

PREDIKSI LAMA STUDI MAHASISWA DENGAN MENGGUNAKAN METODE K-NN PREDICTION OF STUDY TIMING PERIOD USING K-NN METHOD

Deserius Marianus Oenuu¹, Naniek Widyastuti², Uning Lestari³

^{1,2,3}Teknik Informatika, FTI, Institut Sains & Teknologi AKPRIND
semyoenuu@yahoo.co.id¹, naniek_wid@yahoo.com², uning@akprind.ac.id

ABSTRACT

Quality of a university can be seen from study timing period. Development of data mining can be used to make predictions of study timing period. These research predictions of study timing period using data mining. The method used to make predictions is K-Nearest Neighbour. A total 216 students will be predicted into 3 classes: Very On Time, On Time and Not On Time. From the experimental results, the obtained value of the highest accuracy reaches 80% when the data training 186 with a value of $k=30$

Keywords: Data Mining, Classification, K-NN, Prediction of Study Timing Period.

INTISARI

Kualitas suatu lembaga perguruan tinggi dapat dilihat dari waktu studi mahasiswa. Perkembangan data mining dapat digunakan untuk melakukan prediksi lama studi. Penelitian ini melakukan prediksi lama studi menggunakan data mining. Metode yang digunakan untuk melakukan prediksi adalah K-NN. Sebanyak 216 data mahasiswa diprediksi kedalam 3 kelas yaitu Sangat Tepat Waktu (STW), Tepat Waktu (TW) dan Tidak Tepat Waktu (TTW). Dari hasil percobaan diperoleh nilai akurasi tertinggi mencapai 80% ketika data latih 186 dengan nilai $k = 30$.

Kata kunci: Data Mining, Klasifikasi, k-NN, Prediksi Lama Kelulusan Mahasiswa.

PENDAHULUAN

Waktu studi mahasiswa merupakan salah satu faktor yang menentukan kualitas dari suatu lembaga pendidikan tinggi. Data menentukan bahwa jumlah mahasiswa baru lebih banyak dibandingkan dengan mahasiswa yang lulus. Lembaga pendidikan tinggi perlu melakukan analisa terhadap hasil belajar mahasiswa sehingga dapat memprediksi mahasiswa yang mungkin bermasalah. Hasil prediksi dapat digunakan sebagai peringatan dini sehingga dapat dicari solusi untuk membantu mahasiswa lulus tepat waktu. Dalam penelitian ini akan dilakukan prediksi lama studi mahasiswa terhadap atribut yang paling berpengaruh terhadap tingkat kelulusan studi mahasiswa dan melakukan pengelompokan terhadap sejumlah mahasiswa ke dalam *cluster* berdasarkan kemiripan atribut. Dari hasil tersebut dapat ditentukan mahasiswa yang dapat diprediksi lama kelulusannya dengan kriteria sangat tepat waktu, tepat waktu ataupun gagal menyelesaikan studi di perguruan tinggi.

Pokok masalah yang dikaji dalam penelitian ini adalah bagaimana model prediksi waktu studi mahasiswa menggunakan metode K-Nearest Neighbour dan bagaimana tingkat akurasi dari model prediksi yang dikembangkan.

Penelitian ini bertujuan untuk menerapkan metode K-NN dan mengukur kinerja algoritma K-NN dalam prediksi masa studi mahasiswa.

Hasil penelitian ini diharapkan dapat digunakan untuk memprediksi mahasiswa yang memiliki kemungkinan gagal menyelesaikan studi sehingga dapat dicari solusinya.

Data diambil dari database akademik mahasiswa jurusan Teknik Informatika S1, Institut Sains & Teknologi AKPRIND Yogyakarta (Tahun Akademik 2004 s/d 2009 yang telah lulus) sebanyak 216 mahasiswa dengan 34 atribut diantaranya Jenis Kelamin, Usia Saat Mendaftar, Agama, Asal Sekolah, Pekerjaan Orang Tua, IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, Jumlah SKS Semester 1, Jumlah SKS Semester 2, Jumlah SKS Semester 3, Jumlah SKS Semester 4, Nilai Setiap Mata Kuliah Semester 1, Nilai Setiap Mata Kuliah Semester 2, Nilai Setiap Mata Kuliah Semester 3, Nilai Setiap Mata Kuliah Semester 4.

