

Algoritma K-Nearest Neighbor Untuk Peramalan Kelulusan Mahasiswa

Febri Andika Putra^{1*}, Muhammad Harits Pratama²

¹Universitas Citra Mandiri Padangsidempuan, Padangsidempuan, Indonesia

² Institut Teknologi Mitra Gama, Riau, Inodensia

Email: ^{1*}Febriandika9725@gmail.com, ² haritsp3@gmail.com

Email Penulis Korespondensi: ¹ Febriandika9725@gmail.com

Abstrak– Universitas Adzkia salah satu kampus swasta di Kota Padang dengan jumlah mahasiswa mencapai ribuan dan terus meningkat setiap tahunnya, manajemen kampus masih belum sepenuhnya memahami manfaat dari data akademik yang terus bertambah. Kondisi ini menunjukkan bahwa masih ada potensi yang belum tereksplorasi secara maksimal dari data akademik mahasiswa yang ada. Untuk menangani hal tersebut, maka dilakukan peramalan data mahasiswa untuk menentukan kelulusan. Prediksi yang dilakukan bertujuan untuk mendapatkan informasi dan pengetahuan dengan menggunakan Data Mining dengan teknik klasifikasi menggunakan algoritma K-Nearest Neighbor. Pada algoritma K-Nearest Neighbor dengan menghasilkan 2 klasifikasi yaitu lulus tepat waktu dan tidak tepat waktu. Berdasarkan pengujian menggunakan algoritma K-Nearest Neighbor dengan menggunakan 55 data mahasiswa yang sudah dibersihkan dan atribut yang digunakan adalah nilai indeks prestasi (IP) semester 1 sampai semester 6 dan total satuan kredit semester (SKS) lalu memberikan nilai $k=3$, menghasilkan klasifikasi “Tidak Tepat Waktu” untuk data ujinya. Hasil pengukuran accuracy algoritma K-Nearest Neighbor dalam memprediksi kelulusan mahasiswa Universitas Adzkia Angkatan 2016 sebesar 94,55%. Dengan demikian, algoritma K-Nearest Neighbor dapat diterapkan untuk memprediksi kelulusan mahasiswa dengan baik.

Kata Kunci: Data Mining, K-Nearest Neighbor, Peramalan Kelulusan, Mahasiswa.

Abstract– Adzkia University is one of the private campuses in Padang City with thousands of students and continues to increase every year, campus management still does not fully understand the benefits of increasing academic data. This condition shows that there is still potential that has not been maximally explored from the existing student academic data. To deal with this, student data forecasting is carried out to determine graduation. The predictions made aim to obtain information and knowledge by using Data Mining with classification techniques using the K-Nearest Neighbor algorithm. In the K-Nearest Neighbor algorithm, it produces 2 classifications, namely graduating on time and not on time. Based on testing using the K-Nearest Neighbor algorithm using 55 cleaned student data and the attributes used are the achievement index (IP) values from semester 1 to semester 6 and total semester credit units (SKS) then giving a value of $k = 3$, resulting in a classification of "Not On Time" for the test data. The results of measuring the accuracy of the K-Nearest Neighbor algorithm in predicting the graduation of Adzkia University students of the 2016 Class were 94.55%. Thus, the K-Nearest Neighbor algorithm can be applied to predict student graduation well.

Keywords: Data Mining, K-Nearest Neighbor, Graduation Forecasting, Students

1. PENDAHULUAN

Pertumbuhan data akademik mahasiswa menciptakan pola-pola yang beragam pada atribut data mereka. Potensi ini dapat dieksplorasi untuk memperoleh informasi berharga salah satu teknik yang digunakan adalah data mining untuk meramalkan keputusan berdasarkan pola yang terbentuk dari data yang ada. Data mining merupakan proses pengolahan data yang melibatkan penggalian informasi khusus menggunakan metode dan algoritma yang ditentukan [1]. Terdapat berbagai metode untuk menentukan apakah setiap mahasiswa dapat menyelesaikan studi mereka, entah itu sesuai dengan jadwal dan tepat waktu, atau terlambzyat, bahkan melewati batas waktu yang ditetapkan untuk setiap mahasiswa [2].

