**DATA MINING REPORT**

**DECISION TREE, RANDOM FOREST & ARTIFICIAL NEURAL NETWORKS**

SHARON ZACHARIA | 22074407

[Link to Dataset](https://archive.ics.uci.edu/ml/datasets/Mushroom) | [GitHub.](https://github.com/SHARONZACHARIA/DataMining_Report.git)

In this report, Artificial Neural Networks and machine learning techniques like Random Forest and Decision Trees are compared. These algorithms are used on the mushroom dataset, which describes samples that are representative of 23 species of gilled mushrooms belonging to the families Agaricus and Lepiota.

**Artiﬁcial Neural Networks (ANNs)** are powerful classiﬁcation models that can learn highly complex and nonlinear decision boundaries purely from the data. Input node and output node are used to represent the input and output values of the network. A perceptron is a simple type of ANN model, and a multi-layer neural network extends the fundamental idea of a perceptron to more intricate node architectures. The weights and bias terms ( w, b ) of the ANN model are learned during training so that the predictions on training instances match the true labels. This is achieved by using a loss function.

**Decision Trees Classifiers and Random Forest : *Decision Trees*** resemble flowcharts, with each internal node conducting a test on an attribute. The branches represent the outcomes of these tests, leading to leaf nodes that hold class labels. They provide a straightforward way to represent decision-making processes based on feature values.

***Random Forests*** are collections of distinct decision trees aimed at enhancing generalization. Random forests introduce diversity by employing dual randomness. At each internal node, they randomly select a subset of features for splitting, and they use different bootstrap samples for training instances. This approach fosters a more varied set of trees, leading to improved performance and robustness.

**COMPARITIVE ANALYSIS OF MODELS**

* Artificial Neural Networks , Decision Tree Classifiers & Random Forest are capable of classification tasks , where input data is assigned to predefined classes.
* Decision Trees and Random Forests are more scalable and efficient compared to ANNs when dealing with large datasets.
* Due to their inherent nature as sets of if-then-else decision rules, decision trees are naturally interpretable and random forests present a challenge in interpretation due to their aggregation of multiple decision trees.
* The complex and nonlinear nature of ANNs make them considerably more challenging to interpret.