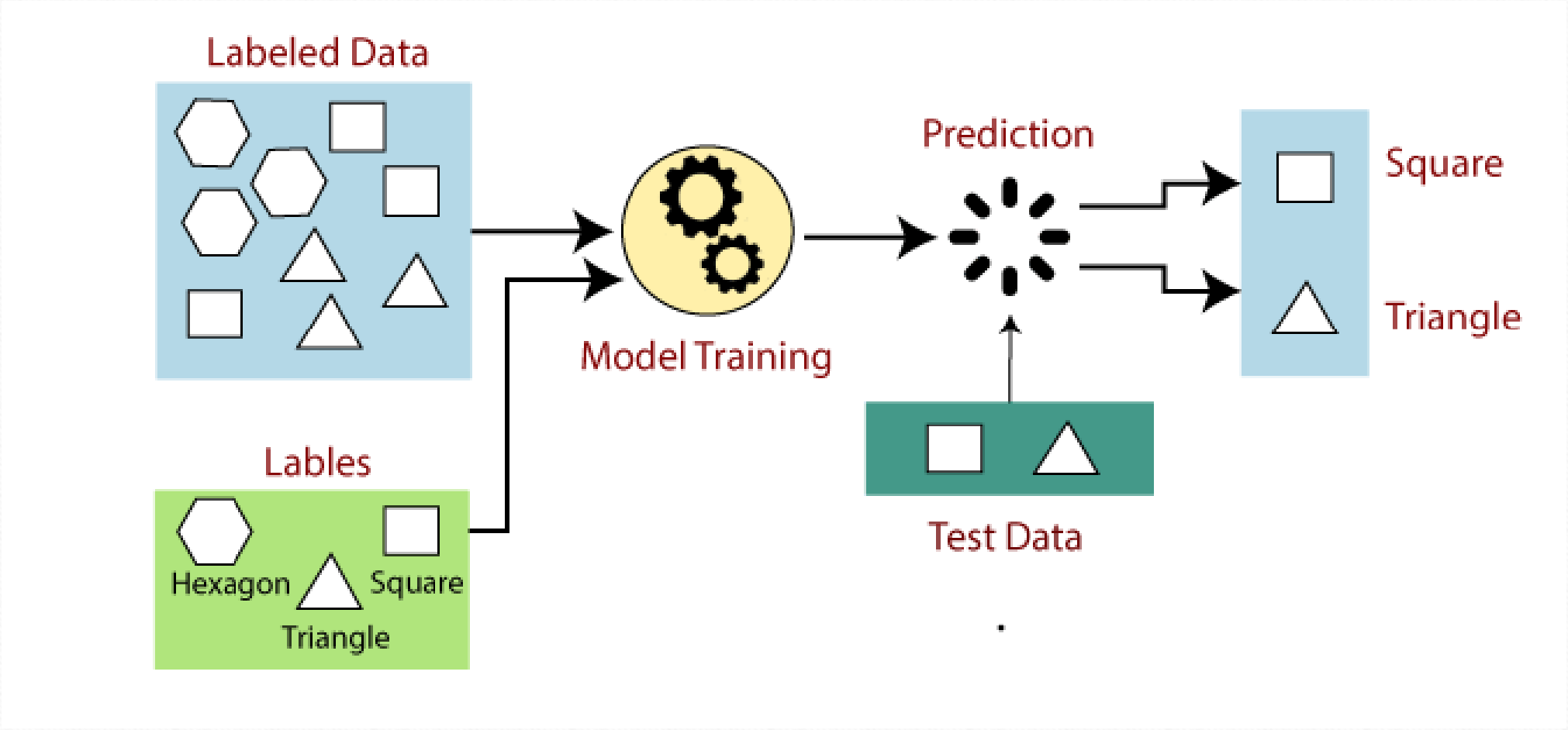
Types of Machine Learning

- Machine Learning
- Supervised Learning
- Un-supervised Learning
- Deep Learning
- Ensemble Learning
- Reinforcement Learning

Supervised Learning

- Supervised learning is the types of machine learning in which **machines are trained using well "labelled" training data**, and on basis of that data, machines predict the output.
- The labelled data means some input data is already **tagged with the correct output**.
- In supervised learning, **the training data** provided to the machines **work as the supervisor** that **teaches the machines to predict the output correctly**. It applies the same concept as a student learns in the supervision of the teacher.
- Supervised learning is a process of providing input data as well as correct output data to the machine learning model.
- The aim of a supervised learning algorithm is to **find a mapping function to map the input variable(x) with the output variable(y)**.

Supervised Learning



Supervised Learning - Steps

- First Determine **the type/domain of training dataset and Collect/Gather the labelled** training data.
- Split the training dataset into training **dataset, training dataset, and test/validation dataset.**
- Determine the **input features** of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- **Execute** the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- **Evaluate the accuracy** of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

Supervised Learning – Popular Techniques

- **Regression**
  - Linear Regression
  - Regression Trees
  - Non-Linear Regression
  - Bayesian Linear Regression
  - Polynomial Regression
- **Classification**
  - Random Forest
  - Decision Trees
  - Logistic Regression
  - Support vector Machines

Supervised Learning – Advantages

- With the help of supervised learning, the model can accurately predict the output on the basis of prior experiences.
- In supervised learning, we can have an exact idea about the classes of objects.
- Supervised learning model helps us to solve various real-world problems such as **fraud detection, spam filtering**, etc.

