

MENENTUKAN PILIHAN SEKOLAH DIDALAM PENERIMAAN PESERTA DIDIK BARU DENGAN MENGUNAKAN METODE NAÏVE BAYES DAN K-NEAREST NEIGHBOR (Studi Kasus: PPDB Online Jenjang SMP Kota Metro)

Toto Andri Puspito
Informatics & Business Institute Darmajaya Magister Teknik Informatika
Jl. Z.A Pagar Alam No. 93 Labuhan Ratu, Bandar Lampung
totoandri@gmail.com

ABSTRACT

Acceptance of New Learners (PPDB) is an important process to attract new students according to the criteria and characteristics of the school. PPDB needs to be implemented in an objective, transparent, accountable and non-discriminatory because it concerns the public interest, that is education. Following the current technological advances, PPDB has been held online. With the online PPDB is expected to provide solutions to parents of students for the realization of PPDB is transparent, accountable and not discriminatory. By using the online PPDB system PPDB results can be seen by anyone and anywhere. In addition, with the online PPDB will provide convenience in the registration process and the selection process because registration and selection process can be done and viewed anywhere so prospective students or parents will be easier to register.

This study aims to provide recommendations to parents in the implementation of PPDB junior high school in order to facilitate parents in determining the choice of schools in accordance with the circumstances of prospective students. In this study, the authors use the Naïve Bayes method and K-Nearest Neighbor which is the method used when determining the selection of schools in the process of acceptance of online PPDB. Naive Bayes will provide the school choice recommendation with the greatest probability of the calculation whereas K-NN sees the closeness of the data to the data of trying and the closest data to be used to provide recommendations to prospective students or parents / guardians. The steps of determining the probability of Bayes is to calculate the probability of a certain domicile, the origin of a particular school, again, the achievement of the data of student tryouts received at each junior high school SMP Negeri 1- SMP Negeri 10 which then the highest probability will be taken as a recommendation. While in K-NN use Euclidean distance equation to calculate distance data to data trying and take data closest to data trying as a recommendation.

Keywords: PPDB, K-NN, Naive Bayes

ABSTRAK

Penerimaan Peserta Didik Baru (PPDB) merupakan proses penting untuk menjaring calon siswa baru sesuai kriteria dan karakteristik sekolah. PPDB perlu dilaksanakan secara objektif, transparan, akuntabel dan tidak diskriminatif karena menyangkut kepentingan masyarakat, yakni pendidikan. Mengikuti kemajuan teknologi sekarang ini PPDB telah diselenggarakan secara online. Dengan adanya PPDB online diharapkan mampu memberikan solusi kepada orang tua siswa untuk terwujudnya PPDB yang transparan, akuntabel dan tidak deskriminatif.

Dengan menggunakan sistem PPDB online hasil PPDB dapat dilihat siapapun dan dimanapun. Selain itu, dengan adanya PPDB online akan memberikan kemudahan dalam proses pendaftaran dan proses seleksi karena pendaftaran dan proses seleksi dapat dilakukan dan dilihat dimana saja sehingga calon siswa atau orang tua akan lebih mudah melakukan pendaftaran.

Penelitian ini bertujuan memberikan rekomendasi kepada orangtua dalam penyelenggaraan PPDB jenjang SMP agar dapat mempermudah orang tua dalam menentukan pilihan sekolah yang sesuai dengan keadaan calon siswa. Pada penelitian ini penulis menggunakan metode Naïve bayes dan K-Nearest Neighbord yang merupakan metode yang digunakan saat menentukan pemilihan sekolah pada proses penerimaan PPDB online. Naive bayes akan memberikan rekomendasi pilihan sekolah dengan probabilitas terbesar hasil perhitungan sedangkan K-NN melihat kedekatan suatu data terhadap data trying dan data terdekat yang akan digunakan untuk memberikan rekomendasi kepada calon siswa atau orangtua/wali siswa. Langkah – langkah penentuan probablitas bayes adalah dengan menghitung probabilitas domisili tertentu, asal sekolah tertentu ,gakin, prestasi terhadap data trying siswa yang diterima pada masing – masing sekolah SMP Negeri 1- SMP Negeri 10 yang kemudian probabilitas tertinggi yang akan diambil sebagai rekomendasi. Sedangkan pada K-NN menggunakan persamaan euclidean distance untuk menghitung jarak data terhadap data trying dan mengambil data yang paling dekat dengan data trying sebagai rekomendasi.

