

Integrasi SMOTE dan Ensemble AdaBoost Untuk Mengatasi Imbalance Class Pada Data Bank Direct Marketing

Amin Nur Rais¹, Agus Subekti²

¹STMIK Nusa Mandiri Jakarta

²Pusat Penelitian Elektronik dan Telekomunikasi
Lembaga Ilmu Pengetahuan Indonesia (LIPI)

e-mail: ¹aminnurraiskleca@gmail.com, ²agus.subekti@lipi.go.id

Abstrak

Kampanye pemasaran produk bank secara langsung dapat dibantu dengan adanya teknologi informasi. Dengan terus bertambahnya data dan penggunaan teknologi informasi, data yang didapatkan dapat dimanfaatkan lebih maksimal untuk membuat keputusan. Tetapi, data yang ada menunjukkan ketidak seimbangan pada kelasnya sehingga dapat menurunkan prediksi dalam mengambil keputusan karena cenderung menghasilkan prediksi kelas mayoritas daripada kelas minoritas. Dengan menyeimbangkan data dengan menggunakan pra pemrosesan SMOTE, kelas data dapat menjadi lebih seimbang. Pada penelitian ini, diusulkan model dengan menggunakan pra proses SMOTE dan ensemble AdaBoost yang dikombinasikan dengan algoritma klasifikasi untuk mendapatkan nilai akurasi yang terbaik. Algoritma klasifikasi yang digunakan yaitu Naïve Bayes, SVM, dan Decision Tree. Hasil penelitian menunjukkan bahwa model klasifikasi dengan model SMOTE dan AdaBoost yang digabungkan dengan algoritma klasifikasi naïve bayes memiliki nilai akurasi terendah dengan nilai akurasi 88,3%, sedangkan SVM memiliki akurasi 89,68%, dan pada model klasifikasi dengan Decision Tree memperoleh hasil klasifikasi terbaik dengan nilai akurasi 92,25%.

Kata kunci: Kelas tidak seimbang, smote, ensemble adaboost, pemasaran langsung

Abstract

Direct bank product marketing campaigns can be helped by the existence of information technology. As data continues to grow and the use of information technology, the data obtained can be utilized more optimally to make decisions. However, existing data shows an imbalance in the class so that it can reduce predictions in making decisions because it tends to produce the prediction of the majority class rather than the minority class. By balancing data using SMOTE pre-processing, data classes can become more balanced. In this study, a proposed model using the SMOTE pre-process and the AdaBoost ensemble is combined with a classification algorithm to get the best accuracy value. The classification algorithm used is Naïve Bayes, SVM, and Decision Tree. The results showed that the classification model with the SMOTE and AdaBoost models combined with the naïve bayes classification algorithm had the lowest accuracy value with an accuracy value of 88.3%, whereas SVM had an accuracy of 89.68%, and the classification model with the Decision Tree obtained the best classification results with an accuracy value of 92.25%.

Kata kunci: Unbalanced class, smote, adaboost ensemble, direct marketing

1. Pendahuluan

Mengenalkan produk secara langsung telah umum digunakan untuk berbagai industri, salah satunya industry perbankan. Dalam mengenalkan produk

secara langsung, bank dapat melakukan analisa pasar dengan memanfaatkan ruang teknologi informasi yang dapat membantu dalam mengambil keputusan (Abbas, 2015). Kampanye pemasaran produk bank secara langsung sangat berguna untuk

menawarkan produk baru kepada calon pelanggan. Dengan melakukan analisa data pemasaran bank, dapat digunakan untuk memilih jenis pemasaran yang dilakukan. Kampanye pemasaran dapat dilakukan melalui email, telepon, dan email langsung kepada calon pelanggan yang memungkinkan calon pelanggan dapat memutuskan untuk mengambil produk yang ditawarkan atau tidak. Proses analisa dapat dilakukan dengan teknologi informasi yang didukung dengan terus bertambahnya informasi data pelanggan untuk mendukung pengambilan keputusan (Moro, Laureano, & Cortez, 2011)

