**KHALIDI HASSAN**

**BSC. COMPUTER SCIENCE**

**UNIVERSITY OF DODOMA**

**Introduction to machine learning: A reading report**

**Brief content:**

-why machine learning

-why python

-essential libraries and tools

**1. Why machine learning**

To come up with efficient and durable solutions for programming problems that used to be solved by hand coding approach or had not any solutions at all.

Example:

-Decision making with if---else structures had a number of disadvantages like the inflexibility incase of task changes or deep human understanding requirement.

-face recognition in digital images used to be unsolvable

* **Problems machine learning can solve**
* **Supervised learning:** learning by pairs of input/outputs

Example:-identifying zip code from handwritten digits on an envelope

-determining whether a tumor is benign based on a medical image

-detecting fraudulent activity in credit card transactions

* **Unsupervised** algorithms: input data is known but no output is provided

Example:-identifying topics in a set of blog posts

-segmenting customers into groups with similar preferences

-detecting abnormal access patterns to a website

* Data should have a specific representation especially in form of a table consisting of columns for each data point and rows called samples.

**Feature extraction and feature engineering** concepts deal with good representation of the data

**2. Why python**

* a lot of related libraries
* ease of use
* general-purpose and powerful language
* direct interaction with the code using terminal or Jupiter notebooks
* provides quick iteration and easy interaction between data and analysis processes
* GUI and web services

Scikit-learn: the most used and open source tool for machine learning algorithms libraries

**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

**Reference**

Machine Learning with Python a guide for data scientists: Andreas C. Müller & Sarah Guido (oreilly)