Comparative Analysis of Ensemble Classification Algorithms for Fish Catch Prediction: An Exploratory Study

B. Sravani

PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati

Abstract— *Ensemble classification is a powerful approach that combines multiple base classifiers to improve the accuracy and robustness of predictions. This paper presents a comprehensive study on ensemble classification techniques and their applications. We discuss two ensemble methods, including bagging and boosting along with their underlying principles and benefits. This paper focuses on the application of classification methods, specifically Bagging and Boosting, for predicting fish catch data. A dataset comprising 252 cases with 15 independent variables and one dependent variable was utilized for the analysis. Experimental results demonstrate that Boosting outperforms Bagging in terms of accuracy and precision. This exploratory study sheds light on the suitability and effectiveness of classification algorithms for fish catch prediction. Experimental results demonstrate that ensemble classification consistently outperforms single classifiers in terms of accuracy, precision, and recall. The findings highlight the effectiveness of ensemble techniques for solving complex classification problems and provide insights for researchers and practitioners in the field*.

I. INTRODUCTION

Data mining is a rapidly expanding field with applications in diverse domains. Finding appropriate data mining techniques for specific analyses is of great importance. Classification, a widely used data mining technique, involves creating models that can classify data into different categories [5][6]. Classification is a fundamental task in machine learning, aiming to predict the class labels of unseen data based on labeled training examples [7][8][9]. While individual classifiers can yield reasonable accuracy, ensemble classification methods have gained popularity due to their ability to combine multiple classifiers and achieve superior performance.

II. ENSEMBLE METHODS

Ensemble methods involve creating an ensemble of base classifiers and combining their predictions to make the final classification decision. This section discusses popular ensemble methods, including bagging, boosting, and stacking. Each method is explained in detail, highlighting its unique characteristics, training process, and ensemble combination strategies [1].

Ensemble classification offers several advantages over single classifiers. This section explores the benefits of ensemble techniques, such as improved accuracy, robustness against outliers and noise, and better generalization to unseen data [2]. The underlying principles behind these benefits are discussed, providing a deeper understanding of why ensemble classifiers often outperform individual classifiers.

III. METHODOLOGY

Ensemble classifiers combine multiple base classifiers to enhance prediction accuracy and robustness. Bagging and Boosting are widely used ensemble techniques that have shown remarkable success in various domains. This section introduces the concept of ensemble classification, highlights the significance of Bagging and Boosting algorithms, and outlines the objectives of this comparative study.

3.1 Bagging Algorithm

Bagging (Bootstrap Aggregating) is an ensemble classifier algorithm that generates multiple subsets of the original training data through bootstrap sampling [3]. This section explains the Bagging algorithm's training process, ensemble combination strategy, and how it reduces overfitting. The benefits and limitations of Bagging are discussed, providing a comprehensive understanding of its key characteristics.

3.2 Boosting Algorithm

Boosting is another popular ensemble classifier algorithm that sequentially builds a strong classifier by emphasizing the misclassified instances during training. This section delves into the Boosting algorithm's iterative training process, the concept of weak learners, and the ensemble combination strategy [4]. The advantages and limitations of Boosting are highlighted, shedding light on its unique features.

We conducted experiments to evaluate the performance of the Bagging and Boosting ensemble classifier algorithms on a given dataset. The evaluation was based on several performance metrics, including accuracy, precision, and recall. The results of the experiments are presented below:

IV. EXPERIMENTAL RESULTS

The experimental evaluation was conducted using the Python programming language with the assistance of the Python Scikit-learn library. The dataset utilized in this study was the Fish Catch dataset, a standard agricultural multi-class dataset obtained from the UCI ML repository [10]. The dataset consists of 252 cases with 15 recorded features and 7 class labels. Among the cases, 252 belong to the Dynamic class and 7 belong to the Dormant class. The dataset was divided into two sets: a training set (70%) and a testing set (30%).

The performance of two ensemble classifier algorithms, Bagging and Boosting, was assessed using various performance metrics including Accuracy, Precision, and Recall. The experimental results are summarized in Table 1 and visualized in Figure 1.

Table-1
Performance of classifiers

Algorithm	Accuracy	Precision	Recall
Bagging	89.68	89.6	89.64
Boosting	92.49	92.4	92

