

**PENINGKATAN PERFORMA NAIVE BAYES (NB) BERBASIS
PARTICLE SWARM OPTIMIZATION (PSO) UNTUK KLASIFIKASI PENENTU
PENERIMAAN KARYAWAN**

Hidayatur Rakhmawati¹, Yuniarti Lestari²

^{1,2}STMik Muhammadiyah Paguyangan Brebes

Email: ¹hidarahmawati@stmikmpb.ac.id, ²yuniartilestari@stmikmpb.ac.id.

Abstrak

Karyawan merupakan unsur yang penting yang berperan bagi kesuksesan perusahaan menghasilkan sumber daya akan tetapi perlu diperoleh untuk melewati proses perekrutan yang sesuai. proses pemilihan karyawan amat dibutuhkan supaya kegiatan dalam melengkapi kriteria dan standar mutu untuk kepentingan terlaksananya sebuah target perusahaan. Oleh karena itu, perusahaan membutuhkan informasi yang benar. Kemudian manfaat untuk memperoleh calon karyawan yang dibutuhkan dengan kondisi perusahaan. Metode pembobotan dapat digunakan untuk mengatasi independensi atribut dalam algoritma NB. Nilai penelitian menunjukkan bahwa model *Naive Bayes*(NB) dan *Particle Swarm Optimization* (PSO), hasilnya *accuracy* lebih besar yaitu 94,10%, dibandingkan dengan model *Naive Bayes* 91,45%. Penelitian ini membuktikan bahwa ketika menggunakan *Particle Swarm Optimization* (PSO) memiliki *accuracy* yang lebih besar. Dengan begitu, ini bisa dijadikan model untuk diterapkan pada kasus lain, dengan dataset dan atribut yang lebih lengkap.

Kata kunci: *Penerimaan karyawan, Naive bayes, Pembobotan atribut, PSO*

Abstract

Employees are an important element that contributes to the success of the company and is a resource and will only be obtained through effective recruiting efforts. An employee selection model is needed as an action to meet qualifications and quality standards in order to achieve company goals. For this reason, companies need accurate and continuous information in order to get prospective employees who are in accordance with the company's qualifications. The weighting method can be used to address attribute independence in the NB algorithm. The research value shows that the Naive Bayes model based on Particle Swarm Optimization (PSO), has a higher accuracy of 94.10%, compared to the Naive Bayes model of 91.45%. This research proves that when using Particle Swarm Optimization (PSO) it has higher accuracy. That way, this can be used as a model to be applied to other cases, with more complete datasets and attributes.

Keywords: Employee acceptance, Naive bayes, Attribute weighting, PSO

1. PENDAHULUAN

Karyawan merupakan bagian dari hal terpenting peranan bagi kesuksesan perusahaan dan menghasilkan sumber daya dan akan tetapi perlu diperoleh untuk melewati proses perekrutan yang sesuai. proses pemilihan karyawan amat dibutuhkan supaya kegiatan dalam melengkapi kriteria dan standar mutu untuk kepentingan terlaksananya sebuah target perusahaan. Oleh karena itu, perusahaan membutuhkan informasi yang benar. Kemudian manfaat untuk memperoleh calon karyawan yang dibutuhkan dengan kondisi perusahaan [1].

Naive Bayes (NB) telah digunakan sebagai metode algoritma yang paling banyak digunakan dalam permasalahan dalam klasifikasinya. sifatnya yang sederhana, dan keefektivannya pada bidang *software defect prediction* [2] *Naive Bayes* (NB) merupakan pendekatan probabilistik berdasarkan pada asumsi bahwa atribut tidak bergantung satu sama lain dan bobotnya sama pentingnya. Namun, dalam praktiknya, atribut mungkin saling berhubungan. Dalam hal ini, asumsi tersebut dapat menyebabkan penurunan kinerja algoritma NB dan memiliki kelemahan pada seleksi atribut dan sebagai pembawaan independen dari sifat *Naive Bayes* tidak konsisten digunakan. akibatnya akan berdampak pada proses penurunan *accuracy* perhitungan [3]. Cara menangani masalah dengan salah satu cara memberikan metode pembobotan atribut agar akurasi NB meningkat.

