

Prediksi Keaktifan Studi Mahasiswa Baru dengan Algoritma Naive Bayes

1st Ciske Mulyadi, 2nd M. Nur Juniadi
 1st Program Studi Manajemen Informatika
 2nd Program Studi Komputerisasi Akuntansi
 STMIK Amikom Surakarta

Solo, Indonesia

1st ciske@dosen.amikomsolo.ac.id, 2nd nurjuniadi@dosen.amikomsolo.ac.id

Abstrak—Masalah banyaknya mahasiswa yang putus studi pada saat menempuh pendidikan masih menjadi momok bagi perguruan tinggi swasta. Masalah ini jika tidak diatasi maka akan mengganggu operasional perguruan tinggi swasta, mengingat ketergantungan operasional perguruan tinggi swasta terhadap pendapatan dari pembayaran biaya kuliah mahasiswa. Salah satu cara mengatasi masalah ini adalah dengan membuat model prediksi keaktifan mahasiswa pada semester berikutnya dengan teknik data mining. Studi ini menggunakan algoritma Naïve Bayes untuk membuat model prediksinya dan adapun hasilnya model tersebut mempunyai tingkat akurasi sebesar 77.2%, nilai presisi sebesar 82.1%, dan nilai recall sebesar 92.2%. Model prediksi ini nantinya dapat digunakan pengelola program studi untuk mengurangi jumlah mahasiswa yang putus studi dengan memberikan perhatian khusus terhadap mahasiswa-mahasiswa baru yang diprediksi akan putus studi pada saat menempuh pendidikan di perguruan tinggi tersebut.

Kata kunci—algoritma *Naïve Bayes*, data mining, prediksi

I. PENDAHULUAN

Perguruan tinggi swasta setiap tahun berusaha memenuhi target penerimaan mahasiswa baru sesuai dengan daya tampung yang mereka miliki. Banyak perguruan tinggi swasta yang tidak bisa memenuhi target jumlah penerimaan mahasiswa baru yang sudah mereka tentukan sendiri. Oleh karena itu penting kiranya menjaga agar para mahasiswa tersebut bisa menyelesaikan studinya, mengingat jika banyak mahasiswa yang putus studi ataupun mengundurkan diri pada saat menempuh pendidikan, maka akan sangat mempengaruhi biaya operasional perguruan tinggi swasta tersebut.

Masalah banyaknya mahasiswa yang putus studi pada saat menempuh pendidikan masih menjadi momok bagi perguruan tinggi swasta. Sudah bukan rahasia lagi bahwa sebagian besar operasional perguruan tinggi swasta sangat bergantung dari pendapatan yang diperoleh pembayaran biaya kuliah mahasiswanya. Jika sebuah perguruan tinggi swasta tidak mampu memenuhi target penerimaan mahasiswa baru maka target pendapatan dari mahasiswa baru tidak tercapai. Masalah ini bisa semakin buruk jika banyak mahasiswa yang putus studi di awal,

pertengahan, maupun akhir pendidikannya di perguruan tinggi.

Masalah banyaknya mahasiswa yang putus studi tersebut dapat diatasi salah satunya dengan memprediksi siapa saja mahasiswa yang nantinya akan mempunyai kecenderungan putus studi saat menempuh pendidikan. Untuk melakukan prediksi berdasarkan data-data masa lalu dapat dilakukan dengan teknik data mining. Data mining adalah proses untuk menemukan pola yang berguna dan kecenderungan di dalam kumpulan data yang besar [1]. Salah satu fungsi data mining yang dapat digunakan untuk memprediksi adalah klasifikasi. Tugas dari klasifikasi adalah memprediksi keluaran variabel/class yang bernilai kategorikal atau polinomial [2]. Beberapa metode data mining yang dapat diterapkan untuk klasifikasi. Algoritma klasifikasi yang populer adalah Decision Trees, Neural Networks, k-Nearest Neighbours, Naive Bayes, dan algoritma Genetik [3].

Studi tentang penggunaan data mining untuk melakukan prediksi dalam dunia pendidikan di perguruan tinggi telah banyak dipublikasikan.

