JURNAL MULTI MEDIA DAN IT
VOL.09 NO.01 (2025), PP. 053-061
P-ISSN : 2548-9534 | E-ISSN : 2548-9550

Penerapan *Synthetic Minority Oversampling Technique* (SMOTE) pada Klasifikasi Kinerja Karyawan Di PT. Indomarco Prismatama Menggunakan Metode C 4.5 dan Naive Bayes

Nur Hasanah [a] *Sriyadi*[b] *Wawan Kurniawan* [c]

[a,b,c] *Universitas Bina Sarana Informatika*

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima Redaksi: 11 Maret 2025
Revisi Akhir: 6 Juli 2025
Diterbitkan *Online*: 7 Juli 2025

KATA KUNCI

C4.5, Kinerja Karyawan, Key Perfomance Indicators, Naïve Bayes, SMOTE

KORESPONDENSI

Nur Hasanah
Program Studi Sistem Informasi
Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 98, Kwitang, Senen, Jakarta Pusat, DKI Jakarta 10450
Email : Nurwarboys@gmail.com

ABSTRACT

Sistem penentuan kinerja karyawan di PT. Indomarco Prismatama hanya menggunakan microsoft excel yang memiliki kelemahan untuk data yang kompleks seperti dataset pada penelitian ini, terdapat 17 atribut dan 696 record yang terdiri dari 520 kelas No Reward dan 179 Reward. Sistem penentuan kinerja karyawan dapat menggunakan pendekatan machine learning yaitu teknik klasifikasi. Klasifikasi pada data yang tidak seimbang menyebabkan akurasi prediksi yang bias. Penelitian ini mengklasifikasikan kinerja karyawan menggunakan algoritma C4.5 dan Naïve Bayes dengan pendekatan teknik SMOTE untuk mengatasi ketidakseimbangan data. Tujuan dari penelitian ini untuk mengetahui faktor yang memengaruhi kinerja karyawan dan mengetahui performa terbaik dari kedua algoritma tersebut. Algoritma C4.5 digunakan untuk menghasilkan pohon keputusan yang membantu mengidentifikasi atribut paling signifikan, sedang Naïve Bayes memanfaatkan pendekatan probabilistik untuk menganalisis pola kinerja berdasarkan indikator utama. Evaluasi menggunakan K-Fold Cross Validation berdasarkan metrik akurasi, presisi, recall dan AUC. Hasil penelitian menunjukan bahwa atribut yang paling berpengaruh terhadap kinerja karyawan adalah variance plus. Performa algoritma C4.5 unggul dibandingkan Naïve Bayes, dengan akurasi 99%, recall 99%, presisi 99%, dan AUC 1.00. Penelitian ini memberikan kontribusi dalam pengembangan sistem penentuan kinerja karyawan yang lebih objektif, terstruktur dan berbasis data yang dapat dijadikan alternatif pengambilan keputusan oleh perusahaan.
DOI: https://doi.org/10.46961/jommit.v9i1

1. PENDAHULUAN

Karyawan merupakan salah satu faktor yang berperan penting dalam memajukan sebuah perusahaan. Kemampuan perusahaan untuk menghasilkan uang juga dipengaruhi oleh kinerja karyawan. Sebuah perusahaan memilih pekerja terbaik setiap periode dan memberi mereka bonus atau kenaikan gaji untuk memotivasi mereka[1]

Dalam meningkatkan kualitas pelayanan prima terhadap karyawan, perusahaan perlu menerapkan strategi untuk beberapa proses penilaian kinerja untuk menentukan karyawan terbaik. Prosedur untuk memperoleh dan mengevaluasi pengembangan kinerja karyawan diperlukan untuk pelaksanaan penilaian kinerja ini. Penilaian kinerja yang dilaksanakan secara berkala ini bertujuan untuk mengidentifikasi kelebihan dan kekurangan karyawan, menentukan kebijakan promosi atau mutasi, serta mengidentifikasi program pelatihan dan pengembangan. Hasil pengembangan juga menjadi dasar dalam pemberian kompensasi dan penghargaan kepada karyawan. Dengan adanya sistem penilaian yang terstruktur, diharapkan dapat meningkatkan motivasi dan mendorong peningkatan kinerja perusahaan secara keseluruhan.

PT Indomarco Prismatama Cabang JKT1 menggunakan standarisasi indikator kinerja untuk proses penilaiannya. Rapat kerja bulanan diadakan di kantor cabang JKT1 untuk membahas isu-isu yang berkaitan dengan pencapaian kinerja, operasional toko, pemilihan karyawan untuk penghargaan sebagai karyawan terbaik, kenaikan gaji dan pemberian insentif. Perusahaan memiliki standar indikator penilaian kinerja untuk menilai performa toko. Toko yang memenuhi target perusahaan berdasarkan penilaian indikator operasional toko dengan nilai performa di atas 70 akan mendapatkan reward yang diberikan kepada karyawan toko tersebut.

Key perfomance indicator operasional toko berdasarkan standarisasi yang diberikan perusahaan yaitu; ceklis *store*

activity, ceklis toko, cek sewa *display*, cek perubahan harga, *quantity* ITT, *cancel sales*, jumlah transaksi pada; tarik tunai, *new member* isaku, *new member* poinku, *new member* klik indomaret, jumlah transaksi pos kasir, pelaksanaan PJR, *variance minus* dan *variance plus*, tertib setor uang sales dan toko prima.

Berdasarkan pengamatan yang penulis lakukan pada PT Indomarco Prismatama cabang JKT1 dalam menentukan hasil dari klasifikasi penilaian karyawan yakni masih menggunakan hitung manual pada *software* microsoft excel. Meskipun metode ini memiliki kelebihan dalam memberikan pemahaman mendalam terhadap data dan fleksibilitas dalam analisis dengan biaya rendah namun juga memiliki kelemahan seperti waktu pengerjaan yang lama dan hanya efektif untuk dataset berukuran kecil serta sulit diterapkan pada data yang kompleks seperti pada dataset untuk klasifikasi kinerja karyawan yang mempunyai 696 *record* dengan 17 indikator penilaian.

