

KLASIFIKASI MAHASISWA PENERIMA PROGRAM BEASISWA BIDIK MISI MENGGUNAKAN *NAIVE BAYES*

Yuniar Farida ¹⁾, Nurissaidah Ulinnuha ²⁾

^{1,2)} Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Ampel
Jl. Ahmad Yani 117 Surabaya

E-mail : yuniar_farida@uinsby.ac.id¹⁾, nuris.ulinnuha@uinsby.ac.id²⁾

Abstrak

Program beasiswa Bidik Misi merupakan program bantuan biaya pendidikan yang diberikan oleh Kementerian Riset Teknologi dan Pendidikan Tinggi kepada mahasiswa yang memiliki tingkat ekonomi menengah ke bawah namun memiliki potensi akademik yang baik. Universitas Islam Negeri Sunan Ampel Surabaya merupakan perguruan tinggi dengan jumlah pendaftar Bidik Misi yang terus meningkat setiap tahunnya namun hanya menerima dengan kuota yang sangat terbatas. Dengan banyaknya persyaratan dalam proses penyeleksian dan melihat banyaknya jumlah pendaftar, maka diperlukan sebuah sistem yang mampu membantu dalam proses penyeleksian mahasiswa pendaftar Bidik Misi dengan cepat, mudah dan tepat sasaran. Dalam penelitian ini, digunakan metode Naive Bayes classifier dengan menggunakan variabel penghasilan orang tua, prestasi akademik, prestasi non akademik, uang kuliah tunggal, daya listrik serta jumlah tanggungan keluarga. Keenam variabel tersebut digunakan dalam proses klasifikasi sebagai indikator penentuan kelayakan mahasiswa mendapatkan beasiswa. Sistem yang dibuat dengan metode Naive Bayes classifier sudah baik, dibuktikan dengan hasil akurasi hingga 83,33 %.

Kata Kunci: Bidik Misi, Parameter Klasifikasi, Naive Bayes Classifier

Abstract

Bidik Misi scholarship program is a tuition assistance program provided by the Ministry of Technology and Higher Education Research to students who have lower middle economic level but have good academic performance. Sunan Ampel State Islamic University Surabaya is a university with a number of Bidik Misi registrants that continues to increase every year but only receives a very limited quota. With so many requirements in the selection process and seeing the large number of registrants, we need a system that is able to assist in the selection process of Bidik Misi students quickly, easily and effectively. In this study, Naive Bayes classifier method was used by using variables of parental income, academic achievement, non-academic achievement, tuition, power-house electrical supply and the number of family dependents. The six variables are used in the classification process as an indicator of determining the eligibility of students to get a scholarship. The implemented system using Naive Bayes classifier method is good, proven by the accuracy up to 83.33%.

Keywords: Bidik Misi, Classification Parameters, Naive Bayes Classifier

1. PENDAHULUAN

Program Beasiswa pendidikan mahasiswa miskin berprestasi atau yang sering dikenal dengan sebutan Bidik Misi merupakan program bantuan biaya pendidikan yang diberikan oleh Kementerian Riset Teknologi dan Pendidikan Tinggi kepada calon mahasiswa yang memiliki tingkat ekonomi menengah ke bawah namun memiliki potensi akademik yang baik. Oleh karena itu, tidak sembarang mahasiswa yang bisa mendapatkan beasiswa tersebut. Hanya mahasiswa tertentu saja yang dinyatakan memenuhi persyaratan dan kriteria yang telah ditetapkan langsung oleh

Kementerian Riset Teknologi dan Pendidikan Tinggi sebagai mahasiswa yang berhak dan layak menerima Bidik Misi [1].

Universitas Islam Negeri (UIN) Sunan Ampel Surabaya merupakan salah satu perguruan tinggi keagamaan Islam negeri dengan pendaftar program beasiswa Bidik Misi yang terus meningkat setiap tahunnya [2], sedangkan kuota penerima beasiswa Bidik Misi di universitas ini sangat terbatas. Melihat indikator dalam penyeleksian berkas pengajuan beasiswa dengan beberapa macam kriteria, namun masih menggunakan cara manual dalam proses penyeleksiannya, maka pemilihan mahasiswa yang berhak menerima beasiswa Bidik

Misi berdasarkan kelengkapan data yang dikumpulkan terkesan objektif, kurang efektif dan membutuhkan waktu yang relatif lebih lama.