Tantangan terbesar yang dihadapi perguruan tinggi adalah mengurangi jumlah peserta didik yang putus studi [4]. Jumlah peserta didik yang putus studi menjadi indikator seberapa baik kinerja akademik maupun pengelolaan seleksi penerimaan mahasiswa baru. Hal ini menyebabkan perguruan tinggi lebih berfokus pada kekuatan siswa daripada kualitas pendidikan. Pada penelitian ini diterapkan aplikasi data mining untuk menghasilkan model prediktif untuk pengelolaan mahasiswa putus studi, sehingga dapat diketahui mana mahasiswa yang perlu mendapatkan dukungan lebih. Hasil penelitian menunjukkan bahwa algoritma mesin pembelajaran mampu membangun model prediksi secara efektif dari data mahasiswa putus studi yang ada.

Penelitian yang dilakukan [5] tentang prediksi kuliah mahasiswa baru, ditemukan hasil bahwa penyebab mahasiswa tidak melanjutkan studi di semester 2 adalah nilai IPK yang dibawah 1. Penelitian ini menggunakan atribut nomor induk mahasiswa, IPK semester 1, grade masuk, gelombang pendaftaran, dan keaktifan selanjutnya.

Kinerja dalam program akademik merupakan salah satu faktor yang paling penting yang mempengaruhi kualitas pendidikan tinggi tersedia untuk mahasiswa [6]. Pada studi ini digunakan teknik data mining khususnya klasifikasi

untuk menganalisis nilai mahasiswa dalam tugas evaluatif yang berbeda untuk mata kuliah data terstruktur. Untuk tujuan ini, dibandingkan tiga pengklasifikasi yang berbeda untuk memprediksi kinerja mahasiswa. Di sini diterapkan teknik klasifikasi untuk kedua atribut numerik dan dikategorikan. Hasilnya menunjukkan bahwa model berdasarkan algoritma Naïve Bayes memberikan prediksi yang paling akurat dengan akurasi 91% untuk memprediksi kegagalan mahasiswa dalam mata kuliah.

Studi lain menunjukkan bahwa faktor yang paling berpengaruh dalam tingkat kelulusan mahasiswa adalah Indeks Prestasi Semester (IPS) dan Jumlah Sistem Kredit Semester (SKS) secara keseluruhan maupun tiap semester [7]. Masa studi mahasiswa dapat diprediksi berdasarkan faktor yang berkaitan dengan akademik mahasiswa, seperti program studi, nilai indeks prestasi semester dan jumlah sks saat di perguruan tinggi. Algoritma *Naïve Bayes* yang dipakai dapat menentukan prediksi masa studi mahasiswa dengan tingkat Akurasi pada pengujian algoritma bernilai 85.17 % pada nilai rata-rata pengujian di lima semester.

Studi lain menyatakan bahwa salah satu tantangan terbesar yang dihadapi pendidikan tinggi saat ini adalah memprediksi jalur akademik mahasiswa [8]. Banyak sistem pendidikan tinggi tidak mampu mendeteksi populasi mahasiswa yang cenderung putus karena kurangnya metode intelijen untuk menggunakan informasi, dan bimbingan dari sistem universitas. Metode data mining untuk mengklasifikasikan dan memprediksi siswa putus sekolah, diusulkan dua pengklasifikasi berbeda, yaitu *Decision Tree* (DT), dan *Naive Bayes* (NB), dan dilatih menggunakan dataset yang telah kumpulan. Hasilnya menunjukkan bahwa akurasi dari DT mencapai 98,14%, sedangkan NB mencapai 96,86%.

Penelitian yang telah dilakukan oleh Sulistiono dan Defiyanti menunjukkan bahwa algoritma *Naïve bayes* memiliki tingkat akurasi yang paling tinggi [9]. Tingkat akurasi algoritma *Naïve bayes* sebesar 93,58% dibandingkan dengan algoritma *C4.5* sebesar 93,05 dan *Neural Network* sebesar 89,56%. Oleh karena itulah penulis menggunakan metode klasifikasi Naïve bayes untuk melakukan penelitian ini. Metode klasifikasi *Naïve bayes* dipilih karena metode *Naïve bayes* merupakan metode probabilitas statistik yang sederhana tetapi menghasilkan hasil yang akurat. Studi ini akan menerapkan algoritma Naïve Bayes untuk mendapatkan model prediksi keaktifan mahasiswa pada semester 2.

