

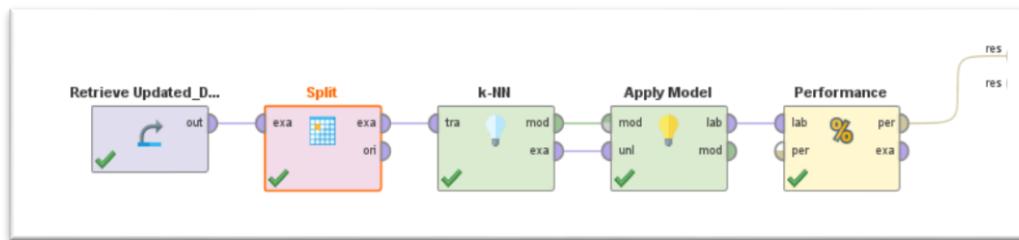
BAB IV

HASIL DAN PEMBAHASAN

Bab ini akan membahas tentang proses, hasil, dan pembahasan penelitian. Software RapidMiner digunakan sebagai alat bantu untuk memproses data. Proses, hasil, dan pembahasan terbagi ke dalam beberapa subbab, termasuk Pengolahan dan distribusi Data serta Implementasi algoritma pada RapidMiner.

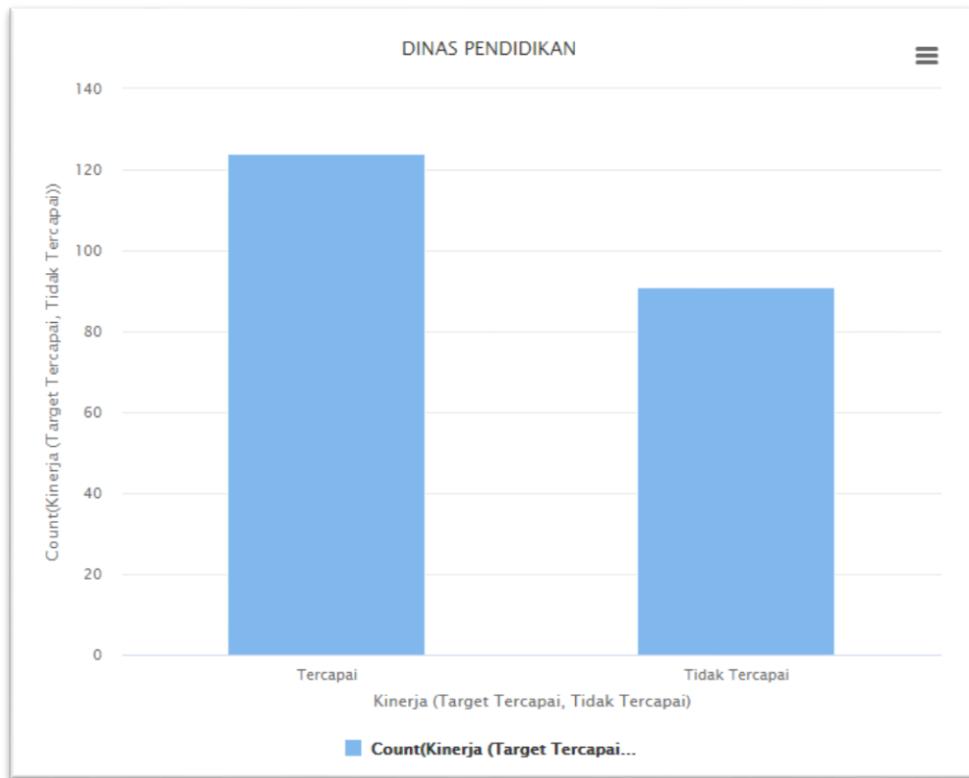
4.1 Hasil Penelitian

Berikut adalah penerapan algoritma KNN yang telah dioptimalkan dengan menggunakan perangkat RapidMiner



