



PROPOSAL TUGAS AKHIR

**ANALISIS DATA PENERIMAAN
MAHASISWA BARU
MENGUNAKAN ALGORITMA
NAÏVE BAYES - STUDI KASUS:
INSTITUT TEKNOLOGI
KALIMANTAN**

Gardian Bimo Lukito
NIM. 10141003

Dosen Pembimbing
Subchan, Ph.D
Yuyun Tri Wiranti, S.Kom., M.MT.

Program Studi Sistem Informasi
Jurusan Matematika dan Teknologi Informasi
Institut Teknologi Kalimantan
Balikpapan 2018

LEMBAR PERSETUJUAN

Proposal Tugas Akhir dengan judul :

“ANALISIS DATA PENERIMAAN MAHASISWA BARU MENGGUNAKAN
ALGORITMA NAÏVE BAYES - STUDI KASUS: INSTITUT TEKNOLOGI
KALIMANTAN ”

Yang disusun oleh :

Gardian Bimo Lukito
NIM. 10141003

Telah diperiksa dan disetujui oleh dosen pembimbing:

Dosen Pembimbing Utama

Dosen Pembimbing Pendamping

Subchan, Ph.D
NIP. 197105131997021001

Yuyun Tri Wiranti, S.Kom., M.MT.
NIPH. 100116058

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat dan anugerah-Nya sehingga penulis dapat menyelesaikan proposal tugas akhir yang berjudul:

**“ANALISIS DATA PENERIMAAN MAHASISWA BARU
MENGUNAKAN ALGORITMA NAÏVE BAYES - STUDI KASUS:
INSTITUT TEKNOLOGI KALIMANTAN ”**

Proposal tugas akhir ini merupakan salah satu syarat yang harus ditempuh untuk menyelesaikan Program Sarjana di Program Studi Sistem Informasi, Jurusan Matematika dan Teknologi Informasi, Institut Teknologi Kalimantan (ITK) Balikpapan. Untuk itu penulis mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Bapak Subchan, Ph.D selaku Dosen Pembimbing Utama dan Ibu Yuyun Tri Wiranti, S.Kom., M.MT selaku Dosen Pembimbing Pendamping.
2. Ibu Vinda Daningrum., S.Kom., M.Kom., M.IM selaku Koordinator Program Studi Sistem Informasi Jurusan Matematika dan Teknologi Informasi ITK.
3. Serta semua pihak yang terlibat dalam penyusunan proposal tugas akhir ini.

Penulis menyadari bahwa penyusunan proposal tugas akhir ini masih jauh dari sempurna, karena itu penulis mengharapkan segala kritik dan saran yang membangun. Semoga tugas akhir ini dapat bermanfaat bagi kita semua. Atas perhatiannya penulis mengucapkan terima kasih.

Balikpapan, Januari 2018

Penulis

ANALISIS DATA PENERIMAAN MAHASISWA BARU MENGUNAKAN ALGORITMA NAÏVE BAYES - STUDI KASUS: INSTITUT TEKNOLOGI KALIMANTAN

Nama Mahasiswa : Gardian Bimo Lukito
NIM : 10141003
Dosen Pembimbing Utama : Subchan, Ph.D
Dosen Pembimbing Pendamping : Yuyun Tri Wiranti, S.Kom., M.MT.

ABSTRAK

Perguruan tinggi sebagai penyelenggara pendidikan setelah pendidikan menengah menerima calon mahasiswa yang berprestasi akademik tinggi dan diprediksi akan berhasil menyelesaikan studi di perguruan tinggi berdasarkan prestasi akademik. Sejak tahun 2011, Kementerian Pendidikan dan Kebudayaan membuka kesempatan bagi siswa Sekolah Menengah Atas (SMA) tingkat akhir untuk mendaftar langsung ke Program Studi (Prodi) dan Perguruan Tinggi Negeri (PTN) yang diminatinya melalui sekolah masing-masing. Hal ini dilakukan oleh semua PTN salah satunya yaitu Institut Teknologi Kalimantan (ITK). Proses seleksi dalam menentukan kelulusan calon mahasiswa baru di ITK masih dilakukan secara manual sehingga membutuhkan waktu yang cukup lama. Sedangkan pemanfaatan mengenai data penerimaan mahasiswa baru melalui jalur ini hanya sebatas sebagai keperluan beasiswa bagi mahasiswa yang bersangkutan dan keperluan akreditasi institut, belum terdapat pemanfaatan untuk keperluan penelitian. Oleh karena itu akan dilakukan penelitian terhadap data penerimaan mahasiswa baru di ITK menggunakan teknik *Knowledge Discovery in Databases* (KDD) dengan algoritma Naïve Bayes. Proses yang dilakukan pada penelitian ini yaitu seluruh tahapan yang terdapat dalam KDD meliputi *data selection*, *pre-processing/cleaning*, *transformation*, *data mining*, dan *interpretation/evaluation*. Hasil yang diharapkan dari penelitian ini adalah dapat menghasilkan suatu pengetahuan atau informasi yang bermanfaat yang dapat mempermudah pimpinan dalam mengambil keputusan menentukan status kelulusan calon mahasiswa, serta dapat mengetahui atribut yang memiliki pengaruh signifikan pada data tersebut.