Algoritma *Particle Swarm Optimization* (PSO) merupakan cara memaksimalkan proses kelompok. akan dioptimalkan oleh Eberhart dan Kennedy pada tahun 1995. PSO meniru sifat dari sekumpulan burung yang mencari makanan. Pada kejadian sekumpulan burung ini berguna untuk menjaga jarak yang terbaik antara makanan dengan burung tersebut dan sekelompok pasukan burung lainnya. [4]. Cara yang banyak dilakukan untuk mendapatkan potongan makanan tersebut yaitu dengan menempuh jalan burung yang jaraknya paling dekat dengan makanan tersebut. *Particle swarm optimization* dapat dipakai kalau cara untuk memaksimalkan subset fitur dan parameter searah. [5]

Cara menangani masalahnya dengan *Naive Bayes* perlu dioptimalkan menggunakan PSO, melalui cara pemberian bobot pada setiap atributnya untuk mendapatkan akurasi lebih tinggi pada data penerimaan karyawan. [6] Metode ini terbukti *accuracy* dan hasilnya dimanfaatkan buat menyelesaikan persoalan optimasi *multidimensi* dan *multiparameter* pada pembelajaran *machine learning* seperti *Neural Network*, SVM, dan algoritma klasifikasi lainnya. algoritma (PSO) dapat digunakan untuk proses pembobotan atribut untuk mengoptimalkan *accuracy Naive Bayes Classifier* [7]. Dataset mencakup fitur-fitur seperti umur, jenis kelamin, pendidikan terakhir, kelengkapan berkas, status sipil, pelamar bagian, tes psikotest, tes wawancara, tes kesehatan dan Atribut status digunakan sebagai label. Pada algoritma *Particle Swarm Optimization* diharapkan akan menambah akurasi dari *Naive Bayes Classifier*.

2. METODE PENELITIAN

2.1. Pengumpulan data

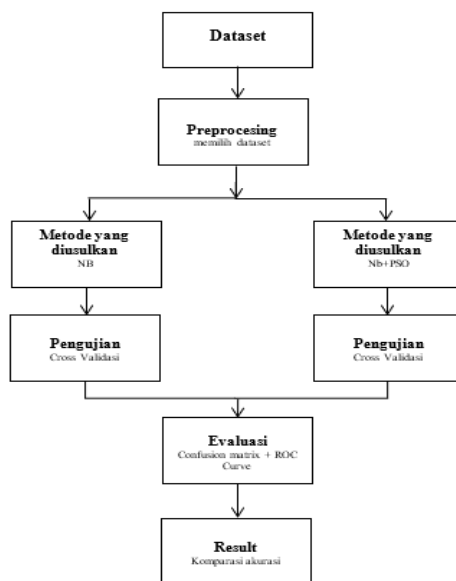
Data preparation atau yang biasa juga disebut dengan preprocessing adalah proses yang dilakukan untuk mendapatkan data yang berkualitas sebelum data tersebut diolah dan memudahkan penerapan algoritma data [8] Dataset dalam Penelitian ini terdiri dari eksperimen dengan data primer berupa dataset penerimaan karyawan, penelitian ini tujuannya akan diujikan pada proses klasifikasi yang terpilih berdasarkan label akhir lulus dan gagal dalam penerimaan karyawan. Dataset penerimaan calon karyawan yang akan digunakan adalah data dalam periode waktu 5 (lima) tahun yaitu dari tahun 2018-2022.

2.2. Pengolahan Data Awal (*Attribute selection*)

Attribute selection dalam tahap *preprocessing* data mining adalah memilih subset atribut yang relevan dari semua atribut pada dataset, kemudian atribut yang terpilih akan diproses dengan algoritma data mining [9]. tahap selanjutnya adalah pengolahan dataset penerimaan calon karyawan . proses ini diujikan dengan *analisis* tergantung *atribut* sesudah diproses selanjutnya, *atribut* yang telah diujikan pada percobaan ini berjumlah *10 atribut* yaitu umur, jenis kelamin, pendidikan terakhir, kelengkapan berkas, status sipil, pelamar bagian, tes psikotest, tes wawancara, tes kesehatan dan Atribut status hasil digunakan sebagai label. Lalu proses selanjutnya itu diproses *replace missing value*, bersumberkan model *average* ialah mengambil alih nilai yang nol beserta nilai rata-rata sehingga akan tampak datasetnya dari penerimaan calon karyawan.

