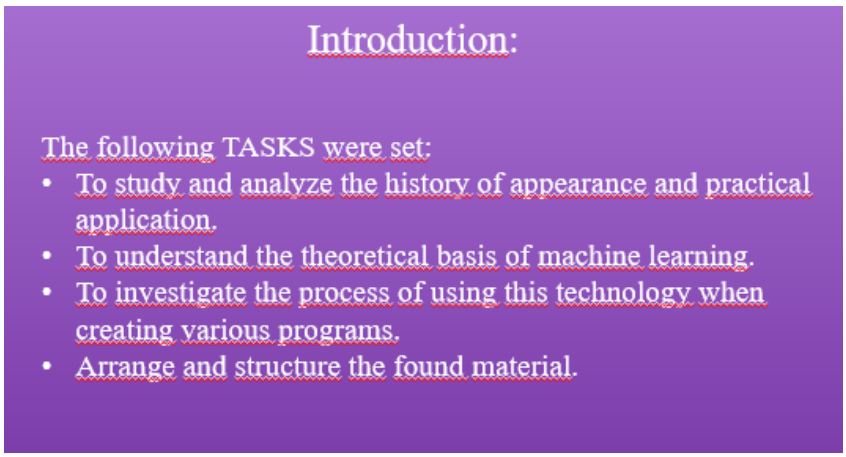
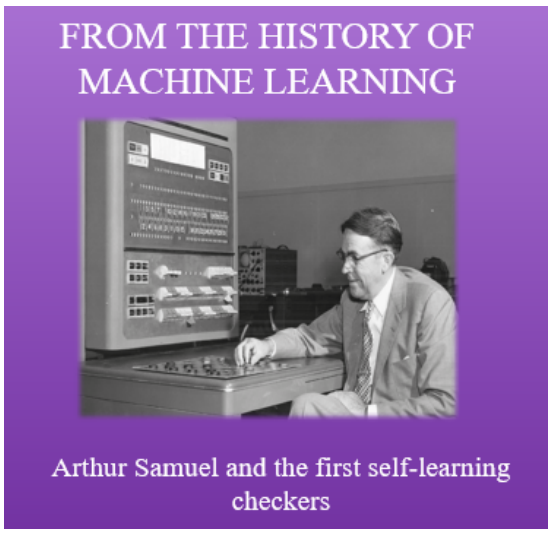
# PYTHON PROGRAMMING: AN INTRODUCTION TO COMPUTER SCIENCE

The following TASKS were set:

- To study and analyze the history of appearance and practical application.
- To understand the theoretical basis of machine learning.
- To investigate the process of using this technology when creating various programs.
- Arrange and structure the found material.



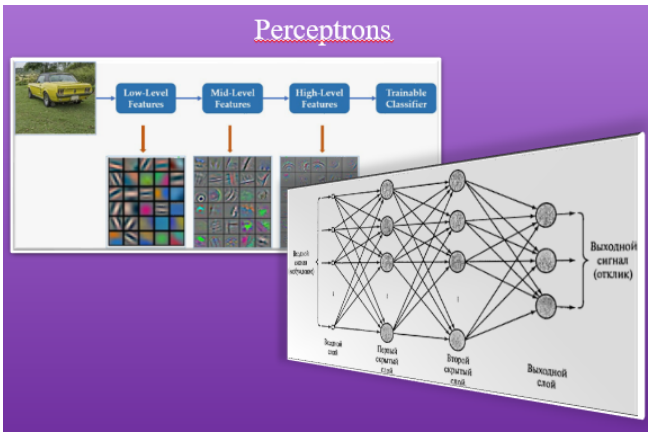

Machine learning (ML) is a field of computer science by which computer systems can make sense of data in much the same way that humans do.Simply put, ML is a type of artificial intelligence that extracts patterns from raw data using an algorithm or method. The main purpose of ML is to allow computer systems to learn from their own experience without explicit programming or human intervention. **Arthur Samuel** first came up with the phrase "Machine Learning" in 1952. In 1957, Frank Rosenblatt – at the Cornell Aeronautical Laboratory – combined Donald Hebb's model of brain cell interaction with Arthur Samuel's Machine Learning efforts and created the perceptron.



