

Implementasi Algoritma K-Nearest Neighbor Untuk Memprediksi Program Studi Bagi Calon Mahasiswa Baru

Ratna Rahmawati Rahayu^{1*}, Lidiawati²

¹Program Studi Sistem Informasi, STMIK Bani Saleh

²Program Studi Teknik Informatika, STMIK Bani Saleh

*ratnaridw4n@gmail.com

Abstrak

Salah satu faktor mahasiswa lulus tepat waktu dengan nilai yang baik adalah karenanya program studi yang diambilnya sesuai dengan minat dan kompetensi. Untuk itu dalam proses penerimaan mahasiswa baru perlu dilakukan seleksi, informasi dan arahan mengenai program studi yang dipilih. Dengan menggunakan data mahasiswa tahun sebelumnya dilakukan pengolahan data mining untuk menghasilkan klasifikasi program studi bagi calon mahasiswa baru. Untuk mendapatkan hasil yang maksimal maka dilakukan *data preprocessing*, setelahnya data dibagi menjadi data training dan data testing. Kedua data tersebut selanjutnya diolah dengan algoritma K-Nearest Neighbor untuk menentukan kesesuaian class Program Studi yang ada di data testing lalu dihitung nilai akurasi pengukurannya. Karena memiliki nilai akurasi tinggi yaitu 74%, maka dengan menggunakan data training tersebut dikembangkan dalam bentuk aplikasi dengan Java NetBeans yang dapat dipakai untuk membantu calon mahasiswa baru dalam memprediksi program studi yang sesuai.

Kata kunci: Data Mining, Klasifikasi, K-Nearest Neighbor, Program Studi

Abstract

One of the factors for students graduating on time with good grades is that the study program they take is in accordance with their interests and competencies. For this reason, in the process of admitting new students, it is necessary to carry out selection, information and direction regarding the chosen study program. By using previous year's student data, data mining processing is carried out to produce classifications of study programs for prospective new students. To get maximum results, preprocessing data is carried out, after which the data is divided into training data and testing data. The two data are then processed with the K-Nearest Neighbor algorithm to determine the suitability of the Study Program class in the testing data and then the measurement accuracy value is calculated. Because it has a high accuracy value of 74%, using this training data it is developed in the form of an application with Java NetBeans which can be used to assist prospective new students in predicting the appropriate study program.

Keywords: Data Mining, Classification, K-Nearest Neighbors, Study Program

1 Pendahuluan

Sekolah Tinggi merupakan Perguruan Tinggi yang menyelenggarakan pendidikan akademik dan dapat menyelenggarakan pendidikan vokasi dalam satu rumpun ilmu pengetahuan dan/atau teknologi tertentu. Sedangkan Program Studi

merupakan kesatuan kegiatan pendidikan dan Pembelajaran yang memiliki kurikulum dan metode Pembelajaran tertentu dalam satu jenis pendidikan akademik, pendidikan profesi, dan/atau pendidikan vokasi. STMIK Bani Saleh merupakan salah satu perguruan tinggi dengan

rumpun bidang Ilmu Komputer dan Manajemen Informatika terdiri dari 5 (lima) program studi, yaitu : Teknik Informatika, Teknik Komputer, Sistem Informasi, Manajemen Informatika dan Komputerisasi Akuntansi. Di dalam deskripsi masing-masing program studi tersebut sudah tertuang kompetensi dan luaran yang dapat dicapai bagi lulusan, sehingga pemilihan program studi bagi calon mahasiswa merupakan hal yang sangat mendasar. Karena mahasiswa akan menuntut ilmu sesuai dengan kompetensi program studi dan diharapkan dapat diselesaikan tepat waktu dengan nilai yang baik. Pemilihan program studi yang tidak sesuai dapat menyebabkan seorang mahasiswa berhenti, lulus tidak tepat waktu atau nilai yang kurang bagus.

Pemilihan program studi yang sesuai sangat ditentukan di dalam proses penerimaan mahasiswa baru. Ada beberapa faktor mahasiswa dalam memilih program studi, diantaranya faktor teman, orang tua atau yang penting kuliah dan lain sebagainya tanpa memperhatikan program studi yang benar-benar sesuai dengan keadaan dirinya. Untuk membantu calon mahasiswa dalam memilih program studi tersebut yang telah dilakukan diantaranya dengan melakukan seleksi wawancara, selain tes seleksi yang terdiri dari pengetahuan umum, matematika dan bahasa Inggris. Berdasarkan hasil tes seleksi tersebut

dan riwayat hidup dari calon mahasiswa tersebut, pada saat wawancara dapat diberikan masukan atau pertimbangan antara lain : bagi yang belum atau masih bingung maka dapat diarahkan program studi yang sesuai dengan keadaan dan latar belakangnya, bagi yang sudah memilih tapi dinilai kurang sesuai maka dapat diberi masukan agar tidak salah dalam memilih dan lain sebagainya.

