Work Plan for Assignment

ARM 210 : Introduction to Machine Learning

Name : Satyam Kumar Jha

Batch : B.Tech AIML (B2)

Enroll : 20219051622

Dataset provided : Acoustic features extracted from syllables of anuran (frogs) calls, including the family, the genus, and the species labels (multilabel).

AIM : To develop an interpretable and trustworthy predictive model that can classify various anuran species accurately and effectively.

Preface : Anuran is any organism that is member of the order Anura of the Amphibia class (the organism of this class are commonly known as frogs and toads. The identification of frog species by their calls has gained attention for its potential use in ecological research. However, many recorded frog species are believed to be unique, posing a challenge for their classification. Here my project comes to play as it aims to implement different algorithms and test them based on the accuracy.

Data Description

The Anuran Species dataset is the dataset assigned to me. Based on the two tables below, we can notice that the dataset contains 22 columns of features and a single column of the target. There is not a very unique name for all the features, they only differ in terms of the numbering, which represents different animals without clarifying what the actual animal is. The target is the 'Species', in which we are expected to get the outcome of different types of Anuran species (10 different species).

Data Splitting:

The dataset comprising frog call recordings will be divided into two subsets: a training set and a test set. The training set, constituting 80% of the data, will be utilized to train the machine learning models. The remaining 20% will be reserved for evaluating the performance of the trained models.

Model Selection and Parameters:

Three primary machine learning models will be explored for this classification task: Decision Tree, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN).

Decision Tree:

- Decision Tree algorithm works by iteratively partitioning the data based on the features to create a tree-like structure of decisions.
- Various tree depths will be tested to determine the optimal depth for accurate classification.

Support Vector Machine (SVM):

- SVM is a powerful classification algorithm that attempts to find the hyperplane that best separates different classes in the feature space.
- Different kernel functions and regularization parameters will be tested to identify the optimal combination for classification accuracy.
- GridSearchCV will be employed for hyperparameter tuning to enhance model performance.

K-Nearest Neighbors (KNN):

- KNN classifies data points based on the majority class of its nearest neighbors in the feature space.
- The 'k' parameter, representing the number of neighbors to consider, will be varied to optimize classification accuracy.
- Different distance metrics, such as Euclidean distance, will be evaluated to determine the most suitable for the dataset.
- Techniques like cross-validation will be utilized to fine-tune the KNN model for improved performance.

Conclusion:

This preparatory report outlines the initial steps and methodologies to be undertaken in the Frog Call Classification project. By exploring the Decision Tree, SVM, and KNN algorithms, along with appropriate parameter tuning techniques, we aim to develop robust models capable of accurately identifying frog species based on their calls. The subsequent stages of the project will involve data preprocessing, model training, evaluation, and optimization to achieve the desired classification accuracy.