Dengan bertambahnya waktu, jumlah data yang masuk terus bertambah. Data yang ada memberikan informasi tentang konsumen yang berguna untuk membuat keputusan di masa mendatang. Akan tetapi, terjadi ketimpangan respon dari konsumen terhadap penawaran yang dilakukan. Sehingga terjadi ketidakseimbangan kelas pada data. Permasalahan ketidakseimbangan kelas (Imbalance Class) menyebabkan hasil klasifikasi lebih condong ke class mayoritas. Dimana jumlah class negative (mayoritas) lebih besar dari class positive (minoritas) yang dapat menurunkan kinerja model prediksi.

Dalam menangani dataset yang tidak seimbang, terdapat 3 pendekatan yang dapat dilakukan, yaitu pendekatan level data, level algoritmik, dan pendekatan dengan menggabungkan metode. Pada pendekatan level data, mencakup berbagai teknik resampling untuk memperbaiki distribusi kelas pada data. Pada level algoritmik, dilakukan proses penyesuaian operasi algoritma yang ada untuk menjalankan pengklasifikasian (classifier) agar lebih konduktif terhadap klasifikasi kelas minoritas. Sedangkan pada pendekatan gabungan (ensemble), terdapat 2 ensemble-learning paling populer, yaitu boosting dan bagging. Pada pendekatan algoritma dan ensemble memiliki tujuan yang sama, yaitu memperbaiki algoritma pengklasifikasi tanpa mengubah data, sehingga dapat dianggap ada 2 pendekatan saja, yaitu pendekatan level data dan pendekatan level algoritma. Dengan membagi menjadi 2 pendekatan dapat mempermudah dalam proses perbaikan.

Untuk mengatasi ketidakseimbangan kelas, dapat digunakan

pendekatan teknik sampling over-sampling, under sampling, dan kombinasi keduanya (Perbankan & No, 2016). Salah satu teknik sampling yang digunakan adalah dengan melakukan pendekatan level data dengan teknik SMOTE (Synthetic Minority Over-sampling Technique) (Putri, 2015).

Dalam mengolah data yang asimetris, digunakan algoritma SMOTE dan Rotation Forest (PCS)-J48 dimana SMOTE digunakan untuk memodifikasi data dan meningkatkan keakuratan prediksi. Dengan menggunakan PCA-J48 mendapatkan nilai akurasi yang tinggi dengan nilai 90.81% dan sensitivitas 81,23%. Namun sensitivitas metode BayesNet dan PCA-RandomTree memiliki nilai sensitivitas lebih tinggi dari PCA-J48 (Ruangthong & Jaiyen, 2015).

Dalam memprediksi respon pelanggan, dengan menerapkan empat pengklasifikasian, yaitu Multilayer Perceptron Neural Network (MLPNN), Decision Tree (C4.5), Regresi Logistik, dan Random Forest (RF). Penelitian ini menggambarkan bahwa klasifikasi dengan RF menjadi klasifikasi paling produktif dalam kemampuan prediksi dengan nilai akurasi 87%. Sedangkan fitur utama dari pelanggan yang kemungkinan besar berlangganan berjangka jika pelanggan menghabiskan lebih lama dalam panggilan dengan pendidikan minimal sekolah menengah (Asare-Frempong & Jayabalan, 2017).

Pada penelitian ini dilakukan dengan pendekatan teknik SMOTE yang dikombinasikan dengan ensemble AdaBoost dengan algoritma J48, SVM, dan Naïve Bayes.