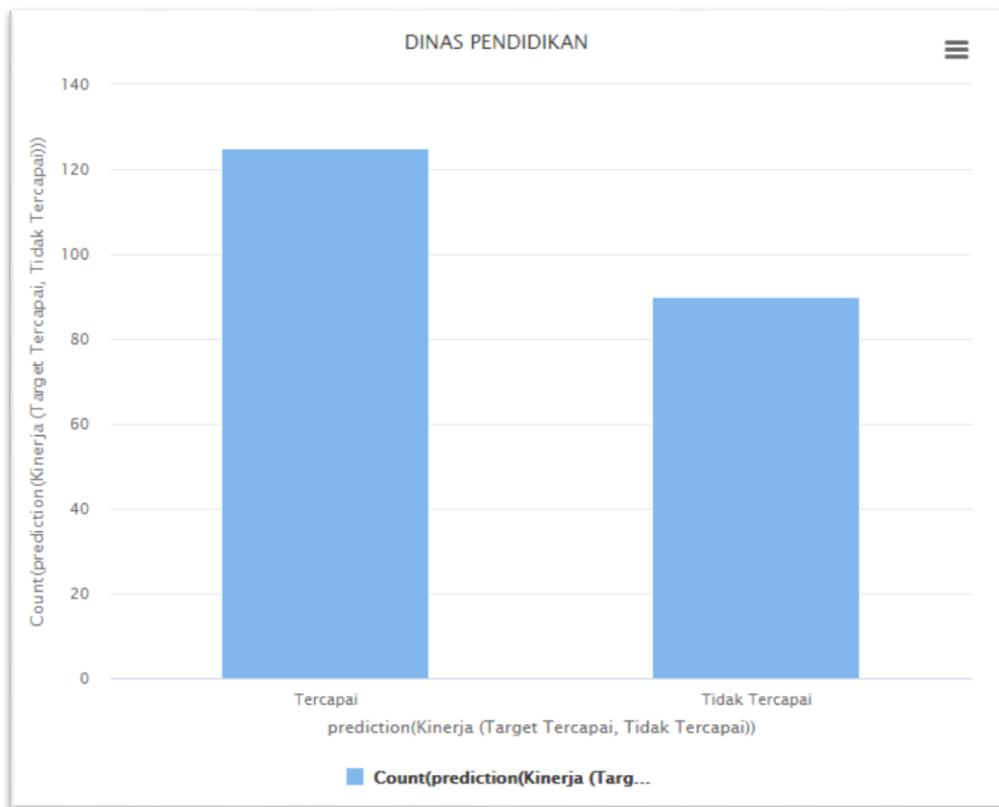
Gambar 4.1 Optimasi Algoritma KNN

Gambar berikut menunjukkan distribusi pencapaian kinerja pada instansi Dinas Pendidikan berdasarkan dua kategori utama, yaitu "Tercapai" dan "Tidak Tercapai. Grafik batang ini menggambarkan jumlah individu atau unit kerja yang berhasil memenuhi target kinerja dibandingkan dengan yang tidak berhasil mencapainya. Dari visualisasi tersebut, terlihat bahwa sebagian besar entitas berada pada kategori "Tercapai", yang menunjukkan bahwa secara umum target kinerja telah berhasil dicapai oleh mayoritas pihak yang terlibat.



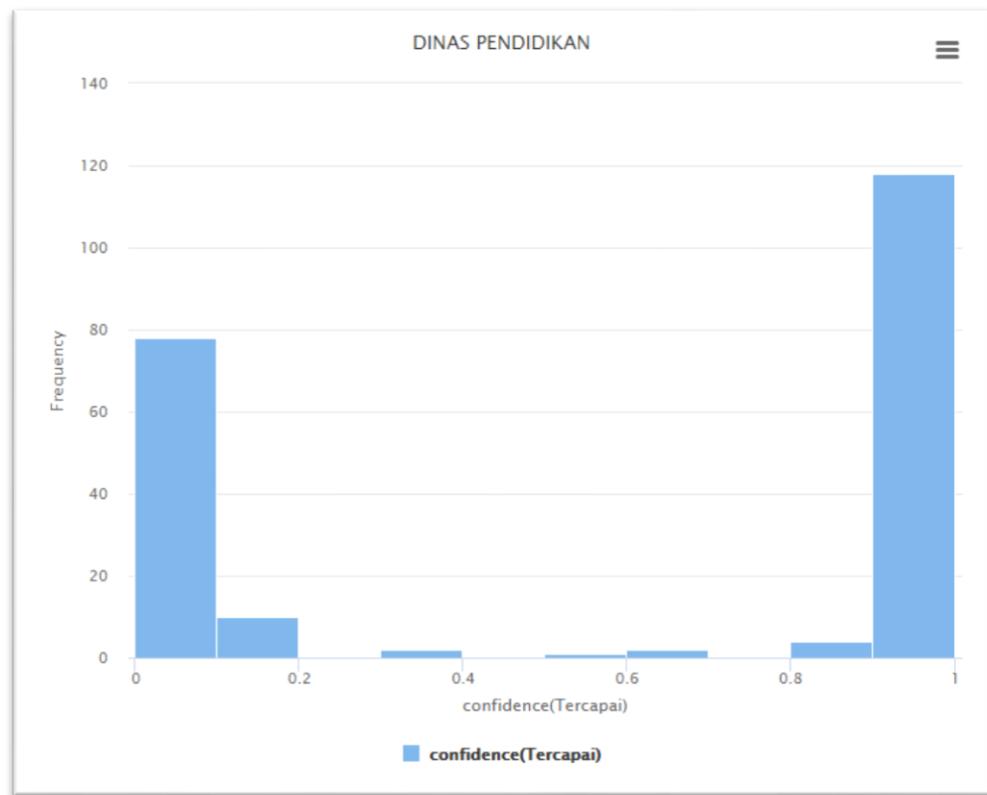
Gambar Grafik 4.2 Kinerja Tercapai dan tidak tercapai

grafik pada gambar ini memperlihatkan hasil prediksi kinerja berdasarkan dua kategori utama, yaitu "Tercapai" dan "Tidak Tercapai", pada instansi Dinas Pendidikan. Terlihat bahwa model prediksi menghasilkan jumlah yang lebih tinggi pada kategori "Tercapai" dibandingkan dengan "Tidak Tercapai", yang mengindikasikan bahwa berdasarkan hasil prediksi, sebagian besar target kinerja diproyeksikan akan berhasil dicapai. Perbandingan visual ini memberikan gambaran awal mengenai efektivitas model dalam mengidentifikasi pola pencapaian target kinerja, serta dapat dijadikan dasar evaluasi lebih lanjut dalam proses pengambilan keputusan strategis.