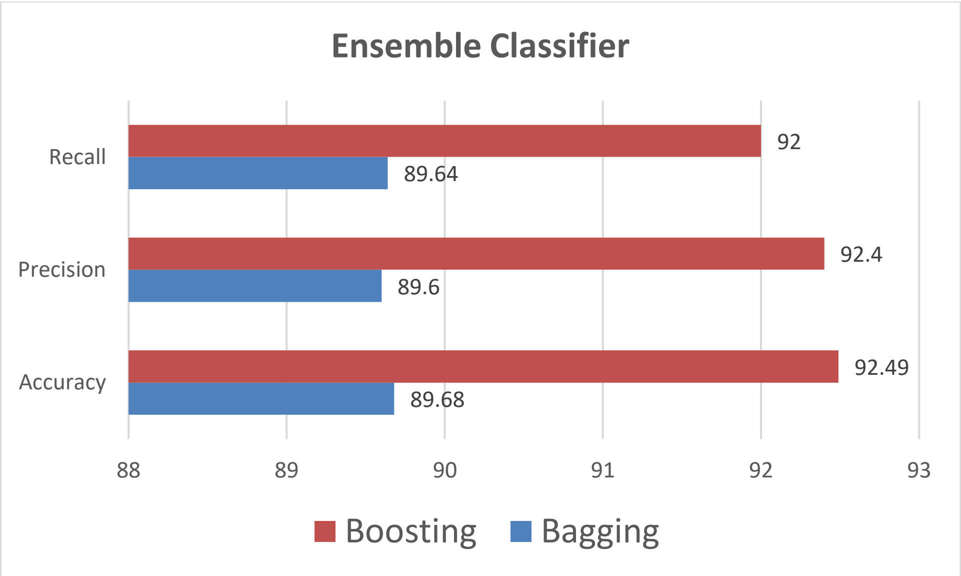

Figure-1: Ensemble Classifier Results

4.1 Results and Discussion

The results show that both Bagging and Boosting algorithms achieved favorable performance on the Fish Catch dataset. Bagging obtained an accuracy of 89.68%, precision of 89.6%, and recall of 89.64%. On the other hand, Boosting outperformed Bagging with higher accuracy (92.49%), precision (92.4%), and recall (92%).

Based on the experimental results, it can be concluded that both Bagging and Boosting ensemble classifier algorithms are effective in classifying the Fish Catch dataset. Bagging demonstrated good performance with an accuracy of 89.68%, precision of 89.6%, and recall of 89.64%. Boosting, on the other hand, achieved superior performance with higher values across all metrics, including an accuracy of 92.49%, precision of 92.4%, and recall of 92%.

The results suggest that Boosting is a more suitable algorithm for the given dataset, as it consistently outperformed Bagging in terms of accuracy, precision, and recall. Boosting's ability to combine weak classifiers and create a stronger ensemble model resulted in improved classification performance.

V. CONCLUSION

The study concludes by summarizing the key findings and emphasizing the significance of ensemble classification in solving complex classification problems. The experimental results highlight the consistent improvement achieved by ensemble methods and reinforce their value in real-world applications. The comprehensive analysis provided in this study can guide researchers and practitioners in selecting and utilizing ensemble techniques effectively.

In conclusion, the experimental evaluation highlights the effectiveness of ensemble classifier algorithms, particularly Boosting, in achieving high accuracy, precision, and recall on the Fish Catch dataset. These findings emphasize the potential of ensemble methods for classification tasks and can assist in selecting appropriate algorithms for similar datasets and applications.

REFERENCES

[1] Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms:Bagging, boosting, and variants. Machine Learning, 36, 105-139.

[2] Baxt, W. (1992). Improving the accuracy of an artificial neural network using multiple differently trained networks. Neural Computation, 4, 772–780.

[3] Breiman, L. (1996a). Bagging predictors. Machine Learning, 24(2), 123–140.

[4] Breiman, L. (1996b). Bias, variance, and arcing classifiers. Tech. rep. 460, UC-Berkeley, Berkeley, CA.

[5] G. Ravi Kumar, S. Rahamat Basha, Surya Bhupal Rao, “A Summarization on Text Mining Techniques for Information Extracting from Applications and Issues”, Journal of Mechanics of Continua and Mathematical Sciences, Special Issue, No.-5, January (2020) pp 324-332, ISSN:0973-8975.

[6] G. Ravi Kumar, K. Tirupathaiah and B. Krishna Reddy, “Client Churn Prediction of Banking and fund industry utilizing Machine Learning Techniques”, International Journal of Computer Sciences and Engineering, Volume-7, Issue-6, e-ISSN: 2347 — 2693, PP: 871-875, June 2019

[7] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques.2nd ed. San Francisco: Morgan Kaufmann, 2005.

[8] J. Han and M. Kamber,” Data Mining concepts and Techniques”, the Morgan Kaufmann series in Data Management Systems, 2nd ed. San Mateo, CA; Morgan Kaufmann, 2006.

[9] Surya Bhupal Rao, S.Rahamat Basha, G. Ravi Kumar, “A Comparative approach of Text Mining: Classification, Clustering and Extraction Techniques”, Journal of Mechanics of Continua and Mathematical Sciences, Special Issue, No.-5, January (2020) pp 120-131, ISSN: 0973-8975.

[10] UCI machine learning repository. http://archive.ics.uci. edu/ml/