**Names: Monei Bakang Motshegwe**

**Email: cse20-018**

**Phone Number: 72711192/ 73065111/ 71851394**

**LEARNING TECHNIQUES**

**INTRODUCTION TO MACHINE LEARNING**

My name is Monei Bakang Motshegwe, a Computer Systems Engineering student at Botswana Accountancy College, I was tasked to conduct a research on about LEARNING TECHNIQUES.

Given the evolution and growth of Machine Learning, as well as the ever-increasing volumes/amounts of data/information, there is reason to believe that smart data analysis will become more prominent as a critical component of technological innovation. Machine learning is an information-driven method, wherein machines learn from data without requiring human intervention. Several areas benefit from mind-boggling machine learning programs. Machine learning allows systems to learn without being expressly designed to do the latter, culminating in more intelligent behavior. It generates information forecasts by constructing a model that discover patterns in past data and utilize those patterns to generate predictions. The general architecture of machine learning consists of several steps, including business understanding (domain comprehension and awareness), data capture and comprehension (trying to collect and interpretation of data), modeling (which includes feature engineering, model training, and evaluation), and mobilization (deploy the model on the cloud).

**UNSUPERVISED LEARNING**

Unsupervised learning algorithms detect trends in data sets that lack classified or labeled data pieces. Consequently, the algorithms can categorize, label, and organize data points inside data sets without any outside aid. The model does not require users to supervise it. In other words, the learning process is not labeled, leaving it to identify structure in the data set on its own. In unsupervised learning, an AI system will categorize unsorted content based on similarities and contrasts even if no categories are given. The Unsupervised learning's primary purpose is to identify hidden and surprising patterns.

**TYPES OF UNSUPERVISED TASKS**

**AUTO ENCODERS** – Auto encoders are an unsupervised learning strategy that knows and understands representations by utilizing neural network capabilities. Auto encoders are a type of feedforward neural network with the same input and output. They reduce the input to a lower-dimensional representation and then reconstruct the result from this representation. An autoencoder consists of three components: encoder, code, and decoder. The encoder compresses the input and creates the code, which is then used by the decoder to reconstruct the input. To link the output to the objective, an auto encoder requires three components: an encoding methodology, a decoding method, and a loss function. At their essence, autoencoders are dimensionality reduction (or compression) algorithms.

**ANOMALLY-DEETECTION -** Any process that discovers outliers in a dataset is referred to as anomaly detection. These anomalies might be caused by unexpected network traffic, a defective sensor, or data that has to be cleaned before analysis. An anomaly occurs when data patterns go beyond or depart from usual patterns. An irregular network traffic pattern, for example, suggests that the hacked system is sending sensitive data to an unauthorized location. Anomalies are discovered by recognizing and anticipating data points that deviate from the normal trend. Anomaly detection may be used for.

**ASSOCIATION** - The association rule is a form of unsupervised learning, wherein the algorithms tries to study and understand alone without the assistance of a teacher, so because data is not labeled. The association rule is a diagnostic, not forecasting, strategy that is extensively used to uncover interesting relationships buried in massive datasets. To determine the link, rules or frequent item sets are typically utilized. The extraction of association rules is used to discover new and fascinating links between items in a collection, a frequent trend in data information, or any database system. Market Basket Analysis (deciding which goods are bought often), Consumer Clustering in Retail (deciding which shops customers like to visit together), Price Bundling, Variety Choices, Cross-Selling, as well as other applications utilize them regularly.

**CLUSTEERING**: is a crucial concept in unsupervised learning. Its primary goal is to find a structure or pattern in a set of data that has no belongings(category). Without Supervision, Learning. Clustering algorithms will hunt for natural clusters (groupings) in your data. You may also tell your methods how several clusters they should find. The precision of these categories can be changed. Clustering is the most important unsupervised issue since it focuses with recognizing structure in a collection of unlabeled data. It is defined as "the technique of classifying products into categories whose members are in a certain sense similarity." Thus, a cluster is a group of items that are "similar" among themselves but "dissimilar" to

**CATEGORIES OF CLUSTERING**

Clustering may be divided into numerous forms, including partitioning, hierarchical, overlapping, and probabilistic. Partitioning groups data so that each data point can just relate to one cluster. It is sometimes referred to as exclusive clustering. Partitioning clustering is demonstrated via K-means. Every piece of data is a cluster in hierarchical clustering. The number of clusters is reduced through repeated unions between the two adjacent clusters. Data is clustered using overlapping fuzzy sets. Each point may be a member of two or more clusters with varying degrees of membership. Data will be paired with a suitable membership value, such as Fuzzy C-Means, in this case. Lastly, the probabilistic creates clusters using probability distribution.

**ALGORITHMS AND APPROACHES IN UNSUPERVISED LEARNING**

**K-MEANS CLUSTERING ALGORITHM** - -K-Means Clustering is an Unsupervised Learning technique. The unlabeled dataset is divided into multiple clusters, with "K" representing the number of pre-defined groupings. K can be any random number, such as 2, which produces two clusters, or 5, which produces five clusters. It is a strategy that splits an unlabeled dataset into K clusters by repeating it. Each dataset is a member of just one group of datasets with comparable characteristics. It allows data to be classified into numerous groups.