Model Algoritma C4.5 dan Naïve Bayes merupakan metode *machine learning* yang dapat membantu dalam pengambilan keputusan untuk menentukan penilaian karyawan dalam mengklasifikasikan indikator kinerja karyawan. Sistem ini dimaksudkan untuk menghasilkan pengetahuan yang akan membantu manajer SDM dalam mengevaluasi kinerja karyawan melalui penerapan algoritma Naïve Bayes dan C4.5. Diharapkan bahwa Algoritma C4.5 akan menghasilkan pohon keputusan dimulai dari yang paling atas yang berfungsi sebagai node akhir, atau node, yang akan mewakili atribut dan daun paling bawah, atau daun, yang berfungsi untuk mewakili kelas [2] . Algoritma Naïve Bayes adalah pengklasifikasi probabilistik langsung yang menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data untuk menentukan serangkaian probabilitas.

Klasifikasi data dengan distribusi kelas yang tidak seimbang menjadi permasalahan utama dalam bidang *machine learning*. Saat menangani data yang tidak seimbang, sebagian besar algoritma klasifikasi cenderung menghasilkan akurasi yang lebih tinggi pada kelas mayoritas dibandingkan kelas minoritas. Dataset yang digunakan dalam penelitian ini memiliki distribusi kelas yang tidak seimbang yaitu 696 record yang terdiri dari 520 kelas No Reward dan 179 kelas Reward. Penelitian ini menerapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) sebagai solusi untuk mengatasi ketidakseimbangan pada dataset ini. SMOTE adalah teknik pengambilan sampel berlebih yang dapat menduplikasi data secara artifisial untuk mengatasi masalah distribusi data yang berbeda [3].

Tujuan dari penelitian ini, yaitu untuk menganalisis faktor-faktor yang memengaruhi kinerja karyawan di PT. Indomarco Prismatama serta mengevaluasi keunggulan dari kedua algoritma yang digunakan dalam meningkatkan kualitas pelayanan kepada konsumen. Penelitian ini juga bertujuan untuk membandingkan performa Algoritma C4.5 dan metode Naïve Bayes dalam analisis KPI guna mengidentifikasi pola dari berbagai atribut yang memengaruhi prediksi kinerja karyawan. Selain itu studi ini berfokus pada pengujian efektivitas penerapan SMOTE yang dikombinasikan dengan kedua metode tersebut dalam mengatasi ketidakseimbangan dataset KPI.

Berdasarkan penelitian yang dilakukan oleh Ridwan, Eni Heni Hermaliani, dan Muji Ernawati bahwa pendekatan Borderline-SMOTE menunjukkan kinerja terbaik dalam menangani data tidak seimbang dibandingkan metode SMOTE lainnya, yang dibuktikan dengan nilai accuracy sebesar 84,09%, recall 85,25%, precision 84,55%, dan F1-Score 81,16%, serta di antara algoritma machine learning yang diuji, Random Forest memiliki performa lebih unggul dibandingkan Naive Bayes, Support Vector Machine, dan Logistic Regression [3]. Penelitian selanjutnya, oleh Galih dan Mindit Eriyadi hasil penelitian membuktikkan bahwasannya algoritma Naïve Bayes Classifier mengungguli model lain yang diuji menggunakan dua alat berbeda dalam hal akurasi dan kinerja. Hasilnya, manajemen PT. Ganda Mady Indotama akan mendapatkan keuntungan di masa mendatang ketika model algoritma NBC diimplementasikan ke dalam aplikasi untuk pengukuran kinerja karyawan. [4]. Penelitian yang dilakukan Tedi Wahyudi, Popon Handayani, Rusdianto temuan penelitian ini diperoleh dari uji perbandingan antara model algoritma klasifikasi C4.5 dengan Naïve Bayes yang selanjutnya akan digunakan untuk menilai kinerja personel PT Catur Sentosa Adiprana. Jika dibandingkan dengan Naïve Bayes, nilai akurasi model algoritma klasifikasi C4.5 (*decision tree*) lebih tinggi. Dengan nilai akurasi sebesar 98,18%, metode klasifikasi C4.5 mengungguli algoritma Naïve Bayes sebesar 3,03%, dengan nilai akurasi sebanyak 95,15%. Kedua algoritma, Naïve Bayes dan C4.5 sebesar 0,994, mencapai tingkat terbaik (*excellent classification*) pada nilai uji ROC. Nilai presisi C4.5 sebanyak 99,38%, > Naïve Bayes sebanyak 98,73%, dan nilai recall-nya sebesar 98,75% masih lebih tinggi dari Naïve Bayes sebanyak 96,25% [5].

Dalam penelitian ini penulis melakukan kombinasi algoritma C4.5 dan Naive Bayes dengan metode SMOTE yang penulis gunakan untuk menangani data *imbalance,* menganalisis dan mengklasifikasikan kinerja karyawan di PT. Indomarco Prismatama. Kombinasi ini dipilih untuk memanfaatkan kekuatan masing-masing algoritma interpretabilitas pohon keputusan dari C4.5 dan efisiensi probabilistik dari Naive Bayes. Beberapa penelitian sebelumnya telah menunjukkan bahwa menggunakan pendekatan ensemble atau hibrida dari beberapa algoritma dapat meningkatkan akurasi dan robustness model klasifikasi. Maka penelitian ini diharapkan dapat memberikan hasil yang lebih komprehensif dan akurat dalam penilaian kinerja karyawan.

2. TINJAUAN PUSTAKA

2.1. Synthetic Minority Oversampling Technique (SMOTE)

SMOTE adalah teknik yang membuat sampel sintetis baru untuk mengambil *oversampling* dari kelas minoritas. Untuk membuat sampel sintetis baru dalam ruang fitur berdasarkan persentase tertentu untuk kelas minoritas, SMOTE menggunakan k- *nearest neighbor*. Untuk mengatasi masalah *overfitting*, SMOTE dapat membuat data sintetis baru berdasarkan data kelas minoritas yang sudah ada tanpa menduplikasinya. Untuk membuat sampel baru, SMOTE pertama-tama secara berurutan memilih setiap sampel dari sampel minoritas untuk dijadikan sampel akar untuk sintesis sampel lebih lanjut. Proses ini diulang n kali. Terakhir, n sampel yang disintesis diproduksi dengan melakukan interpolasi linier antara sampel dan setiap sampel tambahan. Masalah *overfitting* yang disebabkan oleh replikasi buta pendekatan oversampling acak berhasil diselesaikan oleh algoritma SMOTE [3]. Pendekatan resampling yang disebut undersampling mengurangi jumlah data di kelas primer untuk menyeimbangkan kumpulan data. Sebaliknya, *oversampling* adalah pendekatan resampling yang meningkatkan ukuran kelas minor untuk menyeimbangkan kumpulan data. Pendekatan SMOTE merupakan salah satu metode *oversampling* [6].

