TUMAINI STEVEN

T/UDOM/2017/11329

**1. Machine learning**

It is a research field at the intersection of statistics, artificial intelligence, and computer science. Is about extracting knowledge from data

**Why machine learning**

in early days of "intelligent" application, many systems use hand coded rules of "if" and "else" decision to process data or adjust input but there are many problems that are difficult to be solved by hand coded rules. Machine learning provides the flexibility of obtaining the output from a given data even if the task is changed unlike in hand coded if the task changes even slightly it requires a rewrite of the whole system.

Disadvantages of using hand coded

* The logic required to make a decision is specific to a single domain and task.
* Designing rules require a deep understanding of how a decision should be made by a human expert

**Problems machine learning can solve**

Machine learning problems can be categorized into two

* **Supervised learning** - automate decision-making processes by generalizing from known examples. learning by pairs of input/outputs.
* **Unsupervised learning -** only the input data is known, and no known output data is given to the algorithm.

**2.** **Why Python**

Python is used the mostly in machine learning because It combines the power of general-purpose programming languages with the ease of use of domain-specific scripting languages like MATLAB or R. Python has libraries for data loading, visualization, statistics, natural language processing, image processing, and more. This vast toolbox provides data scientists with a large array of general- and special-purpose functionality.

one of the main advantages of Python is the ability to interact directly with the code, using terminal or other tools like jupyter Notebook

**Scikit-learn**

is an open source project, meaning that is free to use and distribute, and anyone can obtain the source code to see what is going on behind the scene

**3.** **Essential libraries and tools**

* **numpy:** packages for scientific computing
* **jupyter notebook**: interaction environment for running code in the browser
* **scipy:** collection of functions for scientific computing in python
* **matplotlib:** the primary scientific plotting library in python with functions like (charts visualization, histograms, scatter plots).
* **pandas:** python library for data wrangling and analysis