

LEVERAGING HEURISTIC SAMPLING AND ENSEMBLE LEARNING FOR ENHANCED INSURANCE BIG DATA CLASSIFICATION

Naresh Kumar Reddy Panga

Virtusa Corporation, New York, USA

ABSTRACT

This paper combines a merging ensemble learning approach with heuristic bootstrap sampling for the analysis of large-scale insurance data. Traditional methods such as logistic regression and support vector machines (SVM) suffer from problems including lack of user knowledge, imbalanced datasets. Good news: We then created an improved ensemble random forest method to leverage Spark's memory-cache and parallel processing capabilities. Results: The method used obtained greater accuracy and efficiency than conventional methods, as evidenced by testing with data from China Life Insurance Company. Metrics like F-Measure and G-Mean illustrate how well the machine-learning algorithm works with imbalanced data, enhancing its effectiveness as a tool for improving insurance marketing campaigns while pinpointing prospective customers.

KEYWORDS: *Big Data, Heuristic Sampling, Ensemble Learning, Random Forest, Spark Optimization, Insurance Data.*

Article History

Received: 16 Jan 2020 | Revised: 21 Jan 2020 | Accepted: 24 Jan 2020
