

## PREDIKSI PENENTUAN PROGRAM STUDI MENGGUNAKAN ALGORITMA K-NN PADA PELAJAR SMAN 6 KOTA DEPOK JURUSAN IPA

Arindiah Rahmawati, Bayu Hananto

Fakultas Ilmu Komputer  
Universitas Pembangunan Nasional Veteran Jakarta  
Email : [arinrah184@gmail.com](mailto:arinrah184@gmail.com), [bayu.hananto86@gmail.com](mailto:bayu.hananto86@gmail.com)  
Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia

### Abstrak

Saat ini pelajar SMA masih banyak yang masih bingung dalam menentukan Program Studi di Perguruan Tinggi Negeri, sedangkan itu merupakan salah satu hal yang penting. Mereka masih bergantung kepada orang tua atau teman, siswa SMA tersebut belum memiliki alasan yang kuat untuk memilih Program Studi. Teknik data mining dapat digunakan untuk memprediksi program studi berdasarkan nilai mereka. Pengolahan dan prediksi nilai berdasarkan nilai pelajar dari semester 1 sampai semester 5. Penelitian menggunakan metode K-Nearest Neighbor yang merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat dengan objek tersebut. Output dari sistem yang dibangun dapat dijadikan sebagai acuan bagi pelajar SMA agar tidak salah dalam memilih program studi dan dapat membantu dalam menentukan program studi di Perguruan Tinggi Negeri. Dari data yang diperoleh sebanyak 208 data siswa, lalu dibagi menjadi data latih dan data uji secara berurutan sebesar 70% dan 30% maka hasil yang diperoleh adalah *cluster*  $k=2$  dengan akurasi 79.03% dengan nilai Kappa 0.713 merupakan akurasi yang paling tinggi dalam penelitian ini.

Kata Kunci : K-NN, Prediksi Program Studi, Klasifikasi Data Mining

### 1. PENDAHULUAN

Dalam kerangka integrasi pendidikan menengah dengan pendidikan tinggi, sekolah diberi peran dalam proses seleksi Nasional Masuk Perguruan Tinggi Negeri atau biasa disingkat SNMPTN dengan asumsi bahwa sekolah sebagai satuan pendidikan dan guru sebagai pendidik selalu menjunjung tinggi kehormatan dan kejujuran sebagai bagian dari prinsip pendidikan karakter (<http://snmptn.ac.id>). SNMPTN adalah salah satu bentuk jalur seleksi penerimaan mahasiswa untuk memasuki perguruan tinggi negeri yang dilaksanakan serentak seluruh Indonesia, SNMPTN merupakan jalur undangan dan tidak membutuhkan tes atau ujian lagi.

Penerimaan mahasiswa baru harus memenuhi prinsip adil, akuntabel, transparan, dan tidak diskriminatif dengan tidak membedakan jenis kelamin, agama, suku, ras, kedudukan sosial, dan tingkat kemampuan ekonomi calon mahasiswa serta tetap memperhatikan potensi calon mahasiswa dan kekhususan perguruan tinggi. Perguruan tinggi sebagai penyelenggara pendidikan, setelah pendidikan menengah, menerima calon mahasiswa yang berprestasi akademik tinggi dan diprediksi akan berhasil menyelesaikan studi di perguruan tinggi tepat waktu. Siswa yang berprestasi tinggi dan konsisten menunjukkan prestasinya di SMA/SMK/MA layak mendapatkan kesempatan untuk menjadi calon mahasiswa melalui SNMPTN (<http://snmptn.ac.id>). Pelajar yang berkeinginan untuk melanjutkan Pendidikan ke perguruan tinggi negeri (PTN) dapat memilih cara pendaftaran melalui ujian mandiri oleh PTN. SNMPTN merupakan seleksi nasional masuk perguruan tinggi negeri berdasarkan nilai rapor dan prestasi-prestasi lain. Kemudahan ini menyebabkan SNMPTN menjadi jalur yang banyak

