

KLASIFIKASI PENERIMAAN TENAGA KERJA TERTUTUP MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (KNN)

Fridy Mandita¹⁾, Rizal Kurnia Pratama²⁾

^{1,2)}Teknik Informatika, Universitas 17 Agustus 1945 Surabaya

Jl. Semolowaru No.45, Menur Pumpungan, Kec. Sukolilo, Surabaya, Jawa Timur 60118

e-mail: fridymandita@untag-sby.ac.id¹⁾, rkurniaprutama@gmail.com²⁾

(Naskah masuk : xx xx xxxx Diterima untuk diterbitkan : xx xx xxxx)

ABSTRAK

Perusahaan yang telah menerapkan pemanfaatan teknologi tidak akan lepas dari pengolahan data atau informasi, akan tetapi tidak semua perusahaan dapat memaksimalkan pengolahan data yang mereka miliki, salah satunya di CV. Arfa Nusantara Teknologi, data calon karyawan yang masuk ke dalam sebuah perusahaan tidak dilakukan pengolahan lebih lanjut. Penelitian ini mencoba untuk melakukan pengolahan terhadap data calon karyawan yang masuk ke sebuah perusahaan dengan tujuan untuk mengklasifikasikan data calon tenaga kerja berdasarkan variabel hasil skor tes teknis dari tim penilai, total nilai bobot keahlian, minimum gaji yang diminta dan jumlah tahun pengalaman bekerja dengan mengimplementasikan algoritma K-Nearest Neighbor (KNN), dengan implementasi algoritma ini nantinya akan dihasilkan beberapa klasifikasi data calon tenaga kerja yaitu paling baik, cukup baik dan kurang baik yang dimana dapat membantu tim penilai dalam perusahaan tersebut untuk membuat keputusan dalam menerima calon tenaga kerja yang terbaik bagi perusahaan dan mengetahui tingkat akurasi dari algoritma tersebut. Selain itu penelitian ini mencoba beberapa jumlah pengambilan ketetanggaan KNN dan menerapkan K-Fold Cross Validation dalam menentukan model optimal klasifikasi. Pada model optimal yang berhasil didapatkan berada di rata-rata nilai akurasi, presisi dan recall adalah 100% pada pembagian $k = 4$ dan $k = 5$ di uji coba pengambilan 10 K-Fold Cross Validation. Sedangkan nilai ketetanggaan algoritma KNN paling baik adalah $k = 7$. Pada pengujian System Usability Scale didapatkan skor sebesar 80,21 dan masuk kedalam kategori nilai B.

Kata kunci: Klasifikasi, K-Nearest Neighbor, K-Fold Cross Validation, Penerimaan Tenaga Kerja

ABSTRACT

Companies that have implemented the use of technology will not be separated from processing data or information, but not all companies can maximize the processing of the data they have, one of which is CV. Arfa Nusantara Teknologi, data on prospective employees entering a company is not subjected to further processing. This research tries to process data on prospective employees entering a company with the aim of classifying prospective workforce data based on the variables of technical test scores from the assessment team, total skill weight scores, minimum salary requested and number of years of work experience by implementing an algorithm. K-Nearest Neighbor (KNN), with the implementation of this algorithm, several data classifications of prospective workers will be produced, namely the best, quite good and not so good, which can help the assessment team in the company to make decisions in accepting the best job candidates for the company. company and know the level of accuracy of the algorithm. Apart from that, this research tries several numbers of KNN neighbor retrievals and applies K-Fold Cross Validation to determine the optimal classification model. In the optimal model that was successfully obtained, the average value of accuracy, precision and recall was 100% in the division $k = 4$ and $k = 5$ in the 10 K-Fold Cross Validation retrieval trials. Meanwhile, the best neighbor value for the KNN algorithm is $k = 7$. In the System Usability Scale test, a score of 80.21 was obtained and it was included in the B value category.

Keywords: Classification, K-Nearest Neighbor, K-Fold Cross Validation, Labor Recruitment

I. PENDAHULUAN

Departemen Sumber Daya Manusia (SDM) sangat penting dalam perusahaan atau organisasi mana pun, karena departemen ini terutama bertanggung jawab untuk mengawasi operasi penting termasuk manajemen proses rekrutmen, penyediaan pendidikan tenaga kerja, analisis kebijakan tenaga kerja, dan kesejahteraan karyawan. Seringkali, departemen SDM bekerja sama dengan departemen terkait yang membutuhkan personel tambahan untuk membentuk tim penilai sendiri ketika merekrut personel baru. Banyak faktor yang dipertimbangkan oleh kelompok ini dalam proses evaluasi calon kandidat. Proses seleksi, yang merupakan komponen penting dalam manajemen personalia, terjadi setelah analisis pekerjaan, yang menghasilkan deskripsi pekerjaan dan prasyarat bagi calon karyawan untuk memenuhi posisi tertentu [1].

Mengidentifikasi dan menarik calon kandidat, mengevaluasi lamaran mereka, dan pada akhirnya mempekerjakan individu yang paling memenuhi syarat merupakan prosedur seleksi. Dalam fase penyaringan lamaran, departemen SDM melakukan pemeriksaan menyeluruh terhadap setiap lamaran untuk menentukan kandidat mana yang memenuhi persyaratan organisasi. Prosedur penyaringan mengalami kesulitan karena banyaknya aplikasi yang masuk. Setelah kandidat berhasil menyelesaikan tahap penyaringan CV, departemen SDM dihadapkan pada tugas menilai kesesuaian setiap individu untuk organisasi, yang dapat mengakibatkan gangguan dan inefisiensi proses perekrutan [2].

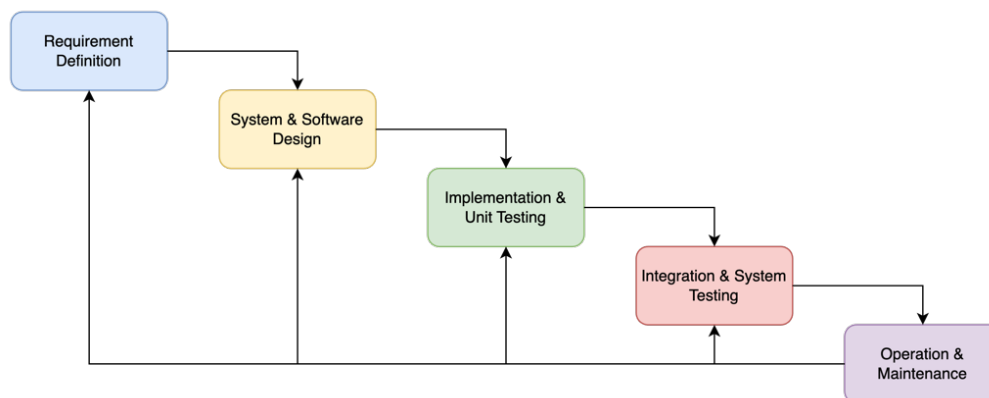
Banyak waktu yang masih terbuang oleh sejumlah tim penilai karena mereka terus mengandalkan proses manual untuk membandingkan data kandidat. Waktu sangat penting dalam mencapai keakuratan evaluasi oleh karena itu, algoritma yang mudah diinterpretasikan dan diimplementasikan sangatlah penting, khususnya dalam klasifikasi. Berbagai teknik klasifikasi, seperti *K-Nearest Neighbor (KNN)*, *Decision Tree Classifier*, *Naive Bayes Classifier*, *Support Vector Machine (SVM)* dan *AdaBoost* telah diselidiki dalam penelitian sebelumnya untuk mengevaluasi kinerjanya dalam mengklasifikasikan kumpulan data yang berkaitan dengan secara akurat pada penerimaan siswa baru [3]. Tujuan evaluasi adalah untuk menilai kualitas algoritma, khususnya kemampuannya dalam memastikan preferensi calon siswa terhadap program studinya. Algoritma *KNN Classifier* menunjukkan tingkat akurasi tertinggi mencapai 72,2%. Sebaliknya, *Decision Tree Classifier* mencapai tingkat akurasi terendah yaitu 33,4%. Algoritma tambahan yang mencapai tingkat akurasi dalam rentang menengah adalah *Naive Bayes Classifier* (60,7%), *Support Vector Machine (SVM)* (61,7%), dan *Adaptive Boosting (AdaBoost)* (62,4%). Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa algoritma *KNN* menunjukkan akurasi yang unggul dalam memodelkan preferensi program studi berdasarkan dataset penerimaan mahasiswa baru, mencapai tingkat akurasi sebesar 72,2%. Menyadari pentingnya pemilihan personel yang tepat dan tepat bagi sebuah organisasi, algoritma *KNN*, yang telah terbukti sangat akurat dalam penelitian sebelumnya, muncul sebagai pendekatan yang berharga untuk memfasilitasi klasifikasi tenaga kerja.