2. Metode Penelitian

Paper ini menggunakan dataset public yang diambil dari kaggle repository tentang pemasaran langsung yang dilakukan oleh bank di Portugis dengan 41.188 data yang terdiri dari 20 atribut dan 2 kelas. dataset ini berkaitan dengan pemasaran secara langsung yang dilakukan oleh lembaga perbankan Portugis kepada calon pelanggannya. Pemasaran ini didasarkan pada panggilan telepon. Biasanya perlu lebih dari satu kali proses penawaran kepada calon konsumen yang sama untuk memastikan apakah akan menggunakan prosuk yang ditawarkan atau tidak.

Dataset ini terbagi menjadi 4 bagian (table 1), yaitu data client, related with the last contact of the current campaign, other atribut, dan social and economic context attributes.

Tabel 1. Ringkasan Dataset

| No | Attribute Name | Attribut Type |
|---|----------------|---------------|
| Data Client | | |
| 1 | Age | Numeric |
| 2 | Job | Categorical |
| 3 | Marital | Categorical |
| 4 | Education | Categorical |
| 5 | Default | Categorical |
| 6 | Housing | Categorical |
| 7 | Loan | Categorical |
| Related With The Last Contact Of The Current Campaign | | |
| 8 | Contact | Categorical |
| 9 | Month | Categorical |
| 10 | Day_of_week | Categorical |
| 11 | Duration | Numeric |
| Other Attributes | | |
| 12 | Campaign | Numeric |
| 13 | Pdays | Numeric |
| 14 | Previous | Numeric |
| 15 | Poutcome | Categorical |
| Social And Economic Context Attributes | | |
| 16 | Emp.var.rate | Numeric |
| 17 | Cons.price.idx | Numeric |
| 18 | cons.conf.idx | Numeric |
| 19 | Euriborn3m | Numeric |
| 20 | Nr.employed | Numeric |
| Output | | |
| 21 | y | Binary |

Dataset yang digunakan akan diuji dengan berbagai pemodelan klasifikasi dengan teknik K-Fold Validation (10). Dimana proses pengujian dilakukan dengan menguji pendekatan level data SMOTE dengan ensemble AdaBoost yang dipadukan dengan algoritma Naïve Bayes, SVM, dan Decision Tree.

Sebagai evaluasi model dilakukan dengan menggunakan akurasi dari model pengklasifikasian yang diajukan. Kemudian

hasil dari masing – masing pengujian disajikan untuk melihat hasilnya dan menentukan hasil yang terbaik. Dari hasil pengujian yang terbaik, akan dibandingkan dengan hasil dari model pada penelitian sebelumnya (Lawi, Velayaty, & Zainuddin, 2018).

3. Hasil dan Pembahasan

Penelitian dilakukan dengan menggunakan laptop dengan processor Intel Core i5-2520M CPU @2.50GHz dengan RAM 4GB, dan system operasi Windos 7 Professional 64bit. Sedangkan dalam melakukan eksperimen didukung dengan tools WEKA 3.8.1.

Penelitian ini menggunakan dataset public yang diambil dari UCI Machine Learning Repository tentang pemasaran langsung yang dilakukan oleh bank di Portugis dengan 41.188 data yang terdiri dari 20 atribut dan 2 kelas. dataset ini berkaitan dengan pemasaran secara langsung yang dilakukan oleh lembaga perbankan Portugis kepada calon pelanggannya. Pemasaran ini didasarkan pada panggilan telepon. Biasanya perlu lebih dari satu kali proses penawaran kepada calon konsumen yang sama untuk memastikan apakah akan menggunakan prosuk yang ditawarkan atau tidak.

Dalam proses penelitian, digunakan 20 feature yang terbagi menjadi 4 bagian (table 1), yaitu data client, related with the last contact of the current campaign, other atribut, dan social and economic context attributes dengan 2 kelas, dimana kelas yes terdiri dari 464 data, dan no 3.655 data.