Universitas Adzkia salah satu kampus swasta di Kota Padang dengan jumlah mahasiswa mencapai ribuan dan terus meningkat setiap tahunnya, manajemen kampus masih belum sepenuhnya memahami manfaat dari data akademik yang terus bertambah. Data yang menumpuk tidaklah terlalu memiliki kegunaan data tersebut hanya diperlukan saat universitas membutuhkan suatu informasi tertentu atau saat proses akreditasi dilakukan [3]. Kondisi ini menunjukkan bahwa masih ada potensi yang belum tereksplorasi secara maksimal dari data akademik mahasiswa yang ada. Peramalan (Forecasting) adalah suatu teknik atau seni yang dipakai untuk memprediksi atau memperkirakan kejadian atau peristiwa dimasa yang akan datang atau yang akan terjadi, dan juga peramalan merupakan suatu ilmu untuk memprediksi suatu peristiwa yang belum terjadi atau yang akan terjadi dimasa yang akan datang [4]. Penelitian terdahulu tentang peramalan telah banyak dilakukan diantaranya [5] untuk menguji kinerja algoritma KNN dalam meramalkan penyakit stroke yaitu $k=9$, dimana didapatkan nilai akurasi sebesar 95% [6] Adapun penelitian dengan menggunakan 134 data testing yang diuji nilai akurasi yang didapatkan mencapai 88,06% [7] Metode KNN dengan menggunakan 208 data yang dibagi menggunakan persentase split dengan rasio 80%:20% antara data training dan data testing diapat produk diklasifikasikan 6 Tidak laris, dan 35 Laris dengan akurasi 80,4% [8] Pada penelitian yang melakukan perbandingan melalui pengujian degan 180 data, dan diperoleh akurasi terendah adalah 40-50 % dan akurasi tertinggi adalah 90-100 % pada training data ke-30, dengan nilai K yang digunakan default atau $K = 5$.

Prediksi kelulusan juga diteliti dengan metode Algoritma Neural Network dengan judul Prediksi Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika Universitas Banten Jaya Menggunakan Algoritma Neural Network dengan sample tahun 2018 dan 2019 dengan nilai presentasi 98,27%. Dari penelitian ini didapatkan sebuah hasil yang dapat meningkatkan jumlah dari parameter node input layer penelitian selanjutnya [9]. Penelitian yang berjudul Analisis Tingkat Kepuasan Siswa Dalam Pembelajaran Hybrid Menggunakan Algoritma K – Nearest Neighbor (KNN) memperoleh hasil rata-rata akurasi 86.78%, dari total data pada penelitian ini adalah 128 data dibagi menjadi data latih dan data uji dengan 5 kali percobaan [10]. Penelitian dengan kombinasi algoritma yang judul Implementasi Data Mining Untuk Memprediksi Penyakit Diabetes Menggunakan Algoritma Naives Bayes Dan K-Nearest Neighbor juga mendapatkan akurasi 99% untuk KNN sedangkan Naïve Bayes memiliki hasil accuracy sejumlah 75% [11], Penelitian dengan judul Prediksi Tindak Pidana UU ITE Dengan Metode K-NN Oleh Personel Polres Metro Bekasi Era Police 4.0. memperoleh akurasi yang cukup tinggi yaitu rata-rata sebesar 98,06% dengan 200 data [12]. Data sebanyak 768 dengan delapan parameter yang sudah lengkap dengan nilai parameternya mendapatkan Akurasi 82% untuk 150 data uji dan pada pengujian kedua dengan 200 data didapatkan nilai Akurasi 82% pada penelitian yang berjudul Supervised Learning Metode K-Nearest Neighbor Untuk Prediksi Diabetes Pada Wanita [13], penelitian pada bidang Kesehatan yang berjudul rediksi Penyebaran Kasus Demam Berdarah Dengue (Dbd) Dengan Metode K-Nearest Neighbor Di Kabupaten Maluku Tengah dengan 256 jenis data yang digunakan adalah data jumlah kasus DBD menghasilkan performance terbaik menggunakan nilai K=3 dan nilai accuracy sebesar 75,27%. Secara keseluruhan, penerapan Algoritma KNN dapat memberikan accuracy dan prediksi yang cukup baik [14].

Berpandangan pada penelitian terdahulu penulis tertarik juga melakukan peralaman data mahasiswa dengan metode yang berbeda yaitu metode K-Nearest Neighbor untuk menentukan kelulusan tepat waktu atau tidak, Algoritma KNN ini pada umumnya digunakan dalam mengklasifikasikan objek berdasarkan pada data pembelajaran yang memiliki nilai selisih kecil dan jarak tetangga terdekat dengan objek [5]. K-NN memiliki beberapa kelebihan, antara lain, memiliki ketahanan terhadap noise pada data latih, tidak memerlukan proses pra-pemrosesan data yang kompleks, mudah diimplementasikan dan dipahami, serta efektif ketika digunakan dengan data latih yang berukuran besar [15]

