

Analisis Seleksi Fitur untuk Optimasi Metode Klasifikasi k-NN pada Studi Kasus Penilaian Kinerja Karyawan

Irene Tangkawarow¹, Dandy Pramana Hostiadi², Nenden Siti Fatonah³, Mohammad Yazdi⁴, Eva Hariyanti⁵

¹Universitas Negeri Manado, ²Institut Teknologi dan Bisnis STIKOM Bali, ³Universitas Esa Unggul,

⁴Universitas Tadulako, ⁵Universitas Airlangga

e-mail: ¹irene.tangkawarow@unima.ac.id, ²dandy@stikom-bali.c.id, ³nenden.siti@esaunggul.ac.id, ⁴yazdi.diyana@gmail.com, ⁵eva.hariyanti@fst.unair.ac.id

Diajukan: 1 Juli 2023; Direvisi: 3 Agustus 2023; Diterima: 7 Agustus 2023

Abstrak

Model Klasifikasi banyak digunakan dalam rangka menganalisis dan menemukan jenis kategori kelas data. Salah satu bentuk pemanfaatan metode klasifikasi adalah mengklasifikasikan hasil penilaian pengukuran kinerja karyawan. Metode klasifikasi yang umum dan dapat digunakan antara lain adalah metode Decision Tree, Naive Bayes, k-NN dan Random Forest. Namun tidak semua metode dapat menghasilkan performa yang baik dalam penilaian kinerja Karyawan. Sehingga perlu dilakukan optimasi misalnya melalui penggunaan seleksi fitur. Beberapa penelitian telah dilakukan optimasi metode klasifikasi melalui penggunaan metode seleksi fitur dalam penilaian kinerja karyawan. Namun optimasi ini dipengaruhi oleh karakteristik data yang digunakan. Tidak semua teknik seleksi fitur sesuai untuk meningkatkan hasil klasifikasi dan jumlah penggunaan fitur dapat mempengaruhi performa model klasifikasi. Penelitian ini mengusulkan teknik analisis penggunaan jumlah fitur pada data kinerja dosen melalui metode seleksi fitur ANOVA untuk meningkatkan performa model klasifikasi metode k-NN. Tujuannya adalah untuk mendapatkan jumlah fitur yang terbaik dalam peningkatan performa metode klasifikasi k-NN. Hasil penelitian menunjukkan bahwa jumlah fitur terbaik dari metode ANOVA adalah sejumlah 5 fitur dengan hasil akurasi klasifikasi k-NN sebesar 0.839, precision 0.8323, recall 0.839 dan F1-score 0.833. Teknik analisis ini dapat digunakan oleh sebuah perusahaan dalam mengutamakan fitur terbaik dalam menilai kualitas kinerja karyawannya.

Kata kunci: Seleksi fitur, k-NN, Kinerja karyawan, Klasifikasi.

Abstract

Classification models are widely used to analyze and find types of data class categories. One form of utilizing the classification method is to classify the results of employee performance measurement assessments. Common classification methods include the Decision Tree, Naive Bayes, k-NN and Random Forest. However, not all methods can produce good performance in employee performance appraisals. So optimization needs to be done, for example, by using selection features. Several studies have optimized classification methods through feature selection methods in employee performance assessment. However, this optimization is influenced by the characteristics of the data used. Not all feature selection techniques are suitable for improving classification results, and the number of features used can affect the performance of the classification model. This research proposes an analysis technique using the number of features on the performance of data provided through the ANOVA feature selection method to improve the performance of the k-NN method classification model. The goal is to get the best number of features to increase the performance of the k-NN classification method. The research results show that the best features from the ANOVA method are five features with k-NN classification accuracy results of 0.839, precision 0.8323, recall 0.839 and F1-score 0.833. A company can use this analysis technique to prioritize the best features in assessing the quality of its employees' performance.

Keywords: feature selection, k-NN, work performance, classification .

1. Pendahuluan

Di era digital saat ini, penggunaan sistem berbasis kecerdasan buatan atau dikenal dengan Artificial Intelligence (AI) banyak dimanfaatkan dalam proses pengolahan data [1, 2]. Pengimplementasian AI yang sering digunakan adalah untuk melakukan menganalisis karakteristik, pola dan data analitik[3]. Teknik ini memiliki irisan dengan teknik penambangan data. Misalnya adalah melakukan klasifikasi, klusterisasi dan asosiasi.

Penggunaan analisis data dengan metode klasifikasi banyak digunakan seperti menggunakan metode Naive Bayes, Decision tree, Random Fores dan k -NN. Misalnya dalam studi kasus penilaian kinerja karyawan[4], metode klasifikasi digunakan untuk dapat mengklasifikasikan penilaian berdasarkan parameter penilaian karyawan secara otomatis melalui metode klasifikasi berbasis *machine learning* dan menghasilkan analisis yang akurat[5]. Jika dibandingkan dengan penilaian secara manual, penilaian kinerja dengan menerapkan metode klasifikasi dapat dilakukan secara cepat dan tepat untuk mengolah data besar.

