

Penerapan *Data Mining* Menggunakan Algoritma *K- Nearest Neighbor* untuk Penentuan Penerimaan Proposal Hibah

Sri Widaningsih^{1*}, Agus Suheri², Tiara Tsaniya Fauziyana³

^{1,2,3} Teknik Informatika Universitas Suryakancana
Jl. Pasir Gede Raya Cianjur

^{1*}sriwida@unsur.ac.id

²agussuheri@unsur.ac.id

³tiaratsaniya@gmail.com

Application of Data Mining Using The K-Nearest Neighbor Algorithm for Determining Acceptance of Grand Proposals

Dikirimkan : 11, 2022. Diterima : 09, 2023. Dipublikasikan : 05, 2024.

Abstract— The grant program for farmers is one of the government programs provided by the Cianjur Regency Agriculture Office. To accept the program, farmers submit proposals which will be reviewed by the proposal reviewer analyst. In making a decision to accept a grant proposal, it is necessary to make a quick and accurate decision. With quite a lot of proposal data that has been submitted, it can be used data patterns that have been stored with data mining methods to predict proposals that are accepted with proposals that are rejected. The algorithm used is K-Nearest Neighbor where the determination of acceptance of new proposal data is based on the closest distance to the acceptance of the previous proposal. There are six predictor variables used, namely assets, capital, business scale, length of business, cost plan and completeness of documents. The stages of data mining use the stages in Knowledge Discovery in Databases (KDD). Calculations use the weight of the assessment for each proposal criteria that have been determined. The results of the evaluation of the calculation of the kNN algorithm obtained a real accuracy of 76% and an error rate of 24%. From the calculation results, the kNN algorithm can be used as a model to determine recipients of agricultural grant programs.

Keywords— Proposal; Data mining; KDD; Classificatio; kNN

Abstrak— Program pemberian dana hibah untuk para petani merupakan salah satu program pemerintah yang diberikan oleh Dinas Pertanian Kabupaten Cianjur. Untuk menerima program tersebut petani mengajukan proposal yang akan diperiksa oleh analis pemeriksa proposal. Dalam mengambil keputusan penerimaan proposal hibah perlu adanya keputusan yang cepat dan akurat. Dengan cukup banyaknya data proposal yang telah diajukan, maka dapat digunakan pola data yang telah tersimpan dengan metode *data mining* untuk memprediksi proposal yang diterima dengan proposal yang ditolak. Algoritma yang digunakan yaitu *K-Nearest Neighbor* di mana penentuan penerimaan data proposal baru berdasarkan pada jarak terdekat dengan penerimaan proposal sebelumnya. Terdapat enam variabel *prediktor* yang digunakan yaitu aset, modal, skala usaha, lama usaha, rencana biaya dan kelengkapan dokumen. Tahapan *data mining* menggunakan tahapan dalam *knowledge Discovery in Databases* (KDD). Perhitungan menggunakan bobot penilaian untuk setiap kriteria proposal yang sudah ditentukan. Hasil evaluasi perhitungan algoritma kNN diperoleh akurasi sebenar 76% dan tingkat *error* sebesar 24%. Dari hasil perhitungan, algoritma kNN ini dapat digunakan sebagai model untuk menentukan penerima program hibah pertanian.

Kata kunci— Proposal; Data mining; KDD; Klasifikasi; kNN

I. PENDAHULUAN

Dinas Pertanian Kabupaten Cianjur merupakan salah satu instansi yang memiliki program yang diberikan oleh Kementerian Pertanian untuk memberikan dana hibah kepada para petani. Program ini dimulai dari tahun 2019 yang bertujuan untuk meningkatkan usaha para petani. Hibah yang diberikan tidak terbatas di sektor pertanian tetapi juga di sektor peternakan, perikanan dan perkebunan seperti budidaya bebek petelur dan pedaging, budidaya tanaman hias, budidaya bunga potong dan lain-lain. Setiap tahunnya sudah cukup banyak petani yang mengikuti program hibah ini. Pengelolaan proposal hibah dilakukan di Seksi Bina Usaha Agrobisnis yang berada di bawah Seksi Bina Usaha Agrobisnis dan Penyuluhan. Proposal hibah dilakukan oleh petani secara mandiri dan bukan kelompok tani. Untuk memperoleh dana hibah tersebut, petani mengajukan proposal mengenai kegiatan serta jumlah dana yang diajukan. Berdasarkan proposal yang masuk pada Seksi Bina Usaha Agrobisnis, pada tahun 2019 terdapat 150 proposal yang diterima untuk pengajuan dana hibah tersebut.

Untuk menentukan penerimaan proposal saat ini yaitu dengan memeriksa dokumen proposal oleh analis proposal pada Seksi Bina Usaha Agrobisnis dilihat dari kelengkapan maupun isi proposal lalu diputuskan berdasarkan beberapa kriteria. Kriteria yang ditetapkan berdasarkan wawancara dengan Kepala Seksi Bina usaha Agrobisnis pada bulan Februari 2019. Berdasarkan hasil wawancara tersebut ditetapkan enam kriteria yang dipertimbangkan yaitu aset, modal, skala usaha, lama usaha, nilai ajuan dan kelengkapan dokumen. Apabila proposal yang diajukan telah memenuhi kriteria yang ditetapkan maka selanjutnya akan ditentukan proposal yang layak dan diterima.