Ketika pemrosesan algoritma mengabaikan ketidakseimbangan data, kelas utama cenderung mendominasi kelas minor. Sementara pendekatan SMOTE menciptakan data palsu untuk meningkatkan jumlah data kelas minor agar sama dengan kelas utama, metode Oversampling didasarkan pada gagasan peningkatan observasi secara acak.[7]

2.2. *Algoritma C4.5*

Algoritma C4.5 yaitu metode klasifikasi yang menggunakan pohon keputusan untuk memproses data dengan atribut kategori atau numerik. Algoritma ini memanfaatkan prinsip entropi dan gain untuk mengidentifikasi kualitas yang paling informatif sebagai dasar pemisahan data. Pohon keputusan atau yang biasa dikenal dengan sebutan *decision tree* yaitu penggambaran hasil yang diterima setelah melakukan perhitungan dengan metode klasifikasi. Pohon keputusan mempunyai label selain node internal dan node akar yang disebut atribut [8].

Langkah-langkah membangun pohon keputusan menggunakan algoritma C4.5 adalah sebagai berikut [9]:

1) Pilih atribut sebagai akar. Pemilihan atribut sebagai akar berdasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai gain tertinggi digunakan persamaan berikut:

$$Gain(S,A) = Entropy(S) - \sum i{=}1\frac{|Si|}{|S|} * Entropy(Si)$$

Keterangan:

S : himpunan kasus
A : atribut
N : jumlah partisi atribut A
|Si| : jumlah kasus pada partisi ke-i
|S| : jumlah kasus dalam S

Nilai entropi dapat dihitung dengan cara sebagai berikut:

$$Entropy(S) = \sum_{i=1}^{n} - pi\ x\ pi$$

Keterangan:

S : himpunan kasus

n : jumlah partisi S

pi : proporsi dari S_i terhadap S

2) Buat cabang untuk tiap-tiap nilai.

3) Bagi kasus dalam cabang.

4) Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

2.3. *Naive Bayes*

Naïve Bayes ialah pengklasifikasi probabilistik langsung yang menghasilkan serangkaian probabilitas dengan menambahkan frekuensi dan kombinasi nilai dari sekumpulan data. Berdasarkan nilai-nilai dalam kelas variabel, Pendekatan ini membuat asumsi bahwa semua atribut bersifat independen satu sama lain atau tidak terkait satu sama lain dengan menerapkan teorema Bayes. Salah satu algoritma yang digunakan dalam pendekatan klasifikasi yang cepat dan mudah digunakan adalah Naïve Bayes. Dengan bobot atribut yang sama yang signifikan dan setiap atribut bersifat independen dari yang lain, semua atribut akan berperan dalam pengambilan keputusan dengan metode ini [10].Dibawah ini yaitu rumus persamaan perhitungan Teorema Bayes:

$$P(P|H) = \frac{P(H)_x P(H)}{P(Y)}$$

Keterangan:

H : hipotesis Y dari kelas spesifik

Y : data dengan kelas yang belum diketahui

P(H|Y) : probabilitas hipotesis H berdasarkan kondisi y (probabilitas posterior)

P(H) : probabilitas hipotesis H (probabilitas prior)

P(Y|H) : probabilitas hipotesis Y berdasarkan kondisi pada hipotesis H

P(Y) : probabilitas Y

Keterangan:

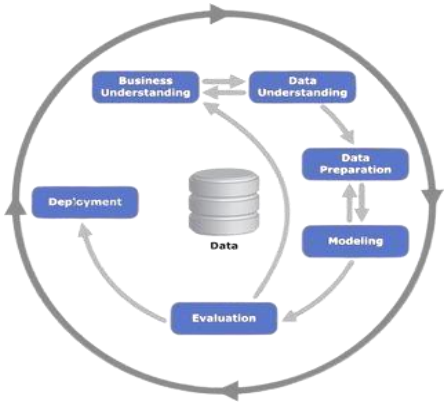
Probabilitasposterior : peluang munculnya kelas H

Probabilitasprior : peluang sampel awal kelas Y [11].

3. KERANGKA PENELITIAN

Data mining ialah sebuah proses, sehingga dalam menjalankan prosesnya harus sesuai dengan prosedur CRISP-DM (*Cross-Industry Standard Process for Data Mining*). CRISP-DM yaitu standarisasi data mining yang disusun oleh tiga pengagas data mining market meliputi: Daimler Chrysler, SPSS, NCR. CRISP-DM tidak menentukan standar atau karakteristik tertentu karena setiap data yang akan dianalisis akan diproses kembali pada fase-fase di dalamnya. Berikut gambar yang menjelaskan tentang siklus hidup CRISP-DM (*Cross Industry Standar Process for data mining*)[12]

Penelitian ini menggunakan analisis data standar yang terdapat pada data mining yaitu CRISP-DM (*Cross Standart Industries for Data Mining*) untuk klasifikasi penilaian kinerja karyawan. Dengan menggunakan dua model metode sebagai perbandingan yakni algoritma C4.5 dan Naïve Bayes untuk klasifikasikan kinerja karyawan dengan dataset KPI operasional toko.



Gambar 1. Diagram CRISP-DM (*Cross Standart Industries for Data Mining*) : [4]

https://doi.org/10.46961/jommit.v9i1

Gambar 1 diagram CRISP-DM (*Cross Standart Industries for Data Mining*) adalah metodologi dalam data mining yang memiliki enam tahap yakni, *business understanding, data understanding, data preparation, modeling, evaluation, deployment.* Tahapan standar CRISP-DM yang biasanya digunakan dalam data mining adalah sebagai berikut:

1) Business Understanding

Fokus utama dari fase pertama ini yaitu memahami tujuan dan persyaratan proyek dari sudut pandang bisnis. Informasi ini kemudian digunakan untuk mendefinisikan masalah penambangan data dan membuat rencana proyek awal yang bertujuan untuk mencapai tujuan tersebut[13].