Kata kunci : PPDB, K-NN ,Naive Bayes

1. PENDAHULUAN

Penerimaan Peserta Didik Baru atau disingkat PPDB adalah kegiatan penerimaan dan seleksi calon peserta didik guna menuju ke jenjang berikutnya. Kegiatan ini rutin dilaksanakan setiap tahun oleh Dinas Pendidikan Provinsi atau kota setempat. Hal ini juga di laksanakan di Kota Metro, Kota Metro yang berdiri sejak 9 Juni 1937 memiliki semboyan dan semangat untuk menjadi Kota pendidikan ini memiliki 52 sekolah negeri jenjang SD,13 sekolah jenjang SMP,10 sekolah jenjang SMA dan 4 sekolah jenjang SMK ditambah dengan jumlah sekolah swasta yang terdiri dari 18 sekolah jenjang SD,23 sekolah jenjang SMP,18 sekolah jenjang SMA dan 18 sekolah jenjang SMK.

Salah satu langkah dalam proses PPDB adalah menentukan pilihan sekolah yang ingin dituju oleh peserta didik. Jumlah pilihan sekolah yang dapat dipilih oleh peserta didik, sudah ditetapkan sesuai dengan aturan dan prosedur yang berlaku sesuai dengan Petunjuk Teknis dari Kepala Dinas Pendidikan yang merujuk pada Peraturan Wali Kota Metro. Pilihan sekolah akan berpengaruh pada proses selanjutnya yaitu seleksi. Pada umumnya Orang tua/Wali dalam menentukan pilihan sekolah, perlu mempertimbangkan beberapa faktor, yaitu nilai NUN calon peserta didik, batas nilai terendah sekolah (passing grade), popularitas sekolah (sekolah favorit), Sarana & Prasarana sekolah dan jarak sekolah.

Dengan adanya kemajuan teknologi, penyelenggaraan PPDB Kota Metro saat ini sudah dilakukan secara online, PPDB Online di Kota Metro diselenggarakan pada jenjang SMP, SMA dan SMK dengan adanya PPDB online sangat membantu orang tua wali dan calon siswa karena dalam prosesnya pendaftaran dapat dilaksanakan dimanapun dan proses seleksi yang dapat dilihat kapan saja dan dimana saja. Orang tua dapat mengetahui setiap proses seleksi yang dilakukan sistem berdasarkan Nilai Akhir (NA) siswa yang merupakan jumlah Nilai Ujian Nasional (NUN) ditambah dengan pembobotan keluarga miskin (GAKIN) dan prestasi, ketika nilai peserta didik sudah melalui ambang batas terkecil dan daya tampung pada suatu sekolah maka siswa tersebut tidak diterima disekolah pilihan tetapi ambang batas nilai terkecil sekolah A belum tentu sama dengan ambang batas terkecil sekolah B untuk itu penentuan pilihan saat pendaftaran sangat menentukan mengingat pemilihan sekolah pada sistem PPDB hanya mengizinkan siswa untuk memilih dua pilihan tempat yaitu pilihan pertama dan pilihan ke dua dan tidak dapat lagi melakukan pendaftaran atau melakukan pembatalan.

Mengacu pada hal tersebut maka dibutuhkan penelitian untuk menentukan pilihan sekolah yang diharapkan dapat membantu dan mempermudah para calon siswa dan Orang tua/Wali dalam menentukan pilihan sekolah atau sebagai pendukung dalam membuat keputusan pada saat memilih pilihan sekolah. Terutama untuk jenjang SMP dimana pada jenjang ini merupakan jenjang pertama pada penyelenggaraan PPB di Kota metro yang tentu saja Calon siswa atau orang tua/

wali siswa baru pertama kali melakukan PPDB secara online.

1.1 Tujuan dan Manfaat

- a. Tujuan dan manfaat memberikan rekomendasi kepada orang tua siswa dalam menentukan pilihan sekolah saat penyelenggaraan PPDB ini adalah :
 1. Memberikan masukan kepada orang tua untuk memilih sekolah.
 2. Mengetahui prediksi siswa untuk menentukan sekolah pilihan.
- b. Manfaat yang diperoleh dari penelitian ini adalah untuk :
 1. Memberikan saran untuk orang tua wali siswa dalam pemilihan Sekolah Menengah Pertama
 2. Memberikan masukan pada orang tua siswa untuk memilih sekolah favorit sesuai dengan kemampuan peserta didik.
 3. Memberikan tambahan pengetahuan penulis tentang algoritma dan sistem PPDB yang sedang berjalan di Kota Metro.

2. LANDASAN TEORI

2.1 Naive Bayes

Merupakan metode klasifikasi *supervised learning* atau statistic yang dapat digunakan untuk klasifikasi suatu class dengan melakukan perhitungan probabilitas. Dasar dari *Naive Bayesian Classification* adalah teorema bayes.

Naive Bayesian Classification (NBC) menyediakan algoritma pembelajaran praktis, pengetahuan sebelumnya (*prior knowledge*) dan data yang diamati dapat dikombinasikan. Teori keputusan Bayes merupakan pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*), teori Bayesian pada

dasarnya adalah kemungkinan kejadian di masa depan yang bisa dihitung dengan menentukan frekuensi pengalaman sebelumnya. Penggunaan algoritma Bayes dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya.

