

***DATA MINING* UNTUK MENINGKATKAN
KELULUSAN SISWA KELAS XII DENGAN
METODE NAÏVE BAYES**

SKRIPSI



Oleh:

Amrin Simanjuntak

150210071

**FAKULTAS TEKNIK DAN KOMPUTER
PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS PUTERA BATAM
TAHUN 2019**

**DATA MINING UNTUK MENINGKATKAN
KELULUSAN SISWA KELAS XII DENGAN
METODE NAÏVE BAYES**

SKRIPSI

Untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana



Oleh:

Amrin Simanjuntak

150210071

**FAKULTAS TEKNIK DAN KOMPUTER
PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS PUTERA BATAM
TAHUN 2019**

PERNYATAAN

Dengan ini saya menyatakan bahwa:

Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana, dan/atau magister), baik di Universitas Putera Batam maupun di perguruan tinggi lain. Skripsi ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan pembimbing. Dalam skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di perguruan tinggi.

Batam, 16 Februari 2019
Yang membuat pernyataan,

Amrin Simanjuntak
150210071

***DATA MINING* UNTUK MENINGKATKAN
KELULUSAN SISWA KELAS XII DENGAN
METODE NAÏVE BAYES**

**Oleh
Amrin Simanjuntak
150210071**

SKRIPSI

**Untuk memenuhi salah satu syarat
guna memperoleh gelar Sarjana**

**Telah disetujui oleh Pembimbing pada tanggal
seperti tertera di bawah ini**

Batam, 16 Februari 2019

**Sunarsan Sitohang, S.Kom., M.TI
Pembimbing**

ABSTRAK

SMA NEGERI 16 BATAM adalah sekolah yang masih akreditasi B dan memiliki penurunan kelulusan siswa masuk PTN (Perguruan Tinggi Negeri) lewat jalur SNMPTN (Seleksi Nasional Masuk Perguruan Tinggi Negeri). Hal tersebut akan menjadi penunjang bagi pihak sekolah untuk meningkatkan mutu dan kualitas proses belajar mengajar. Penelitian ini untuk mengetahui tingkat kelulusan siswa masuk PTN lewat jalur SNMPTN. Data nilai siswa diklasifikasikan kedalam prediksi siswa lulus dan tidak lulus. Input dari sistem ini adalah nilai raport siswa dari semester 1 sampai 5 sehingga memiliki atribut sebanyak 70 atribut. Data nilai siswa tahun 2017 yang dinyatakan lulus sebagai data training. Sedangkan data siswa tahun 2018 sebagai data testing. Data training akan diproses menggunakan teknik *Data Mining* algoritma *naïve bayes classifier (NBC)* untuk membentuk tabel probabilitas sebagai dasar proses klasifikasi kelulusan siswa. Ouput dari sistem ini adalah berupa tingkat akurasi siswa yang diprediksi kelulusannya. Data nilai siswa dirata-ratakan permata pelajaran dalam satu semester, selanjutnya dikategorikan. Data nilai siswa akan dilakukan dalam 2 kali percobaan yaitu berdasarkan rangking dan drop. Pengujian pada data siswa 2018 jurusan IPS, dengan algoritma naïve bayes menghasilkan nilai akurasi, *recall*, dan *precision* masing-masing 80%, 100% dan 80%. Hasil pengujian yang menunjukkan bahwa faktor yang paling berpengaruh dalam penentuan klasifikasi kelulusan siswa yaitu rata-rata nilai siswa setiap mata pelajaran dalam 5 semester. Sehingga faktor-faktor tersebut dapat digunakan sebagai bahan evaluasi dan pertimbangan bagi perguruan tinggi negeri.

Kata Kunci: Kelulusan SNMPTN, *Data Mining*, *Naïve Bayes Clasifier*

ABSTRACT

SMA 16 BATAM is a school that is still accredited B and has decreased graduation of students entering PTN (State Universities) through the SNMPTN (National Selection to Enter State Universities). This will be a support for the school to improve the quality of the teaching and learning process. This research is to determine the graduation rate of students entering PTN through the SNMPTN pathway. Classified into predictions of students graduate and not graduate. Input from this system is the value of student report cards from semester 1 to 5 and have 70 attributes. Data report students of 2017 who were said to had passed as training data. While the student's data on 2018 become testing data. Training data will be processed using naïve bayes classifier (NBC) Data Mining techniques to form probability tables as the basis for the student graduation classification. The output of this system is the level of accuracy from the students predicted for graduation. The student's report will be averaged per subject in one semester and then be categorized. Training data and testing data will be done in twice trial that is by ranking and full student. Trial for students majoring in Social Sciences, using naïve Bayes algorithm have result that is accuracy, recall, and precision values of 80%, 100% and 80% respectively. The test results show that the most influential factor in determining student graduation classification is the average student score for each subject in 5 semesters. So that these factors can be used as material for evaluation and consideration for state universities.

Keywords: SNMPTN Graduation, Data Mining, Naïve Bayes Classifier

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati.

Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Ibu Dr. Nur Elfi Husda S.Kom M.Kom selaku Rektor Universitas Putera Batam.
2. Bapak Andi Maslan S.T M.Si Ketua Program Studi teknik informatika
3. Sunarsan Sitohang S.Kom., M.TI selaku pembimbing Skripsi dan pembimbing akademik pada Program Studi teknik informatika Universitas Putera Batam.
4. Dosen dan Staff Universitas Putera Batam
5. Orang Tua yang telah mendukung baik secara materi ataupun non materi.
6. saudara kakak dan abang Erni Simanjuntak, Yani Simanjuntak, Arnold Simanjuntak dan Erisna Simanjuntak yang telah mendukung baik secara materi ataupun nonmateri.
7. Cindy, Karolina, Nurmanti. Meliani, Vera selaku sahabat yang yang membantu dan mendukung secara materi ataupun nonmateri.
8. teman-teman seperjuangan dan rekan kerja

Semoga Tuhan membalas kebaikan dan selalu mencurahkan berkah serta anugerahNYA, Amin.