Saat ini merupakan era data, di mana data berukuran besar sangat mudah dan cepat untuk dikumpulkan. Pertumbuhan volume data yang tersedia secara eksplisif ini merupakan hasil komputerisasi dari masyarakat dan pesatnya perkembangan alat pengumpulan dan penyimpanan data yang canggih. Untuk itu dibutuhkan suatu alat yang ampuh yang secara otomatis dapat mengungkap informasi berharga dari sejumlah besar data dan melakukan transformasi data tersebut menjadi pengetahuan yang terorganisir. Metode yang dapat digunakan untuk menggali pengetahuan dari sekumpulan besar data yaitu dengan *data mining* [1]

Banyaknya petani yang mengajukan proposal maka menyebabkan data yang tersimpan di Seksi Bina Usaha Agrobisnis tersebut juga semakin banyak. Data yang diperoleh dapat diolah lebih lanjut sehingga menjadi suatu pengetahuan yang bermanfaat untuk mengambil keputusan penentuan penerimaan proposal hibah. Dengan metode *data*

mining dapat diketahui suatu pola penerimaan proposal dari data yang telah tersimpan sehingga proses penentuan petani yang menerima hibah lebih cepat, mudah dan tepat. Hal tersebut juga akan mempermudah seorang analis proposal untuk mengambil keputusan. Untuk proposal yang tidak diterima maka akan diberitahukan kembali kepada petani yang mengajukan bahwa proposalnya tidak lolos dalam penerimaan hibah.

Data mining adalah proses pencarian pola dan *trend* yang berguna dari sekumpulan data dengan jumlah besar [2]. Terdapat beberapa tahapan dalam penerapan *data mining*. *Data mining* adalah studi untuk mengumpulkan, membersihkan, mengorganisir, menganalisis, dan memperoleh pengetahuan bermanfaat yang dalam dari data [3]. *Data mining* juga bisa digunakan untuk dasar pengambilan keputusan. Menurut [4] *data mining* adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan. Terdapat beberapa fungsi dari *data mining* sesuai dengan tujuan dari penggunaannya. Fungsi-fungsi *data mining* di antaranya yaitu deskripsi, estimasi, prediksi, klasifikasi, pengelompokan dan asosiasi [2][5]. Dalam penelitian mengenai penentuan penerimaan proposal hibah, fungsi *data mining* yaitu fungsi klasifikasi.

Klasifikasi adalah bentuk analisis data yang mengekstrak model yang menggambarkan kelas data [6]. Klasifikasi berfungsi menentukan item-item ke dalam kategori target atau kelas. Tujuan klasifikasi adalah memprediksi secara akurat target kelas untuk setiap *record* data [7]. Ada beberapa algoritma dalam teknik klasifikasi yaitu *decision tree*, *naïve bayes*, *support vector machine*, *K-Nearest Neighbor* (kNN), dan *rule base classifier* [2] [6][7]. Algoritma *decision tree*, *naïve bayes*, SVM dan *rule base classifier* termasuk ke dalam *eager learner* sedangkan algoritma kNN termasuk ke dalam *lazy learner*. Pada *eager learner* ketika diberikan sekelompok data latih, akan membentuk model generalisasi sebelum menerima data baru atau data uji untuk diklasifikasi. Sedangkan pada *lazy learner* ketika diberikan data latih akan menyimpannya dan menunggu hingga diberikan data uji. Data uji secara umum diklasifikasi berdasarkan kesamaan pada data latih yang disimpan [6].

Dalam penelitian ini algoritma yang digunakan yaitu kNN karena data yang digunakan merupakan data yang bersifat non linier sehingga tidak dibutuhkan asumsi dalam perhitungannya. Selain itu jumlah data yang tidak terlalu besar menjadi alasan pemilihan algoritma ini sehingga waktu perhitungan menjadi lebih singkat. Algoritma kNN merupakan algoritma yang paling sering digunakan dalam klasifikasi [2][8][9] dapat digunakan juga

untuk estimasi dan prediksi. kNN merupakan algoritma klasifikasi non *parametrik* di mana tidak membutuhkan *pra* asumsi apa pun pada kelompok data. Digunakan untuk mengklasifikasi data berdasarkan pada kedekatan dengan data latih yang diberikan. Untuk data yang bersifat kontinu untuk menghitung kedekatan dapat menggunakan jarak *euclidean*. Algoritma kNN merupakan algoritma yang paling sederhana di antara algoritma dalam mesin pembelajaran [7]. Kelebihan lainnya yaitu kesederhanaan, komprehensif dan skala *bilitas*. Algoritma ini mudah diinterpretasi dan sangat efektif untuk data latih yang besar. Perhitungan matematis mudah dipahami, berguna untuk data yang bersifat non linier.

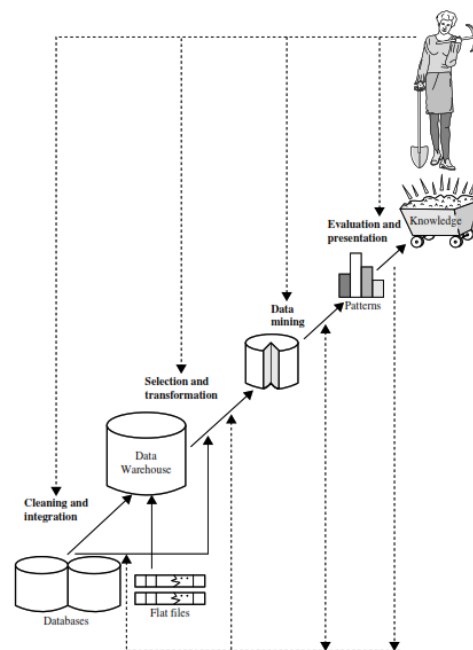
Sedangkan kelemahan algoritma ini yaitu kNN dapat menjadi mahal dalam menentukan nilai K jika *dataset* berukuran besar, membutuhkan penyimpanan yang lebih besar sehingga pada proses prediksi menjadi lambat. Keakuratan penghitungan jarak menentukan peranan penting dalam keakuratan algoritma. Karena termasuk ke dalam *lazy learner* maka tidak ada pembelajaran dari data latih[7][8][10][11].