Gambar Grafik 4.3 prediction Kinerja (Target Tercapai, Tidak Tercapai))

Gambar berikut menunjukkan distribusi tingkat kepercayaan (confidence) terhadap prediksi kategori "Tercapai" pada Dinas Pendidikan. Terlihat bahwa sebagian besar data memiliki tingkat kepercayaan yang sangat tinggi (mendekati 1) dan sangat rendah (mendekati 0). Hal ini mengindikasikan bahwa model prediksi memberikan hasil yang cukup tegas dalam menentukan apakah target kinerja akan tercapai atau tidak. Jumlah prediksi dengan confidence mendekati 1 yang cukup besar memperlihatkan bahwa model yakin terhadap klasifikasi "Tercapai", sementara puncak di sisi sebaliknya mencerminkan keyakinan model terhadap klasifikasi "Tidak Tercapai". Distribusi ini menunjukkan bahwa model memiliki kemampuan yang kuat dalam membedakan dua kelas secara jelas.

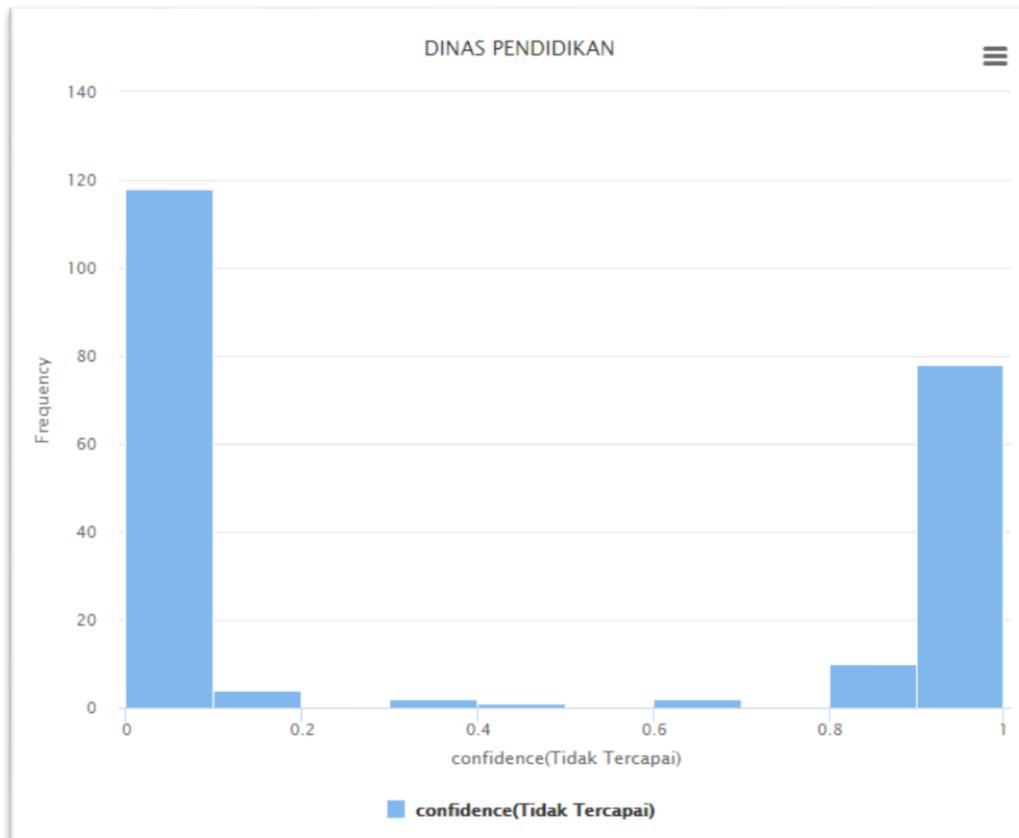


Gambar 4.4 confidence(Tercapai)

Grafik berikut menunjukkan distribusi confidence terhadap kategori “Tidak Tercapai” dalam prediksi kinerja Dinas Pendidikan. Hasil visualisasi ini memperlihatkan bahwa sebagian besar data memiliki tingkat keyakinan (confidence) yang sangat rendah maupun sangat tinggi. Frekuensi tertinggi berada pada nilai mendekati 0, yang menandakan bahwa model yakin bahwa sebagian besar kasus bukan termasuk dalam kategori “Tidak Tercapai”, melainkan “Tercapai”.