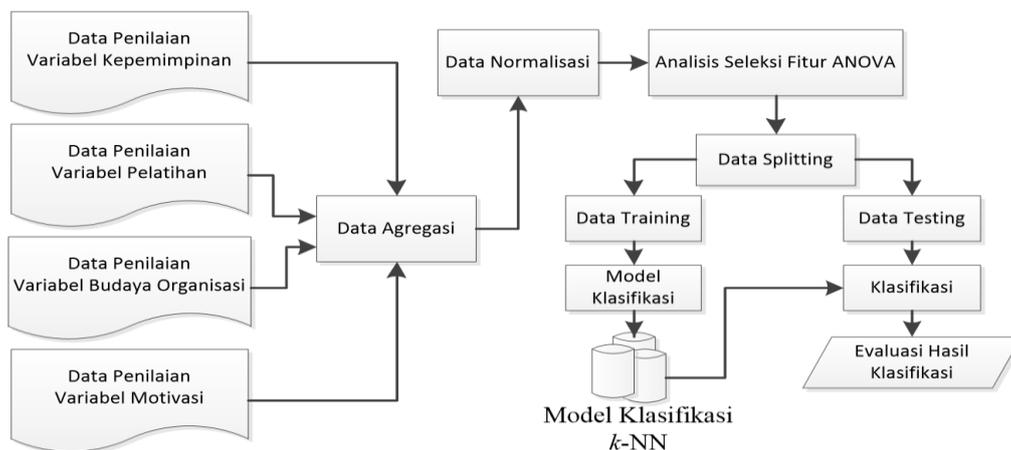
Data kinerja karyawan dapat dianalisis dari 4 variabel utama yaitu motivasi [6–8], pelatihan [4, 9, 10], budaya organisasi [11] dan kepemimpinan [11–14]. Variabel budaya organisasi terdiri dari parameter keyakinan, perilaku, sistem kontrol, norma-norma, kepercayaan, nilai kebersamaan dan pembagian kerja. Variabel kepemimpinan terdiri dari parameter konsultasi, manajemen demokratis, desentralisasi, delegasi, bertanggungjawab, kecerdasan, evaluasi rutin. Variabel motivasi memiliki parameter keinginan, tujuan, kebutuhan, usaha, sikap, fasilitas dan *team work*. Sedangkan pada variabel pelatihan terdapat sub-parameter: Kemampuan, Pendidikan, Latihan, Pengetahuan, Keterampilan, Sikap dan Fleksibilitas. Beberapa penelitian mengenai analisis kinerja karyawan telah dikenalkan pada penelitian sebelumnya. Misalnya pada studi kasus penilaian kinerja dosen yang dikenalkan pada [6, 7, 11]. Penelitian ini digunakan untuk menganalisis dan mendapatkan nilai pengaruh di antara parameter pengukuran. Pengukuran pengaruh diukur dengan penggunaan SEM (*Structural Equation Modelling*) dan secara sekuensial. Penelitian ini mampu menganalisis hubungan setiap parameter pengukuran dengan hasil nilai yang optimal.

Pengolahan data pada analisis kinerja karyawan, tidak hanya dilihat dari jumlah data, namun dilihat dari dimensi data. Dalam konsep metode data analitic, pengolahan dimensi data adalah terhadap penanganan jumlah fitur yang banyak dan relevan terhadap hasil metode analisis data[15–18]. Teknik yang umum digunakan dalam penanganan masalah jumlah fitur dalam model analisis data dikenal dengan istilah seleksi fitur[9, 19, 20]. Seleksi fitur sering dan dapat digunakan dalam mengoptimasi kinerja metode pengolahan data mining, misalnya adalah metode klasifikasi k -NN[19, 20]. Penggunaan optimasi metode seleksi fitur pada metode k -NN bergantung pada jenis atau karakteristik data yang digunakan. Tidak semua seleksi fitur sesuai dengan metode k -NN, dan tidak semua jenis data dapat menghasilkan diolah secara optimal oleh metode klasifikasi. Sehingga perlu dilakukan analisis yang mendalam terhadap penggunaan metode seleksi fitur dalam hal perankingan dan menentukan jumlah fitur yang tepat dalam sebuah metode klasifikasi, untuk emnghasilkan meodel klasifikasi yang optimal.

Dalam penelitian ini mengusulkan teknik analisis terhadap fitur dan jumlah fitur yang relevan dalam model klasifikasi k -NN pada data pengukuran kinerja karyawan. Terdapat 4 parameter pengukuran terhadap penilaian kinerja karyawan dengan setiap parameter terdiri dari 7 sub-parameter. Metode seleksi fitur yang dianalisis adalah ANOVA. Metode ANOVA dipilih karena sering digunakan dalam penelitian berbasis machine learning. Tujuannya adalah untuk mendapatkan fitur dan jumlah fitur terbaik dalam pemodelan klasifikasi berbasis *machine learning*. Manfaat dari analisis penggunaan dan jumlah fitur terbaik dapat membantu perusahaan dalam mengutamakan parameter penilaian karyawannya.

2. Metode Penelitian

Penelitian ini melakukan analisis dari metode seleksi fitur ANOVA terhadap performa model klasifikasi k -NN. Alur penelitian dapat ditunjukkan pada Gambar 1.



Gambar 1. Alur Model.

Tujuan dari penelitian ini adalah melakukan analisis terhadap fitur yang paling berpengaruh dan jumlah fitur yang paling optimal untuk mendapatkan hasil terbaik dari metode klasifikasi. Dalam penelitian ini, menggunakan data pada studi kasus penilaian kinerja karyawan yang didapatkan dalam proses kuesioner.

2.1. Data Penilaian Kinerja Karyawan

Pada penelitian ini data yang digunakan adalah data kinerja karyawan yang terdiri dari 4 variabel utama yaitu kepemimpinan, budaya organisasi, motivasi dan pelatihan. Variabel budaya organisasi terdiri dari parameter keyakinan, perilaku, sistem kontrol, norma-norma, kepercayaan, nilai kebersamaan dan pembagian kerja. Variabel kepemimpinan terdiri dari parameter konsultasi, manajemen demokratik, desentralisasi, delegasi, bertanggungjawab, kecerdasan, evaluasi rutin. Variabel motivasi memiliki parameter keinginan, tujuan, kebutuhan, usaha, sikap, fasilitas dan team-work. Sedangkan pada variabel pelatihan terdapat sub-parameter: Kemampuan, Pendidikan, Latihan, Pengetahuan, Keterampilan, Sikap dan Fleksibilitas. Data penilaian ini diambil melalui penyebaran kuesioner dan dilakukan penilaian dalam bentuk skala nilai 0 sampai dengan 100. Kemudian data ini diproses pada tahap berikutnya yaitu data agregasi.