2.3. Metode Yang di usulkan

Kerangka metode penelitian yang digunakan sebagai proses dari pengujian metode *Naive Bayes(NB)* dan PSO yang diproses untuk menyelesaikan permasalahan dipenelitian, pada gambar 3.1.



Gambar 3.1 Tahapan yang diusulkan

2.4. Eksperimen Metode

Tahapan dalam penelitian ini prosesnya berikut ini : yang pertama disiapkan *dataset* akan penggunaan, Setelah itu menjalankan *preprocessing* menggabungkan *naive bayes* standar dan *dataset* penerimaan karyawan ,kemudian melakukan *preprocessing* menggabungkan klasifikasi *naive bayes* yang telah dilakukan pembobotan *attribut* dengan algoritma PSO menggunakan dataset penerimaan karyawan, Melakukan pengujian menggunakan *cross validasi*, Evaluasi menggunakan Model *Confusion Matrix* + *ROC curve* dan Menggabungkan komparasi hasil pengujian *Naive bayes* standar dan hasil pengujian *naive baye* + PSO.

2.5. Evaluasi dan Hasil Validation

Keseluruhan akurasi pada pengujian dataset umumnya digunakan untuk mengevaluasi kinerja pengklasifikasi[10]. semua akurasi dan rata-rata akurasi untuk kelas minoritas. Untuk melakukan evaluasi dan validasi terhadap model yang diusulkan, maka dilakukan beberapa pengujian menggunakan *confusion matrix*. Apasaja metode yang akan di hasilkan pada pengujian ini. digunakan *cross validation* yang alan diproses untuk mengetahui tingkat akurasi, *cross validation* diproses untuk mengatasi *overlapping* didata *testing*. Proses dari *cross validation* terbagi pada pemrosesan data dari *k* subset beserta ukuran yang sama lalu digunakan tahap *subset* untuk data *testing* kemudian hasilnya data *training*. lebih tinggi nilai *accuracy* maka lebih baik juga hasil model yang digunakan . hasil dari percobaan *accuracy* yang telah dilangsungkan mendapatkan hasil metode *confusion matrix* (*Accuracy*, *Recall* dan *Precision*).

2.6. K-Fold Cross Validation

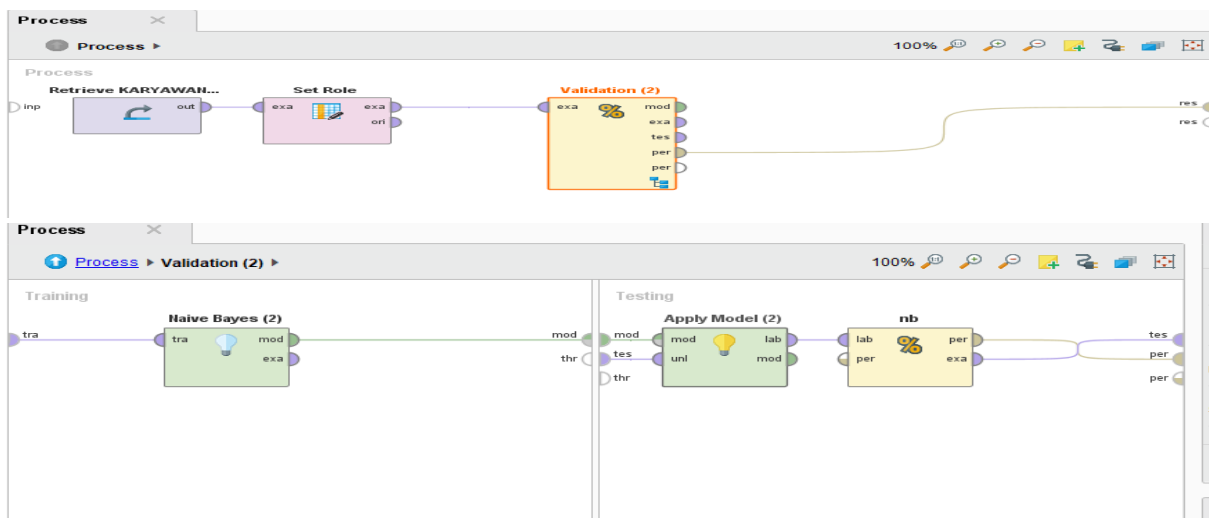
Cross validation adalah *metode statistik* untuk mengevaluasi dan membandingkan algoritma pembelajaran (*learning algorithms*) dengan membagi data menjadi dua segmen, satu segmen digunakan untuk belajar atau data latih, dan yang lain digunakan untuk memvalidasi model [11]. *K-fold cross validation* adalah teknik umum untuk memperkirakan kinerja pengklasifikasi. *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 [12].

