***T/UDOM/2017/02680***

**INTRODUCTION TO MACHINE LEARNING**

Machine learning is about extracting knowledge from data. The computer perfoms intelligent tasks. We can say the computer learns. In the early days of “intelligent” applications, many systems used handcoded rules of “if” and “else” decisions to process data or adjust to user input. The handcoded rules have got some disadvantages. 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.

The most successful kinds of machine learning algorithms are those that automate decision-making processes by generalizing from known examples. 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. In 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. Many people spend a lot of time building complex

machine learning solutions, only to find out they don’t solve the right problem.

Python is 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. 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.

We have Scikit-learn. Scikit-learn is an open source project tool used in machine learning , It depends on two other libraries the numpy and Scipy, it is widely used in industry and academic.

Python has got two versions which are Python2 and Python3. These versions 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.