

Prediksi Prestasi Mahasiswa dengan Jalur Siswa Berprestasi (PSB) Menggunakan Metode Naïve Bayers

Harwati^{*1)}, Hanna Miratama²⁾

1) 2) Industrial Engineering Universitas Islam Indonesia

Abstract

The purpose of this research is to predict the academic performance of students who submit university by special way (PSB) based on their GPA using naïve Bayes methods. The predicted results will be used to provide recommendations what attributes can be used as indicators of the selection through the PSB. From the calculation, it is resulted PSB students with satisfying GPA is they that have the attributes from Central Java, majoring in science, from public schools, parents' work of civil servants, have academic achievement during high school, and male. The level of accuracy for this calculation is equal to 51.69%.

Kata kunci: *Seleksi, Mahasiswa, Naïve Bayes, Akurasi, Prediksi, Keputusan*

1. Pendahuluan

Pendidikan merupakan salah satu syarat dalam meningkatkan kemajuan suatu bangsa. Terwujudnya pendidikan yang berkualitas diharapkan Perguruan Tinggi mampu mengantarkan bangsa menuju kesuksesan (Ningrum, 2014). Bagi lembaga pendidikan tinggi yang tujuannya untuk memberikan kontribusi pada peningkatan kualitas pendidikan tinggi, keberhasilan penciptaan modal manusia adalah subjek analisis yang terus menerus (Osmanbegovic dan Suljic, 2012). Mahasiswa merupakan salah satu aspek penting dalam evaluasi keberhasilan penyelenggaraan program studi pada suatu perguruan tinggi (Ridwan, et.al. 2013). Menurut Ibrahim (dalam Eko et al 2014), memprediksi prestasi akademik siswa sangat penting bagi suatu lembaga pendidikan karena program strategis tersebut dapat direncanakan dalam meningkatkan atau mempertahankan kinerja siswa selama masa studi di lembaga. Prestasi belajar dibidang pendidikan adalah hasil dari pengukuran terhadap peserta didik yang meliputi faktor kognitif, afektif, dan psikomotor setelah mengikuti proses pembelajaran yang diukur dengan menggunakan instrumen tes (Sutrisno, et.al 2014). Mahasiswa merupakan *raw input* yang diproses dalam suatu pendidikan diterima melalui seleksi penerimaan mahasiswa baru. Sistem seleksi penerimaan mahasiswa baru merupakan bagian terpenting dalam rangka meningkatkan mutu masukan bagi suatu Universitas. Semua proses dalam penerimaan mahasiswa ini sangat penting terutama pada tes masuk yang merupakan sebagai filter untuk dapat menerima mahasiswa sesuai dengan kompetensinya. (Jati Sasongko Wibowo, 2012).

Universitas Islam Indonesia adalah universitas swasta di Indonesia terletak di Yogyakarta. Dalam system penerimaan mahasiswa baru ada beberapa pilihan untuk calon mahasiswa dalam mengikuti seleksi penerimaan mahasiswa baru di Universitas Islam Indonesia yaitu *Computer Based Test (CBT)*, *Paper Based Test (PBT)*, dan *Penelusuran Siswa Berprestasi (PSB)*. Pada penelitian ini difokuskan untuk memprediksi penerimaan mahasiswa jalur *Penelusuran Siswa Berprestasi (PSB)* pada Fakultas Teknik Industri, Prodi Teknik Industri. Dimana *Penelusuran Siswa Berprestasi (PSB)* adalah pola penerimaan calon mahasiswa melalui

* *Correspondance* : harwati@uii.ac.id

seleksi prestasi Bidang Akademik dan Minat-Bakat. Dalam jalur penelusuran siswa berprestasi (PSB) ini terdiri dari 4 (empat) kategori, yaitu Siswa Berprestasi Bidang Akademik, Siswa Berprestasi Bidang Olah Raga dan Seni, Beasiswa Bibit Unggul, serta Hafidz/Hafidzah Al-Qur'an.

