

## Predictive Analysis for Big Data: Extension of Classification and Regression Trees Algorithm

**Authors :** Ameer Abdelkader, Abed Bouarfa Hafida

**Abstract :** Since its inception, predictive analysis has revolutionized the IT industry through its robustness and decision-making facilities. It involves the application of a set of data processing techniques and algorithms in order to create predictive models. Its principle is based on finding relationships between explanatory variables and the predicted variables. Past occurrences are exploited to predict and to derive the unknown outcome. With the advent of big data, many studies have suggested the use of predictive analytics in order to process and analyze big data. Nevertheless, they have been curbed by the limits of classical methods of predictive analysis in case of a large amount of data. In fact, because of their volumes, their nature (semi or unstructured) and their variety, it is impossible to analyze efficiently big data via classical methods of predictive analysis. The authors attribute this weakness to the fact that predictive analysis algorithms do not allow the parallelization and distribution of calculation. In this paper, we propose to extend the predictive analysis algorithm, Classification And Regression Trees (CART), in order to adapt it for big data analysis. The major changes of this algorithm are presented and then a version of the extended algorithm is defined in order to make it applicable for a huge quantity of data.

**Keywords :** predictive analysis, big data, predictive analysis algorithms, CART algorithm

**Conference Title :** ICAOSS 2019 : International Conference on Aspect-Oriented Software Systems

**Conference Location :** Paris, France

**Conference Dates :** September 19-20, 2019