

INTISARI

IMPLEMENTASI *EXTREME GRADIENT BOOSTING* (XGBOOST) DALAM MENANGANI *IMBALANCED CLASS* PADA ANALISIS KLASIFIKASI

oleh

Aniendyatri Dwi Marhaening Kusumastuti

18/424280/PA/18385

Teknologi membawa solusi atas permasalahan klasifikasi dalam data analisis, salah satunya adalah permasalahan *imbalanced class* dalam analisis klasifikasi. *Imbalanced class* merupakan keadaan di mana salah satu kelas dalam variabel dependen memiliki jumlah yang lebih banyak secara signifikan dibanding kelas yang lain. Keadaan ini mampu menimbulkan model yang bias akibat kurangnya kemampuan model untuk memprediksi kelas dengan jumlah observasi lebih sedikit (kelas minor). Penanganan dengan pendekatan algoritma berupa penggunaan algoritma *ensemble learning* dan modifikasinya diterapkan untuk mengatasi permasalahan tersebut. *Extreme Gradient Boosting* (XGBoost) merupakan metode *ensemble learning* yang membentuk beberapa *decision tree* secara berurutan untuk memperbaiki kesalahan yang dibentuk model sebelumnya. Modifikasi algoritma XGBoost dengan mengaplikasikan pembobotan, *Weighted-Loss*, dan *Focal-Loss*. Pada skripsi ini dilakukan analisis pada *Bank Marketing* dengan menggunakan XGBoost dan modifikasinya, kemudian dibandingkan dengan model dasarnya, *decision tree*, dan model *ensemble learning* lain, yaitu AdaBoost dan *Random Forest*. Dari analisis yang dilakukan diperoleh kesimpulan bahwa *Weighted-XGBoost* menghasilkan performa yang paling baik.

Kata kunci : klasifikasi, data tidak seimbang, XGBoost, Weighted Loss, Focal-Loss

ABSTRACT

IMPLEMENTATION OF EXTREME GRADIENT BOOSTING (XGBOOST) FOR HANDLING IMBALANCED CLASS IN CLASSIFICATION ANALYSIS

by

Aniendyaratri Dwi Marhaening Kusumastuti

18/424280/PA/18385

Technology brings solutions to classification problem in data analysis, such as the problem of imbalanced class in classification analysis. Imbalanced class is a condition where one of the classes in the dependent variable has significantly more numbers than the other classes. This situation can lead to a biased model due to the model's lack ability to predict classes with fewer observations (minority class). With an algorithmic approach, ensemble learning and algorithm modification, are applied to overcome these problems. *Extreme Gradient Boosting* (XGBoost) is an ensemble learning method that build several decision trees in sequence to correct errors made by the previous model. For handling the imbalanced class, modification of the XGBoost algorithm is done by applying weighting, Weighted-Loss, dan Focal-Loss. In this thesis we analyze Bank Marketing data using XGBoost and its modifications, then compared with basic model, decision tree, and other ensemble learning models, namely AdaBoost and Random Forest. From the analysis, it was concluded that Weighted XGBoost has the best performance.

Keywords : classification, imbalanced class, XGBoost, Weighted-Loss, Focal-Loss