Kata kunci :

Penerimaan Mahasiswa Baru, *Knowledge Discovery in Databases*, Naïve Bayes

DAFTAR ISI

Lembar Persetujuan	ii
Kata Pengantar	iii
Abstrak	iv
Daftar Isi	v
Daftar Gambar.....	vi
Daftar Tabel	vii
BAB 1 Pendahuluan	
1.1. Latar Belakang	1
1.2. Perumusan Masalah.....	4
1.3. Tujuan Penelitian.....	4
1.4. Manfaat Penelitian.....	5
1.5. Kerangka Pemikiran Penelitian	5
BAB 2 Tinjauan Pustaka	
2.1. Seleksi Nasional Masuk Perguruan Tinggi Negeri	6
2.1.1. Tujuan SNMPTN	6
2.1.2. Ketentuan Umum dan Persyaratan SNMPTN	6
2.2. <i>Knowledge Discovery in Databases</i>	8
2.3. <i>Data mining</i>	9
2.4. Metode Klasifikasi	10
2.5. Algoritma Klasifikasi	10
2.6. Algoritma Naïve Bayes	11
2.7. Pengukuran Kinerja Klasifikasi	13
2.8. Penelitian Terdahulu	14
BAB 3 Metode Penelitian	
3.1. Metodologi	17
3.2. Rencana Jadwal Penelitian	24
Daftar Pustaka	
Lampiran A <i>Interview Protocol</i>	

DAFTAR GAMBAR

Gambar 1.1 Bagan kerangka pemikiran penelitian	5
Gambar 2.1 Tahap penemuan <i>knowledge</i> pada KDD	8
Gambar 3.1 Bagan metodologi bagian 1	17
Gambar 3.2 Bagan metodologi bagian 2	18
Gambar 3.3 Diagram alir tahap <i>data selection</i>	19
Gambar 3.4 Diagram alir tahap <i>pre-processing/cleaning</i>	20
Gambar 3.5 Diagram alir tahap <i>transformation</i>	21
Gambar 3.6 Diagram alir algoritma Naïve Bayes	22
Gambar 3.7 Diagram alir tahap <i>data mining</i>	23
Gambar 3.8 Diagram alir tahap <i>interpretation/evaluation</i>	24

DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu	15
Tabel 3.1 Rencana jadwal penelitian	24

BAB 1

PENDAHULUAN

Bab ini merupakan deskripsi singkat dari bab 1 meliputi latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian dan kerangka penelitian. Pada bagian latar belakang akan dijelaskan masalah pentingnya penerapan *knowledge discovery in databases* (KDD) pada sebuah data khususnya data penerimaan mahasiswa baru jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) di Institut Teknologi Kalimantan (ITK). Berdasarkan latar belakang tersebut dirumuskan beberapa masalah yang akan diteliti tentang penerapan KDD pada data tersebut. Selanjutnya dijelaskan pula tujuan yang akan dicapai pada penelitian ini serta manfaat yang dapat diberikan kepada ITK.

1.1. Latar Belakang

Penerimaan mahasiswa baru harus memenuhi prinsip adil, akuntabel, transparan, dan tidak diskriminatif dengan tidak membedakan jenis kelamin, agama, suku, ras, kedudukan sosial, dan tingkat kemampuan ekonomi calon mahasiswa serta tetap memperhatikan potensi calon mahasiswa dan kekhususan perguruan tinggi (Karsidi, 2017). Perguruan tinggi sebagai penyelenggara pendidikan setelah pendidikan menengah menerima calon mahasiswa yang berprestasi akademik tinggi dan diprediksi akan berhasil menyelesaikan studi di perguruan tinggi berdasarkan prestasi akademik.

Sejak tahun 2011, Kementerian Pendidikan dan Kebudayaan membuka kesempatan bagi siswa SMA tingkat akhir untuk mendaftar langsung ke Program Studi (Prodi) dan Perguruan Tinggi Negeri (PTN) yang diminatinya melalui sekolah masing-masing (Setiawan dkk, 2015). Kesempatan ini diwujudkan melalui skema jalur undangan Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN). Hal ini dilakukan oleh banyak PTN, salah satunya yaitu Institut Teknologi Kalimantan (ITK).