II. METODE PENELITIAN

Makalah Metode yang digunakan pada studi ini adalah eksperimen yang mencakup investigasi hubungan sebab-akibat menggunakan pengujian yang dikontrol sendiri [10]. Studi ini bertujuan untuk mendapatkan model prediksi keaktifan mahasiswa baru di semester 2. Karena penelitian yang diakui/diterima harus mengikuti aturan yang diakui [10], maka pada penelitian ini dilakukan dengan mengikuti tahapan dalam data mining yang memiliki enam fase CRISP-DM (*Cross Industry Standard Process for Data Mining*) [11]. Tahapan-tahapannya sebagai berikut:

1. *Business Understanding*

Tahap pertama adalah memahami tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menterjemahkan

pengetahuan ini ke dalam pendefinisian masalah dalam data mining. Selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut.

2. *Data Understanding*

Tahap ini dimulai dengan pengumpulan data yang kemudian akan dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi masalah kualitas data, atau untuk mendeteksi adanya bagian yang menarik dari data yang dapat digunakan untuk hipotesa untuk informasi yang tersembunyi.

3. *Data Preparation*

Tahap ini meliputi semua kegiatan untuk membangun dataset akhir (data yang akan diproses pada tahap pemodelan/modeling) dari data mentah. Tahap ini dapat diulang beberapa kali. Pada tahap ini juga mencakup pemilihan tabel, record, dan atribut-atribut data, termasuk proses pembersihan dan transformasi data untuk kemudian dijadikan masukan dalam tahap pemodelan (*modelling*).

4. *Modelling*

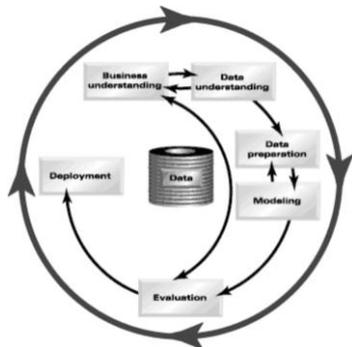
Dalam tahap ini akan dilakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Secara khusus, ada beberapa teknik berbeda yang dapat diterapkan untuk masalah data mining yang sama. Di pihak lain ada teknik pemodelan yang membutuhkan format data khusus. Sehingga pada tahap ini masih memungkinkan kembali ke tahap sebelumnya.

5. *Evaluation*

Pada tahap ini, model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisa data. Pada tahap ini akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal (*Business Understanding*). Kunci dari tahap ini adalah menentukan apakah ada masalah bisnis yang belum dipertimbangkan. Di akhir dari tahap ini harus ditentukan penggunaan hasil proses data mining.

6. *Deployment*

Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna. Tahap deployment dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses data mining yang berulang dalam perusahaan. Dalam banyak kasus, tahap deployment melibatkan konsumen, di samping analisis data, karena sangat penting bagi konsumen untuk memahami tindakan apa yang harus dilakukan untuk menggunakan model yang telah dibuat.



Gambar 1 Proses CRISP-DM [12]

III. HASIL DAN PEMBAHASAN

Pembahasan Adapun hasil studi ini dibahas sesuai tahapan dalam data mining yang memiliki enam fase CRISP-DM.

A. Business Understanding

Pemahaman bisnis merupakan tahap pertama dalam proses CRISP-DM yang lebih tepat disebut sebagai tahap pemahaman studi. Tahapan ini bisa dipecah menjadi beberapa tahapan sebagai berikut.

1. Menentukan Tujuan Bisnis

Pada tahapan ini yang dimaksud dengan tujuan bisnis adalah tujuan studi ini. Tujuan studi ini adalah mendapatkan model prediksi keaktifan mahasiswa pada semester 2. Data yang digunakan adalah data mahasiswa baru tahun 2016-2017 di STMIK Amikom Surakarta. Tujuannya agar nantinya dapat mengurangi jumlah mahasiswa yang putus studi pada tahun-tahun berikutnya. Hal ini dilakukan dikarenakan banyaknya mahasiswa yang putus studi pada saat menempuh pendidikan di STMIK Amikom Surakarta. Hasil studi ini nantinya dapat dijadikan sebagai salah satu dasar pengambilan keputusan dalam menentukan kebijakan program studi.