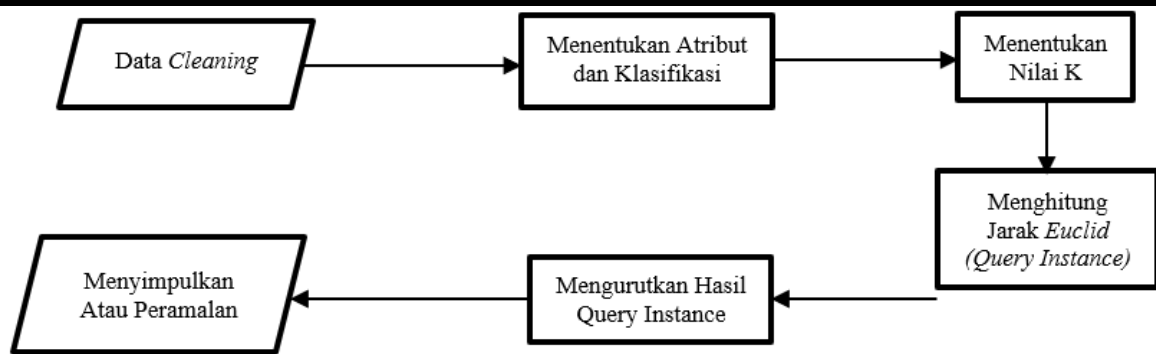
Metode ini mudah diimplementasikan dan cukup efektif untuk klasifikasi data dengan jumlah sampel yang kecil hingga menengah [16]. Dalam metode KNN, langkah awalnya adalah menentukan nilai K, misalnya jika k=3, berarti kita akan mengambil 3 dokumen terdekat dengan dokumen uji. Kelas yang paling banyak muncul di antara dokumen-dokumen terdekat ini akan menjadi kelas di mana data evaluasi tersebut berada [17]. Misalkan untuk mencari waktu kelulusan mahasiswa semester 6, digunakan data dari mahasiswa yang sudah lulus. Kedekatan kasus mahasiswa yang akan lulus dihitung dengan semua kasus mahasiswa yang sudah lulus. Solusi untuk kasus mahasiswa yang akan lulus diambil dari kedekatan terbesar antara kasus tersebut dengan kasus yang sudah lulus.

Algoritma K-Nearest Neighbor merupakan algoritma yang digunakan dalam dunia Machine Learning [18] Dalam penghitungan nilai KNN digunakan rumus perhitungan jarak (euclidean), K-Nearest Neighbor adalah suatu metode pembelajaran untuk memecahkan kasus dengan menghitung antara kasus baru dengan kasus lama dengan memperhatikan jarak yang paling dekat [19]. K-Nearest Neighbor merupakan metode klasifikasi Data Mining dimana KNN mengklasifikasikan suatu dataset berdasarkan pembelajaran dari data yang diklasifikasikan atau diberi label [20].

Berdasarkan referensi yang dijadikan penulis sebagai referensi dalam menjalankan penelitian ini maka penelitian dengan menggunakan metode KNN merupakan salah satu metode yang tepat dalam mengeksplorasi data kelulusan mahasiswa untuk peramalan kelulusan. Data yang digunakan dalam penelitian ini adalah data mahasiswa Universitas Adzkie angkatan 2016 yang mencakup atribut seperti indeks prestasi mahasiswa semester 1-6, jumlah sks hingga semester 6, dan keterangan lulus. Metode k-nearest neighbor digunakan karena metode ini tangguh terhadap training data yang noise, tidak memerlukan pre-processing data yang rumit, mudah diimplementasikan dan dipahami dan efektif apabila data latihnya besar, secara spesifik penelitian ini juga berfokus pada perhitungan menggunakan euclidean untuk menentukan pola data, euclidean merupakan perhitungan jarak terdekat yang mengambil nilai mayoritas untuk menentukan pola dari sebuah data.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan proses k-nearest neighbor dengan tujuan untuk mendapatkan hasil yang optimal dalam melakukan peramalan kelulusan mahasiswa. Proses penelitian ini melakukan analisis terhadap data mahasiswa yang diperoleh dari bagian akademik Universitas Adzkie Angkatan 2016, Adapun tahapan penelitian dapat disajikan pada Gambar.1



Gambar 1. Tahap tahap penelitian

2.1 Data Cleaning

Penerapan metode K-Nearest Neighbor memerlukan data yang bersih maka data yang telah didapat akan dilakukan cleaning atau pembersihan data. Data pada tabel 4.1 peneliti akan menghilangkan data akademik mahasiswa yang Drop Out dan data yang tidak lengkap karena akan berpengaruh terhadap akurasi dari hasil peramalan dengan metode K-Nearest Neighbor. Berikut hasil data cleaning dapat dilihat pada Tabel 1.