Dari hasil penelitian ini akan dikembangkan sebuah sistem berbasis web dengan menerapkan proses klasifikasi terhadap data calon tenaga kerja menggunakan algoritma *KNN* berdasarkan variabel hasil skor tes teknis dari tim penilai, total nilai bobot keahlian, minimum gaji yang diminta dan jumlah tahun pengalaman bekerja. Diharapkan dengan adanya penelitian ini sistem yang akan akan dikembangkan dapat memberikan daftar calon tenaga kerja kedalam beberapa kategori dan mengetahui tingkat akurasi dari algoritma *KNN*.

II. METODE PENELITIAN

2.1. Metode Pengembangan Waterfall

Terdapat berbagai metode dalam proses pengembangan perangkat lunak salah satunya adalah *waterfall* yang banyak digunakan dalam proses pengembangan. Metode *waterfall* memiliki pendekatan desain air terjun menggunakan prosedur yang secara bertahap bergerak dari atas ke bawah, menyerupai penampilan air terjun.



Gambar 2. 1 Metode Pengembangan *Waterfall*

Tahapan metode pengembangan *waterfall* memiliki 5 tahapan sebagai berikut:

1. *Requirement Definition*
 Pada tahap ini dilakukan proses pengumpulan informasi terkait kebutuhan fungsional maupun non-fungsional sistem dengan melibatkan pengguna dan pemangku kepentingan. Dalam tahap ini juga dilakukan proses pengamatan dilapangan untuk mengidentifikasi masalah dan kebutuhan apa saja yang nantinya akan diakomodir oleh sistem yang akan dibangun.
2. *System & Software Design*
 Pada tahap ini dilakukan proses pembuatan desain sistem yang lebih spesifik yang nanti akan ditujukan untuk memudahkan tim pengembang dalam mengembangkan sistem. Desain sistem mencakup *User Interface, ERD, Use Case, Activity Diagram* dan lain-lain.
3. *Implementation & Unit Testing*
 Pada tahap ini tim pengembang akan mulai melakukan proses pengkodean sistem berdasarkan desain sistem yang telah dibuat pada proses sebelumnya dan melakukan pengujian secara kecil untuk memastikan kode yang dibuat berjalan sesuai dengan desain yang dibuat.
4. *Integration & System Testing*
 Pada tahap ini dilakukan pengujian secara menyeluruh terhadap sistem yang telah dibangun untuk memastikan apakah sistem yang dibangun telah memenuhi kebutuhan yang telah didefinisikan di tahap sebelumnya, sehingga apabila terdapat kebutuhan yang belum terpenuhi maka akan dilakukan pengkajian ulang terhadap sistem.
5. *Operation & Maintenance*
 Pada tahap ini sistem sudah bisa digunakan oleh pengguna, akan tetapi sistem akan butuh *maintenance* secara rutin untuk memastikan agar sistem yang telah dibangun berjalan dengan optimal dan bahkan memungkinkan untuk terjadinya penambahan fitur baru seiring dengan peningkatan kebutuhan.

2.2. K-Nearest Neighbor

Data yang telah dihitung jarak kedekatannya dan diurutkan berdasarkan yang terdekat ke yang terjauh nantinya hanya akan diambil data sejumlah nilai dari k . Dimana nilai k ialah jumlah dari tetangga terdekat yang akan diambil sebagai penentu klasifikasi dengan melihat kelas dominan yang muncul pada data yang diambil sejumlah nilai k tersebut. Salah satu cara menghitung jarak adalah menggunakan *Euclidean Distance*.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.1)$$

Keterangan:

- $d(x, y)$ = Jarak
- x_i = Data testing
- y_i = Data training
- i = Variabel data

n = Dimensi data

2.3. Confusion Matrix

Confusion Matrix adalah statistik yang digunakan untuk menilai efektivitas dan kualitas model yang dibangun [4]. Terutama diterapkan dalam domain penambangan data, ini menilai akurasi dengan menyediakan tabel yang mengilustrasikan data pengujian yang benar dan salah diklasifikasikan. Nilai-nilai dalam matriks konfusi kemudian digunakan untuk menghitung akurasi, presisi, dan *recall*.

Tabel 2. 1 *Confusion Matrix*

	Aktual	
Prediksi	Positive	Negative
Positive	TP (<i>True Positive</i>)	FP (<i>False Positive</i>)
Negative	FN (<i>False Negative</i>)	TN (<i>True Negative</i>)

Keterangan:

- TP = Jumlah di mana data sebenarnya positif dan prediksi juga menunjukkan positif.
- FP = Jumlah di mana data sebenarnya negatif tapi prediksi menunjukkan positif.
- FN = Jumlah di mana data sebenarnya positif tapi prediksi menunjukkan negatif.
- TN = Jumlah di mana data sebenarnya negatif dan prediksi juga menunjukkan negatif

1. Akurasi

Akurasi berfungsi sebagai metrik penilaian dengan membandingkan jumlah kelas yang diprediksi positif dan negatif pada kenyataannya dibandingkan dengan seluruh kumpulan data yang digunakan. Perhitungan nilai akurasi dapat dilakukan melalui rumus berikut:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.2)$$

2. Presisi

Presisi berfungsi sebagai metrik penilaian dengan membandingkan jumlah kelas yang diprediksi positif pada kenyataannya dengan jumlah total kelas yang diprediksi positif. Perhitungan nilai presisi dapat dilakukan melalui rumus berikut:

$$precision = \frac{TP}{TP + FP} \quad (2.3)$$

3. Recall

Recall berfungsi sebagai metrik penilaian dengan mengontraskan jumlah kelas yang diprediksi positif pada kenyataannya dengan jumlah total kelas yang diprediksi negatif. Perhitungan nilai *recall* dapat dilakukan melalui rumus berikut:

$$recall = \frac{TP}{TP + FN} \quad (2.4)$$

2.4. K-Fold Cross Validation

Dengan menggunakan metode ini untuk mempartisi data ke dalam set pelatihan dan pengujian, peneliti sering kali mengurangi bias pengambilan sampel. Dengan mempartisi kumpulan data secara berulang menjadi subset pelatihan dan pengujian, *K-Fold Cross Validation* memberikan setiap titik data peluang untuk disertakan dalam kumpulan pengujian. Metodologi ini melibatkan pembagian stokastik kumpulan data menjadi beberapa partisi. Setiap eksperimen menggunakan partisi ke-k sebagai set pengujian dan partisi yang tersisa sebagai set pelatihan di seluruh k iterasi data yang dipartisi.

2.5. System Usability Scale

Pada tahun 1986, John Brooke mengembangkan *System Usability Scale (SUS)*. *SUS* adalah metode yang digunakan untuk mengukur seberapa efektif suatu sistem dengan mengajukan pertanyaan khusus kepada sekelompok peserta [5]. *SUS* terdiri dari sepuluh pertanyaan, dan setiap pertanyaan dilengkapi dengan lima pilihan jawaban, mulai dari sangat setuju hingga sangat tidak setuju dan nilai maksimal skor berkisar 0 hingga 100. Adapun aturan untuk menentukan skor *SUS*:

1. Setiap pertanyaan yang memiliki nomor ganjil nantinya untuk skor pertanyaan yang dipilih dikurangi 1 ($X-1$).
2. Setiap pertanyaan yang memiliki nomor genap nantinya nilai 5 dikurangi skor pertanyaan yang dipilih ($5-X$).
3. Setelah semua skor pertanyaan didapatkan dan dijumlahkan secara keseluruhan kemudian hasilnya dikalikan dengan 2,5.

$$\text{Skor } SUS = ((X1 - 1) + (5 - X2) + (X3 - 1) + (5 - X4) + (X5 - 1) + (5 - X6) + (X7 - 1) + (5 - X8) + (X9 - 1) + (5 - X10)) \times 2,5 \quad (2.5)$$

Rata-rata skor *SUS* yang diperoleh dari beberapa investigasi adalah 68. Oleh karena itu, Skor *SUS* di atas 68 dianggap di atas rata-rata, sedangkan skor di bawah 68 dianggap di bawah rata-rata. Namun, penting untuk diperhatikan bahwa kategori penilaian pada *SUS* telah ditetapkan oleh penelitian sebelumnya [6].