2. Melakukan Penilaian Situasi

Pada tahapan ini dibutuhkan pemahaman terhadap tujuan studi dan menerjemahkan ke dalam tujuan *data mining*. Berdasarkan tujuan studi yang sudah ditentukan pada tahapan sebelumnya, maka diperlukan pemahaman hal-hal apa saja yang dianggap mempengaruhi keaktifan mahasiswa pada semester 2. Selanjutnya berdasarkan pemahaman terhadap studi-studi yang sudah dilakukan sebelumnya diperoleh data tentang atribut-atribut yang mempengaruhi keaktifan mahasiswa. Data Atribut-atribut mahasiswa tersebut nantinya yang akan diolah menggunakan metode data mining.

3. Menentukan strategi awal *data mining*

Strategi awal dalam menerapkan tujuan dilakukannya data mining adalah melakukan permintaan data mahasiswa baru terlebih dahulu ke Bagian Administrasi Akademik dan informasi Kemahasiswaan (BAIK) STMIK Amikom Surakarta.

B. Data Understanding

Dataset mahasiswa yang didapatkan dari BAIK STMIK Amikom Surakarta berupa dokumen excel sejumlah 251 record.

1. Pengumpulan data awal

Adapun sumber data utama yang digunakan dalam studi ini adalah *dataset* mahasiswa baru 2016-2017 di STMIK Amikom Surakarta dengan tipe file Microsoft Excel Worksheet (.xls).

2. Mendeskripsikan data

Dataset alumni 2016-2017 tersebut terdiri dari beberapa atribut antara lain NIM, nama, peminatan program studi, jenis kelamin, agama, asal sekolah, tempat & tanggal lahir, alamat, dan gelombang pendaftaran.

3. Evaluasi kualitas data

Hasil evaluasi terhadap kualitas data yaitu menemukan banyak nilai kosong / *null* yang disebut dengan *missing value* pada atribut dalam *dataset* mahasiswa.

4. Pemilihan Atribut

Atribut yang digunakan adalah NIM, jenis kelamin, asal sekolah, program studi, dan status mahasiswa di semester 2 (aktif atau tidak). Penggunaan atribut-atribut tersebut dengan pertimbangan 4 atribut awal adalah data asli mahasiswa baru sebelum menempuh semester pertama kuliah. Penelitian ini tidak menggunakan atribut IP semester 1 karena jika atribut tersebut digunakan dalam pembuatan model maka otomatis atribut tersebut juga digunakan untuk melakukan prediksi. Jika hal ini dilakukan maka ketika akan melakukan prediksi keaktifan mahasiswa baru di semester 2 harus menunggu data IP semester 1. Ketika semester 1 berakhir baru bisa dilakukan prediksi, hal ini justru tidak efektif karena jaraknya terlalu dekat dengan semester 2. Otomatis malah tidak sempat melakukan tindakan terhadap mahasiswa baru yang diprediksi akan putus studi di semester 2.

C. Persiapan Data (*Data Preparation*)

Persiapan data mencakup semua kegiatan untuk membangun *dataset* mahasiswa yang akan diterapkan ke dalam alat pemodelan, dari data mentah awal berupa *dataset* mahasiswa dan selanjutnya akan melakukan proses *data mining*.

1. Seleksi data

Atribut yang digunakan adalah NIM, jenis kelamin, asal sekolah, program studi, dan status mahasiswa di semester 2 (aktif atau tidak).

2. Preprocessing Data

Data yang sudah dikumpulkan diolah untuk mengurangi data yang tidak relevan, atau data dengan atribut yang hilang. Pengolahan juga berupa konversi nilai-nilai redundan (berlebihan), atau nilai yang terlalu beragam ke dalam kelompok yang lebih kecil untuk mempermudah pembentukan model. Pada tahap ini merupakan tahap untuk memastikan data mahasiswa yang dipilih telah layak untuk dilakukan proses pengolahan.