Batam, 16 Februari 2019

Amrin simanjuntak

DAFTAR ISI

HALAMAN SAMPUL DEPAN	
HALAMAN JUDUL	ii
HALAMAN PERNYATAAN.....	iii
HALAMAN PENGESAHAN	iv
ABSTRAK.....	v
ABSTRACT	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	ii
DAFTAR RUMUS	ii
 BAB 1 PENDAHULUAN.....	 1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah Penelitian	4
1.3 Batasan Masalah.....	4
1.4 Rumusan Masalah	5
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	5
 BAB II LANDASAN TEORI.....	 7
2.1 <i>Knowledge Discovery In Database (KDD)</i>	7
2.2 <i>Data Mining</i>	9
2.2.1 Definisi <i>Data Mining</i>	9
2.2.2 Manfaat <i>Data Mining</i>	10
2.2.3 Pengelompokan <i>Data Mining</i>	11
2.2.4 Tahapan <i>Data Mining</i>	13
2.3 SNMPTN.....	14
2.4 Metode Naïve Bayes	15
2.5 Software Pendukung	17
2.6 Penelitian Terdahulu	17
2.7 Kerangka Pemikiran.....	21
 BAB III METODE PENELITIAN	 23
3.1 Desain Penelitian.....	23
3.2 Pengumpulan Data	26
3.2.1 Teknik Pengumpulan Data	27
3.3 Operasional Variabel.....	29
3.4 Metode Perancangan Sistem	30
3.5 Lokasi Dan Jadwal Penelitian	32
3.5.1 Lokasi Penelitian.....	32
3.5.2 Jadwal Penelitian.....	32

BAB IV HASIL DAN PEMBAHASAN	33
4.1 Analisa Data	33
4.1.1 Pengumpulan Data.....	33
4.1.3 Kategorik Nilai	38
4.2 Proses Import Data	42
4.2.1 Pola Naïve Bayes Pada Rapidminer	44
4.3 Pengujian Algoritma.....	48
4.3.1 Studi Kasus Rangking	48
4.2.3 Pengujian Sampel Studi Kasus Rangking.	51
4.2.4 Studi Kasus Sistem <i>Drop</i>	56
4.2.5 Pengujian Sampel Kasus Sistem <i>Drop</i>	58
 BAB V KESIMPULAN DAN SARAN	 64
5.1 Kesimpulan	64
5.2 Saran.....	64
DAFTAR PUSTAKA.....	66
 RIWAYAT HIDUP	
SURAT KETERANGAN PENELITIAN	
LAMPIRAN	

DAFTAR TABEL

	Halaman
Tabel 3.1 Jadwal Penelitian	32
Tabel 4.1 Daftar Siswa Jurusan IPS.....	34
Tabel 4.2 Siswa 45 Rangking Teratas.....	35
Tabel 4.3 <i>Training Data</i>	37
Tabel 4.4 Tabel Kategori	38
Tabel 4.5 Nilai Siswa Kategori	39
Tabel 4.6 Data Rangking	40
Tabel 4.7 Data <i>Training</i>	41
Tabel 4.8 Proses Hitung Manual Sampel Lulus.....	51
Tabel 4.9 Proses hitung Manual Sampel Tidak Lulus	53
Tabel 4.10 Proses Hitung Manual Sampel Tidak Lulus	58
Tabel 4.11 Proses Hitung Manual Sampel Tidak Lulus	60

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Ilustrasi Proses Tahapan KDD	7
Gambar 2.2 Gambaran Tahap-Tahap <i>Data Mining</i>	13
Gambar 2.3 Logo Rapidminer.....	17
Gambar 2.4 Kerangka Pemikiran	21
Gambar 3.1 Desain Penelitian.....	24
Gambar 3.2 Flowchart Naïve Bayes	30
Gambar 4.1 Halaman Depan Rapidminer	43
Gambar 4.2 Proses <i>Import data</i>	43
Gambar 4.3 Hasil Data <i>Trainig</i> Dan Data <i>Testing</i>	44
Gambar 4.4 <i>Input Data Training Dan Testing</i>	45
Gambar 4.5 Naïve Bayes.....	45
Gambar 4.6 <i>Apply model</i>	46
Gambar 4.7 <i>Performances</i>	46
Gambar 4.8 Pola Pemodelan Naïve Bayes.....	47
Gambar 4.9 Pola Naïve Bayes Sistem Rangking	48
Gambar 4.10 <i>Result</i> Naïve Bayes.....	49
Gambar 4.11 Hasil <i>Performance Ranking</i>	50
Gambar 4.12 Hasil Sampel Hitungan Rapidminer.....	56
Gambar 4.13 Pola Naïve Bayes Sistem Rangking	57
Gambar 4.14 <i>Result</i> Naïve Bayes.....	57
Gambar 4.15 Hasil Hitungan Sampel Drop	63

DAFTAR RUMUS

	Halaman
Rumus 2.1 Naïve Bayes.....	15
Rumus 2.2 Probabilitas.....	16
Rumus 2.3 Hitung Probabilitas.....	16