

BAB II

LANDASAN TEORI

2.1. Konsep Dasar Kinerja

Konsep Dasar Kinerja Pegawai Kinerja pegawai adalah tingkat keberhasilan seseorang dalam melaksanakan tugas-tugas yang menjadi tanggung jawabnya sesuai dengan standar yang telah ditetapkan oleh organisasi. Menurut Mangkunegara (2005), kinerja pegawai dipengaruhi oleh dua faktor utama, yaitu kemampuan (ability) dan motivasi (motivation). Kemampuan mencakup pengetahuan, keterampilan, dan kompetensi individu, sedangkan motivasi berkaitan dengan dorongan internal yang memengaruhi semangat kerja seseorang untuk mencapai tujuan tertentu. Faktor lingkungan kerja, seperti dukungan atasan, budaya organisasi, dan fasilitas, juga memiliki pengaruh signifikan terhadap kinerja pegawai[7].

Dalam konteks organisasi pemerintahan, kinerja pegawai menjadi salah satu indikator utama dalam menilai efektivitas pelayanan publik. Kinerja yang optimal tidak hanya mencerminkan keberhasilan individu, tetapi juga menjadi cerminan keberhasilan organisasi dalam mencapai target strategisnya. Oleh karena itu, pengelolaan kinerja pegawai melalui metode yang efektif sangat diperlukan untuk memastikan peningkatan produktivitas dan efisiensi dalam pelayanan publik[7][8].

2.2. Teknologi dalam Evaluasi Kinerja Pegawai

Teknologi telah memberikan dampak signifikan dalam proses evaluasi kinerja pegawai. Sistem berbasis data analytics memungkinkan pengumpulan, pengolahan, dan analisis data kinerja secara otomatis, sehingga memberikan hasil yang lebih cepat dan akurat dibandingkan metode manual. Selain itu, teknologi machine learning telah menjadi alat yang populer dalam memprediksi kinerja pegawai dengan memanfaatkan data historis untuk mengidentifikasi pola dan tren tertentu.

Implementasi teknologi dalam evaluasi kinerja memberikan berbagai manfaat, seperti mengurangi bias subjektif, meningkatkan transparansi, dan mendukung pengambilan keputusan berbasis bukti (*evidence-based decision-making*). Dalam organisasi pemerintahan, penggunaan teknologi ini dapat meningkatkan efisiensi pengelolaan sumber daya manusia (SDM) dan memberikan masukan yang akurat untuk perencanaan strategis[9].

2.3. Data Mining

Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu[10]. Kaka's data mining meramalkan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting. Analisis yang diotomatisasi yang dilakukan oleh data mining melebihi yang dilakukan oleh sistem pendukung keputusan tradisional yang sudah banyak digunakan. Secara khusus, koleksi metode yang dikenal sebagai 'data mining' menawarkan metodologi dan solusi teknis untuk mengatasi analisis data medis dan konstruksi prediksi model [3]

Berdasarkan tugas dan tujuan analisis, proses data mining dapat dibagi menjadi dua kategori utama, Tergantung pada adanya target variabel dan metode belajar (*learning*) yaitu antara proses belajar yang diawasi (*supervised*) dan tanpa pengawasan (*unsupervised*) [11]

2.4. Klasifikasi.

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri dapat berupa aturan jika-maka (*if-then*), berupa pohon keputusan (*decision tree*), jaringan saraf tiruan (*neural network*).

Klasifikasi adalah urutan yang sangat penting dalam data komunitas pertambangan. Klasifikasi adalah salah satu prediksi teknik data mining yang membuat prediksi tentang data nilai menggunakan hasil yang diketahui yang ditemukan dari kumpulan data yang berbeda. Masalah akurasi dari banyak algoritma klasifikasi adalah diketahui mengalami penurunan informasi saat dihadapi dengan data yang tidak seimbang, misalnya ketika distribusi sampel lintas kelas sangat miring[12]

Dalam klasifikasi, ada variabel kategoris target, seperti braket pendapatan, yang, misalnya, dapat dipartisi menjadi tiga kelas atau kategori: berpenghasilan tinggi, menengah pendapatan, dan pendapatan rendah. Model data mining memeriksa satu set besar catatan, masing-masing catatan yang berisi informasi tentang variabel target serta satu set input atau prediktor variable. Contoh tugas klasifikasi dalam bisnis dan penelitian meliputi:

1. Menentukan apakah transaksi kartu kredit tertentu adalah penipuan
2. Menempatkan siswa baru pada jalur tertentu yang berkaitan dengan kebutuhan khusus

3. Menilai apakah aplikasi hipotek adalah risiko kredit yang baik atau buruk
4. Mendiagnosis apakah ada penyakit tertentu
5. Menentukan apakah surat wasiat ditulis oleh almarhum yang sebenarnya, atau curang oleh orang lain
6. Mengidentifikasi apakah perilaku keuangan atau pribadi tertentu menunjukkan kemungkinan ancaman teroris

Klasifikasi yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari algoritma cerdas komputer. Sedangkan klasifikasi yang dilakukan dengan bantuan teknologi, memiliki beberapa algoritma, diantaranya *Naïve Bayes*, *Support Vector Machine*, *Decision Tree*, Fuzzy dan Jaringan Saraf Tiruan [13].