Karena merupakan algoritma yang paling banyak digunakan, maka telah banyak penelitian penerapan kNN dalam berbagai bidang seperti *text mining*, keuangan, pertanian, pengobatan dan lain-lain [8][10][11]. Penerapannya dalam *text mining* adalah mengklasifikasikan jurnal ilmu komputer [12] atau mengukur kinerja suatu instansi melalui klasifikasi teks pada *twitter* [13]. Pada bidang pendidikan algoritma kNN digunakan untuk memprediksi nilai dan waktu kelulusan [14][15], selain itu algoritma ini dapat memprediksi siswa berprestasi [16] maupun penentuan beasiswa [17]. Pada bidang penjualan dapat memprediksi penjualan produk seperti tekstil [18]. Penerapan algoritma kNN tidak terbatas pada data yang bersifat struktur, tetapi juga pada data tidak terstruktur seperti proses klasifikasi bunga berdasarkan ekstraksi fitur warna pada gambar [19].

Berdasarkan dari latar belakang di atas maka tujuan penelitian ini yaitu memanfaatkan salah satu algoritma data *mining* yaitu kNN untuk menentukan penerimaan proposal hibah pertanian diajukan petani kepada Seksi Bina Usaha Agrobisnis.

II. METODOLOGI PENELITIAN

Dalam penelitian ini, penerapan data *mining* untuk penentuan proposal hibah mengikuti tahapan *Knowledge Discovery in Databases* (KDD). Proses penemuan pengetahuan merupakan urutan *iteratif* dari beberapa langkah [6]. Gambar 1 menunjukkan tahapan dalam proses KDD.



Gambar 1. Langkah-langkah proses KDD [1]

Berikut ini langkah-langkah proses KDD dalam penelitian [1]:

1. *Data cleaning*
Pada tahap ini dilakukan pembersihan data untuk menghilangkan *noise* dan data yang tidak konsisten. Data proposal dibersihkan dari *redudansi* data, data yang *outlier*, maupun data yang kosong.
2. *Data integration*
Pada tahap ini dilakukan penggabungan data yang berasal dari beberapa sumber.
3. *Data selection*
Data yang relevan dengan penelitian diambil dari *database* untuk dianalisis. Kriteria-kriteria yang digunakan untuk menentukan penerimaan proposal hibah termasuk ke dalam variabel *prediktor* yaitu modal, aset, skala usah, lama usaha, Rencana biaya dan kelengkapan dokumen. Pada tabel I ditunjukkan variabel-variabel *prediktor* yang menjadi dasar penentuan diterima atau tidaknya proposal hibah.
4. *Data transformation*
Transformasi dengan cara *diskritisasi* dilakukan pada beberapa variabel *prediktor* yaitu modal usaha, lama usaha, dan rencana biaya. *Diskritisasi* yaitu mengubah data dari atribut kontinu menjadi atribut kategoris. Proses *diskritisasi* variabel-variabel *prediktor* yang bersifat kontinu ditunjukkan pada tabel II.

TABEL I
VARIABEL-VARIABEL PREDIKTOR

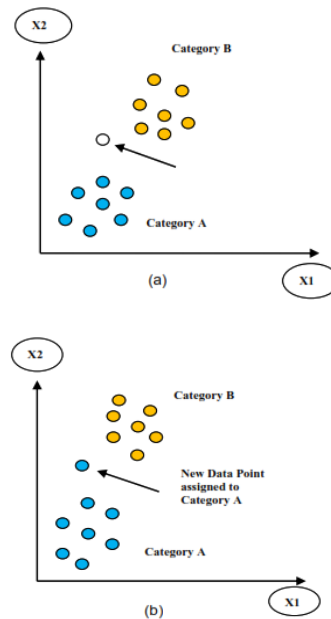
Jenis	Nilai	Keterangan
Aset	Ada	Dana awal yang dimiliki oleh petani.
	Tidak ada	
Modal	Jumlah dalam rupiah	Total dana awal yang dikeluarkan oleh petani pada saat membangun usahanya.
Skala usaha	Mikro : Lingkup usaha yang dijalankan masih kecil, tenaga kerja masih sedikit dan <i>asset</i> yang dimiliki tidak terlalu besar	Luas ruang lingkup usaha yang sedang dijalani.
	Makro : Lingkup usaha yang dijalankan masih kecil, tenaga kerja masih sedikit dan <i>asset</i> yang dimiliki tidak terlalu besar	
Lama usaha	Lamanya usaha dalam tahun	Untuk mengetahui lamanya petani dalam menjalankan usaha
Rencana biaya	Jumlah dalam rupiah	Untuk mengetahui nilai biaya yang diajukan.
Dokumen	Lengkap,	Kelengkapan dokumen-dokumen dari proposal yang diajukan.
	Tidak lengkap	