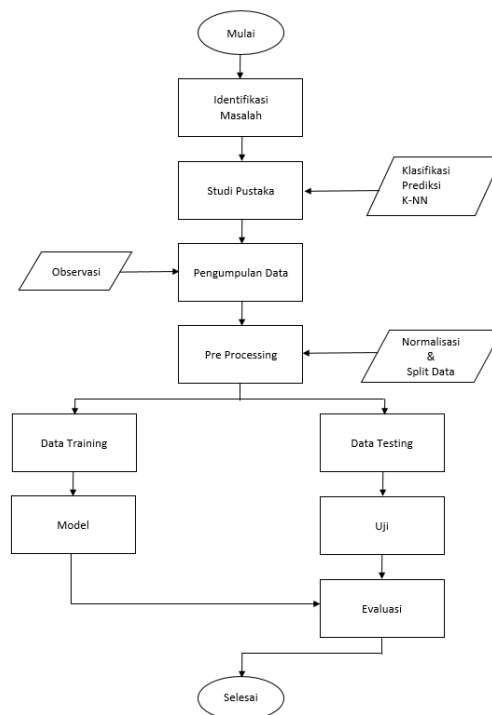
diminati oleh pelajar, tetapi daya tampung yang terbatas menyebabkan pelajar harus menentukan pilihan jurusan dan universitas dengan tepat agar dapat diterima melalui jalur SNMPTN.

Dengan menggunakan nilai rata-rata raport dari semester 1 sampai 5, untuk menghitung nilai raport pelajar akan menggunakan metode K-Nearest Neighbor (K-NN) nilai yang akan diproses untuk jurusan IPA adalah Bahasa Indonesia, Bahasa Inggris, Matematika, Fisika, Kimia, dan Biologi dan pilihan Universitas hanya Universitas Indonesia yang lebih banyak dirujuk oleh siswa. Algoritma K-Nearest Neighbor (K-NN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Termasuk dalam supervised learning, dimana hasil query instance yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam K-NN (informatikalogi.com).

## 2. METODOLOGI PENELITIAN

### 2.1. Kerangka Pikir

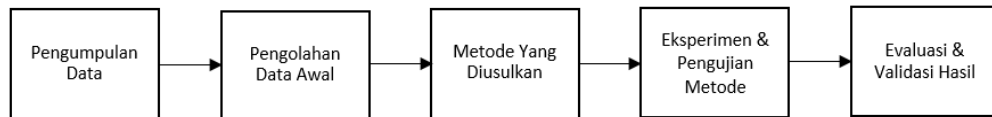
Dalam penelitian ini dibutuhkan data-data serta informasi untuk mendukung kebenaran dalam materi dan pembahasan. Data didapatkan dengan observasi langsung ke SMAN 6 Depok dan mengambil data kelas XII. Data tersebut diambil dari semester 1 sampai dengan semester 5 mereka. Dengan jumlah data sebanyak 208 siswa, data tersebut dibagi menjadi 70% untuk data training dan didapatkan 146 data siswa, dan 30% untuk data testing didapatkan 62 data siswa.



Gambar 7 Metodologi Penelitian

### 2.2. Desain Pengujian

Pada penelitian ini data yang digunakan adalah data nilai pelajar yang didapatkan dari observasi ke SMAN 6 Kota Depok. Selanjutnya data tersebut akan diolah menggunakan algoritma K-NN (K-Nearest Neighbor) dengan desain pengujian penelitian sebagai berikut :



Gambar 8 Desain Pengujian

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Persiapan Data

Pada tahap analisis, persiapan data merupakan kebutuhan utama untuk mencapai suatu sasaran yang ingin dicapai dari proses pembuatan aplikasi ini, yaitu tentang input dan output yang dihasilkan. Data – data yang menjadi input dalam aplikasi ini merupakan nilai – nilai dari mata pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika, Fisika, Kimia, dan Biologi dari pelajar SMAN 6 Kota Depok dari semester 1 sampai dengan semester 5 pada tahun 2018.

Untuk pilihan Universitas diambil dari Universitas Indonesia, dengan beberapa program studi pilihan yang lebih banyak dirujuk oleh siswa SMAN 6. Untuk hasilnya dapat dilihat pada Tabel 1 berikut :