2.5. *Support Vector Machine*

Support Vector Machine (SVM) adalah algoritma *machine learning* yang digunakan untuk tugas klasifikasi dan regresi dengan pendekatan berbasis *hyperplane*. SVM bertujuan untuk menemukan *hyperplane* optimal yang memisahkan data dari dua kelas dengan margin maksimum. Margin yang lebih besar memberikan generalisasi yang lebih baik pada data baru.

Keunggulan utama SVM adalah kemampuannya menangani data dengan dimensi tinggi dan data yang tidak terpisahkan secara linear melalui penggunaan kernel trick. Kernel trick memungkinkan transformasi data ke dalam ruang dimensi yang lebih tinggi untuk membuatnya lebih mudah dipisahkan. Namun, kelemahan SVM meliputi kompleksitas komputasi yang tinggi dan sensitivitas terhadap pemilihan parameter, seperti jenis kernel dan nilai regularisasi[8], adapun rumus dari SVM bisa dilihat pada persamaan 2.1 berikut.

$$w \cdot x + b = 0 \quad (2.1)$$

di mana:

1. w adalah vektor bobot
2. x adalah vektor data
3. b adalah bias

2.6. Algoritma K-Nearest Neighbors (KNN)

Algoritma K-Nearest Neighbors (KNN) adalah salah satu metode supervised learning yang digunakan untuk tugas klasifikasi dan regresi. KNN bekerja dengan cara membandingkan data baru dengan data yang telah ada berdasarkan jarak terdekat. Algoritma ini menentukan kelas data baru berdasarkan mayoritas kelas dari data tetangga terdekat (nearest neighbors).

Keunggulan KNN adalah kesederhanaannya dalam implementasi dan kemampuannya memberikan hasil yang baik pada dataset dengan pola sederhana. Namun, KNN memiliki beberapa kelemahan, seperti sensitivitas terhadap outlier, ketergantungan pada skala data, dan kebutuhan komputasi yang tinggi untuk dataset berukuran besar. Oleh karena itu, diperlukan teknik preprocessing data, seperti normalisasi, untuk meningkatkan performa algoritma ini.[7] Untuk menghitung jarak euclidean dan jarak manhattan menggunakan persamaan 2.2 dan 2.3 berikut.

Jarak Euclidean (umum digunakan):

$$d_{Euclidean}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.2)$$

Jarak Manhattan (alternatif):

$$d_{Manhattan}(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2.3)$$

Prediksi kelas dalam KNN:

$$\hat{y} = \text{mode}(\{y_i | x_i \in k\text{-tetangga terdekat}\})$$

Kelas yang paling sering muncul di antara k tetangga dipilih sebagai hasil klasifikasi.

2.7. Confusion Matrix

Pengujian dengan Confusion Matrix Pada tahap ini pengujian model penelitian dilakukan dengan metode Confusion Matrix yang mempresentasikan hasil evaluasi model dengan menggunakan tabel matrik, Jika dataset terdiri dari 2 kelas, kelas pertama dianggap positif dan kelas kedua dianggap negatif. Evaluasi menggunakan confusion matrix menghasilkan nilai Akurasi, Precision, Recall, serta F-Measure. Akurasi dalam klasifikasi merupakan presentasi ketepatan record data diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi. Precision merupakan proposisi yang diprediksi positif yang juga positif benar pada data sebenarnya. Recall merupakan proporsi kasus positif yang sebenarnya diprediksi positif secara benar. True Positive (TP) merupakan jumlah record positif dalam dataset yang diklasifikasikan positif. True Negative (TN) merupakan jumlah record negative dalam dataset yang diklasifikasikan positif[14].

False Positive (FP) merupakan jumlah record negatif dalam dataset yang diklasifikasikan positif. False Negative (FN) merupakan jumlah record positif dalam dataset yang diklasifikasikan negatif. Berikut adalah persamaan model Confusion Matrix.

Akurasi merupakan salah satu metrik evaluasi yang digunakan untuk mengukur tingkat ketepatan prediksi model. Untuk menghitung akurasi menggunakan persamaan 2.4 berikut

$$:Akurasi = \frac{(TN+TP)}{(TN+FN+FP+TP)} \quad (2.4)$$

Precision digunakan untuk mengukur seberapa besar proporsi dari kelas data positif yang berhasil diprediksi dengan benar dari keseluruhan hasil prediksi kelas positif.