Dengan sistem pengolahan data secara manual dalam proses penyeleksian seperti yang terjadi sekarang, maka akan sangat sulit untuk menentukan siapa saja yang berhak dan layak mendapatkan Bidik Misi tersebut. Dengan demikian, diperlukan suatu metode khusus dan sistem tertentu yang dapat memberikan suatu keputusan yang tepat, efektif dan efisien dalam proses penyeleksian untuk menentukan penerima beasiswa Bidik Misi bagi seluruh pendaftar berdasarkan data yang masuk. Sistem ini diharapkan dapat membantu bagian kemahasiswaan UIN Sunan Ampel Surabaya dalam proses penyeleksian mahasiswa pendaftar beasiswa Bidik Misi dengan cepat dan tepat sasaran.

Naive Bayes merupakan salah satu metode yang memiliki tingkat keefektifan dan keefisienan yang cukup tinggi untuk diterapkan dalam pembelajaran mesin dan data mining [3]. Metode Naive Bayes ini, sangat cocok digunakan untuk menyelesaikan berbagai kasus klasifikasi. Berbagai problem masalah klasifikasi dalam kehidupan nyata, telah terbukti dapat terselesaikan dengan baik menggunakan Naive Bayes. Hal tersebut dikarenakan metode Naive Bayes ini memiliki performa klasifikasi yang tinggi [4]. Selain itu, metode Naive Bayes ini mampu mengakomodasi berbagai jenis data. Baik data kuantitatif maupun data diskrit dengan hanya membutuhkan data testing yang sedikit, namun mampu mengestimasi parameter yang dibutuhkan dalam proses klasifikasi dengan cepat dan efisien. Metode ini sangat mudah digunakan dalam proses pengklasifikasian. Karena pada data training, Naive Bayes membutuhkan *scan* hanya cukup satu kali saja dan dapat menangani dengan baik beberapa nilai atribut yang kosong. Selain itu, Naive Bayes juga sangat cocok digunakan untuk data yang berkelanjutan [5].

2. LANDASAN TEORI

2.1 Beasiswa Bidik Misi

Bidik Misi merupakan bantuan biaya pendidikan bagi calon mahasiswa yang tidak mampu secara ekonomi namun memiliki potensi akademik yang baik dengan tujuan agar dapat menempuh pendidikan di suatu perguruan tinggi hingga dapat lulus dengan tepat waktu. Pada tahun 2010, Pemerintah melalui Direktorat Jendral Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan mulai menjalankan program Bidik Misi ini untuk seluruh perguruan tinggi negeri di Indonesia. Kemudian, pada tahun 2012, Kementerian Agama melalui Direktorat

Pendidikan Tinggi Islam ikut serta memberikan program Bidik Misi untuk kalangan Mahasiswa Perguruan Tinggi Agama Islam Negeri (PTKIN). Adapun beberapa persyaratan yang harus dipenuhi untuk Mahasiswa calon penerima Bidik Misi adalah sebagai berikut [1]:

1. Mahasiswa merupakan lulusan dari MA/MAK/SMA/SMK sederajat pada tahun ajaran yang ditentukan
2. Mahasiswa harus memiliki prestasi akademik maupun non akademik yang baik, namun memiliki keadaan ekonomi menengah kebawah.
3. Jumlah penghasilan antara ayah dan ibu tidak lebih dari Rp.3.000.000/bulan.
4. Jumlah penghasilan setiap bulan antara ayah dan ibu ketika dibagi dengan banyaknya seluruh anggota keluarga tidak lebih dari Rp.750.000 /bulan
5. Memiliki potensi akademik maupun non akademik yang baik dibuktikan dengan surat rekomendasi dan keterangan siswa berprestasi dari sekolah asal.
6. Tidak terlibat dan tergabung kedalam suatu kegiatan maupun organisasi anti pancasila dan NKRI., yang dibuktikan dengan penandatanganan pakta integritas
7. Pendidikan orang tua setinggi-tingginya adalah S1 / Diploma
8. Memiliki kartu KPS, BSM dan sejenisnya

2.2 Klasifikasi

Klasifikasi merupakan pengelompokan beberapa parameter kedalam salah satu kategori yang sudah ditetapkan sebelumnya. Klasifikasi sering disebut dengan *supervised learning* hal tersebut disebabkan karena kelas yang digunakan dalam proses pengklasifikasian telah ditentukan sebelumnya [6]. Pada umumnya, klasifikasi dipakai untuk melakukan prediksi suatu label terhadap kelas tertentu. Hal tersebut dilakukan dengan membangun suatu model klasifikasi menggunakan sejumlah *data training*. Kemudian dari model klasifikasi yang dihasilkan, akan digunakan untuk mengklasifikasi data yang baru. Metode dalam data mining yang biasa digunakan dalam proses klasifikasi adalah C45, K-nearest *neighbor classification* (KNN), Naive Bayes *classification* dan pohon regresi (CART) dan masih banyak lainnya.

