

## Intisari

Perusahaan ataupun organisasi yang bergerak di bidang bisnis pada umumnya memiliki strategi pemasaran tersendiri untuk dapat menarik perhatian pelanggan dan mendapatkan keuntungan darinya. Salah satu strategi pemasaran, yaitu *direct marketing*, bertujuan untuk menargetkan pelanggan yang secara spesifik berada dalam kriteria atau golongan tertentu. Dalam penggunaan strategi pemasaran *direct marketing*, pemilihan dan pengelompokan pelanggan berpotensi sebagai target pasar dilakukan berdasarkan data dan informasi yang tersedia dalam *database* perusahaan. Kebutuhan inilah yang menyebabkan *data mining* dibutuhkan sebagai teknik komputasi sekaligus analisis statistik untuk mendapatkan target pasar yang tepat melalui data dan informasi tersebut. Penelitian berkaitan dengan *data mining* pada *direct marketing* pernah dilakukan sebelumnya, yaitu dengan teknik *uplift modelling*. Akan tetapi, dalam penelitian tersebut tidak terdapat penanganan terhadap distribusi jumlah sampel berdasarkan label kelas yang tidak merata (*imbalance*) pada *dataset*. Selain itu, prediksi target pelanggan pada *direct marketing* menggunakan teknik *uplift modelling* hanya cocok pada kasus dimana terdapat *control group* dengan jumlah tertentu pada *dataset*.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk memperbaiki penelitian sebelumnya. Pada penelitian ini, teknik *sampling Synthetic Minority Oversampling Technique* (SMOTE) yang dikombinasikan dengan teknik *undersampling* diterapkan pada tahap *pre-processing* pada *data mining*. Hal ini dilakukan sebagai bentuk penanganan terhadap distribusi jumlah sampel berdasarkan label kelas yang bersifat *imbalance*. Kemudian, prediksi target pelanggan berpotensi dilakukan menggunakan metode klasifikasi dengan model klasifikasi *Neural Network* (NN) dan *Support Vector Machine* (SVM). Model NN dan SVM digunakan karena lebih fleksibel dibandingkan dengan teknik *uplift modelling* yang membutuhkan adanya *control group*. Selain itu, model NN dan SVM dipilih karena memiliki performa yang lebih baik dibandingkan model konvensional lainnya.

Evaluasi hasil prediksi oleh model NN dan SVM dilakukan dengan pengukuran performa menggunakan *accuracy* dan *Area Under the Curve* (AUC) pada masing-masing model. Dari hasil evaluasi, didapatkan bahwa model NN memiliki performa yang lebih baik dari model SVM. Analisis lebih lanjut dilakukan dengan melakukan perhitungan *variable importance* untuk mengetahui faktor-faktor yang mempengaruhi respon pelanggan terhadap *direct marketing*. Faktor yang didapatkan yaitu pelanggan baru, pelanggan wanita, sejarah pembelian, dan promosi produk pria. Faktor-faktor tersebut dapat dijadikan pertimbangan oleh manajer di divisi pemasaran dalam membuat strategi pemasaran berikutnya.

**Kata kunci :** *Direct marketing, data mining, Neural Network, Support Vector Machine, SMOTE*

## ***Abstract***

*Any company or organization whose area is focused on business generally has a marketing strategy to gain interest from its customers and thus receive benefits from them. Direct marketing, as one of many marketing strategies, has an objective to target customers who have specific criteria or are in a specific group of people. When using this marketing strategy, the company decides and classifies its customers based on their data and information stored in a database. This necessity is the reason why data mining is needed as a computation and a statistical analysis technique to obtain the proper target market for direct marketing. A research focusing on implementation of data mining in direct marketing had been conducted using uplift modelling technique. However, this preceding research did not handle the imbalanced distribution of the instances based on their label class in the dataset. Moreover, predicting the potential customers in direct marketing using uplift modelling technique is only proper when the dataset contains a right amount of control group.*

*This research aims to improve the preceding one with respect to the issues mentioned before. In this research, Synthetic Minority Oversampling Technique (SMOTE) technique combined with undersampling technique is applied in the pre-processing step of data mining. This action is conducted as a handling of the imbalanced distribution of instances based on their label class in the dataset. Subsequently, potential customers is predicted using classification method with a Neural Network (NN) and a Support Vector Machine (SVM) model. NN and SVM are used because of their better flexibility compared to uplift modelling technique that requires a control group in its dataset. Moreover, NN and SVM are chosen for their better performance compared with other conventional models.*

*The NN and SVM model's prediction results are evaluated by accuracy and Area Under the Curve (AUC) performance measurement. The evaluation resulted in NN having a better performance compared to SVM. For further analysis, variable importance calculation is conducted to discover factors which affected customers' response towards direct marketing. These factors are new customers, women customers, purchase history, and men's product promotion. These factors can then be a consideration by the marketing division manager to build the next marketing strategy.*

**Keywords :** *Direct marketing, data mining, Neural Network, Support Vector Machine, SMOTE*