3.1. Preprocessing Data

Preprocessing data dilakukan untuk menangani ketidak seimbangan kelas dengan menggunakan teknik SMOTE. Dimana teknik SMOTE berguna untuk menghasilkan data yang lebih baik dan efektif untuk menangani ketidak seimbangan kelas yang mengalami over-fitting pada proses teknik over-sampling untuk kelas minoritas (positif). SMOTE menciptakan sebuah contoh dari kelas minoritas sintetis yang beroperasi di ruang fitur daripada ruang data. Dengan menduplikasi contoh kelas minoritas, teknik SMOTE menghasilkan contoh sintetis baru dengan melakukan ekstrapolasi sampel minoritas yang ada dengan sampel acak

yang diperoleh dari nilai k tetangga terdekat. Dengan hasil sintesis pada contoh yang lebih dari kelompok minoritas, sehingga mampu memperluas area keputusan mereka untuk minoritas.

Prinsip dari Metode SMOTE adalah dengan menambah jumlah data kelas minor agar setara dengan kelas mayor dengan cara membangkitkan data buatan. Data buatan atau sintesis tersebut dibuat berdasarkan k-tetangga terdekat (k-nearest neighbor) (Terhadap, Tidak, Pada, Model, & Jamu, 2013). Jumlah k-tetangga terdekat ditentukan dengan mempertimbangkan kemudahan dalam melaksanakannya. Pembangkitan data buatan yang berskala numerik berbeda dengan kategorik. Data numerik diukur jarak kedekatannya dengan jarak Euclidean sedangkan data kategorik lebih sederhana yaitu dengan nilai modus. Perhitungan jarak antar contoh kelas minor yang peubahnya berskala kategorik dilakukan dengan rumus Value Difference Metric (VDM) yaitu :

$$\Delta(X, Y) = W_x W_y \sum_{i=1}^N \delta(X_i, Y_i)^r \dots\dots\dots (1)$$

Dengan :

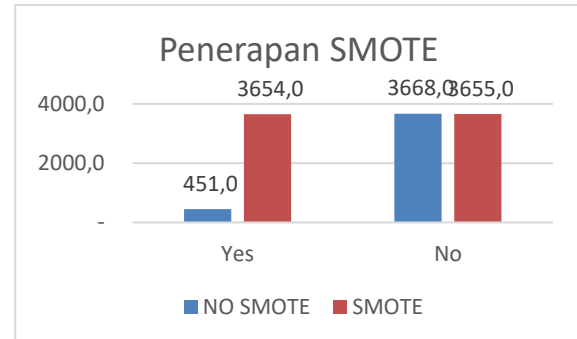
- $\Delta(X, Y)$: Jarak antara amatan X dengan Y
- $W_x W_y$: Bobot amatan (dapat diabaikan)
- N : Banyaknya peubah penjelas
- R : Bernilai 1 (jarak Manhattan) atau 2 (jarak Euclidean)
- $\delta(X_i, Y_i)^r$: Jarak antar kategori, dengan rumus

$$\delta(V_1, V_2) = \sum_{i=1}^n \left| \frac{c_{1i}}{c_1} - \frac{c_{2i}}{c_2} \right|^k \dots\dots\dots (2)$$

Dengan :

- $\delta(V_1, V_2)$: Jarak antara nilai V1 dan V2
- C1i : Banyaknya V1 yang termasuk kelas i
- C2i : Banyaknya V2 yang termasuk kelas i
- I : Banyaknya kelas; I = 1, 2, ..., m
- C1 : Banyaknya nilai 1 terjadi
- C2 : Banyaknya nilai 2 terjadi
- N : Banyaknya kategori
- K : Konstanta (biasanya 1)

Dengan meneapkan SMOTE, meminimalisir ketidak seimbangan kelas dengna membangkitkan data sintesis kelas minoritas dari 451 menadi 3.654 (gambar 2)