Menentukan prediksi lama studi mahasiswa dengan hasil prediksi sebagai berikut : Studi mahasiswa <5 Tahun dinyatakan “Sangat Tepat Waktu”, Studi mahasiswa > 5s/d6 Tahun dinyatakan “Tepat Waktu”, Studi mahasiswa >6 Tahun dinyatakan “Tidak Tepat Waktu”

Sambani dkk (2014), mengklasifikasi lama kelulusan menggunakan algoritma *Backpropagation*. Tujuannya adalah untuk menguji keakuratan *Artificial Neural Network (ANN)* dengan menggunakan algoritma *Backpropagation* serta meningkatkan akurasi dalam memprediksi kelulusan mahasiswa dengan menggunakan metode ANN sehingga dapat diketahui prediksi kelulusan mahasiswa dan meningkatkan akurasi prediksi kelulusan mahasiswa. Hadjratie (2014) dan Hanafidan Burhanudin (2014) juga menggunakan metode ANN yang dapat dilatih untuk mengenali pola Indeks Prestasi dan melakukan generalisasi terhadap pola yang dilatihkan tersebut. Dari proses tersebut, ANN mampu mengenali pola dan memprediksi keberhasilan studi dengan tingkat akurasi mencapai 85% sampai dengan 90%.

Jananto (2013), dengan menggunakan teknik data mining khususnya klasifikasi untuk prediksi dengan algoritma *Naive Bayes* dapat dilakukan prediksi terhadap ketepatan waktu studi dari mahasiswa berdasarkan data training yang ada. Fungsi untuk prediksi ini dibuat dengan menggunakan *Query* pada *MySql* dalam bentuk *function*. Dengan teknik data mining tersebut algoritma *Naive Bayes* mampu memprediksi keberhasilan studi mahasiswa dengan tingkat akurasi 80%.

Jananto (2010), dikatakan bahwa algoritma *Nearest Neighbour (NN)* lebih sederhana dan membutuhkan waktu yang lebih lama untuk melakukan proses trining secara berulang untuk setiap data yang diprediksi. Dalam penelitian ini, dia membandingkan antara algoritma SLIQ dengan algoritma NN. Algoritma SLIQ dapat mengklasifikasikan dan memprediksi kinerja mahasiswa dengan hasil tingkat akurasi yang masih rendah, hal ini dikarenakan banyaknya pembatasan pada jumlah kemunculan atribut. NN mampu memprediksi kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada dengan ketepatan akurasi 88%.

PEMBAHASAN

Model prediksi hasil tahap perancangan diimplementasikan menggunakan bahasa pemrograman Delphi 7. Komponen pada model prediksi diimplementasikan dalam bentuk prosedur. Berikut ini hasil implementasi dari prosedur utama sistem.

- Load Data, data yang digunakan dalam percobaan dapat dipilih melalui combobox. Setelah data dipilih, untuk menjalankan proses load digunakan tombol Load Data. Kode program yang digunakan seperti pada kode program 1.

Program 1

```
//load data dari database --> variabel data
maxData := adoquery1.RecordCount;
for idData := 1 to maxData do
begin
DataAsli[idData].nim := adoquery1.Fields.Fields[0].AsString;
DataAsli[idData].kelas := adoquery1.Fields.Fields[35].AsFloat;
for idxAtrib := 1 to 34 do
DataAsli[idData].atrib[idxAtrib]:=
adoquery1.Fields.Fields[idxAtrib].AsFloat;
adoquery1.Next;
end;
```

- Perhitungan Jarak, dilakukan dengan konsep *Euclidean₁*. Dari konsep tersebut maka akan peroleh hasil perhitungan antara jarak data yang akan diuji terhadap data yang dilatih. Untuk menjalankan proses hitung jarak maka kode program yang digunakan seperti tampak pada program 2.
$$d(vu, vl) = \sqrt{\sum_{k=1}^m (vu - vl_k)^2} \dots \dots \dots (1)$$

Program2

```
function hitungJarak(dtUji, dtLatih: array of real): real;
var
```

```

i,j: word;
tmpJum,hasil : real;
begin
tmpJum:= 0;
for i:= 1 to 34 do
tmpJum:= tmpJum + sqr(dtUji[i] - dtLatih[i]);
hasil:= sqrt(tmpJum);
hitungJarak:= hasil;
end;

```

- Urut Data Jarak, setelah mendapat hasil jarak antara V_{uji} terhadap V_{latih} , maka akan diurutkan data jarak dari nilai V_{latih} yang terkecil sampai dengan nilai yang terbesar. Untuk menjalankan proses urut jarak maka kode program yang digunakan seperti tampak pada program 3.

