

Komparasi Adaboost dan Bagging Dengan Naïve Bayes Pada Dataset Bank Direct Marketing

Rousyati¹, Amin Nur Rais², Noor Hasan³, Richky Faizal Amir⁴, Warjiyono⁵

Universitas Bina Sarana Informatika

rousyati.rou@bsi.ac.id¹, amin.arv@bsi.ac.id², noor.nhs@bsi.ac.id³, richky.cfi@bsi.ac.id⁴,
warjiyono.wrj@bsi.ac.id⁵

Abstrak - Kampanye pemasaran produk bank secara langsung memiliki kegunaan bagi marketer untuk menawarkan produk baru kepada calon pelanggan yang ditargetkan. Dengan menggunakan data yang sudah ada yang bersumber dari pengalaman melakukan pemasaran dapat digunakan sebagai data yang akan diolah untuk membuat keputusan dimasa depan. Dengan menggunakan algoritma klasifikasi naïve bayes yang digabungkan dengan ensemble adaboost dan bagging, dilakukan percobaan untuk mendapatkan hasil komparasi akurasi dan presisi terbaik. Penelitian ini dilakukan dengan melakukan 4 kali percobaan dengan skema hanya dengna algoritma naïve bayes, naïve bayes dengan adaboost, naïve bayes dengan bagging, dan naïve bayes dengan adaboost dan bagging. Hasil penelitian menunjukkan bahwa penggunaan algoritma naïve bayes dengan adaboost dan bagging menghasilkan nilai akurasi terbaik sebesar 90,29%. Sedangkan pada presisi, diketahui bahwa nilai presisi terbaik pada percobaan algoritma naïve bayes saja dan naïve bayes dengna bagging dengan nilai presisi yang sama sebesar 94,99%.

Kata Kunci : *adaboost, bagging, naïve bayes, imbalance class*

Abstract - *Direct bank product marketing campaigns have uses for marketers to offer new products to targeted potential customers. By using existing data that comes from experience doing marketing can be used as data that will be processed to make decisions in the future. By using the naïve Bayes classification algorithm combined with the adaboost ensemble and bagging, experiments were carried out to get the best accuracy and precision comparison results. This research was conducted by conducting 4 experiments with the scheme only with the naïve bayes algorithm, naïve bayes with adaboost, naïve bayes with bagging, and naïve bayes with adaboost and bagging. The results showed that the use of the naïve Bayes algorithm with adaboost and bagging resulted in the best accuracy value of 90.29%. While for precision, it is known that the best precision value is in the naïve Bayes algorithm experiment only and naïve Bayes with bagging with the same precision value of 94.99%.*

Keywords: *adaboost, bagging, naïve bayes, imbalance class*

I. PENDAHULUAN

Kampanye pemasaran produk bank secara langsung memiliki kegunaan bagi *marketer* untuk menawarkan produk baru kepada calon pelanggan yang ditargetkan. Dengan melakukan analisa data pada data data sebelumnya tentang pemasaran bank, dapat digunakan untuk memilih jenis pemasaran yang dilakukan yang efektif. Kampanye pemasaran dapat dilakukan melalui baik melalui telepon, email, maupun email langsung kepada calon pelanggan yang memungkinkan calon pelanggan dapat memutuskan untuk mengambil produk yang ditawarkan atau tidak sesuai kriteria yang dibutuhkan. Proses analisa data dapat dilakukan dengan teknologi informasi yang didukung dengan terus bertambahnya informasi data pelanggan untuk mendukung sehingga dalam pengambilan keputusan dapat dilakukan lebih akurat (Moro et al., 2011). Mengenalkan produk secara langsung telah banyak digunakan oleh industri, seperti pada industri perbankan. Proses pengenalan produk secara langsung, bank dapat melakukan analisa pasar dengan memanfaatkan data data sebelumnya yang

dikombinasikan dengan ruang teknologi informasi yang dapat membantu dalam mengambil keputusan (Abbas, 2015).

Semakin bertambahnya waktu, jumlah data yang masuk dan dapat diolah terus bertambah. Data yang ada memberikan informasi yang merepresentasikan tentang karakter konsumen yang dapat berguna untuk membuat keputusan di masa mendatang. Akan tetapi, terjadi ketimpangan respon dari konsumen terhadap penawaran yang dilakukan. Sehingga terjadi perbandingan yang tinggi antara konsumen yang menerima penawaran dan yang menolaknya. Permasalahan hasil respon yang memiliki perbedaan yang banyak menyebabkan hasil klasifikasi lebih condong ke haisl mayoritas. Dimana jumlah mayoritas menolak lebih besar dari hasil yang menerima yang dapat menurunkan kinerja model prediksi.