Klasifikasi kinerja karyawan dapat membantu perusahaan untuk mengidentifikasi karyawan operasional mana yang memerlukan pelatihan dan karyawan yang mendapatkan penghargaan serta promosi. Dengan data yang terklasifikasi manajemen perusahaan dapat membuat keputusan yang lebih spesifik seperti kegiatan operasional yang berjalan sesuai ketentuan yang ada pada klasifikasi KPI Proses operasional toko. Dalam indikator penilaian kinerja (KPI Proses) ditentukan nilai poin jika diatas 6 akan mendapatkan insentif berupa tunjangan hal ini bertujuan untuk meningkatkan motivasi kinerja karyawan guna mencapai tujuan perusahaan yakni pelayanan yang optimal kepada konsumen. Transparansi data ini membuat sistem klasifikasi akan lebih objektif karena dapat memastikan bahwa penilaian dilakukan secara adil. Penunjang keputusan algoritma C4.5 dan Naïve Bayes akan membantu perusahaan dalam perencanaan jangka panjang untuk pengelolaan sumber daya manusia. Dengan mengetahui kinerja karyawan saat ini melalui klasifikasi perusahaan dapat merencanakan kebutuhan rekrutmen dan pengembangan beberapa klasifikasi guna mencapai visi dan misi perusahaan.

2) Data understanding

Sumber data yang digunakan pada penelitian yaitu data primer, penulis menggunakan dataset *Key Perfomance Indicator* tipe proses operasional toko dalam bentuk excel tahun 2024 periode bulan september dari total 696 data operasional toko cabang JKT1. Data KPI operasional toko ini diperoleh dari *area supervisor* setiap bulan dibagikan kepada setiap toko untuk membahas evaluasi kinerja karyawan. Dalam tahapan ini penulis memperoleh data KPI operasional toko yang telah berisikan nilai pada atribut atau variabel yang ada sebanyak 19 atribut dari total 25 atribut. Pada tahapan ini, data yang telah diterima dari subjek penelitian berbentuk *softcopy* berekstensi xls dan sudah memiliki hasil penilaian dari proses yang sudah dilaksanakan dari seluruh toko. Hasil penilaian dan atribut ini meliputi ceklis toko, jumlah transaksi tarik tunai, perubahan harga, jumlah *new member* I-saku, jumlah *new member* poinku, jumlah *new member* klik, *quantity* item tidak terdisplay, jumlah transaksi pos kasir diluar CPU utama, pelaksanaan penanggung jawab rak, ceklis *store activity,* cek sewa *display, variance minus, variance plus*, tertib setor uang *sales*, toko prima dan *cancel sales*.

3) Data Preparation

Pada tahapan ini, penulis melakukan persiapan data dengan mengolah data yang ada dengan memeriksa kolerasi setiap atribut untuk mendapatkan hasil analisis yang akurat dan memastikan bahwa model yang dibangun memiliki kemampuan generalisasi yang baik. Dari dataset yang diperoleh ada 17 atribut yang akan digunakan. Berikut di bawah ini yang menjadi tahapan-tahapan untuk persiapan data, antara lain[12]:

a. *Data Cleaning* yang bertujuan untuk menemukan, memeriksa dan membersihkan nilai-nilai atau data yang tidak konsisten.
b. *Data integration* adalah menggabungkan dataset excel KPI Proses ke dalam *software* RapidMiner untuk meningkatkan kualitas akurasi dan analisis data yang lebih komprehensif.
c. Data Reduction dikarenakan tidak semua atribut digunakan untuk atribut penentu sehingga penulis melakukan reduksi dengan mengurangi jumlah atribut agar lebih sedikit tetapi masih memiliki data yang informatif. Setelah direduksi jumlah atribut yang digunakan penulis hanya menggunakan 17 atribut dari total 25 atribut.

Tabel 1. Proses *Data Cleaning*

No.	Atribut	Tipe Data	Parameter
1	*Variance Minus*	*Integer*	(0-100)
2	*Variance Plus*	*Integer*	(0-100)
3	*Cancel sales*	*Integer*	(0-100)
4	Tertib setor uang *sales*	*Integer*	(0-100)
5	Jumlah transaksi diluar cpu utama	*Integer*	(0-100)
6	Transaksi tarik tunai	*Integer*	(0-100)
7	AVG *New Member* I-Saku	*Integer*	(0-100)
8	AVG *New Member* Poinku	*Integer*	(0-100)
9	AVG *New Member* Klik	*Integer*	(0-100)
10	Ceklis Toko	*Integer*	(0-100)
11	Cek Lis *Store Activity*	*Integer*	(0-100)
12	Cek sewa *display*	*Integer*	(0-100)
13	Toko Prima	*Integer*	(0-100)
14	Cek Perubahan Harga	*Integer*	(0-100)
15	Pelaksanaan PJR	*Integer*	(0-100)
16	QTY ITT	*Integer*	(0-100)
17	*REMARK*	Binominal	*Reward, No Reward*

Sumber: Hasil penelitian, 2024

Dataset direduksi dengan mengurangi jumlah atribut yang tidak relevan, atribut tersebut meliputi kode toko, nama toko, AS dan AM.

4) *Modeling*

Pada tahap ini, berbagai teknik pemodelan dipilih dan diterapkan, dan parameternya dikalibrasi ke nilai yang optimal [14]. Pada tahap *data preparation* sebelumnya menghasilkan 17 atribut yang akan digunakan untuk diuji dengan dua algoritma sebagai komparasi yaitu algoritma C4.5 dan Naïve Bayes. Dalam menangani *Imbalanced Data* penulis menggunakan Teknik SMOTE yang dimasukan saat uji data menggunakan *software* RapidMiner.