Program 3

```

procedure urut(maks:word;var dtJarak: array of TJarak);
var
i,j,k: word;
tmp : Tjarak;
begin
for i:= 0 to maks-2 do
begin
k:= i;
for j:= i+1 to maks-1 do
if (dtJarak[k].jarak > dtJarak[j].jarak) then
k := j;
tmp.nim:= dtJarak[i].nim;
tmp.kelas:= dtJarak[i].kelas;
tmp.jarak:= dtJarak[i].jarak;

dtJarak[i].nim:= dtJarak[k].nim;
dtJarak[i].kelas:= dtJarak[k].kelas;
dtJarak[i].jarak:= dtJarak[k].jarak;

dtJarak[k].nim:= tmp.nim;
dtJarak[k].kelas:= tmp.kelas;
dtJarak[k].jarak:= tmp.jarak;
end;
end;

```

- Hitung Kelas, menghitung kemunculan masing-masing kelas. Untuk menjalankan proses hitung kelas maka kode program yang digunakan dapat dilihat pada program 4.

Program 4

```

for i:=0 to k-1 do
begin
if (jarakTmp[i].kelas = 0) then
jmlKls1:= jmlKls1 + 1
else
if (jarakTmp[i].kelas = 1) then
jmlKls2:= jmlKls2 + 1
else
jmlKls3:= jmlKls3 + 1;

```

end;

- Penentuan Kelas, menerima masukan dari hasil perhitungan kemunculan kelas. Untuk menjalankan proses hitung kelas maka kode program yang digunakan dapat dilihat pada program 5.

Program 5

```
function kelas(a,b,c: word):String;
var hasil: string;
begin
  if (a > b) then
    if (a > c) then
      hasil:='STW'
    else
      hasil:= 'TTW'
    else
      if (b > c) then
        hasil:='TW'
      else
        hasil:='TTW';
        kelas:= hasil;
      end;
    end;
```

- Prediksi Benar yaitu proses prediksi terhadap data uji. Pada setiap proses prediksi nilai k akan diinput dari k = 1 hingga maksimal k latih. Untuk menjalankan hasil prediksi maka kode program yang digunakan dapat dilihat pada program 6.

Program 6

```
jumBenar=0;
for i:= 1 to maksUji do
begin
  if (DataHasil[i-1].kelas = DataHasil[i-1].prediksi) then
    jumBenar:= jumBenar + 1;
  end;
```

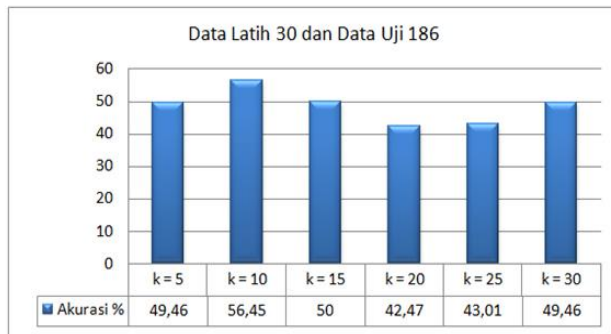
- Akurasi, Laporan dari proses pengujian adalah tingkat akurasi dari setiap percobaan untuk setiap nilai k. Untuk dapat menghitung akurasi maka kode program yang digunakan dapat dilihat pada program 7.

Program 7

```
akurasi:= (jumBenar/maksUji) * 100;
lbBenar.Caption := intToStr(jumBenar);
lbAkurasi.Caption := formatfloat('#,##0.00',akurasi)+' %';
```

Dalam penelitian ini, penulis menggunakan 8 data percobaan dengan masing-masing data percobaan diberi nilai k = 5, k = 10, k = 15, k = 20, k = 25 dan k = 30. Tujuannya adalah untuk mengetahui akurasi kinerja algoritma K-NN.