Untuk menghitung *precision* menggunakan persamaan 2.5 berikut:

$$precision = \frac{TP}{(TP+FP)} \dots (2.5)$$

Recall mengukur proporsi data positif yang berhasil dikenali dengan benar oleh model. Metrik ini penting untuk melihat seberapa baik model menangkap seluruh kasus positif. Untuk menghitung *recall* menggunakan persamaan 2.6 berikut:

$$Recall = \frac{TP}{(TP + FN)} \quad (2.6)$$

2.8. Kurva ROC dan AUC

Dalam Machine Learning, pengukuran kinerja adalah tugas penting. Jadi dalam masalah klasifikasi, kita dapat mengandalkan Kurva AUC - ROC. Ketika kita perlu memeriksa atau memvisualisasikan kinerja masalah klasifikasi multi-kelas, kita menggunakan kurva AUC (Area Under The Curve) ROC (Receiver Operating Characteristics). Ini adalah salah satu metrik evaluasi terpenting untuk memeriksa kinerja model klasifikasi apa pun. Itu juga ditulis sebagai AUROC (Area Di Bawah Karakteristik Operasi Penerima) [6]. Metode umum untuk menghitung daerah dibawah kurva ROC yaitu Area Under Curve (AUC) dimana bidang yang berada dibawah kurva mempunyai nilai yang selalu berada pada nilai 0,0 dan 1,0. Namun yang menarik untuk dihitung adalah yang mempunyai luas diatas 0,5, semakin tinggi luasnya maka akan semakin baik seperti yang disajikan berikut ini [15]

0.9-1.0 = klasifikasi yang sangat baik (*Excellent Classification*)

0.8-0.9 = klasifikasi baik (*Good Classification*)

0.7-0.8 = klasifikasi rata-rata (*Fair Classification*)

0.6-0.7 = klasifikasi rendah (*Poor Classification*)

0.5-0.6 = kegagalan (*Failure Classification*)

2.9. Tinjauan Studi

Berikut adalah ringkasan dari beberapa penelitian sebelumnya yang terkait dengan klasifikasi *data mining*

Tabel 2.1. Ringkasan Tinjauan Studi

No	Judul	Penulis	Objek	Algoritma	Hasil
1	IMPLEMENTASI DATA MINING DALAM PREDIKSI KINERJA KEUANGAN DAN OPERASIONAL PADA PERUSAHAAN AVIASI MENGGUNAKAN LINEAR REGRESSION	Andika ¹ Juwita ² Nugroho ³ ,Entis Sutrisna ⁴	Dalam pengelolaannya, perusahaan ini memiliki kinerja keuangan dan operasional untuk memprediksi kinerja keuangan dan operasional di masa yang akan datang, sehingga memudahkan perusahaan dalam menilai kinerja keuangan dan operasional agar tetap stabil. Hal ini berguna untuk meminimalisir kerugian.	LINEAR REGRESSION	sistem yang dapat memprediksi informasi keuangan dan operasional sebagai sumber untuk mendukung penguatan dalam pengambilan keputusan, khususnya dari sisi keuangan perusahaan.
2	Penerapan Data Mining Dalam Menentukan Kinerja Karyawan Terbaik Dengan Menggunakan	Alfi Novia Zahrotul Hidayah ¹ , Anief Fauzan Rozi ²	Pada penelitian ini dilakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat	Decision Tree 4.5	dalam penilaian dosen terbaik di Universitas Mercu Buana Yogyakarta dalam

	Metode Algoritma C4.5 (Studi Kasus : Universitas Mercu Buana Yogyakarta)		prediktif yang digunakan untuk membentuk pohon keputusan (Decision Tree)		penelitian ini memiliki tingkat akurasi yang termasuk dalam klasifikasi sangat baik yaitu sebesar 85,52% yang didapat dari hasil uji coba menggunakan tools Rapid Miner dengan 80% data sebagai data training dan 20% data uji.
3	ANALISIS KINERJA PEGAWAI MENGGUNAKAN ALGORITMA KMEANS PADA DINAS PENDIDIKAN DAN KEBUDAYAAN KABUPATEN BENGKULU TENGAH	Panggi Oktara ¹ , Liza Yulianti ² , Jhoanne Fredricka ³	Penilaian kinerja dengan form yang sudah ditetapkan. Setiap tahunnya akan dilakukan evaluasi kinerja pegawai untuk mengambil suatu keputusan dan bahan pertimbangan dalam mengetahui kinerja yang diperoleh pegawai tersebut.	ALGORITMA KMEANS	pengelompokan terhadap data penilaian kinerja pegawai, didapatkan bahwa jumlah pegawai yang berada pada cluster tinggi (Cluster I) 22% dan jumlah pegawai yang berada pada cluster rendah (Cluster II) 82%. Fungsional dari aplikasi sudah berjalan dengan baik dan aplikasi dapat membantu pihak Dinas Pendidikan dan Kebudayaan Kabupaten Bengkulu

					Tengah dalam mengetahui kinerja para pegawainya berdasarkan 2 kelompok yaitu cluster tinggi dan cluster rendah menggunakan confusion matrix
--	--	--	--	--	---

Berdasarkan temuan-temuan dari penelitian-penelitian sebelumnya, dapat disimpulkan bahwa penerapan Algoritma KNN dan SVM memiliki potensi untuk diimplementasikan dalam berbagai kebutuhan klasifikasi. Oleh karena itu, hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pemanfaatan Algoritma KNN dan SVM secara optimal untuk memprediksi kinerja pegawai di Dinas Pendidikan Kota Bandar Lampung