2.3 Naive Bayes Classifier

Naive Bayes merupakan salah satu teknik prediksi yang memanfaatkan teorema probabilistik sederhana yang bersumber dari pengaplikasian aturan bayes dengan pengansumsian tingkat independensi yang sangat kuat. Artinya bahwa sebuah parameter pada suatu data tidak bergantung

atau berpengaruh dengan keberadaan parameter-parameter lainnya dalam suatu data yang sama [7]–[9]. Penggunaan Naive Bayes, dilakukan dengan pengklasifikasian sejumlah data training secara tepat, efektif dan efisien [10]–[12].

Prediksi dengan menggunakan metode Naive Bayes Didasarkan pada teorema Bayes dengan formula umum sebagai berikut :

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \tag{1}$$

Keterangan :

$P(H|E)$ = Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika terdapat bukti E yang terjadi

$P(E|H)$ = Probabilitas sebuah bukti E yang terjadi akan mempengaruhi hipotesis H

$P(H)$ = Probabilitas awal (*priori*) hipotesis H terjadi tanpa melihat bukti apa pun
 $P(E)$ = Probabilitas awal (*priori*) bukti E terjadi tanpa melihat hipotesis / bukti yang lain.

Beberapa hal yang perlu diketahui terkait aturan bayes adalah sebagai berikut :

1. Probabilitas awal ($P(H)$) merupakan probabilitas dari suatu hipotesis sebelum bukti diamati
2. Probabilitas akhir H ($P(H|E)$) merupakan probabilitas dari suatu hipotesis setelah bukti diamati.

Kelebihan dari Teorema Bayes ini tidak hanya dapat menyelesaikan bukti tunggal saja. Namun juga dapat menyelesaikan lebih dari satu bukti lainnya. Misal, diketahui beberapa bukti E_1 , E_2 dan E_3 . Maka, probabilitas akhir dari Teorema Bayes dapat ditulis sebagai berikut :

$$P(H|E_1, E_2, E_3) = \frac{P(E_1, E_2, E_3|H) \times P(H)}{P(E_1, E_2, E_3)} \tag{2}$$

Dalam aturan bayes, hipotesis merupakan sebuah label kelas yang nantinya akan dipetakan kedalam beberapa kelas klasifikasi. Sedangkan bukti, merupakan beberapa parameter yang berperan penuh dalam proses pembentukan model klasifikasi. Misal, diketahui bahwa X merupakan sebuah vektor masukan yang didalamnya berisi beberapa fitur, Sedangkan Y adalah sebuah label kelas, maka aturan Naive Bayes dapat dinyatakan dengan $P(Y|X)$ atau biasa disebut dengan hasil probabilitas akhir untuk label kelas Y, sehingga $P(Y)$ merupakan probabilitas awal dari label kelas Y. Dalam notasi tersebut, dari beberapa parameter X sebagai vektor masukan yang diamati, maka akan didapatkan sebuah probabilitas label kelas Y. Dari model yang telah terbentuk, dapat diketahui bahwa suatu data uji X' dapat diklasifikasikan hanya dengan mencari nilai dari Y', hal tersebut dilakukan dengan memaksimalkan nilai dari $P(Y'|X')$ yang diperoleh. Sehingga, persamaan Naive Bayes yang telah diklasifikasi dapat ditulis

dengan :

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \tag{3}$$

Keterangan :

$P(Y|X)$ = Probabilitas data dengan vektor X pada kelas Y

$P(Y)$ = Probabilitas awal kelas Y

$\prod_{i=1}^q P(X_i|Y)$ = Probabilitas independen kelas Y dari semua fitur dalam vektor X