Dalam memprediksi penerimaan mahasiswa baru jalur Penelusuran Siswa Berprestasi (PSB) bidang ilmu yang dapat digunakan untuk memprediksi suatu keputusan berdasarkan beberapa parameter adalah bidang ilmu *Data Mining* dengan teknik klasifikasi. *Data Mining* adalah proses menggali dan menganalisa sejumlah data yang sangat besar untuk memperoleh sesuatu yang benar, baru, sangat bermanfaat dan akhirnya dapat dimengerti suatu corak atau pola dalam data tersebut (Han & Kamber, 2006). Klasifikasi adalah sebuah proses untuk menemukan model yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui (Pang-NingTan, et.al. 2004).

Pada teknik klasifikasi terdapat beberapa algoritma yang bisa digunakan antara lain pohon keputusan (*decisiontree*), *rulebased*, *neuralnetwork*, *support vector machine*, *naivebayes*, dan *nearest neighbor*. Penelitian ini menggunakan metode algoritma *naïve bayes* dengan menggunakan *software* Rapid Miner, karena *naïve bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema atau aturan bayes dengan asumsi independensi yang kuat pada fitur, artinya bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama (Yuda Septian Nugroho, 2014). Naive Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. (Kusrini, 2009). Implementasi menggunakan *software* Rapid Miner digunakan untuk membantu menemukan nilai yang akurat (Nugroho, 2014). Maka diharapkan dengan metode algoritma *naïve bayes* menggunakan *software* Rapid Miner dapat membantu menemukan variabel atau nilai yang akurat dalam seleksi penerimaan mahasiswa baru jalur penelusuran siswa berprestasi (PSB) pada Fakultas Teknik Industri, Prodi Teknik Industri.

2. Studi Literatur

Banyak penelitian mengenai aplikasi data mining khususnya teknik klasifikasi diberbagai bidang seperti bidang medis, marketing, manufaktur dan di bidang akademik. Chaurasia dan Pal (2013) memprediksi kemungkinan pasien jantung berdasarkan teknik klasifikasi dengan menggunakan tiga algoritma klasifikasi sekaligus yaitu Naïve Bayes, J48 Decision Tree, dan algoritma Bagging. Dengan menggunakan 11 atribut seperti jenis kelamin, usia, tekanan darah, detak jantung dan kondisi fisik lainnya, penelitian ini berhasil mengklasifikasi pasien dengan penyakit jantung dengan keakuratan sekitar 85% dan waktu prediksi hanya sekitar 0,05 detik. Hal ini sangat bermanfaat untuk memprediksi kemungkinan seseorang terkena penyakit jantung hanya dengan melihat *record* data sebelumnya dalam waktu singkat dan keakuratan yang relatif tinggi.

Klasifikasi juga bisa dimanfaatkan untuk memprediksi kemungkinan menjadi nasabah di suatu bank, asuransi maupun ritel (Karim dan Rahman, 2013). Dengan menggunakan data individu seperti status, pekerjaan, pendidikan, status kepemilikan rumah penelitian tersebut dapat memprediksi kecenderungan seseorang membuka tabungan, memiliki kartu kredit atau menjadi member di sebuah tempat perbelanjaan. Algoritma yang dipakai adalah *decision tree* C4.5 dan Naïve Bayes sebagai algoritma klasifikasi yang mudah dipalिकासikan dan dipahami dengan tingkat akurasi yang relative tinggi.

Bidang manufaktur juga memanfaatkan Teknik klasifikasi ini dalam pengambilan keputusan. Seperti halnya sebuah studi di Taiwan yang dilakukan pada 66 data untuk memprediksi kegagalan (*failure*) dimana 43 data digunakan sebagai data training set dan 23 data digunakan sebagai data testing (Yeh, dkk, 2011). Dengan tingkat keakuratan 97.6% untuk

training set dan 86.9% untuk data testing, penelitian tersebut menghasilkan rules klasifikasi kegagalan proses molding berdasarkan beberapa atribut yang terjadi selama proses manufaktur antara lain: Suhu injeksi, velocity, waktu pengemasan, tekanan injeksi, waktu injeksi dan sebagainya.