Dalam banyak tugas klasifikasi, terdapat kelas data yang tidak merata sehingga masalah ini memiliki dampak potensi untuk mengakibatkan kecenderungan hasil pelatihan ke kelas mayoritas (Aridas et al., 2020). Dalam penelitiannya, teknik klasifikasi dengan algoritma naïve bayes digunakan untuk memilih

contoh paling informative dari dataset yang tersedia secara acak (Aridas et al., 2020). Dengan menggunakan naïve bayes dalam pengklasifikasian data mining penentuan penerimaan kartu Indonesia sehat diperoleh hasil bahwa menggunakan algoritma naïve bayes memiliki nilai hasil yang lebih baik dari sisi presisi (Y. I. Kurniawan, 2018). 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.

Pada penelitian *credit risk*, algoritma *support vector machine* (svm) diusulkan oleh defri dkk sebagai algoritma data mining (D. Kurniawan & Supriyanto, 2013). Dalam penelitiannya, terdapat kesulitan dalam menentukan parameter ideal dan mengusulkan penggunaan algoritma adaboost untuk mengatasinya. 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) Hasil penelitiannya mendapatkan hasil bahwa algoritma adaboost dapat meningkatkan nilai akurasi yang lebih baik. Algoritma adaboost juga digunakan dengan mengkombinasikan dengan algoritma c4.5 untuk meningkatkan nilai akurasi. Hasilnya mengungkapkan bahwa adaboost jika diterapkan dengan algoritma c4.5 juga dapat meningkatkan nilai akurasi (Nurzahputra & Muslim, 2017).

Algoritma bagging digunakan untuk menghasilkan sebuah nilai prediksi dari hasil agregasi beberapa predictor (Zulhanif, 2018). Algoritma ensemble bagging digunakan untuk mengatasi kelemahan data yang memiliki kelas tidak seimbang dan memiliki dimensi yang tinggi sehingga menurunkan kinerja dari model prediksi. Dalam melakukan prediksinya, algoritma ensemble bagging yang dikombinasi dengan greedy forward dan naivebayes menjadi model terbaik (Fitriyani & Wahono, 2015).

Oleh sebab itu, dalam penelitian ini akan dilakukan komparasi penggunaan ensemble adaboost dan bagging dengan algoritma naïve bayes untuk mengatasi imbalance class pada dataset bank direct marketing.

II. METODOLOGI PENELITIAN

Pendekatan pada penelitian ini menggunakan pendekatan kuantitatif atau disebut juga metode discovery. Dengan pendekatan kuantitatif, dapat memberikan penekanan pengukuran sebagai dasarnya, karena memberikan hubungan formalisasi model dan

observasi, teori, dan hipotesis. Pada pendekatan kuantitatif juga dapat ditemukan dan dikembangkan informasi dan pengetahuan baru. Disebut metode kuantitatif karena data penelitian berupa angka-angka dan analisis menggunakan statistik (Sugiyono, 2016). Dalam melakukan penelitian, dilakukan tahap – tahap dengan kerangka untuk melakukan penelitian yang dilakukan sebagaimana digambarkan pada gambar 1 sebagai acuan penelitian.



Gambar 1. Kerangka Penelitian

Objek penelitian ini menggunakan public dataset yang berasal dari UCI Machine Learning Dataset dengan menggunakan dataset data bank direct marketing pada halaman

<https://archive.ics.uci.edu/ml/datasets/bank+marketing>. Dataset ini berkaitan dengan kampanye pemasaran langsung dari lembaga perbankan Portugis, dimana model pemasarannya berdasarkan pada panggilan telepon. Dalam pemasarannya, seringkali menghubungi kontak klien yang sama lebih dari satu kali untuk mengakses produknya akan berlangganan atau tidak berlangganan.