- Percobaan I Data Latih 30 dan Data Uji 186

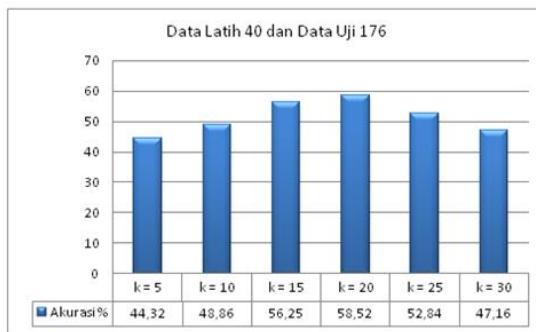


Nilai k	Akurasi (%)	Nilai k	Akurasi (%)
1	41,94	16	51,61
2	43,55	17	49,46
3	47,31	18	44,62
4	45,70	19	48,92
5	49,46	20	42,47
6	48,92	21	45,70
7	51,61	22	45,16
8	48,92	23	47,85
9	56,45	24	45,16
10	56,45	25	43,01
11	54,30	26	39,78
12	54,30	27	39,78
13	56,45	28	40,32
14	52,69	29	29,03
15	50,00	30	49,46

Gambar 1.Data Latih 30 dan Data Uji 186

Hasil percobaan I yang tampak pada gambar 1.menunjukkan bahwa akurasi maksimal diperoleh ketika nilai k = 10 yaitu sebesar 56,45%. Terjadi penurunan akurasi ketika nilai k ≤ 8, k = 11, k = 12 dan k ≥ 14 diantaranya: k = 5 yaitu sebesar 49,46%, k = 15 yaitu sebesar 50%, k = 20 yaitu sebesar 42,47%, k = 25 yaitu sebesar 43,01 dan k = 30 yaitu sebesar 49,46%. Dari hasil percobaan I, ketepatan akurasi maksimal mencapai 56,45% dengan nilai k=10. Pada tabel 4.6.akurasi maksimal 56,45% lainnya terdapat pada k = 9 dan 13.

- Percobaan II Data Latih 40 dan Data Uji 176

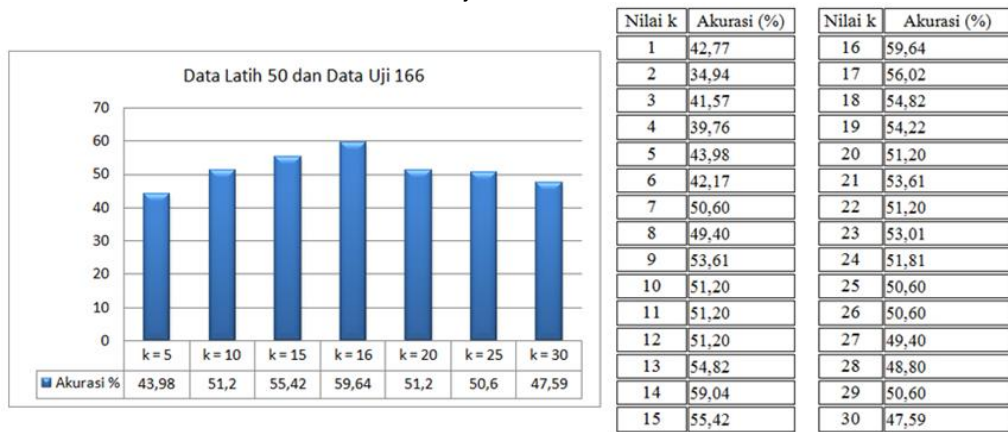


Nilai k	Akurasi (%)	Nilai k	Akurasi (%)
1	38,64	16	57,95
2	34,66	17	53,98
3	42,05	18	53,41
4	41,48	19	57,39
5	44,32	20	58,52
6	41,48	21	55,68
7	44,32	22	55,68
8	46,02	23	55,68
9	47,16	24	53,41
10	48,86	25	52,48
11	52,84	26	49,43
12	50,57	27	49,43
13	53,98	28	56,59
14	55,11	29	47,73
15	56,25	30	47,16

Gambar 2.Data Latih 40 dan Data Uji 176

Hasil percobaan II pada gambar 2.menunjukkan bahwa akurasi maksimal diperoleh ketika nilai k = 20 yaitu sebesar 58,52%. Terjadi penurunan akurasi ketika nilai k ≤ 19 dan k ≥ 21 diantaranya: k = 5 yaitu sebesar 44,32%, k = 10 yaitu sebesar 48,86%, k=15 yaitu sebesar 56,25%, k = 25 yaitu sebesar 52,84%, dan k = 30 yaitu sebesar 47,16%. Dari hasil percobaan II, ketepatan akurasi maksimal mencapai 58,52% dengan nilai k = 20.

