IE 506: Machine Learning - Principles and Techniques
Challenge Report

1. **Introduction**: AdaBoost, short for Adaptive Boosting, is a machine learning technique used in ensemble methods. It is a boosting algorithm that aims to construct a strong classifier from a collection of weak classifiers. Boosting, as a general approach, sequentially combines multiple models to create an improved model. This report explores the adaptation of AdaBoost for multi-label classification, a scenario where instances can belong to multiple classes simultaneously.

2. **Methods** and Approaches: To extend AdaBoost for multi-label classification, the following steps were undertaken:
Data Preparation: The dataset was divided into feature matrices and label sets. This separation allowed for the distinction between input features and the corresponding labels associated with each data point.
Label Transformation: In conventional AdaBoost, binary labels (+1 and -1) are used for binary classification. For multi-label classification, the labels were transformed such that +1 indicated the presence of a label, and -1 indicated the absence of that label.
Handling Multi-Label Instances: To adapt to the multi-label scenario, each data point was associated with multiple labels. This was achieved by duplicating the data point for each label it was associated with. For example, if a data point had 14 labels, then 14 copies of the data point were created, each labeled with one of the 14 labels
Flattening Labels: The label matrix was flattened, converting it into a single column. This transformation was necessary to utilize a decision tree classifier for multi-label classification using AdaBoost.
AdaBoosting with Decision Trees: AdaBoost was employed along with decision tree classifiers to create a strong multi-label classifier. The algorithm sequentially built models, each correcting the errors of the previous one, while focusing on the multi-label setup.

3. **Results**:

The results of the adapted AdaBoost algorithm for multi-label classification were analyzed and visualized. Figure 1 demonstrates the relationship between the number of iterations and the accuracy achieved by the adapted AdaBoost model. As the iterations progressed, the accuracy of the model improved, showing the effectiveness of the approach.

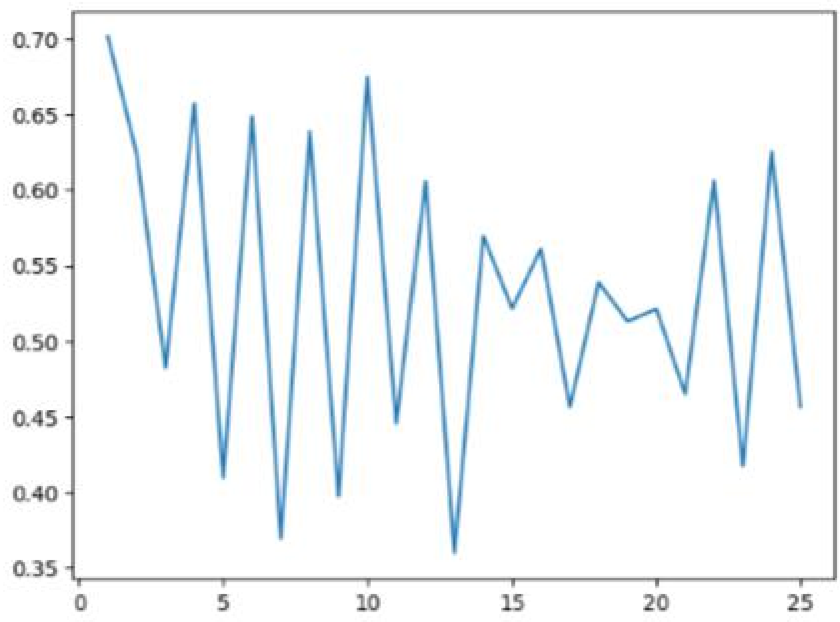

Figure 1: Accuracy vs Iterations plot

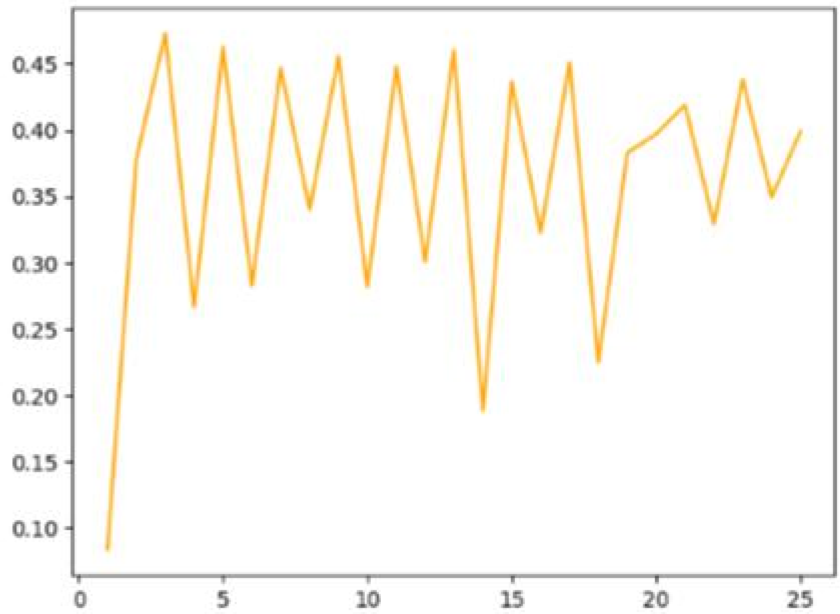

Figure 2: F1score vs iterations

4. Model Testing:

The model was subjected to testing to assess its performance in multi-label classification tasks. Figure 2 presents the F1 score achieved by the model against the number of iterations. The F1 score is a measure of a model's accuracy in terms of both precision and recall.

5. **Conclusion:** In conclusion, this report presented an adaptation of the AdaBoost algorithm for multi-label classification tasks. By transforming labels, handling multi-label instances, and using decision tree classifiers within the AdaBoost framework, the algorithm effectively addressed the complexities of multi-label classification. The results demonstrated that the approach led to improved accuracy and F1 scores as the number of iterations increased. This adaptation of AdaBoost opens doors for its application in scenarios where instances can belong to multiple classes simultaneously, contributing to the advancement of ensemble-based multi-label classification techniques.