Literature Review: Boosting Techniques in Classification, with Emphasis on C4.5 Algorithm

Data mining classification approaches have become integral in handling massive databases, providing solutions to complex problems. Among these approaches, the C4.5 algorithm, Naive Bayes, Neural Network, and K-NN are widely utilized. The C4.5 algorithm, renowned for its research history in classification, is not without limitations, particularly in terms of over-fitting and convergence. To address these challenges, boosting algorithms, such as Adaboost, have emerged, aiming to enhance accuracy by combining multiple classifiers and reducing variance.

Boosting techniques have demonstrated effectiveness in various applications, extending beyond traditional classification problems. Previous research has successfully applied boosting in diverse fields, including brain tumor detection and the classification of Parkinsonian disorders, showcasing its adaptability and robustness. The key strength of boosting lies in its ability to mitigate classification errors by assigning higher weights to misclassified instances, thus balancing class representation and improving overall data spread.

The focus of this literature review centers on a recent article by Sunarto et al. (2023), titled "Application of Boosting Technique with C4.5 Algorithm to Reduce the Classification Error Rate in Online Shoppers Purchasing Intention." The primary goal of their research is to leverage boosting techniques, specifically Adaboost, in conjunction with the C4.5 algorithm, to reduce the classification error rate in predicting online shoppers' purchasing intention.

The rationale behind incorporating boosting in this context is to give more weight to instances with classification errors, thereby addressing imbalances in class representation. The authors argue that this approach can lead to a more accurate and reliable prediction of online shoppers' purchasing behavior.

To substantiate their findings, Sunarto et al. (2023) employ the Online Shoppers Purchasing Intention dataset from the UCI Machine Learning Repository. This dataset serves as a practical testbed to evaluate the effectiveness of the boosting technique, particularly in combination with the C4.5 algorithm, in the domain of e-commerce and online shopping.

In summary, the literature surrounding boosting techniques, with a specific focus on the application of Adaboost and the C4.5 algorithm, reveals a growing interest in overcoming the limitations of individual classifiers. The Sunarto et al. (2023) study contributes to this body of knowledge by applying boosting to the domain of online shoppers' purchasing intention, demonstrating the potential for improved predictive accuracy and model robustness in e-commerce applications.