Supervised Learning – Disadvantages

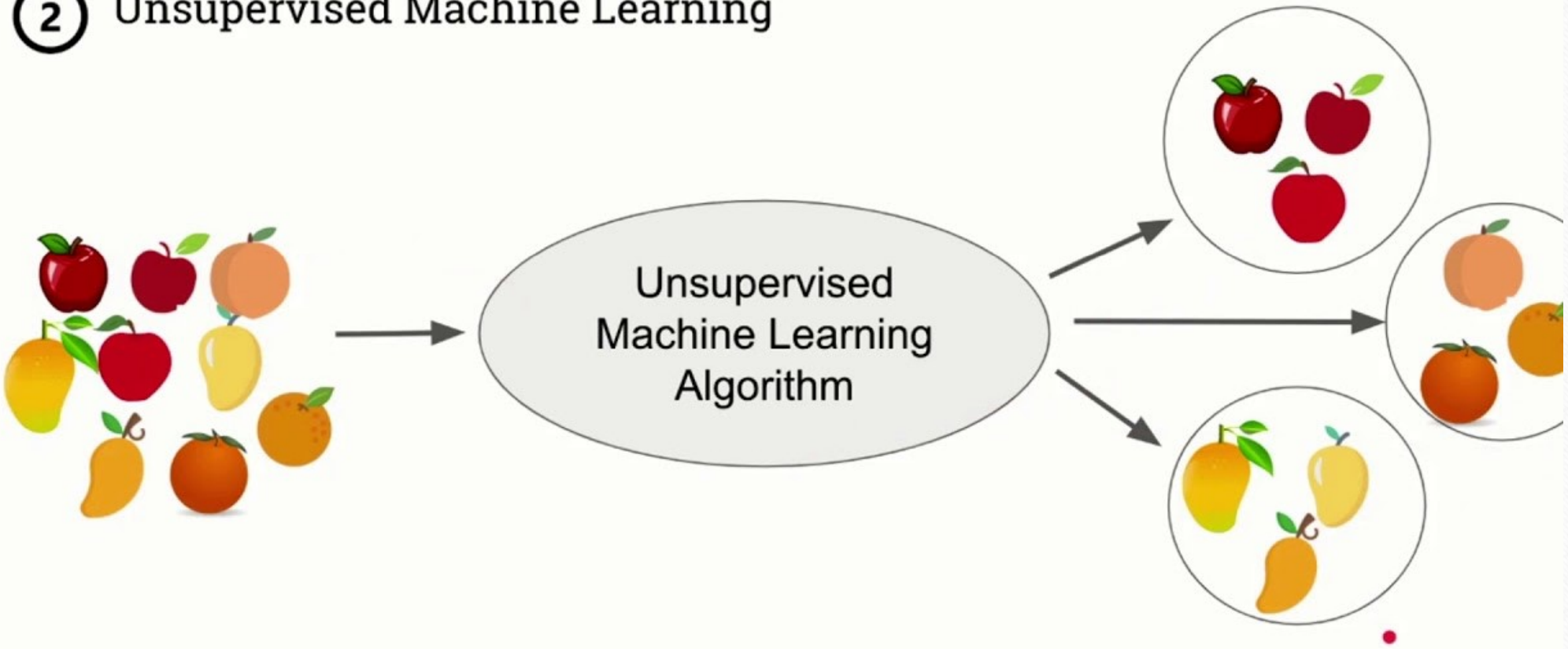
- Supervised learning models **are not suitable for handling the complex tasks**.
- Supervised learning cannot **predict the correct output if the test data is different from the training dataset**.
- Training required **lots of computation times**.
- In supervised learning, we need **enough knowledge about the classes** of object.

Unsupervised Learning

- Unsupervised learning is a machine learning technique in which models are **not supervised** using training dataset.
- Instead, **models itself find the hidden patterns** and insights from the given data.
- It can be **compared to learning which takes place in the human brain** while learning new things. It can be defined as:
- *Unsupervised learning is a type of machine learning in which* ***models are trained using unlabeled dataset and are allowed to act on that data without any supervision***.

Types of Machine Learning Algorithms

(2) Unsupervised Machine Learning



Unsupervised Learning

- Unsupervised learning is helpful for **finding useful insights** from the data.
- Unsupervised learning is **much similar as a human learns** to think by their own experiences, which makes it closer to the real AI.
- Unsupervised learning works on **unlabeled and uncategorized data** which make unsupervised learning more important.
- In real-world, we **do not always have input data with the corresponding output** so to solve such cases, we need unsupervised learning.

Unsupervised Learning : Popular Techniques

- K-means clustering
- KNN (k-nearest neighbors)
- Hierarchal clustering
- Anomaly detection
- Neural Networks
- Principle Component Analysis
- Independent Component Analysis
- Apriori algorithm
- Singular value decomposition

Unsupervised Learning : Advantages

- Unsupervised learning is **used for more complex tasks** as compared to supervised learning because, in unsupervised learning, we don't have labeled input data.
- Unsupervised learning is preferable as **it is easy to get unlabeled data** in comparison to labeled data.
- More **near to real AI.**

Unsupervised Learning : Disadvantages

- Unsupervised learning is intrinsically **more difficult** than supervised learning as it does not have corresponding output.

- The result of the unsupervised learning algorithm might be **less accurate** as input data is not labeled, and algorithms do not know the exact output in advance.

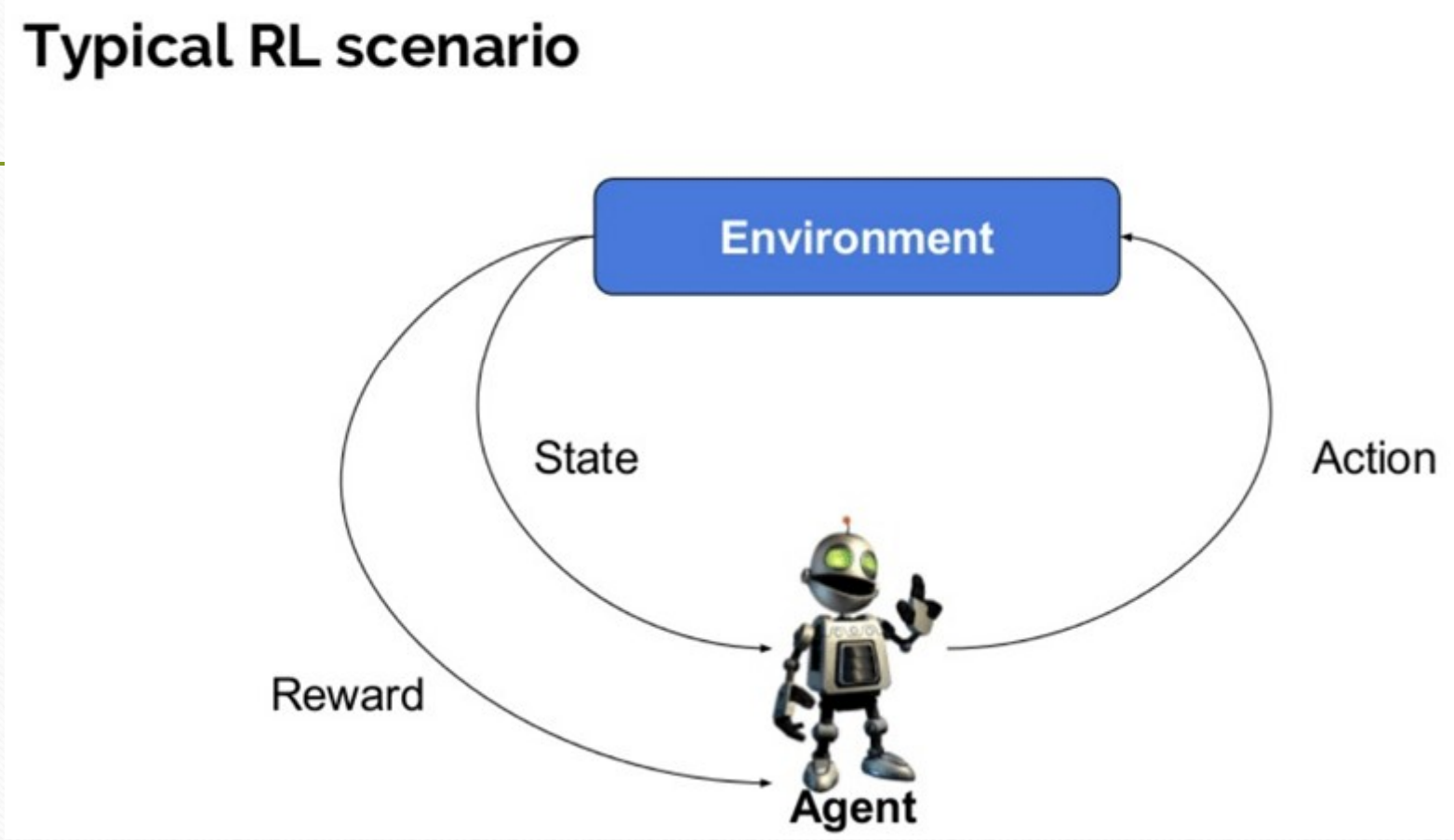
Supervised Learning	Unsupervised Learning
• Supervised learning algorithms are trained using **labeled data**.	• Unsupervised learning algorithms are trained using unlabeled data.
• Supervised learning model **takes direct feedback** to check if it is predicting correct output or not.	• Unsupervised learning model does not take any feedback.
• Supervised learning model **predicts the output**.	• Unsupervised learning model finds **the hidden patterns** in data.
• In supervised learning, **input data** is provided to the model along with **the output**.	• In unsupervised learning, only **input data** is provided to the model.
• The goal of supervised learning is to train the model so that it can **predict the output** when it is given new data.	• The goal of unsupervised learning is to find the **hidden patterns** and useful insights from the unknown dataset.
• Supervised learning **needs supervision** to train the model.	• Unsupervised learning does not need any supervision to train the model.
• Supervised learning can be categorized in **Classification and Regression** problems.	• Unsupervised Learning can be classified in **Clustering and Associations** problems.