Metode Naïve Bayes menggunakan prinsip teorema Bayes, yaitu menghitung probabilitas suatu kejadian berdasarkan suatu kondisi tertentu, dengan menggunakan persamaan :

Karena $P(X)$ Konstan untuk setiap C maka bisa ditulis :

$$P(C|X) = P(X|C)P(C)$$

Kemudian persamaan diatas dikembangkan lagi agar mendapatkan probabilitas maksimal dengan persamaan berikut :

$$P(X|C_i) = P(C_i) \prod_{k=1}^n P(x_k|c_i)$$

X = Data dengan *class* yang belum diketahui

C = Hipotesis data X merupakan suatu *class* spesifik

$P(C|X)$ = Probabilitas hipotesis H berdasar kondisi X

$P(C)$ = Probabilitas hipotesis C

$P(X|C)$ = Probabilitas X berdasar kondisi pada hipotesis C

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

$P(X)$ = Probabilitas dari X

2.2 K-Nearest Neighbor

KNN dilakukan dengan mencari kelompok k objek dalam data *training* yang paling dekat (mirip) dengan objek pada data baru atau data *testing*. Algoritma *K-Nearest Neighbor* adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. *Nearest Neighbor* adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dan kasus lama yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada.

Klasifikasi K-NN bekerja berdasarkan pada analogi, dimana data uji dan training dikomparasikan dan ditarik kesimpulan berdasarkan kesamaan data yang dihasilkan oleh komparasi tersebut. Perhitungan dilakukan berdasarkan jarak (kedekatan data) yang kemudian dikenal dengan nama *Euclidean Distance*.

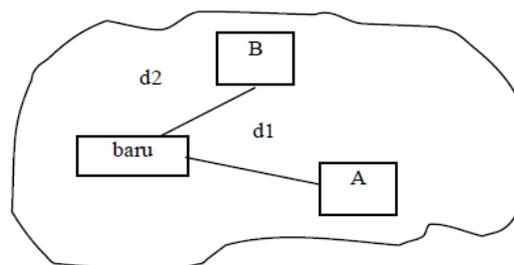
$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2}$$

Dimana :

C_{ij} = sampel data

C_{kj} = data uji

K-Nearest Neighbor (KNN) merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga terdekatnya. KNN termasuk algoritma supervised learning, dimana hasil dari query instance yang baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi. Nearest



Gambar 2.3 Ilustrasi Kedekatan Kasus

Neighbor adalah suatu pendekatan untuk menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada. Ilustrasi kedekatan kasus pada Gambar 2.3 memberikan gambaran tentang proses mencari solusi terhadap seorang pasien baru dengan

mengacu pada solusi dari pasien terdahulu. Untuk mencari kasus pasien mana yang akan digunakan, maka dihitung kedekatan anantara kasus pasien baru dengan semua kasus pasien lama. Kasus pasien lama dengan kedekatan terbesar-lah yang akan diambil solusinya untuk digunakan pada kasus pasien baru [13].

2.3 Pemodelan data

Sebelum melakukan proses klasifikasi, dilakukan pemodelan data, karena terdapat perbedaan satuan untuk digunakan pada metode Naive Bayes dan KNN. Untuk pemodelan yang dilakukan pada atribut jenis kelamin, asal daerah, gakin dan prestasi yang masing – masing :

Tabel 3.1 Pemodelan data Domisili dan Gakin

Kualifikasi	Kode
Dalam daerah	200
Luar daerah	100
Gakin	200
Bukan gakin	100
Sekolah Dalam Kota	200
Sekolah Luar Kota	100
Dalam Daerah	200
Luar Daerah	100

Pemodelan untuk jenis prestasi :

Tabel 3.2 Pemodelan data Prestasi

Kualifikasi	Kode
Tidak memiliki prestasi	100
Juara 3 Tk. Kab/Kota	200
Juara 2 Tk. Kab/Kota	300
Juara 1 Tk. Kab/Kota	400
Juara 3 Tk. Provinsi	500
Juara 2 Tk. Provinsi	600
Juara 1 Tk. Provinsi	700
Juara 3 Tk. Nasional	800
Juara 2 Tk. Nasional	900
Juara 1 Tk. Nasional	1000
Juara 3 Tk. Internasional	1100
Juara 2 Tk. Internasional	1200
Juara 1 Tk. Internasional	1300

Pemodelan data nilai rata-rata nilai akhir dilakukan dengan cara menghitung mean dan standar deviasi pada data nilai rata-rata nilai akhir. Tahapan pengkategorian data yang dilakukan adalah sebagai berikut:

1. Menghitung mean (nilai rata-rata) dari seluruh data nilai akhir dengan rumus:

$$\bar{x} = \frac{\sum x}{n}$$

Keterangan:

\bar{x} : mean
 $\sum x$: hasil penjumlahan nilai akhir
 n : jumlah data PPDB

2. Menghitung standar deviasi dari seluruh data nilai akhir dengan rumus:

$$\sigma = \sqrt{\frac{\sum (xi - \bar{x})^2}{n}}$$

Keterangan :

σ : Standar deviasi
 xi : nilai individu data nilai akhir
 \bar{x} : nilai rata-rata/mean
 n : jumlah data PPDB

3. Setelah mendapatkan nilai mean dan standar deviasi, maka data dapat dimodelkan menggunakan aturan seperti yang ditunjukkan pada Tabel 3.3

Tabel 3.3. Tabel Pemodelan Data Nilai Rata-Rata Nilai Akhir

Kategori	Kualifikasi	Kode
1	$NA < (\bar{x} - \sigma)$	4
2	$(\bar{x} - \sigma) \leq NA < \bar{x}$	3
3	$\bar{x} \leq NA < (\bar{x} + \sigma)$	2
4	$NA \geq (\bar{x} + \sigma)$	1

Setelah mendapatkan mean dan standar deviasi nilai rata-rata nilai akhir maka data dikodekan dengan aturan sebagai berikut:

Tabel 3.4. Tabel Pemodelan Data Nilai Rata-Rata Nilai Akhir

Kategori	Kualifikasi	Aturan Kodifikasi	Kode
1	$NA < (\bar{x} - \sigma)$	$NA < 21,23151$	4
2	$(\bar{x} - \sigma) \leq NA < \bar{x}$	$21,23151 \leq NA < 23,69578$	3
3	$\bar{x} \leq NA < (\bar{x} + \sigma)$	$23,69578 \leq NA < 26,16006$	2
4	$NA \geq (\bar{x} + \sigma)$	$NA \geq 26,16006$	1

3. PEMBAHASAN

3.1 Proses Klasifikasi

Pada penelitian ini akan digunakan dua klasifikasi, yaitu metode Naive bayes dan K-Nearest Neighbor. Analisis klasifikasi dilakukan untuk mendefinisikan termasuk di dalam class mana data yang terbentuk menggunakan metode Naive bayes dan K-Nearest Neighbor. Pada kasus ini klasifikasi yang dilakukan menjadi 10 class yaitu seluruh sekolah SMP Negeri Kota Metro.

3.2 Klasifikasi menggunakan Naive Bayes

Pada tahap ini klasifikasi dilakukan dengan menggunakan data siswa PPDB online jenjang SMP Kota Metro dengan jumlah data sebanyak 1689 menggunakan 6 atribut yaitu Domisili, Asal Sekolah, Siswa Miskin, Siswa prestasi dan Nilai Akhir (NA) dimana NA diperoleh dari jumlah Total NUN ditambah poin Prestasi dan Gakin. Berikut gambaran penggunaan algoritma Naive bayes :

Tabel 4.2 Tabel PPDB SMP Kota Metro tahun 2015

Tabel 4.2 Tabel PPDB SMP Kota Metro tahun 2015

NO	DOMISILI	ASAL SEKOLAH	GAKIN	PRESTASI	NILAI AKHIR	NA	SEKOLAH
1	100	100	100	100	26,2	4	SMP N 1
2	200	200	100	100	26,2	4	SMP N 1
3	200	200	200	100	27,6	4	SMP N 2
206	200	200	200	100	27,6	4	SMP N 2
207	200	200	200	100	26,95	4	SMP N 2
208	200	200	200	100	31	4	SMP N 3
383	200	200	200	100	28,05	4	SMP N 3
384	200	200	200	100	27,6	4	SMP N 3
385	100	100	100	100	23,3	2	SMP N 3
562	100	100	100	700	29,7	4	SMP N 4
563	100	100	100	700	29,3	4	SMP N 4
564	200	200	200	100	28,9	4	SMP N 4
741	200	200	100	100	26,9	4	SMP N 5
742	100	100	200	100	26,7	4	SMP N 5
743	200	200	100	100	26,6	4	SMP N 5
920	200	200	200	100	29,95	4	SMP N 6