2.7. Confusion matrix

Untuk mengukur peningkatan akurasi dari hasil masing-masing validasi, digunakan *confusion matrix*. *Confusion matrix* memberikan keputusan yang diperoleh dalam pelatihan dan pengujian [13]. *Confusion matrix* memberikan penilaian kinerja klasifikasi berdasarkan objek dengan benar atau salah [14]. *Confusion matrix* merupakan matrik 2 dimensi yang menggambarkan perbandingan antara hasil prediksi dengan kenyataan.

3. HASIL DAN PEMBAHASAN

Pada pengujian kesatu mengaplikasikan algoritma Naive Bayes terdahulu tanpa menggabungkann PSO untuk seleksi fitur. Pengujian ini digunakan jumlah data 678 *record* dan 10 *atribut*. Dengan menerapkan *10 – fold cross validation* dari data *confution matrix* yang akan diketahui hasil dari tingkat akurasi, diujikan selama 10 kali percobaan dan hasil pengujian mempunyai hasil *accuracy* yang tidak sama hasilnya dari pengambilan data secara *random* pada setiap pengujian. Bisa amati digambar 3.1.



Gambar 3.1 Model Training dan Testing Metode Naïve Bayes

Pada Tabel 3.1 menunjukkan bahwa hasil percobaan klasifikasi data penentu seleksi karyawan tersebut algoritma *Naive Bayes* menggunakan *10 fold Cross Validation* menunjukkan hasil yang tidak sama. dan pengujian mendapatkan akurasi paling tinggi berada pada pengujian yang ke-1 yaitu pada *fold* ke-2 dan nilai akurasi sebesar 90.71 %, precision 99.46 %, recall 75.45 % dan nilai AUC sebesar 0.996.

Tabel 3.1 pengujian klasifikasi Naive Bayes Standart dengan 10 kali pengujian

NO	Number of validations	Akurasi	Precision	Recall	AUC
1	2	90.71%	99.46%	75.45%	0.996
2	3	90.41%	99.44%	74.91%	0.997
3	4	90.26%	98.23%	75.83%	0.991
4	5	89.38%	98.27%	73.47%	0.991
5	6	89.68%	99.44%	73.10%	0.985
6	7	90.11%	99.47%	74.14%	0.986
7	8	89.83%	98.26%	74.13%	0.992
8	9	89.37%	98.31%	73.28%	0.991
9	10	89.68%	98.41%	73.97%	0.990
10	11	90.12%	99.43%	74.18%	0.997

