Life Expectancy Prediction using Ensemble Methods (Bagging and Boosting)

Author: Adel.Ahmadi

1. Bagging Method Overview

Bagging is an ensemble learning method for enhancing model stability and accuracy by training multiple base models on bootstrapped subsets of the dataset. Each model provides its prediction, and the final output is determined by voting. It helps reduce variance and prevents overfitting.

In this section, we used the BaggingClassifier with multiple SVM models as base learners to classify life expectancy categories.

2. Boosting Method Overview

Boosting is an ensemble method that sequentially trains models, with each new model correcting errors from the previous ones. This process enhances weak learners into a strong and accurate classifier.

We employed AdaBoost algorithm, where multiple Decision Trees were trained iteratively that reduces bias and variance and improves prediction performance in general.

3. Code Implementation

Let's review our main steps in this project:

a. Bagging with SVM

   A BaggingClassifier was used with SVM as the base. 6 SVM models were trained in this ensemble method and predictions from all models were aggregated using majority voting.

b. Boosting with Decision Trees

An AdaBoostClassifier was used with Decision Trees as base. 30 week learners were trained iteratively and each tree focused on correcting previous tree errors.

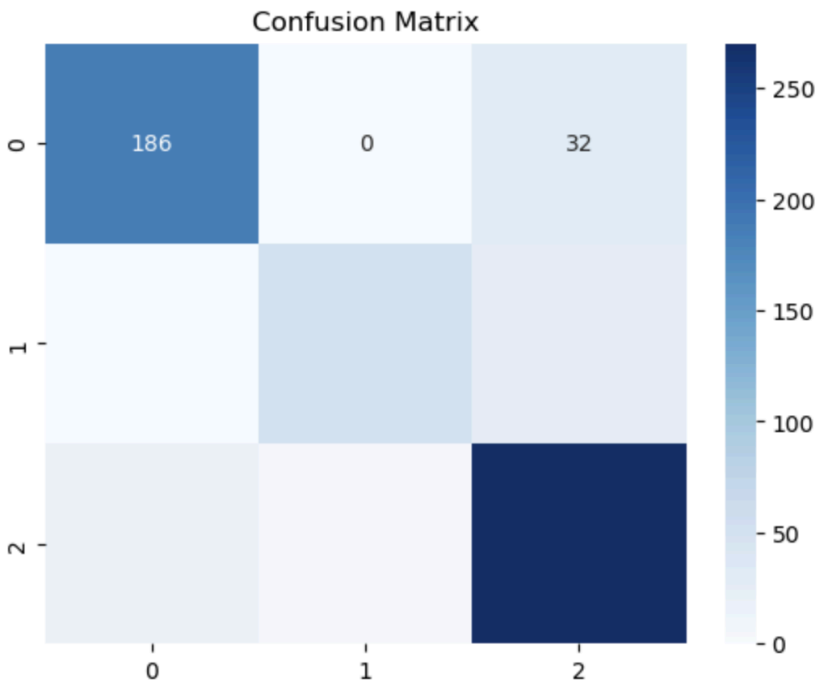
c. Evaluation

For both methods, we evaluated model performance with Classification Reports and Confusion Matrices.

4. Results

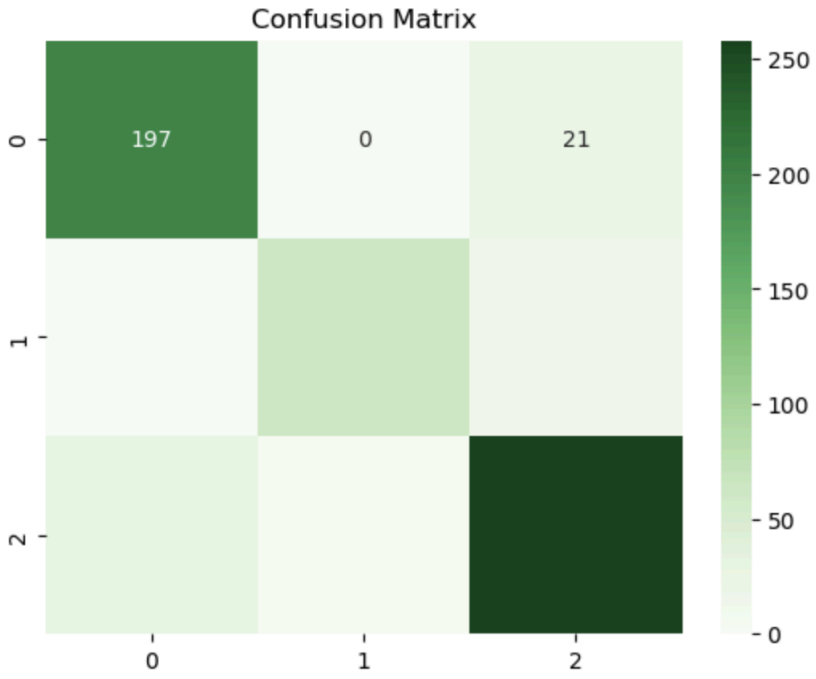
- **Bagging with SVM: Achieved an overall accuracy of 86%.**

Model Accuracy	Precision	Recall	F1-Score	Support
High	0.90	0.85	0.88	218
Low	0.91	0.65	0.76	75
Medium	0.82	0.92	0.87	295
Accuracy			0.86	588
Macro avg	0.88	0.81	0.83	588
Weighted avg	0.86	0.86	0.86	588



- **Boosting with Decision Trees and AdaBoost: Achieved an overall accuracy of 88%.**

Model Accuracy	Precision	Recall	F1-Score	Support
High	0.86	0.90	0.88	218
Low	0.93	0.83	0.87	75
Medium	0.88	0.87	0.88	295
Accuracy			0.88	588
Macro avg	0.89	0.87	0.88	588
Weighted avg	0.88	0.88	0.88	588


Confusion Matrix
197
0
21
0
1
2
0
1
2
250
200
150
100
50
0