TABEL II
TRANSFORMASI DATA KONTINU

Variabel	Range	Nilai
Modal usaha	1 – 35 juta	Kecil
	36 – 65 juta	Sedang
	66 – 100 juta	Besar
Lama usaha	0 – 1 tahun	Sebentar
	1 – 2 tahun	Sedang
	> 2 tahun	Lama
Rencana Biaya	1 – 35 juta	Kecil
	36 – 65 juta	Sedang
	66 – 100 juta	Besar

5. Data mining

Pada tahap ini diterapkan algoritma kNN untuk menentukan penerimaan proposal hibah oleh Dinas Pertanian. Ilustrasi mengenai algoritma kNN ditunjukkan pada gambar 2[9], di mana data baru dimasukkan ke dalam kategori A karena memiliki jarak yang lebih dekat dengan kategori A yang berwarna biru dibandingkan dengan kategori B.



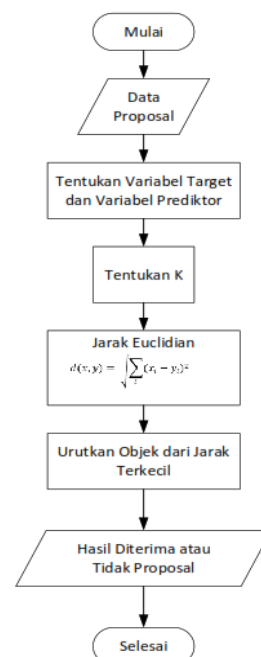
Gambar 2. Sebelum kNN (a) dan sesudah kNN (b)

Nilai k yang digunakan yaitu diambil $k = 3$. Untuk menghitung jarak kedekatan menggunakan rumus *euclidean* sebagai berikut :

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (1)$$

Di mana $x = x_1, x_2, \dots, x_m$ dan $y = y_1, y_2, \dots, y_m$ menunjukkan m nilai atribut dari 2 data.

Flowchart algoritma kNN ditunjukkan pada gambar 3.



Gambar 3. Flowchart algoritma kNN

Untuk menentukan kedekatan antara variabel yang bersifat kategori, digunakan bobot kedekatan sebagai berikut. Bobot kedekatan untuk setiap variabel aset dapat dilihat pada tabel III.

TABEL III
BOBOT KEDEKATAN VARIABEL ASSET

Nilai 1	Nilai 2	Bobot
Ada	Ada	0
Ada	Tidak Ada	1
Tidak Ada	Tidak Ada	0

Bobot kedekatan variabel modal usaha ditunjukkan pada tabel IV.

TABEL IV
BOBOT KEDEKATAN VARIABEL MODAL USAHA

Nilai 1	Nilai 2	Bobot
Kecil	Kecil	0
Kecil	Sedang	0,5
Kecil	Besar	1
Sedang	Sedang	0
Sedang	Besar	0,5
Besar	Besar	0

Bobot kedekatan variabel skala usaha ditunjukkan pada tabel V.

TABEL V
BOBOT KEDEKATAN VARIABEL SKALA USAHA

Nilai 1	Nilai 2	Bobot
Mikro	Mikro	0
Mikro	Makro	1
Makro	Makro	0

Bobot kedekatan variabel lama usaha ditunjukkan pada tabel VI.

TABEL VI
BOBOT KEDEKATAN VARIABEL LAMA USAHA

Nilai 1	Nilai 2	Bobot
Sebentar	Sebentar	0
Sebentar	Sedang	0,5
Sebentar	Lama	1
Sedang	Sedang	0
Sedang	Lama	0,5
Lama	Lama	0

Bobot kedekatan variabel lama usaha ditunjukkan pada tabel VII.

TABEL VII
BOBOT KEDEKATAN VARIABEL RENCANA BIAYA

Nilai 1	Nilai 2	Bobot
Kecil	Kecil	0
Kecil	Sedang	0,5
Kecil	Besar	1
Sedang	Sedang	0
Sedang	Besar	0,5
Besar	Besar	0

Bobot kedekatan variabel lama usaha ditunjukkan pada tabel VI.

TABEL VIII
BOBOT KEDEKATAN KELENGKAPAN DOKUMEN

Nilai 1	Nilai 2	Bobot
Lengkap	Lengkap	0
Lengkap	Tidak Lengkap	1
Tidak Lengkap	Tidak Lengkap	0

6. Evaluasi pola

Pada tahap ini dilakukan pola yang muncul dari hasil perhitungan algoritma kNN yang digunakan untuk penentuan penerimaan proposal hibah. Evaluasi hasil perhitungan algoritma dengan algoritma kNN menggunakan *confusion matrix*. *Confusion matrix* digunakan untuk mengukur keakuratan model yang dibuat. *Confusion matrix* adalah tabel yang digunakan untuk menunjukkan jumlah prediksi yang salah dan benar pada permasalahan klasifikasi ketika nilai sebenarnya dari data uji diketahui. Jika data set hanya terdiri dari dua kelas maka kelas satu dianggap positif dan kelas lainnya dianggap negatif [20]. Gambar 4 merupakan bentuk dari *confusion matrix* [6]

		Predicted class		
		yes	no	
Actual class	yes	TP	FN	P
	no	FP	TN	N
Total		P'	N'	P + N