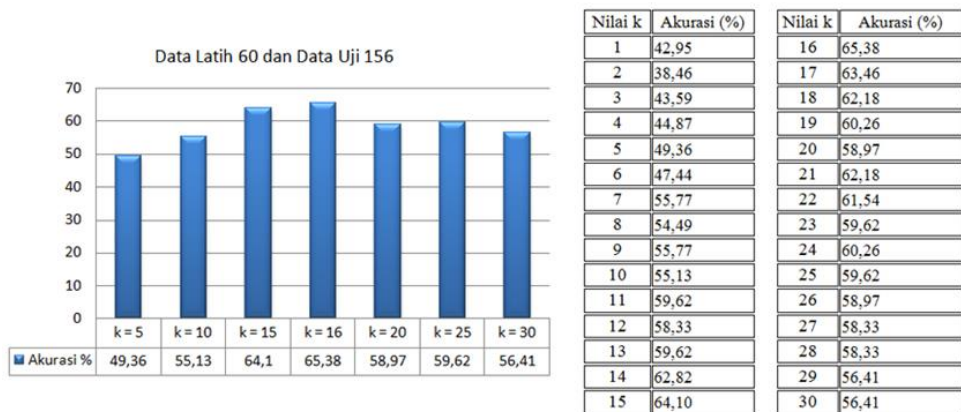
- Percobaan III Data Latih 50 dan Data Uji 166



Gambar 3. Data Latih 50 dan Data Uji 166

Hasil percobaan III yang tampak pada gambar 3. menunjukkan bahwa akurasi maksimal diperoleh ketika nilai k = 16 yaitu sebesar 59,04%. Terjadi penurunan akurasi ketika nilai k ≤ 15 dan k ≥ 17 diantaranya: k = 5 yaitu sebesar 43,98%, k = 10 yaitu sebesar 51,20%, k = 20 yaitu sebesar 51,20%, k = 25 yaitu sebesar 50,60%, dan k = 30 yaitu sebesar 47,59%. Dari hasil percobaan III, ketepatan akurasi maksimal mencapai 59,64% dengan nilai k = 16.

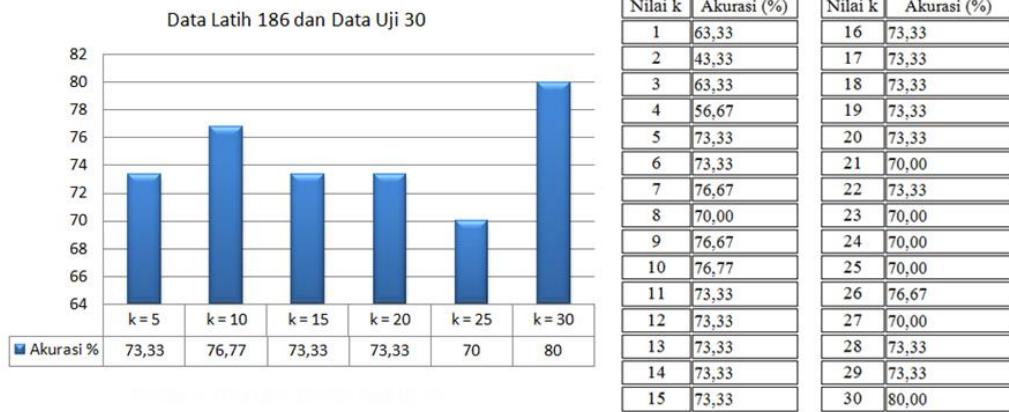
- Percobaan IV Data Latih 60 dan Data Uji 156



Gambar 4. Data Latih 60 dan Data Uji 156

Hasil percobaan IV yang tampak pada gambar 4. menunjukkan bahwa akurasi maksimal diperoleh ketika nilai k = 16 yaitu sebesar 65,38%. Terjadi penurunan akurasi ketika nilai k ≤ 15 dan k ≥ 17 diantaranya : k = 5 yaitu sebesar 49,36%, k=10 yaitu sebesar 55,13%, k = 15 yaitu sebesar 64,10%, k = 20 yaitu sebesar 58,97%, k=25 yaitu sebesar 59,62%, dan k = 30 yaitu sebesar 56,41%. Dari hasil percobaan IV, ketepatan akurasi maksimal mencapai 65,38% dengan nilai k = 16.