Tabel 2. 2 Kategori Nilai *SUS*

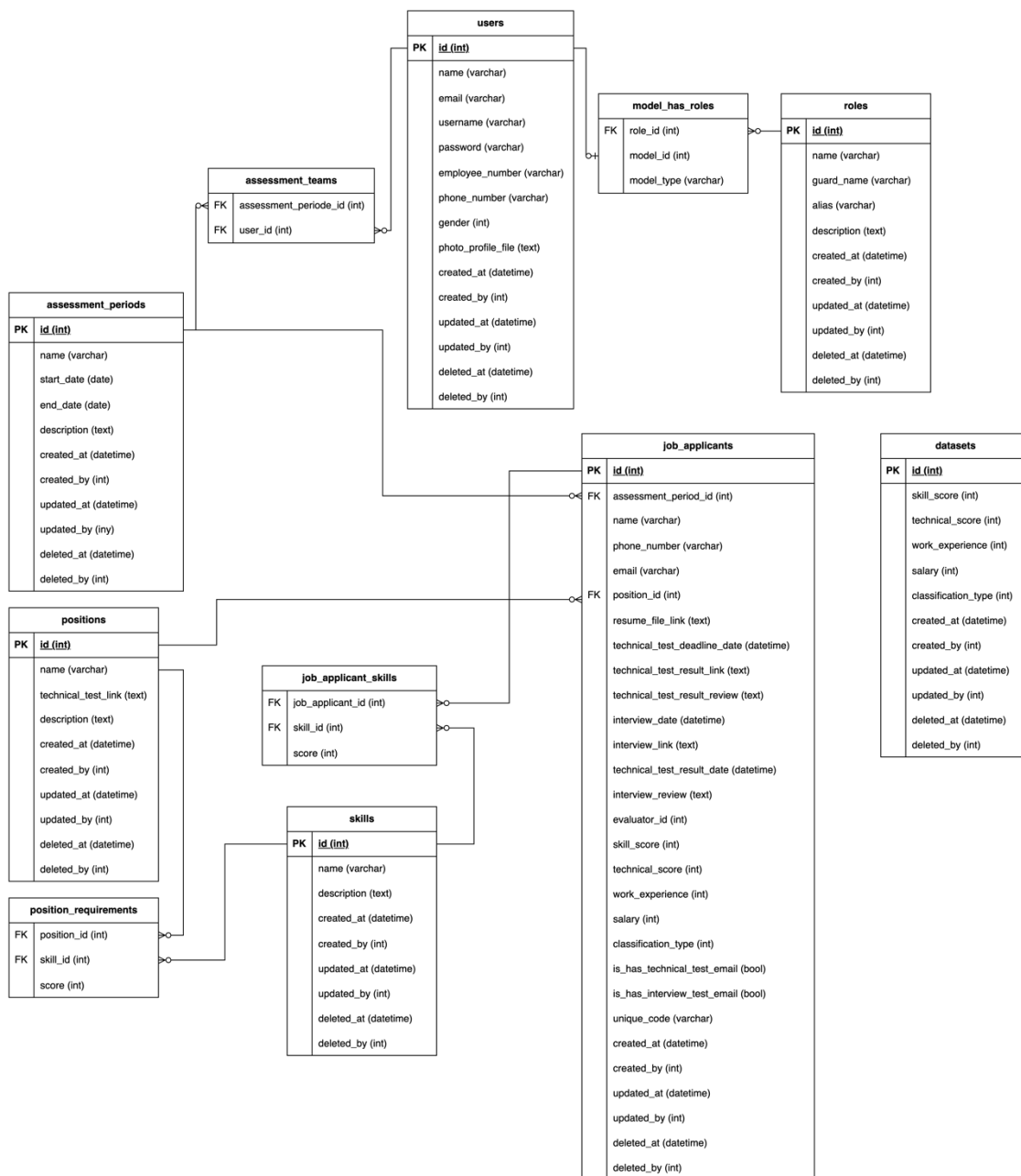
Kategori	Skor Penilaian
A	90 - 100
B	80 - 90
C	70 - 80
D	60 - 70
F	< 60

III. HASIL DAN PEMBAHASAN

Penelitian ini mengembangkan suatu sistem berbasis web yang dapat memberikan bantuan tim penilai saat melakukan proses rekrutmen dengan memprediksi calon tenaga kerja yang melamar diperusahaan ke dalam tiga kategori: sangat baik, cukup baik dan kurang baik menggunakan algoritma *KNN* dengan perhitungan jarak *euclidean distance* dengan nilai k ketetanggaan 5, 7 dan 10. Untuk pengujian algoritma *KNN* agar mendapatkan model terbaik yang nanti akan digunakan oleh sistem akan mengimplementasikan *K-Fold Cross Validation* menggunakan nilai k pembagian 5 dan 10. Pada sistem yang telah dibuat nantinya juga akan dilakukan penilaian *SUS* untuk mendapatkan penilaian dari responden secara langsung terkait dengan fungsionalitas sistem dan kemudahan dalam menggunakan sistem.

3.1. Entity Relationship Diagram

Entity Relationship Diagram (ERD) merupakan diagram visual yang dapat menjelaskan tentang hubungan tabel satu dengan lainnya yang akan menjadi tempat penampungan data dari proses yang akan terjadi pada sistem.



Gambar 3. 1 Entity Relationship Diagram

3.2. Hasil Evaluasi K-Nearest Neighbor dan K-Fold Cross Validation

Berdasarkan penelitian ini akan digunakan data calon tenaga kerja yang pernah melamar di perusahaan yang dimana akan dilakukan proses klasifikasi berdasarkan variabel hasil skor tes teknis dari tim penilai, total nilai bobot keahlian, minimum gaji yang diminta dan jumlah tahun pengalaman bekerja.

Tabel 3. 1 Dataset Calon Tenaga Kerja

Indeks	Nilai Tes	Bobot Keahlian	Pengalaman Kerja	Min. Gaji	Klasifikasi
1	75	55	0	4000000	3
2	85	75	0	4500000	2
3	80	65	0	4500000	3
4	80	75	0	6000000	3
5	70	60	0	4000000	3

Indeks	Nilai Tes	Bobot Keahlian	Pengalaman Kerja	Min. Gaji	Klasifikasi
6	85	90	0	4000000	1
7	70	65	0	3500000	3
8	80	75	0	6000000	3
9	75	75	0	3500000	2
10	80	80	0	4000000	1
11	85	85	1	3500000	1
...
106	75	50	0	4000000	3
107	80	70	0	4000000	2
108	85	75	0	4500000	2
109	70	70	0	6000000	3
110	80	75	0	6000000	3

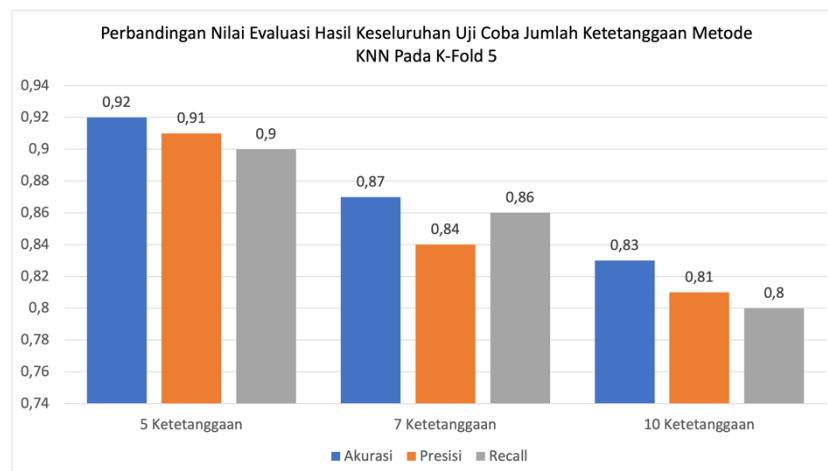
Dalam investigasi ini, sejumlah eksperimen dilakukan agar model prediktif optimal untuk klasifikasi data rekrutmen tenaga kerja dapat dikembangkan. Eksperimen yang dilakukan meliputi evaluasi algoritma *KNN* melalui pemanfaatan rumus (2.1) untuk menghitung jarak *euclidean distance*, dengan memperhatikan berbagai nilai *k* (5, 7, dan 10), dan mengeksekusi *K-Fold Cross Validation* dengan nilai *k* yang bervariasi (5 dan 10). Investigasi tersebut melibatkan pembagian kumpulan data lengkap, yang terdiri dari 110 data yang dapat dilihat pada Tabel 3. 1, menjadi dua bagian berbeda: kumpulan pelatihan, yang dipilih secara acak, dan kumpulan pengujian. Pembagian ini ditentukan oleh *K-Fold Cross Validation*, dengan *K-Fold 5* menerima pembagian 80%:20% dan *K-Fold 10* menerima pembagian 90%:10%. Hasilnya, rasio data pelatihan-pengujian untuk setiap *K-Fold* adalah sebagai berikut: 88:22 untuk *K-Fold 5* dan 99:11 untuk *K-Fold 10*. Evaluasi akurasi, presisi dan *recall* dihitung menggunakan rumus (2.2) untuk akurasi, menggunakan rumus (2.3) untuk presisi dan menggunakan rumus (2.4) untuk *recall*. Hasil percobaan yang dilakukan dirinci pada Tabel 3. 2.