2.2. Data Agregasi

Pada tahap ini dilakukan agregasi data menjadi satu data vektor. Jika setiap variabel penilaian kinerja karyawan memiliki 7 sub parameter, dimana budaya organisasi dinotasikan sebagai O , Kepemimpinan dinotasikan sebagai G , motivasi dinotasikan sebagai V dan pelatihan dinotasikan sebagai T , maka $o \in O, O = \{o_1, o_2, o_3, o_4, o_5, o_6, o_7\}$, $g \in G, G = \{g_1, g_2, g_3, g_4, g_5, g_6, g_7\}$, $v \in V, V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ dan $t \in T, T = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7\}$. Penilaian kinerja karyawan dinotasikan sebagai K_{kar} maka proses data agregasi dilakukan terhadap variabel kinerja menjadi $K_{kar} = \{O_{(i,7)}, G_{(i,7)}, V_{(i,7)}, T_{(i,7)}\}$.

2.3. Data Normalisasi

Pada tahap data nilai dari K_{kar} dilakukan proses normalisasi data. Tujuan dari proses data normalisasi adalah melakukan standarisasi skala nilai setiap nilai yang terdapat pada subparameter menjadi terstandarisasi dalam nilai skala 0 hingga 1. Di penelitian ini, teknik normalisasi menggunakan linear scalling, dimana persamaan normalisasi data ditunjukkan pada persamaan (1).

$$c' = \frac{(c - c_{min})}{(c_{max} - c_{min})}, \tag{1}$$

dimana, c' adalah hasil normalisasi, c adalah nilai data pada setiap fitur, c_{min} adalah nilai minimum dalam seluruh nilai fitur dan c_{max} adalah nilai maksimum pada seluruh nilai fitur.

2.4. Analisis Seleksi Fitur ANOVA

Analisis seleksi fitur dalam penelitian ini menggunakan metode ANOVA. Setiap fitur yang ada pada K_{kar} akan dihitung dan dianalisis seberapa banyak jumlah dan fitur terbaik yang dapat digunakan dalam pengoptimalan performa model klasifikasi. Pengukuran seleksi fitur ANOVA ditunjukkan pada algoritma 1.

```

Algorithm 1. ALGORITMA ANOVA
FUNCTION ANOVA_feature_selection( $X, y, threshold\_signifikansi$ ):
    num_features = jumlah fitur dalam  $X$ 
    num_samples = jumlah sampel dalam  $X$ 

    for each fitur in  $X$ :
        Hitung nilai  $F$  ANOVA untuk fitur tersebut:
        - Hitung rata-rata kuadrat antar kelompok ( $MSB$ )
        - Hitung rata-rata kuadrat dalam kelompok ( $MSW$ )
        - Hitung nilai  $F = MSB / MSW$ 

    Urutkan fitur berdasarkan nilai  $F$ -nya secara menurun

    fitur_terpilih = []
    untuk setiap fitur dalam fitur_terurut:
        jika nilai  $F$  dari fitur >  $threshold\_signifikansi$ :
            fitur_terpilih.tambahkan(fitur)

    RETURN fitur_terpilih
    
```

2.5. Data Splitting

Pada tahap ini data kinerja karyawan K_{kar} dibagi menjadi dua bagian data yaitu data training dan data testing. Data training adalah data yang digunakan pada proses model klasifikasi k -NN. Sedangkan data testing adalah data yang digunakan dalam proses pengujian data. Komposisi data splitting yang digunakan adalah 70% untuk data training dan 30% digunakan dalam data testing.

2.6. Proses Pemodelan Klasifikasi

Pada proses klasifikasi, data training dilatih dengan menggunakan metode k -NN yang ditunjukkan pada persamaan (2). Klasifikasi dilakukan dengan menentukan k sebesar 5.

$$\hat{y} = \arg \max_b \sum_{i=1}^{k=5} 1(y_i = b), \tag{2}$$

dimana \hat{y} adalah kelas yang diprediksi untuk titik data pada sebaran data X . b beriterasi nelalui semua label kelas yang mungkin terjadi. $1(y_i = b)$ adalah fungsi indikator yang mengembalikan 1 jika y_i sama dengan c , dan 0 sebaliknya. k adalah jumlah tetangga terdekat yang akan dipertimbangkan.

2.7. Proses Evaluasi Model

Pada tahap ini dilakukan evaluasi terhadap performa kerja dari model klasifikasi k -NN. Dalam evaluasi, pengujian dilakukan melalui analisis bertahap dari penggunaan 1 fitur hingga 28 fitur. Pada model klasifikasi evaluasi dilakukan dengan melihat akurasi, precision dan recall yang dilakukan terhadap seleksi fitur bertahap. Nilai akurasi, precision, F1-score dan recall didapatkan dari penelusuran nilai confusion matrix, yaitu true positif, false negative, true negative dan false positive.

3. Hasil dan Pembahasan

Di penelitian ini, model diolah dengan spesifikasi komputer dengan processor core i5, RAM 8 GB dan storage SSD 256GB. Bahasa pemrograman yang digunakan adalah Python versi 3. Data yang digunakan dalam penelitian ini adalah data kualitas kinerja karyawan yang didalamnya terdapat 28 fitur dan 1 kelas label penilaian. Jumlah data karyawan yang digunakan dalam penelitian ini adalah sejumlah 285 karyawan. Data yang digunakan sudah dalam bentuk data numerik yang memiliki skala nilai 0 hingga 1. Contoh tampilan data yang digunakan ditunjukkan pada Tabel 1.

Tabel 1. Contoh Data Penilaian Kinerja Karyawan.