Saat ini sudah di era digital, perkembangan teknologi yang sudah semakin modern, maka untuk memudahkan dalam proses wawancara atau calon mahasiswa dalam memilih program studi dapat dikembangkan sebuah model yang berisikan mengenai pengolahan data yang sudah ada sebelumnya (data training) digunakan untuk memprediksi, data baru (data testing) merupakan data yang akan diprediksi, kemudian kedua data tersebut diolah dengan menggunakan algoritma K-Nearest Neighbor sehingga dari data testing tersebut akan menghasilkan model untuk memprediksi program studi yang sebaiknya dipilih berdasarkan data yang telah ada sebelumnya.

2 Tinjauan Pustaka

2.1 Penelitian Terkait

Dalam penelitian ini menggunakan beberapa referensi penelitian, diantaranya :

1. Penelitian yang dilakukan oleh Muhammad Kurniawan Khamdani, Nurul Hidayat dan

- Ratih Kartika Dewi dengan judul "Implementasi Metode K-Nearest Neighbor Untuk Mendiagnosis Penyakit Tanaman Bawang Merah". Dari penelitian tersebut dapat disimpulkan bahwa sumber data penelitian terhadap penyakit bawang merah diperoleh dari Dinas Pertanian Kabupaten Malang. Jenis penyebab penyakit tanaman bawang merah yang diteliti sebanyak 5 (lima) yaitu : ulat bawang, lalat penggorok, moler, trotol dan antrakosa dengan gejala yang dianalisis sebanyak 15 gejala. Setiap gejala memiliki 5 keluaran yaitu : tidak ada gejala, gejala sedikit tapi bukan gejala utama, gejala banyak tapi bukan gejala utama, gejala sedikit dan merupakan gejala utama, dan gejala banyak dan merupakan gejala utama. Berdasarkan data tersebut dilakukan uji coba terhadap 30 data testing dengan $K = 1, 4, 7$ dan 10 menghasilkan nilai akurasi yang tidak jauh berbeda. Karenanya jumlah K tidak terlalu mempengaruhi nilai akurasi.
2. Penelitian yang dilakukan oleh Yahya dan Winda Puspita Hidayati yang berjudul "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penuaan Vape (Rokok Elektrik) pada Lombol Vape On". Yang dapat disimpulkan pada penelitian tersebut adalah dengan menggunakan data hasil penjualan tahun 2019 dilakukan klasifikasi berdasarkan jumlah item atau jenis barang yang mampu terjual dengan klas mencapai target dan tidak mencapai target. Apabila rata-rata jumlah satu jenis barang penjualan 10 pcs dapat disebut sebagai kategori mampu mencapai target dan apabila kurang dari 10 yang terjual maka akan disebut dengan kategori tidak mampu mencapai target. Pengujian menggunakan nilai $K = 1, 2, 3, 4, 5, 6, 7, 8, 9$ dan 10 . Semakin tinggi jumlah K yang digunakan maka akan semakin bagus hasil nilai akurasi.
 3. Penelitian yang dilakukan oleh Rio Adi Arnomo, Wawan Laksito Yuly Saptomo dan Paulus Harsadi dengan judul "Implementasi Algoritma K-Nearest Neighbor Untuk Identifikasi Kualitas Air". Yang dapat disimpulkan dari penelitian tersebut adalah penelitian dilakukan berdasarkan laporan uji lab kualitas air bulan Maret dan April 2016 di PDAM kota Surakarta sebanyak 111 data, yang terdiri dari 71 data training dan 40 data testing. Data dikelompokkan dalam 2 (dua) parameter yaitu parameter fisik dengan atribut bau, rasa, suhu, warna dan kekeruhan, dan parameter kimia dengan atribut sisa clor, pH, kesadahan, besi, mangan, ammonium dan nitrit. Dengan variable *class* yang ditentukan MS (Memenuhi Syarat) dan TMS (Tidak Memenuhi Syarat). Setelah dilakukan pengolahan data dengan algoritma K-NN

dimana $K=7$ menghasilkan nilai akurasi sebesar 82,5%.

4. Penelitian yang dilakukan oleh Ni Luh Gede Pivin Suwirmayanti dengan judul “Penerapan Metode K-Nearest Neighbor Untuk Sistem Rekomendasi Pemilihan Mobil”. Pada penelitian ini yang dapat disimpulkan berupa perancangan pembuatan model yang terdiri dari *data flow diagram*, *entity relationship diagram*, struktur file dan desain sistem. Kriteria-kriteria yang digunakan dalam rekomendasi pemilihan mobil yaitu : tujuan pembelian, harga, tahun pembuatan, kapasitas penumpang, warna, kapasitas mesin, jenis transmisi.
5. Penelitian yang dilakukan oleh Muhammad Rivki dan Adam Mukharil Bachtiar yang berjudul “Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian *Follower Twitter* Yang Menggunakan Bahasa Indonesia”. Yang dapat disimpulkan dari penelitian tersebut adalah bahwa data penelitian diperoleh dengan memanfaatkan REST-ful APIs yang sudah disediakan oleh *twitter* dan dapat digunakan secara publik. Kemudian dilakukan *preprocessing*, dan pembobotan. Selanjutnya dilakukan klasifikasi menggunakan algoritma K-Nearest Neighbor dengan nilai $K=3$, dan *class* terdiri dari kuliner, travel dan olahraga. Setelah dilakukan 4 (empat) kali pengujian dengan

data uji 25, 50, 75 dan 100, diperoleh nilai akurasi terbesar adalah 68%.