Tabel 1. Spesifikasi Dan Atribut Dataset

No	Attribute Name	Attribute Type
1	Age	Numeric
2	Job	Categorical
3	Marital	Categorical
4	Education	Categorical
5	Default	Categorical
6	Housing	Categorical
7	Loan	Categorical
8	Contact	Categorical
9	Month	Categorical
10	Day_of_week	Categorical
11	Duration	Numeric
12	Campaign	Numeric
13	Pdays	Numeric
14	Previous	Numeric
15	Poutcome	Categorical
16	Emp.var.rate	Numeric
17	Cons.price.idx	Numeric
18	cons.conf.idx	Numeric
19	Euriborn3m	Numeric
20	Nr.employed	Numeric
21	Y	Binary

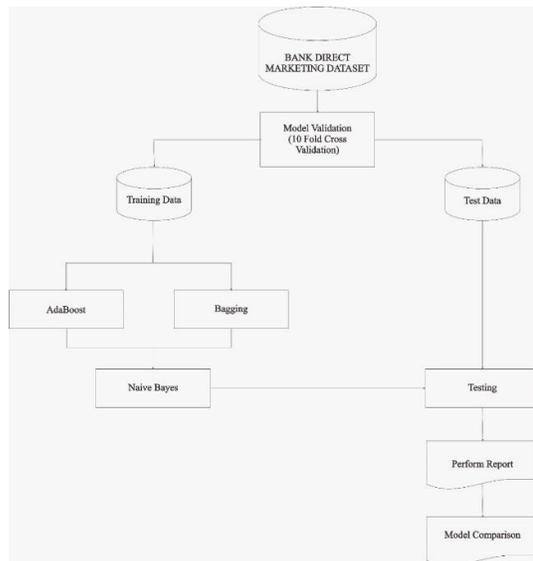
Eksperimen model penelitian mengacu pada gambar 2 menggunakan ensemble Adaboost dan bagging dengan algoritma Naïve Bayes. Pada eksperimen penelitian ini dilakukan dengan proses uji coba yang dapat dilakukan

dengan berbagai percobaan, mulai dari percobaan langsung dengan algoritma naïve bayes, adaboost, atau bagging dan dengan percobaan kombinasi dari algoritma – algoritma tersebut.

Hasil proses pengujian akan menghasilkan table confusion matrix yang nantinya dapat digunakan sebagai bahan evaluasi percobaan untuk melihat hasil komparasi percobaan kombinasi algoritmanya. Komparasi yang dilakukan pada penelitian ini dengan melihat hasil perhitungan confusion matrix pada nilai akurasi dan presisi nya.

III. HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan eksperimen untuk melakukan pengujian terhadap model yang diusulkan. Penelitian ini menggunakan komputer dengan spesifikasi perangkat keras prosesor Intel(R) Core (TM) i5-6200U CPU @ 2.30 GHz, memori (RAM) 16,00 GB, monitor standar dan menggunakan sistem operasi Windows 10 Profesional 64bit. Dan Aplikasi mechine learning yang digunakan adalah WEKA versi 3.8 untuk software data mining. Penelitian ini menggunakan dataset bank direct marketing yang bersifat public dari uci machine learning (<https://archive.ics.uci.edu/ml/datasets/bank+marketing>). Dataset bank direct marketing yang digunakan pada penelitian ini dengan menggunakan model eksperimen akan menghasilkan nilai akurasi dan presisi.



Gambar 2. Eksperimen Model

Berdasarkan model yang diusulkan, terdapat proses eksperimen pengolahan dataset dengan menggunakan AdaBoost, Bagging, dan tanpa penggunaan keduanya. Penggunaan algoritma AdaBoost untuk melakukan boosting kinerja klasifikasi dari

sebuah algoritma pembelajaran. Sedangkan Bagging bekerja dengan cara melakukan replikasi bootstrap terhadap peubah penjelas secara bersama sama. Dan algoritma naïve bayes menjadi metode klasifikasi yang berakar pada teorema bayes dimana mengansumsikan bahwa atribut objek bersifat independedn. Dengan menggunakan algoritma naïve bayes, setiap kelas keputusan akan dihitung probabilitasnya dengan syarat bahwa kelas keputusan bernilai benar.

Hasil pengujian ini akan dievaluasi menggunakan cross validation dengan membandingkan algoritma pembelajaran (learning algorithms) dengan membagi datanya menjadi dua bagian, bagian pertama digunakan sebagai data latih, dan bagian kedua akan digunakan sebagai data validasi. Untuk memperkirakan kinerja pengklasifikasian, cross validation mengumpulkan pelatihan dan validasi yang bersifat crossover berturut turut sehingga setiap data memiliki kesempatan tervalidasi. K-fold cross validation dilakukan dengan menggunakan kembali dataset yang sama, sehingga menghasilkan k perpecahan dari kumpulan data menjadi non-overlapping dengan proporsi pelatihan (k-1)/k dan 1/k untuk pengujian.