Tabel 3. 2 Hasil Evaluasi Klasifikasi

Jumlah K-Fold	K (K-Fold)	K (Tetangga KNN)	Akurasi	Presisi	Recall
5	1	5	0,86	0,86	0,89
		7	0,82	0,83	0,84
		10	0,73	0,8	0,75
	2	5	0,95	0,97	0,89
		7	0,91	0,86	0,86
		10	0,77	0,73	0,74
	3	5	0,82	0,78	0,78
		7	0,73	0,67	0,67
		10	0,91	0,92	0,89
4	5	1	1	1	
	7	0,95	0,92	0,98	
	10	0,95	0,92	0,98	
10	5	5	0,95	0,93	0,93
		7	0,95	0,93	0,93
		10	0,77	0,68	0,63
	1	5	0,82	0,87	0,78
		7	0,82	0,9	0,78
		10	1	1	1
	2	5	0,91	0,93	0,93
		7	0,91	0,93	0,93
		10	0,82	0,85	0,85
3	5	0,82	0,83	0,85	
	7	0,82	0,83	0,85	
	10	0,82	0,81	0,85	
4	5	1	1	1	
	7	1	1	1	
	10	1	1	1	
5	5	1	1	1	
	7	1	1	1	
	10	1	1	1	

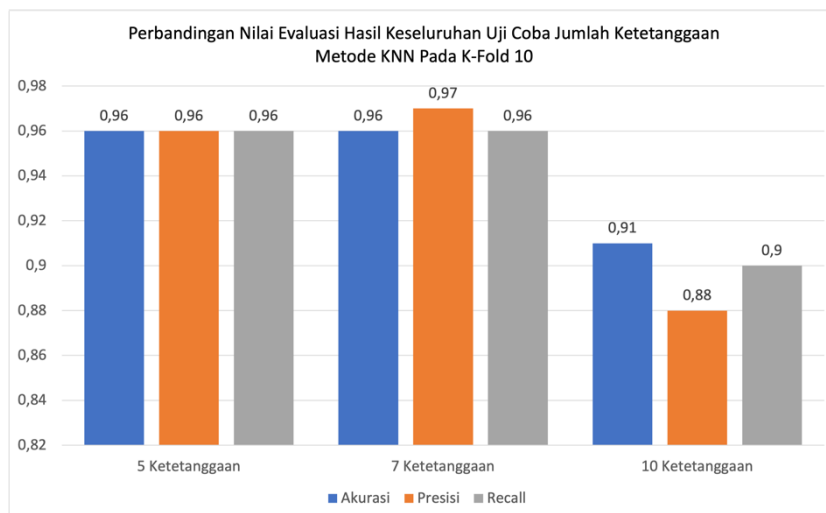
Jumlah K-Fold	K (K-Fold)	K (Tetangga KNN)	Akurasi	Presisi	Recall
		5	1	1	1
	6	7	1	1	1
		10	0,91	0,92	0,89
		5	1	1	1
	7	7	1	1	1
		10	0,91	0,89	0,92
		5	1	1	1
	8	7	1	1	1
		10	0,91	0,58	0,67
		5	1	1	1
	9	7	1	1	1
		10	0,91	0,92	0,92
		5	1	1	1
	10	7	1	1	1
		10	0,82	0,83	0,89

Berdasarkan hasil evaluasi pada **Tabel 3. 2**, terlihat bahwa terdapat beberapa uji coba yang menghasilkan nilai evaluasi yang sangat baik dan berhasil memprediksi semua data *testing* dengan tepat. Jika diamati lebih detail, penerapan 10 pembagian *K-Fold Cross Validation* lebih banyak memprediksi semua data *testing* dengan tepat dengan rentang nilai akurasi antara 82% hingga 100% jika dibandingkan pada 5 pembagian *K-Fold Cross Validation* yang memiliki rentang nilai akurasi 73% hingga 100%. Semakin tinggi nilai pembagi *K-Fold* maka semakin tinggi pula variasi data yang dapat terjadi pada data *training* dan *testing*. Sehingga semakin banyak kemungkinan model akan jauh lebih baik maupun sebaliknya. Hal tersebut juga sangat dipengaruhi oleh proses *random* data pada *K-Fold Cross Validation* yang memungkinkan data akan terbagi menjadi sangat baik, cukup baik, maupun kurang baik. Dengan beberapa kombinasi variasi data ini, penulis akan mendapatkan urutan data yang paling tepat, yang dapat merepresentasikan data *training* dan *testing* dengan baik, sehingga model yang dihasilkan jauh lebih optimal. Selanjutnya dapat dilihat beberapa grafik sebagai analisis hasil penerapan uji coba pada penelitian ini.



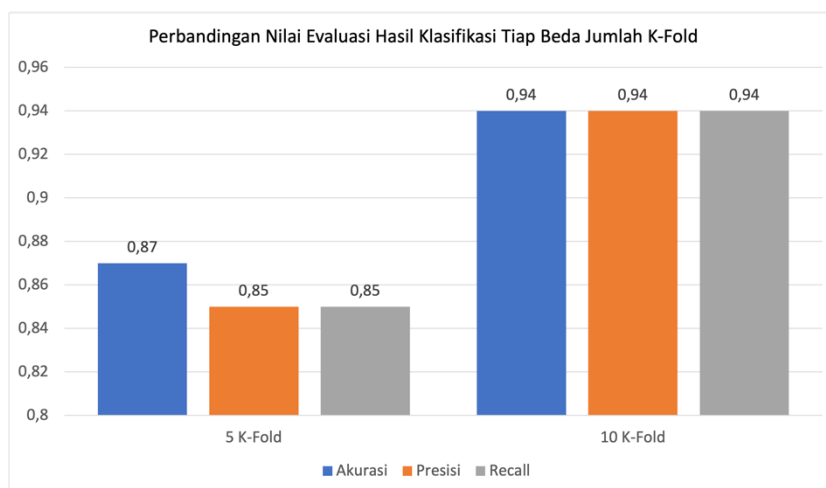
Gambar 3. 2 Grafik Hasil Evaluasi Ketetanggaan Pada K-Fold 5

Berdasarkan grafik **Gambar 3. 2** menunjukkan bahwa rata-rata nilai akurasi, presisi, dan *recall* pada pemilihan jumlah ketetanggaan algoritma *KNN* pada *K-Fold Cross Validation* 5 yang memiliki rata-rata nilai paling baik ada pada ketetanggaan 5 dengan rata-rata nilai akurasi 92%, presisi 91% dan *recall* 90%. Sedangkan pada ketetanggaan 7 memiliki rata-rata nilai akurasi 87%, presisi 84% dan *recall* 86% dimana nilai ini lebih kecil dibandingkan ketetanggaan 5. Sedangkan pada ketetanggaan 10 memiliki rata-rata nilai akurasi 83%, presisi 81% dan *recall* 80% dimana nilai ini yang paling kecil diantara semua ketetanggaan.