Tabel 1. Data Cleaning Nilai Akademik Mahasiswa Angkatan 2016

No	NIM	Nama	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5	Sem 6	Total SKS	Masa Studi
1	1620001	Ayu Putri	3.8	3.81	3.45	3.83	3.5	3.75	129	4 Tahun, 0 Bulan
2	1620002	Antika Yolanda Putri	3.7	3.9	3.59	3.86	3.92	3.65	126	4 Tahun, 0 Bulan
3	1620008	Ayu Azarita	3.5	3.14	3.55	3.77	3.38	3.47	126	4 Tahun, 3 Bulan
4	1620015	Andre Oklanda Gessa	2.95	3.14	2.95	3.42	2.9	3.05	125	4 Tahun, 3 Bulan
5	1620017	Dio Firrahardi	0.1	0	2.55	3	2.65	1.52	123	5 Tahun, 4 Bulan
---	---	---	---	---	---	---	---	---	---	---
51	1620232	Adhya Muqarramah Rahayu	3.7	4	3.79	3.41	3.54	3.76	128	4 Tahun, 0 Bulan
52	1620234	Diki Wahyudi. S	3	3	3.23	2.87	2.71	3.1	130	4 Tahun, 1 Bulan
53	1620241	Ayu Widia Agustin	3.65	3.24	3.23	2.96	2.75	3.42	135	4 Tahun, 7 Bulan
54	1620243	Amrizal Efendi	3.6	3.1	3.32	1.83	2.79	2.44	129	4 Tahun, 0 Bulan
55	1620245	Ermita Susandri	3.8	3.52	3.71	3.62	3.58	3.67	131	4 Tahun, 3 Bulan

2.2 Menentukan Atribut Data dan Klasifikasi

Atribut pada analisa metode K-Nearest Neighbor ini dapat dilihat pada Tabel 2 dan untuk atribut masa studi akan diklasifikasikan menjadi 2 kategori yang bisa dilihat pada Tabel 3.

Tabel 2 Tabel Atribut Data Mahasiswa

Atribut	Keterangan
NIM	Nomor Induk Mahasiswa
Nama	Nama Mahasiswa
Sem	Indeks Prestasi dari semester 1-6 yang akan digunakan untuk proses <i>k-nearest neighbor</i>
Total SKS	Total SKS sampai semester 6 yang akan digunakan untuk proses <i>k-nearest neighbor</i>
Masa Studi	Keterangan untuk Klasifikasi

Tabel 3 Klasifikasi Kelulusan Berdasarkan Masa Studi

Lulus	Klasifikasi
Lulus < 5 Tahun	Tepat Waktu
Lulus > 5 Tahun	Tidak Tepat Waktu

Data yang telah di cleaning dan selanjutnya akan di klasifikasikan berdasarkan masa studi dapat dilihat pada Tabel 3 diatas. Hasil dari klasifikasi data dapat dilihat pada Tabel 4 dan data testing yang akan diuji pada Tabel 4.

Tabel 4 Data Training Lengkap

No	NIM	Nama	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5	Sem 6	Total SKS	Masa Studi	Klasifikasi
1	1620001	Ayu Putri	3.8	3.81	3.45	3.83	3.5	3.75	129	4 tahun, 0 bulan	Tepat Waktu
2	1620002	Antika Yolanda Putri	3.7	3.9	3.59	3.86	3.92	3.65	126	4 tahun, 0 bulan	Tepat Waktu

3	1620008	Ayu Azarita	3.5	3.14	3.55	3.77	3.38	3.47	126	4 tahun, 3 bulan	Tepat Waktu
4	1620015	Andre Oklanda Gessa	2.95	3.14	2.95	3.42	2.9	3.05	125	4 tahun, 3 bulan	Tepat Waktu
5	1620017	Dio Firrahardi	0.1	0	2.55	3	2.65	1.52	123	5 tahun, 4 bulan	Tidak Tepat Waktu
---	---	---	---	---	---	---	---	---	---	---	---
51	1620232	Adhya Muqarramah Rahayu	3.7	4	3.79	3.41	3.54	3.76	128	4 tahun, 0 bulan	Tepat Waktu
52	1620234	Diki Wahyudi. S	3	3	3.23	2.87	2.71	3.1	130	4 tahun, 1 bulan	Tepat Waktu
53	1620241	Ayu Widia Agustin	3.65	3.24	3.23	2.96	2.75	3.42	135	4 tahun, 7 bulan	Tepat Waktu
54	1620243	Amrizal Efendi	3.6	3.1	3.32	1.83	2.79	2.44	129	4 tahun	Tepat Waktu
55	1620245	Ermita Susandri	3.8	3.52	3.71	3.62	3.58	3.67	131	4 tahun, 3 bulan	Tepat Waktu

Tabel 5 Data Uji

No	NIM	Nama	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5	Sem 6	Total SKS	Y = Klasifikasi
1	2010010	Miza Yulia Nofika	3.25	3.65	3.51	3.3	3.63	3.76	118	?