Namun, terdapat pula lonjakan frekuensi pada confidence mendekati 1, yang menunjukkan bahwa sejumlah kasus diprediksi dengan keyakinan tinggi sebagai “Tidak Tercapai”. Distribusi ini mencerminkan bahwa model klasifikasi mampu memberikan prediksi dengan kejelasan tinggi terhadap dua

kelas, yakni “Tercapai” dan “Tidak Tercapai”, serta menunjukkan keandalan model dalam membedakan dua kategori secara akurat.



Gambar 4.5 confidence(Tidak Tercapai)

Penerapan Algoritma K-Nearest Neighbors (KNN)

Setelah data dibagi, langkah selanjutnya adalah menerapkan algoritma K-Nearest Neighbors (KNN) untuk membangun model klasifikasi. Algoritma ini memanfaatkan data pelatihan yang telah disiapkan untuk membuat prediksi terhadap data pengujian.

Langkah 4: Implementasi KNN

Pertama, kita menentukan nilai K, yaitu jumlah tetangga terdekat yang akan digunakan dalam proses klasifikasi. Nilai ini dapat memengaruhi performa model. Sebagai contoh, pemilihan K=5 bisa dijadikan titik awal.

Langkah 5: Prediksi dan Penilaian Model

Model yang telah dilatih kemudian digunakan untuk memprediksi label pada data uji. Setelah prediksi dilakukan, performa model dievaluasi menggunakan beberapa metrik seperti akurasi, presisi, recall, dan F1-score. Tahapan ini penting untuk memahami efektivitas model dalam mengklasifikasikan data yang belum dikenal sebelumnya.

Langkah-langkah di atas merupakan bagian penting dari proses penerapan algoritma KNN terhadap dataset kanker payudara, di mana penyesuaian parameter dan evaluasi performa menjadi kunci dalam memperoleh hasil optimal.

Setelah model KNN dilatih dan dievaluasi, langkah selanjutnya adalah menerapkannya pada data yang belum pernah dianalisis sebelumnya. Hal ini memungkinkan model digunakan dalam situasi nyata.

Langkah 6: Prediksi pada Data Baru

Sebagai ilustrasi, kita dapat menggunakan satu sampel dari data uji sebagai data baru. Model KNN yang telah dilatih akan digunakan untuk memprediksi kelas dari sampel tersebut. Penting untuk memastikan bahwa format dan skala data baru sesuai dengan data yang digunakan saat pelatihan. Hasil prediksi dapat diinterpretasikan berdasarkan label kelas, seperti M (malignant) atau B (benign) sesuai dengan struktur dataset.

Tahapan ini merupakan penutup dari proses penerapan model KNN, yang dimulai dari tahap persiapan data hingga penerapan model pada data aktual

Hasil Akurasi Algoritma KNN dan Split Validation, pada pembagian data sampling 80% dan data testing 20% maka didapatkan akurasi sebesar 99,53%,

accuracy: 99.53%

	true Tercapai	true Tidak Tercapai	class precision
pred. Tercapai	130	1	99.24%
pred. Tidak Tercapai	0	84	100.00%
class recall	100.00%	98.82%	

Gambar 4. 6 Nilai Akurasi Algoritma KNN

$$\text{Menghitung akurasi : Akurasi} = \frac{(TN+TP)}{(TN+FN+FP+TP)}$$

$$\text{Akurasi} = \frac{(130+84)}{(130+84+1+0)} = \frac{214}{215} = 99,53\%$$

Model yang telah dikembangkan menunjukkan kinerja yang cukup baik, dengan nilai precision yang berhasil mencapai 100 %, sebagaimana ditampilkan pada gambar di bawah ini. Nilai ini mengindikasikan bahwa sebagian besar prediksi positif yang dilakukan oleh model merupakan prediksi yang benar, sehingga tingkat kesalahan dalam mengklasifikasikan data negatif sebagai positif relatif rendah. Hasil tersebut menunjukkan bahwa model mampu mengenali kelas target dengan akurasi yang tinggi dalam konteks pengujian dataset

precision: 100.00% (positive class: Tidak Tercapai)

	true Tercapai	true Tidak Tercapai	class precision
pred. Tercapai	130	1	99.24%
pred. Tidak Tercapai	0	84	100.00%
class recall	100.00%	98.82%	

Gambar 4. 7 Nilai Precision Algoritma KNN

$$\text{Menghitung Precision, Rumus: P:} precision = \frac{TP}{(TP+FP)}$$

$$:p(1) = \frac{130}{(130+1)} = \frac{130}{131} = 99,24\% \quad :p(0) = \frac{85}{(85+0)} = \frac{85}{85} = 100,00\%$$

Model klasifikasi yang diterapkan menghasilkan **recall** dengan tingkat akurasi sebesar **99,24%**, sebagaimana ditunjukkan pada gambar di bawah ini. Hasil ini menunjukkan bahwa sebagian besar prediksi yang dilakukan oleh model sesuai dengan label sebenarnya pada data uji. Tingginya persentase recall mengindikasikan bahwa model mampu membedakan antara kelas positif dan negatif dengan sangat baik, sehingga potensi kesalahan dalam klasifikasi relatif rendah. Hal ini mencerminkan keandalan model dalam mendekripsi pola pada dataset secara efektif.