ITK adalah perguruan tinggi yang menyelenggarakan program pendidikan akademik dan dapat menyelenggarakan program pendidikan vokasi dalam berbagai rumpun ilmu pengetahuan dan/atau teknologi serta jika memenuhi syarat dapat menyelenggarakan pendidikan profesi sesuai dengan ketentuan perundang-undangan (Nasir, 2017). ITK merupakan perguruan tinggi negeri di lingkungan Kementerian Riset, Teknologi, dan Pendidikan Tinggi yang berkedudukan di Balikpapan, Provinsi Kalimantan Timur. ITK didirikan berdasarkan Peraturan Presiden Nomor 125 Tahun 2014 tentang Pendirian Institut Teknologi Kalimantan.

Berdasarkan hasil wawancara dengan Ketua Pelaksana Penerimaan Mahasiswa Baru Jalur SNMPTN pada LAMPIRAN A, diketahui bahwa penerimaan mahasiswa baru di ITK melalui jalur SNMPTN dimulai sejak tahun 2016 hingga sekarang. Setiap tahunnya, jumlah mahasiswa baru yang diterima di ITK melalui jalur SNMPTN semakin meningkat. Namun proses seleksi dalam menentukan kelulusan peserta SNMPTN masih dilakukan secara manual sehingga membutuhkan waktu yang cukup lama. Sedangkan pemanfaatan mengenai data penerimaan mahasiswa baru melalui jalur ini hanya sebatas sebagai keperluan beasiswa bagi mahasiswa yang bersangkutan dan keperluan akreditasi institut, belum terdapat pemanfaatan untuk keperluan penelitian.

Penelitian yang memanfaatkan data penerimaan mahasiswa baru melalui jalur SNMPTN telah dilakukan oleh Tambun dan Anofrizen. Tambun dan Anofrizen menggunakan data SNMPTN Universitas Islam Negeri Sultan Syarif Kasim Riau (UIN Suska Riau) pada penelitiannya untuk menemukan hubungan tingkat kelulusan dengan jalur masuk ke UIN Suska Riau menggunakan *data mining* (Tambun dan Anofrizen, 2015). Tujuan dari penelitian tersebut adalah untuk mengetahui sejauh mana tingkat keberhasilan jalur masuk yang ada di UIN Suska Riau agar pimpinan membuka peluang yang lebih besar terhadap jalur dengan lulusan tepat waktu terbanyak. Hernawati dkk juga menggunakan data SNMPTN Universitas Negeri Yogyakarta pada penelitiannya untuk mengetahui pola hubungan data dasar SNMPTN dengan Indeks Prestasi Mahasiswa (IPK) (Hernawati dkk, 2015). Mereka menyebutkan bahwa data dengan jumlah yang besar dapat membuat suatu permasalahan bagi lembaga pendidikan jika tidak bisa dimanfaatkan. Semakin banyak data, semakin banyak pula usaha yang diperlukan

untuk memilah data mana yang dapat diolah menjadi informasi. Sehingga mereka menerapkan *data mining* pada data tersebut agar menghasilkan informasi yang bermanfaat untuk mempermudah pimpinan dalam membuat keputusan.

Berdasarkan dua penelitian yang telah diuraikan, kedua penelitian tersebut menerapkan *data mining* yang merupakan salah satu dari proses *Knowledge Discovery in Databases* (KDD). KDD merupakan serangkaian proses yang terdiri dari *data selection*, *pre-processing*, *transformation*, *data mining*, dan *interpretation* (Kusrini dan Luthfi, 2009). Dengan menerapkan seluruh proses KDD akan dihasilkan suatu nilai tambah berupa pengetahuan baru yang selama ini tidak diketahui secara manual dari sekumpulan data (Saputra dkk, 2016).

Data mining merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan metode tertentu (Kusrini dan Luthfi, 2009). *Data mining* terdiri dari beberapa metode, salah satunya adalah metode klasifikasi. Metode klasifikasi merupakan pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu (Defiyanti dan Jajuli, 2015).

Pada metode klasifikasi sendiri terdapat berbagai algoritma antara lain C4.5, *K-Nearest Neighbor* (KNN), dan Naïve Bayes. Penelitian yang dilakukan oleh Kamagi dan Hansun menyebutkan bahwa algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan dan memiliki tingkat akurasi yang dapat diterima, namun terdapat kelemahan yaitu algoritma tersebut hanya dapat digunakan jika *data training* dapat disimpan secara keseluruhan dan pada waktu yang bersamaan di memori karena dalam mengkonstruksi pohon algoritma C4.5 membaca seluruh sampel data training dari storage dan memuatnya ke memori (Kamagi dan Hansun, 2014). Selanjutnya pada penelitian yang dilakukan oleh Nisa menyebutkan bahwa algoritma KNN dapat menghasilkan data yang lebih akurat dan lebih efektif pada *data training* yang besar, namun terdapat kelemahan yaitu memiliki nilai komputasi yang tinggi dan perlu menentukan parameter 'k' yang optimal untuk menyatakan jumlah tetangga yang terdekat (Nisa, 2017). Pada penelitian yang dilakukan oleh Natalius menyebutkan bahwa algoritma Naïve Bayes dapat dipersonalisasi sedemikian rupa sehingga dapat disesuaikan dengan sifat dan kebutuhan masing-masing data (Natalius, 2010). Serta pada penelitian