Di bidang akademik, Thomas dan Galambos (2004) menggunakan teknik klasifikasi decision tree untuk memprediksi kepuasan pelajar terhadap sekolah. Hal ini dilakukan untuk memprediksi apakah pelajar yang bersangkutan melanjutkan jenjang pendidikan ke tingkat selanjutnya di sekolah yang sama atau tidak, sekaligus untuk menentukan atribut apa saja yang membuat pelajar tersebut memutuskan untuk melanjutkan pendidikan di sekolah yang sama. Decision tree juga dapat memprediksi dan mengevaluasi performansi mahasiswa di suatu universitas (Baradwaj dan Pal, 2011). Hasilnya dapat membantu universitas untuk mengidentifikasi lebih dini mahasiswa yang terancam dropout sehingga dapat diberikan penanganan khusus dan memberikan kesempatan bagi dosen pembimbing untuk menyediakan bimbingan dan konseling.

Osmanbegović dan Suljić (2012) telah memanfaatkan Teknik Naïve Bayes ini dibandingkan dengan dua teknik klasifikasi lainnya yaitu decision tree dan Multilayer Perceptron (MLP) untuk memprediksi performansi mahasiswa berdasarkan 12 atribut antara lain yaitu: jenis kelamin, asal SMA, jumlah keluarga, jarak rumah, beasiswa, waktu belajar, nilai ujian masuk dan sebagainya. Hasilnya menunjukkan bahwa algoritma decision tree lebih mudah untuk diaplikasikan dan dipahami dibandingkan algoritma lain dalam memprediksi performansi mahasiswa meskipun tingkat akurasi masih dibawah dua algoritma lain.

Lillyan Hadjaratie melakukan penelitian tentang “Prediksi dan Pemetaan Data Mahasiswa Fakultas Teknik Menggunakan Pendekatan Data Mining” Dalam penelitian ini bertujuan untuk membangkitkan informasi dan pengetahuan dari data akademik kemahasiswaan dengan: (1) Mengklasifikasi data mahasiswa aktif dan lulusan untuk memprediksi kategori IPK dan kategori Lama Studi; (2) Mengklaster data mahasiswa aktif dan lulusan berdasarkan kategori IPK dan Lama_Studi. Pada penelitian ini menggunakan metode *Decision Tree*, Artificial Neural Network dan K-Nearest Neighbour, sedangkan metode yang digunakan untuk mengklaster data mahasiswa aktif dan lulusan adalah metode Hirarkis (*Average Between Linkage*) dan Non-Hirarkis (*K-Means*). Penelitian ini menghasilkan informasi akademik berupa hasil klasifikasi data mahasiswa aktif dan lulusan untuk memprediksi kategori IPK dan Lama_Studi serta pemetaan data mahasiswa aktif dan lulusan berdasarkan sasaran mutu IPK.

3. Tujuan Penelitian

Tujuan utama dari penelitian ini adalah untuk mengidentifikasi pola berdasarkan hasil teknik klasifikasi pada prediksi prestasi akademik mahasiswa yang masuk melalui jalur Penelusuran Siswa Berprestasi (PSB) terhadap IPK di jurusan Teknik Industri Universitas Islam Indonesia berikut dengan tingkat akurasi yang dihasilkan.

4. Metode

Pengumpulan data dilakukan dengan mengekstrak data admisi mahasiswa baru sejumlah lebih dari 200 data selama 3 tahun untuk kemudian dilakukan analisis relevansi untuk memilih atribut yang relevan berdasarkan logika pengambilan keputusan. IPK adalah atribut prediksi dengan pembagian IPK ≥ 3.5 disebut terpuji, $2.75 \leq \text{IPK} < 3.5$ disebut memuaskan, $2.00 \leq \text{IPK} < 3.00$ disebut cukup, sedangkan analisis relevansi ini menghasilkan 8 atribut yang digunakan sebagai dasar klasifikasi seperti yang disajikan dalam table berikut:

Tabel 1. Atribut dan Kategori

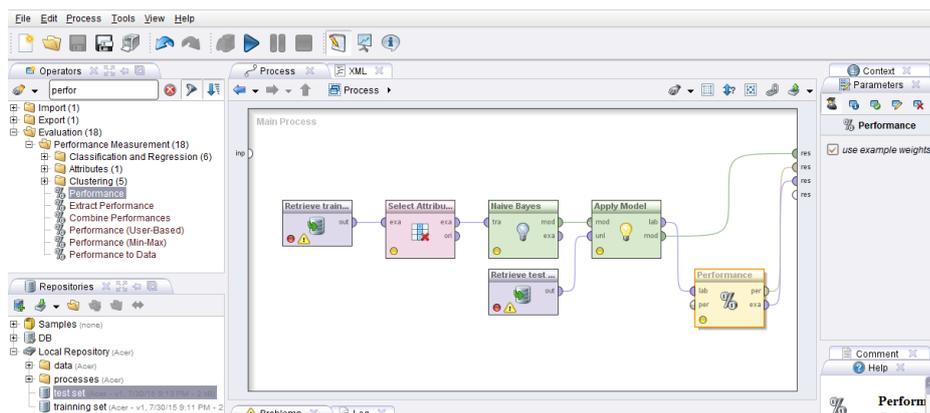
No	Atribut	Kategori
1	Rata-rata nilai rapot	Sangat baik, jika nilai PSB/ Rata-rata Nilai Rapot ≥ 85 , Baik, jika $70 \leq \text{Nilai PSB} < 85$
2	Asal Wilayah/daerah	Jawa Barat, Jawa Tengah, Jawa Timur dan Luar Jawa
3	Status Sekolah	Negeri, Swasta, MA dan lainnya.

4	Jurusan sekolah	IPA dan lainnya.
5	Pekerjaan orang tua	PNS, Pegawai Swasta, Wiraswasta, Petani dan lainnya
6	Prestasi yang diraih	Ya atau Tidak
7	Jenis Kelamin	Laki-laki dan Perempuan
8	Jalur PSB	Akademik, Beasiswa Bibit Unggul dan Olah Raga Dan Seni.

Data cleaning dilakukan terhadap lebih dari 200 data meliputi data *Incomplete*: berisi nilai yang salah yang disebabkan karena tidak terecord (*not recorded*), tidak tersedia (*not available*), sengaja dihapus karena dianggap salah atau tidak penting, data *Noisy*: berisi data salah yang tidak wajar (*Anomalies/outliers*) karena instrumen pengumpulan data yang digunakan mungkin salah, kesalahan manusia atau komputer yang terjadi pada saat memasukkan data dan kesalahan dalam transmisi data, dan data *Inconsistent*: berisi data yang tidak konsisten dan berbeda dari sewajarnya disebabkan oleh perubahan kode di sistem yang menjadikan data tidak konsisten

5. Hasil dan Pembahasan

Perhitungan prediksi menggunakan *software* aplikasi data mining Rapidminer dengan *interface* seperti pada gambar berikut:



Gambar 1. Interface Prediksi dengan RapidMiner

Akurasi

Akurasi merupakan tingkat ketepatan antara data di lapangan dengan data hasil perhitungan Software Rapid Miner. Maka dapat dilihat hasil atau nilai akurasi adalah sebesar 51.69%. Dari 9 data yang diprediksikan dengan nilai IPK terpuji terdapat 3 data yang benar dengan nilai IPK terpuji dan 6 data mendapatkan nilai IPK memuaskan. Dari 77 data yang diprediksikan mendapatkan nilai IPK memuaskan, sebanyak 42 data yang mendapatkan nilai IPK memuaskan dan 35 data tidak mendapatkan nilai IPK memuaskan yang terdiri dari 26 data yang mendapatkan nilai IPK terpuji dan 9 data yang mendapatkan nilai IPK cukup. Selanjutnya, dari 3 data yang diprediksi mendapatkan nilai IPK cukup dengan 1 data yang mendapatkan nilai IPK cukup dan sisanya mendapatkan nilai IPK memuaskan.