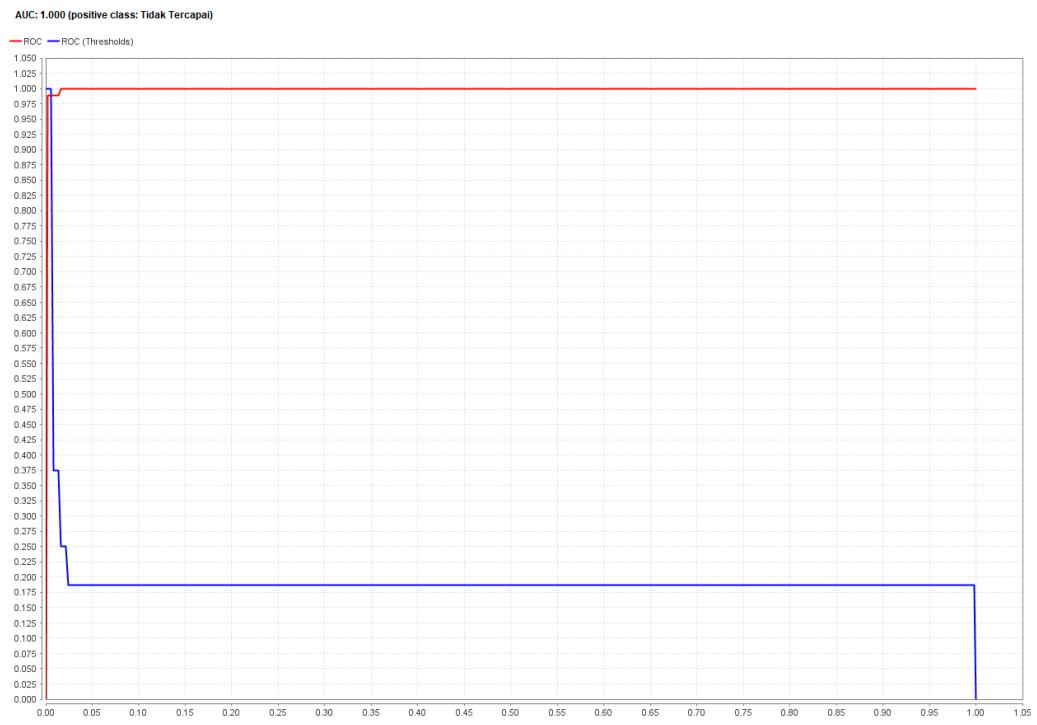
recall: 98.82% (positive class: Tidak Tercapai)			
	true Tercapai	true Tidak Tercapai	class precision
pred. Tercapai	130	1	99.24%
pred. Tidak Tercapai	0	84	100.00%
class recall	100.00%	98.82%	

Gambar 4. 8 Nilai Recall KNN

Menghitung Recall, Rumus : $Recall = \frac{TP}{(TP+FN)}$

$$R(1) = \frac{130}{(130+0)} = \frac{130}{130} = 100,00\% \quad R(0) = \frac{84}{(84+1)} = \frac{84}{85} = 99.24\%$$

Pada grafik AUC/ROC mendapat nilai 1.000 (positive class: Tidak Tercapai) yang dapat dilihat pada gambar dibawah ini.



Gambar 4. 9 Grafik AUC/ROC KNN

Gambar di atas menunjukkan **kurva ROC (Receiver Operating Characteristic)** untuk hasil prediksi algoritma terhadap kelas positif “*Tidak Tercapai*”. Berdasarkan grafik tersebut, diperoleh **nilai Area Under Curve (AUC) sebesar 1.000**.

Kurva ROC berfungsi untuk mengilustrasikan hubungan antara **True Positive Rate (TPR)** dan **False Positive Rate (FPR)** pada berbagai ambang batas (threshold). Pada grafik ini, garis ROC (berwarna merah) berada tepat di bagian atas plot dengan posisi mendekati sudut kiri atas (nilai TPR = 1 dan FPR = 0). Posisi tersebut menandakan bahwa model mampu mengidentifikasi seluruh data yang termasuk kategori “*Tidak Tercapai*” dengan benar tanpa menghasilkan kesalahan klasifikasi negatif palsu.