2.2 Landasan Teori

1. Data Mining

Data mining merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database yang berukuran besar. Data mining dibagi menjadi beberapa kelompok, yaitu : deskripsi, estimasi, prediksi, klasifikasi, pengklusteran dan asosiasi.

Untuk pengolahan data mining terbagi dalam beberapa tahapan, yaitu :

- *Data Cleaning*, untuk menghilangkan noise atau yang tidak relevan.
- *Data Integration*, menggabungkan beberapa sumber data.
- *Data Selection*, mengambil data yang relevan untuk dianalisis.
- *Data Transformation*, mengubah data kedalam bentuk yang sesuai dengan model yang digunakan.
- *Data Mining*, proses aplikasi dengan suatu metode untuk mengekstrak pola data.
- *Pattern Evaluation*, mengidentifikasi pola yang benar-benar menarik yang mewakili

pengetahuan yang didasarkan pada beberapa pengukuran.

- *Knowledge Presentation*, teknik representasi *knowledge* yang digunakan untuk menyajikan *knowledge* kepada pengguna.

2. Preprocessing

Sekumpulan data yang ada di dalam penyimpanan data cenderung tidak sempurna dikarenakan beberapa hal, yaitu :

- *Incomplete*, nilai atribut tidak lengkap, nilai atribut yang seharusnya ada tetapi tidak ada.
- *Noisy*, mengandung *error* atau *outliers* (data yang secara nyata berbeda dengan data-data yang lain)
- *Inconsistent*, terjadi perbedaan dalam pengkodean dan nilai.

3. Klasifikasi

Klasifikasi adalah suatu proses menempatkan suatu objek tertentu dalam set kategori berdasarkan properti objek yang bersangkutan. Sehingga hasil dari klasifikasi ini berupa target variabel kategori (*class/label*). Klasifikasi merupakan proses pembelajaran secara terbimbing (*supervised learning*), sehingga untuk melakukan klasifikasi dibutuhkan training set sebagai data pembelajaran, dan setiap sampel dari training set memiliki atribut *class/label*.

4. Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor merupakan metode klasifikasi yang menentukan label (*class*)

dari suatu objek baru berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam K-Neighbors. Salah satu teknik menghitung pengukuran jarak dalam K-Nearest Neighbors adalah *Euclidean Distance*. *Euclidean distance* antara titik X_1 and X_2 merupakan panjang dari X_1 X_2 . Segmen dalam koordinat Cartesian, jika $X_1 = (X_{11}, X_{12}, \dots, X_{1n})$ dan $X_2 = (X_{21}, X_{22}, \dots, X_{2n})$ merupakan dua titik dalam *Euclidean n-space*, maka jarak dari X_1 ke X_2 dapat dihitung sebagai berikut:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Keterangan

p : nilai atribut data training

q : nilai atribut data testing

5. Pengukuran Akurasi

Untuk mengukur akurasi hasil pengujian data testing pada metode klasifikasi dapat dihitung dengan :

$$Akurasi = \frac{\text{Jumlah klasifikasi benar}}{\text{Total data testing yang diuji}} \quad (2)$$

2.3 Tahapan Penelitian

Tahapan-tahapan yang dilakukan pada penelitian ini adalah :

- Pengumpulan data dengan cara mengambil data yang telah ada di database.
- *Preprocessing*, melakukan *data cleaning* dan *data transformation*.

- Menyelompokkan data menjadi data training dan data testing.
- Menghitung *euclidean distance* untuk setiap data testing dengan data training untuk menentukan *class* Program Studi untuk setiap data testing.
- Membandingkan *class* Program Studi yang dihasilkan dari algoritma K-NN, dengan *class* Program Studi yang ada di data testing.
- Menghitung pengukuran tingkat akurasi data testing.
- Mengimplementasikan prediksi program studi bagi calon mahasiswa baru dalam bentuk perangkat lunak.

3 Metode Penelitian

Pada penelitian ini, data diambil dari data mahasiswa tahun masuk 2019 yang berjumlah 379 record, dengan atribut terdiri dari :

- No. Register, merupakan nomor register pendaftaran.
- Asal Sekolah, nama sekolah tingkat SMU/SMK/MA.
- Jurusan, jurusan yang diambil pada saat SMU/SMK/MA.
- Nilai Tes, merupakan grade dari nilai rata-rata hasil tes seleksi umum, matematika dan bahasa Inggris.

