

INTISARI

PERBANDINGAN PERFORMA REGRESI LOGISTIK BINER DAN DECISION TREE C4.5 DALAM KLASIFIKASI MENGGUNAKAN METODE *BOOTSTRAP AGGREGATING (BAGGING)*

Oleh

Bella Vebrya Arifin

18/427692/PA/18652

Data mining adalah proses memodelkan data dalam jumlah besar serta mencari pola untuk memperoleh informasi yang berguna. Salah satu teknik dalam *data mining* adalah klasifikasi. Klasifikasi digunakan untuk mengelompokkan data secara sistematis. Beberapa metode dalam klasifikasi adalah regresi logistik dan *decision tree* C4.5. Mayoritas metode dalam teknik klasifikasi mempunyai kelemahan dalam menangani *dataset* yang memiliki ketidakseimbangan kelas sehingga dapat menurunkan performa klasifikasi. Namun dalam *data mining* sering kali muncul ketidakseimbangan kelas. Metode *bootstrap aggregating (bagging)* merupakan salah satu metode *ensemble* untuk membentuk klasifikasi yang lebih stabil dan meningkatkan performa klasifikasi. Pada penelitian ini digunakan metode *bagging* untuk klasifikasi pada regresi logistik dan *decision tree* C4.5 dengan beberapa kali replikasi *bootstrap*. Metode regresi logistik dan *decision tree* C4.5 tanpa *bagging* digunakan sebagai pembanding untuk mengetahui metode mana yang memiliki performa terbaik. Dari hasil analisis menggunakan data *Pima Indians Diabetes* diperoleh kesimpulan bahwa penerapan metode *bagging* pada regresi logistik dan *decision tree* C4.5 mengalami peningkatan performa dibandingkan metode yang tidak menggunakan *bagging*. Dan penerapan *bagging* pada *decision tree* C4.5 mendapatkan hasil analisis dengan performa terbaik dibanding yang lain.

Kata kunci: Klasifikasi, Ketidakseimbangan Kelas, Regresi Logistik, *Decision Tree* C4.5, *Bootstrap Aggregating (Bagging)*

ABSTRACT

THE COMPARISON OF BINARY LOGISTIC REGRESSION AND DECISION TREE C4.5 PERFORMANCE IN CLASSIFICATION USING BOOTSTRAP AGGREGATING (BAGGING) METHOD

By

Bella Vebrya Arifin

18/427692/PA/18652

Data mining is the process of modeling large amounts of data and looking for patterns to obtain useful information. One of the techniques in data mining is classification. Classification is used to group data systematically. Several methods in the classification are logistic regression and decision tree C4.5. The majority of methods in the classification technique have weaknesses in handling datasets that have class imbalance which can reduce classification performance. However, in data mining, class imbalance often arises. The bootstrap aggregating (bagging) method is one of the ensemble methods to form a more stable classification and improve classification performance. In this study, the bagging method was used for classification in logistic regression and decision tree C4.5 with several bootstrap replications. The logistic regression method and decision tree C4.5 without bagging were used as comparisons to find out which method had the best performance. From the results of the analysis using Pima Indians Diabetes data, it was concluded that the application of the bagging method to logistic regression and decision tree C4.5 experienced an increase in performance compared to methods that did not use bagging. And the application of bagging in decision tree C4.5 gets the analysis results with the best performance compared to others.

Keywords: Classification, Imbalanced Class, Logistic Regression, Decision tree C4.5, Bootstrap Aggregating (Bagging)