Dublin Business School

Student Declaration:

By submitting this assignment, I confirm that I am aware of DBS's policy regarding cheating, plagiarism and all other forms of academic impropriety. The coursework submitted is my own or my group's work, and all other sources consulted have been appropriately acknowledged. I am aware that in the case of doubt an investigation will be held.

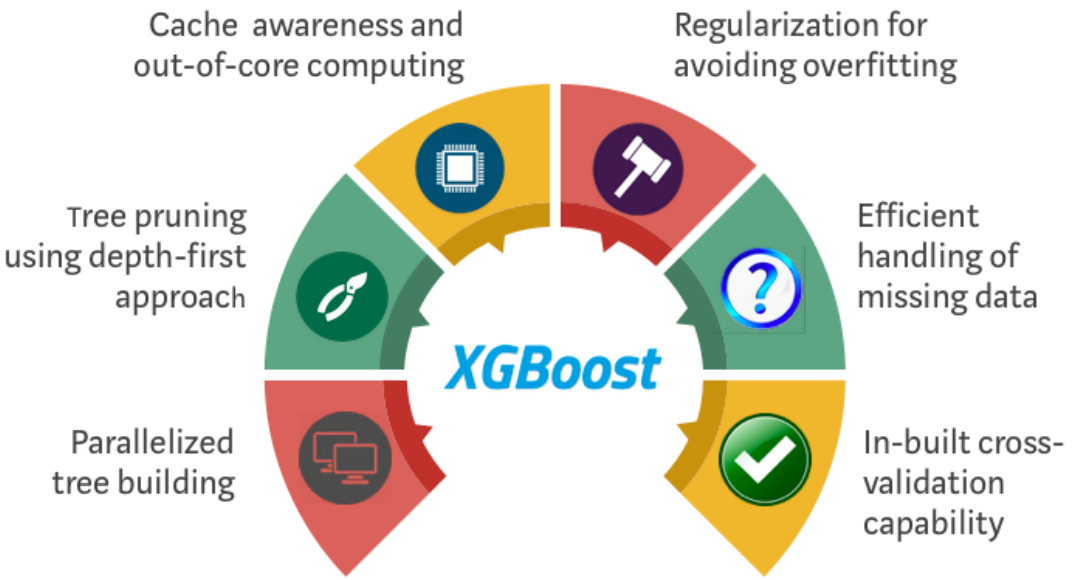
Student's Name: Advika kavya Parlakoti Student No. (from student card): 20003401
Pradyumna Shivanandaiah : 20006386
Satya Kishore Reddy Duvvi : 10639744

Assignment Title: - Critique of the XGBoost Algorithm
Professor: Kunwar Madan

Date of Submission: 03/11/2023

Signed: 

Critique of the XGBoost Algorithm



Images Reference: XGBoost Algorithm: Long May She Reign! | by Vishal Morde | Towards Data Science

Contents

Executive Summary

XGBoost has emerged as a powerhouse in the machine learning landscape, particularly in the realm of classification problems. This paper critically evaluates the application of XGBoost in predicting customer satisfaction, examining its performance, identifying its key strengths and weaknesses, and discussing the implications of its application in a business context. The model's remarkable accuracy (0.94) and ROC AUC (0.99) underscore its potential for robust predictive capabilities.

Introduction to XGBoost

XGBoost, an optimized distributed gradient-boosting library, is designed to be highly efficient, flexible, and portable. It implements machine learning algorithms under the Gradient Boosting framework, offering a solution that is both powerful in prediction and versatile in handling various types of data (Chen & Guestrin, 2016). In this critique, XGBoost's application in predicting customer satisfaction is dissected, highlighting the algorithm's strengths in handling a complex dataset with multiple variables that span demographic and service-related factors.

Working of the XGBoost Algorithm

XGBoost builds decision trees sequentially, where each tree corrects the errors of its predecessors. It integrates methods to prevent overfitting while maintaining model complexity. The algorithm stands out for its ability to manage sparse data and use parallel processing, which accelerates computational time.

Key Features:

Gradient Boosting: XGBoost applies this technique for optimization.
Regularization: It includes L1 and L2 regularization, which prevents overfitting.
Cross-Validation: Built-in cross-validation aids in determining the efficacy of the model.

Application of XGBoost Algorithm on Dataset:

XGBoost has garnered acclaim across diverse industries, serving pivotal roles from fraud detection in finance to disease forecasting in healthcare, optimizing customer retention in retail, enhancing predictive maintenance in manufacturing, and improving traffic flow in smart cities. Its prowess in handling complex datasets with speed and accuracy makes it indispensable for businesses seeking data-driven insights for strategic decision-making, thereby driving profitability and efficiency.

On Related to our Dataset

The binary target, customer satisfaction, was analyzed against a host of predictors. Numerical variables such as age and flight distance were normalized, while categorical variables like gender and class were encoded appropriately to ensure model compatibility.

Strengths of XGBoost:

Model Performance

At the core of XGBoost's strengths is its performance metrics. The model achieved a high accuracy of 0.94 and an ROC AUC of 0.99 in the present study, indicative of its superior predictive power These metrics suggest that the model not only predicts the correct outcomes most of the time but also distinguishes between the classes with a high degree of certainty.

Feature Importance

XGBoost provides a built-in way to examine the importance of features within the model. In this study, feature importance analysis revealed that variables such as 'Flight Distance' and 'Type of Travel' were significant predictors of customer satisfaction. This aligns with industry understanding and provides a validation of the model's practical utility.

Weaknesses and Challenges

Parameter Tuning

XGBoost's parameter space is vast, making the process of tuning the model to achieve optimal performance challenging and time-consuming. The hyperparameter tuning process, often requiring a methodical approach like grid search or random search, can become a significant bottleneck, especially for practitioners with limited computational resources

Scalability

Although XGBoost is engineered for efficiency, it can struggle with scalability when dealing with extremely large datasets. The computational demand can become a limiting factor, necessitating advanced hardware or distributed computing frameworks to handle the workload.

Conclusion:

In conclusion, the XGBoost algorithm stands out as a robust, efficient, and versatile machine learning technique that has significantly impacted the field of data science. With its ability to manage large and complex datasets, deliver high performance with speed, and offer flexibility through hyperparameter tuning, XGBoost has proven to be an invaluable asset in predictive analytics. Its widespread application across various domains has not only led to enhanced operational efficiencies but also to innovative solutions to complex problems. The continuous development and community support behind XGBoost promise further advancements, making it a go-to algorithm for researchers and practitioners seeking to extract deep insights and predictive power from their data.

References

- XGBoost Documentation, XGBoost Developers. :https://xgboost.readthedocs.io/
- XGBoost: A Comprehensive Guide,Towards Data Science, Medium. https://towardsdatascience.com/understanding-xgboost-4f3aaf82d8b7
- Hyperparameter Tuning in XGBoost, Machine Learning Mastery. https://machinelearningmastery.com/hyperparameter-optimization-with-random-search-and-grid-search/
- Dataset: https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction/data