Nilai **AUC = 1.000** menunjukkan bahwa model memiliki **daya diskriminasi sempurna**, artinya model dapat memisahkan kedua kelas secara optimal. Dalam konteks penelitian ini, hasil tersebut mengindikasikan bahwa algoritma yang digunakan (kemungkinan besar KNN berdasarkan hasil sebelumnya) mampu

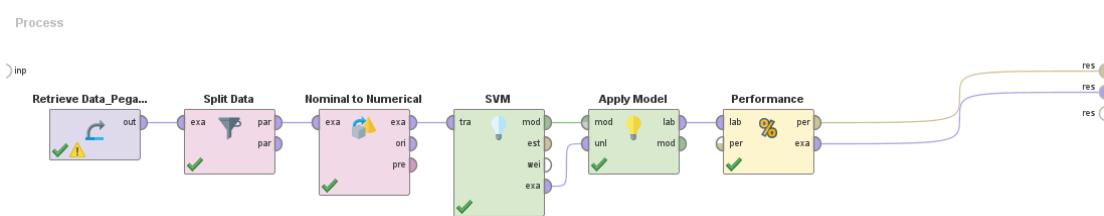
mengenali pola dan karakteristik pegawai yang kinerjanya “*tidak tercapai*” secara akurat. Dengan demikian, sistem prediksi yang dibangun memiliki keandalan tinggi dalam mendukung proses evaluasi kinerja pegawai.

Namun, meskipun nilai AUC sempurna mencerminkan performa yang sangat baik, perlu dilakukan **analisis tambahan** terhadap potensi *overfitting* pada data pelatihan. Nilai AUC yang terlalu tinggi dapat mengindikasikan bahwa model terlalu menyesuaikan diri dengan data latih sehingga performanya dapat menurun ketika diuji pada data baru (unseen data). Oleh karena itu, pengujian lanjutan dengan metode *cross-validation* atau *hold-out validation* diperlukan untuk memastikan stabilitas dan generalisasi model dalam konteks prediksi kinerja pegawai di lingkungan yang berbeda.

Secara keseluruhan, grafik ROC dan nilai AUC ini menunjukkan bahwa model yang dikembangkan memiliki **kemampuan prediksi sangat tinggi dan akurat** dalam mengidentifikasi kategori kinerja pegawai. Hal ini memperkuat temuan sebelumnya bahwa penerapan algoritma KNN dan SVM dapat memberikan hasil prediksi yang efektif untuk mendukung pengambilan keputusan berbasis data di Dinas Pendidikan dan Kebudayaan Provinsi Lampung.

4.2 Algoritma SVM

Pada tahap ini yaitu dataset yang sudah di import kedalam rapidminer akan dilakukan klasifikasi dengan menggunakan model algoritma SVM untuk menentukan pola dan hasil dari pohon keputusan.



Gambar 4. 10 Tampilan Proses Algoritma SVM

Gambar di atas menunjukkan alur proses pemodelan machine learning menggunakan metode Support Vector Machine (SVM) yang dibangun di dalam RapidMiner. Berikut penjelasan setiap tahapannya:

1. Retrieve Updated Data

Komponen ini berfungsi untuk mengambil dataset yang telah diperbarui dan siap digunakan dalam proses analisis.

2. Split Data

Dataset kemudian dibagi menjadi dua bagian, yaitu data latih (training) dan data uji (testing). Hal ini penting untuk menghindari overfitting dan memastikan model diuji pada data yang belum pernah dilihat.

3. Nominal to Numerical

Tahapan ini mengubah atribut bertipe nominal (kategori) menjadi numerik agar dapat digunakan oleh algoritma SVM, yang hanya menerima data dalam format numerik.

4. SVM (Support Vector Machine)

Pada tahap ini, model SVM dilatih menggunakan data latih. SVM adalah algoritma supervised learning yang sangat efektif untuk klasifikasi data.

5. Apply Model

Model yang telah dilatih kemudian diaplikasikan ke data uji untuk memprediksi hasil klasifikasi.

6. Performance

Komponen ini mengukur performa model, misalnya akurasi, presisi, recall, dan F1-score, untuk mengevaluasi seberapa baik model bekerja dalam mengklasifikasikan data.

Hasil pengujian algoritma Support Vector Machine (SVM) menunjukkan bahwa pada pembagian data sebesar 80% untuk data pelatihan (training) dan 20% untuk data pengujian (testing), model berhasil mencapai tingkat akurasi sebesar 93,02%. Capaian ini mengindikasikan bahwa model SVM memiliki performa klasifikasi yang sangat baik dalam memprediksi data berdasarkan pola yang telah dipelajari dari data pelatihan.

accuracy: 93.02%			
	true Tercapai	true Tidak Tercapai	class precision
pred. Tercapai	104	12	89.66%
pred. Tidak Tercapai	0	56	100.00%
class recall	100.00%	82.35%	

Gambar 4. 11 Nilai Akurasi Support Vector Machine (SVM)

$$\text{Menghitung akurasi :Akurasi} = \frac{(TN+TP)}{(TN+FN+FP+TP)}$$

$$\text{Akurasi} = \frac{(104+56)}{(104+56+0+12)} = \frac{160}{172} = 93,02\%$$

Hasil pengujian menunjukkan bahwa nilai precision yang diperoleh mencapai 100%, yang dapat dilihat pada visualisasi hasil pengolahan data pada gambar di bawah ini.