Supervised Learning (contd\..)	Unsupervised Learning (contd\..)
• Supervised learning can be used for those cases where we **know the input as well as corresponding outputs.**	• Unsupervised learning can be used for those cases where **we have only input data** and no corresponding output data.
• Supervised learning model produces an **accurate** result.	• Unsupervised learning model may give less accurate result as compared to supervised learning.
• Supervised learning is not close to **true Artificial intelligence** as in this, we first train the model for each data, and then only it can predict the correct output.	• Unsupervised learning is more close to the true Artificial Intelligence as it learns **similarly as a child learns daily routine things b**y his experiences.
• It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc.	• It includes various algorithms such as Clustering, KNN, and Apriori algorithm.

Reinforcement Learning

- Reinforcement Learning is defined as a Machine Learning method that is concerned with how software agents should take actions in an environment.

Reinforcement Learning



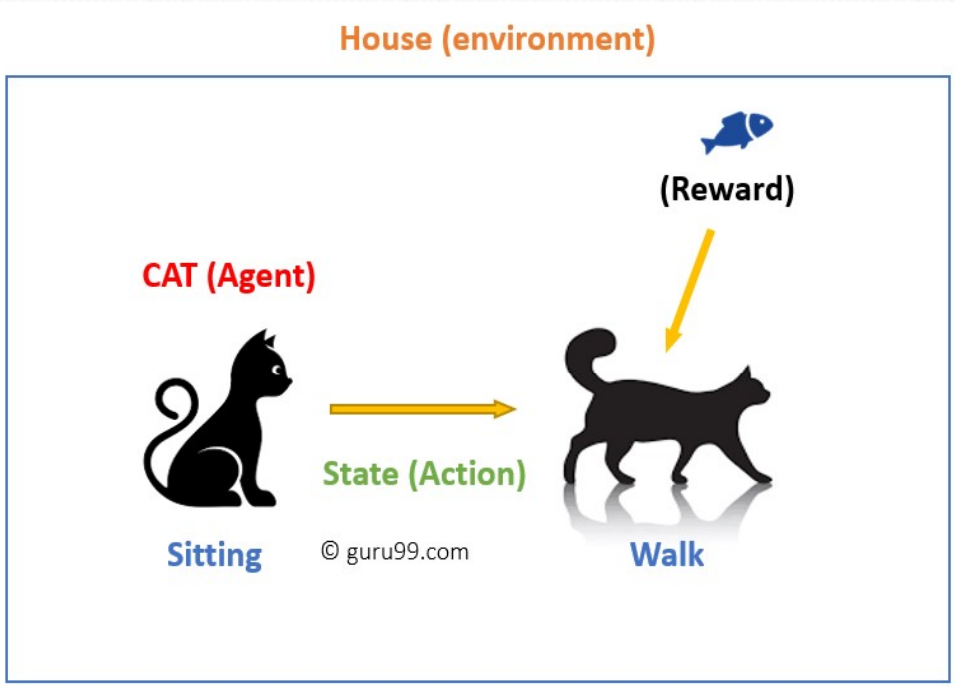
Reinforcement Learning

- **Agent:** It is an assumed entity which performs actions in an environment to gain some reward.
- **Environment (e):** A scenario that an agent has to face.
- **Reward (R):** An immediate return given to an agent when he or she performs specific action or task.
- **State (s):** State refers to the current situation returned by the environment.
- **Policy (π):** It is a strategy which applies by the agent to decide the next action based on the current state.
- **Value (V):** It is expected long-term return with discount, as compared to the short-term reward.
- **Value Function:** It specifies the value of a state that is the total amount of reward. It is an agent

Reinforcement Learning

- Consider the scenario of teaching new tricks to your cat
- As **cat doesn't understand English** or any other human language, we can't tell her directly what to do. Instead, we follow a different strategy.
- We emulate a situation, and the cat tries to respond in many different ways. If the cat's **response is the desired way, we will give her fish**.
- Now whenever the cat is exposed to the same situation, the cat executes a similar action with even more enthusiastically in expectation of getting more reward(food).
- That's like learning that **cat gets from "what to do" from positive experiences**.
- At the same time, the **cat also learns what not do when faced with negative experiences**.

Reinforcement Learning



- Your cat is an agent that is exposed to the environment, which is the house.
- An example of a state could be your cat sitting, and you use a specific word in for cat to walk.
- Our agent reacts by performing an action transition from one "state" to another "state."
- For example, your cat goes from sitting to walking.
- The reaction of an agent is an action, and the policy is a method of selecting an action given a state in expectation of better outcomes.
- After the transition, they may get a reward or penalty in return.