Dari persamaan diatas, nilai dari $P(X)$ akan terus tetap. Sehingga untuk menghitung nilai prediksi dalam beberapa waktu tertentu, peneliti hanya melakukan perhitungan pada bagian $P(Y) \prod_{i=1}^q P(X_i|Y)$ saja. Yaitu dengan mengambil nilai yang paling besar sebagai kelas yang nantinya akan dipilih sebagai nilai dari hasil prediksi. Sedangkan pada probabilitas independen $\prod_{i=1}^q P(X_i|Y)$ merupakan pengaruh dari semua parameter data terhadap setiap kelas Y. Oleh karena itu, dapat diformulasikan sebagai berikut :

$$P(X|Y = y) = \prod_{i=1}^q P(X_i|Y=y) \tag{4}$$

pada setiap set fitur $X = \{X_1, X_2, X_3, \dots, X_q\}$ terdiri dari q dimensi.

2.4 Validasi dan Evaluasi Hasil

Evaluasi dan validasi hasil klasifikasi dilakukan untuk mengetahui tingkat keakuratan model klasifikasi. Dalam penelitian ini, digunakan metode *confussion matrix* untuk mengevaluasi hasil dari klasifikasi yang didapat. Evaluasi menggunakan metode *confussion matrix* menghasilkan nilai *akurasi* [13].

Confussion Matrix menyatakan banyaknya data yang diklasifikasikan benar dan banyaknya data yang diklasifikasikan salah sesuai dengan data aktual. Tabel 1 merupakan contoh klasifikasi biner menggunakan *confussion matrix*.

Tabel 1
Contoh klasifikasi biner menggunakan *confussion matrix*

| | Kelas Hasil Prediksi | |
|-------------------|----------------------|-------|
| | 1 | 2 |
| Kelas Data Aktual | 1 | TP FN |
| | 2 | FP TN |

Keterangan :

TP (True Positive) : jumlah data dari kelas 1 yang benar dan diklasifikasikan kedalam kelas 1,

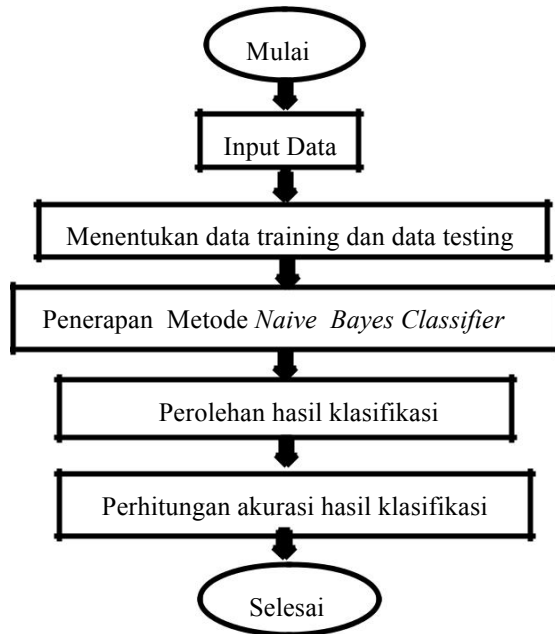
TN (True Negative) : jumlah data dari kelas 0 yang benar dan diklasifikasikan kedalam kelas 0,

FP (False Positive) : jumlah data dari kelas 0 yang salah dan diklasifikasikan kedalam kelas 1,

FN (False Negative) : jumlah data dari kelas 1 yang salah dan diklasifikasikan kedalam kelas 0.

3. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode Naive Bayes. Gambar 1 adalah diagram alir pemrosesan data dalam proses pengklasifikasian.



Gambar 1. Diagram Alir Pemrosesan Data

Langkah awal yang harus dilakukan adalah penginputan data. Dari data yang telah diinputkan tersebut, akan dilakukan pembagian data secara acak menjadi data *training* sebesar 75% dan data *testing* sebesar 25%. Selanjutnya, akan dilakukan *training* model data dengan menerapkan metode *Naive Bayes Classifier* dengan melakukan perhitungan probabilitas tiap-tiap kelas. Dari hasil perhitungan tersebut, akan didapatkan hasil prediksi dari data *testing*. Selanjutnya, untuk mengetahui tingkat keakurasian hasil pengklasifikasian, diterapkan metode *confusion matriks* dan melakukan perbandingan data aktual dengan data hasil prediksi.