Bentuk data penelitian dapat dilihat pada berikut ini :

Tabel 1. Data Mahasiswa Tahun 2019

NO. REGISTER	ASAL SEKOLAH	NILAI TES	JURUSAN	PROGRAM STUDI	IPK
2019010002	PGRI RAWALIMBU	C	TKJ	SISTEM INFORMASI	3,08
2019010004	NEGERI 3 BEKASI	B	TEKNIK KOMPUTER DAN JARINGAN	TEKNIK INFORMATIKA	3,47
2019010005	NEGERI 1 BENDO KAB. MAGETAN	C	MULTIMEDIA	TEKNIK INFORMATIKA	-
2019010008	TAMAN HARAPAN	C	TEKNIK KENDARAAN RINGAN	TEKNIK INFORMATIKA	1,92
2019010010	KARYA GUNA BHAKTI 2		TEKNIK KOMPUTER DAN JARINGAN	TEKNIK INFORMATIKA	3,00
2019010011	BINAKARYA MANDIRI 2	C	TEKNIK KOMPUTER DAN JARINGAN	TEKNIK INFORMATIKA	-
2019010012	BINA KARYA MANDIRI		TEKNIK KOMPUTER DAN JARINGAN	TEKNIK INFORMATIKA	3,21
2019010014	NEGERI 1 ABUNG BARAT	C	IPA	TEKNIK INFORMATIKA	-
2019010015	TELEKOMUNIKASI TELESANDI BEKASI	A	TEKNIK KOMPUTER DAN JARINGAN	TEKNIK INFORMATIKA	3,32
2019010016	TUNAS HARAPAN PULUPUH	C	TEKNIK OTOMOTIF	MANAJEMEN INFORMATIKA	3,10
2019010017	PGRI 11 JAKARTA	C	TEKNIK OTOMOTIF	TEKNIK INFORMATIKA	2,29
2019010019	GLOBAL PRIMA ISLAMIC SCHOOL	C	TEKNIK KOMPUTER DAN JARINGAN	SISTEM INFORMASI	3,00
2019010020	NEGERI 2 CIKARANG BARAT KAB. BEKASI	C	PEMASARAN	SISTEM INFORMASI	2,49
2019010023	YPIA ALFALAH	C	AKUNTANSI	TEKNIK INFORMATIKA	2,82
2019010024	NEGERI 1 NGLIPAR GUNUNGKIDUL	B	AKUNTANSI	KOMPUTERISASI AKUNTANSI	-
2019010025	KARYA GUNA BHAKTI 1	C	TEKNIK KENDARAAN RINGAN	SISTEM INFORMASI	1,13
2019010027	NEGERI 1 LURAGUNG	C	MULTIMEDIA	TEKNIK INFORMATIKA	2,06
2019010030	TELEKOMUNIKASI TELESANDI BEKASI	C	TEKNIK KOMPUTER DAN JARINGAN	TEKNIK INFORMATIKA	-
2019010031	NEGERI 2 KOTA BEKASI	C	TEKNIK ELEKTRONIKA INDUSTRI	SISTEM INFORMASI	2,61
2019010032	BINA PRESTASI	C	TEKNIK KENDARAAN RINGAN	TEKNIK INFORMATIKA	2,87

Kemudian pada data tersebut dilakukan proses *cleaning* dengan mengabaikan data yang memiliki kriteria :

- Nilai atributnya tidak lengkap, data yang seharusnya ada tetapi tidak ada.
- IPK kurang dari 3.00. Karena untuk yang IPK lebih dari sama dengan 3.00 dapat dipastikan bahwa mahasiswa tersebut dapat mengikuti matakuliah di program studi. Sehingga dapat dinyatakan mahasiswa tersebut dalam memilih program studi sudah sesuai. Karena itu yang digunakan dalam penelitian ini hanya mahasiswa yang memiliki IPK lebih dari sama dengan 3.00.

Sehingga dari 379 record yang ada, setelah dilakukan *data cleaning* menghasilkan data sebanyak 173 record.

Untuk memudahkan penelitian, data diubah dengan kriteria sebagai berikut :

- Asal Sekolah, dikelompokkan menjadi

NEGERI dan SWASTA. Untuk Asal Sekolah yang terdapat kata NEGERI maka dikelompokkan dalam kategori NEGERI selain itu SWASTA.

2. Jurusan, dikelompokkan menjadi 6 kategori :
 - a. Ekonomi, terdiri dari jurusan Akuntansi, Penjualan dan Perbankan.
 - b. IPA
 - c. IPS
 - d. RPL (Rekayasa Perangkat Lunak)
 - e. Teknik, terdiri dari semua jurusan bidang teknik : Teknik Mesin, Teknik Otomotif, Teknik Elektronika Industri dll, kecuali TKJ.
 - f. TKJ (Teknik Komputer Jaringan)

Hasil data yang telah dikelompokkan dapat dilihat pada tabel 2.