5) Evaluation

Evaluasi adalah proses menilai keakuratan prediktif suatu model; validasi adalah proses membuat prediksi menggunakan model yang sudah ada dan kemudian membandingkan hasilnya dengan hasil yang diketahui [9]

Setelah pemrosesan data, tingkat akurasi dapat diperiksa pada langkah evaluasi untuk menilai seberapa baik pendekatan algoritma C4.5 dan teorema Naive Bayes bekerja. Tujuan dari studi ini adalah untuk mengevaluasi seberapa akurat analis data penyimpanan menentukan evaluasi kinerja staf. *Confussion matrix* dan kurva ROC/AUC (*Area Under Cover*) digunakan untuk mengukur tingkat akurasi. Kurva ROC mempunyai tingkat diagnosa yakni akurasi bernilai 0.90-1.00 = *excellent classification*, 0.80-0.90 = *good classification*, 0.70-0.80 = *fair classification*, 0.60-0.70 = *poor classification*, dan 0.50-0.60 = *failure classification*[15]

6) Deployment

Penulis melakukan penerapan model algoritma klasifikasi terhadap sistem operasional PT Indomarco Prismatama berdasarkan hasil data yang diperoleh. Penulis mengikuti perubahan atribut klasifikasi berdasarkan data yang diperoleh menyebabkan berubahnya karakteristik dari hasil uji data hal ini mempengaruhi dari hasil uji data, model dibangun akan mengikuti perkembangan dari data pada waktu tertentu. Maka PT Indomarco Prismatama perlu pemahaman yang lebih dalam apakah penerapan model perlu diganti atau diperbaharui.

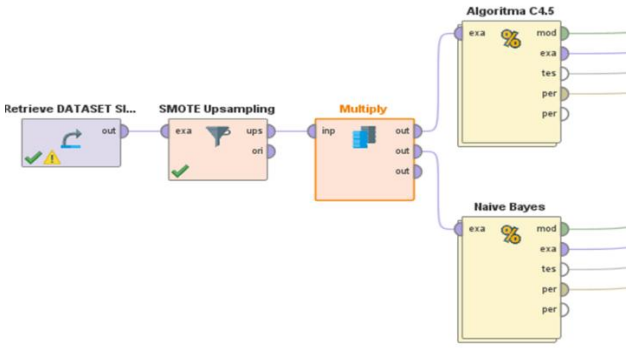
4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data sample dari data KPI Proses dan diperoleh 696 *record* yang terdiri 520 *reward* dan 179 *no reward* serta 17 atribut yang termasuk label yang diuji tingkat akurasinya untuk melihat kinerja metode algoritma C4.5 dan Naïve Bayes. Sebelum menguji dataset pada model penulis menangani ketidakseimbangan kelas yang mengalami *over-fitting* pada proses teknik *over-sampling* untuk kelas minoritas menggunakan teknik SMOTE selanjutnya masuk ke tahap data training yang dijadikan sebagai alat yang digunakan untuk menguji kebenaran dari hasil pengolahan menggunakan *software rapidminer studio* versi 10.1 dengan menggunakan dua *subdata* yaitu *data training* dan *data testing* untuk menguji kebenaran dari hasil pengolahan data training.

Pada tahapan *data selection*, perubahan *set role* pada atribut *remark* diubah menjadi label agar mempercepat proses seleksi data. Selanjutnya mebambahkan SMOTE Upsampling dari operator untuk menangani kelas data minoritas menjadi seimbang, menambahkan *multiply* dari operator yang digunakan sebagai pembuat salinan objek.

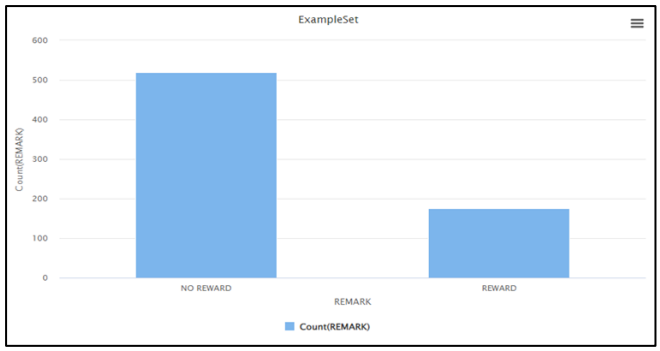
Gambar 2. Penambahan Operator *Set Role*, SMOTE dan *Multiply* (Sumber : Hasil penelitian, 2024)

Tahapan selanjutnya adalah menambahkan cross validation untuk dua model yang dibangun berdasarkan dataset. Penulis menggunakan *cross-validation* K-Fold Cross Validation.



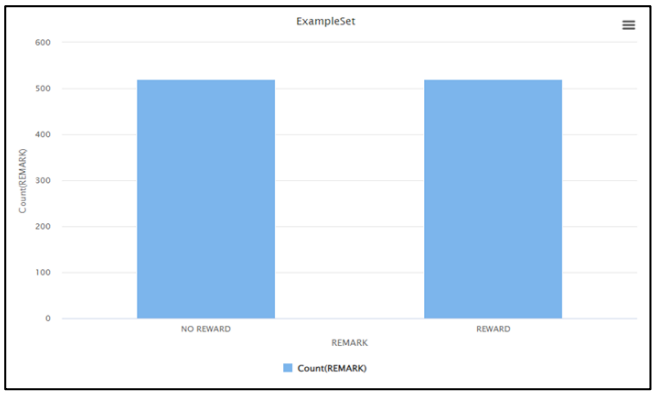
Gambar 3. Penambahan *cross validation* kedua model (Sumber : Hasil penelitian, 2024)

Setelah memasukan semua langkah Setelah memasukan semua langkah yang telah ditambahkan maka selanjutnya penulis memastikan semua garis terhubung dan tidak ada yang error, dimulai dari garis retrive dataset uji keluaran (*out*), ke masukan (*inp*) *multiply*, keluaran (*out*) *multiply* ke *training* (*tra*) cross validation Naïve Bayes dan Decision Tree sampai keluaran model dan *perfomance* (per) ke *result* (*res*). Setelah itu melakukan eksekusi dengan menekan tombol F11 atau dapat klik *Process*. Akan muncul hasil *result* berupa pohon keputusan dan nilai perolehan *simple distribution* dari model klasifikasi Naïve Bayes. Hasil distribusi kelas pada atribut label REMARK memiliki jumlah data yang tidak seimbang yakni 520 *no reward* dan 176 *reward* sehingga perlu dilakukan proses untuk mengatasi *imbalanced data*.