Reinforcement Learning - Characteristics

- There is **no supervisor**, only a **real** number or **reward signal**
- **Sequential decision making**
- **Time plays a crucial role** in Reinforcement problems
- Agent's **actions are dynamic** based on rewards and other data

Reinforcement Learning – Applications

- **Gaming**
- Robotics for **industrial automation.**
- Business **strategy planning**
- It helps you to create training systems that provide **custom instruction and materials** according to the requirement of students.
- **Aircraft control** and robot motion control

Reinforcement Learning – Why?

- It helps you to find which situation needs an action
- Helps you to discover which action yields the highest reward over the longer period.
- Reinforcement Learning also provides the learning agent with a reward function.
- It also allows it to figure out the best method for obtaining large rewards.

Reinforcement Learning – Why Not?

- When you have enough data to solve the problem with a supervised learning method
- You need to remember that Reinforcement Learning is computing-heavy and time-consuming. in particular when the action space is large.

Reinforcement Learning – Challenges

- Feature/reward design which should be very involved
- Parameters may affect the speed of learning.
- Realistic environments can have partial observability.
- Too much Reinforcement may lead to an overload of states which can diminish the results.
- Realistic environments can be non-stationary.

Ensemble Learning

- In statistics and machine learning, ensemble methods use **multiple learning algorithms** to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone.

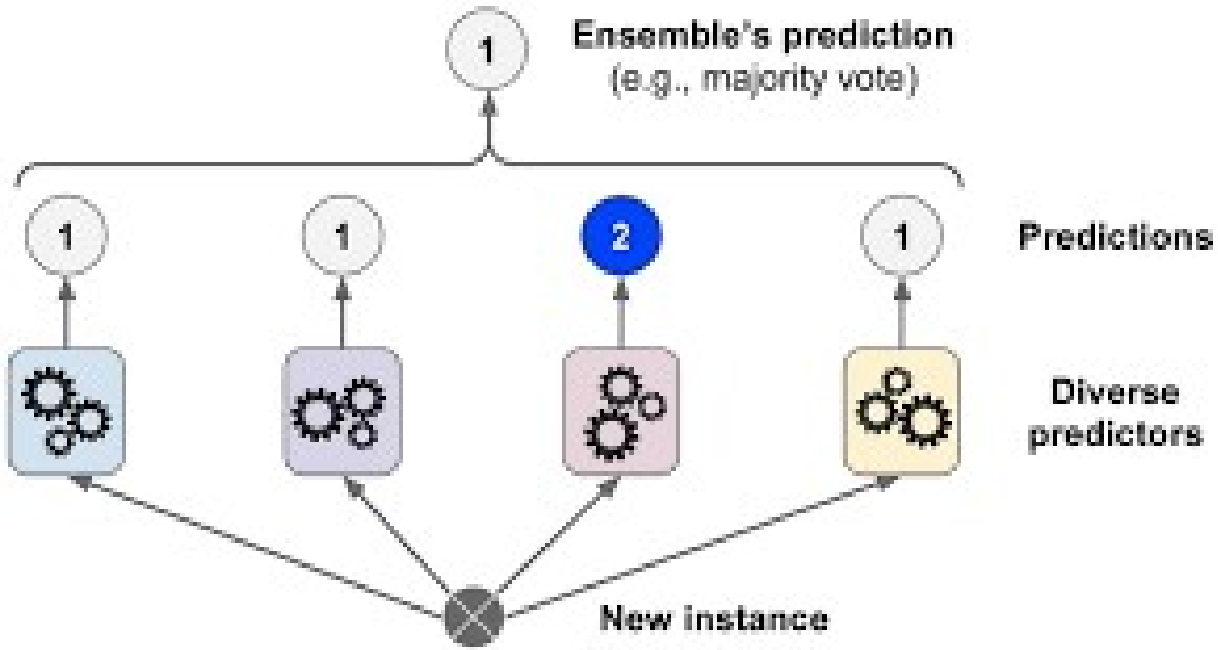
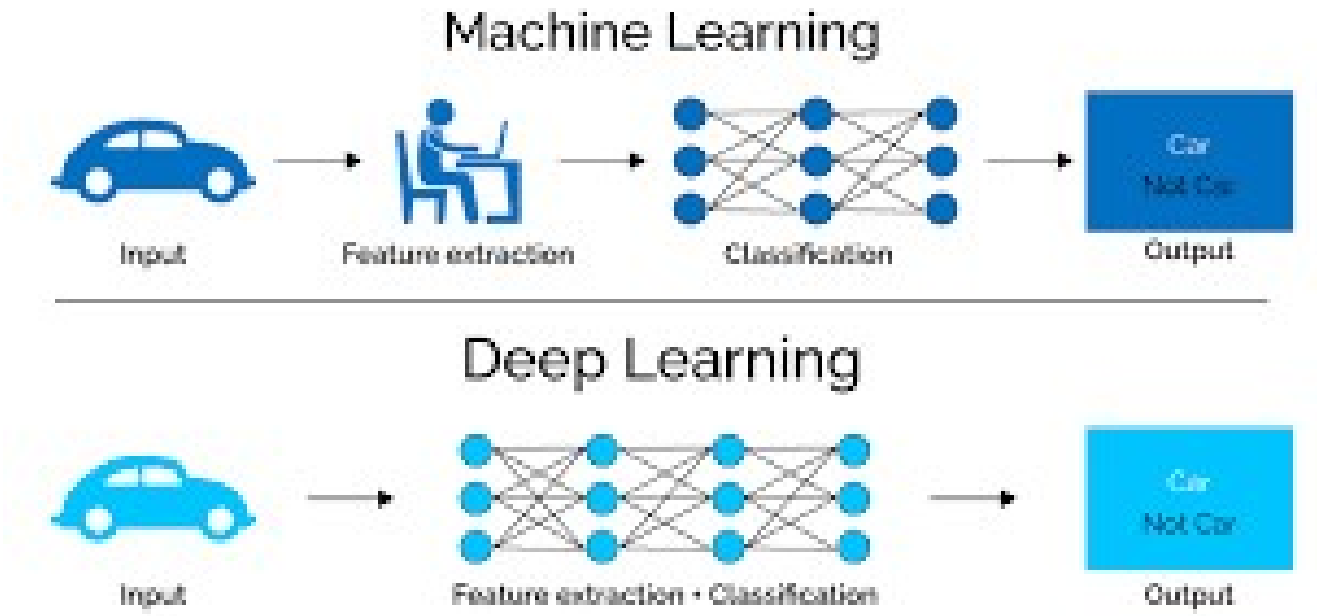


Figure 7-2. Hard voting classifier predictions

Deep Learning

- Deep learning is subset of **artificial intelligence (AI)** that imitates the workings of the human brain in processing data and creating patterns for use in decision making. ... Also known as deep neural learning or deep neural network.