Keputusan

Naive Bayes digunakan untuk menghitung probabilitas terjadinya suatu peristiwa berdasarkan pengaruh yang didapat dari hasil observasi. Nilai probabilitas digunakan sebagai dasar untuk memprediksi prestasi akademik mahasiswa jalur penelusuran siswa berprestasi (PSB) dapat dilihat dari pola atribut yang terpilih.

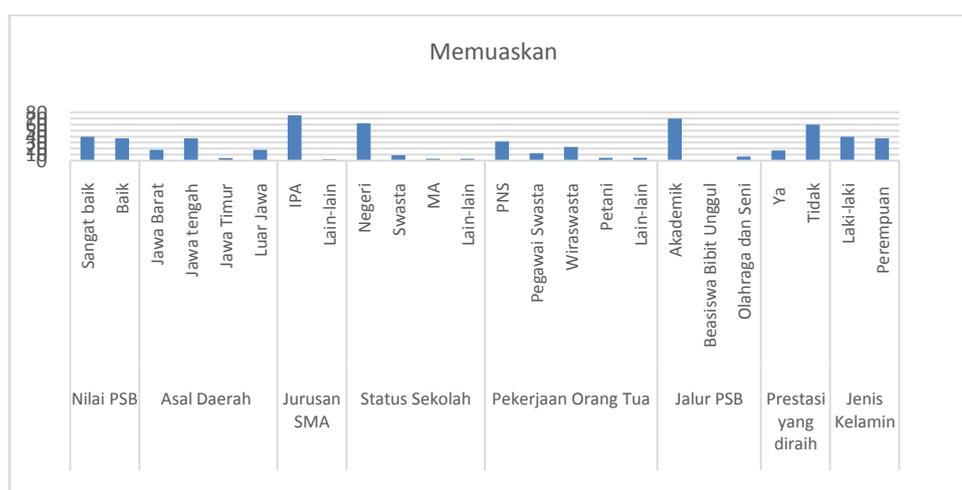
Tabel 2.Contoh Nilai probabilitas Masing-masing Atribut

Atribut	Parameter	Terpuji	Memuaskan	Cukup
NILAI PSB	value=Sangat Baik	0.57105	0.50838	0.39978
NILAI PSB	value=Baik	0.42842	0.49144	0.59911
NILAI PSB	value=unknown	0.00053	0.00019	0.00111
ASAL DAERAH	value=Jawa Barat	0.04802	0.32192	0.20000
ASAL DAERAH	value=Luar Jawa	0.42797	0.32192	0.20000
ASAL DAERAH	value=Jawa Timur	0.14301	0.06792	0.20000