Tabel 2. Data Hasil Pengelompokkan

NO. REGISTER	ASAL SEKOLAH	NILAI TES	JURUSAN	IPK	PROGRAM STUDI
2019010002	SWASTA	C	TKJ	3,08	SISTEM INFORMASI
2019010016	SWASTA	C	TEKNIK	3,10	MANAJEMEN INFORMATIKA
2019010040	SWASTA	C	TKJ	3,11	TEKNIK INFORMATIKA
2019010058	SWASTA	B	TKJ	3,13	TEKNIK INFORMATIKA
2019010060	SWASTA	A	TKJ	3,68	TEKNIK INFORMATIKA
2019010072	SWASTA	C	TKJ	3,61	TEKNIK INFORMATIKA
2019010073	SWASTA	C	TKJ	3,24	TEKNIK INFORMATIKA
2019010077	SWASTA	B	TKJ	3,71	TEKNIK INFORMATIKA
2019010079	SWASTA	C	TKJ	3,26	SISTEM INFORMASI
2019010080	NEGERI	C	IPA	3,18	TEKNIK INFORMATIKA
2019010086	SWASTA	A	TKJ	2,97	TEKNIK INFORMATIKA
2019010097	NEGERI	B	IPA	3,55	TEKNIK INFORMATIKA
2019010103	SWASTA	A	TKJ	3,55	TEKNIK KOMPUTER
2019010108	NEGERI	B	TKJ	3,47	TEKNIK INFORMATIKA
2019010114	SWASTA	A	TKJ	3,00	TEKNIK INFORMATIKA

Untuk atribut penelitian selanjutnya yang digunakan adalah Asal Sekolah, Nilai Tes, dan Jurusan. Dengan variable target adalah Program

Studi. Sedangkan IPK hanya digunakan sebagai filter bahwa data yang digunakan sebagai data training dan data testing yang memiliki IPK lebih dari sama dengan 3.00.

Pada tahapan preprocessing selanjutnya dilakukan transformation, mengubah data sesuai dengan bentuk model yang digunakan. Untuk itu diberikan nilai pembobotan yang dapat dilihat pada masing-masing tabel di bawah ini.

Tabel 3. Pembobotan Asal Sekolah

ASAL SEKOLAH	BOBOT
NEGERI	1
SWASTA	2

Tabel 4. Pembobotan Nilai Tes

NILAI TES	BOBOT
A	3
B	2
C	1

Tabel 5. Pembobotan Jurusan

JURUSAN	BOBOT
Ekonomi	1
IPA	2
IPS	3
RPL	4
Teknik	5
TKJ	6

Dari 173 data yang ada sebanyak 130 digunakan untuk data training yang dapat dilihat di tabel 6,

Tabel 6. Data Training

NO. REGISTER	ASAL SEKOLAH	BOBOT 1	NILAI TES	BOBOT 2	JURUSAN	BOBOT 3	PROGRAM STUDI
2019010002	SWASTA	2	C	1	TKJ	6	SISTEM INFORMASI
2019010016	SWASTA	2	C	1	TEKNIK	5	MANAJEMEN INFORMATIKA
2019010040	SWASTA	2	C	1	TKJ	6	TEKNIK INFORMATIKA
2019010058	SWASTA	2	B	2	TKJ	6	TEKNIK INFORMATIKA
2019010060	SWASTA	2	A	3	TKJ	6	TEKNIK INFORMATIKA
2019010072	SWASTA	2	C	1	TKJ	6	TEKNIK INFORMATIKA
2019010073	SWASTA	2	C	1	TKJ	6	TEKNIK INFORMATIKA
2019010077	SWASTA	2	B	2	TKJ	6	TEKNIK INFORMATIKA
2019010079	SWASTA	2	C	1	TKJ	6	SISTEM INFORMASI
2019010080	NEGERI	1	C	1	IPA	2	TEKNIK INFORMATIKA
2019010086	SWASTA	2	A	3	TKJ	6	TEKNIK INFORMATIKA
2019010097	NEGERI	1	B	2	IPA	2	TEKNIK INFORMATIKA
2019010103	SWASTA	2	A	3	TKJ	6	TEKNIK KOMPUTER
2019010108	NEGERI	1	B	2	TKJ	6	TEKNIK INFORMATIKA
2019010114	SWASTA	2	A	3	TKJ	6	TEKNIK INFORMATIKA

dan sisanya sebanyak 43 digunakan untuk data testing dapat dilihat di tabel 6.