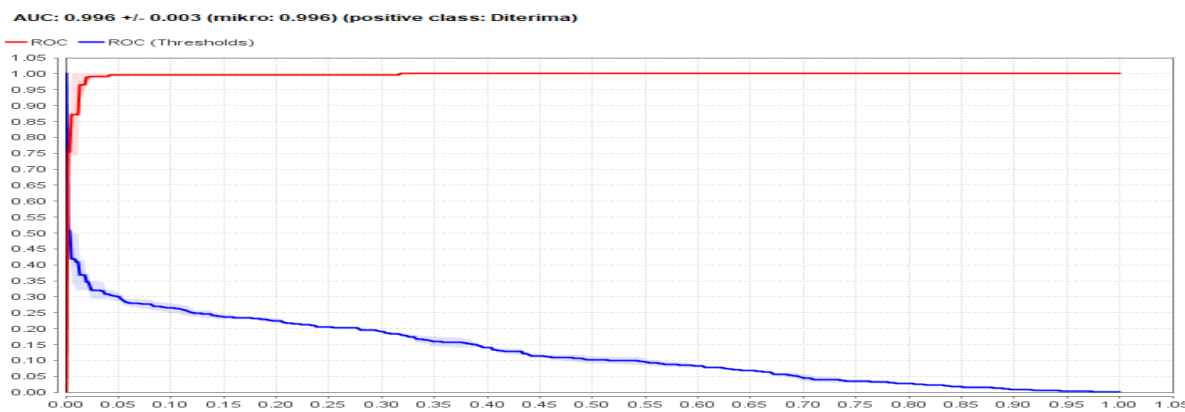
Pada gambar 3.2 Total True Positif (TP) adalah 424 record diklasifikasikan sebagai segmen penerimaan karyawan (yes/no) sesuai dengan prediksi yang dilakukan dengan metode Naive Bayes, lalu *False Negatif (FN)* sebanyak 62 record yang diprediksi sebagai segmen penerimaan karyawan (diterima/gagal) , kemudian *True Negatif (TN)* sebanyak 191 record sebagai diterima sebagai prediksi, dan *False Positif (FP)* sebanyak 1 record di prediksi diterima tetap Gagal. Tingkat akurat yang dihasilkan dengan menggunakan algoritma Naive Bayes adalah 90.71 %

accuracy: 90.71% +/- 0.74% (mikro: 90.71%)

	true Gagal	true Diterima	class precision
pred. Gagal	424	62	87.24%
pred. Diterima	1	191	99.48%
class recall	99.76%	75.49%	

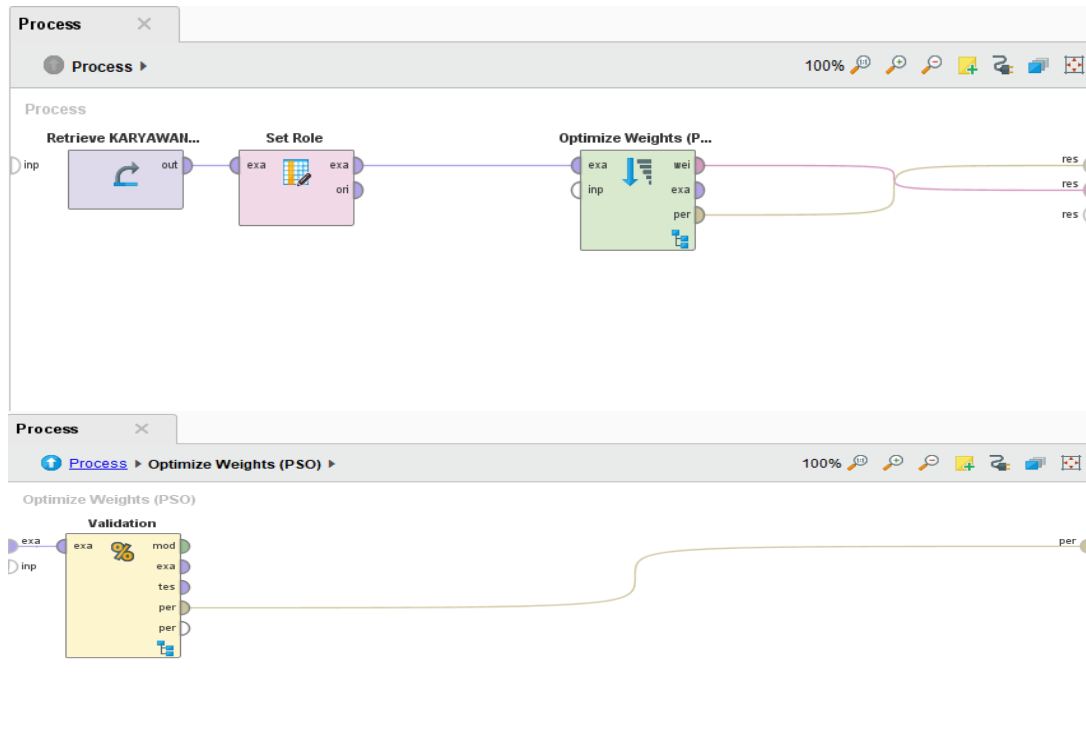
Gambar 3.2 Nilai Accuracy Naive Bayes(NB)