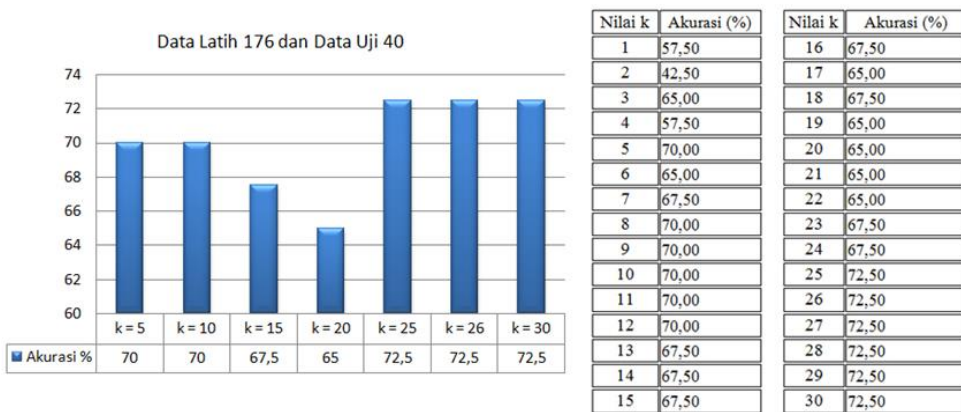
- Percobaan V Data Latih 186 dan Data Uji 30



Gambar 5.Data Latih 186 dan Data Uji 30

Hasil percobaan V yang tampak pada gambar 5.menunjukkan bahwa akurasi maksimal diperoleh ketika nilai k = 30 yaitu sebesar 80%. Terjadi penurunan akurasi ketika nilai $k \leq 29$ dan $k \geq 31$ diantaranya : k = 5 yaitu sebesar 73,33%, k = 10 yaitu sebesar 76,77%, k = 15 yaitu sebesar 73,33%, k = 20 yaitu sebesar 73,33%, dan k=25 yaitu sebesar 70%. Dari hasil percobaan V, ketepatan akurasi maksimal mencapai 80% dengan nilai k = 30. Akurasi dengan 80% lainnya terdapat pada k=33 dan k = 35 s/d 48.

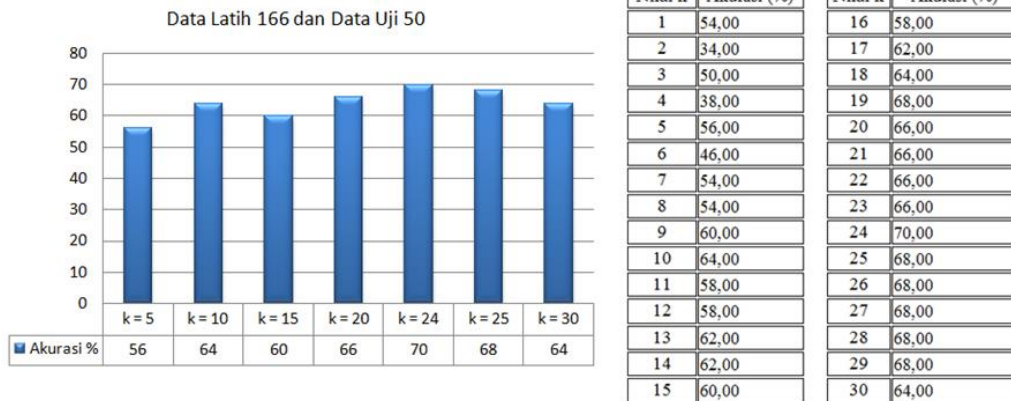
- Percobaan VI Data Latih 176 dan Data Uji 40



Gambar 6. Data Latih 176 dan Data Uji 40

Hasil percobaan VI yang tampak pada gambar 6.menunjukkan bahwa akurasi maksimal diperoleh ketika nilai k = 25 sampai 30 yaitu sebesar 72,50%. Terjadi penurunan akurasi ketika nilai $k \leq 24$ diantaranya : k = 5 dan 10 yaitu sebesar 70%, k = 15 yaitu sebesar 67,50%, k = 20 yaitu sebesar 65%. Dari hasil percobaan VI, ketepatan akurasi maksimal mencapai 72,50% dengan nilai k = 25 sampai 30.