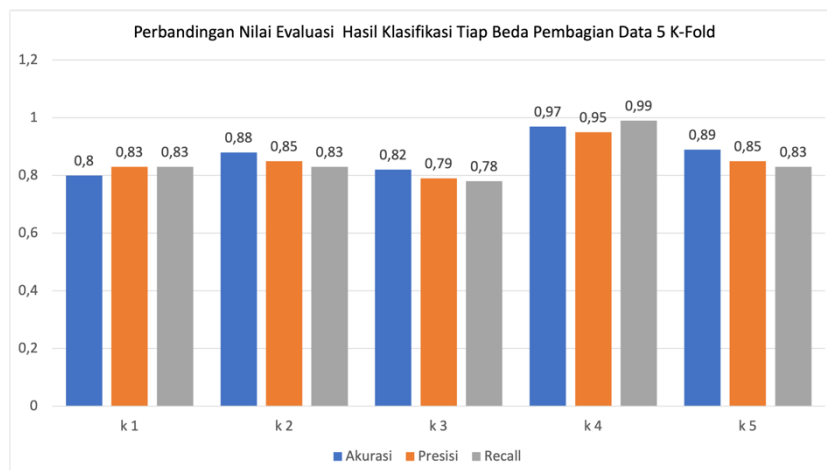
Gambar 3.3 Grafik Hasil Evaluasi Ketetangaan Pada K-Fold 10

Berdasarkan grafik **Gambar 3.3** menggambarkan bahwa rata-rata nilai dari akurasi, presisi, dan *recall* berdasarkan pemilihan jumlah ketetangaan algoritma *KNN* di *K-Fold Cross Validation* 10 yang memiliki rata-rata nilai paling baik ada pada ketetangaan 7 dengan hasil dari rata-rata nilai akurasi 96%, presisi 97% dan *recall* 96%. Sedangkan pada ketetangaan 5 memiliki rata-rata nilai akurasi, presisi dan *recall* yang sama yaitu 96% dimana nilai ini hampir sama dengan ketetangaan 7 hanya saja pada bagian presisi ketetangaan 5 lebih rendah 1%. Sedangkan pada ketetangaan 10 memiliki rata-rata nilai akurasi 91%, presisi 88% dan *recall* 90% dimana nilai ini yang paling kecil diantara semua ketetangaan. Jika dibandingkan dengan rata-rata nilai tertinggi ketetangaan 5 pada *K-Fold 5* untuk rata-rata nilai ketetangaan 7 pada *K-Fold 10* masih lebih besar. Dapat disimpulkan bahwa jumlah ketetangaan *KNN* sangat menentukan hasil prediksi akhir, semakin jauh ketetangaan yang diperhitungkan, semakin bermacam-macam hasil kelas yang diprediksi, sedangkan probabilitas suara penentuan kelas data *testing* sama antar tetangga. Sehingga semakin tinggi jumlah ketetangaan yang diperhitungkan, banyak kemungkinan hasil prediksi semakin tidak akurat.



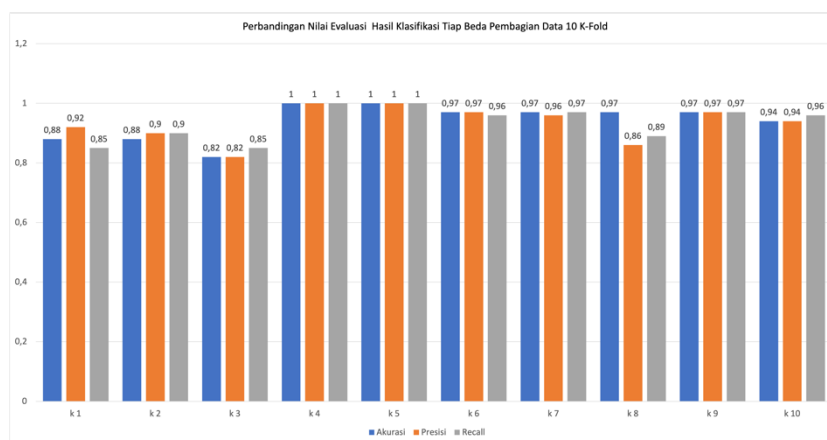
Gambar 3.4 Grafik Hasil Evaluasi K-Fold

Selanjutnya berdasarkan grafik **Gambar 3.4** sama dengan penjelasan sebelumnya yaitu hasil prediksi pada penerapan 10 jumlah *K-Fold Cross Validation* lebih tinggi dibandingkan dengan 5 *K-Fold Cross Validation*. Nilai rata-rata akurasi, presisi dan *recall* mencapai 94%. Sedangkan penerapan 5 *K-Fold Cross Validation* memiliki nilai rata-rata akurasi 87%, presisi 85% dan *recall* 85%. Nilai-nilai tersebut membuktikan pernyataan diatas bahwa semakin tinggi nilai pembagi *K-Fold* maka semakin tinggi pula variasi data. Tidak hanya mendapatkan variasi data yang paling baik, bisa juga terdapat kemungkinan mendapatkan variasi data yang paling buruk. Sebaliknya semakin kecil nilai pembagi *K-Fold* maka semakin kecil pula variasi data dan tidak menjamin variasi data yang didapatkan menjadi yang paling baik.



Gambar 3. 5 Grafik Hasil Evaluasi Pembagian 5 K-Fold

Selanjutnya berdasarkan grafik yang ditunjukkan pada **Gambar 3. 5** terlihat bahwa rata-rata nilai evaluasi pada $k = 4$ merupakan pembagian data yang paling stabil dibandingkan pembagian data dengan nilai k yang lain pada *5 K-Fold Cross Validation* dengan rata-rata hasil nilai akurasi 97%, presisi 95% dan *recall* 99%. Dapat disimpulkan bahwa $k = 4$ merupakan pembagian yang paling baik dengan random data *training* dan *testing* paling optimal. Dengan demikian seluruh data *training* ini dapat merepresentasikan keadaan data *testing* dengan baik. Sebaliknya nilai $k = 3$ terlihat paling rendah nilai evaluasinya dengan nilai hasil rata-rata akurasi 82%, presisi 79% dan *recall* 78%. Oleh sebab itu, pembagian pada $k = 3$ merupakan variasi random data *training* dan *testing* yang buruk, sehingga prediksi data *testing* banyak yang gagal atau salah prediksi.

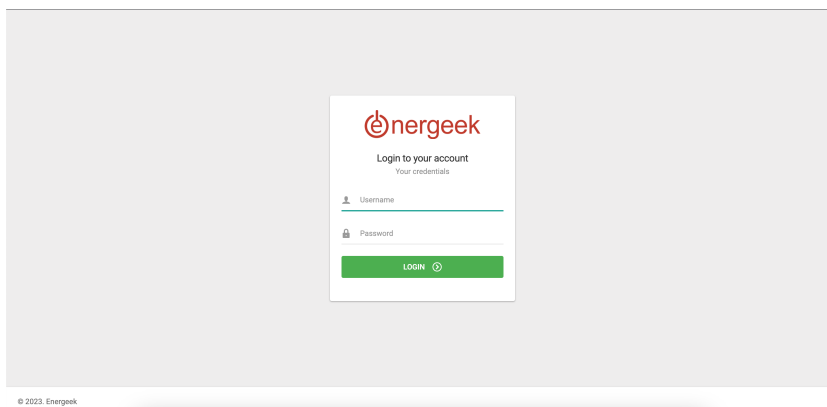


Gambar 3. 6 Grafik Hasil Evaluasi Pembagian 10 K-Fold

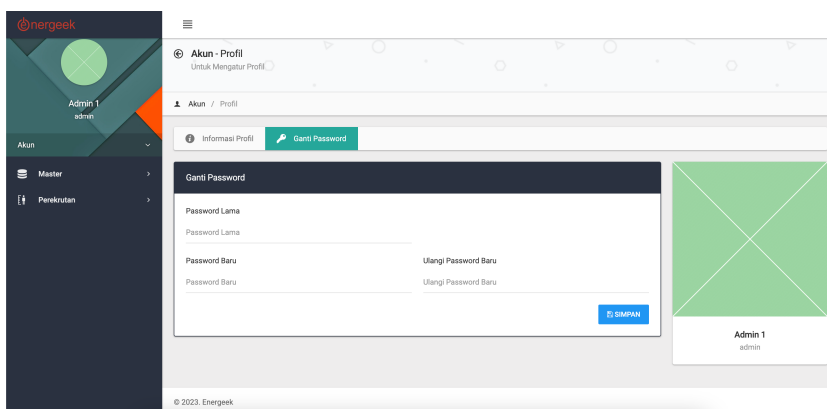
Berdasarkan grafik **Gambar 3. 6** menunjukkan bahwa rata-rata nilai evaluasi pada $k = 4$ dan $k = 5$ merupakan pembagian data yang paling stabil dibandingkan pembagian data dengan hasil dari k yang lain pada *10 K-Fold Cross Validation* dengan hasil nilai rata-rata akurasi, presisi dan *recall* adalah 100%. Dapat disimpulkan bahwa $k = 4$ dan $k = 5$ merupakan pembagian yang paling baik dengan random data *training* dan *testing* paling optimal. Artinya bahwa seluruh data *training* dapat merepresentasikan keadaan data *testing* dengan baik. Sebaliknya nilai $k = 3$ terlihat paling rendah nilai-nilai evaluasinya dengan hasil nilai rata-rata adalah akurasi 82%, presisi 82% dan *recall* 85%. Oleh sebab itu, pembagian pada $k = 3$ merupakan variasi random data *training* dan *testing* yang buruk, sehingga prediksi data *testing* banyak yang gagal atau salah prediksi. Dapat dilihat pada grafik pembagian *K-Fold Cross Validation* dengan pembagian 10 memiliki rata-rata nilai akurasi, presisi dan *recall* diatas 80% bahkan terdapat k yang menyentuh nilai 100%, jika dibandingkan dengan pembagian 5 terdapat rata-rata nilai dibawah 80%.