Algorithm or pseudocode for k-means clustering:

1. Enter the num 'K' to choose the number of choice for the clusters.
2. Select a random K points
3. To build predetermined clusters, assign all data points to their nearest centroid.
4. Determine a new variance and generate a new centroid
5. Reassign each data point to the centroid of the most recent cluster.
6. End

PCA – Principal Component Analyst

**PCA** is a technique for reducing the dimensionality of a data set with numerous interconnected variables. This is accomplished by changing the key characteristics to a new collection of variables (PCs). In general, PCA computes the eigenvalue decomposition of a data estimate's covariance matrix and utilizes the feature representation is projected to a lesser space using relevant eigenvectors. In plenty of other words, the PCA theory's physical interpretation would be that it projects to a subspace that preserves the bulk of the data's volatility. PCA is used in both exploration predictive modeling and data analysis. It is a technique for detecting hidden patterns in a dataset by minimizing variances. It makes use of a feature extraction approach. Regardless of the fact that it is widely used in many other fields, it is a statistical method. Analysts have made substantial contributions to its progress. It entails conserving the variability in order to discover fresh variables that are linear functions of the original dataset's data points. Typically, PCA attempts to represent the lower-dimensional surface in order to project the high-dimensional data. The variance of each feature is determined through PCA. The characteristic with a large variance demonstrates an excellent separation across classes and hence minimizes dimensionality.

Pseudocode for PCA

1. Arrange the data in a logical order.
2. Complete the regularization.
3. Determine Z's Covariance
4. Calculate the Eigen-Vectors and Eigen-Values.
5. Sort the Eigen-Vectors that have been computed.
6. Evaluate the main components or additional features
7. Remove irrelevant characteristics first from given data.

**ADVANTAGES OF UNSUPERVISED TECHNIQUE**

-In compared to supervised learning, there is less intricacy. Unlike supervised algorithms, no individual is obligated to interpret and then label the data inputs in unsupervised learning. This reduces the complexity of unsupervised learning and illustrates why several people favor unsupervised approaches.

-It tends to happen in real time, so that all input data may be examined and categorized while learners are present. This enables them to comprehend various models of learning and raw data sorting.

-Unlabeled data is frequently simpler to get from a computer than labeled data, which requires human interaction. This is also a significant distinction among both supervised and unsupervised learning.

**CONSEQUENCIES OF UNSUPERVISED TECHNIQUE**

-Less accuracy of data.

-The description of the data sorting and output is not entirely exact. This is due to the fact that the data utilized in unsupervised learning is labeled and unknown. The machine's role is to categorize and group the raw data before discovering the hidden patterns.

-The analysis findings cannot be determined.

**SUPERVISED LEARNING TECHNIQUE**

Algorithms for machine learning regard each instance of a dataset as a collection of features. These attributes could be binary, category, or continuous. When the instances are marked, this is known as supervised learning. The model is trained using labeled data and then tested using unlabeled data. Its primary design begins with dataset collecting, followed by dataset splitting into testing and training data, and ultimately data preparation. Features that are extracted are sent into an algorithm, which is then trained to figure out which feature correlates to which label. Finally, the test data is given into the model, and the model predicts the test data by supplying the expected labels.

**APPLICATIONS OF SUPERVISED LEARNING IN REAL WORLD**

-For text sentiment analysis in marketing, a variety of text mining algorithms are utilized (happy, not happy).

-Fraudulent transaction detection in finance and banking.

-As for medical science, for forecasting patient risk (such as high-risk vs low-risk patients) or the likelihood of congestive heart failure.

**BENEFITS OF SUPERVISED TECHNIQUE**

-The findings provided by the supervised approach are more precise and dependable than the results obtained by unsupervised machine learning techniques. This is mostly due to the fact that the supervised algorithm's input data is quite well prominent and labeled.

-It enables you to be quite exact regarding the actual label definition. In other words, you may train the algorithm to discriminate between distinct classes and define an optimal decision limit.

-Because all of the classes employed are widely recognized, the outcomes in the analysis and the output of your method are likely to be established.

-The input data is classified popularly and highly.

**SELECTING THE BEST APPROACH FOR AN ORGANIZATION**

Regardless of the fact that we discussed the pros and downsides of supervised and unsupervised learning, it is not entirely correct to conclude that one way is superior to the other. Keeping this in mind, it is incorrect to suggest that unsupervised and supervised approaches are mutually exclusive. The fundamental objectives and issues that can be solved using supervised and unsupervised approaches differ. When to employ which strategy relies upon your needs and the challenges you need to tackle.

**REFERENCES**

S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classiﬁcation Techniques," Informatica, Vol.31,No.3, pp.249-268, 2007

Hastie, T., Tibshirani, R., & Friedman, J.,"Unsupervised learning," The elements of statistical learning, Springer,pp.485-585, 2009

Cios, K. J., Swiniarski, R. W., Pedrycz, W., & Kurgan, L. A., "Unsupervised learning: association rules," DataMining, Springer, pp.289-306, 2007

Baldi, P., "Autoencoders, unsupervised learning, and deep architectures," Proceedings of ICML workshop onunsupervised and transfer learning, pp.37-49, June 2012

Alloghani, M., Al-Jumeily, D., Mustaﬁna, J., Hussain, A., & Aljaaf, A. J., "A Systematic Review on Supervisedand Unsupervised Machine Learning," Supervised and Unsupervised Learning for Data Science, Vol.3, 2019

Grira, N., Crucianu, M., & Boujemaa, N., "Unsupervised and semi-supervised clustering: a brief survey," A reviewof machine learning techniques for processing multimedia content, Vol.1, pp.9-16, 2004