2.3 Menentukan Nilai K

Dalam metode K-Nearest Neighbors (KNN), penentuan nilai parameter k didasarkan pada pertimbangan yang berkaitan dengan kebutuhan khusus, yang diarahkan untuk memperoleh hasil peramalan yang optimal. Pada konteks peramalan status kelulusan, pendekatan ini dilaksanakan dengan memilih mayoritas label tetangga terdekat. Dalam penelitian ini nilai k yang digunakan adalah k=3.

2.4 Menghitung Jarak Euclid (Query Instance)

Metode K-Nearest Neighbor merupakan metode untuk mencari jarak terdekat antara data training dengan data testing. Untuk menghitung jarak terdekat dapat menggunakan rumus euclidean distance, Rumus untuk menghitung Jarak Euclid ini disajikan pada Persamaan (1).

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

Dimana: x merupakan sampel data, y adalah data uji. i adalah variabel data dan d adalah jarak.

Query instance data training no=1

$$\begin{aligned}
 &= \sqrt{(3,8 - 3,25)^2 + (3,81 - 3,65)^2 + (3,45 - 3,51)^2 + (3,83 - 3,3)^2 +} \\
 &\quad \sqrt{(3,5 - 3,63)^2 + (3,75 - 3,76)^2 + (129 - 118)^2} \\
 &= \sqrt{(0,55)^2 + (0,16)^2 + (-0,06)^2 + (0,53)^2 + (-0,13)^2 + (-0,01)^2 + (11)^2} \\
 &= \sqrt{0,3025 + 0,0256 + 0,0036 + 0,2809 + 0,0169 + 0,0001 + 121} \\
 &= \sqrt{121,63} = 11,038676
 \end{aligned}$$

Query instance data training no=2

$$\begin{aligned}
 &= \sqrt{(3,7 - 3,25)^2 + (3,9 - 3,65)^2 + (3,59 - 3,51)^2 + (3,86 - 3,3)^2 +} \\
 &\quad \sqrt{(3,92 - 3,63)^2 + (3,65 - 3,76)^2 + (126 - 118)^2} \\
 &= \sqrt{(0,45)^2 + (0,25)^2 + (0,08)^2 + (0,56)^2 + (0,29)^2 + (-0,11)^2 + (8)^2} \\
 &= \sqrt{0,2025 + 0,0625 + 0,0064 + 0,3136 + 0,0841 + 0,0121 + 64} \\
 &= \sqrt{64,6813} = 8,04246231
 \end{aligned}$$

Demikian seterusnya perhitungan dilakukan sampai data training ke – 55 yang dapat dilihat pada Tabel 6.

Tabel 6 Hasil Query Instance

No.	NIM	Nama	Query Instance
1	1620001	Ayu Putri	11.02858105
2	1620002	Antika Yolanda Putri	8.04246231
3	1620008	Ayu Azarita	8.043115068
4	1620015	Andre Oklanda Gessa	7.121453503
5	1620017	Dio Firrahardi	7.431998385
---	---	---	---
51	1620232	Adhya Muqarramah Rahayu	10.02115762
52	1620234	Diki Wahyudi. S	12.08429973
53	1620241	Ayu Widia Agustin	17.04148174
54	1620243	Amrizal Efendi	11.22808978

2.5 Mengurutkan Hasil Query Instance

Setelah Query Instance didapatkan dari masing-masing data, langkah selanjutnya adalah mengurutkan data-data tersebut kedalam kelompok yang mempunyai jarak euclid terkecil hingga terbesar yang dapat dilihat pada Tabel 7.

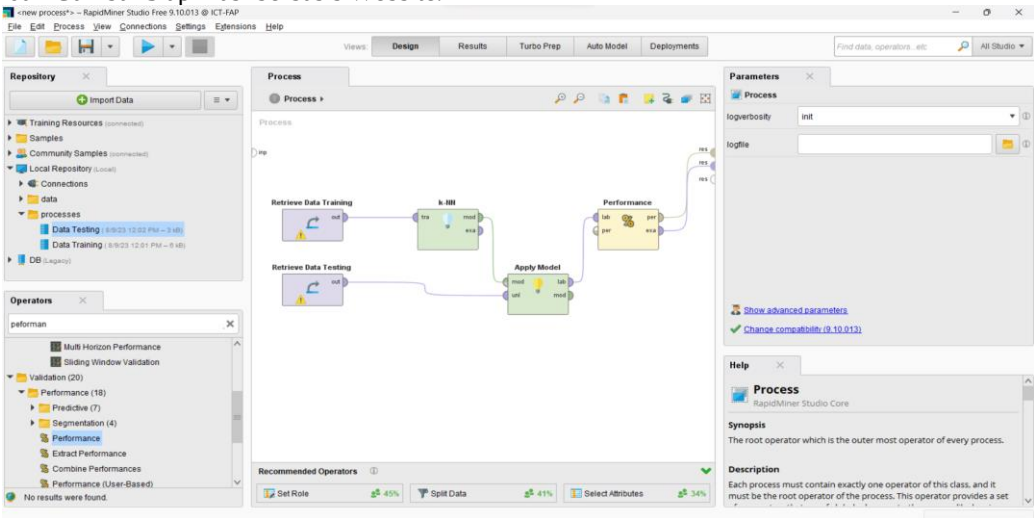
Tabel 7 Hasil Query Instance yang telah diurutkan