Deep Learning

- Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning.
- Learning can be supervised, semi-supervised or unsupervised.
- Deep-learning popular architectures are
  - deep neural networks,
  - deep belief networks,
  - deep reinforcement learning,
  - recurrent neural networks and
  - convolutional neural networks

Deep Learning

- Deep Learning can be Applied to
  - Computer vision,
  - Speech recognition,
  - Natural language processing,
  - Machine translation,
  - Bioinformatics,
  - Drug design,
  - Medical image analysis, etc.

Tools and Technologies for Machine Learning

- **Python, Java, R, Julia, etc Programming tools**

- **Accord.NET:** Accord.NET is a framework **written in C#** and provides machine learning facilities for **audio and image processing**.
- It also supports **statistical data processing, pattern recognition**, and other performance measurement techniques.
- It is used for the scientific and machine learning based computing on **.NET framework**. However, it support **only .NET** supported languages.
- http://accord-framework.net/

Tools and Technologies for Machine Learning

- **AutoML**

- AutoML by the Google clouds supports computer vision, NLP, translation, and other model building application using the machine learning algorithms. It considers the **labelled training data** and outputs an optimized model. It is not a completely free package. However, its **trail version is free**.

- https://cloud.google.com/automl/

Tools and Technologies for Machine Learning

- **CoLab**
- CoLab is a short form of **Colaboratory**.
- It is an environment that **executes directly on cloud** without any additional set ups.
- It is also called **Jupyter Notebook environment**. One can directly call (import) all the packages and libraries without installing them manually.
- The CoLab is a **free tool** by the Google to employ various machine learning algorithms. According to the claim of the Google, it requires **zero set up**.
- https://research.google.com/colaboratory/

Tools and Technologies for Machine Learning

- **Hadoop and Spark**
- Apache Haddop is a **framework based on the java programming language** for the distributed processing of big data. Typically the Hadoop framework has a **computation layer and a storage layer**. The processing layer uses a parallel programming approach called **MapReduce**. The storage layer offers distributed file system based on the **Google File System (GFS)**.
- **Apache Spark** is considered as an **extension of the Hadoop** to handle batch as well as real time (streaming) data with efficient built in models for big data handling. It is advertised as a '**lightning fast unified analytical engine**'.
- https://hadoop.apache.org/
- https://spark.apache.org/

Tools and Technologies for Machine Learning

- **KNIME**
- KNIME is another machine learning tool that helps in generating the data science solutions/workflows.
- This tool is visual, user friendly and **eliminates writing codes**. Instead, it uses **drag and drop** kind of visual interface.
- This tool allows the practitioner to access and transform the data, offers modelling and visualization, and deployment of the model.
- https://www.knime.com/

Tools and Technologies for Machine Learning

- **LibROSA**

- LibROSA is also a **python package**, especially developed to manage **audio and music analysis**. In most of the speech to text conversion applications and music retrieval systems, this tool is used.

- https://librosa.org/

Tools and Technologies for Machine Learning

- **Machine Learning Studio**
- Machine Learning Studio by the Microsoft Azure is a visual drag and drop type of tool to experiment machine learning models. The trail version for the tool is free.
- **MLFlow**
- The tool MLFlow manages the complete life cycle of a machine learning application till deployment of a model and its use. According to the claim of its developers, the tool can be used in conjunction with any machine learning library.
- https://azure.microsoft.com/
- https://mlflow.org/

Tools and Technologies for Machine Learning

- **Neo4j**
- Neo4j is a native graph platform that handles big data through the graphical relationship between the data. That is, it manages not only data, but also concentrates on their relationship graphically. The relationships between the data entities are stored as connections and presented visually. At the time of any query, these connections are used. In various graph related problem such as social networking, scheduling, and path finding applications, this tool is useful.
- https://neo4j.com/

Tools and Technologies for Machine Learning

- **OpenAIGym**
- The OpenAIGym is a consortium of techniques that helps in implementation of reinforcement learning based applications via teaching agents. As it supports the reinforcement learning, it can be effectively used for games and robotic systems development. Besides these applications, it can be also used in various domains such as business management, energy, education, finance, and transportation, etc.
- https://gym.openai.com/

Tools and Technologies for Machine Learning

- **RapidMiner**
- RapidMiner helps in all the phases of machine learning practices such as setting databases, pre-processing data, applying machine learning algorithms, and visualization of data. It is used by non-programmers and researchers from various domains for data science and big data related machine learning based projects. Rapidminer's graphical interface makes it easy for non-computer professionals and researchers to develop machine learning and data science based applications.
- https://rapidminer.com/

Tools and Technologies for Machine Learning

- **SimpleCV**
- SimpleCV is a open source python based framework. This tool helps in building models that deal computer vision. Here, input images or videos can come from variety of devices and interfaces.
- **StanfordNLP**
- StanfordNLP is a package that helps in working with natural language. StanfordNLP, as per the claim of the developers, support more than 70 languages.
- http://simplecv.org/
- https://stanfordnlp.github.io/stanfordnlp/

Tools and Technologies for Machine Learning

- **TensorFlow and TensorFlowLite**
- TensorFlow is an open source library that can be used as an efficient tool for various machine learning applications through a browser. This tool is also able to import readily available and trained models for use or retrain with new data.
- TensorFlowLite is a collection of utilities that helps deployment and execution of various TensorFlow models on mobile. Besides mobile application, this tool is also used for embedded systems and IoT/IoE devices.
- https://www.tensorflow.org/
- https://www.tensorflow.org/lite

Tools and Technologies for Machine Learning

- **Uber Ludwig**
- Uber Ludwig is a code free toolbox for deep learning applications. It is a toolbox built on the TensorFlow to build, train, and test deep learning models. One major advantage of the tool is that it avoids programming and offers code free development.
-
- **UnityMLAgents**
- The Unity Machine Learning Agent is also used for development of gaming applications through reinforcement learning and interactive intelligent agents. This is also a python based toolkit. It is to be noted that the toolkit helps in converting a utility scene into a learning environment to train the model or a character. It takes learning algorithm support from the python.
- https://eng.uber.com/introducing-ludwig/
- https://docs.unity3d.com/Packages/com.unity.ml-agents@2.0/