Tabel 7. Data Testing

NO. REGISTER	ASAL SEKOLAH	BOBOT 1	NILAI TES	BOBOT 2	JURUSAN	PROGRAM STUDI
2019010004	NEGERI	1	B	2	TKJ	TEKNIK INFORMATIKA
2019010015	SWASTA	2	A	3	TKJ	TEKNIK INFORMATIKA
2019010019	SWASTA	2	C	1	TKJ	SISTEM INFORMASI
2019010035	SWASTA	2	C	1	TKJ	TEKNIK INFORMATIKA
2019010045	SWASTA	2	B	2	TKJ	TEKNIK KOMPUTER
2019010048	NEGERI	1	C	1	TEKNIK	TEKNIK INFORMATIKA
2019010051	SWASTA	2	B	2	TKJ	TEKNIK INFORMATIKA
2019010053	NEGERI	1	C	1	IPA	TEKNIK INFORMATIKA
2019010054	SWASTA	2	C	1	TEKNIK	TEKNIK INFORMATIKA
2019010067	SWASTA	2	C	1	TEKNIK	TEKNIK KOMPUTER

Kemudian dari masing-masing data training dan data testing diolah menggunakan algoritma K-Nearest Neighbor dengan K=7, dan dihitung nilai akurasi.

4 Hasil dan Pembahasan

Pada penelitian ini perhitungan masih dilakukan dengan menggunakan Ms. Excel. Hal ini dilakukan agar peneliti lebih memahami proses perhitungan tersebut agar lebih mudah dalam pembuatan aplikasinya.

Perhitungan mencari jarak antara data data testing ke-1 dengan data training ke-1 dengan cara :

$$d(p, q) = \sqrt{(2 - 1)^2 + (1 - 2)^2 + (6 - 6)^2} = \sqrt{2} = 1,41$$

Kemudian dihitung data testing-1 dengan data training ke-2 sampai data training ke-130, diperoleh hasil secara keseluruhan ditunjukkan pada tabel 8.

Tabel 8. Jarak data testing ke-1 dengan seluruh data training

NO. REGISTER	BOBOT 1	BOBOT 2	BOBOT 3	PROGRAM STUDI	JARAK
2019010002	2	1	6	SISTEM INFORMASI	1,41
2019010016	2	1	5	MANAJEMEN INFORMATIKA	1,73
2019010040	2	1	6	TEKNIK INFORMATIKA	1,41
2019010058	2	2	6	TEKNIK INFORMATIKA	1,00
2019010060	2	3	6	TEKNIK INFORMATIKA	1,41
2019010072	2	1	6	TEKNIK INFORMATIKA	1,41
2019010073	2	1	6	TEKNIK INFORMATIKA	1,41
2019010077	2	2	6	TEKNIK INFORMATIKA	1,00
2019010079	2	1	6	SISTEM INFORMASI	1,41
2019010080	1	1	2	TEKNIK INFORMATIKA	4,12
2019010086	2	3	6	TEKNIK INFORMATIKA	1,41
2019010097	1	2	2	TEKNIK INFORMATIKA	4,00
2019010103	2	3	6	TEKNIK KOMPUTER	1,41
2019010108	1	2	6	TEKNIK INFORMATIKA	0,00
2019010114	2	3	6	TEKNIK INFORMATIKA	1,41

Data diurutkan secara *ascending* berdasarkan jarak. Karena ditentukan K=7 maka untuk 7 (tujuh) tetangga terdekat adalah data kesatu sampai ketujuh seperti yang terlihat pada tabel 10.

Tabel 10. Penentuan jarak terdekat dan *class*

NO. REGISTER	BOBOT 1	BOBOT 2	BOBOT 3	PROGRAM STUDI	JARAK
2019010108	1	2	6	TEKNIK INFORMATIKA	0,00
2019010133	1	2	6	TEKNIK INFORMATIKA	0,00
2019010353	1	2	6	TEKNIK INFORMATIKA	0,00
2019010353	1	2	6	TEKNIK INFORMATIKA	0,00
2019010415	1	2	6	SISTEM INFORMASI	0,00
2019010058	2	2	6	TEKNIK INFORMATIKA	1,00
2019010077	2	2	6	TEKNIK INFORMATIKA	1,00
2019010123	1	3	6	TEKNIK KOMPUTER	1,00
2019010135	2	2	6	SISTEM INFORMASI	1,00
2019010191	2	2	6	TEKNIK INFORMATIKA	1,00
2019010295	1	3	6	TEKNIK INFORMATIKA	1,00
2019010323	2	2	6	TEKNIK INFORMATIKA	1,00
2019010331	2	2	6	MANAJEMEN INFORMATIKA	1,00
2019010344	1	1	6	TEKNIK INFORMATIKA	1,00
2019010344	1	1	6	TEKNIK INFORMATIKA	1,00

Dari 7 (tujuh) tetangga terdekat tersebut diperoleh *class* Teknik Informatika sebanyak 6 sedangkan *class* Sistem Informasi sebanyak 1 seperti yang dapat dilihat pada tabel 10. Karena mayoritas adalah *class* Teknik Informatika, sehingga dapat disimpulkan untuk data testing calon mahasiswa dengan data Asal Sekolah = Negeri, Nilai Tes = B dan dari Jurusan = TKJ maka prediksi calon mahasiswa tersebut akan memilih Program Studi Teknik Informatika.

