**Name : Vishal Kumar Paswan**

**Roll Number : 24MA10068**

**Task-2 [AI/ML]**

**(KodeInKGP : Technology Web3.0 Society)**

* The aim of this experiment is the construction of the suitable decision tree for categorization of the given dataset and provide the status as either "Alive" or "Dead". This was achieved, as mentioned, through the implementations of various decision tree algorithms in the models like "Gini Impurity" model, the "Entropy/Information Gain" model, and the "Random Forest" model and then evaluating their accuracy rates over varied epochs.
* The pre-processing was carried out by splitting the data in an 80:20 ratio of training data and testing data, then fed this data into the above three decision tree algorithms.
* The "Gini Impurity" model, which measures the probability of a randomly chosen element to be wrongly labeled given that they were labeled randomly and independently, reached an accuracy rate of 86.7% in the test set.
* The "Entropy Gain" model, or the "Information Gain" model, is a model that determines which feature to split, while building the tree, in an attempt to maximize overall "information gain". This model reached 86.1% accuracy rate on the test set.
* One of the decision tree models is the "Random Forest," which combines multiple decision trees together, having a single output; for this case, it improves generalization and reduces overfitting. Model Accuracy achieved 91.3% on test set.
* In this process of passing the pre-processed data set to the decision tree algorithms, it was required that the model doesn't overfit the data which it is trained upon. Therefore the "random forest" is one of the most efficient models in this respect since it can quite easily overcome this very phenomenon by taking the average of results from different trees.
* Finally, we may have also been able to have improved further performance on the given models by improving on such cases. This may have had to be done through hyperparameter tuning, such as:

1. adjusting the depth of the trees in the random forest model,
2. adjusting the number of trees in the random forest model, etc.

**Thank You**