Gambar 2. Penerapan SMOTE

3.2. Eksperimen dan Pengujian Model

3.2.1. Metode Ensemble Adaptive Boosting (AdaBoost)

Boosting merupakan pendekatan pada machine learning untuk meningkatkan peraturan prediksi yang akurat dengan menggabungkan banyak peraturan yang relatif lemah dan tidak akurat (Nurzahputra & Muslim, 2017). Adaptive boosting (adaboost) merupakan salah satu dari beberapa varian pada algoritma boosting (Liu, Tian, Li, & Zhang, 2015). AdaBoost menjadi ensemble learning yang sering digunakan pada algoritma AdaBoost. AdaBoost dan variannya telah berhasil diterapkan pada beberapa bidang karena dasar teorinya yang kuat, prediksi yang akurat, dan kesederhanaan yang besar dengan langkah - langkah sebagai berikut :

- a. Input : Suatu kumpulan kumpulan sample penelitian dengan lable $\{(x_i, y_i), \dots, (X_n, X_n)\}$, suatu component learn algoritma dengan jumlah perputaran T
- b. Initialize : Bobot suatu sample pelatihan $W_i^1 = 1/N$, untuk semua $i = 1, \dots, N$
- c. Do for $t=1, \dots, T$
 1. Gunakan component learn algoritma untuk melatih suatu komponen klasifikasi h_t , pada sample bobot pelatihan.
 2. Hitung kesalahan pelatihannya pada $h_t: \varepsilon_t = \sum_{i=1}^N w_i^t, y_i \neq h_t(x_i)$
 3. Tetapkan bobot untuk component classifier $h_t == a_t = \frac{1}{2} \ln \left(\frac{1-\varepsilon_t}{\varepsilon_t} \right)$

4. Update bobot sample pelatihan

$$w_i^{t+1} = \frac{w_i^t \exp\{-a_t y_t h_t(x_i)\}}{c_t}, i = 1, \dots, N$$

C_t adalah suatu konstanta normalisasi

$$\text{Output } f(x) = \text{sign}\left(\sum_{t=1}^T a_t h_t(x)\right)$$

3.2.2. Algoritma Naïve Bayes

Naïve bayes merupakan teknik probabilistik klasifikasi berdasarkan teorema Bayes dengan asumsi independensi diantara variabel predictor (Anam & Santoso, 2018). Secara sederhana, pengelompokan Naïve Bayes menganggap adanya suatu fitur tertentu dalam sebuah kelas tidak terkait dengan adanya fitur lainnya. Teorema Bayes memberikan cara untuk menghitung posterior probability $P(C|X)$ dari $P(C)$, $P(X)$ dan $P(X|C)$ dengan persamaan berikut:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \dots\dots\dots (3)$$

Dengan :

$P(C|X)$: Posterior Probability kelas (C, target) yang diberikan predictor (X, atribut)

$P(C)$: Probabilitas kelas sebelumnya

$P(X|C)$: Kemungkinan yang merupakan probabilitas predictor kelas yang diberikan

$P(X)$: Probabilitas predictor sebelumnya

3.2.3. Algoritma SVM

Support vector machine (SVM) adalah classifier biner yang secara linier memisahkan kedua kelas dengan menemukan margin maksimum yang memisahkan bidang hyper antara dua kelas (Lin, Yang, & Zuo, 2010). Hyper plane pemisah margin maksimum ditemukan dengan menyelesaikan masalah optimisasi quadratic programming (QP). Keuntungan terbesar SVM datang ketika data dipisahkan secara nonlinier. Dalam hal ini, SVM membuat data dapat dipisahkan secara linear dengan bantuan fungsi kernel. Fungsi kernel adalah pemetaan pola input data ke beberapa ruang dimensi tinggi sehingga titik data menjadi terpisah secara linear. Dalam praktiknya, tidak mendefinisikan pemetaan titik-titik data

secara implisit, melainkan secara eksplisit didefinisikan sebagai produk dalam antara titik-titik data menurut dipisahkan dalam ruang dimesi tinggi. Berbagai jenis fungsi kernel dapat digunakan sebagai Gaussian kernel, polynomial kernel dll. Makalah ini kami ambil dari kernel polynomial. Polynomial kernel didefinisikan sebagai

$$k(\vec{x}, \vec{z}) = (\vec{z}^T \vec{x} + C)^n \dots\dots\dots (4)$$

Dimana n adalah urutan dari kernel, dan c adalah konstanta yang memungkinkan untuk menukarkan pengaruh urutan yang lebih tinggi dan urutan yang lebih rendah.