Selected attribute	
Name: NIM	Type: Numeric
Missing: 1 (0%)	Distinct: 250
	Unique: 250 (100%)
Statistic	Value
Minimum	1611031802
Maximum	1712030319
Mean	1651858492.404
StdDev	49074860.215

Gambar 2 Data Hilang (Missing: 1 (0%))

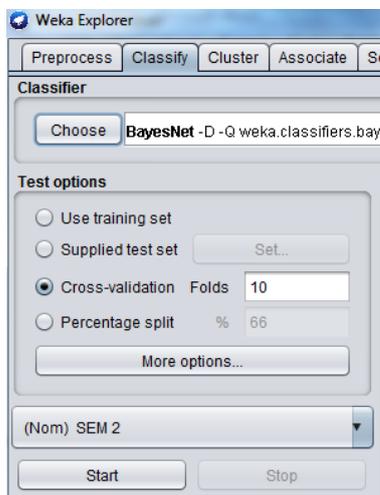
Pada gambar 2 terlihat ada data yang hilang, dan hanya 1. Langkah yang dilakukan pada tahap *preprocessing* data adalah dengan menghapus *instance* tersebut sehingga jumlah *instance* sekarang menjadi 250.

3. Transformasi data

Data yang berjenis numerikal dilakukan proses inialisasi data terlebih dahulu ke dalam bentuk nominal.

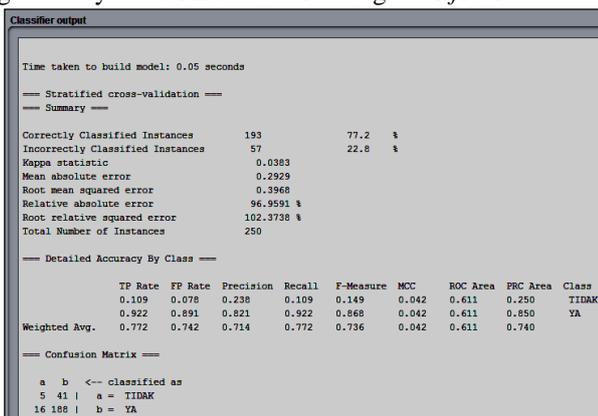
D. *Pemodelan (Modeling)*

Pemodelan adalah fase yang secara langsung melibatkan teknik *data mining* yaitu dengan melakukan pemilihan teknik *data mining* dan menentukan algoritma yang akan digunakan. Pada studi ini diusulkan model prediksi keaktifan mahasiswa pada semester 2 menggunakan teknik *data mining*, yaitu dengan algoritma Naïve Bayes. Tahapan *Modelling* dilakukan terhadap 250 dataset mahasiswa baru 2016-2017 dengan tools Weka 3.8.3 seperti ditunjukkan pada gambar 3.



Gambar 3 Modelling dengan Naïve Bayes

Selanjutnya dilakukan pembuatan model Naïve bayes tersebut dengan menggunakan terhadap 250 dataset mahasiswa baru 2016-2017 dengan test options yang umum digunakan yaitu *cross validation* dengan 10 *folds*.



Gambar 4 Classifier Ouput

E. *Evaluation*

Evaluasi adalah fase lanjutan terhadap tujuan *data mining*. Evaluasi dilakukan secara mendalam dengan tujuan agar hasil pada tahap pemodelan sesuai dengan sasaran yang ingin dicapai dalam tahap *business understanding*.