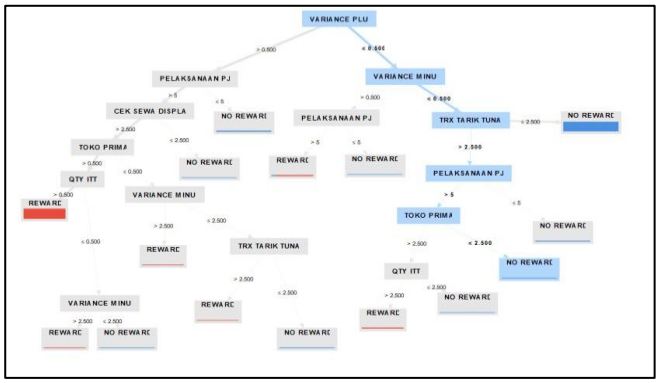
Gambar 4. Distribusi kelas sebelum menggunakan SMOTE (Sumber : Hasil penelitian, 2024)

https://doi.org/10.46961/jommit.v9i1

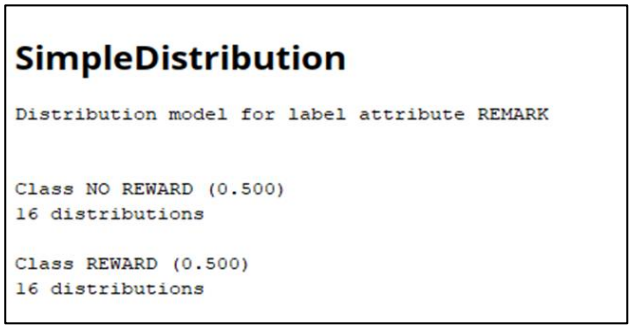


Gambar 5. Distribusi kelas sesudah menggunakan SMOTE (Sumber : Hasil penelitian, 2024)

penerapan teknik SMOTE menghasilkan distribusi kelas yang seimbang yakni 520 *Reward* dan 520 *No Reward* dari sebelum menggunakan SMOTE yakni 520 *no reward* dan 176 *reward*.



Gambar 6. Hasil *tree decision* C4.5 (Sumber : Hasil penelitian, 2024)



Gambar 7. Hasil proses SimpleDistribution Naïve Bayes (Sumber : Hasil penelitian, 2024)

Hasil proses yang menjadi keluaran dari data training dengan pemilihan model menggunakan cross validation 10 K-Fold. Bentuk yang ditampilkan C4.5 (*decision tree*) merupakan pohon keputusan yang terbentuk dari akar (*node*) yang mempunyai nilai *gain* tertinggi, sedangkan algoritma Naïve Bayes menampilkan hasil perhitungan peluang dari kelas atau label yang maksimal. Analisis yang ditampilkan oleh aplikasi RapidMiner dengan model decision tree didapatkan hasil dengan tingkat akurasi 99.42%.

accuracy: 99.42% +/- 0.67% (micro average: 99.42%)

	true NO REWARD	true REWARD	class precision
pred. NO REWARD	517	3	99.42%
pred. REWARD	3	517	99.42%
class recall	99.42%	99.42%	

Gambar 8. *Confusion Matrix* algoritma C4.5 (Sumber : Hasil penelitian, 2024)

Dapat dilihat dari 696 data training terdapat 517 data diprediksikan *NO REWARD* sesuai yaitu 517 data *NO REWARD* dan 3 data yang diprediksikan *NO REWARD* tapi ternyata *REWARD*. Sebanyak 3 data diprediksikan *REWARD* ternyata termasuk klasifikasi *NO REWARD* dan sebanyak 517 data diprediksi sesuai yaitu *REWARD*.

Pada model Naïve Bayes menghasilkan akurasi 73,.37%.

accuracy: 73.37% +/- 3.17% (micro average: 73.37%)

	true NO REWARD	true REWARD	class precision
pred. NO REWARD	245	2	99.19%
pred. REWARD	275	518	65.32%
class recall	47.12%	99.62%	

Gambar 9. *Confusion Matrix* Naive Bayes 5 (Sumber : Hasil penelitian, 2024)

diprediksikan *NO REWARD* ternyata *REWARD*. Sebanyak prediksi 275 data *REWARD t*ernyata masuk klasifikasi *NO REWARD* dan sebanyak 618 data diprediksi sesuai yaitu *REWARD*.

Hasil presisi telah menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model C4.5 (*decision tree*) adalah sebagai berikut:

precision: 99.43% +/- 0.91% (micro average: 99.42%) (positive class: REWARD)

	true NO REWARD	true REWARD	class precision
pred. NO REWARD	517	3	99.42%
pred. REWARD	3	517	99.42%
class recall	99.42%	99.42%	

Gambar 10. Hasil *Precision* Algoritma C4.5 (Sumber : Hasil penelitian, 2024)

Hasil presisi algoritma C4.5 menghasilkan 99.43% per 0.91% dengan *micro average*.

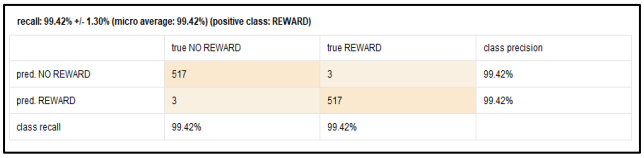
Adapun nilai presisi untuk klasifikasi Naïve Bayes, tetap memiliki nilai yang lebih rendah dibandingkan klasifikasi C4.5 (*decision tree*)dengan nilai presisi 65.40% per 2.65% dengan *micro average* 65.32%.

precision: 65.40% +/- 2.65% (micro average: 65.32%) (positive class: REWARD)

	true NO REWARD	true REWARD	class precision
pred. NO REWARD	245	2	99.19%
pred. REWARD	275	518	65.32%
class recall	47.12%	99.62%	

Gambar 11. Hasil *precision* Naive Bayes (Sumber : Hasil penelitian, 2024)

Pada *recall* algoritma C4.5 menghasilkan 99.42% per 1.30% dengan *micro average* 99.42% namun untuk prediksi *reward* dan *no reward* menghasilkan nilai yang sama dengan akurasi *confusion matrix* dan *precisison* C4.5.