3.1. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data yang didapatkan dari Badan Akademik dan Kemahasiswaan UIN Sunan Ampel Surabaya berupa data pendaftar Bidik Misi, baik pendaftar beasiswa yang dinyatakan lolos maupun yang dinyatakan tidak lolos tahun periode 2017/2018. Informasi data yang diperoleh, telah sesuai dengan syarat-syarat yang diajukan oleh Direktorat Jendral Pendidikan Tinggi bagi mahasiswa pendaftar Bidik Misi, yang terdiri atas :

1. Nama mahasiswa
2. Fakultas/jurusan
3. Penghasilan orang tua
4. Uang kuliah tunggal
5. Jumlah tanggungan keluarga
6. Daya listrik rumah (watt)
7. Prestasi akademik dan non akademik

Tabel 2
Sampel Data Penelitian

| Ukt | Penghasilan | Prestasi Akademik | Prestasi Non Akademik | Jumlah Keluarga | Listrik | Klasifikasi |
|--------------|-------------------------------|-------------------|-----------------------|-----------------|---------|-------------|
| Rp 2.100.000 | > 3.000.000 s.d. 5.000.000 | 6 | 0 | 3 | 900 | 0 |
| Rp 1.700.000 | Rp.500.001 - Rp.1.000.000 | 4 | 0 | 4 | 1200 | 1 |
| Rp 2.600.000 | < 1.000.000 | 6 | 5 | 5 | 450 | 1 |
| Rp 4.400.000 | 1.000.000 s.d. 2.000.000 | 0 | 1 | 4 | 450 | 1 |
| Rp 2.600.000 | < 1.000.000 | 6 | 2 | 5 | 900 | 1 |
| Rp 2.800.000 | 1.000.000 s.d. 2.000.000 | 5 | 1 | 3 | 900 | 1 |
| Rp 4.400.000 | > 3.000.000 s.d. 5.000.000 | 6 | 0 | 2 | 900 | 1 |
| Rp 400.000 | Rp. 500.001 - Rp. 1.000.000 | 0 | 1 | 3 | 450 | 1 |
| Rp 2.600.000 | < 1.000.000 | 4 | 4 | 4 | 900 | 1 |
| Rp 2.600.000 | > 2.000.000 s.d. 3.000.000 | 6 | 4 | 5 | 450 | 1 |
| Rp 3.700.000 | Rp. 3.000.001 - Rp. 4.000.000 | 1 | 0 | 6 | 450 | 1 |
| Rp 3.300.000 | > 2.000.000 s.d. 3.000.000 | 0 | 3 | 2 | 450 | 0 |

Data yang digunakan berjumlah 99 data. Dengan 28 data pendaftar yang tidak lolos dan 71 data dari pendaftar yang lolos. Dari 99 data tersebut, 75% diantaranya digunakan sebagai data *training* dan 25% diantaranya digunakan sebagai data *testing*. Variabel yang digunakan dalam penelitian ini terdiri dari 7 variabel. 6 variabel sebagai variabel faktor dan 1 variabel lainnya sebagai variabel kategori. 6 variabel faktor tersebut terdiri dari data penghasilan orang tua, uang kuliah tunggal (UKT), prestasi akademik, prestasi non akademik, jumlah tanggungan keluarga dan daya listrik (watt). 1 variabel kategori tersebut adalah variabel keputusan, yang terdiri dari 2 keputusan yaitu mahasiswa yang layak menerima Bidik Misi atau mahasiswa yang tidak layak menerima Bidik Misi. Tabel 2 merupakan sampel dari data penelitian yang digunakan.

4. HASIL DAN PEMBAHASAN

Tahap awal yang harus dilakukan adalah melakukan pembagian seluruh data menjadi data *training* dan data *testing* dengan pengambilan secara acak menggunakan perbandingan 3:1, sehingga diperoleh jumlah data *training* sebanyak 75 data dan data *testing* sebesar 24 data.

4.1 Perhitungan Probabilitas Tiap Kelas

Perhitungan probabilitas tiap kelas dapat dihitung dengan rumus seperti dibawah ini :

$$P(A_i) = \frac{\text{Banyak kelas } i}{\text{jumlah seluruh data}}, \text{ dengan } i \text{ 1 dan } 0$$

Sehingga, diperoleh probabilitas dari kelas 0 (kelas yang dinyatakan tidak diterima) sebesar 0.2929292929, serta probabilitas kelas 1 (kelas yang dinyatakan diterima) sebesar 0.7070707071.