1. Menentukan Data Testing

Dari tabel diatas untuk melakukan testing dengan analogi yaitu :

a. Domisili : dalam kota kode : 100

b. Asal sekolah : dalam kota kode : 100

c. Gakin : Keluarga gakin Kode :100

d. Prestasi : Tidak Memiliki Prestasi

Kode : 100

e. Total Nilai Ujian Nasional : 22

Menghitung probabilitas dari data testing berdasarkan data training

$P(C) =$

Probabilitas (SMP NEGERI 1) = $205/1689 = 0,21374$

Probabilitas (SMP NEGERI 2) = $177/1689 = 0,104796$

Probabilitas (SMP NEGERI 3) = $179/1689 = 0,10598$

Probabilitas (SMP NEGERI 4) = $179/1689 = 0,10598$

Probabilitas (SMP NEGERI 5) = $179/1689 = 0,10598$

Probabilitas (SMP NEGERI 6) = $179/1689 = 0,10598$

$$\text{Probabilitas (SMP NEGERI 7)} = 153/1689 = 0,090586$$

$$\text{Probabilitas (SMP NEGERI 8)} = 128/1689 = 0,075784$$

$$\text{Probabilitas (SMP NEGERI 9)} = 182/1689 = 0,107756$$

$$\text{Probabilitas (SMP NEGERI 10)} = 128/1689 = 0,075784$$

Menghitung probabilitas atribut terhadap masing-masing kelas pada data testing berdasarkan data training.