Data Set									
Split 1	Split 2	Split 3	Split 4	Split 5	Split 6	Split 7	Split 8	Split 9	Split 10
Test	Training								
Training	Test	Training							
Training		Test	Training						
Training			Test	Training					
Training				Test	Training				
Training					Test	Training			
Training						Test	Training		
Training							Test	Training	
Training								Test	Training
Training									Test

Gambar 3. Pembagian Dataset untuk 10-Fold Cross Validation

Kinerja model yang diusulkan menggunakan dataset bank direct marketing diukur berdasarkan akurasi dan presisi yang dihasilkan. Akurasi hasil pengujian dapat dilihat dengan membandingkan hasil penjumlahan antara true positive dan true negative dengan jumlah nilai confusion matrix yang dihasilkan. Sedangkan presisi dapat dilihat dari nilai true positif dibagi dengan jumlah true positive dan false positive.

Dengan menggunakan confusion matrix, dapat dilakukan analisa seberapa baik hasil klasifikasi dapat mengenali fitur dari kelas yang berbeda

dan menunjukkan rincian kinerja pengklasifikasian dengan memberikan informasi jumlah fitur suatu kelas yang diklasifikasikan dengan tepat dan tidak tepat. Confusion matriks digambarkan dengan matriks 2 dimensi yang merepresentasikan hasil prediksi dengan kenyataan, dan memberikan hasil penilaian kinerja model klasifikasi berdasarkan jumlah objek yang diprediksi dengan benar dan salah.

Tabel 2. Confusion Matrix

Kelas		Nilai Sebenarnya	
		Benar	Salah
Nilai Prediksi	Benar	TP (True Positive)	FP (False Positive)
	Salah	FN (False negative)	TN (True Negative)

Akurasi = $(TP+TN) / (TP+TN+FP+FN)$
 Precision /PPV = $TP / (TP+FP)$

Teknik analisis dilakukan dengan praktek uji coba dilakukan untuk membandingkan kinerja semua model skema yang di kembangkan dengan ensemble AdaBoost dan Bagging dengan pengklasifikasi Naïve Bayes. Analisis ini ditujukan untuk mengetahui model mana yang terbaik untuk menghasilkan nilai akurasi dan presisi yang dihasilkan dari imbalance dataset bank direct marketing. Pada percobaan penitilian ini dibagi menjadi 4 percobaan, percobaan tanpa ensemble, percobaan ensemble adaboost dengan algoritma naïve bayes, percobaan ensemble bagging dengan algoritma naïve bayes, dan percobaan dengan penggabungan ensemble adaboost dan bagging dengan algoritma naïve.

Percobaan pertama dilakukan dengan menggunakan model tanpa ensemble, dan didapatkan hasil seperti pada table 3 dengan nilai akurasi 87,52% dan presisi 94,99%

Tabel 3. Confusion Matrix Tanpa Ensemble

Kelas		Nilai Sebenarnya	
		Benar	Salah
Nilai Prediksi	Benar	3.316	175
	Salah	339	289

Akurasi = $(TP+TN) / (TP+TN+FP+FN)$ = 0,8752
 Precision /PPV = $TP / (TP+FP)$ = 0,9499

Percobaan kedua dilakukan dengan menggunakan model algoritma naïve bayes yang digabungkan dengan ensemble adaboost, dan didapatkan hasil seperti pada table 4 dengan nilai akurasi 89,88% dan presisi 94,23%

Tabel 4. Confusion Matrix Algoritma Naïve Bayes Dengan Ensemble AdaBoost

Kelas		Nilai Sebenarnya	
		Benar	Salah
Nilai Prediksi	Benar	3.449	211
	Salah	206	253

Akurasi = $(TP+TN) / (TP+TN+FP+FN)$ = 0,8988
 Precision /PPV = $TP / (TP+FP)$ = 0,9423

Percobaan ketiga dilakukan dengan menggunakan model algoritma naïve bayes yang digabungkan dengan ensemble bagging, dan didapatkan hasil seperti pada table 5 dengan nilai akurasi 87,59% dan presisi 94,99%

Tabel 5. Confusion Matrix Algoritma Naïve Bayes Dengan Ensemble Bagging

Kelas		Nilai Sebenarnya	
		Benar	Salah
Nilai Prediksi	Benar	3.319	175
	Salah	336	289

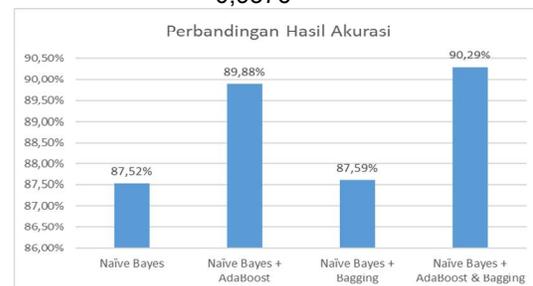
Akurasi = $(TP+TN) / (TP+TN+FP+FN)$ = 0,8759
 Precision /PPV = $TP / (TP+FP)$ = 0,9499