3.3. Antarmuka Sistem

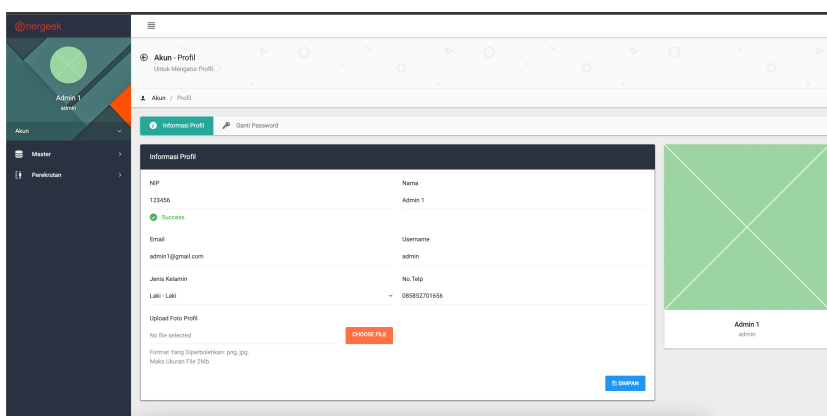
Sistem yang dibangun menggunakan konsep *material design*, berikut antarmuka sistem masing-masing menu yang ada pada sistem.