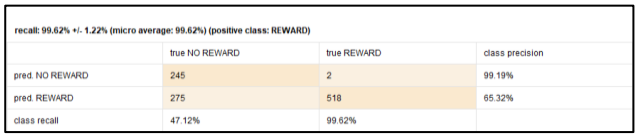


recall: 99.42% +/- 1.30% (micro average: 99.42%) (positive class: REWARD)

	true NO REWARD	true REWARD	class precision
pred. NO REWARD	517	3	99.42%
pred. REWARD	3	517	99.42%
class recall	99.42%	99.42%	

Gambar 12. Hasil recall Algoritma C4.5 (Sumber : Hasil penelitian, 2024)

Pada recall algoritma Naive Bayes menghasilkan 99.62% per 1.22% dengan micro average 99.62% hasil ini lebih besar dibandingkan dengan akurasi confusion matrix dan precision, namun unruk prediksi reward dan no reward menghasilkan nilai yang sama dengan akurasi confusion matrix dan precision Naive Bayes.

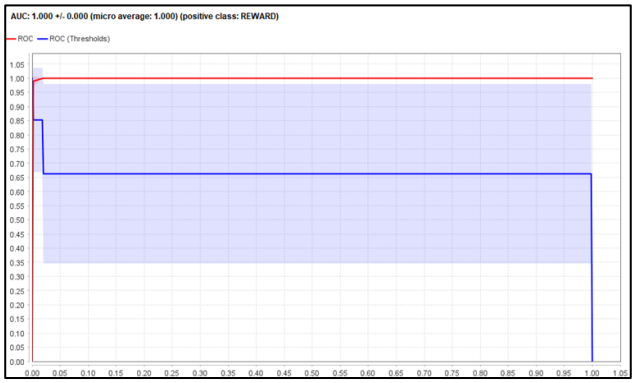


recall: 99.62% +/- 1.22% (micro average: 99.62%) (positive class: REWARD)

	true NO REWARD	true REWARD	class precision
pred. NO REWARD	245	2	99.19%
pred. REWARD	275	518	65.32%
class recall	47.12%	99.62%	

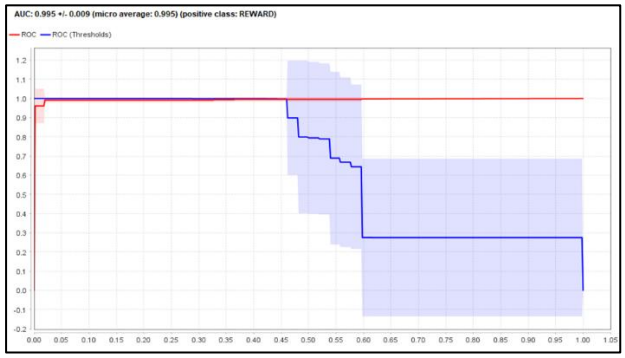
Gambar 13. Hasil *recall* Naive Bayes (Sumber : Hasil penelitian, 2024)

Pada Algoritma C4.5 menghasilkan AUC dengan nilai 1.000 atau 1,00 dengan *micro avarage* 1.000 dengan kelas positif *REWARD*. Terbukti bahwa nilai sensitivitas (*true positive rate*) dan spesifisitas (*false positive rate*) tidak melebihi 1. Nilai AUC 1,00 menunjukkan tingkat pengklasifikasikan berkualitas sangat baik atau *Excellent Classification*.



Gambar 14. Grafik ROC klasifikasi C4.5 (Sumber : Hasil penelitian, 2024)

Pada grafik klasifikasi AUC Naïve Bayes menghasilkan nilai AUC lebih kecil yakni 0.995 dengan mincro average 0.995 dan kelas positif *REWARD*. Terbukti bahwa nilai sensitivitas (*true positive rate*) dan spesifisitas (*false positive rate*) tidak melebihi 1. Nilai AUC 0,99 menunjukkan tingkat pengklasifikasikan berkualitas sangat baik atau *Excellent Classification.*



Gambar 15. Grafik ROC klasifikasi Naive Bayes (Sumber : Hasil penelitian, 2024)

Perbandingan perfomance kedua model usulan sebelum menggunakan teknik SMOTE berikut:

Tabel 2. Hasil kinerja sebelum menggunakan SMOTE

Model	**C4.5**	**Naive Bayes**
Accuracy	99.43%	59.75%
Recall	99.44%	99.44%
Precision	98.36%	38.80%
AUC	0,95	0,92

(Sumber : Hasil penelitian, 2024)

Perbandingan kinerja C4.5 dan Naïve Bayes yang telah menerapkan SMOTE dapat dilihat pada Tabel III.5 menggambarkan bahwa algoritma C4.5 menunjukkan performa yang lebih tinggi dibandingkan Naïve Bayes dalam memprediksi dataset KPI dengan nilai *Accuracy* 99,42%, *Recall* 99.42%, *Precision* 99.43% dan pada AUC 1,00.

Tabel 3. Hasil kinerja setelah menggunakan SMOTE

Model	**C4.5**	**Naive Bayes**
Accuracy	99.42%	73.37%
Recall	99.42%	99.62%
Precision	99.43%	65.40%
AUC	1,00	0,99

(Sumber : Hasil penelitian, 2024)

SMOTE terbukti dapat menangani ketidakseimbangan kelas dengan menyeimbangkan distribusi data, model dapat belajar secara lebih baik, dan mengurangi bias terhadap kelas mayoritas dan meningkatkan kinerja model. Implementasi SMOTE pada metode C4.5 dan Naïve Bayes dapat mengoptimalkan akurasi, recall, presisi dan AUC (Area Under Cover) yang menunjukkan performa model yang lebih baik. Penelitian ini menunjukkan bahwa metode yang diusulkan tersebut sangat efektif dalam meningkatkan kemampuan klasifikasi, terutama dalam kasus data yang kompleks dan tidak seimbang.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat diketahui bahwa algoritma C4.5 dan Naïve Bayes mampu menghasilkan performa yang baik dalam memprediksi kinerja karyawan dengan menganalisis pola dari atribut pada dataset KPI

https://doi.org/10.46961/jommit.v9i1

di PT. Indomarco Prismatama Cabang JKT 1. Hasil decision tree menunjukkan bahwa atribut yang menjadi thresold paling berpengaruh terhadap kinerja karyawan adalah variance plus. Penerapan teknik SMOTE merupakan pendekatan yang efektif untuk permasalahan imbalanced data terbukti dapat meningkatkan akurasi kedua algoritma tersebut. Hasil C4.5 dengan akurasi, recall, presisi 99% dan AUC 1,00, sedangkan hasil Naïve Bayes akurasi 73% recall 99%, presisi 65% dan AUC 0,99. Performa C4.5 dan Naïve Bayes dengan mengintegrasikan SMOTE lebih unggul dibandingkan tanpa penerapan SMOTE. Klasifikasi penilaian kinerja karyawan menggunakan machine learning, khususnya dengan algoritma C4.5 menjadi referensi bagi perusahaan untuk mengembangkan platform evaluasi penilaian kinerja karyawan yang lebih efektif dan objektif.