Perhitungan untuk mencari *class* Program Studi dilakukan untuk semua data testing yang hasilnya dapat dilihat pada tabel 11.

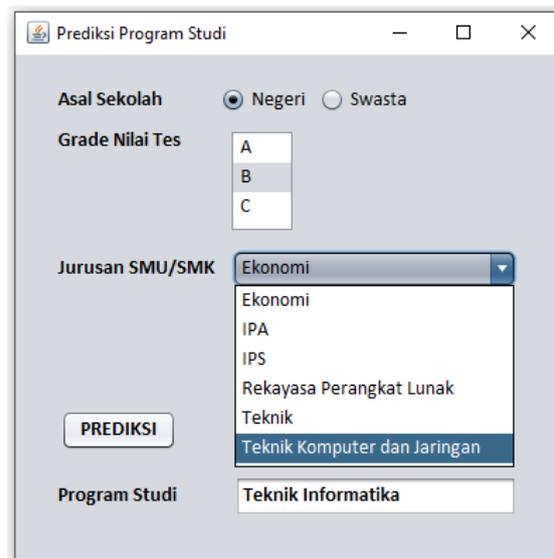
Tabel 11. Hasil klasifikasi antara data testing dengan data hasil K-NN

NO. REGISTER	BOBOT 1	BOBOT 2	BOBOT 3	PROGRAM STUDI (Data Testing)	PROGRAM STUDI (K-NN; K=7)
2019010004	1	2	6	TEKNIK INFORMATIKA	TEKNIK INFORMATIKA
2019010015	2	3	6	TEKNIK INFORMATIKA	TEKNIK INFORMATIKA
2019010019	2	1	6	SISTEM INFORMASI	TEKNIK INFORMATIKA
2019010035	2	1	6	TEKNIK INFORMATIKA	TEKNIK INFORMATIKA
2019010045	2	2	6	TEKNIK KOMPUTER	TEKNIK INFORMATIKA
2019010048	1	1	5	TEKNIK INFORMATIKA	KOMPUTERISASI AKUNTANSI
2019010051	2	2	6	TEKNIK INFORMATIKA	TEKNIK INFORMATIKA
2019010053	1	1	2	TEKNIK INFORMATIKA	SISTEM INFORMASI
2019010054	2	1	5	TEKNIK INFORMATIKA	TEKNIK KOMPUTER
2019010067	2	1	5	TEKNIK KOMPUTER	TEKNIK KOMPUTER

Berdasarkan hasil pada tabel 11 secara keseluruhan diperoleh *class* Program Studi (Data Testing) yang hasilnya sama dengan *class* Program Studi (K-NN; K=7) berjumlah 32 data, sehingga diperoleh nilai akurasi sebesar :

$$Akurasi = \frac{32}{43} = 74\%$$

Dari rangkaian proses menentukan *class* Program Studi dengan algoritma K-NN yang menggunakan pengukuran jarak *Euclidean Distance* diimplementasikan ke Java NetBeans dengan rancangan seperti pada gambar 1.



Gambar 1. Rancangan Form

Pada penelitian ini, aplikasi yang dibangun hanya terdiri dari satu form saja, yaitu untuk membantu memprediksi program studi bagi calon mahasiswa baru yang masih bingung dalam menentukan program studi. Sehingga

rancangan database hanya terdiri satu tabel untuk menyimpan data training. Sedangkan penyimpanan data lainnya menggunakan variable dan array yang didefinisikan dalam program. Pengolahan data algoritma K-NN ada di proses tombol Prediksi yang hasilnya akan ditampilkan pada teks Program Studi. Karena aplikasi ini hanya membantu bagi calon mahasiswa dalam memilih program studi maka data calon mahasiswa tidak disimpan dalam database.

5 Kesimpulan

Dari penelitian yang telah dilakukan, maka dapat disimpulkan beberapa hal sebagai berikut :

1. Dalam klasifikasi dengan algoritma K-NN diperlukan data training yang digunakan sebagai data pembelajaran yang diambil dari data mahasiswa tahun masuk 2019 dengan jumlah data yang digunakan sebanyak 173 data.
2. Pengolahan data terhadap 43 data testing dan 130 data training dengan algoritma K-NN menggunakan rumus pengukuran jarak *Euclidean Distance* diperoleh nilai akurasi sebesar 74%.
3. Dengan nilai akurasi yang cukup tinggi maka dengan data training tersebut layak untuk digunakan sebagai data pendukung untuk memprediksi calon mahasiswa baru dalam memilih program studi.