1. *Evaluation Results*

Tahap ini menilai sejauh mana hasil pemodelan *data mining* memenuhi tujuan *data mining* yang telah ditentukan pada tahap *business understanding*. Pada tahap evaluasi ini digunakan metode confusion matrix sebagai metode yang akan menguji precision, recall maupun accuracy dari pola yang terbentuk. Nilai recall, presisi maupun accuracy seperti terlihat pada gambar 5 berikut.

```

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.109	0.078	0.238	0.109	0.149	0.042	0.611	0.250	TIDAK
	0.922	0.891	0.821	0.922	0.868	0.042	0.611	0.850	YA
Weighted Avg.	0.772	0.742	0.714	0.772	0.736	0.042	0.611	0.740	

Gambar 5 Detailed Accuracy By Classes

Hasilnya pengolahan ditampilkan pada gambar 5, dimana nilai akurasi yang diperoleh sebesar 77.2%, nilai presisi sebesar 82.1%, dan nilai recall sebesar 92.2%.

Sedangkan nilai sensitifity sebesar 92.2%, nilai specificity sebesar 10.9%, nilai PPV sebesar 82.1%, dan NPV sebesar 23.8% seperti terlihat pada gambar 5.

2. *Review Process*

Tahapan ini digunakan untuk memastikan bahwa semua tahapan atau faktor penting yang telah dilakukan dalam pengolahan data tidak ada yang terlewatkan. Berdasarkan hasil pemeriksaan dipastikan semua tahapan telah dilakukan dan tidak ada yang terlewatkan.

3. *Determine Next Steps*

Pada tahap ini adalah tahapan dalam menentukan langkah selanjutnya yang dilakukan. Terdapat 2 pilihan yaitu kembali pada tahap awal (*business understanding*) atau melanjutkan ke tahap akhir (*deployment*). Mengingat hasil pemodelan *data mining* memenuhi tujuan *data mining* yang telah ditentukan pada tahap *business understanding* maka dilanjutkan ke tahapan *deployment*.

F. *Deployment*

Deployment merupakan tahapan akhir dalam pembuatan laporan hasil kegiatan *data mining*. Laporan akhir yang berisi mengenai pengetahuan yang diperoleh atau pengenalan pola pada data dalam proses *data mining*. Berdasarkan studi yang dilakukan, telah dihasilkan suatu pola, informasi, dan pengetahuan baru dalam proses *data mining* untuk menentukan model prediksi keaktifan mahasiswa pada semester 2 berdasarkan mahasiswa baru tahun 2016-2017. Dalam rangka menilai kinerja model prediksi keaktifan mahasiswa dilakukan dengan metode confusion matrix. Berikut ditampilkan kembali hasil confusion matrix (gambar 6) dan pada tabel 1.