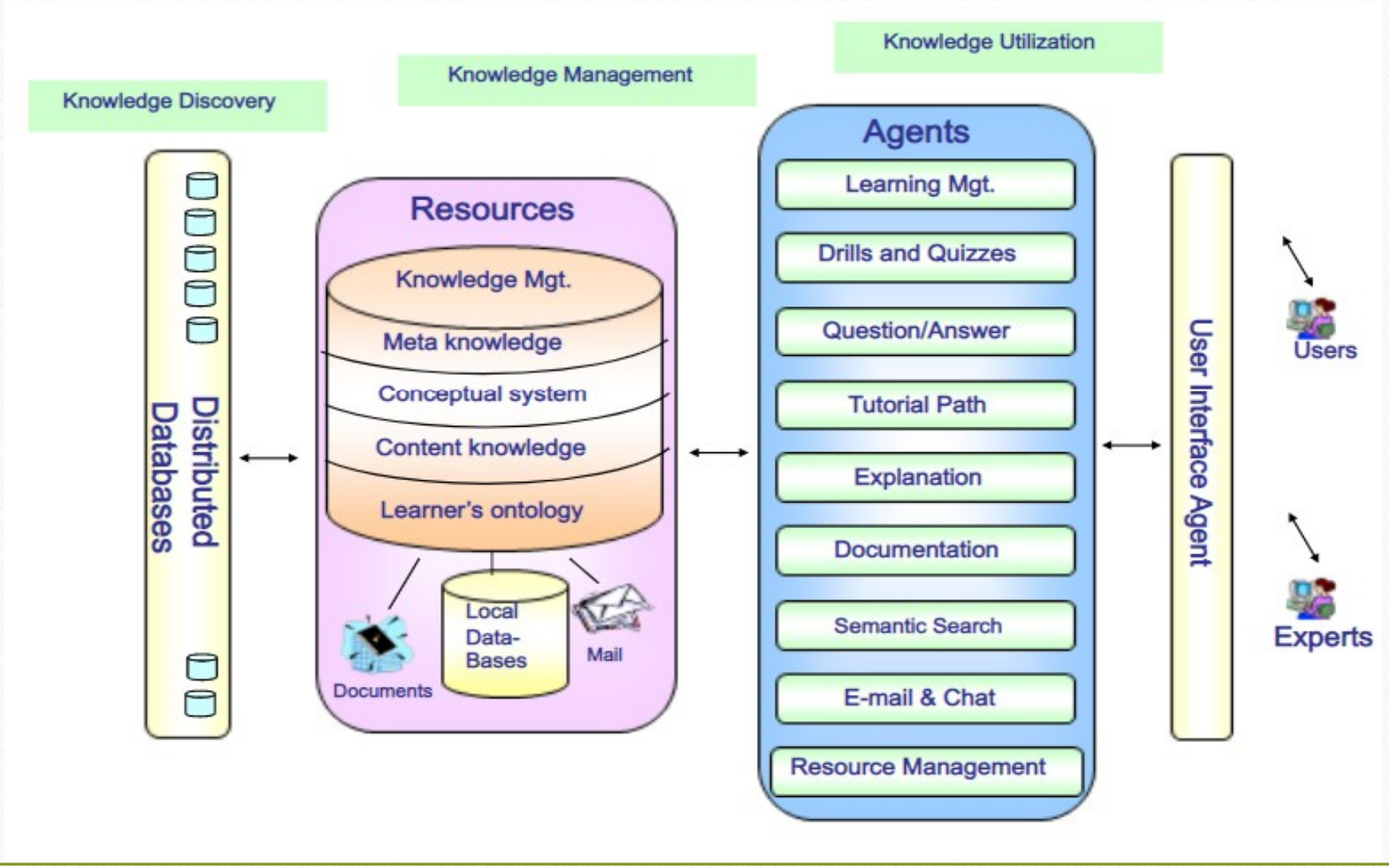
Tools and Technologies for Machine Learning

- **WEKA**
- WEKA is an abbreviation of 'Waikato Environment for Knowledge Analysis'. This is an open source java based tool that supports various machine learning algorithms to perform data preparation, classification, clustering, mining, exploration and visualization via efficient graphical interfaces for its various functionalities. Weka is also used for deep learning applications. Weka is extensively used in the industry, academic and research domains.
- https://www.cs.waikato.ac.nz/ml/weka/

Applications of Machine Learning

- Image Recognition
- Speech Recognition
- Traffic prediction
- Product recommendations
- Self-driving cars
- Email Spam and Malware Filtering
- Virtual Personal Assistant
- Online Fraud Detection
- Stock Market trading
- Medical Diagnosis
- Automatic Language Translation
- Etc.