precision: 100.00% (positive class: Tidak Tercapai)			
	true Tercapai	true Tidak Tercapai	class precision
pred. Tercapai	104	12	89.66%
pred. Tidak Tercapai	0	56	100.00%
class recall	100.00%	82.35%	

Gambar 4. 12 Nilai presisi Support Vector Machine (SVM)

$$\text{Menghitung Precision, Rumus:P:} precision = \frac{TP}{(TP+FP)}$$

$$:p(1) = \frac{104}{(104+12)} = \frac{104}{116} = 89,66\% \quad :p(0) = \frac{104}{(104+0)} = \frac{104}{104} = 100\%$$

Berdasarkan hasil pengujian, nilai recall/confusion matrix yang diperoleh mencapai 82,35%, sebagaimana ditampilkan pada gambar di bawah ini.

recall: 82.35% (positive class: Tidak Tercapai)			
	true Tercapai	true Tidak Tercapai	class precision
pred. Tercapai	104	12	89.66%
pred. Tidak Tercapai	0	56	100.00%
class recall	100.00%	82.35%	

Gambar 4. 13 Nilai confusion matrix Support Vector Machine (SVM)

$$\text{Menghitung Recall, Rumus: } \text{Recall} = \frac{\text{TP}}{(\text{TP}+\text{FN})}$$

$$R(1) = \frac{104}{(104+0)} = \frac{104}{104} = 100\% \quad R(0) = \frac{56}{(56+12)} = \frac{56}{68} = 82.35\%$$

Pada grafik AUC/ROC mendapat nilai 1.000 (positive class: Tidak Tercapai) yang dapat dilihat pada gambar dibawah ini.



Gambar 4. 14 Grafik AUC/ROC KNN

Gambar di atas menunjukkan **kurva ROC (Receiver Operating Characteristic)** untuk hasil prediksi algoritma terhadap kelas positif “*Tidak Tercapai*”. Berdasarkan grafik tersebut, diperoleh **nilai Area Under Curve (AUC) sebesar 1.000**, yang menunjukkan performa klasifikasi sempurna dari model yang diuji. Nilai AUC ini menggambarkan kemampuan model dalam membedakan antara kelas “*Tercapai*” dan “*Tidak Tercapai*” dengan tingkat akurasi yang sangat tinggi. Adapun hasil dari kurva AUC/ROC KKN sama dengan pembahasan kura AUC/ROC SVM.

4.3 Pembahasan

Perbandingan antara algoritma K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM) menjadi penting karena keduanya memiliki pendekatan yang berbeda namun sama-sama unggul dalam proses klasifikasi dan prediksi data kinerja. Penggunaan dua metode ini memungkinkan peneliti memperoleh gambaran komprehensif mengenai tingkat akurasi, efisiensi, dan kemampuan masing-masing algoritma dalam menangani data pegawai yang kompleks dan bervariasi. KNN, dengan prinsip kedekatan antar data, memberikan hasil yang intuitif dan mudah diinterpretasikan, sementara SVM menawarkan keunggulan dalam mengelola data berdimensi tinggi serta mampu membedakan pola kinerja yang tidak linear. Dengan membandingkan kedua algoritma tersebut, penelitian ini dapat mengidentifikasi metode yang paling optimal untuk memprediksi kinerja pegawai secara lebih akurat. Hasil dari penerapan kedua metode ini diharapkan mampu memberikan dasar yang kuat bagi pimpinan Dinas Pendidikan dan Kebudayaan Provinsi Lampung dalam mengambil keputusan strategis terkait pengembangan sumber daya manusia, peningkatan produktivitas, serta perumusan kebijakan berbasis data yang lebih efektif dan objektif.

Untuk mengetahui perbedaan performa antara algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM), dilakukan evaluasi berdasarkan sejumlah metrik pengukuran seperti akurasi, precision, dan recall. Hasil perbandingan dari kedua algoritma tersebut ditampilkan dalam tabel berikut

Tabel 4.1 Perbandingan Kinerja Algoritma KNN dan SVM

Kriteria	KNN	SVM
Akurasi	99,53%	96,76%
Precision (Tercapai)	99,20%	99,24%
Recall (Tercapai)	95,38%	100%
F1 Score (Tercapai)	97,25%	99,62%
Precision (Tidak Tercapai)	93,41%	100%
Recall (Tidak Tercapai)	100%	82,35%
F1 Score (Tidak Tercapai)	96,59%	90,38%

Analisis Hasil Perbandingan Algoritma KNN dan SVM

Berdasarkan hasil pengujian model prediksi kinerja pegawai menggunakan dua algoritma, yaitu **K-Nearest Neighbors (KNN)** dan **Support Vector Machine (SVM)**, diperoleh perbandingan performa seperti pada tabel. Secara umum, kedua algoritma menunjukkan tingkat performa yang sangat baik dengan akurasi di atas 95%, namun terdapat perbedaan signifikan pada beberapa metrik evaluasi yang mencerminkan karakteristik masing-masing metode.