3.2.4. Algoritma Decision Tree

Decision tree dengan algoritma J48 merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (tree) dimana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas (Ardiansyah Sembiring, Fitri Larasati Sibuea, Sapta, Studi Sistem Informasi, & Royal, 2018). Node yang paling atas dari decision tree disebut sebagai root. Dimana dilakukan langkah – langkah sebagai berikut untuk membangun pohon keputusan ;

1. Membentuk decision system yang terdiri dari atribut kondisi dan atribut keputusan. Memperlihatkan sebuah contoh decision system yang pada penelitian ini. Ia hanyaterdiri dari n objek, $E_1, E_2, E_3, E_4, \dots, E_n$ dan attribute kondisi yaitu penjualan, pembelian, stok gudang, serta beban usaha. Sedangkan profit adalah decision attribute
2. Hitung jumlah data kolom, jumlah data berdasarkan anggota atribut hasil dengan syarat tertentu. Untuk proses pertama syaratnya masih kosong
3. Pilih atribut sebagai Node
4. Buat cabang untuk tiap tiap anggota dari Node
5. Periksa apakah nilai entropy dari anggota Node ada yang bernilai nol. Jika ada, tentukan daun yang terbentuk. Jika seluruh nilai entropy
6. Jika ada anggota Node yang memiliki nilai entropy lebih besar dari nol, ulangi lagi proses dari awal dengan Node sebagai syarat sampai semua anggota dari Node bernilai nol

Node menjadi atribut dengan nilai gain tertinggi dari atribut yang ada. Untuk menghitung nilai gain suatu atribut digunakan rumus :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * Entropy(S_i) \dots\dots\dots (5)$$

- Dengan :
- S : Himpunan kasus
 - A : Atribut
 - N : Jumlah partisi atribut A
 - |Si| : Proporsi Si terhadap S
 - |S| : Jumlah kasus dalam S

Sementara untuk menghitung nilai Entropy dengan persamaan :

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots\dots\dots (6)$$

- Dengan :
- S : Himpunan Kasus
 - N : Jumlah partisi S
 - Pi : Proporsi Si terhadap S

3.3. Evaluasi Hasil Kinerja Model Usulan

Dalam pengevaluasian model digunakan k-fold validation (10 fold) dan confusion matrix dengan membandingkan hasil klasifikasi yang dilakukan oleh system dengan hasil klasifikasi yang sebenarnya. Pengukuran akurasi dengan confusion matrix dapat dilihat pada table 1 (Bisri, 2015).

Table 2. Confusion Matrix

| Class Aktual | Class Prediksi | | |
|--------------|----------------|-----|----|
| | | Yes | no |
| Yes | TP | FN | |
| No | FP | TN | |

Dengan formula penghitungan sebagai berikut :

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Sensitivity = Recall = TPRate = \frac{TP}{TP + FN}$$

$$Specificity = TNRate = \frac{TN}{FP + TN}$$

$$FPRare = \frac{FP}{FP + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$G - mean = \sqrt{Sensitivity \times Specificity}$$