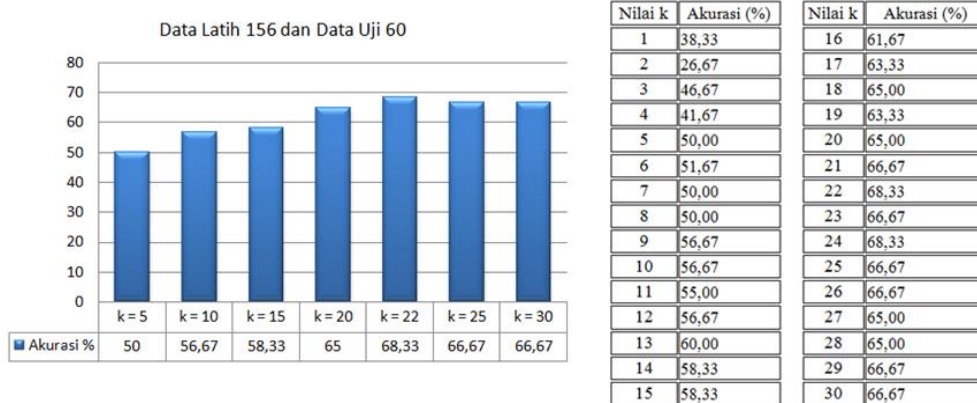
- Percobaan VII Data Latih 166 dan Data Uji 50



Gambar 7. Data Latih 166 dan Data Uji 50

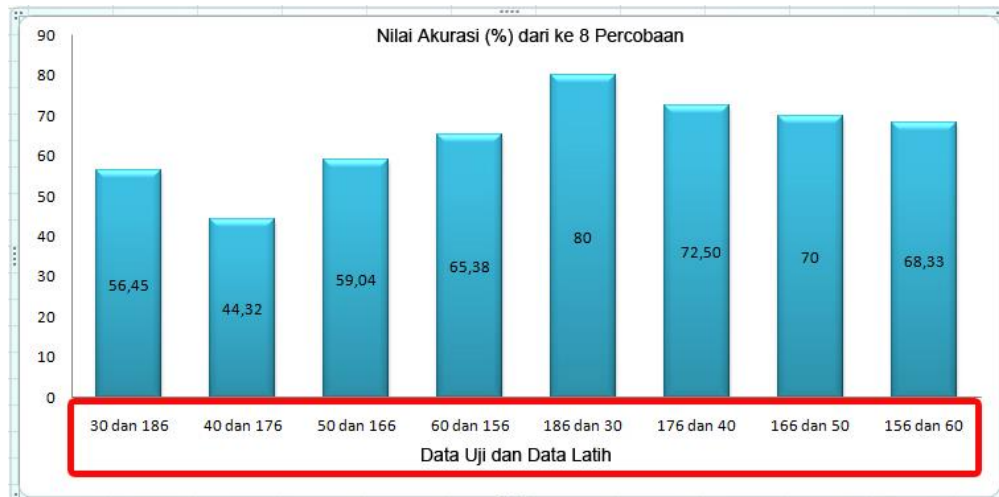
Hasil percobaan VII yang tampak pada gambar 7. menunjukkan bahwa akurasi maksimal diperoleh ketika nilai $k = 24$ yaitu sebesar 70%. Terjadi penurunan akurasi ketika nilai $k \leq 23$ dan $k \geq 25$ diantaranya : $k = 5$ yaitu sebesar 56%, $k = 10$ yaitu sebesar 64%, $k = 15$ yaitu sebesar 60%, $k = 20$ yaitu sebesar 66%, $k = 25$ yaitu sebesar 68% dan $k = 30$ yaitu sebesar 64%. Dari hasil percobaan VII, ketepatan akurasi maksimal mencapai 70% dengan nilai $k = 24$. Akurasi dengan 70% lainnya terdapat pada $k = 34$ dan 35 .