Pada hasil perhitungan di visualisasi dengan *kurva ROC*. Perbandingan dari dua class atau segmen untuk algoritma Naive Bayes. Kurva ROC merupakan plot yang mengekspresikan *confusion matrix* dari gambar 4.3 Line horisontal berupa *false* positif and Line vertikal true positif. Menghasilkan nilai AUC (*Area Under Curce*) sebesar 0.996 dengan nilai akurasi klasifikasi baik (*Excellent classification*).



Gambar 3.3 Kurva ROC Naive Bayes (NB)

Berikut adalah gambar penguraian *algoritma Naive Bayes* menggunakan *K-Fold Cross Validation* dengan *RapidMiner*. Proses pengujian pembobotan atribut dengan *PSO* menghasilkan beberapa langkah awal yaitu menentukan barometer fitur pembobotan pada aplikasi *rapidminer* pada *Population Size* dan *Maximum Number Off Generation*.



Gambar 3.4 Algoritma *Naive Bayes(NB)* pada *Particle Swarm Optimization (PSO)*

Gambar 3.5 merupakan jumlah *True Positif (TP)* adalah 425 record di *classification* sebagai segmen penerimaan karyawan (diterima/gagal) dengan prediksi yang dilakukan dengan metode *Naive Bayes* dan *PSO*, lalu *False Negatif (FN)* sebanyak 40 record yang diprediksi sebagai segmen penerimaan karyawan (Terima/gagal), kemudian *True Negatif (TN)* sebanyak 213 record sebagai *yes* sesuai dengan prediksi, dan *False Positif (FP)* sebanyak 0 record diprediksi *keterima* tetap gagal. Tingkat akurat yang dihasilkan dengan menggunakan *PSO* adalah 94.10%

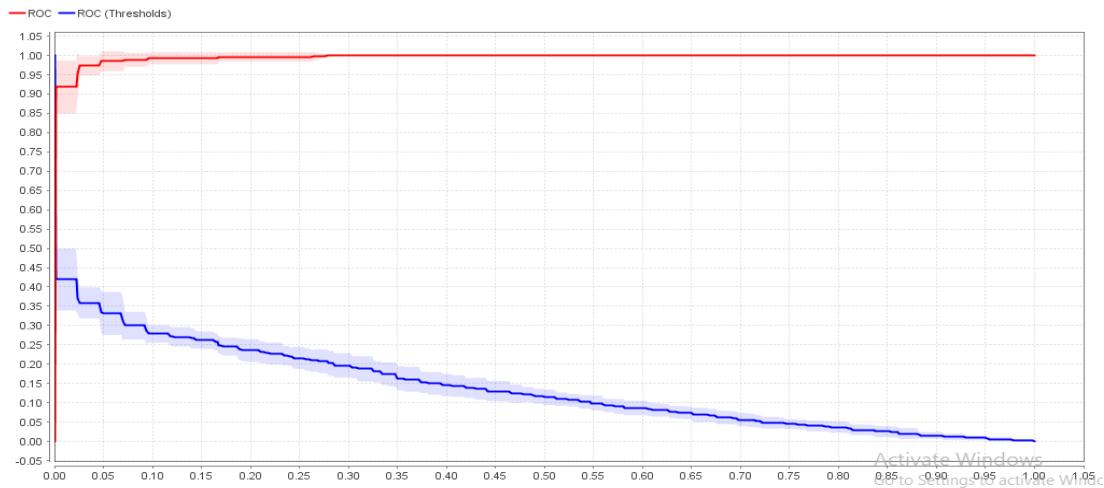
accuracy: 94.10% +/- 5.60% (mikro: 94.10%)

	true Gagal	true Diterima	class precision
pred. Gagal	425	40	91.40%
pred. Diterima	0	213	100.00%
class recall	100.00%	84.19%	