Probabilitas Domisili

$$P(X1=100|SMP NEGERI 1) = 37/205 = 0,180487805$$

$$P(X1=100|SMP NEGERI 2) = 67/177 = 0,378531073$$

$$P(X1=100|SMP NEGERI 3) = 41/179 = 0,229050279$$

$$P(X1=100|SMP NEGERI 4) = 52/179 = 0,290502793$$

$$P(X1=100|SMP NEGERI 5) = 43/179 = 0,240223464$$

$$P(X1=100|SMP NEGERI 6) = 51/179 = 0,284916201$$

$$P(X1=100|SMP NEGERI 7) = 57/153 = 0,37254902$$

$$P(X1=100|SMP NEGERI 8) = 48/128 = 0,375$$

$$P(X1=100|SMP NEGERI 9) = 67/182 = 0,368131868$$

$$P(X1=100|SMP NEGERI 10) = 17/128 = 0,1328125$$

Probabilitas Asal Sekolah

$$P(X2=100|SMP NEGERI 1) = 37/205 = 0,180487805$$

$$P(X2=100|SMP NEGERI 2) = 67/177 = 0,378531073$$

$$P(X2=100|SMP NEGERI 3) = 41/179 = 0,229050279$$

$$P(X2=100|SMP NEGERI 4) = 52/179 = 0,290502793$$

$$P(X2=100|SMP NEGERI 5) = 43/179 = 0,240223464$$

$$P(X2=100|SMP NEGERI 6) = 51/179 = 0,284916201$$

$$P(X2=100|SMP NEGERI 7) = 57/153 = 0,37254902$$

$$P(X2=100|SMP NEGERI 8) = 48/128 = 0,375$$

$$P(X2=100|SMP NEGERI 9) = 67/182 = 0,368131868$$

$$P(X2=100|SMP NEGERI 10) = 37/128 = 0,1328125$$

Probabilitas Gakin

$$P(X3=200|SMP NEGERI 1) = 9/205 = 0,043902439$$

$$P(X3=200|SMP NEGERI 2) = 14/177 = 0,079096045$$

$$P(X3=200|SMP NEGERI 3) = 24/179 = 0,134078212$$

$$P(X3=200|SMP NEGERI 4) = 5/179 = 0,027932961$$

$$P(X3=200|SMP NEGERI 5) = 24/179 = 0,134078212$$

$$P(X3=200|SMP NEGERI 6) = 29/179 = 0,162011173$$

$$P(X3=200|SMP NEGERI 7) = 22/153 = 0,14379085$$

$$P(X3=200|SMP NEGERI 8) = 17/128 = 0,1328125$$

$$P(X3=200|SMP \text{ NEGERI } 9) = 15/182 = 0,082417582$$

$$P(X3=200|SMP \text{ NEGERI } 10) = 16/128 = 0,125$$

Probabilitas Prestasi

$$P(X4=100|SMP \text{ NEGERI } 1) = 181/205 = 0,882926829$$

$$P(X4=100|SMP \text{ NEGERI } 2) = 175/177 = 0,988700565$$

$$P(X4=100|SMP \text{ NEGERI } 3) = 176/179 = 0,983240223$$

$$P(X4=100|SMP \text{ NEGERI } 4) = 169/179 = 0,944134078$$

$$P(X4=100|SMP \text{ NEGERI } 5) = 179/179 = 1$$

$$P(X4=100|SMP \text{ NEGERI } 6) = 178/179 = 0,994413408$$

$$P(X4=100|SMP \text{ NEGERI } 7) = 153/153 = 1$$

$$P(X4=100|SMP \text{ NEGERI } 8) = 127/128 = 0,9921875$$

$$P(X4=100|SMP \text{ NEGERI } 9) = 182/182 = 1$$

$$P(X4=100|SMP \text{ NEGERI } 10) = 128/128 = 1$$

Probabilitas Nilai Akhir

$$P(X5=3|SMP \text{ NEGERI } 1) = 3/205 = 0,014634146$$

$$P(X5=3|SMP \text{ NEGERI } 2) = 163/177 = 0,920903955$$

$$P(X5=3|SMP \text{ NEGERI } 3) = 140/179 = 0,782122905$$

$$P(X5=3|SMP \text{ NEGERI } 4) = 138/179 = 0,770949721$$

$$P(X5=3|SMP \text{ NEGERI } 5) = 23/179 = 0,12849162$$

$$P(X5=3|SMP \text{ NEGERI } 6) = 63/179 = 0,351955307$$

$$P(X5=3|SMP \text{ NEGERI } 7) = 16/153 = 0,104575163$$

$$P(X5=3|SMP \text{ NEGERI } 8) = 17/128 = 0,1328125$$

$$P(X5=3|SMP \text{ NEGERI } 9) = 14/182 = 0,076923077$$

$$P(X5=3|SMP \text{ NEGERI } 10) = 19/128 = 0,1484375$$

Menghitung perkalian probabilitas dengan probabilitas atribut pada masing-masing kelas

$$P(X|C_i) = P(C_i) \prod_{k=1}^n P(x_k|c_i)$$

$$P(X|SMP \text{ NEGERI } 1) = 0,21374 * 0,180487805 * 0,180487805 * 0,043902439 * 0,882926829 * 0,014634146 = 0,000002243$$

$$P(X|SMP \text{ NEGERI } 2) = 0,104796 * 0,378531073 * 0,378531073 * 0,079096045 * 0,988700565 * 0,920903955 = 0,001081386$$

$$P(X|SMP \text{ NEGERI } 3) = 0,10598 * 0,229050279 * 0,229050279 * 0,134078212 * 0,983240223 * 0,782122905 = 0,000573295$$

$$P(X|SMP \text{ NEGERI } 4) = 0,10598 * 0,290502793 * 0,290502793 * 0,027932961 * 0,944134078 * 0,770949721 = 0,000181845$$

$$P(X|SMP \text{ NEGERI } 5) = 0,10598 * 0,240223464 * 0,240223464 * 0,134078212 * 1 * 0,12849162 = 0,000105363$$

$$P(X|SMP \text{ NEGERI } 6) = 0,10598 * 0,284916201 * 0,284916201 * 0,162011173 * 0,994413408 * 0,351955307 = 0,000487817$$

$$P(X|SMP NEGERI 7) = 0,090586 * 0,37254902 * 0,37254902 * 0,14379085 * 1 * 0,104575163 = 0,000189055$$

kode : 100

$$P(X|SMP NEGERI 8) = 0,075784 * 0,375 * 0,375 * 0,1328125 * 0,9921875 * 0,1328125 = 0,000186515$$

$$P(X|SMP NEGERI 9) = 0,107756 * 0,368131868 * 0,368131868 * 0,082417582 * 1 * 0,076923077 = 0,000092582$$

$$P(X|SMP NEGERI 10) = 0,075784 * 0,1328125 * 0,1328125 * 0,125 * 1 * 0,1484375 = 0,000024803$$

5) Mencari nilai maksimal probabilitas.

Setelah melakukan perhitungan dari data testing diatas diketahui probabilitas tertinggi SMP Negeri 2 dan probabilitas kedua pada SMP Negeri 3 Metro sehingga dari kasus diatas sekolah pilihan yang akan diambil adalah SMP Negeri 2 Metro dengan alternatif SMP Negeri 3 Metro.

3.3 Klasifikasi dengan K-Nearest Neighbor

struktur data, arsitektur perangkat lunak, representasi interface, dan detail algoritma prosedural.

Untuk klasifikasi yang kedua dilakukan tetap menggunakan data siswa PPDB online jenjang SMP Kota Metro dengan jumlah data sebanyak 1689 menggunakan atribut yaitu Domisili, Asal Sekolah, Siswa Miskin, Siswa prestasi dan Nilai Akhir (NA) Data Testing :

- a. Domilisi : dalam kota kode : 100
- b. Asal sekolah : dalam kota kode : 100
- c. Gakin : Keluarga gakin Kode :200