Tabel 2 Passing Grade dan Kisaran Nilai

No	Program Studi	Passing Grade	Kisaran Nilai
1	UI Teknik Industri	51.86	78.90 – 82.90
2	UI Pendidikan Kedokteran Gigi	36.92	80.10 – 82.70
3	UI Teknik Sipil	39.87	79.40 – 82.80
4	UI Teknik Perkapalan	33.73	80.50 – 84.40
5	UI Teknik Lingkungan	35.78	79.90 – 82.50
6	UI Kimia	51.60	78.30 – 81.00
7	UI MIPA Geologi	49.65	79.00 – 80.80
8	UI Geofisika	44.09	79.70 – 82.20
9	UI Fisika	36.54	80.30 – 82.20
10	UI Matematika	49.65	79.40 – 82.70
11	UI Biologi	36.83	81.00 – 83.70
12	UI Statistika	35.52	82.10 – 83.70
13	UI Teknik Kimia	40.27	82.70 – 83.60
14	UI Kesehatan Masyarakat	30.58	81.70 – 84.60
15	UI Teknik Mesin	29.24	80.10 – 82.30
16	UI Teknik Elektro	25.34	81.10 – 83.40

Sumber : <http://halokampus.com/snmptn/nilai-raport-siswa-diterima-snmptn/>

Pada Tabel 1 kisaran nilai didapat dari nilai rata-rata pelajar dari semester 1 sampai semester 5 yang mengacu pada *passing grade* dan telah diasumsikan dengan mencari nilai rata-rata *minimum* dan nilai *maximum*. Sebagai contoh target yang diberikan adalah UI Teknik Industri dengan *passing grade* 51.86, maka asumsi nilai yang diberikan adalah nilai minimal rata-rata pelajar adalah 78.90 dan nilai maksimal pelajar adalah 82.90. Data tersebut telah disesuaikan dari data yang didapat melalui halokampus.com.

Untuk penentuan Prediksi *Class*, peneliti telah mengubah untuk mempermudah proses penelitian menggunakan variabel sebagai berikut:

- a. Nilai Bahasa Indonesia = X1
- b. Nilai Bahasa Inggris = X2
- c. Nilai Matematika = X3
- d. Nilai Fisika = X4
- e. Nilai Kimia = X5
- f. Nilai Biologi = X6
- g. Nilai Rata-Rata Keseluruhan = X7

Terdapat perubahan pada tipe role dalam NISN menjadi ID, dan Class menjadi Label, dan pada statistik Min adalah nilai yang terkecil pada variabel X1 yaitu 77.600 dan Max adalah nilai terbesar pada variabel X1 yaitu 86. Berikut ini adalah tampilan hasil data nilai yang sudah di input dapat dilihat pada Tabel 2 Data Nilai Pelajar.

Tabel 3 Data Nilai Pelajar

No	NISN	X1	X2	X3	X4	X5	X6	X7	Target
1	9987251471	77,6	80,6	79,0	79,6	81,0	81,2	79,8	UI Teknik Industri
2	9991255217	78,2	77,4	76,4	79,4	77,6	82,6	78,6	UI Teknik Industri
3	9991150668	78,2	78,6	79,0	82,2	82,0	83,4	80,6	UI Teknik Industri
4	9994694815	78,4	78,2	78,8	81,0	81,6	82,8	80,1	UI Teknik Industri
5	9992015122	78,4	78,6	77,8	81,0	82,2	83,6	80,3	UI Teknik Industri
...	...	...	...	...	...	...	...	...	...
208	9984578460	86,0	84,6	80,4	83,4	83,0	89,2	84,4	UI Teknik Elektro

### 3.2. Pengolahan Data (*Preprocessing*)

Proses pengolahan data dalam penelitian ini adalah melakukan *preprocessing* data yang dilakukan untuk melihat apakah data yang akan digunakan mengalami suatu masalah terhadap mengilangnya data atau biasa yang disebut *missing value* yang jika data tersebut dibiarkan akan menghambat pada saat pengakurasian data. Dan pemeriksaan pada data apakah data yang akan di gunakan tersebut sudah sesuai untuk dilakukan analisa dengan model yang digunakan.

Menurut Sugiyono (2012) data yang sudah terkumpul di dalam tahap pengumpulan data, kemudian perlu diolah kembali. Pengolahan data tersebut memiliki tujuan agar data lebih sederhana, sehingga semua data yang telah terkumpul dan menyajikannya sudah tersusun dengan baik dan rapi baru data tersebut dapat dianalisis. Untuk mendapatkan data yang berkualitas, maka dilakukan beberapa proses sebagai berikut:

#### 3.2.1. Transformasi Data

Tahap selanjutnya dalam proses pengolahan data adalah melakukan transformasi data, pada penelitian ini menggunakan normalisasi *min-max* pada data yang akan dianalisis. Pada tahap ini peneliti melakukan transformasi data atau melakukan pengubahan skala pengukuran terhadap data asli menjadi bentuk lain yang lebih mudah dipahami. Pada penelitian ini peneliti menggunakan normalisasi pada data menggunakan rentan nilai interval yang kecil yaitu [0.1-0.9].