6 Daftar Pustaka

- [1] A. Q. Sesilia Novita R, Prihastuti Harsani, "Penerapan K-Nearest Neighbor (KNN) untuk Klasifikasi Anggrek Berdasarkan Karakter Morfologi Daun dan Bunga," *J. Ilm. Ilmu Komput. dan Mat.*, vol. 15, no. 1, pp. 118–125, 2018, [Online]. Available: <https://journal.unpak.ac.id/index.php/komputasi/article/view/1267/1074>.
- [2] A. S. Budiman and X. A. Parandani, "Uji Akurasi Klasifikasi Dan Validasi Data Pada Penggunaan Metode Membership Function Dan Algoritma C4.5 Dalam Penilaian Penerima Beasiswa," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 565–578, 2018, doi: 10.24176/simet.v9i1.2021.
- [3] A. Q. Sesilia Novita R, Prihastuti Harsani, "Penerapan K-Nearest Neighbor (KNN) untuk Klasifikasi Anggrek Berdasarkan Karakter Morfologi Daun dan Bunga," *J. Ilm. Ilmu Komput. dan Mat.*, vol. 15, no. 1, pp. 118–125, 2018, [Online]. Available: <https://journal.unpak.ac.id/index.php/komputasi/article/view/1267/1074>.
- [4] A. Yobioktabera and A. W. Wibowo, "Penerapan Data Mining Untuk Memprediksi Penerimaan Calon Mahasiswa Baru Fakultas Kedokteran," *J. Tek. Elektro Terap.*, vol. 10, no. 1, pp. 16–19, 2021, [Online]. Available: <https://jurnal.polines.ac.id/index.php/jtet/article/view/2550/pdf>.
- [5] Luh Gede Pivin Suwirmayanti, "Penerapan Metode K-Nearest Neighbor Untuk Sistem Rekomendasi Pemilihan Mobil Implementation of K-Nearest Neighbor Method for Car Selection Recommendation System," *Techno.COM*, vol. 16, no. 2, pp. 120–131, 2017.
- [6] M. K. Khamdani, N. Hidayat, and R. K. Dewi, "Implementasi Metode K-Nearest Neighbor Untuk Mendiagnosis Penyakit Tanaman Bawang Merah," *J. Pengemb.*

- Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 1, pp. 11–16, 2021.
- [7] M. Reza Noviansyah, T. Rismawan, and D. Marisa Midyanti, "Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data Aws (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya)," *J. Coding, Sist. Komput. Untan*, vol. 06, no. 2, pp. 48–56, 2018, [Online]. Available: <http://jurnal.untan.ac.id/index.php/jcskommi pa/article/view/26672>.
- [8] M. Rivki and A. M. Bachtiar, "Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia," *J. Sist. Inf.*, vol. 13, no. 1, p. 31, 2017, doi: 10.21609/jsi.v13i1.500.
- [9] M. Saiful and Syamsuddin, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Predikat Ketuntasan Belajar Siswa Pasca Pandemi Covid 19," *J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 96–104, 2021.
- [10] R. A. Amomo, W. L. Y. Saptomo, and P. Harsadi, "Implementasi Algoritma K-Nearest Neighbor Untuk Identifikasi Kualitas Air (Studi Kasus : Pdam Kota Surakarta)," *J. Teknol. Inf. dan Komun.*, vol. 6, no. 1, 2018, doi: 10.30646/tikomsin.v6i1.345.
- [11] R. R. Rahayu, "Mplementasi algoritma c4.5 untuk menentukan aturan rekomendasi calon penerima beasiswa," *INFOKOM*, vol. 7, no. 2, pp. 37–43, 2019, [Online]. Available: <http://journal.piksi.ac.id/index.php/INFOKOM/article/view/163>.
- [12] S. Sanjaya, M. L. Pura, S. K. Gusti, F. Yanto, and F. Syafria, "K-Nearest Neighbor for Classification of Tomato Maturity Level Based on Hue, Saturation, and Value Colors," *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 2, p. 101, 2019, doi: 10.24014/ijaidm.v2i2.7975.
- [13] Yahya and W. Puspita Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada 'Lombok Vape On,'" *Infotek J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 104–114, 2020, doi: 10.29408/jit.v3i2.2279.
- [14] B. Liu, *Web Data Mining Exploring Hyperlinks, Content, and Usage Data*, Second. Springer London, 2011.
- [15] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, Third. Morgan Kaufman Publishers, 2012.
- [16] M. Saiful and S. Aris, "Penerapan Sistem Informasi Tracer Study untuk Mengetahui Tingkat Kontribusi Perguruan Tinggi dengan Kompetensi Lulusan (Studi Kasus Fakultas Teknik Universitas Hamzanwadi)," *J. Inform. dan Teknol.*, vol. 2, no. 1, pp. 43–52, 2019.