$$F - Measure = \frac{2 \times Recall \times Precision}{(Recall + Precision)}$$

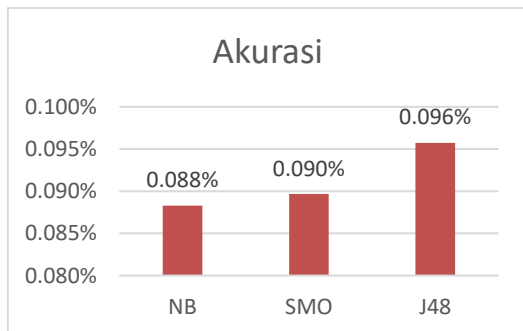
$$AUC = \frac{1 + TPRate + FPRate}{2}$$

Pengujian dilakukan dengan menguji pendekatan level data SMOTE dengan ensemble AdaBoost yang dipadukan dengan algoritma Naïve Bayes, SVM, dan Decision Tree (table 3).

Table 3. Hasil Pengujian

| No | Kriteria | NB | SMO | J48 |
|----|---|------------|--------------|------------|
| 1 | Akurasi | 88,3 0% | 89,6 8% | 95,7 3% |
| 2 | Komputasi | 7,38 | 1.07 4,06 | 886, 79 |
| 3 | SENSITIVIT Y = RECALL = TPrate | 86,3 8% | 88,4 5% | 94,8 0% |
| 4 | SPECIFITY = TNrate | 90,2 2% | 90,9 1% | 96,6 6% |
| 5 | FPrate | 9,78 % | 9,09 % | 3,34 2% |
| 6 | Precision | 89,8 4% | 90,6 9% | 96,6 0% |
| 7 | G-Mean | 88,2 8% | 89,6 7% | 95,7 2% |
| 8 | F-Measure | 88,0 7% | 89,5 6% | 95,6 9% |
| 9 | AUC | 88,3 0% | 89,6 8% | 95,7 3% |

Dari hasil klasifikasi dengan mengintegrasikan pendekatan level data SMOTE dan ensemble AdaBoost dari sample dataset bank-marketing diketahui bahwa algoritma J48 memiliki nilai akurasi (gambar 3) dan presisi terbaik dari algoritma SVM dan Naïve Bayes. Tetapi dalam melakukan komputasi, algoritma SVM paling lama, dan Naïve Bayes paling cepat.



Gambar 3. Perbandingan akurasi

3.4. Perbandingan Dengan Penelitian Terkait

Hasil perbandingan dengan penelitian sebelumnya menunjukkan bahwa kinerja dari model yang diusulkan memiliki nilai akurasi lebih baik. Dimana pada model yang diusulkan dengan menggunakan SMOTE + AdaBoost + J48 menunjukkan nilai akurasi yang lebih baik. Hasil penelitian ini selanjutnya dapat dijadikan bahan referensi bagi penelitian selanjutnya.

Table 4. Perbandingan Penelitian Terkait

| No | Model | Akurasi |
|----|---|---------------|
| 1 | SMOTE + Rotation Forest (PCA)-J48 (Ruangthong & Jaiyen, 2015) | 90,81% |
| 2 | partitioning based method - C4.5 DT (Mitik, Korkmaz, Karagoz, Toroslu, & Yucel, 2017) | 74,61% |
| 3 | MLPNN, C4.5, Regresi Logistik and RF (Asare-Frempong & Jayabalan, 2017) | 86,8%, |
| 4 | SMOTE + J48 (Perbankan & No, 2016) | 93,21% |
| 5 | SMOTE + AdaBoost + J48 | 95,73% |

4. Kesimpulan

Penelitian dilakukan dengan tools WEKA pada dataset pemasaran langsung pada bank untuk membandingkan kinerja dari algoritma Naïve Bayes, SVM, dan Decision Tree dengan pendekatan level data SMOTE dan ensemble AdaBoost. Hasil penelitian menunjukkan bahwa klasifikasi dengan J48 memiliki kemampuan

prediksi lebih baik dari Naïve Bayes dan SVM. Metode yang diusulkan dengan pendekatan level data SMOTE dan ensemble AdaBoost dengan algoritma J48 memiliki kemampuan lebih baik dari PCA J48 dan SMOTE dengan SVM.