d. Prestasi : Tidak Memiliki Prestasi

No	DOMISILI	ASAL SEKOLAH	GAKIN	PRESTASI	NUN	SEKOLAH	
						Pilihan 1	Pilihan 2
1	200	200	100	700	27,8	SMP N 1 METRO	SMP N 4 METRO
2	200	200	100	400	27,7	SMP N 1 METRO	SMP N 4 METRO
12	100	100	100	100	28	SMP N 1 METRO	SMP N 4 METRO
13	200	200	200	100	24,95	SMP N 1 METRO	SMP N 3 METRO
14	100	100	100	100	27,8	SMP N 1 METRO	SMP N 4 METRO
15	200	200	100	400	25	SMP N 1 METRO	SMP N 4 METRO
16	100	100	100	100	27,55	SMP N 1 METRO	SMP N 4 METRO
17	200	200	200	100	24,45	SMP N 1 METRO	SMP N 3 METRO
18	200	200	200	100	26,15	SMP N 1 METRO	SMP N 3 METRO
19	100	100	100	100	25,15	SMP N 2 METRO	SMP N 4 METRO
20	100	100	100	100	25,05	SMP N 2 METRO	SMP N 4 METRO
53	100	100	100	100	26,85	SMP N 1 METRO	SMP N 4 METRO
54	200	200	200	100	23,8	SMP N 1 METRO	SMP N 3 METRO
55	100	100	100	100	26,7	SMP N 1 METRO	SMP N 4 METRO
111	100	100	100	100	22,05	SMP N 7 METRO	SMP N 8 METRO
112	100	100	100	100	22	SMP N 7 METRO	SMP N 8 METRO
115	200	200	100	100	20,8	SMP N 1 METRO	SMP N 5 METRO
116	100	100	100	100	20,8	SMP N 9 METRO	SMP N 8 METRO
117	200	200	200	100	17,8	SMP N 5 METRO	SMP N 9 METRO
118	200	200	100	100	20,7	SMP N 9 METRO	SMP N 10 METRO
119	200	200	200	100	17,65	SMP N 5 METRO	SMP N 9 METRO
120	100	100	100	100	20,6	SMP N 9 METRO	SMP N 8 METRO
127	100	100	100	100	19,5	SMP N 9 METRO	SMP N 8 METRO
128	200	200	100	100	19,25	SMP N 9 METRO	SMP N 10 METRO
129	200	200	100	100	19,05	SMP N 9 METRO	SMP N 10 METRO
130	200	200	100	100	18,95	SMP N 9 METRO	SMP N 10 METRO
131	200	200	100	100	18,85	SMP N 9 METRO	SMP N 10 METRO
132	200	200	100	100	18,75	SMP N 9 METRO	SMP N 10 METRO
133	200	200	100	100	18,65	SMP N 9 METRO	SMP N 10 METRO
162	200	200	200	100	24,65	SMP N 1 METRO	SMP N 3 METRO
163	200	200	200	100	23,4	SMP N 1 METRO	SMP N 3 METRO
164	200	200	200	100	19,2	SMP N 10 METRO	SMP N 6 METRO

e. Total Nilai Ujian Nasional :22

Menghitung jarak data tersting terhadap jarak setiap data peserta UN dengan menggunakan Euclidean distance :

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2}$$

D1=

$$\sqrt{(200 - 100)^2 + (200 - 100)^2 + (100 - 200)^2 + (700 - 100)^2 + (30,8 - 25)^2} = 600,0280327$$

D2=

$$\sqrt{(100 - 100)^2 + (100 - 100)^2 + (200 - 200)^2 + (400 - 100)^2 + (30,02 - 25)^2} = 300,0450633$$

D100=

$$\sqrt{(100 - 100)^2 + (100 - 100)^2 + (200 - 200)^2 + (100 - 100)^2 + (27,05 - 25)^2} = 2,05$$

D300=

$$\sqrt{(100 - 100)^2 + (100 - 100)^2 + (200 - 200)^2 + (100 - 100)^2 + (24,05 - 25)^2} = 0,45$$

D500=

$$\sqrt{(100 - 100)^2 + (100 - 100)^2 + (200 - 200)^2 + (100 - 100)^2 + (24,1 - 25)^2} = 0,9$$

Dari perhitungan diatas dicari data yang paling dekat dengan data testing sehingga

No	DOMISILI	ASAL SEKOLAH	GAKIN	PRESTASI	NUN	SEKOLAH	
						Pilihan 1	Pilihan 2
1	200	200	100	700	27,8	SMP N 4 METRO	SMP N 1 METRO
2	200	200	100	400	27,7	SMP N 1 METRO	SMP N 2 METRO
3	200	200	100	400	27,55	SMP N 1 METRO	SMP N 4 METRO
4	200	200	200	100	26,6	SMP N 1 METRO	SMP N 4 METRO
30	100	100	100	100	23,85	SMP N 3 METRO	SMP N 2 METRO
31	200	200	200	100	28	SMP N 1 METRO	SMP N 4 METRO
32	200	200	200	100	25,05	SMP N 1 METRO	SMP N 4 METRO
33	200	200	200	100	24,6	SMP N 1 METRO	SMP N 4 METRO
34	200	200	200	100	24,35	SMP N 1 METRO	SMP N 4 METRO
39	200	200	200	100	23,4	SMP N 1 METRO	SMP N 4 METRO
40	200	200	200	100	22,9	SMP N 2 METRO	SMP N 3 METRO
41	200	200	200	100	22,75	SMP N 2 METRO	SMP N 3 METRO
56	100	100	100	100	26,55	SMP N 1 METRO	SMP N 3 METRO
57	100	100	100	100	26,4	SMP N 3 METRO	SMP N 1 METRO
58	100	100	100	100	26,35	SMP N 3 METRO	SMP N 1 METRO
59	100	100	100	100	26,2	SMP N 3 METRO	SMP N 1 METRO
60	100	100	100	100	26,15	SMP N 3 METRO	SMP N 2 METRO
61	100	100	100	100	25,85	SMP N 2 METRO	SMP N 3 METRO
62	200	200	100	400	23,35	SMP N 4 METRO	SMP N 1 METRO
63	100	100	100	100	25,8	SMP N 2 METRO	SMP N 8 METRO
83	200	200	200	100	25,45	SMP N 1 METRO	SMP N 3 METRO
144	200	200	100	100	17,4	SMP N 10 METRO	SMP N 9 METRO
145	200	200	100	100	17,35	SMP N 10 METRO	SMP N 9 METRO