4.2. Perhitungan Probabilitas H Bersyarat (P(H|E_i))

Perhitungan probabilitas H bersyarat E, dimana H merupakan jenis persyaratan pendaftaran Bidik Misi. Diantaranya yaitu H.UKT, H.Prestasi akademik, H. Prestasi non akademik, H.Penghasilan orang tua, H.Daya listrik, H.Jumlah tanggungan keluarga. Perhitungan H bersyarat E dapat ditulis dengan (P(H.UKT, H.Prestasi akademik, H.Prestasi non akademik, H.Daya listrik, H.Penghasilan orang tua, H.Jumlah tanggungan keluarga|E). Dilakukan perhitungan kemungkinan setiap H terhadap E. Pertama, akan dilakukan perhitungan untuk setiap kemungkinan dari hasil parameter uang kuliah tunggal untuk setiap probabilitas H yang termasuk ke dalam kelas E_i untuk (P(H.UKT|E_i), (P(H.Penghasilan orang tua | E_i)), (P(H.Jumlah prestasi akademik | E_i)), (P(H.Jumlah prestasi non akademik | E_i)), (P(H.Jumlah tanggungan keluarga | E_i)) dan (P(H.Daya listrik | E_i)) berturut-turut ditunjukkan di Tabel 3, Tabel 4, Tabel 5, Tabel 6, Tabel 7 dan Tabel 8.

Tabel 3
Hasil Perhitungan Nilai Probabilitas UKT (Uang Kuliah Tunggal)

| UKT | Banyaknya Kejadian yang Terjadi | | Probabilitas (P) | |
|--------------|---------------------------------|----------------|------------------|--------------------|
| | Diterima | Tidak Diterima | Diterima(1) | Tidak Diterima (0) |
| Rp 400.000 | 9 | 1 | 0,12676056 | 0,035714286 |
| Rp 900.000 | 19 | 5 | 0,26760563 | 0,178571429 |
| Rp 1.000.000 | 2 | 2 | 0,02816901 | 0,071428571 |
| Rp 1.200.000 | 0 | 1 | 0 | 0,035714286 |
| Rp 1.300.000 | 5 | 4 | 0,07042254 | 0,142857143 |
| Rp 1.400.000 | 1 | 3 | 0,01408451 | 0,107142857 |
| Rp 1.700.000 | 5 | 1 | 0,07042254 | 0,035714286 |
| Rp 1.900.000 | 3 | 0 | 0,04225352 | 0 |
| Rp 2.100.000 | 8 | 4 | 0,11267606 | 0,142857143 |
| Rp 2.400.000 | 0 | 1 | 0 | 0,035714286 |
| Rp 2.600.000 | 16 | 4 | 0,22535211 | 0,142857143 |
| Rp 2.800.000 | 1 | 0 | 0,01408451 | 0 |
| Rp 3.300.000 | 0 | 1 | 0 | 0,035714286 |
| Rp 4.400.000 | 1 | 1 | 0,01408451 | 0,035714286 |
| Rp 4.900.000 | 1 | 0 | 0,01408451 | 0 |
| Jumlah | 71 | 28 | 1 | 1 |

Tabel 4
Hasil Perhitungan Nilai Probabilitas Jumlah Penghasilan Orangtua Per Bulan

| PENGHASILAN | Banyaknya Kejadian yang Terjadi | | Probabilitas (P) | |
|------------------|---------------------------------|----------------|------------------|--------------------|
| | Diterima | Tidak Diterima | Diterima(1) | Tidak Diterima (0) |
| >1500000-2500000 | 22 | 8 | 0,309859155 | 0,285714286 |
| >2500000-3500000 | 0 | 4 | 0 | 0,142857143 |
| >3500000-4500000 | 0 | 2 | 0 | 0,071428571 |
| >4500000 | 0 | 1 | 0 | 0,035714286 |
| 500000-1500000 | 49 | 13 | 0,690140845 | 0,464285714 |
| Jumlah | 71 | 28 | 1 | 1 |

Tabel 5
Hasil Perhitungan Nilai Probabilitas Jumlah Prestasi Akademik

| Daya Listrik | Banyaknya Kejadian yang Terjadi | | Probabilitas (P) | |
|--------------|---------------------------------|----------------|------------------|--------------------|
| | Diterima | Tidak Diterima | Diterima(1) | Tidak Diterima (0) |
| 0 | 2 | 4 | 0,028169014 | 0,142857143 |
| 1 | 1 | 1 | 0,014084507 | 0,035714286 |
| 2 | 3 | 1 | 0,042253521 | 0,035714286 |
| 3 | 3 | 1 | 0,042253521 | 0,035714286 |
| 4 | 2 | 2 | 0,028169014 | 0,071428571 |
| 5 | 9 | 4 | 0,126760563 | 0,142857143 |
| 6 | 50 | 14 | 0,704225352 | 0,5 |
| 7 | 0 | 0 | 0 | 0 |
| 8 | 0 | 1 | 0 | 0,035714286 |
| 9 | 1 | 0 | 0,014084507 | 0 |
| Jumlah | 71 | 28 | 1 | 1 |