Application 1: Customized and Dynamic Material Presentation



Application 2: To Issue Credit Card or Not: A Case of Single Perceptron

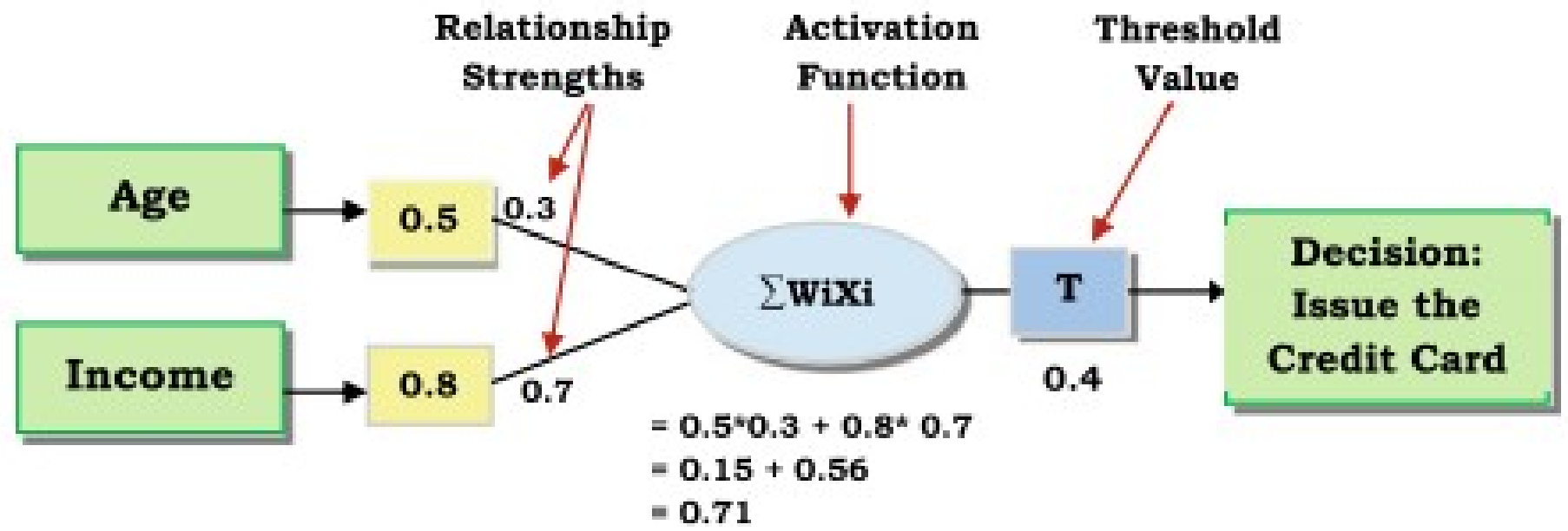
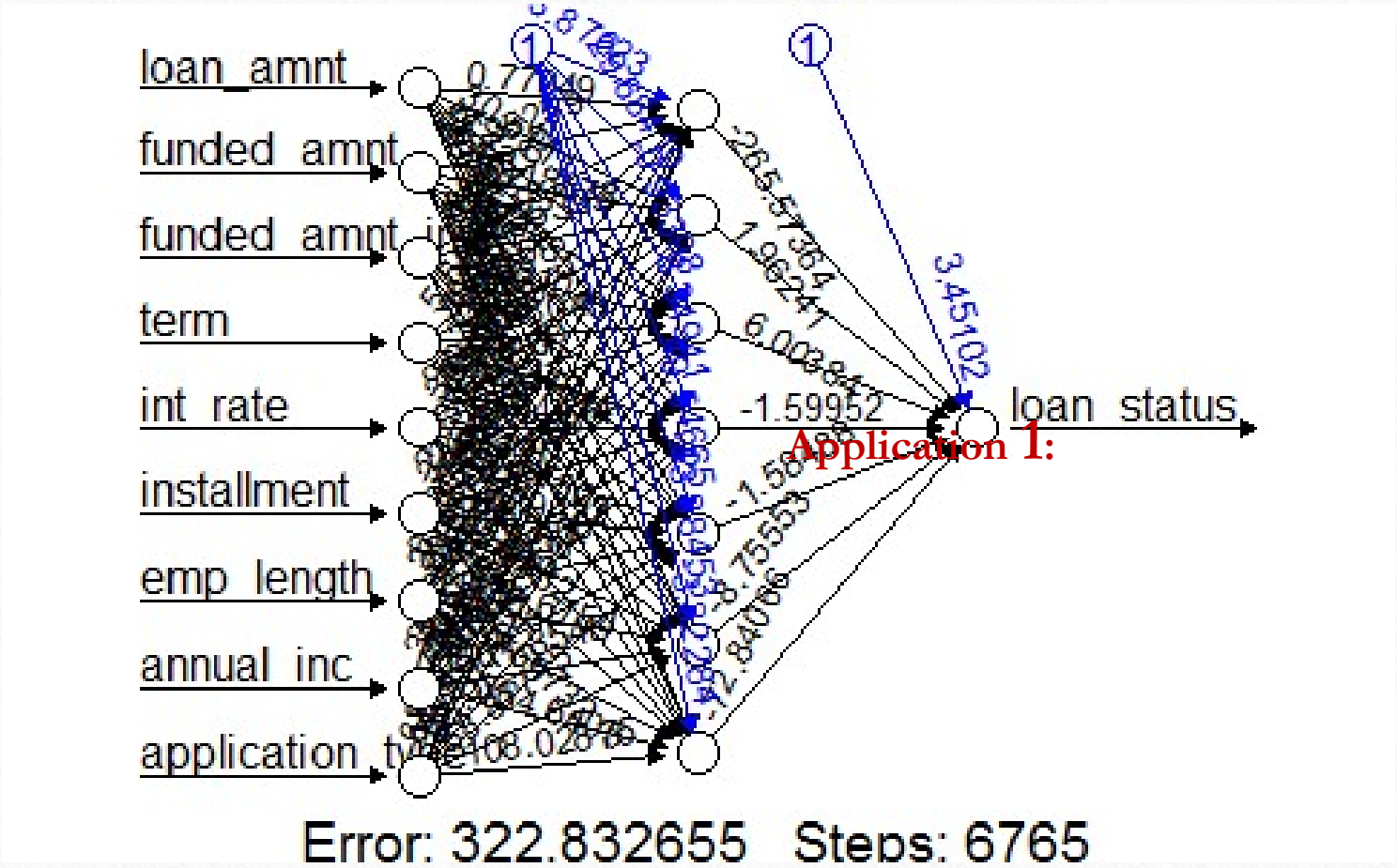