- Percobaan VIII Data Latih 156 dan Data Uji 60



Gambar 8. Data Latih 156 dan Data Uji 60

Hasil percobaan VIII yang tampak pada gambar 8. menunjukkan bahwa akurasi maksimal diperoleh ketika nilai $k = 22$ yaitu sebesar 68,33%. Terjadi penurunan akurasi ketika nilai $k \leq 21$ dan $k \geq 25$ diantaranya : $k = 5$ yaitu sebesar 50%, $k = 10$ yaitu sebesar 56,67%, $k = 15$ yaitu sebesar 58,33%, $k = 20$ yaitu sebesar 65%, $k = 25$ dan 30 yaitu sebesar 66,67%. Dari hasil percobaan VIII, ketepatan akurasi maksimal mencapai 68,33% dengan nilai $k = 22$. Pada gambar 8. akurasi dengan 68,33% lainnya terdapat pada $k = 24$.



Gambar 9. Nilai Akurasi (%) dari ke 8 percobaan

Dari gambar 9. Dapat diketahui bahwa akurasi terbesar diperoleh ketika data latih 186 terhadap data uji 30 yaitu sebesar 80%, sedangkan akurasi terendah diperoleh ketika data latih 40 terhadap data uji 176 yaitu sebesar 44,32%.

KESIMPULAN

Dari ke 8 percobaan tersebut maka dapat disimpulkan sebagai berikut:

- Semakin besar jumlah data latih, maka semakin besar tingkat akurasi yang didapat. Ketika jumlah data latih sebanyak 30 akurasi yang diperoleh sebesar 56,45%. Ketika jumlah data latih ditambah sebanyak 186 maka akurasi meningkat menjadi 80%.
- Nilai k mempengaruhi tingkat akurasi dimana semakin besar nilai k maka akurasi semakin tinggi. Hal ini dapat dilihat pada delapan percobaan yang dilakukan dimana tingkat akurasi cenderung meningkat ketika Nilai $k > 16$.
- Nilai akurasi tertinggi hanya mencapai 80%. Hal ini disebabkan karena atribut Nilai IPK, Jumlah SKS dan Matakuliah yang digunakan adalah semester awal yaitu semester I sampai dengan semester IV.
- Penelitian ini dapat diimplementasikan oleh pengelola kampus sehingga dapat diambil tindakan preventif terhadap mahasiswa yang diprediksi kelulusannya *Tidak Tepat Waktu* (TTW).
- Selain itu, kiranya penelitian ini dapat dikembangkan oleh peneliti berikut kaitannya dengan prediksi lama studi mahasiswa dengan atribut tambahan semester V, IV dan VII agar supaya mendapat hasil penghitungan tingkat akurasi yang lebih baik lagi.
- Pada penelitian ini, program yang dihasilkan menggunakan tiga kelas dengan batasan waktu yang sudah ditentukan sejak awal. Pada penelitian selanjutnya agar program menjadi lebih dinamis maka penentuan jumlah kelas dan batasan waktu dapat ditentukan oleh pengguna.

DAFTAR PUSTAKA

- Hadjaratie, L., 2014, *Model Prediksi Masa Studi Mahasiswa Menggunakan Algoritma Projective Art*, Prosiding Konferensi Nasional Sistem Informasi 2014 (KNSI 2014), Februari – Maret 2014, hal. 1111 – 1114.
- Hanafi, M. dan Burhanudin, A., 2014, *Prediksi Keberhasilan Studi Mahasiswa Menggunakan Artificial Neural Network Berbasis Indeks Prestasi Akademik*, Prosiding Konferensi Nasional Sistem Informasi 2014 (KNSI 2014), Februari – Maret 2014, hal. 1085 – 1091.

- Jananto, A., 2010, *Perbandingan Performansi Algoritma Nearest Neighbour Dan Sliq Untuk Prediksi Kinerja Akademik Mahasiswa Baru*, Jurnal Teknologi Informasi DINAMIK, 15(2), Juli 2010, hal. 157 – 169.
- Jananto, A., 2013, *Algoritma Naive Bayes untuk Mencari Perkiraan Waku Studi Mahasiswa*, Jurnal Teknologi Informasi DINAMIK, 18(1), Januari 2013, hal. 9 – 16.
- Sambani, E.B., Saputra, R.D. dan Mulyani, E.D.S., 2014, *Prediksi Lama Kelulusan Mahasiswa Menggunakan Meetode Jaringan Syaraf Tiruan Study Case STMIK Tasikmalaya*, Prosiding Konferensi Nasional Sistem Informasi 2014 (KNSI 2014), Februari – Maret 2014, hal. 1071 – 1077.