**Assignment (Machine Learning Topics – Decision Tree, Naïve Bayes & NLP )**

**Q1** Write a Python program using Scikit-learn to split the iris dataset into 80% train data and 20% test data. Out of total 150 records, the training set will contain 120 records and the test set contains 30 of those records. Train or fit the data into the model and calculate the accuracy of the model using the Decision Tree Algorithm.

**Q2** For the dataset given below calculate Entropy an Information game, required to create a decision tree?

Table

Description automatically generated

**Q3** What are the advantages & Disadvantages of Decision Tree Algorithms?

**Q4** Describe at least one way to overcome the problem of overfitting when constructing decision trees?

**Q5** What are the advantages of Naïve Bayes Classifier?

**Q6** What is the difference between Parametric & Non-Parametric Machine Learning Algorithms?

**Q7** What is count vectorizer & TDF-IDF vectorizer? Explain with Examples?

**Q8** What is semantic analysis? Explain different techniques that are used for Semantic analysis?

**Q9** Using disaster-tweet.csv, Write a code to classify tweets as Relevant or not Relevant using TF-IDF vectorization method.

**Q10** What is meant by Unigram, Bigram, Trigram & Multigram model in NLP. Explain with Examples?