Referensi

- Abbas, S. (2015). Deposit subscribe Prediction using Data Mining Techniques based Real Marketing Dataset. *International Journal of Computer Applications*, 110(3), 1–7. <https://doi.org/10.5120/19293-0725>
- Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *Jurnal Ilmiah Ilmu-Ilmu Teknik*, 8(1), 13–19.
- Ardiansyah Sembiring, M., Fitri Larasati Sibuea, M., Sapta, A., Studi Sistem Informasi, P., & Royal, S. (2018). Analisa Kinerja Algoritma C.45 Dalam Memprediksi Hasil Belajar. *Journal of Science and Social Research*, 1(February), 73–79. Retrieved from <http://jurnal.goretanpena.com/index.php/JSSR>
- Asare-Frempong, J., & Jayabalan, M. (2017). Predicting customer response to bank direct telemarketing campaign. *2017 International Conference on Engineering Technology and Technopreneurship, ICE2T 2017, 2017-Janua*, 1–4. <https://doi.org/10.1109/ICE2T.2017.8215961>
- Bisri, A. (2015). *Penerapan Adaboost untuk Penyelesaian Ketidakseimbangan Kelas pada Penentuan Kelulusan Mahasiswa dengan Metode Decision Tree*. 1(1).
- Lawi, A., Velayaty, A. A., & Zainuddin, Z. (2018). On identifying potential direct marketing consumers using adaptive boosted support vector machine. *Proceedings of the 2017 4th International Conference on Computer Applications and Information Processing Technology, CAIPT 2017, 2018-Janua*, 1–4. <https://doi.org/10.1109/CAIPT.2017.8320691>
- Lin, L., Yang, S., & Zuo, R. (2010). Protein secondary structure prediction based on multi-SVM ensemble. *Proceedings*

- of 2010 International Conference on Intelligent Control and Information Processing, ICICIP 2010, (PART 2), 356–358.
<https://doi.org/10.1109/ICICIP.2010.5564201>
- Liu, H., Tian, H. Q., Li, Y. F., & Zhang, L. (2015). Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions. *Energy Conversion and Management*, 92, 67–81.
<https://doi.org/10.1016/j.enconman.2014.12.053>
- Mitik, M., Korkmaz, O., Karagoz, P., Toroslu, I. H., & Yucel, F. (2017). Data Mining Based Product Marketing Technique for Banking Products. *IEEE International Conference on Data Mining Workshops, ICDMW*, 552–559.
<https://doi.org/10.1109/ICDMW.2016.0085>
- Moro, S., Laureano, R. M. S., & Cortez, P. (2011). Using Data Mining for Bank Direct Marketing: An application of the CRISP-DM methodology. *25th European Simulation and Modelling Conference- ESM'2011*, (Figure 1), 117–121.
- Nurzahputra, A., & Muslim, M. A. (2017). Peningkatan Akurasi Pada Algoritma C4.5 Menggunakan Adaboost Untuk Meminimalkan Resiko Kredit. *Prosiding SNATIF Ke-4*, (1), 243–247.
- Perbankan, I., & No, J. G. (2016). *Metode Klasifikasi Data Mining dan Teknik Sampling Smote ... (Hairani dkk.)*. 168–172.
- Putri, S. A. (2015). *Integrasi SMOTE dan Information Gain pada Naive Bayes untuk Prediksi Cacat Software*. 1(2).
- Ruangthong, P., & Jaiyen, S. (2015). Bank direct marketing analysis of asymmetric information based on machine learning. *Proceedings of the 2015 12th International Joint Conference on Computer Science and Software Engineering, JCSSE 2015*, 93–96.
<https://doi.org/10.1109/JCSSE.2015.7219777>
- Terhadap, S., Tidak, D., Pada, S., Model, P., & Jamu, K. (2013). *PENERAPAN SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE*. 1(1).