K_{kar} ke- i	1	2	3	4	5	6	7	8	9	10	11	...	
o_1	94	68	70	75	69	78	84	77	67	93	89	...	
Budaya	o_2	62	68	79	86	91	69	75	91	87	83	81	...
Organisasi (O)	o_3	90	82	94	83	79	62	91	92	78	62	86	...
	o_4	62	81	66	77	79	81	87	86	90	91	94	...

	o_5	87	68	71	77	71	68	67	89	79	68	86	...
	o_6	75	68	71	91	64	87	70	89	73	94	90	...
	o_7	76	85	79	71	77	90	76	84	66	72	86	...
Kepemimpinan (G)	g_1	89	89	81	79	68	76	70	79	91	88	77	...
	g_2	78	73	76	72	93	70	79	63	80	71	79	...
	g_3	66	70	88	80	94	77	82	84	67	94	74	...
	g_4	70	73	80	85	89	73	92	80	82	74	88	...
	g_5	64	74	83	87	85	93	66	88	80	86	86	...
	g_6	77	77	64	83	74	72	79	87	88	66	73	...
	g_7	80	91	73	88	85	70	62	89	64	93	80	...
Motivasi (V)	v_1	94	70	90	76	72	82	71	83	81	80	80	...
	v_2	63	81	66	78	65	91	91	86	81	70	89	...
	v_3	69	80	88	86	94	84	67	63	69	72	87	...
	v_4	89	83	94	67	72	84	82	73	78	64	64	...
	v_5	66	85	68	83	73	67	80	78	83	68	70	...
	v_6	87	76	69	82	76	75	93	91	68	74	79	...
	v_7	90	88	85	71	92	72	69	63	81	78	83	...
Pelatihan (T)	t_1	88	71	92	80	75	80	74	70	66	91	82	...
	t_2	92	88	90	93	71	67	79	78	86	94	63	...
	t_3	93	78	94	80	69	73	69	93	78	90	63	...
	t_4	65	71	74	86	79	94	94	89	87	82	90	...
	t_5	88	77	76	89	93	76	92	78	82	76	67	...
	t_6	88	82	82	77	67	69	67	62	91	91	89	...
	t_7	91	63	87	91	64	92	75	91	78	66	84	...
Kelas Kinerja	Cukup Baik	Cukup Baik	Baik	Sangat Baik	Baik	Cukup Baik	Cukup Baik	Sangat Baik	Baik	Baik	Baik	...	

Tahap awal dalam penelitian ini adalah data agregasi, dimana data kinerja karyawan yang terdiri dari 4 variabel dan 7 sub-variabel diurutkan menjadi satu data vektor menjadi $K_{kar} = \{(o_1, o_2, o_{...}, o_7), (g_1, g_2, g_{...}, g_7), (v_1, v_2, v_{...}, v_7), (t_1, t_2, t_{...}, t_7)\}$. Kemudian dilanjutkan dengan proses data normalisasi, yaitu menstandarisasi nilai data dengan skala 0 hingga 1. Contoh hasil standarisasi nilai ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil Data Normalisasi.

K_{kar} $ke-i$	Budaya Organisasi (O)			Kepemimpinan (G)			Motivasi (V)			Pelatihan (T)			
	o_1	o_2	$o_{...}$	g_1	g_2	$g_{...}$	v_1	v_2	$v_{...}$	t_1	t_2	$t_{...}$	t_7
1	0.000	0.9062	...	0.3125	0.75	...	0.50	0.7812	...	0.2812	0.7812	...	0.3125
2	0.00352	0.1875	...	0.2188	0.6875	...	0.6250	0.9062	...	0.9375	0.5312	...	0.0938
3	0.00704	0.2812	...	0.9062	1.00	...	0.5938	0.75	...	0.50	0.9062	...	1.00
4	0.01056	0.2812	...	0.8750	0.1250	...	0.6875	0.5312	...	0.8438	0.8125	...	0.7812
5	0.01408	0.7188	...	0.7812	0.8750	...	0.3125	0.6562	...	0.3750	0.1875	...	0.4688
6	0.01761	0.8438	...	0.50	0.50	...	0.4062	0.5625	...	0.8125	0.5938	...	0.5938
7	0.02113	0.6250	...	0.1562	0.3125	...	0.6250	0.4688	...	0.6250	0.3750	...	0.1250
8	0.02465	0.5312	...	0.7812	0.8125	...	0.6250	0.4375	...	0.8750	0.2188	...	1.00
...

Setelah dilakukan proses normalisasi, dilakukan tahap analisis data seleksi fitur dengan ANOVA. Pada penelitian ini, analisis jumlah fitur dan perangkanan fitur dilakukan secara bertahap dengan menggunakan melakukan seleksi satu persatu dari 28 fitur yang ada. Di tahap awal, perangkanan 28 fitur ditunjukkan pada Tabel 3.

Tabel 3. Hasil Analisis Rank Fitur.

Nama Fitur	Notasi	Score Rank	Rank	Nama Fitur	Notasi	Score Rank	Rank
Perilaku	o_2	8.37551	1	Desentralisasi	g_3	2.37154	15
Sikap	t_6	5.97719	2	Fleksibilitas	t_7	2.25455	16
Delegasi	g_4	5.27504	3	Keinginan	v_1	1.88691	17

Konsultasi	g_1	5.22171	4	Kepercayaan	o_5	1.8501	18
Manajemen Demokratik	g_2	4.4113	5	Kecerdasan	g_6	1.82454	19
Latihan	t_3	4.32347	6	Team-Work	v_7	1.80168	20
Kemampuan	t_1	4.09063	7	Pengetahuan	t_4	1.64786	21
Fasilitas	v_6	3.89262	8	Sikap	v_5	1.17719	22
Kebersamaan	o_6	3.61897	9	Bertanggung-Jawab	g_5	1.145	23
Keyakinan	o_1	3.38875	10	Pembagian Kerja	o_7	0.925532	24
Usaha	v_4	2.88256	11	Norma-Norma	o_4	0.829875	25
Kebutuhan	v_3	2.81706	12	Keterampilan	t_5	0.808741	26
Evaluasi Rutin	g_7	2.59423	13	Sistem Kontrol	o_3	0.190138	27
Pendidikan	t_2	2.4202	14	Tujuan	v_2	0.0942743	28