Percobaan ke empat dilakukan dengan menggunakan model algoritma naïve bayes yang digabungkan dengan ensemble AdaBoost dan bagging, dan didapatkan hasil seperti pada table 6 dengan nilai akurasi 90,29% dan presisi 93,76%

Tabel 6. Confusion Matrix Algoritma Naïve Bayes Dengan Ensemble AdaBoost dan Bagging

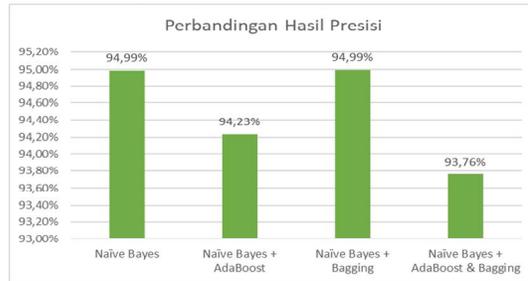
Kelas		Nilai Sebenarnya	
		Benar	Salah
Nilai Prediksi	Benar	3.487	232
	Salah	168	232

Akurasi = $(TP+TN) / (TP+TN+FP+FN)$ = 0,9029
 Precision /PPV = $TP / (TP+FP)$ = 0,9376



Gambar 4. Perbandingan Hasil Akurasi

Dari hasil percobaan pertama, kedua, ketiga, dan keempat menggunakan cross validation, nilai confusion matriks diolah untuk menghasilkan nilai akurasi dan presisi. Hasil pengolahan confusion matriks menghasilkan perbandingan nilai akurasi seperti pada gambar 4 dimana akurasi penggunaan ensemble adaboost dan bagging menghasilkan akurasi terbaik dengan nilai akurasi 90,29%.



Gambar 5. Perbandingan Hasil Presisi

Sedangkan jika melihat hasil pengolahan data confusion matrix pada gambar 5 dapat diketahui bahwa nilai presisi hasil pengujian dengan naïve bayes dan pengujian naïve bayes + bagging memiliki hasil tertinggi dengan nilai yang sama, yaitu 94,99%.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan dengan menggunakan tools WEKA pada dataset bank direct marketing dengan membandingkan kinerja ensemble adaboost dan bagging yang dikombinasikan dengan algoritma naïve bayes, dapat diketahui bahwa penggunaan ensemble baik adaboost maupun bagging dapat mempengaruhi nilai akurasi dan presisi hasil pengolahannya. Pada hasil akurasi, penggunaan bagging menghasilkan kenaikan akurasi meski nilainya kecil. Berbeda dengan penggunaan adaboost yang mendapatkan kenaikan akurasi yang lebih baik. Namun, hasil kenaikan akurasi mendapatkan nilai terbaiknya ketika ensemble adaboost dan bagging digunakan dengan algoritma naïve bayes sebesar 90,29%. Sedangkan jika dilihat dari presisi yang dihasilkan, penggunaan ensemble adaboost dan bagging yang digunakan secara bersamaan dengan algoritma naïve bayes memiliki nilai presisi terkecil. Dan nilai presisi terbesar diketahui dengan menggunakan algoritma naïve bayes

tanpa ensemble dan dengan ensemble bagging memiliki nilai presisi yang sama sebesar 94,99%.

V. 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.
- Aridas, C. K., Karlos, S., Kanas, V. G., Fazakis, N., & Kotsiantis, S. B. (2020). Uncertainty Based Under-Sampling for Learning Naive Bayes Classifiers under Imbalanced Data Sets. *IEEE Access*, 8, 2122–2133. <https://doi.org/10.1109/ACCESS.2019.2961784>
- Fitriyani, & Wahono, R. S. (2015). Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naive Bayes. *IlmuKomputer.Com Journal of Software Engineering*, 1(2), 101–108.
- Kurniawan, D., & Supriyanto, D. C. (2013). Optimasi Algoritma Support Vector Machine (Svm) Menggunakan Adaboost Untuk Penilaian Risiko Kredit. *Jurnal Teknologi Informasi*, 9(1), 1414–9999.
- Kurniawan, Y. I. (2018). Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 455. <https://doi.org/10.25126/jtiik.201854803>
- 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.
- Zulhanif. (2018). Algoritma Bagging. *Biostatistics*, 1(1), 68–72.