

## ABSTRAK

Nama : Rizha Fauzi  
Program Studi : Magister Teknik Industri  
Judul : **FORMULASI MODEL PREDIKSI KUALITAS PADA KASUS IMBALANCED DATASET DENGAN PENDEKATAN LEVEL DATA**  
Pembimbing : Fahmi Arif, ST., MT., Ph.D

Dalam mengontrol dan meningkatkan kualitas produk banyak teknik yang diimplementasikan ke dalam proses pembuatan. Prediksi kualitas menjadi salah satu cara yang paling efektif untuk mengendalikan dan meningkatkan kualitas manufaktur dan telah dikembangkan menggunakan berbagai teknik data mining. Proses produksi di suatu perusahaan, pada umumnya menghasilkan produk cacat dalam jumlah yang sedikit dibanding produk yang berkualitas baik. Hal ini menyebabkan formulasi model prediksi kualitas biasanya diterapkan pada dataset yg tidak seimbang. Pada *imbalanced data* itu menunjukkan ketidakakuratan dikarenakan ada kelas mayoritas mendominasi yang menjadikan kelas minoritas tidak terdeteksi. Formulasi model prediksi juga menghadapi masalah-masalah sebagai berikut: 1) Belum diketahuinya teknik preprocessing yang efektif dalam mengatasi ketidakseimbangan data pada dataset dengan atribut numerik dan kelas nominal; dan 2) Hasil performansi yang beragam di setiap parameter yang diujikan pada penerapan teknik yang dilakukan ini menjadikan formulasi model prediksi kualitas ini bersifat kompleks. Metode yang digunakan pada penelitian ini yaitu diadaptasi dari *Data Mining Methodology for Engineering* sehingga proses *data mining* menjadi lebih sistematis dan terdokumentasi dengan baik. Hasil penelitian ini menunjukkan bahwa performansi dari kombinasi *random oversampling* dengan *random forest* memiliki performansi yang tinggi dibandingkan metode lainnya. Nilai akurasi yang didapat yaitu 0,9798, nilai *g-mean* yang didapat 0,9798 begitu pula nilai *f-measure* yakni 0,9798. Hal tersebut menunjukkan bahwa kombinasi metode tersebut dapat menghasilkan sebuah prediksi kualitas yang sangat baik. Kemudian hasil model prediksi yang dilakukan pun memperlihatkan nilai yang serupa antara data awal dengan hasil prediksi yang dilakukan.

Kata kunci : Prediksi Kualitas, *Imbalanced Data*, *Random Forest*, *Random Oversampling*, Akurasi, *F-Measure*, *G-Mean*.

## **ABSTRACT**

Name : Rizha Fauzi  
Study Program : Magister Teknik Industri  
Title : **QUALITY PREDICTION MODEL FORMULATION IN THE CASE OF IMBALANCED DATASET WITH DATA LEVEL APPROACH**  
Counsellor : Fahmi Arif, ST., MT., Ph.D

*In controlling and improving product quality, many techniques are implemented into the manufacturing process. Quality prediction is one of the most effective ways to control and improve manufacturing quality and has been developed using various data mining techniques. The production process in a company, in general, produces defective products in small quantities compared to good quality products. This results in the formulation of quality prediction models usually applied to imbalanced datasets. The imbalanced data shows inaccuracy because there is a majority class that dominates which makes the minority class undetectable. The prediction model formulation faces the following problems: 1) There is no known effective preprocessing technique in overcoming the imbalanced dataset with numeric attributes and nominal classes; and 2) The various performance results in each parameter tested in the application of this technique make the formulation of this quality prediction model complex. The method used in this research is adapted from the Data Mining Methodology for Engineering so that the data mining process becomes more systematic and well documented. The results of this study indicate that the performance of the combination of random oversampling and random forest has a higher performance than other methods. The accuracy value obtained is 0.9798, the g-mean value obtained is 0.9798 as well as the f-measure value which is 0.9798. This shows that the combination of these methods can produce a very good prediction of quality. Then the results of the prediction model carried out show similar values between the production data and the results of the predictions made.*

*Key words : Quality Prediction, Imbalanced Data, Random Forest, Random Oversampling, Accuracy, F-Measure, G-Mean.*