Gambar 3.5 *Accuracy Naive Bayes(NB)* dengan *PSO*

Gambar 3.6 Hasil perhitungan di metode *Naive Bayes(NB)* berbasis *Particle Swarm Optimization* divisualisasi dengan kurva *ROC*. Perbandingan dari dua class atau segmen untuk algoritma *Naive Bayes*. Kurva *ROC* merupakan plot yang menghasilkan *confusion matrix* dari tabel 3.5. Garis horisontal merupakan *false positif* dan garis vertikal *true positif*. Menghasilkan nilai *AUC (Area Under Curve)* sebesar 0.996 dengan nilai *accuracy* klasifikasi Terbaik (*Excellent classification*).

AUC: 0.996 +/- 0.003 (mikro: 0.996) (positive class: Diterima)



gambar 3.6 Kurva ROC *Naive Bayes*(NB) dengan PSO

Berikut hasil *confusion matrix* dengan uji coba NB berbasis PSO pada table 3.2 berikut ini :

Tabel 3.2 *confusion matrix* dengan pengujian *Naive Bayes*(NB) berbasis *particle swarm optimization* (PSO)

	True diterima	True gagal	<i>Class precision</i>
Pred. Gagal	425	40	91.40 %
Pred. Diterima	0	213	100.00 %
Class recall	100.00 %	84.19 %	

Keterangan tabel :

1. True positif (TP) = 425, yang berarti jumlah data sesuai dengan yang diprediksi dilakukan yaitu kelas Gagal
2. False negatif (FN) = 40, yang berarti jumlah data ini, yang diprediksi masuk dikelas layak “tidak”, ternyata berada diterima “Ya”.
3. True negatif (TN) = 213, Jumlah ini maksudnya sesuai dengan yang diprediksi yaitu dikelas diterima.
4. False positif (FP) = 0, yang berarti diprediksi masuk dikelas diterima ternyata berada dikelas Gagal.

Berikut nilai *Accurasi*, *Precision*, dan *recall*.

$$a. \text{ Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{425+213}{425+213+0+40} = \frac{638}{678} = 0.94$$

$$b. \text{ Precision (diterima)} = \frac{TP}{TP+FN} = \frac{425}{425+40} = 0.914$$

$$c. \text{ Precision (Gagal)} = \frac{TN}{TP+FP} = \frac{213}{425+0} = 0.501$$

$$d. \text{ Recall "diterima"} = \frac{TP}{TP+FP} = \frac{425}{425+0} = 1$$

$$e. \text{ Recall "gagal"} = \frac{TN}{TP+FN} = \frac{213}{425+40} = 0.458$$

Pada penelitian ini dapat menghasilkan analisis dari metode *algoritma Naive Bayes*(NB) berbasis *Particle Swarm Optimization* (PSO) dengan evaluasi nilai *accuracy* dan *AUC*.

Tabel 3.3 Hasil Perbandingan *Naive Bayes*(NB) dan NB + PSO

Hasil Pengujian	<i>Naive Bayes</i> (NB)	NB + PSO
Akurasi	91.45%	94.10%
AUC	0.997	0.996

Tingkat akurasi *Naive Bayes* 91.45% dibandingkan 94.10% jika optimasi dengan *particle swarm optimization*. Berdasarkan pengujian metode diperoleh tabel 3.3 yang dapat dideskripsikan sebagai berikut : Pada eksperimen yang pertama menggunakan *naive bayes (NB)* menghasilkan nilai akurasi 91.45% dengan nilai *area under curvenya* 0.997 yang termasuk dalam katagori “*Excellent classification*”. Sedangkan eksperimen kedua dengan menggunakan *naive bayes (NB)* berbasis PSO menghasilkan nilai *accuracy* yang tinggi daripada *naive bayes (NB)* dari 91.45% menjadi 94.10%. Terjadi Kenaikan dalam akurasi dan nilai *area under curvenya* mempunyai nilai yang sama yang termasuk dalam katagori “*Excellent classification*”. Sehingga Penelitian ini dapat peningkatan nilai akurasi.