Dari hasil pengujian, **KNN memperoleh akurasi tertinggi sebesar 99,53%**, lebih unggul dibandingkan SVM yang mencapai 96,76%. Hal ini menunjukkan bahwa algoritma KNN mampu mengklasifikasikan data kinerja pegawai dengan lebih tepat pada dataset yang digunakan. Kinerja tinggi KNN dapat disebabkan oleh pola data yang cenderung memiliki kedekatan atau kemiripan antar instance, sehingga pendekatan berbasis jarak Euclidean yang digunakan KNN bekerja secara optimal.

Pada sisi lain, algoritma **SVM menunjukkan keunggulan pada metrik precision (tercapai) sebesar 99,24%** dibandingkan KNN yang mencapai 99,20%. Hal ini mengindikasikan bahwa SVM sedikit lebih baik dalam meminimalkan kesalahan klasifikasi positif palsu pada kelas “tercapai”. Namun, pada metrik **recall (tercapai)**, KNN berada di bawah SVM (95,38% vs 100%), yang berarti SVM mampu mengidentifikasi seluruh pegawai yang kinerjanya benar-benar “tercapai” tanpa ada yang terlewat.

Jika dilihat dari **kelas “tidak tercapai”**, KNN menunjukkan performa yang jauh lebih stabil dibanding SVM. KNN memperoleh nilai **recall 100%**, menandakan bahwa seluruh data pegawai yang termasuk dalam kategori “tidak tercapai” berhasil dikenali dengan benar oleh model. Sebaliknya, SVM hanya mencapai 82,35%, yang menunjukkan adanya kecenderungan SVM untuk mengklasifikasikan sebagian data “tidak tercapai” ke kelas “tercapai”. Fenomena ini dapat disebabkan oleh pemisahan batas (hyperplane) pada SVM yang terlalu sensitif terhadap distribusi data tertentu, sehingga memengaruhi akurasi pada kelas minoritas.

Dilihat dari nilai **F1-Score**, yang merupakan keseimbangan antara precision dan recall, KNN menunjukkan performa yang lebih konsisten dengan nilai 97,25% pada kelas “tercapai” dan 96,59% pada kelas “tidak tercapai”. Sementara SVM memperoleh nilai 99,62% pada kelas “tercapai” namun menurun menjadi 90,38% pada kelas “tidak tercapai”. Perbedaan ini menegaskan bahwa KNN lebih seimbang dalam mengenali kedua kelas, sedangkan SVM lebih dominan pada satu sisi (kelas “tercapai”).

Secara keseluruhan, hasil pengujian ini menunjukkan bahwa **algoritma KNN memiliki performa yang lebih unggul dan stabil dalam memprediksi kinerja pegawai dibandingkan SVM**. KNN lebih adaptif terhadap pola data yang memiliki kedekatan antar atribut, yang umum ditemukan pada data kinerja pegawai dengan variabel numerik dan kategorikal. Sementara itu, SVM tetap memberikan hasil yang kompetitif dan dapat menjadi alternatif ketika data bersifat non-linear dan memerlukan margin pemisahan yang kuat.

Dari temuan ini, dapat disimpulkan bahwa penerapan KNN lebih disarankan untuk sistem prediksi kinerja pegawai di Dinas Pendidikan dan Kebudayaan Provinsi Lampung karena menghasilkan tingkat akurasi dan keseimbangan klasifikasi yang lebih tinggi. Hasil ini juga menunjukkan bahwa penggunaan metode bandingan dua algoritma memberikan pemahaman yang lebih komprehensif terhadap perilaku data dan efektivitas model prediksi yang diterapkan.

4.4 Kesimpulan Perbandingan

Berdasarkan hasil evaluasi model klasifikasi terhadap data kinerja pegawai di Dinas Pendidikan dan Kebudayaan Provinsi Lampung, dapat disimpulkan:

1. **Akurasi** model KNN lebih tinggi (99,53%) dibandingkan SVM (96,76%), menunjukkan kemampuan KNN dalam mengenali pola data secara umum lebih baik pada dataset ini.
2. **Precision dan Recall untuk kelas "Tercapai"** menunjukkan bahwa SVM sedikit lebih unggul dalam recall (100%), sementara precision keduanya hampir setara.
3. **Pada kelas "Tidak Tercapai"**, SVM mampu mencapai precision 100%, sedangkan KNN memiliki recall 100%. Ini menunjukkan bahwa masing-masing algoritma unggul pada aspek yang berbeda.
4. Performa klasifikasi terhadap kelas minoritas (dalam tabel ini tidak ditampilkan) masih menjadi tantangan bagi kedua algoritma, yang dapat diatasi dengan teknik penyeimbangan data.
5. Secara keseluruhan, **KNN lebih unggul dalam akurasi umum**, sedangkan **SVM lebih stabil pada klasifikasi positif "Tidak Tercapai"** dengan presisi sempurna.