Merupakan proses normalisasi pada data sampel nilai pelajar. Proses normalisasi menggunakan metode Min-Max, yang mengubah nilai pada range [0-1]. Perhitungan normalisasi dengan metode range transformation dengan range nilai maximal 0.9 dan nilai minimal 0.1.

Berikut ini adalah hasil dari data yang sudah di normalisasi menggunakan rumus normalisasi diatas:

Tabel 4 Data Nilai Normalisasi

No	NISN	X1	X2	X3	X4	X5	X6	X7	Target
1	9987251471	0,100	0,379	0,321	0,220	0,468	0,348	0,280	UI Teknik Industri
2	9991255217	0,157	0,176	0,100	0,200	0,100	0,445	0,132	UI Teknik Industri
3	9991150668	0,157	0,252	0,321	0,480	0,576	0,500	0,368	UI Teknik Industri
4	9994694815	0,176	0,227	0,304	0,360	0,532	0,459	0,316	UI Teknik Industri
5	9992015122	0,176	0,252	0,219	0,360	0,597	0,514	0,332	UI Teknik Industri
...	...	...	...	...	...	...	...	...	...
208	9984578460	0,900	0,633	0,440	0,600	0,684	0,900	0,832	UI Teknik Elektro

### 3.2.2. Pembagian Data

Setelah data sudah di normalisasi dalam penelitian ini tahap selanjutnya adalah melakukan *split* data yaitu membagi data menjadi data training dan data testing. Dari data *sampel* yang diperoleh sebanyak 208 data. Pembagian data tersebut membagi partisi data 70% untuk data training dan didapatkan 146 sampel data, dan ratio 30% untuk data testing didapatkan 62 data, pembagian sampel secara linear dari data nilai awal. Data training digunakan untuk menguji apakah hasil output sesuai pada target dan data testing sebagai data latih.

### 3.2.3. Hasil

Proses dari alur pemodelan tahap pertama adalah dimulai dari hasil data yang sudah di normalisasi sebanyak 208 sampel, yang dilanjutkan dengan Splitdata, split data adalah proses pembagian data untuk mencari *error rate* secara keseluruhan secara umum pengujian dilakukan sebanyak 5 kali untuk memperkirakan akurasi estimasi dan jumlah data testing sebanyak 62 sampel data, selanjutnya untuk pemodelan K-NN (K-Nearest Neighbor) menggunakan operator k-NN, pada tahap pemodelan tersebut ditentukan besarnya nilai k pada K-NN dalam proses ini percobaan dilakukan sebanyak 5 kali dengan nilai k yang berbeda-beda dengan format bilangan prima yaitu k=2, k=3, k=5, k=7, dan k=9 dan akan dipilih dengan tingkat akurasi yang paling tinggi. Berikut ini adalah tabel hasil pengujian akurasi:

*Tabel 5 Hasil Pengujian*

No	Nilai <i>k</i>	Akurasi Testing	Kappa
1	2	79.03%	0.713
2	3	74.19%	0.646
3	5	70.96%	0.597
4	7	69.35%	0.569

Dari hasil pengujian yang telah dilakukan pada data testing dengan nilai *k* yang berbeda-beda maka hasil akurasinya pun berbeda-beda. Berdasarkan hasil pengujian pada Tabel 4 hasil akurasi terbesar didapatkan dengan percobaan untuk nilai *k*=2 dengan nilai akurasi yang didapat adalah 79.03% dengan nilai kappa 0.713, sedangkan untuk nilai *k*=7 mendapatkan nilai akurasinya terendah sebesar 69.35% dengan nilai kappa 0.569.

Untuk klasifikasi data mining, nilai tingkatan akurasi dapat dibagi menjadi beberapa kelompok sebagai berikut:

*Tabel 6 Tingkat Akurasi*

No	Akurasi	Keterangan
1	0.90-1.00	<i>Excellent classification</i>
2	0.80-0.90	<i>Good classification</i>
3	0.70-0.80	<i>Fair classification</i>
4	0.60-0.70	<i>Poor classification</i>
5	0.50-0.60	<i>Failure</i>

Sumber: Han J and Kamber M. Data Mining “Concepts, Models, and Techniques”, 2011.