```

=== Confusion Matrix ===

```

a	b	<-- classified as
5	41	a = TIDAK
16	188	b = YA

Gambar 6 Confusion Matrix

Tabel 1: Hasil *Confusion Matrix*

n =250	Prediction	
	TIDAK	YA
Aktual TIDAK	5 (TN)	41 (FP)
Aktual YA	16 (FN)	188 (TP)

Catatan: n= sum of test data, TN= True negatives, FP= False positives, FN= False negatives, TP= True positives

Dalam kasus studi ini jika dijelaskan lebih rinci sebagai berikut:

- True Negatives (TN) berarti jumlah mahasiswa diprediksi tidak aktif di semester 2 dan kenyataannya tidak aktif di semester 2
- False Positives (FP) berarti jumlah mahasiswa diprediksi aktif di semester 2 tapi kenyataannya tidak aktif di semester 2
- False Negatives (FN) berarti jumlah mahasiswa diprediksi lulus tidak aktif di semester 2 tapi kenyataannya aktif di semester 2
- True Positives (FP) berarti jumlah mahasiswa diprediksi aktif di semester 2 dan kenyataannya aktif di semester 2

Berikut adalah hasil evaluasi model *confusion matrix*:

- Nilai akurasi (acc) sebesar 77.2% menunjukkan proporsi jumlah prediksi yang benar. Dari 251 test data yang benar prediksinya (baik lulus tepat waktu maupun tidak tepat waktu) sejumlah 193.
- Nilai sensitivity atau recall sebesar 92.2% menunjukkan proporsi TP terhadap (TP+FN).
- Nilai specificity sebesar 10.9% menunjukkan proporsi TN terhadap (TN+FP).
- Nilai PPV (*positive predictive value*) atau *precision* sebesar 82.1% menunjukkan proporsi kasus dengan hasil diagnosa positive atau proporsi TP terhadap (TP+FP).
- Nilai NPV (*negative predictive value*) sebesar 23.8% menunjukkan proporsi kasus dengan hasil diagnosa negative atau proporsi TN terhadap (TN+FN).

IV. KESIMPULAN DAN SARAN

Berdasarkan pembahasan sebelumnya dapat disimpulkan bahwa algoritma Naïve Bayes mampu menghasilkan model prediksi keaktifan mahasiswa dengan tingkat akurasi sebesar 77.2%, nilai presisi sebesar 82.1%, dan nilai recall sebesar 92.2%. Model ini dapat menghasilkan prediksi mahasiswa yang lanjut dan putus studi pada semester 2. Data mahasiswa yang diprediksi putus studi pada semester 2 menjadi masukan bagi pengelola program studi untuk memberikan penanganan khusus terhadap mahasiswa-mahasiswa tersebut agar nantinya mereka bisa lanjut studi di semester 2.

Adapun saran untuk penelitian berikutnya sebagai berikut:

- Studi ini masih bisa dikembangkan dengan penambahan atribut data mahasiswa baru yang dianggap dapat memperbaiki prediksi seperti jumlah nilai ujian

nasional (UN) atau rata-rata nilai UN sehingga hasilnya lebih akurat.

- Apabila model prediksi dengan algoritma Naïve Bayes ini dianggap belum optimal dikarenakan nilai akurasinya kurang tinggi maka bisa dioptimalkan dengan membuat model prediksi dengan algoritma lain seperti Decision Trees, Neural Networks, k-Nearest Neighbours dan lain-lain.
- Selain itu studi ini bisa dikembangkan dengan membuat model prediksi keaktifan mahasiswa di semester 3, 4, 5, 6, 7, dan 8. Adapun untuk prediksi keaktifan di semester 3, atribut IPK semester 1 bisa digunakan untuk menyempurnakan model. Sedangkan untuk prediksi keaktifan di semester 4, atribut IPK semester 2 bisa digunakan dan seterusnya. Penggunaan data IPK selisih 2 semester tersebut bertujuan agar diketahui hasil prediksi 1 semester sebelumnya sehingga pengawasan bisa dilakukan lebih dini.

REFERENSI

- [1] D. T. Larose and C. D. Larose, "Data Mining and Predictive Analytics", 2nd ed., New Jersey: John Wiley & Sons, Inc, 2015.
- [2] V. Kotu and B. Deshpande, "Predictive Analytics and Data Mining. Concepts and Practice with RapidMiner", Massachusetts: Elsevier Inc, 2015.
- [3] E. Yukselturk, S. Ozekes, and Y. K. Türel, Predicting Dropout Student: An Application of Data Mining Methods in an Online Education Program, *European Journal of Open, Distance and e-Learning*. [Online] 17(1), 118-133, 2014. Available from: doi:10.2478/eurodl-2014-0008 [Accessed 18th July 2019].
- [4] S. Pal, Mining Educational Data Using Classification to Decrease Dropout Rate of Students, *International journal of multidisciplinary sciences and engineering*, 3(5), pp. 35-39, 2012.
- [5] Mardiani, "Penerapan Klasifikasi dengan Algoritma CART untuk Prediksi Kuliah bagi Mahasiswa Baru", SNATI, Juni 2012.
- [6] M. A. Al-Barrak and M. S. Al-Razgan, Predicting Students' Performance Through Classification: A Case Study, *Journal of Theoretical and Applied Information Technology*, pp. 167-175, 2015.
- [7] Mongan Winny Amelia, Arie S. M. Lumenta, and Agustinus Jacobus, "Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes", *E-Journal Teknik Informatika*, Vol 11, No.1 , 2016, ISSN : 2301 – 8364.
- [8] G. S. Abu-Oda and A. M. El-Halees, Data Mining in Higher Education: University Student Dropout Case Study, *International Journal of Data Mining & Knowledge Management Process (IJDKP)*. [Online] 5 (1), pp. 15-27, 2015. Available from: doi:10.5121/ijdkp.2015.5102 [Accessed 15th July 2019].
- [9] Heru Sulistiono, "Kajian Penerapan Algoritma C4.5, Neural Network dan Naïve Bayes untuk Klasifikasi Mahasiswa yang Bermasalah dalam Registrasi". *Faktor Exacta*, 2015.
- [10] C. W. Dawson, "Projects in Computing and Information Systems A Student's Guide", 2nd ed., Great Britain: Pearson Education, 2009.
- [11] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinart, C. Shearer, and R. Wirth, *CRISP-DM Step-by-step Data Mining Guide*, 2000. [Online]. Available from: www.crisp-dm.org/CRISPWP-0800.pdf [Accessed 6th August 2019].
- [12] Daniel T. Larose, "Data Mining Methods and Models. Hoboken", New Jersey : Jhon Wiley & Sons, Inc, 2006.