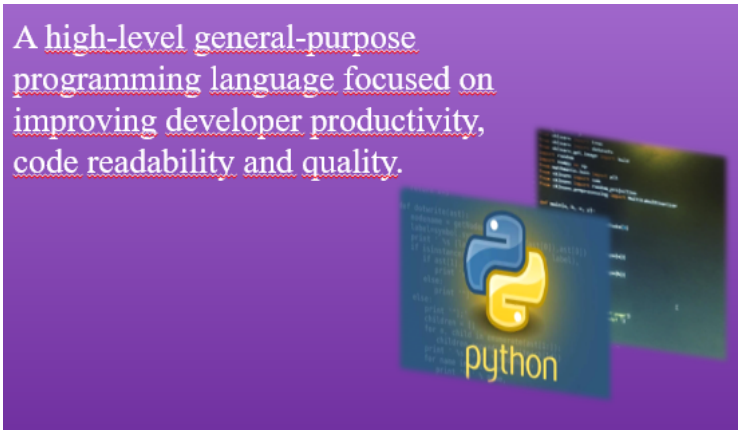

**Arthur Samuel and the first self-learning checkers**

This document has been produced with the support of the EUROPEAN COMMISSION under the ERASMUS+ Programme, KA2 – Capacity Building in the Field of Higher Education: 598092-EPP-1-2018-1-BG-EPPKA2-CBHE-SP.
It reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Warren McCulloch, together with Walter Pitts, created a possible model of a mathematical neuron in 1943. Also in 1958, Frank Rosenblatt created a program based on a neuron, and then a physical device - a perceptron (perceptron). In machine learning, the **perceptron** is an algorithm for supervised learning of binary classifiers. A binary classifier is a function which can decide whether or not an input, represented by a vector of numbers, belongs to some specific class.[1] It is a type of linear classifier, i.e. a classification algorithm that makes its predictions based on a linear predictor function combining a set of weights with the feature vector.




Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.



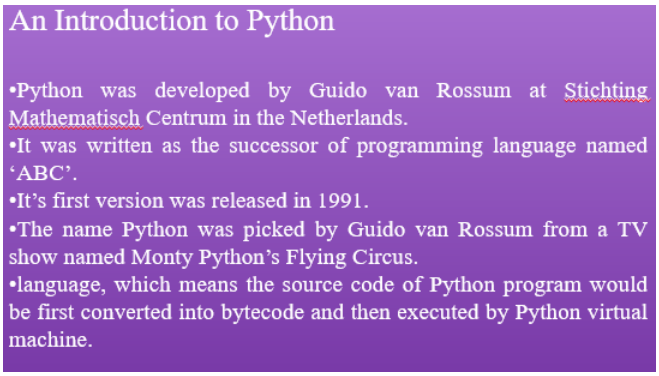

This document has been produced with the support of the EUROPEAN COMMISSION under the ERASMUS+ Programme, KA2 – Capacity Building in the Field of Higher Education: 598092-EPP-1-2018-1-BG-EPPKA2-CBHE-SP.
It reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

**A high-level general-purpose programming language focused on improving developer productivity, code readability and quality.**

An Introduction to Python

Python is a popular object-oriented programing language having the capabilities of high-level programming language. Its easy to learn syntax and portability capability makes it popular these days. The followings facts gives us the introduction to Python −

- Python was developed by Guido van Rossum at Stichting Mathematisch Centrum in the Netherlands.
- It was written as the successor of programming language named 'ABC'.
- It's first version was released in 1991.
- The name Python was picked by Guido van Rossum from a TV show named Monty Python's Flying Circus.
- It is an open source programming language which means that we can freely download it and use it to develop programs. It can be downloaded from www.python.org.
- Python programming language is having the features of Java and C both. It is having the elegant 'C' code and on the other hand, it is having classes and objects like Java for object-oriented programming.
- It is an interpreted language, which means the source code of Python program would be first converted into bytecode and then executed by Python virtual machine.

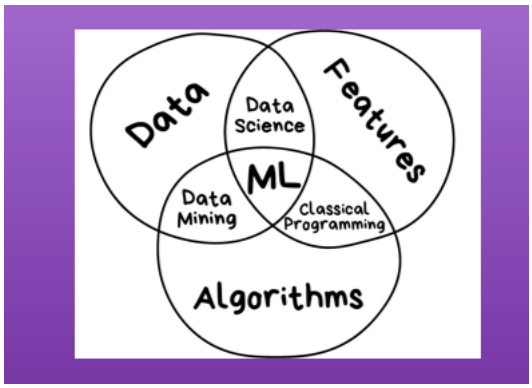



**Supervised Learning**

Supervised learning algorithms or methods are the most commonly used ML algorithms. This method or learning algorithm take the data sample i.e. the training data and its associated output i.e. labels or responses with each data samples during the training process.

The main objective of supervised learning algorithms is to learn an association between input data samples and corresponding outputs after performing multiple training data instances.