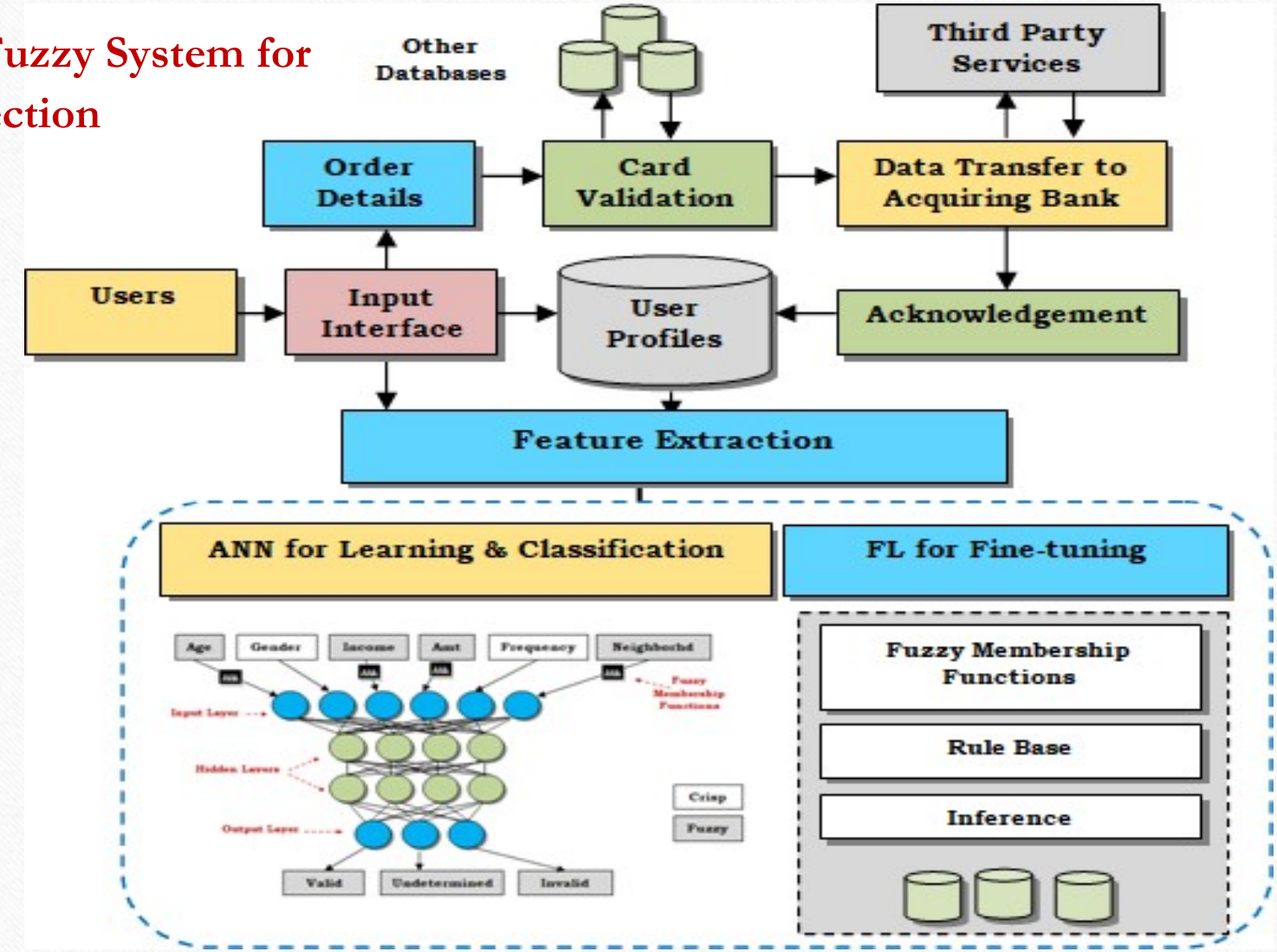
Fig. 4.1 To issue a credit card or not

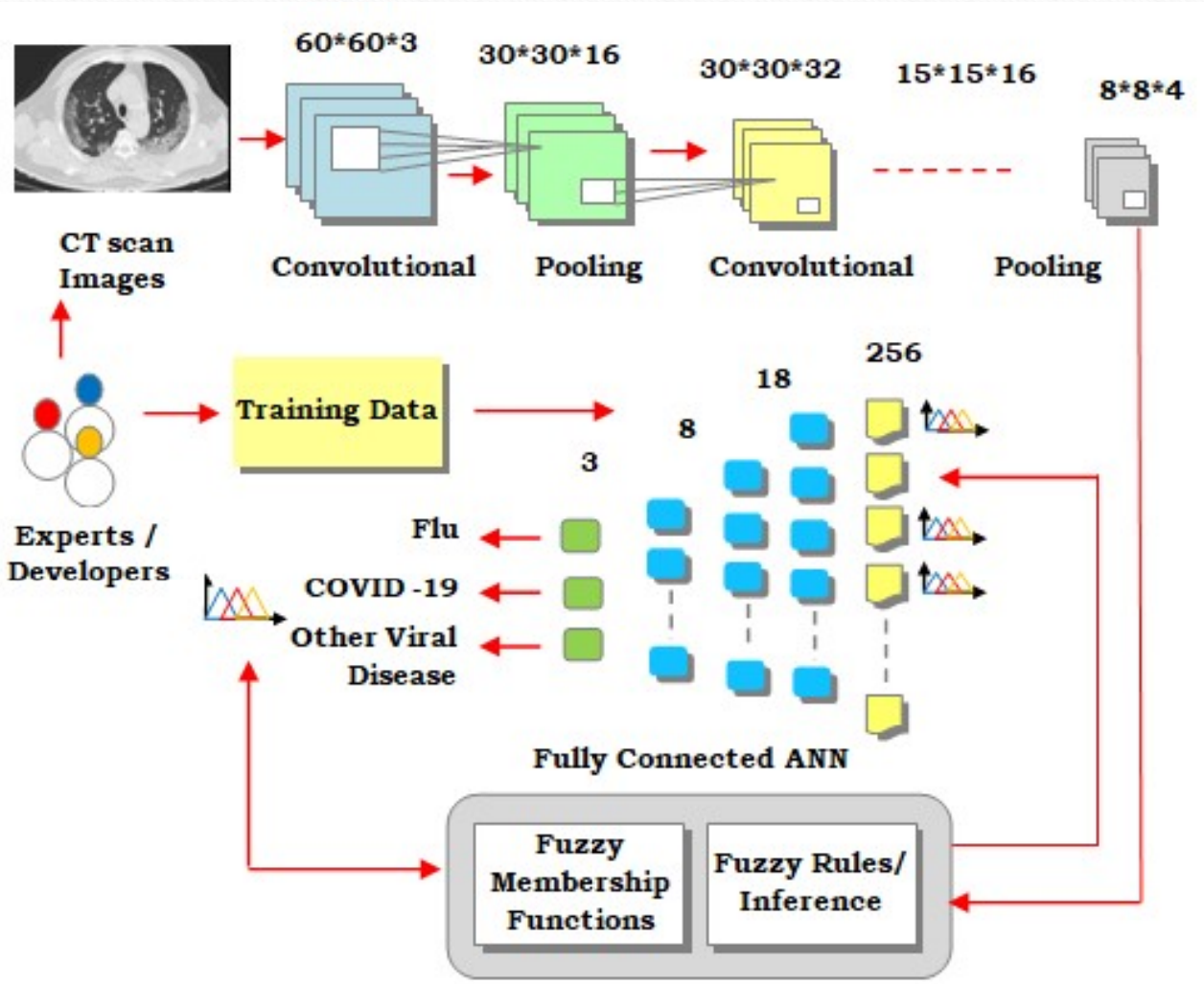
Application 3: To Check/Approve Loan Status



https://www.datasciencecentral.com/profiles/blogs/credit-risk-prediction-using-artificial-neural-network-algorithm

Application 4: Neuro-Fuzzy System for Credit Card Fraud Detection





Application 5:

Fuzzy Convolutional Neural Network to Classify CT-scan Images For COVID -19 Diagnosis

(https://www.springer.com/gp/book/9789811595882)

References

- *Shutterstocks.com, Activegarage.com*
- *Microsoft.com, Kingsoftstore.com,*
- *Amazinganimations.com*
- *Presentationpro.com, Presentermedia.com,*
- *Presentationmagazine.com,*
- *illustrationof.com. Hitton.com,*
- *Clipartof.com*
-
- Sajja, P.S. "Illustrated computational intelligence: Examples and applications", Springer International Publishing, Singapore (Nov'20)
- Sajja, P.S. and Shah, K. "Hybrid intelligent system for credit card fraud detection", SEMCOM Management and Technology Review, vol.8, no.2 (Mar'21)
- *etc.*