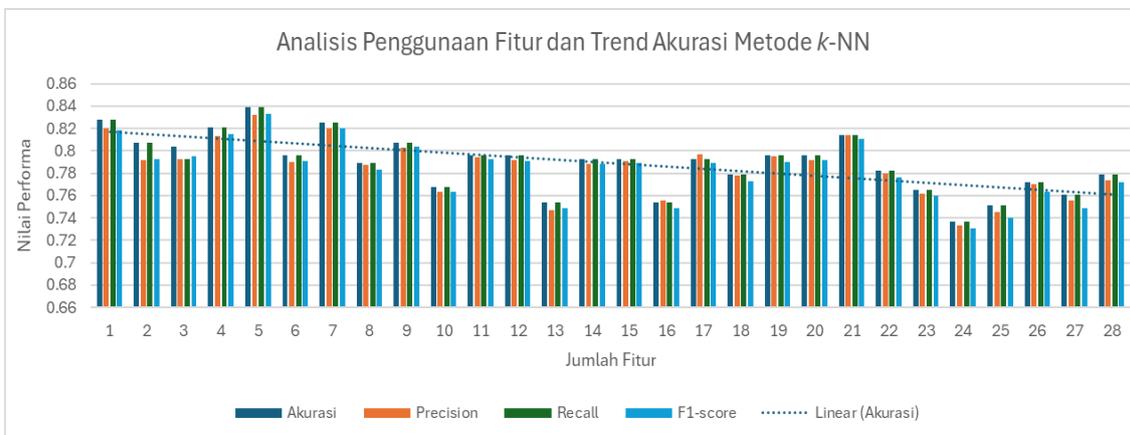
Dari hasil perankingan didapatkan bahwa fitur terbaik ranking pertama adalah perilaku yang merupakan bagian dari variabel budaya organisasi. Kemudian fitur terbaik pada variabel pelatihan adalah fitur sikap dalam ranking fitur kedua. Fitur terbaik delegasi merupakan bagian dari variabel kepemimpinan memiliki ranking ketiga. Sedangkan pada variabel motivasi fitur terbaik adalah fasilitas yang berada pada ranking fitur ke delapan. Selanjutnya dilakukan analisis performa model klasifikasi dengan menyeleksi satu-persatu fitur yang ada dari 28 fitur menjadi 1 fitur. Hasil klasifikasi pada analisis penggunaan jumlah fitur ditunjukkan pada Tabel 4.

Tabel 4. Hasil Analisis Penggunaan jumlah Fitur Pada model Klasifikasi k -NN.

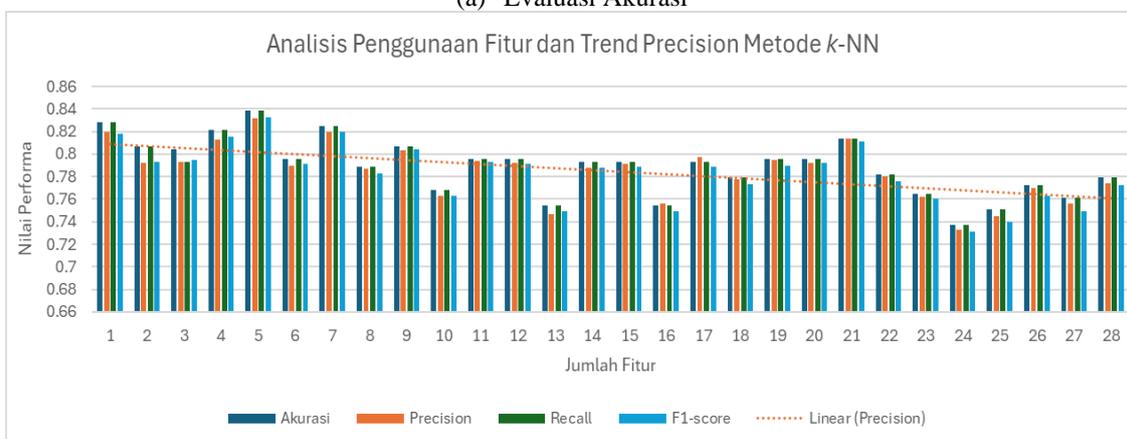
Jumlah Fitur	Akurasi	Precision	Recall	F1-score	Jumlah Fitur	Akurasi	Precision	Recall	F1-score
1	0.828	0.82	0.828	0.818	15	0.793	0.791	0.793	0.789
2	0.807	0.792	0.807	0.793	16	0.754	0.756	0.754	0.749
3	0.804	0.793	0.793	0.795	17	0.793	0.797	0.793	0.789
4	0.821	0.813	0.821	0.815	18	0.779	0.778	0.779	0.773
5	0.839	0.832	0.839	0.833	19	0.796	0.795	0.796	0.79
6	0.796	0.79	0.796	0.791	20	0.796	0.792	0.796	0.792
7	0.825	0.82	0.825	0.82	21	0.814	0.814	0.814	0.811
8	0.789	0.787	0.789	0.783	22	0.782	0.78	0.782	0.776
9	0.807	0.803	0.807	0.804	23	0.765	0.762	0.765	0.76
10	0.768	0.763	0.768	0.763	24	0.737	0.733	0.737	0.731
11	0.796	0.794	0.796	0.793	25	0.751	0.745	0.751	0.74
12	0.796	0.792	0.796	0.791	26	0.772	0.77	0.772	0.763
13	0.754	0.747	0.754	0.749	27	0.761	0.756	0.761	0.749
14	0.793	0.788	0.793	0.788	28	0.779	0.774	0.779	0.772