Gambar 4. *Confusion matrix*

TP : *True Positive*

Data positif diklasifikasikan secara benar

TN : *True Negative*

Data negatif diklasifikasikan secara benar

FP : *False Positive*

Data positif diklasifikasikan secara salah

FN : *False Negative*

Data negatif diklasifikasikan secara salah

Beberapa ukuran evaluasi yaitu akurasi dan error

$$Akurasi = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$Error = \frac{FP + FN}{P + N} = \frac{FP + FN}{TP + FP + TN + FN} \quad (3)$$

7. Presentasi pengetahuan

Presentasi pengetahuan yang dihasilkan berupa penentuan proposal mana yang

diterima atau ditolak yang akan diinformasikan oleh Dinas Pertanian

III. HASIL PENELITIAN

A. Hasil Perhitungan Algoritma kNN

Berikut ini adalah simulasi hasil perhitungan algoritma kNN untuk penentuan penerimaan proposal hibah. Data proposal yang telah ditransformasi ditunjukkan pada tabel IX.

Contoh :

Jika ada jenis usaha F yang mengajukan proposal hibah tentukan apakah proposal tersebut diterima berdasarkan kedekatan dengan data sebelumnya.

Hasil Perhitungan :

$$\begin{aligned} D_{A-F} &= \sqrt{0^2 + 0,5^2 + 0^2 + 0,5^2 + 0,5^2 + 0^2} \\ &= \sqrt{0 + 0,25 + 0 + 0,25 + 0,25 + 0} \\ &= \sqrt{0,75} = 0,867 \end{aligned}$$

$$D_{B-F} = \sqrt{0,5^2 + 0^2 + 0^2 + 0^2 + 0,5^2 + 0,5^2}$$

$$= \sqrt{0,25 + 0 + 0 + 0 + 0,25 + 0,25}$$

$$= \sqrt{0,75} = 0,876$$

$$\begin{aligned} D_{C-F} &= \sqrt{0^2 + 0,5^2 + 0,5^2 + 0,5^2 + 0,5^2 + 0^2} \\ &= \sqrt{0 + 0,25 + 0,25 + 0,25 + 0,25 + 0} \end{aligned}$$

$$= \sqrt{1} = 1$$

$$D_{D-F} = \sqrt{0^2 + 0^2 + 0,5^2 + 0,5^2 + 0,5^2 + 0^2}$$

$$= \sqrt{0 + 0 + 0,25 + 0,25 + 0,25 + 0}$$

$$= \sqrt{0,75} = 0,876$$

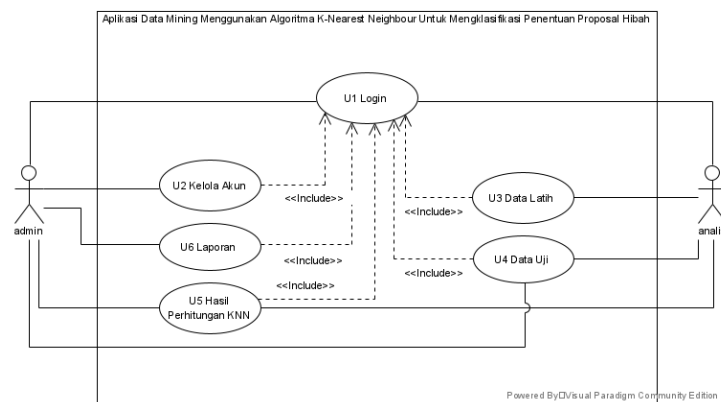
$$D_{E-F} = \sqrt{1^2 + 0^2 + 0^2 + 0^2 + 0,5^2 + 1^2}$$

$$= \sqrt{1 + 0 + 0 + 0 + 0,25 + 1}$$

$$= \sqrt{2,25} = 1,5$$

TABEL IX
CONTOH DATA PROPOSAL HIBAH

Nama Usaha	Aset	Modal	Skala_usaha	Lama_usaha	Rencana_biaya	Dokumen	Keterangan
A	Ada	Sedang	Mikro	Lama	Besar	Lengkap	Diterima
B	Tidak Ada	Kecil	Mikro	Sedang	Kecil	Tidak Lengkap	Diterima
C	Ada	Sedang	Makro	Lama	Besar	Lengkap	Tidak Diterima
D	Ada	Kecil	Makro	Lama	Kecil	Lengkap	Diterima
E	Tidak Ada	Kecil	Mikro	Sedang	Kecil	Tidak Lengkap	Tidak Diterima
F	Ada	Kecil	Mikro	Sedang	Sedang	Lengkap	?



Gambar 5. Use case diagram

Dari perhitungan jarak antara kasus baru dengan kasus lama, jarak terdekat adalah antara kasus baru dengan kasus A, B, dan D dengan jarak 0,867 di mana semua proposal hibah diterima. Ini berarti proposal F juga diprediksi akan diterima.

B. Pemodelan Aplikasi

Berikut ini adalah tahapan-tahapan dalam pembuatan aplikasi *data mining* dengan menggunakan algoritma kNN untuk penentuan proposal hibah. Data proposal hibah yang diterima

dan yang akan diolah menjadi data latih sebanyak 150 proposal.