Tabel 3. Contoh Hasil Prediksi menggunakan 8 atribut

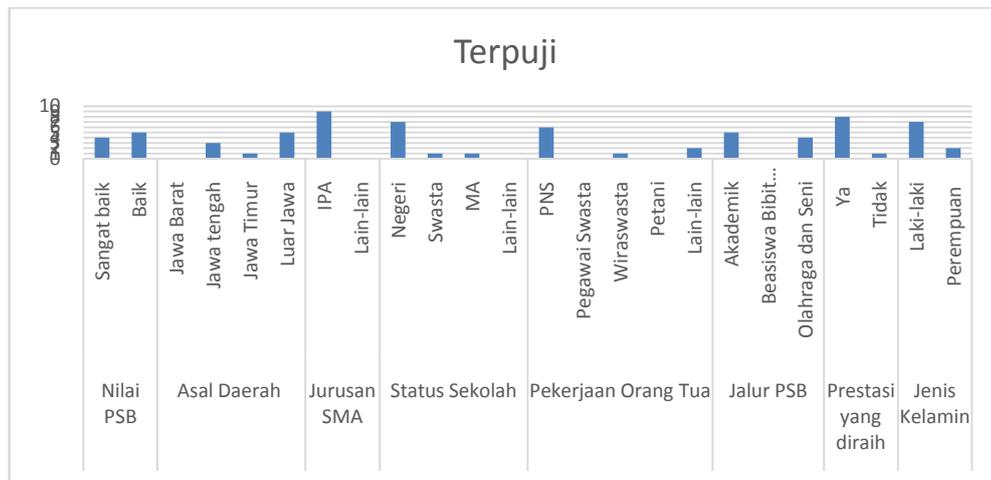
No	Nilai PSB	Asal Daerah	Jurusan SMA	Status Sekolah	Pek. Ortu	Jalur PSB	Pres	Jenis Kelamin	Terpuji	Memuaskan	Cukup	IPK
1	Baik	Jawa Tengah	IPA	Swasta	Lain-Lain	Aka	Tidak	Laki-Laki	0.13767	0.3959	0.4664	Cukup
2	Baik	Jawa Tengah	IPA	Swasta	Lain-Lain	Aka	Tidak	Laki-Laki	0.13767	0.3959	0.4664	Cukup
3	Baik	Jawa Tengah	IPA	Swasta	Pns	Aka	Tidak	Laki-Laki	0.15541	0.3784	0.4661	Cukup
4	Baik	Jawa Tengah	IPA	Negeri	Pns	Aka	Tidak	Pr	0.19447	0.604	0.2014	Memuaskan
5	Baik	Jawa Tengah	IPA	Negeri	Swasta	OR dan Seni	Ya	Pr	0.49047	0.5091	0.76E	Memuaskan

Dari Table diatas sebagai contoh dapat dilihat pola prediksi keputusan prestasi akademik mahasiswa terhadap atribut yang terpilih. Pada baris pertama, mahasiswa dengan nilai PSB Baik yaitu $70 \leq \text{Nilai PSB} < 85$, asal daerah Jawa Tengah, jurusan SMA IPA, status sekolah swasta, pekerjaan orang tua lain-lain, Jalur PSB akademik, prestasi yang diraih tidak, dan jenis kelamin laki-laki akan mendapatkan indeks prestasi akademik (IPK) dengan nilai cukup yaitu $2,00 \leq \text{IPK} < 2,75$. Berikut disajikan hasil prediksi keputusan prestasi akademik mahasiswa jalur penelusuran siswa berprestasi (PSB) dalam bentuk grafik untuk prediksi IPK terpuji dan memuaskan:

**Gambar 2** Nilai Indeks Prestasi Kumulatif (IPK) dengan keputusan Memuaskan

Dari gambar grafik diatas ini ditunjukkan bahwa prestasi mahasiswa jalur penelusuran siswa berprestasi (PSB) berdasarkan performansi variabel dengan parameter yang paling tinggi untuk nilai prediksi keputusan dengan nilai IPK memuaskan adalah variabel nilai PSB dengan nilai sangat baik, variabel asal daerah rata-rata yang berasal dari Jawa Tengah, variabel jurusan SMA dengan jurusan IPA, variabel status sekolah dengan tamatan dari sekolah negeri, variabel pekerjaan orang tua dengan pekerjaan PNS, variabel jalur PSB dengan jalur masuk nilai

akademik, variabel prestasi yang diraih tidak ada prestasi, dan variabel jenis kelamin yaitu laki-laki.



Gambar 4. Nilai Indeks Prestasi Kumulatif (IPK) dengan keputusan Terpuji

Dari gambar grafik diatas ini ditunjukkan bahwa prestasi mahasiswa jalur penelusuran siswa berprestasi (PSB) berdasarkan performansi variabel dengan parameter yang paling tinggi untuk nilai prediksi keputusan dengan nilai IPK terpuji adalah variabel nilai PSB dengan nilai baik, variabel asal daerah rata-rata yang berasal dari Luar Jawa, variabel jurusan SMA dengan jurusan IPA, variabel status sekolah dengan tamatan dari sekolah negeri, variabel pekerjaan orang tua dengan pekerjaan PNS, variabel jalur PSB dengan jalur masuk nilai akademik, variabel prestasi yang diraih iya ada prestasi, dan variabel jenis kelamin yaitu laki-laki.