Dari hasil analisis didapatkan bahwa hasil akurasi tertinggi adalah pada jumlah fitur 5 fitur dengan akurasi klasifikasi adalah sebesar 0.839, precision 0.832, recall sebesar 839 dan F1-score sebesar 0.833. nilai akurasi terendah adalah 0.737, precision 0.733, recall 0.737 dan F1-score sebesar 0.731 dengan menggunakan fitur sebanyak 24 fitur. Secara keseluruhan model memiliki rata rata akurasi deteksi sebesar 0.789, precision 0.785, recall sebesar 0.789 dan F1-score sebesar 0.783. Dari analisis trend, ditemukan bahwa semakin banyak fitur yang digunakan menunjukkan performa model dari evaluasi akurasi, precision, recall dan F1-score semakin menurun. Puncak performa tertinggi terjadi pada penggunaan jumlah fitur sebanyak lima, dan nilai performa terendah terjadi di antara penggunaan fitur sebanyak 10 hingga 16. Penggunaan fitur di atas 16 mampu meningkatkan performa hingga jumlah fitur yang digunakan sebanyak 21 fitur. Walaupun meningkat pada penggunaan fitur sebanyak 21, ternyata kembali performa model klasifikasi menurun dan mencapai puncak rendah dengan penggunaan fitur sebanyak 24 fitur. Hal ini menunjukkan bahwa metode optimasi seleksi fitur berdasarkan jumlah penggunaan dapat mempengaruhi

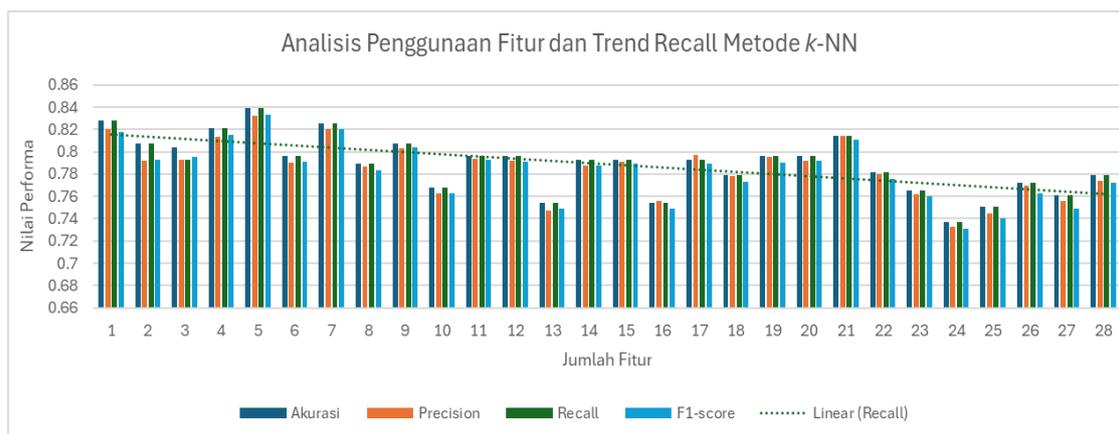
performa model klasifikasi. Dapat disimpulkan juga relatif penggunaan jumlah fitur yang rendah dapat menghasilkan performa optimal. Trend dari performa dari akurasi, precision, recall dan F1-score ditunjukkan pada Gambar 2.



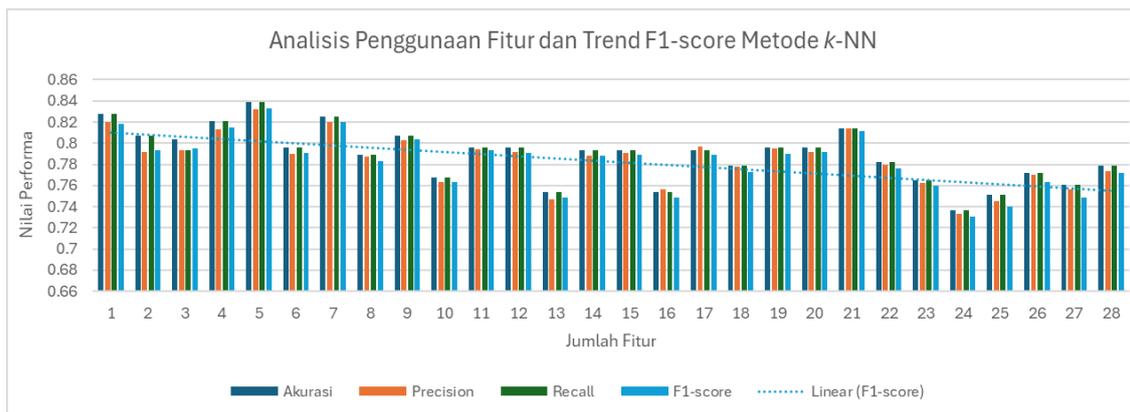
(a) Evaluasi Akurasi



(b) Evaluasi Precision



(c) Evaluasi Recall



(d) Evaluasi F1-score

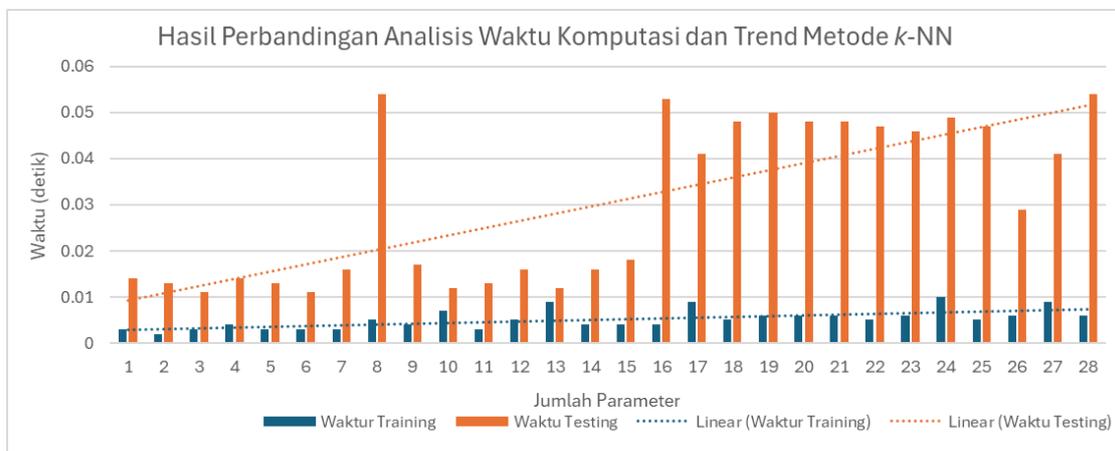
Gambar 2. Trend Performa Model Deteksi.