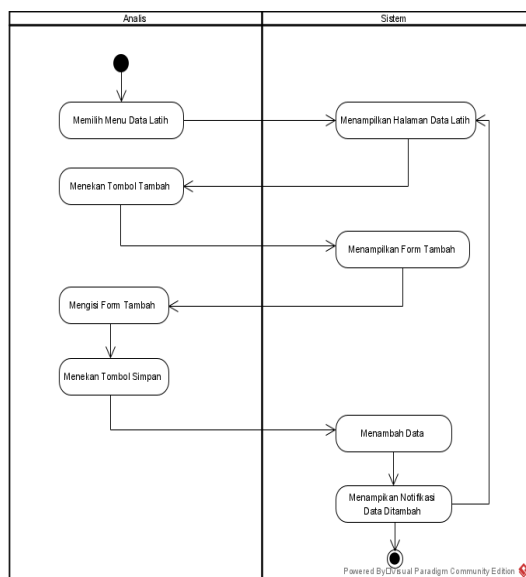
1) Use Case Diagram

Untuk menggambarkan interaksi antara pengguna aplikasi yaitu admin dan analis pemeriksa proposal ditunjukkan pada *use case diagram*. *Use case* ini mengikuti tahapan perhitungan algoritma kNN di mana terdapat fasilitas untuk memasukkan data latih maupun data uji di mana pada algoritma kNN data uji merupakan data baru yang akan ditentukan kelasnya. *Use case diagram* ditunjukkan pada gambar 5.

2) Swimlane Diagram

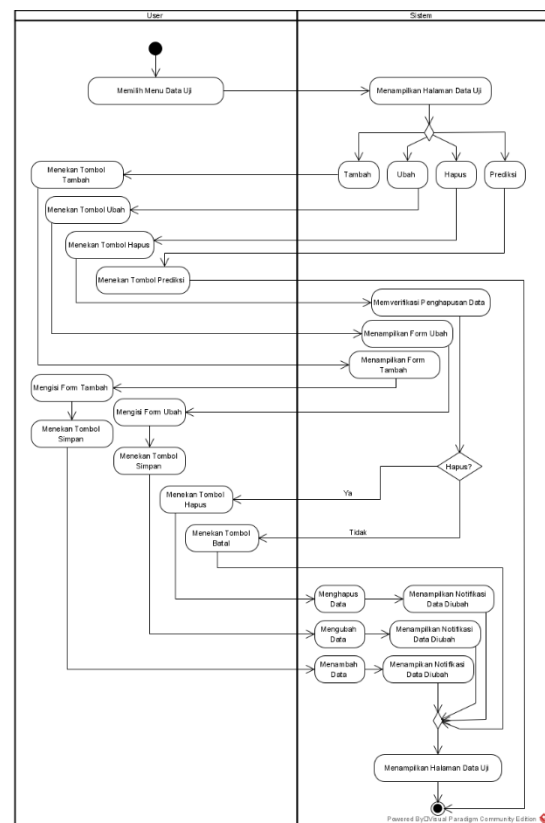
Swimlane diagram menggambarkan aktivitas yang dilakukan oleh admin dan analis saat berinteraksi dengan aplikasi *data mining* ini. Diagram ini berdasarkan dari *use case diagram* yang dibuat sebelumnya.

Swimlane diagram Data latih menunjukkan aktivitas admin memasukkan data proposal yang telah masuk yang telah diketahui status penerimaannya. *Swimlane diagram* ini ditunjukkan gambar 6.

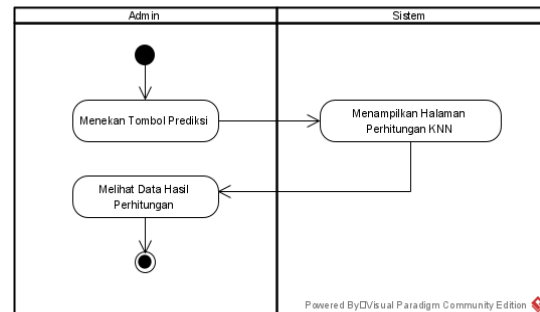


Gambar 6. *Swimlane diagram* input data latih

Swimlane Diagram Data Uji menunjukkan aktivitas analis untuk melakukan proses pada data uji yang dilakukan di dalam sistem. Data uji pada algoritma ini merupakan data baru yang akan ditentukan kelasnya. *Swimlane diagram* uji ditunjukkan pada gambar 7.



Gambar 7. *Swimlane diagram* input data uji

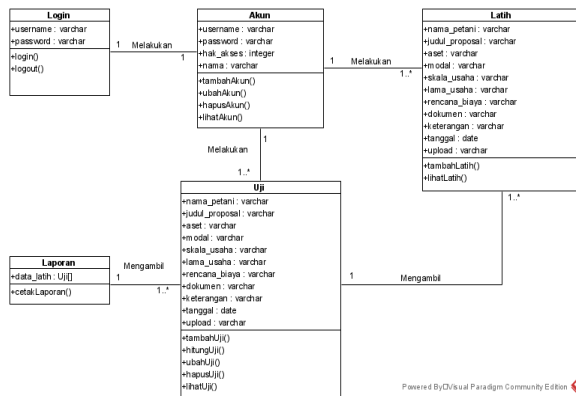


Gambar 8. *Swimlane diagram* perhitungan kNN

Proses perhitungan dan hasil perhitungan algoritma kNN yang dilakukan oleh admin dapat dilihat pada *swimlane diagram* perhitungan kNN pada gambar 8 .

