

## **REPORT ON INTRODUCTION TO MACHINE LEARNING**

Machine learning is the extracting of knowledge from data. The computer performs intelligent tasks. Machine learning came because the handcoded rules of “if” and “else” failed for some action such as face detection. In handcoded rules changing of task even slightly needs a rewrite of the whole system also designing rules requires a deep understanding of how a decision should be made by a human expert.

Machine learning can solve different problems can either Supervised learning or unsupervised learning .Supervised learning user provides the algorithm with a pair of input and desired output. For example of supervised learning are Identifying the zip code from handwritten digits on an envelope, Determining whether the tumor is benign based on a medical image. For unsupervised learning here the input is known and no known output data and is not given to the algorithm. For example Identifying topics in a set of blog posts, Segmenting customer into groups with similar preferences and Detecting abnormal access patterns to a website.

Knowing the task and the data to use is the important thing in machine learning, it will not be effective to choose an algorithm and to throw your data at it. It is important to know what is going in your dataset before you begin building a model. Each algorithm is different in terms of what kind of data and what problem settling it works best for.

Why do we study python and why is it used in machine learning ? Python is seemed to be simple to learn compared to other languages and combines the power of general purpose programming language with the ease of use of domain specific scripting language like MATLAB or R. Python has got libraries for data loading, visualization, statistics, natural language processing and image processing. It interacts directly with the code using terminal or other tools like jupyter notebook. The libraries in python they different activities such as matplotlib helps in data visualization, numpy is used for mathematical and logical operation, pandas is used for data modeling and data analysis, Scipy is used for scientific computation and provides other functionality mathematical optimization, signal processing, statistical distribution.

Scikit-learn is an open source project tool for data mining and data analysis, it is a popular tool and most prominent python library for machine learning. It depends on two other libraries the numpy and Scipy, it is widely used in industry and academic.

There are two version of Python Python2 and Python3. This version have got some differences python2 is the older version and python3 is the later version. When writing any new code ,it is for the most part quite easy to write code that runs under python2 and python3 . If you don't have if you don't have to interface with legacy software, you should definitely use python3.

When dealing with machine learning problem firstly you should understand the problem and choose the right data. The next step is to choose the model of the type of problem can either be k-nearest neighbor, , After that is measuring the success by training and testing data ,after we make prediction after then we do the evaluating of the model.