Pada penelitian ini juga melakukan analisis waktu komputasi dari sisi pelatihan model dan pengujian model. Di penelitian ini, pengujian proses pelatihan dan pengujian menggunakan komposisi 70% sebagai data latih dan 30% sebagai data uji dengan teknik random sampling. Berdasarkan analisis waktu komputasi, didapatkan bahwa model klasifikasi memiliki waktu komputasi tercepat adalah selama 0.002 detik dengan menggunakan 2 fitur pada model klasifikasi. Sedangkan waktu tercepat dalam pengujian adalah selama 0.011 detik pada penggunaan 3 dan 6 fitur. Rata rata pengujian penggunaan satu hingga 28 fitur adalah selama 0.0052 pada proses pelatihan dan 0.0303 detik pada proses pengujian total sebanyak 285 data. Dalam penelitian ini hasil analisis waktu komputasi ditunjukkan pada Tabel 5.

Tabel 5. Analisis Perbandingan Waktu Komputasi.

Jumlah Fitur	Waktur Training	Waktu Testing	Jumlah Fitur	Waktur Training	Waktu Testing
1	0.003	0.014	15	0.004	0.018
2	0.002	0.013	16	0.004	0.053
3	0.003	0.011	17	0.009	0.041
4	0.004	0.014	18	0.005	0.048
5	0.003	0.013	19	0.006	0.05
6	0.003	0.011	20	0.006	0.048
7	0.003	0.016	21	0.006	0.048
8	0.005	0.054	22	0.005	0.047
9	0.004	0.017	23	0.006	0.046
10	0.007	0.012	24	0.01	0.049
11	0.003	0.013	25	0.005	0.047
12	0.005	0.016	26	0.006	0.029
13	0.009	0.012	27	0.009	0.041
14	0.004	0.016	28	0.006	0.054

Waktu terlama dalam proses pelatihan terjadi saat menggunakan 24 fitur dengan waktu 0.010 detik dan waktu pengujian terlama adalah pada jumlah 28 fitur selama 0.054 detik. Tren analisis waktu komputasi pelatihan dan pengujian ditunjukkan pada Gambar 3.



Gambar 3. Analisis Waktu Komputasi dan Trend Metode Klasifikasi *k*-NN.

Hasil analisis trend menunjukkan bahwa semakin banyak fitur yang digunakan, maka menunjukkan waktu komputasi dari sisi pelatihan dan pengujian menghasilkan waktu yang lebih lama. Namun berdasarkan karakteristik data dan teknik random sampling dalam pengujian *k*-NN, terdapat waktu komputasi yang lebih cepat pada saat penggunaan fitur sebanyak 26 fitur. Kemudian waktu komputasi kembali bertambah saat jumlah fitur yang digunakan semakin banyak pada penggunaan 27 dan 28 fitur. Hal ini menyimpulkan bahwa performa waktu komputasi berdasarkan jumlah fitur yang digunakan oleh model klasifikasi dapat terpengaruhi secara signifikan, antara proses pelatihan dan pengujian.

4. Kesimpulan

Penelitian ini mengusulkan pendekatan analisis teknik seleksi fitur dengan metode ANOVA pada model klasifikasi *k*-NN dengan pengujian pada studi kasus data penilaian kinerja karyawan. Analisis bertujuan untuk mengetahui performa dari model klasifikasi dan mendapatkan jumlah fitur terbaik yang paling optimal. Proses analisis diawali dengan mengumpulkan data dari 4 variabel penilaian kinerja karyawan, yaitu budaya organisasi, kepemimpinan, motivasi dan pelatihan. Di mana setiap variabel memiliki 7 parameter penilaian. Kemudian model melakukan proses agregasi data dengan membentuk vektor data dan dilakukan proses normalisasi data yang bertujuan untuk menstandarisasi skala rentang nilai data. Proses selanjutnya analisis fitur rank dan seleksi fitur dengan metode ANOVA dan didapatkan bahwa fitur terbaik ranking pertama adalah perilaku yang merupakan bagian dari variabel budaya organisasi. Kemudian fitur terbaik pada variabel pelatihan adalah fitur sikap dalam rangking fitur kedua. Fitur terbaik delegasi merupakan bagian dari variabel kepemimpinan memiliki ranking ketiga. Sedangkan pada variabel motivasi fitur terbaik adalah fasilitas yang berada pada ranking fitur ke delapan. Pada analisis penggunaan jumlah fitur pada model klasifikasi, didapatkan akurasi tertinggi adalah pada jumlah fitur 5 fitur dengan akurasi klasifikasi adalah sebesar 0.839, precision 0.832, recall sebesar 839 dan F1-score sebesar 0.833. nilai akurasi terendah adalah 0.737, precision 0.733, recall 0.737 dan F1-score sebesar 0.731 dengan menggunakan fitur sebanyak 24 fitur. Secara keseluruhan model memiliki rata rata akurasi deteksi sebesar 0.789, precision 0.785, recall sebesar 0.789 dan F1-score sebesar 0.783. Analisis waktu komputasi menunjukkan model klasifikasi memiliki waktu komputasi tercepat adalah selama 0.002 detik dengan menggunakan 2 fitur pada model klasifikasi. Sedangkan waktu tercepat dalam pengujian adalah selama 0.011 detik pada penggunaan 3 dan 6 fitur. Hasil analisis ini dapat dimanfaatkan bagi sebuah perusahaan atau lembaga atau institusi untuk mengukur kinerja karyawan berdasarkan fitur parameter terbaik. Di Penelitian selanjutnya, dilakukan analisis perbandingan metode seleksi fitur yang diujikan pada beberapa

metode klasifikasi. Harapannya adalah mendapatkan performa yang lebih tinggi dari sisi analisis akurasi, precision dan recall.