Berdasarkan pada Tabel 5 maka dapat disimpulkan bahwa tingkat akurasi tertinggi pada klastering k=2 yaitu dengan akurasi 79.03% menunjukkan akurasi tergolong *Fair Classification*. Dari perbandingan diatas, dapat disimpulkan jika semakin tinggi nilai k maka nilai akurasi akan semakin rendah. Jika semakin kecil nilai k maka nilai akurasinya akan semakin tinggi.

## 4. KESIMPULAN

Dalam penelitian ini dilakukan prediksi program studi dengan menggunakan algoritma K-NN (K-Nearest Neighbor) dengan menggunakan data dari hasil observasi di SMAN 6 Kota Depok, data nilai adalah Bahasa Indonesia, Bahasa Inggris, Matematika, Fisika, Kimia, dan Biologi masing-masing dari semester 1 sampai dengan 5. Pengujian data menggunakan k=2 dengan algoritma K-NN dan akurasi yang didapat adalah 79.03%, dengan cluster data k=3

akurasinya adalah 74.19%, dengan cluster data  $k=5$  akurasinya adalah 70.96%, dengan cluster data  $k=7$  akurasinya 69.35%, dan dengan cluster data  $k=9$  akurasinya adalah 61.29%.

Maka hasil penelitian ini dapat disimpulkan bahwa tingkat akurasi pengujian prediksi program studi dengan menggunakan algoritma K-NN (K-Nearest Neighbor) dipengaruhi oleh jumlah klustering data. Semakin rendah nilai  $k$  maka akurasi semakin tinggi sedangkan jika semakin tinggi nilai  $k$  maka akurasinya akan semakin rendah. Akurasi dan nilai paling tinggi adalah dengan  $k=2$  sebesar 79.03%.

Saran untuk penelitian selanjutnya dapat menggunakan jumlah data yang lebih banyak sehingga akurasi yang diperoleh akan lebih terjamin, dan untuk penelitian selanjutnya sebaiknya gunakan atau bandingkan dengan metode lainnya untuk mendapat akurasi yang paling optimal.

## Referensi

- Banjarsari, Mutiara A. 2015. *Pencarian k-Optimal pada Algoritma kNN untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Berdasarkan IP Sampai Dengan Semester 4*. FMIPA Unlam: Banjarbaru.
- Darmawan, Astrid. 2012. *Pembuatan Aplikasi Data mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighborhood*. Fakultas Teknik Dan Ilmu Komputer Universitas Komputer Indonesia: Bandung.
- Han J and Kamber M. 2006. *Data Mining: Concept and Techniques*. New York: Morgan Kaufmann Publisher.
- Hermawati, Fajar Astuti. 2013. *Data Mining*. Yogyakarta: Andi Offset.
- Indriyawan, Eko. 2005. *Pemrograman Database meningkatkan kemampuan database Menggunakan Delphi dan MSSQL Server*. Yogyakarta: Andi.
- Jananto, A.. 2010, *Perbandingan Performansi Algoritma Nearest Neighbor dan SLIQ untuk Prediksi Kinerja Akademik Mahasiswa Baru*. Jurnal Teknologi Informasi DINAMIK, vol XV, no 2, hal 157-169.
- Kusrini, L. ET. 2009. *Algoritma Data Mining*. Yogyakarta: Andi Publishing.
- Nobertus Krisandi, 2013. *Algoritman K-Nearest Neighbor dalam Klasifikasi data Hasil Produksi Kelapa Sawit pada PT. MINAMAS Kecamatan Parindu*. 47
- Pandie, Emerensye S. Y. 2012. *Implementasi Algoritma Data mining K-Nearest Neighbour (KNN) Dalam Pengambilan Keputusan Pengajuan Kredit*. Jurusan Ilmu Komputer, Fakultas Sains dan Teknik, Universitas Nusa Cendana: Kupang
- Suyanto. 2017. *Data Mining Untuk Klasifikasi Dan Klasterisasi Data*. Bandung: Informatika Bandung
- <http://snmptn.ac.id> (diakses 28 Maret 2018).
- <http://halokampus.com> (diakses 28 Maret 2018)
- <https://informatikalogi.com>