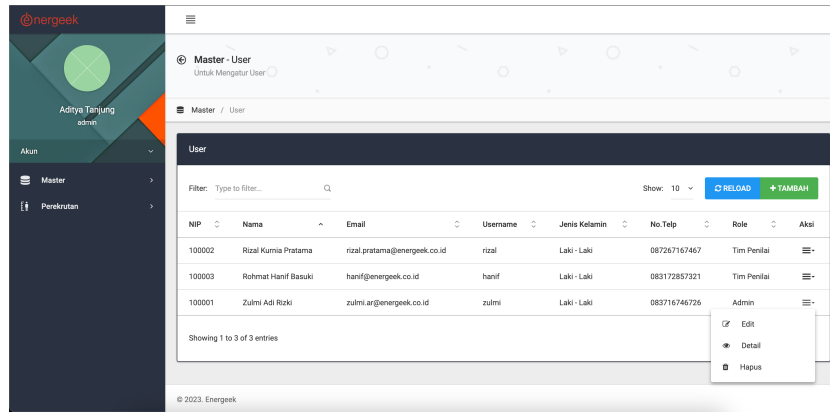
Gambar 3. 7 Tampilan Login



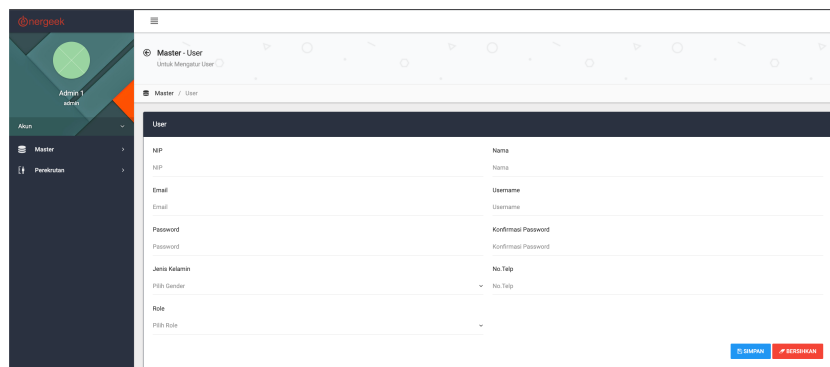
Gambar 3. 8 Tampilan Form Ubah Password



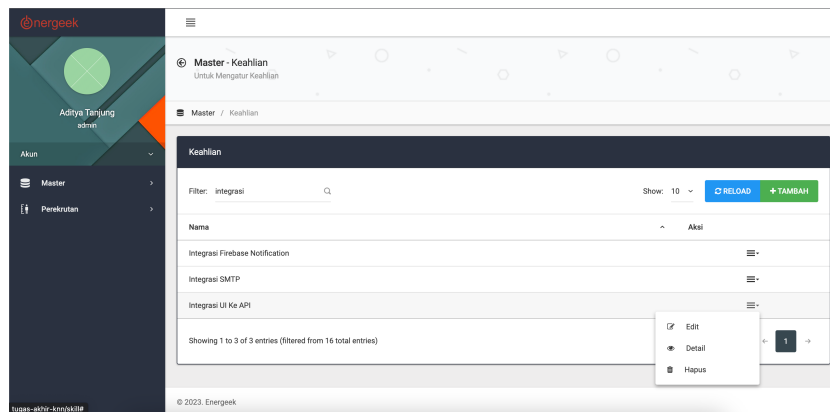
Gambar 3. 9 Tampilan Form Ubah Profil



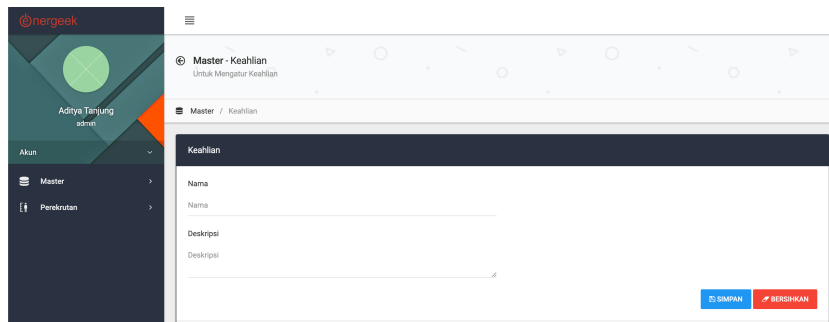
Gambar 3. 10 Tampilan Tabel User



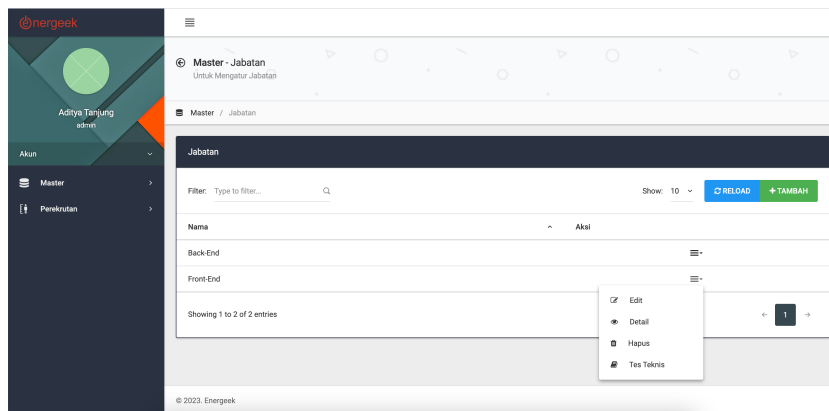
Gambar 3. 11 Tampilan Form User



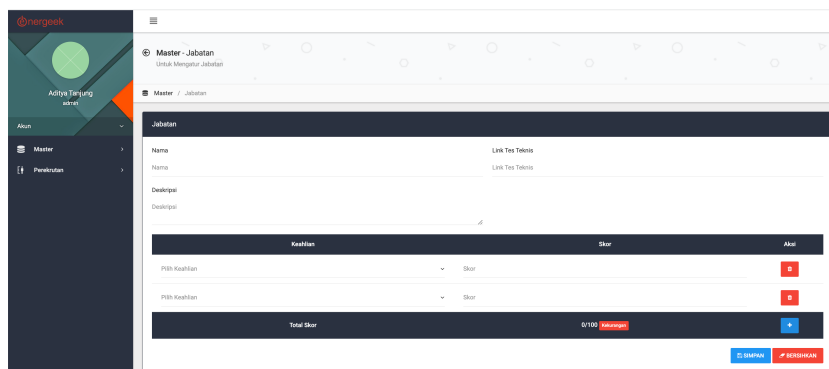
Gambar 3. 12 Tampilan Tabel Keahlian



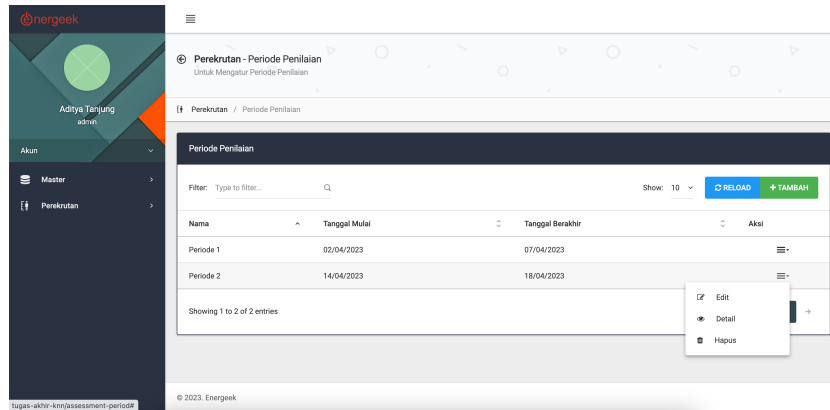
Gambar 3. 13 Tampilan Form Keahlian



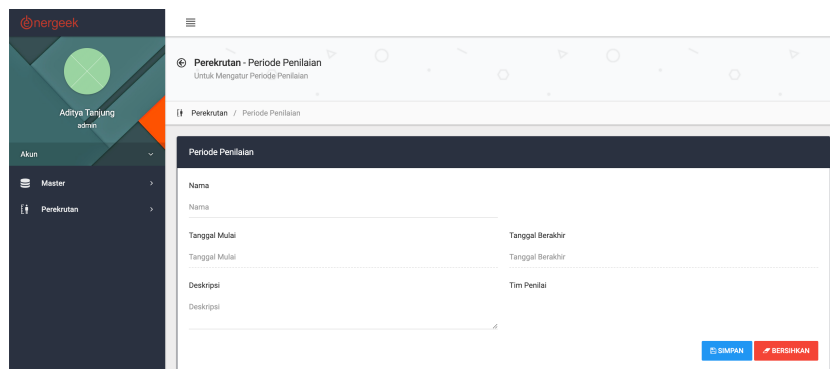
Gambar 3. 14 Tampilan Tabel Jabatan



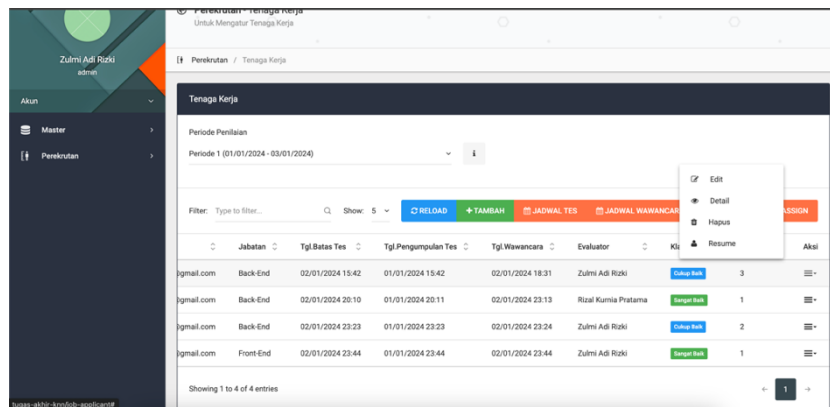
Gambar 3. 15 Tampilan Form Jabatan



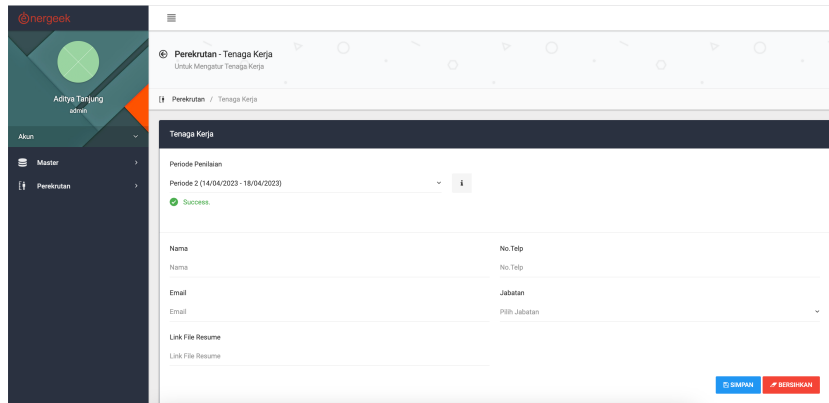
Gambar 3. 16 Tampilan Tabel Periode Penilaian



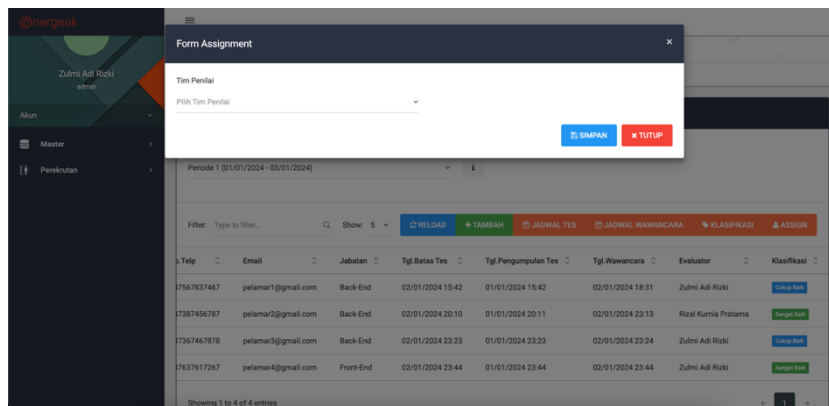
Gambar 3. 17 Tampilan Form Periode Penilaian



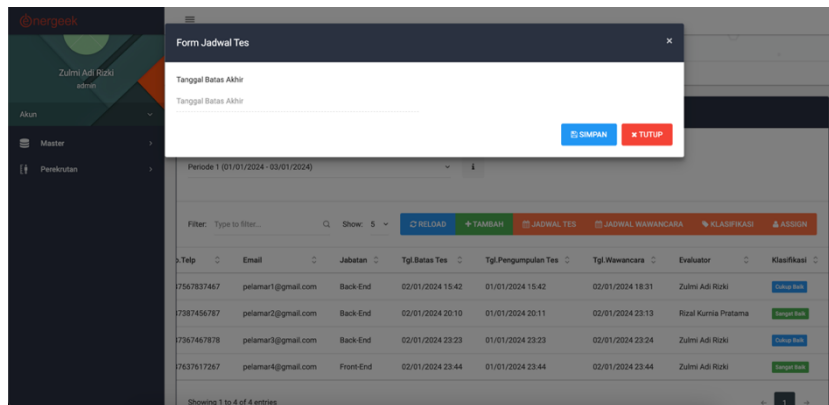
Gambar 3. 18 Tampilan Tabel Tenaga Kerja