Daftar Pustaka

- [1] V. V, “Comparison of Some Classification Algorithms for the Analysis of Students Academic Performance in Educational Data Mining Using Orange,” *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 6, no. 1, pp. 318–324, 2021, doi: 10.48175/ijarsct-1394.
- [2] I. Khan, A. R. Ahmad, N. Jabeur, and M. N. Mahdi, “An artificial intelligence approach to monitor student performance and devise preventive measures,” *Smart Learn. Environ.*, vol. 8, no. 1, 2021, doi: 10.1186/s40561-021-00161-y.
- [3] M. Milkhatun, A. A. F. Rizal, N. W. W. Asthiningsih, and A. J. Latipah, “Performance Assessment of University Lecturers: A Data Mining Approach,” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 6, no. 2, pp. 73–81, 2020, doi: 10.23917/khif.v6i2.9069.
- [4] S. Kim and E. Usui, “Employer learning, job changes, and wage dynamics,” *Econ. Inq.*, vol. 59, no. 3, pp. 1286–1307, 2021, doi: 10.1111/ecin.12980.
- [5] D. S. Gunawan, Suwardi; Sirajang, Ratna Sari; Utomo, “Pengaruh Motivasi, Kepuasan Kerja, dan Komitmen Organisasional Terhadap Kinerja Karyawan (Studi Kasus PT. PLN Sektor Mahakam Samarinda),” *J. Din. Tek.*, vol. 13, no. 1, pp. 14–23, 2020.
- [6] M. Indrasari, “The Effect Of Organizational Culture, Environmental Work, Leadership Style On The Job Satisfaction And Its Impact On The Performance Of Teaching In State Community Academy Bojonegoro,” *Sinergi J. Ilm. Ilmu Manaj.*, vol. 7, no. 1, pp. 58–73, 2017, doi: 10.25139/sng.v7i1.30.
- [7] S. Langgeng Ratnasari, . R., G. Sutjahjo, and D. Yana, “Lecturer’s Performance: Leadership, Organizational Culture, Work Motivation, and Work Behavior,” *KnE Soc. Sci.*, vol. 3, no. 10, pp. 1–7, 2018, doi: 10.18502/kss.v3i10.3416.
- [8] S. Langgeng Ratnasari, . R., G. Sutjahjo, and D. Yana, “Lecturer’s Performance: Leadership, Organizational Culture, Work Motivation, and Work Behavior,” *KnE Soc. Sci.*, vol. 3, no. 10, 2018, doi: 10.18502/kss.v3i10.3416.
- [9] E. KARAAHMETOĞLU, S. ERSÖZ, A. K. TÜRKER, V. ATEŞ, and A. F. İNAL, “Evaluation of Profession Predictions for Today and the Future with Machine Learning Methods : Emperical Evidence From Turkey,” *Politek. Derg.*, vol. 26, no. 1, pp. 107–124, 2021, doi: 10.2339/politeknik.985534.
- [10] A. M. Fevolden, “Skills for the future – forecasting firm competitiveness using machine learning methods and employer–employee register data,” no. April, 2021.
- [11] M. Bakare and Y. C. Ojeleye, “Participative Leadership Style and Employee Commitment in Federal College of Education (Technical) Gusau: Moderating role of Organizational Culture,” *Int. J. Intellect. Discourse*, vol. 3, no. 1, pp. 17–31, 2020.
- [12] M. H. Bhatti, Y. Ju, U. Akram, M. H. Bhatti, Z. Akram, and M. Bilal, “Impact of participative leadership on organizational citizenship behavior: Mediating role of trust and moderating role of continuance commitment: Evidence from the pakistan hotel industry,” *Sustain.*, vol. 11, no. 4, 2019, doi: 10.3390/su11041170.
- [13] M. Saĝnak, “Participative Leadership and Change-Oriented Organizational Citizenship: The Mediating Effect of Intrinsic Motivation,” *Eurasian J. Educ. Res.*, vol. 16, no. 62, pp. 181–194, 2016, doi: 10.14689/ejer.2016.62.11.
- [14] F. Sinani, “The Effects of Participative Leadership Practices on Job Satisfaction for Highly Skilled Virtual Teams,” *ProQuest Diss. Theses*, p. 173, 2016.
- [15] N. Bandhaso, Mira Labi dan Paranoan, “Pengaruh Kepuasan Kerja dan Motivasi Kerja Terhadap Kinerja Dosen Fakultas Ekonomi di Perguruan Tinggi Swasta di Makassar,” *J. Akunt. Netral*, vol. 1, no. 2, pp. 20–30, 2019.
- [16] Fizia Nurul et al, “Pengaruh Pelatihan dan Pengembangan Karir Terhadap Kinerja Karyawan pada PT. Tri Megah Makmur,” *Din. UMT*, vol. 3, no. 1, pp. 2477–1546, 2018, doi: 10.31000/dinamika.v3i1.1091.
- [17] K. S. Akpoviroro, B. Kadiri, and S. O. Owotutu, “Effect of participative leadership style on employees productivity,” *Trendy v Podn.*, vol. 8, no. 2, pp. 48–58, 2018, doi: 10.24132/jbt.2018.8.2.48_58.
- [18] M. Indrasari, *Kepuasan Kerja dan Kinerja Karyawan*. Yogyakarta: Indomedia Pustaka, 2017.
- [19] J. Straub, “Machine learning performance validation and training using a ‘perfect’ expert system,”

-
- MethodsX*, vol. 8, no. August, p. 101477, 2021, doi: 10.1016/j.mex.2021.101477.
- [20] J. G. Choi, I. Ko, J. Kim, Y. Jeon, and S. Han, "Machine Learning Framework for Multi-Level Classification of Company Revenue," *IEEE Access*, vol. 9, no. June, pp. 96739–96750, 2021, doi: 10.1109/ACCESS.2021.3088874.