No.	NIM	Nama	Query Instance
1	1620202	Sri Tini	4.024909937
2	1620127	Septi Eka Nova Wahyuni	4.06123134
3	1620165	Dia Silvila	4.433711763
4	1620029	Ella Oliffvia	5.062163569
5	1620148	Sari Rahma Dewi	6.05389957
---	---	---	---
51	1620131	Nalpi Rahma Yirlanda	13.06182223
52	1620173	Echa Martha Yasmita	13.10046564
53	1620175	Imelda Pebriza	14.17354225
54	1620090	Melly Afindo Liska Syuria	16.0250741

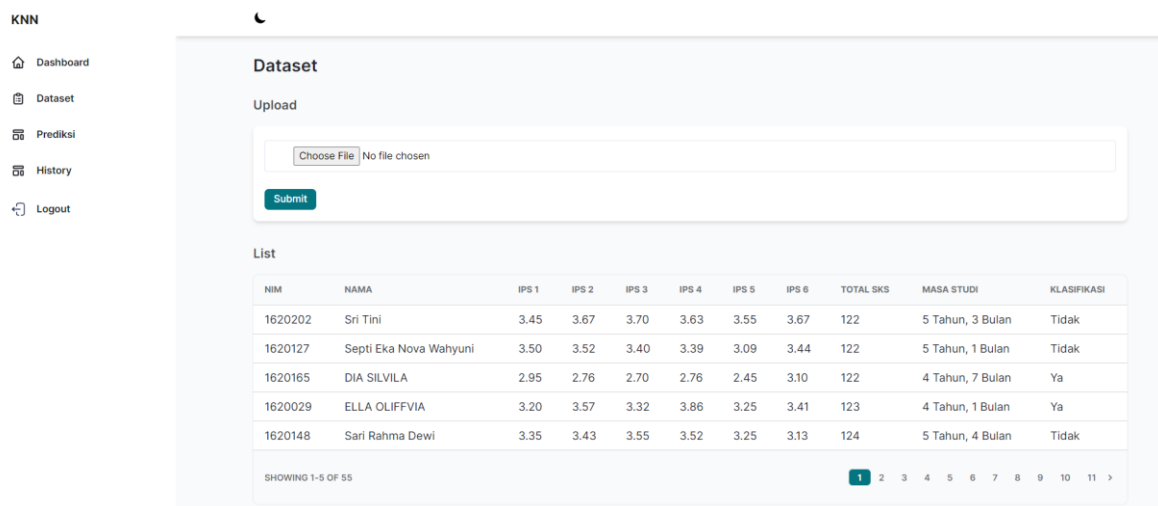
3. HASIL DAN PEMBAHASAN

3.1 Analisa Proses Algoritma

Data Training yang diimplementasikan menggunakan aplikasi RapidMiner Studio 9.10 dan aplikasi berbasis Website sebagai masukan untuk memberikan peengetahuan, dapat dilihat pada Gambar 2 aplikasi RapidMiner Studio 9.10 dan Gambar 3 aplikasi berbasis Website.



Gambar 2. Tampilan Halaman Proses Drag Model K-Nearest Neighbor, Apply Model dan Performance Rapid Miner 9.10.



Gambar 3. Tampilan Halaman Data Training Aplikasi Berbasis Website

Ketika user memasukan data uji pada aplikasi, maka proses pada sistem adalah menghitung untuk mendapatkan nilai Query Instance dan menarik kesimpulan dari klasifikasi yang terbanyak berdasarkan nilai K. Penghitungan atau rumus berdasar ketentuan algoritma KNN sehingga menghasilkan peramalan.