Tabel 6
Hasil Perhitungan Nilai Probabilitas Jumlah Prestasi Non Akademik

| Jumlah Prestasi Non Akademik | Banyaknya Kejadian yang Terjadi | | Probabilitas (P) | |
|------------------------------|---------------------------------|----------------|------------------|--------------------|
| | Diterima | Tidak Diterima | Diterima(1) | Tidak Diterima (0) |
| 0 | 13 | 7 | 0,183098592 | 0,25 |
| 1 | 27 | 11 | 0,38028169 | 0,392857143 |
| 2 | 8 | 3 | 0,112676056 | 0,107142857 |
| 3 | 7 | 3 | 0,098591549 | 0,107142857 |
| 4 | 4 | 4 | 0,056338028 | 0,142857143 |
| 5 | 5 | 0 | 0,070422535 | 0 |
| 6 | 4 | 0 | 0,056338028 | 0 |
| 7 | 2 | 0 | 0,028169014 | 0 |
| 8 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 |
| 10 | 1 | 0 | 0,014084507 | 0 |
| jumlah | 71 | 28 | 1 | 1 |

Tabel 7
Hasil Perhitungan Nilai Probabilitas Jumlah Tanggungan Keluarga

| Jumlah Keluarga | Banyaknya Kejadian yang Terjadi | | Probabilitas (P) | |
|-----------------|---------------------------------|----------------|------------------|--------------------|
| | Diterima | Tidak Diterima | Diterima(1) | Tidak Diterima (0) |
| 1 | 13 | 5 | 0,183098592 | 0,178571429 |
| 2 | 17 | 12 | 0,23943662 | 0,428571429 |
| 3 | 15 | 6 | 0,211267606 | 0,214285714 |
| 4 | 11 | 2 | 0,154929577 | 0,071428571 |
| 5 | 13 | 1 | 0,183098592 | 0,035714286 |
| 6 | 2 | 2 | 0,028169014 | 0,071428571 |
| jumlah | 71 | 28 | 1 | 1 |

Tabel 8
Hasil Perhitungan Nilai Probabilitas Daya Listrik

| Daya Listrik | Banyaknya Kejadian yang Terjadi | | Probabilitas (P) | |
|--------------|---------------------------------|----------------|------------------|--------------------|
| | Diterima | Tidak Diterima | Diterima(1) | Tidak Diterima (0) |
| 450 | 33 | 7 | 0,46478873 | 0,25 |
| 900 | 31 | 15 | 0,43661972 | 0,535714286 |
| 1200 | 7 | 6 | 0,09859155 | 0,214285714 |
| jumlah | 71 | 28 | 1 | 1 |

4.3 Perolehan Hasil Klasifikasi

Setelah pengimplementasian metode Naive Bayes dalam proses klasifikasi penerima program beasiswa Bidik Misi seperti yang dijelaskan diatas, diperoleh hasil klasifikasi dari data *testing* yang nantinya akan dibandingkan dengan data aktual. Perbandingan hasil klasifikasi menggunakan Naive Bayes dan data aktual dapat dilihat pada Tabel 8.

Tabel 8
Perbandingan Data Aktual Dengan Data
Prediksi Naive Bayes

| | | | | | | | | | | | | | | | | | | | | | | | |
|---------------|----|---|---|---|---|---|----|---|---|---|---|---|----|---|---|---|---|---|---|---|----|---|---|
| Data Aktual | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| Data Prediksi | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Keterangan | TS | S | S | S | S | S | TS | S | S | S | S | S | TS | S | S | S | S | S | S | S | TS | S | S |

Keterangan :

0 : Kelas yang dinyatakan tidak lolos seleksi

1 : Kelas yang dinyatakan lolos seleksi

TS : Tidak sesuai dengan data aktual

S : Sesuai dengan data actual

Dari data klasifikasi yang dihasilkan dari metode Naive Bayes, diketahui bahwadari 24 data *testing*, 20 diantaranya menghasilkan hasil prediksi yang benar, artinya sama dengan data aktual. Sedangkan 4 lainnya menghasilkan hasil prediksi yang salah.