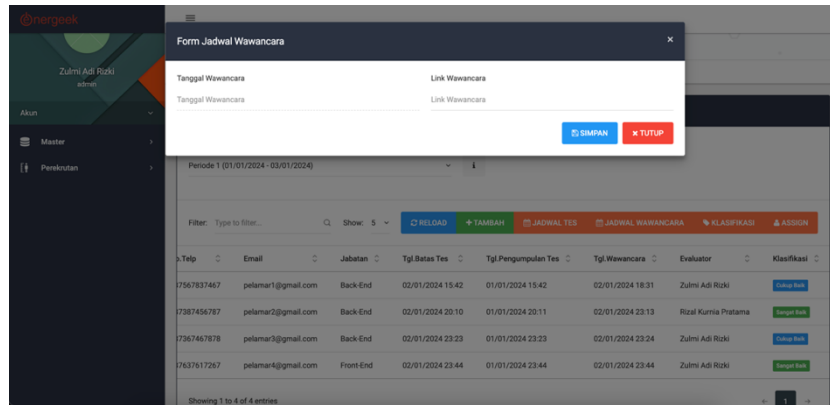
Gambar 3. 19 Tampilan Form Tenaga Kerja



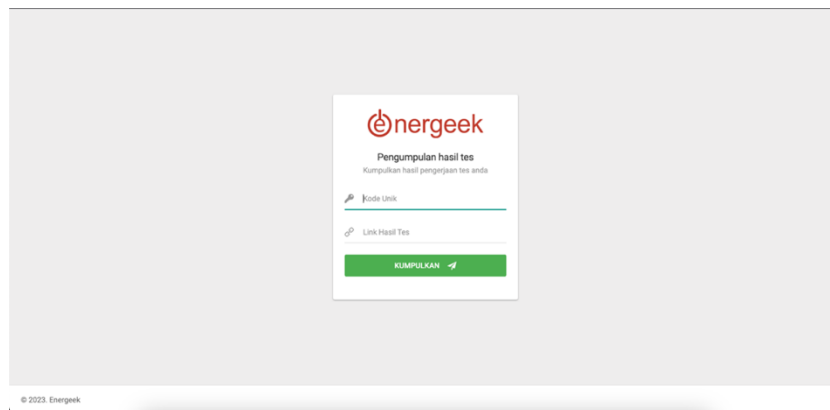
Gambar 3. 20 Tampilan Form *Assignment* Tim Penilai



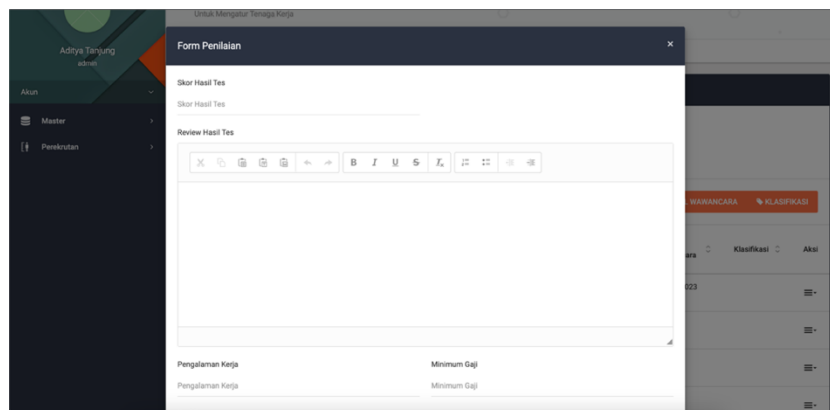
Gambar 3. 21 Tampilan Form Jadwal Tes



Gambar 3. 22 Tampilan Form Jadwal Wawancara



Gambar 3. 23 Tampilan Form Pengumpulan Hasil Tes



Gambar 3. 24 Tampilan Form Penilaian

3.4. Pengujian System Usability Scale

Pada pengujian *SUS* ini guna menilai seberapa besar kegunaan dari sebuah sistem yang telah dibuat maka penulis telah mengumpulkan beberapa jawaban dari responden melalui formulir kuesioner yang berisi 10 pertanyaan.

Tabel 3. 3 Komponen Pertanyaan SUS

No	Komponen Pertanyaan
1	Saya berpikir akan menggunakan sistem ini lagi.
2	Saya merasa sistem ini rumit untuk digunakan.
3	Saya merasa sistem ini mudah digunakan.
4	Saya membutuhkan bantuan dari orang lain atau teknisi dalam menggunakan sistem ini.
5	Saya merasa fitur-fitur sistem ini berjalan dengan semestinya
6	Saya merasa ada banyak hal yang tidak konsisten (tidak serasi pada sistem ini).
7	Saya merasa orang lain akan memahami cara menggunakan sistem ini dengan cepat.
8	Saya merasa sistem ini membingungkan.
9	Saya merasa tidak ada hambatan dalam menggunakan sistem ini.
10	Saya perlu membiasakan diri terlebih dahulu sebelum menggunakan sistem ini.

Jawaban dari setiap responden nantinya akan dilakukan perhitungan menggunakan rumus (2. 5). Berikut data skor akhir jawaban dari responden yang berhasil di dapatkan yang berjumlah 12 responden yang selanjutnya dilakukan rata-rata terhadap semua skor akhir responden untuk mendapatkan kategori nilai *SUS*.

Tabel 3. 4 Jawaban Kuesioner *SUS*

No	Nama	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Skor
1	Riska	5	1	5	1	5	1	5	1	5	3	95
2	Chafid	5	1	5	1	5	1	5	1	5	1	100
3	Farida	4	5	5	2	5	1	5	1	4	4	75
4	Diiky	5	1	5	1	5	1	5	1	5	1	100
5	Yuri	5	1	5	5	5	1	4	2	4	5	72,5
6	Gama	4	1	5	5	5	2	4	1	4	4	72,5
7	Rizky	1	1	1	2	5	1	5	2	5	2	72,5
8	Ikbar	5	1	5	3	4	2	5	2	5	3	82,5
9	Tama	4	2	4	1	5	2	4	2	4	2	80
10	Hayu	5	1	5	4	5	1	4	5	5	5	70
11	Hafied	4	3	4	4	4	3	4	3	4	4	57,5
12	Aloy	4	2	4	4	4	2	4	2	4	4	85
Rata - Rata											80,21	

Dari hasil rata-rata diatas diperoleh nilai sebesar 80,21 yang jika dilihat pada **Tabel 2. 2** maka kategori nilai yang di dapat untuk kegunaan sistem yang telah dibuat adalah B.

IV. KESIMPULAN

Penelitian ini memiliki beberapa kesimpulan dari hasil penelitian yang telah dilakukan. Aplikasi dari hasil uji *SUS* mendapatkan rata-rata nilai dari keseluruhan responden sebesar 80,21 yang dimana skor tersebut masuk kedalam kategori nilai B yang bisa dikatakan dari segi kegunaan dan kemudahan yang dirasakan oleh responden cukup memenuhi. Pada pengujian algoritma *KNN* menggunakan perhitungan jarak *euclidean distance* didapatkan model terbaik dengan nilai rata-rata akurasi, presisi, *recall* sebesar 100% pada pembagian $k = 4$ dan $k = 5$ di uji coba pengambilan 10 *K-Fold Cross Validation*, sedangkan nilai ketetangaan metode *KNN* paling baik ada di $k = 7$. Hasil penelitian menunjukkan bahwa masih ada beberapa kekurangan, seperti fitur yang dimiliki oleh sistem masih sangat sedikit dikarenakan dalam pengembangannya disesuaikan dengan kebutuhan saat penelitian berlangsung, akan tetapi tidak menutup kemungkinan akan terjadi peningkatan kebutuhan dalam perusahaan yang mengharuskan terjadinya penambahan fitur atau penyesuaian terhadap fitur sebelumnya. Dataset

yang digunakan dalam penelitian ini juga sedikit sehingga dibutuhkan penambahan jumlah dataset agar tingkat akurasi dalam proses klasifikasi lebih tinggi dari sebelumnya.

DAFTAR PUSTAKA

- [1] H. Chandra dan S. Tinggi Ilmu Ekonomi YPUP Makassar, “ANALISIS SISTEM SELEKSI PENERIMAAN KARYAWAN TERHADAP KINERJA KARYAWAN PADA PT ANUGRAH BUSANA INDAH DI MAKASSAR,” 2020.
- [2] A. Rahmat, K. Auliasari, dan Y. A. Pranoto, “IMPLEMENTASI METODE K-NEAREST NEIGHBOR (KNN) UNTUK SELEKSI CALON KARYAWAN BARU (Studi Kasus : BFI Finance Surabaya),” 2020.
- [3] I. G. T. Isa dan F. Elfaladonna, “Penilaian Kinerja Akurasi Metode Klasifikasi dalam Dataset Penerimaan Mahasiswa Baru Universitas XYZ,” *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 8, no. 2, hlm. 292–298, 2022.
- [4] R. B. Anshori *dkk.*, “Klasifikasi Citra Kanker Serviks Menggunakan Deep Residual Network,” vol. 8, no. 6, hlm. 3163–3170, 2022.
- [5] I. Listiowarni dan N. P. Dewi, “Pemanfaatan Klasifikasi Soal Biologi Cognitive Domain Bloom’s Taxonomy Menggunakan KNN Chi-Square Sebagai Penyusunan Naskah Soal,” *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, vol. 11, no. 2, hlm. 185–195, 2020.
- [6] A. Bangor, P. Kortum, dan J. Miller, “Determining what individual SUS scores mean: Adding an adjective rating scale,” *J Usability Stud*, vol. 4, no. 3, hlm. 114–123, 2009.