3.2 Pengambilan data berdasarkan Nilai K Berisi hasil implementasi ataupun pengujian.

Setelah Hasil Query Instance diurutkan, maka diambil nilai jarak yang terkecil dengan ketentuan $k = 3$ yang dapat dilihat pada Tabel 8

Tabel 8 Query Instance yang diurutkan dengan $k = 3$

No	NIM	Nama	Query Instance	Masa Studi	Y = Klasifikasi
1	1620202	Sri Tini	4.024909937	5 Tahun, 3 Bulan	Tidak Tepat Waktu
2	1620127	Septi Eka Nova Wahyuni	4.06123134	5 Tahun, 11 Bulan	Tidak Tepat Waktu
3	1620165	Dia Silvila	4.433711763	4 Tahun, 7 Bulan	Tepat Waktu

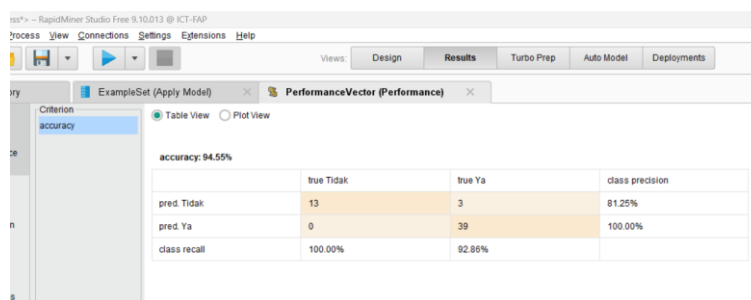
Berdasarkan kategori Nearest Neighbor yang paling mayoritas maka dapat meramalkan nilai query instance yang telah dihitung. Pada urutan jarak yang terdekat dari satu sampai tiga (nilai $k=3$), maka berdasarkan tabel 3 untuk klasifikasi kelulusan, diketahui ada 2 yang lulus > 5 tahun dan 1 yang lulus < 5 tahun pada Tabel 9.

3.3 Hasil Peramalan

Hasil peramalan dari metode K-Nearest Neighbor dengan nilai $K = 3$ pada Tabel 9 terhadap data testing yang menghasilkan peramalan bahwa mahasiswa bernama Miza Yulia Nofika akan lulus diatas 5 tahun, untuk menguji akurasi peramalan akan digunakan aplikasi RapidMiner Studio 9.10 hasilnya dapat dilihat pada Gambar 4.

Tabel 9 Hasil Data Testing

No	N I M	Nama Mahasiswa	Sem 1	Sem 2	Sem 3	Sem 4	Sem 5	Sem 6	Total SKS	Y=klasifikasi
1	2010010	Miza Yulia Nofika	3.25	3.65	3.51	3.3	3.63	3.76	118	Tidak Tepat Waktu



Gambar 4 Hasil Peramalan akurasi kelulusan dengan RapidMiner Studio 9.10

Untuk mengetahui akurasi atau tingkat keberhasilan pada sistem ini maka digunakan rumus di bawah ini: Rumus untuk menghitung akurasi ini disajikan pada Persamaan (2).

Akurasi =
$$\frac{\sum \text{hasil pengujian bernilai benar (kecocokan)}}{\sum \text{Banyak Data Training}} \times 100\% \tag{2}$$

Tabel 10. Hasil Akurasi Peramalan K-Nearest Neighbor

	Peramalan (Tepat Waktu)	Peramalan (Tidak Tepat Waktu)	Peramalan (Jumlah)
Total Data	39	16	55
Cocok	39	13	52
Tidak Cocok	3	0	3

Akurasi =
$$\frac{52}{55} \times 100\% = 94,55\%$$

Dari Tabel 10, dapat disimpulkan bahwa hasil peramalan tingkat kelulusan mahasiswa mencapai tingkat akurasi 94,55% dengan k=3 menggunakan 55 data akademik mahasiswa angkatan 2016.

4. KESIMPULAN

Berdasarkan hasil dan analisis yang telah dilakukan dalam penelitian ini, dapat disimpulkan bahwa dengan menerapkan model data mining menggunakan algoritma K-Nearest Neighbor (KNN) pada data kelulusan mahasiswa Universitas Adzkia yang terdiri dari 55 data dengan parameter k=5, diperoleh tingkat akurasi sebesar 94.55%. Hasil dari proses majority voting pada data uji menunjukkan bahwa mahasiswa tersebut kemungkinan besar tidak akan lulus tepat waktu..

REFERENCES

[1]. R. Sepriansyah, S. D. Purnamasari, K. R. N. Wardani, and N. Halim, “Prediksi Kelulusan Mahasiswa Fakultas Teknik Universitas Bina Darma Menggunakan Algoritma Naïve Bayes,” JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform., vol. 8, no. 1, pp. 313–322, 2023, doi: 10.29100/jipi.v8i1.3459.

[2]. M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, “Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa,” G-Tech J. Teknol. Terap., vol. 7, no. 1, pp. 21–28, 2023, doi: 10.33379/gtech.v7i1.1850.

[3]. R. Hikmah Tanjung, Y. Yunus, and G. Widi Nurcahyo, “Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa,” J. Comput. Sci. Inf. Technol. (CoSciTech), vol. 4, no. 1, pp. 193–199, 2023, doi: <https://doi.org/10.37859/coscitech.v4i1.4755>.