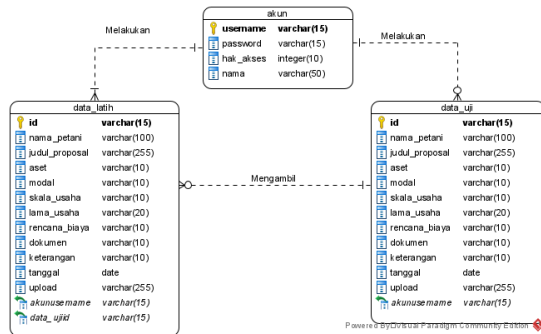
3) Class Diagram

Pada *class diagram* digambarkan kelas-kelas yang berisi atribut dan operasi yang akan digunakan dalam pembuatan aplikasi *data mining* ini. *Class diagram* juga dapat menggambarkan struktur dari pemrograman yang akan dibangun untuk aplikasi. *Class diagram* ditunjukkan pada gambar 9.



Gambar 9. Class diagram

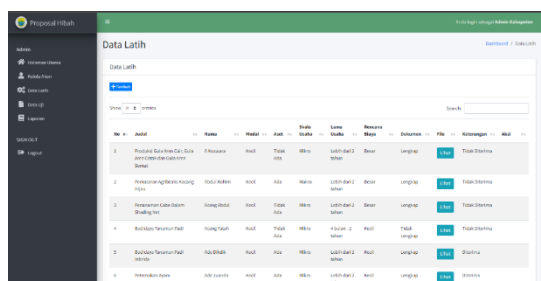
4) Entitiy Relationship Diagram
Entitiy relationship diagram (ERD) menggambarkan hubungan antar entitas atau tabel pada database aplikasi data mining yang akan dibangun. Gambar 10 menunjukkan gambar ERD.



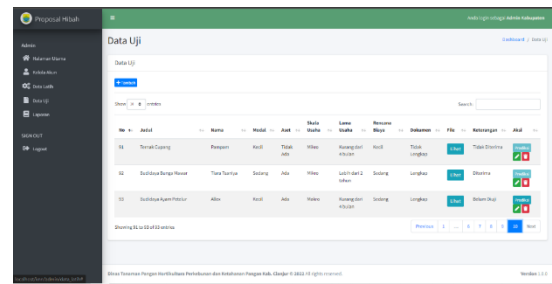
Gambar 10. ERD

C. Pembuatan Aplikasi

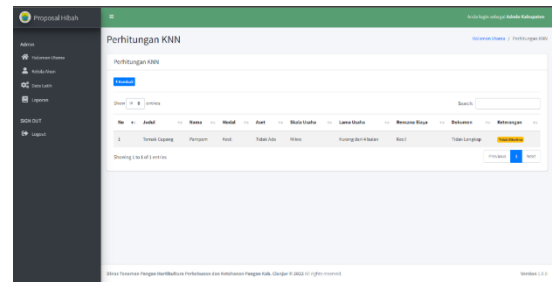
Aplikasi data mining dengan algoritma kNN dibuat setelah proses pemodelan dan perancangan. Hasil aplikasi dapat dilihat pada gambar-gambar berikut. Di mana gambar 11 adalah halaman data latih di mana analis memasukkan data-data proposal hibah. Gambar 12 menunjukkan halaman data uji pada aplikasi sedangkan gambar 13 menunjukkan halaman hasil perhitungan kNN.



Gambar 11. Halaman data latih pada aplikasi



Gambar 12. Halaman data uji pada aplikasi



Gambar 13. Halaman hasil perhitungan kNN pada aplikasi

IV. PEMBAHASAN

Untuk mengetahui performansi dari model yang digunakan dilakukan pengujian terhadap data yang sama. Hasil yang diperoleh dari hasil aplikasi sama dengan perhitungan secara manual. Dari hasil perhitungan kNN untuk 150 data proposal sebagai data latih, diperoleh *confusion matrix* sebagai berikut :

TABEL X
CONFUSION MATRIX PERHITUNGAN

Aktual	Prediksi	
	Ya	Tidak
Ya	105	13
Tidak	23	9
Jumlah	128	22

Dari tabel *confusion matrix* dapat dilihat bahwa :

- Jumlah data proposal sebenarnya yang DITERIMA dengan jumlah data proposal yang diprediksi DITERIMA sebanyak 105 data.
- Jumlah data proposal sebenarnya yang DITERIMA dengan jumlah data proposal yang diprediksi TIDAK DITERIMA sebanyak 13 data.
- Jumlah data proposal sebenarnya yang TIDAK DITERIMA dengan jumlah data proposal yang diprediksi DITERIMA sebanyak 23 data.
- Jumlah data proposal sebenarnya yang TIDAK DITERIMA dengan jumlah data proposal yang diprediksi TIDAK DITERIMA sebanyak 9 data.

Selanjutnya adalah menghitung nilai akurasi dan nilai *error* dari algoritma kNN yang digunakan.

$$\begin{aligned} \text{Akurasi} &= \frac{TP + TN}{TP + FP + TN + FN} \\ &= \frac{105 + 9}{105 + 23 + 13 + 9} = 0,76 \end{aligned}$$

Akurasi merupakan persentase jumlah data uji yang diklasifikasikan dengan benar. Dari perhitungan nilai akurasi yaitu 0,76 atau 76%, ini berarti sebanyak 76% data uji diklasifikasikan dengan benar oleh algoritma kNN.

$$\begin{aligned} \text{Error} &= \frac{FP + FN}{TP + FP + TN + FN} \\ &= \frac{23 + 13}{105 + 23 + 13 + 9} = 0,24 \end{aligned}$$

Error adalah tingkat kesalahan dari klasifikasi yaitu banyaknya data yang diklasifikasikan salah oleh algoritma kNN. Dari perhitungan, tingkat *error* sebesar 0,24 atau 24% data salah diklasifikasikan oleh algoritma kNN. Beberapa cara untuk dapat mengurangi nilai *error* pada algoritma kNN yaitu dengan mencari nilai *k* atau kedekatan optimum. Selain itu dapat dilakukan split data antara data uji dengan data latih sehingga dapat diperoleh hasil yang lebih akurat.