6. Simpulan

Dari hasil penelitian yang dilakukan dengan menggunakan metode naïve bayes untuk prediksi prestasi akademik mahasiswa jalur penelusuran siswa berprestasi (PSB) di Universitas Islam Indonesia (UII), didapatkan kesimpulan bahwa:

1. Hasil pola prediksi keputusan prestasi akademik mahasiswa jalur Penelusuran Siswa Berprestasi (PSB) yang mendapatkan IPK diatas 3.5 adalah mahasiswa dengan kategori: nilai rata-rata rapot ≥ 85 , mahasiswa berasal dari Luar Jawa, jurusan IPA, berasal dari sekolah negeri, pekerjaan orang tua PNS, jalur masuk PSB melalui jalur nilai akademik, memiliki prestasi dan jenis kelamin laki-laki.
2. Tingkat akurasi yang terbentuk adalah sebesar 51.69% artinya model yang terbentuk kurang representatif, hal ini disebabkan penggunaan seluruh atribut tanpa dilakukan uji independensi pada proses perhitungan dan data yang didapatkan terlalu kecil sehingga menghasilkan tingkat akurasi yang kecil.

Daftar Pustaka

- Baradwaj K Brijesh, Pal Saurabh, 2011, *Mining Educational Data to Analyze Students' Performance*, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, page 63
- Chaurasia Vikas, Pal Saurabh, 2013, *Data Mining Approach to Detect Heart Diseses*, *International Journal of Advanced Computer Science and Information Technology (IJACSIT)* Vol. 2, No. 4, Page: 56-66, ISSN: 2296-1739
- Fithri, Diana Laily & Darmanto, Eko.(2014). *Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes*. Program Studi Sistem Informasi Fakultas Teknik. Universitas Muria. Kudus
- Han, J. dan Kamber, M. (2000). *Data Mining: Concepts And Techniques*. New York: Morgan-Kaufman.
- Hadjaratie, Lillyan "Prediksi dan Pemetaan Data Mahasiswa Fakultas Teknik Menggunakan Pendekatan Data Mining". Jurusan Teknik Informatika Universitas Negeri Gorontalo

- Karim Masud, Rahman M Rashedur , 2013, *Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing*, Journal of Software Engineering and Applications, 6, 196-206.
- Nugroho, Yuda Septian. (2014). *Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro*. Jurusan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
- Osmanbegovic, Edin dan Suljic, Mirza. 2012. *Data Mining Approach For Predicting Student Performance*. Journal of Economics and Business
- Ningrum, Tita Pavita, 2014. *Prestasi Akademik Mahasiswa Ditinjau Dari Sitem Jalur Penerimaan Mahasiswa Baru Di Universitas Gajah Mada*. Universitas Gajah Mada. Yogyakarta
- Osmanbegović Edin, Suljić Mirza, 2012, *Data Mining Approach For Predicting Student Performance*, Economic Review – Journal of Economics and Business, Vol. X, Issue 1.
- Ridwan, Mujib. 2013. *Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. Jurnal EECCIS
- Tan, Pang-Ning. (2004). *Introduction To Data Mining*.
- Thomas H Emily, Galambos Nora, 2004, *What Satisfies Students? Mining Student-Opinion Data with Regression and Decision Tree Analysis* *Research in Higher Education*, Vol. 45, No. 3
- Wibowo, Jati Sasongko. (2012). *Rancang Bangun Program Koreksi Lembar Jawab Komputer Untuk Penerimaan Mahasiswa Baru Di Universitas Stikubank*. Fakultas Teknologi Informasi, Universitas Stikubank
- Yeh Duen-Yian, Cheng Ching-Hsue, Hsiao Shih-Chuan, 2011, *Classification Knowledge Discovery In Mold Tooling Test Using Decision Tree Algorithm*,) 22:585–595 DOI 10.1007/s10845-009-0321-7