[4]. S. Sutrisman, H. Syafwan, and R. Rohminatin, “Implementation of Trend Moment Method in Forecasting Regional Income,” Build. Informatics, Technol. Sci., vol. 4, no. 2, pp. 749–758, 2022, doi: 10.47065/bits.v4i2.2090

[5]. M. N. Maskuri, K. Sukerti, and R. M. H. Bhakti, “Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke,” vol. 4, no. 1, pp. 130–140, 2022.

[6]. M. Yunus and N. kadek ari pratiwi, “JTIM: Jurnal Teknologi Informasi dan Multimedia Prediksi Status Gizi Balita Dengan Algoritma K-Nearest,” vol. 4, no. 4, pp. 221–231, 2023, doi: 10.35746/jtim.v4i4.328.

[7]. R. Iriane, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Pangan Hewan Menggunakan Metode K-Nearest Neighbor,” vol. 3, no. 5, pp. 509–515, 2023.

[8]. H. Nisya, C. Setianingsih, and W. Harjupa, “Prediksi Curah Hujan Dari Data Satelit Himawari-8 Menggunakan Metode K-Nearest Neighbor (KNN),” vol. 10, no. 1, pp. 729–735, 2023.

[9]. R. Rudianto, R. Kania, and T. I. Solihati, “Prediksi Kelulusan Mahasiswa Teknik Informatika Universitas Banten Jaya Menggunakan Algoritma Neural Network,” J. Sist. Inf. dan Inform., vol. 5, no. 2, pp. 193–200, 2022, doi: 10.47080/simika.v5i2.2123.

[10]. R. I. Sulthoni, L. A. Muharom, and M. Rahman, “Jurnal Smart Teknologi Analisis Tingkat Kepuasan Siswa Dalam Pembelajaran Hybyrid Menggunakan Algoritma K – Nearest Neighbor (KNN) Student Satisfaction Level Analysis In Hybyrid Learning Using K – Nearest Neighbor (KNN) Algorithm Jurnal Smart Technolog,” vol. 4, no. 4, pp. 406–411, 2023

-
- [11]. F. N. Ikhromr et al., “Implementasi data mining untuk memprediksi penyakit diabetes menggunakan algoritma naives bayes dan k-nearest neighbor implementation of data mining to predict diabetes disease using naives bayes and k-nearest neighbor algorithms,” vol. 6, pp. 416–428, 2023.
 - [12]. A. Firmansyah and J. Prianggono, “PREDIKSI TINDAK PIDANA UU ITE DENGAN METODE K-NN OLEH PERSONEL POLRES METRO BEKASI ERA POLICE 4.0,” vol. 2, no. 5, pp. 1240–1248, 2023.
 - [13]. A. P. Silalahi and H. G. Simanullang, “Supervised Learning Metode K-Nearest Neighbor Untuk Prediksi Diabetes Pada Wanita,” METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt., vol. 7, no. 1, pp. 144–149, 2023, doi: 10.46880/jmika.vol7no1.pp144-149.
 - [14]. S. Latuconsina and M. Y. Matdoan, “PREDIKSI PENYEBARAN KASUS DEMAM BERDARAH DENGUE (DBD) DENGAN METODE K-NEAREST NEIGHBOR DI KABUPATEN MALUKU TENGAH,” vol. 19, no. 2, pp. 254–261, 2022.
 - [15]. R. N. Sukmana, Abdurrahman, and Y. Wicaksono, “IMPLEMENTASI K-NEAREST NEIGHBOR UNTUK MENENTUKAN PREDIKSI PENJUALAN (STUDI KASUS : PT MAKSIPLUS UTAMA INDONESIA) Rini,” vol. 8, no. 2, pp. 31–38, 2020.
 - [16]. Rismala, I. Ali, and A. Rizki Rinaldi, “PENERAPAN METODE K-NEAREST NEIGHBOR UNTUK PREDIKSI,” JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 1, pp. 585–590, 2023.
 - [17]. J. Supriyanto, D. Alita, and A. R. Isnain, “Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring,” vol. 4, pp. 74–80, 2023.
 - [18]. M. Faisal, W. S. Utami, and S. Parmica, “Implementasi Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Indeks Kemiskinan,” vol. 09, no. 01, pp. 11–23, 2023.
 - [19]. E. Nasri and A. S. AW, “Aplikasi Seleksi Penentuan Nasabah Untuk Penjualan Barang Secara Kredit Dengan Algoritma K-Nearest Neighbor,” J. Ilm. Sains dan Teknol. Univ. Baten Jaya, vol. 4, no. 1, pp. 1–11, 2020.
 - [20]. R. Fahrur and M. Bagoes Satria Junianto, “Penerapan Machine Learning Untuk Prediksi Harga Saham PT . Telekomunikasi Indonesia Tbk Menggunakan Algoritma,” vol. 1, no. 1, pp. 18–24, 2023.