4.4 Pengujian Akurasi Sistem

Setelah proses implimentasi pengklasifikasian penerima beasiswa menggunakan metode Naive Bayes dengan uji coba sebanyak 94 data, dimana 71 data diantaranya digunakan sebagai data latih, dan 24 lainnya digunakan sebagai data uji. Setelah dilakukan uji coba serta membandingkan antara data aktual dengan data hasil klasifikasi, sudah pasti akan ada perbedaan diantara keduanya. Dari perbedaan yang relah diketahui tersebut, akan dihitung nilai galat (*error*) dan akurasinya. Hasil dari keakurasian inilah yang akan menentukan seberapa besar kualitas dari sistem yang dibuat. Hasil uji coba data *testing* ditunjukkan oleh *confusion matrix* pada Tabel 9.

Tabel 9
Confusion Matrix

| | Kelas Hasil Prediksi | |
|-------------------|----------------------|---|
| | 0 | 1 |
| Kelas Data Aktual | 0 | 3 |
| | 1 | 4 |

Dari Tabel 9, dapat diketahui bahwa hasil klasifikasi penerima program beasiswa Bidikmisi menggunakan metode Naive Bayes memiliki tingkat keakurasian sistem sebesar 83,33%

5. KESIMPULAN

Berdasarkan hasil penelitian diatas, dapat disimpulkan bahwa metode *Naive Bayes* dapat digunakan sebagai metode pengklasifikasian kelayakan mahasiswa pada seleksi beasiswa Bidik Misi UIN Sunan Ampel Surabaya. Sistem yang dibuat sudah baik dengan dengan hasil klasifikasi yang mendekati data aktual dengan akurasi sebesar

83,33 %.

DAFTAR PUSTAKA

- [1] Ristekdikti, *Pedoman Bidik Misi Tahun 2017*. Jakarta: Ristekdikti, 2017.
- [2] Jawa Pos, "UINSA sediakan 180 kursi BIDIKMISI."
- [3] M.A. Shadiq, "Keoptimalan Naive Bayes dalam Klasifikasi.," presented at the Universitas Pendidikan Indonesia, Bandung, 2017.
- [4] Al Riza Khadafy and Romi Satria Wahono, "Penerapan Naive Bayes Untuk Mengurangi Data Noise pada Klasifikasi Multi Kelas dengan Decision Tree.," *J. Intell. Syst.*, 2015.
- [5] Jiawei Han, Micheline Kamber, and Jian Pei, *Data Mining : Concept and Technique*. New York: Elsevier, 2012.
- [6] Pramita M. and Ely R., "Prediksi Nilai Proyek Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining.," presented at the Seminar Nasional Sistem Informasi Indonesia, pp. 2-3.
- [7] Michael Steinbach, Vipin Kumar, and Pang-Ning Tan, *Introduction to Data Mining*. Boston: Pearson Education, 2006.
- [8] Slamet B. and Setyoningsih W., "Data Mining dalam Kajian Kualitas Aspal Beton Menggunakan Forward Selection Berbasis Naive Bayes.," *J. Inform. Upgris*, vol. 3, no. 1, 2017.
- [9] Harwati and Hanna M., "Prediksi Prestasi Mahasiswa dengan Jalur Siswa Berprestasi (PSB) Menggunakan Metode Naive Bayes.," *Performa*, vol. 15, no. 1, pp. 44-50, 2016.
- [10] Eka Pratiwi Hadiyani, "Rancang Bangun Sistem Pendukung Keputusan Untuk Pemilihan Anggota Terbaik AIESEC Surabaya menggunakan Metode Naive Bayes.," Universitas Airlangga, 2013.
- [11] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi.," *J. Inform.*, vol. 8, no. 1, Jan. 2014.
- [12] Rusmawardhani R., "Implementasi Metode Naive Bayes Pada Penyeleksian Narasumber Program Acara Talkshow Pendidikan (Studi Kasus Tvri Stasiun Kaltim).," presented at the Prosiding Seminar Ilmu Komputer dan Teknologi Informasi, 2016, vol. 1.
- [13] Mujib Ridwan, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier.," *J. EECCIS*, vol. 7, 2013.