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka terdapat beberapa kesimpulan bahwa, penentuan penerimaan proposal dengan menerapkan metode *data mining* dapat dilakukan lebih cepat, mudah dan akurat karena mengambil pola penerimaan dari data-data sebelumnya. Dari hasil evaluasi terhadap algoritma kNN diperoleh nilai akurasi yang cukup tinggi yaitu 76% dan tingkat kesalahan 24%. Untuk memperoleh akurasi yang lebih tinggi dan tingkat kesalahan yang lebih rendah dapat digunakan algoritma klasifikasi lainnya sebagai bahan pembandingan. Variabel-variabel *prediktor* lainnya dapat dicari dengan teknik seperti analisis faktor sehingga dapat ditemukan variabel-variabel yang paling berpengaruh terhadap keputusan penerimaan proposal.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Teknik Universitas Suryakencana atas dukungan dana hibah penelitian yang diberikan

dan pihak-pihak lain untuk saran dan masukannya selama proses penelitian.

REFERENSI

- [1] J. Han, Jiawei; Kamber, Micheline; Pei, *Data Mining Concepts and Techniques*, Thrid. UK, 2012.
- [2] D. T. Larose and C. D. Larose, *Data Mining and Predictive Analytics*. New Jersey: Jhon Wiley & Sons, 2015.
- [3] C. C. Aggarwal, *Data Mining : The Textbook*. London: Springer, 2015.
- [4] J. Suntoro, *Data Mining : Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: Elex Media Komputindo.
- [5] S. Susanto and D. Suryadi, *Pengantar Data Mining : menggali pengetahuan dari bongkahan data*. Yogyakarta: Andi, 2010.
- [6] J. Han, Jiawei; Kamber, Micheline; Pei, *Data Mining Concepts and Techniques*, Thrid. UK, 2012.
- [7] N. Satyanarayana, C. Ramalingaswamy, and Y. Ramadevi, "Survey of Classification Techniques in Data Mining," *IJISET -International Journal of Innovative Science, Engineering & Technology*, vol. 1, no. 9, pp. 268–278, 2014.
- [8] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," *2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019*, no. January 2021, pp. 1255–1260, 2019.
- [9] M. Suyal and P. Goyal, "A Review on Analysis of K-Nearest Neighbor Classification Machine Learning Algorithms based on Supervised Learning," *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 43–48, 2022.
- [10] J. Sun, W. Du, and N. Shi, "A Survey of kNN Algorithm," *Information Engineering and Applied Computing*, vol. 1, no. 1, pp. 1–10, 2018.
- [11] S. B. Imandoust and M. Bolandraftar, "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events : Theoretical Background," *Int. Journal of Engineering Research and Applications*, vol. 3, no. 5, pp. 605–610, 2013.
- [12] S. Widaningsih and A. Suheri, "Klasifikasi Jurnal Ilmu Komputer Berdasarkan Pembagian," *Seminar Nasional Teknologi Informasi dan Komunikasi 2018 (SENTIKA 2018)*, vol. 2018, no. Sentika, pp. 320–328, 2018.
- [13] O. S. Y. Prakasa and K. M. Lhaksmana, "Klasifikasi Teks Dengan Menggunakan Algoritma K-nearest Neighbor Pada Kasus Kinerja Pemerintah Di Twitter," *eProceedings of Engineering*, vol. 5, no. 3, pp. 8237–8248, 2018.
- [14] S. Widaningsih, "PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI NILAI DAN WAKTU KELULUSAN MAHASISWA PRODI TEKNIK INFORMATIKA," vol. 13, no. 1, pp. 16–25, 2019.
- [15] T. Nasution, "Implementasi Algoritma K-Nearest Neighbor Untuk Penentuan Kelulusan Mahasiswa Tepat Waktu," *Jurnal Perangkat Lunak*, vol. 2, no. 1, pp. 1–14, 2020.
- [16] S. Widaningsih, "Penerapan Data Mining untuk Memprediksi Siswa Berprestasi dengan Menggunakan Algoritma K Nearest Neighbor," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 3, pp. 2598–2611, 2022.
- [17] L. Qoiriyah, H. L. Purwanto, and W. Setyaningsih, "Rancang Bangun Sistem Pendukung Keputusan Penentuan Jenis Beasiswa Menggunakan Knn," *Jurnal Terapan Sains & Teknologi Fakultas*, vol. 1, no. 2, pp. 64–72, 2019.

- [18] W. Yusuf, R. Witri, and C. Juliane, "Model Prediksi Penjualan Jenis Produk Tekstil Menggunakan Algoritma KNN," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 7, no. 1, pp. 1–6, 2022.
- [19] L. Farokhah, "Implementasi K-Nearest Neighbor untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna RGB," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 6, p. 1129, 2020.
- [20] M. Bramer, *Principles of Data Mining*, Third Edit. London: Springer, 2016.