didapatkan bahwa jarak terdekat data testing terhadap data berada pada D259 yaitu data pada SMP Negeri 2 Metro sebagai pilihan pertama dan alternatif D440 sekolah pilihan adalah SMP Negeri 3 Metro.

Untuk lebih melihat cara kerja kedua metode dilakukan pengujian dengan menggunakan 175 data testing yang diambil secara acak dari data trying masing – masing sekolah.

Tabel 2.3 Penggunaan algoritma Naive Bayes pada PPDB SMP Kota Metro .:

Tabel 2.4 Penggunaan algoritma K-Nearest Neighbor pada PPDB SMP

Kota Metro .:

4. KESIMPULAN

Berdasarkan pembahasan yang telah diuraikan, tentang menentukan pilihan sekolah didalam penerimaan Peserta Didik

Baru dengan menggunakan metode Naive Bayes dan K Nearest Neighbor dan telah dilakukan testing dengan menggunakan 175 data maka disimpulkan beberapa hal sebagai berikut :

1. Dengan menggunakan algoritma Naive Bayes maupun K Nearest Neighbor menghasilkan 159 data prediksi yang mendekati sama hanya berbeda antara pilihan pertama dan pilihan kedua.
2. Dengan menggunakan algoritma Naive Bayes maupun K Nearest Neighbor didapatkan 16 data yang memiliki prediksi sama antara pilihan pertama dan pilihan kedua.

PUSTAKA

- [1] Connolly, dan Begg. 2005. Database Systems: A Practical Approach To Design,Implementation And Management.PearsonEducation.Amerika: Addison wesley
- [2] Duston, Yager.2008. Biometric System and data Analysis: Design, Evaluion, and Data Mining.New York:Springer
- [3] Fajar, Sandi 2013. Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification, IJCCS, Vol.7, No.1, January 2013, pp. 13~22 ISSN: 1978-1520 yogyakarta : FMIPA UGM
- [4] F. Gorunescu, 2011.Data Mining Concept Model Technique, Craiova, Romania: Springer.
- [5] Guha, S., Rastogi, R., Shim, K., 1997, CURE:An Efficient Clustering Algorithmfor Large Database, Korea Advanced Institute of Science and Technology: Taejon

- [6] Jiawei, H., Kamber, M. 2001. Data Mining Concepts and Techniques, Amerika: Morgan Kaufmann Publishers. 2354-5771 Makasar:STMIK Dipanegara Makasar
- [7] Kusrini & Emha Taufiq Luthfi.2009.Algoritma Data Mining. Yogyakarta:Andi Offset.
- [8] Mining. Canada:JohnWilley & Sons, Inc.Pang-Ning Tan, Michael Steinbach, Vipin Kumar.(2006). Introduction to Data Mining.Boston: Pearson Education, Inc.
- [9] Poniah,Pluraj.2011.Data Warehousing Fundamentals, New York:John Wiley & Sons, Inc.
- [10] Prasetyo, E., 2012.Data Mining: Konsep dan Aplikasi menggunakan MATLAB, Yogyakarta.: Andi,
- [11] Rismawan ,Tedy 2008.Sistem Pendukung Keputusan Berbasis Pocket PC Sebagai Penentu tatus Gizi Menggunakan Metode KNN (K-Nearest Neighbor),volume 13, nomor 2, desember 2008, 18-23, issn: 0853-8697 yogyakarta: Universitas Islam Indonesia
- [12] Santosa, B. 2007. Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis.Yogyakarta: Graha Ilmu.
- [13] Syukri, Muhammad.2014. Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus: Data Akademik Mahasiswa STMIK Dipanegara Makassar). ISSN:
- [14] Wu X, Kumar V. The Top Ten Algorithms in Data Mining. New York: CRC Press;2009