Untuk meningkatkan akurasi prediksi dan pemahaman terhadap faktor-faktor yang memengaruhi klasifikasi penilaian kinerja karyawan penggunaan dataset yang lebih luas dan beragam, baik dari sisi jumlah data maupun indikator kinerja untuk meningkatkan akurasi model klasifikasi sangat disarankan utuk penelitian selanjutnya. Selain itu, uji coba permodelan menggunakan dataset cabang lain perlu dilakukan untuk memastikan geralisasi model sekaligus menilai dampak penerapan algoritma terhadap efesiensi operasional dan motivasi karyawan. Kombinasi teknik seperti ensembel algoritma dan penguatan metode SMOTE juga dapat diterapkan untuk meningkatkan performa klasifikasi dan hasil evaluasi kinerja secara kesluruhan. Menguji algoritma lain seperti Random Forest atau Support Vector Machine (SVM) sebagai diintegrasikan ke dalam sistem manajemen kinerja berbasis digital agar hasil dapat langsung digunakan dalam pengambilan PT. Indomarco Prismatama

DAFTAR PUSTAKA

[1] I. Ardhiyanto, V. Lusiana, and N. Mariana, *Implementasi Metode (WP) Weighted Product Pada Sistem Pendukung Keputusan Penilaian Karyawan Terbaik di Pandanaran Hotel Semarang*. unisbank.ac.id, 2019. [Online]. Available: https://www.unisbank.ac.id/ojs/index.php/sintak/article/view/7573

[2] A. A. Aldino and H. Sulistiani, "DECISION TREE C4.5 ALGORITHM FOR TUITION AID GRANT PROGRAM CLASSIFICATION (CASE STUDY: DEPARTMENT OF INFORMATION SYSTEM, UNIVERSITAS TEKNOKRAT INDONESIA)," 2020.

[3] K. Ujaran, K. Ridwan, E. Heni Hermaliani, M. Ernawati, and C. Author, "Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada," 2024. [Online]. Available: http://jurnal.bsi.ac.id/index.php/co-science

[4] M. Eriyadi, "Perbandingan Model NBC, SVM, dan C4.5 dalam Mengukur Kinerja Karyawan Berprestasi Pasca Pandemi Covid-19," *JURNAL INFORMATIKA*, vol. 9, no. 2, 2022, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ji

[5] T. Wahyudi and P. Handayani, "PERBANDINGAN AKURASI C4.5 DAN NAIVE BAYES UNTUK EVALUASI KINERJA KARYAWAN PT CATUR SENTOSA ADIPRANA," *Nusa Mandiri Tower Jl. Jatiwaringin Rya*, vol. 5, no. 2, 2022.

[6] A. Syukron, E. Saputro, and P. Widodo, "Penerapan Metode Smote Untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung," 2023. [Online]. Available: https://doi.org/10/25047/jtit.v10i1.312

[7] K. Alamat, W. Nugraha, D. Risdiansyah, D. Purwaningtias, T. Hidayatulloh, and S. Suhada, "KOMBINASI TOMEK-LINK DAN SMOTE UNTUK MENGATASI KETIDAKSEIMBANGAN KELAS PADA CREDIT CARD FRAUD," 2022. [Online]. Available: http://jurnal.bsi.ac.id/index.php/larik

[8] I. Puspitorini, D. Novita, and W. D. Septiani, "Penentuan Jenis Benang Untuk Pembuatan Produk Pada Konveksi Menggunakan Metode Algoritma C4. 5 Dengan Rapid Miner," *Jurnal Minfo Polgan*, 2023, [Online]. Available: https://jurnal.polgan.ac.id/index.php/jmp/article/view/12555

[9] F. Fatmawati and N. Narti, "Perbandingan Algoritma C4.5 dan Naive Bayes Dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 1, pp. 1–12, May 2022, doi: 10.35746/jtim.v4i1.196.

[10] V. Vibrianti, E. Wahyudin, K. Kaslani, and ..., "KLASIFIKASI BARANG PRODUKSI PADA TNT. GUITAR WORKSHOP DENGAN METODE NAIVE BAYES MENGGUNAKAN RAPID MINER," *JATI (Jurnal ...*, 2024, [Online]. Available: https://www.ejournal.itn.ac.id/index.php/jati/article/view/8966

[11] L. Elvitaria and E. Malau, "Penerapan Algoritma C4. 5 Dalam Mengukur Tingkat Kepuasan Pengguna E-Learning Pada Mahasiswa Universitas Abdurrab," *JEKIN-Jurnal Teknik Informatika*, 2024, [Online]. Available: https://rumahjurnal.or.id/index.php/JEKIN/article/view/898

[12] A. Pambudi, "Penerapan Crisp-Dm Menggunakan Mlr K-Fold Pada Data Saham Pt. Telkom Indonesia (Persero) Tbk (Tlkm)(Studi Kasus: Bursa Efek Indonesia Tahun 2015-2022)," *Jurnal Data Mining dan Sistem Informasi*, 2023, [Online]. Available: https://ejurnal.teknokrat.ac.id/index.php/JDMSI/article/view/2462

[13] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining."

[14] S. Navisa, L. Hakim, and A. Nabilah, "Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM," *Jurnal Sistem Cerdas*, 2021, [Online]. Available: https://apic.id/jurnal/index.php/jsc/article/view/162

[15] S. Wahyudi, *Data mining dengan decision tree C4. 5 dan apriori (Konsep dan implementasi menggunakan rapid miner dan WEKA)*. digilib.itbwigalumajang.ac.id, 2022. [Online]. Available: https://digilib.itbwigalumajang.ac.id/index.php?p=show_detail&id=22761

https://doi.org/10.46961/jommit.v9i1