This document has been produced with the support of the EUROPEAN COMMISSION under the ERASMUS+ Programme, KA2 – Capacity Building in the Field of Higher Education: 598092-EPP-1-2018-1-BG-EPPKA2-CBHE-SP.
It reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.




Based on the ML tasks, supervised learning algorithms can be divided into following two broad classes −

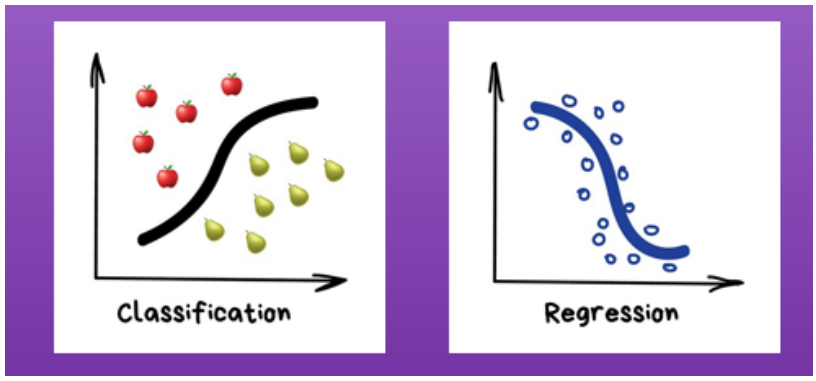
- Classification
- Regression

**Classification**

The key objective of classification-based tasks is to predict categorial output labels or responses for the given input data. The output will be based on what the model has learned in training phase. As we know that the categorial output responses means unordered and discrete values, hence each output response will belong to a specific class or category. We will discuss Classification and associated algorithms in detail in the upcoming chapters also.

**Regression**

The key objective of regression-based tasks is to predict output labels or responses which are continues numeric values, for the given input data. The output will be based on what the model has learned in its training phase. Basically, regression models use the input data features (independent variables) and their corresponding continuous numeric output values (dependent or outcome variables) to learn specific association between inputs and corresponding outputs. We will discuss regression and associated algorithms in detail in further chapters also.




**Based on the ML tasks, supervised learning algorithms can be divided into following two broad classes −**

- **Classification**
- **Regression**

This document has been produced with the support of the EUROPEAN COMMISSION under the ERASMUS+ Programme, KA2 – Capacity Building in the Field of Higher Education: 598092-EPP-1-2018-1-BG-EPPKA2-CBHE-SP.
It reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Based on the ML tasks, unsupervised learning algorithms can be divided into following broad classes −

- Clustering
- Association
- Dimensionality Reduction

**Clustering**

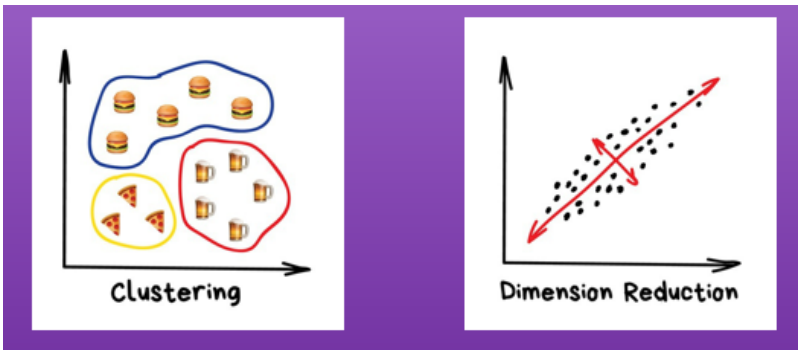
Clustering methods are one of the most useful unsupervised ML methods. These algorithms used to find similarity as well as relationship patterns among data samples and then cluster those samples into groups having similarity based on features. The real-world example of clustering is to group the customers by their purchasing behavior.

**Association**

Another useful unsupervised ML method is **Association** which is used to analyze large dataset to find patterns which further represents the interesting relationships between various items. It is also termed as **Association Rule Mining** or **Market basket analysis** which is mainly used to analyze customer shopping patterns.

**Dimensionality Reduction**

This unsupervised ML method is used to reduce the number of feature variables for each data sample by selecting set of principal or representative features. A question arises here is that why we need to reduce the dimensionality? The reason behind is the problem of feature space complexity which arises when we start analyzing and extracting millions of features from data samples. This problem generally refers to "curse of dimensionality". PCA (Principal Component Analysis), K-nearest neighbors and discriminant analysis are some of the popular algorithms for this purpose.