4. KESIMPULAN

Berdasarkan peneltian diatas bisa di simpulkan sebagai berikut: (1) Metode yang diusulkan yaitu *Particle Swarm Optimization* telah berhasil diimplementasikan menggunakan metode *Naive Bayes (NB)* untuk mengatasi kasus *classification* penentu penerimaan karyawan. (2) Nilai penelitian menunjukkan bahwa model *Naive Bayes (NB)* berbasis *Particle Swarm Optimization (PSO)*, yang hasil dari *accuracy* lebih tinggi 94.10%, dibandingkan dengan model *Naive Bayes* 91.45%. Penelitian ini terbukti, bahwa saat menggunakan *Particle Swarm Optimization (PSO)* memiliki hasil yang *accuracy* lebih tinggi. (3) ini bisa dijadikan pemodelan untuk bisa diaplikasikan ke kasus yang lain, dengan dataset dan atribut yang lebih komplit. (4) Diharapkan untuk penelitian selanjutnya akan ada pengembangan dengan penggunaan metode algoritma penentu yang lain, seperti contoh *Super Vector Machine (SVM)*, *Algoritma Bipolar Slope Network (GRNN)*, dan yang lainnya, dengan penggunaan algoritma optimasi MSO (*Multy Swarm Optimization*) atau dengan *Gentic Algoritma (GA)*.

DAFTAR PUSTAKA

- [1] Ike Rachmawati Kusdyah. 2008. Manajemen Sumber Daya Manusia, Yogyakarta : ANDI
- [2] Ö. F. Arar and K. Ayan, “A Feature Dependent Naive Bayes Approach and Its Application to the Software Defect Prediction Problem,” *Appl. Soft Comput. J.*, 2017.
- [3] N. A. Zaidi, M. J. Carman, and G. I. Webb, “*Alleviating Naive Bayes Attribute Independence Assumption by Attribute weighting*,” vol. 14, pp. 1947-1988, 2006.
- [4] Kennedy, J., & Eberhart, R. (1995, November/December). “Particle Swarm Optimization”. *Proceedings of IEEE International Conference on Neural Networks* , 1942-1948
- [5] Zainudin.2022. Penerapan Algoritma C4.5 Berbasis Particle Swarm Optimization pada Fuzzy Mamdani untuk Prediksi Penerimaan Karyawan Baru. *Jurnal Sains, Teknologi dan Masyarakat* Vol. 2 No. 2, Juni 2022
- [6] T.Sousa, A. Silva, and A. Neves, “Particle Swarm based Data Mining Algorithms for classification tasks,”*Parallel Comput.*, vol. 30, no. 5–6, pp. 767–783, 2019.
- [7] Suamanda Ika Novichasari, “*Particle Swarm Optimization* Meningkatkan Akurasi *Naive Bayes Classifier* ,” Udinus ., 2018.
- [8] Dr. Alex A. Freitas, *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Leiden: Springer, 2002.
- [9] G. H. John, R. Kohavi, and K. Pfleger, “Irrelevant Features and the Subset Selection Problem,” *Mach. Learn. Proc. 1994*, pp. 121–129, 1994.
- [10] Pendahuluan and A. L. B. Masalah, “NEURAL NETWORK DENGAN ALGORITMA GENETIKA SEBAGAI PEMILIHAN,” pp. 78–91, 2018.
- [11] A. Mudry and A. Tjellström, “Historical background of bone conduction hearing devices and bone conduction hearing aids,” *Adv. Otorhinolaryngol.*, vol. 71, pp. 1–9, 2018.
- [12] U. Topics, *Principles of Data Mining*, Second edi. London: Springer London. pp.36-38,2018
- [13] Han, J., dan Kamber, M. (2007). *Data Mining Concepts and Technique*.Morgan Kaufmann publisher.
- [14] J. Lin and J. Yu, “Weighted Naive Bayes classification algorithm based on particle swarm optimization,” *2011 IEEE 3rd Int. Conf. Commun. Softw. Networks*, pp. 444–447, 2018.