**Based on the ML tasks, unsupervised learning algorithms can be divided into following broad classes −**

- **Clustering**
- **Association**
- **Dimensionality Reduction**

**Anomaly Detection**

This unsupervised ML method is used to find out the occurrences of rare events or observations that generally do not occur. By using the learned knowledge, anomaly detection methods would be able to differentiate between anomalous or a normal data point. Some of the unsupervised algorithms like clustering, KNN can detect anomalies based on the data and its features.

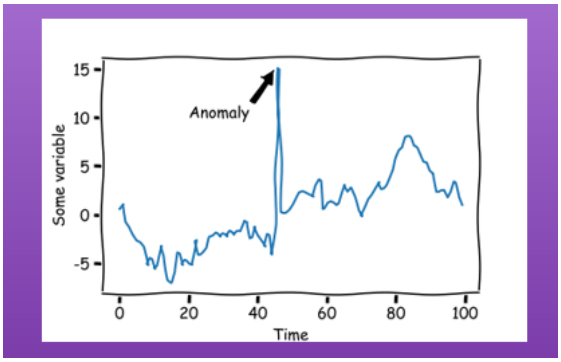
Semi-supervised Learning

Such kind of algorithms or methods are neither fully supervised nor fully unsupervised. They basically fall between the two i.e. supervised and unsupervised learning methods. These kinds of algorithms generally use small supervised learning component i.e. small amount of pre-labeled annotated data and large unsupervised

This document has been produced with the support of the EUROPEAN COMMISSION under the ERASMUS+ Programme, KA2 – Capacity Building in the Field of Higher Education: 598092-EPP-1-2018-1-BG-EPPKA2-CBHE-SP.
It reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

learning component i.e. lots of unlabeled data for training. We can follow any of the following approaches for implementing semi-supervised learning methods −

- The first and simple approach is to build the supervised model based on small amount of labeled and annotated data and then build the unsupervised model by applying the same to the large amounts of unlabeled data to get more labeled samples. Now, train the model on them and repeat the process.
- ,p>The second approach needs some extra efforts. In this approach, we can first use the unsupervised methods to cluster similar data samples, annotate these groups and then use a combination of this information to train the model.




**Such kind of algorithms or methods are neither fully supervised nor fully unsupervised.**

Reinforcement Learning

These methods are different from previously studied methods and very rarely used also. In this kind of learning algorithms, there would be an agent that we want to train over a period of time so that it can interact with a specific environment. The agent will follow a set of strategies for interacting with the environment and then after observing the environment it will take actions regards the current state of the environment. The following are the main steps of reinforcement learning methods −

- Step1 − First, we need to prepare an agent with some initial set of strategies.
- Step2 − Then observe the environment and its current state.
- Step3 − Next, select the optimal policy regards the current state of the environment and perform important action.
- Step4 − Now, the agent can get corresponding reward or penalty as per accordance with the action taken by it in previous step.
- Step5 − Now, we can update the strategies if it is required so.
- Step6 − At last, repeat steps 2-5 until the agent got to learn and adopt the optimal policies.

This document has been produced with the support of the EUROPEAN COMMISSION under the ERASMUS+ Programme, KA2 – Capacity Building in the Field of Higher Education: 598092-EPP-1-2018-1-BG-EPPKA2-CBHE-SP.
It reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

## Reinforcement Learning

- **Step1** – First, we need to prepare an agent with some initial set of strategies.
- **Step2** – Then observe the environment and its current state.
- **Step3** – Next, select the optimal policy regards the current state of the environment and perform important action.
- **Step4** – Now, the agent can get corresponding reward or penalty as per accordance with the action taken by it in previous step.
- **Step5** – Now, we can update the strategies if it is required so.
- **Step6** – At last, repeat steps 2-5 until the agent got to learn and adopt the optimal policies.

## References

https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_quick_guide.htm

https://towardsdatascience.com/anomaly-detection-in-python-part-2-multivariate-unsupervised-methods-and-code-b311a63f298b

https://www.analyticsvidhya.com/blog/2021/06/univariate-anomaly-detection-a-walkthrough-in-python/

https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_quick_guide.htm
https://towardsdatascience.com/anomaly-detection-in-python-part-2-multivariate-unsupervised-methods-and-code-b311a63f298b
https://www.analyticsvidhya.com/blog/2021/06/univariate-anomaly-detection-a-walkthrough-in-python/

This document has been produced with the support of the EUROPEAN COMMISSION under the ERASMUS+ Programme, KA2 – Capacity Building in the Field of Higher Education: 598092-EPP-1-2018-1-BG-EPPKA2-CBHE-SP.
It reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.