yang dilakukan oleh Saleh menyebutkan bahwa algoritma Naïve Bayes hanya membutuhkan *data training* yang kecil sehingga komputasi yang dilakukan juga relatif rendah (Saleh, 2015).

Berdasarkan permasalahan yang terjadi di ITK dan pertimbangan metode yang telah dilakukan, maka penulis akan melakukan penelitian terhadap data penerimaan mahasiswa baru jalur SNMPTN di ITK menggunakan teknik KDD dengan algoritma Naïve Bayes. Pengolahan data tersebut diharapkan dapat menghasilkan suatu pengetahuan atau informasi yang bermanfaat yang dapat mempermudah pimpinan dalam mengambil keputusan menentukan status kelulusan calon mahasiswa, serta dapat mengetahui atribut yang berpengaruh pada data tersebut.

1.2. Perumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan sebelumnya, ditemukan bahwa belum terdapat pemanfaatan terhadap data penerimaan mahasiswa baru di ITK. Sehingga akan dilakukan KDD terhadap data tersebut dengan harapan dapat memberikan informasi yang bermanfaat untuk dijadikan sebagai dasar dalam pengambilan keputusan menggunakan metode klasifikasi dengan algoritma *Naïve Bayes*. Dari penjelasan tersebut, dapat dirumuskan permasalahan yang dibahas dalam penelitian ini yaitu:

1. Bagaimana mengklasifikasi status penerimaan calon mahasiswa pada data penerimaan mahasiswa baru menggunakan metode klasifikasi dengan algoritma *Naïve Bayes* di ITK?
2. Bagaimana menentukan atribut yang berpengaruh terhadap data penerimaan mahasiswa baru di ITK?
3. Bagaimana tingkat akurasi klasifikasi pada data penerimaan mahasiswa baru di ITK?

1.3. Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengetahui klasifikasi status penerimaan calon mahasiswa pada data penerimaan mahasiswa baru menggunakan metode klasifikasi dengan algoritma *Naïve Bayes* di ITK.
2. Menentukan atribut yang berpengaruh terhadap data penerimaan mahasiswa baru di ITK.
3. Mengetahui tingkat akurasi klasifikasi pada data penerimaan mahasiswa baru di ITK.

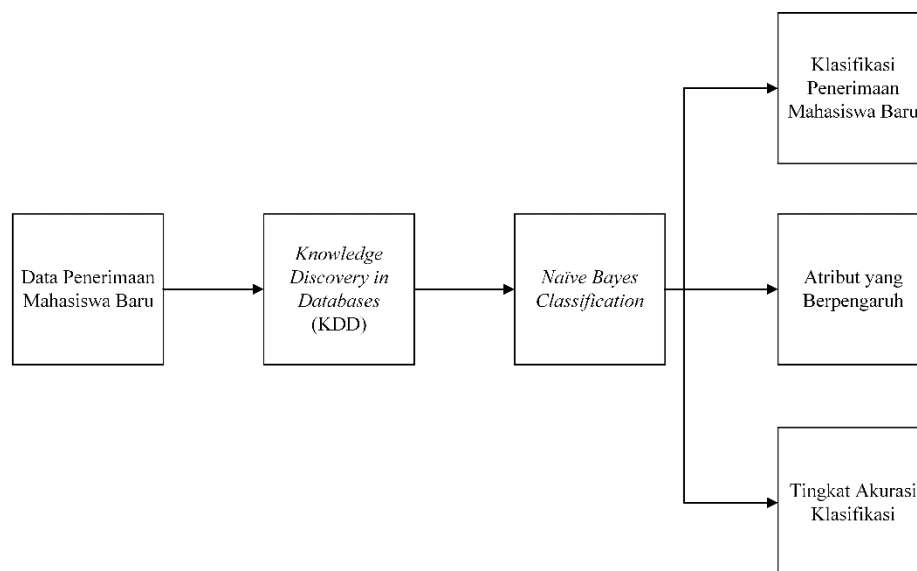
1.4. Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Membantu ITK dalam membuat keputusan terhadap penerimaan mahasiswa baru.
2. Merekomendasikan algoritma *Naïve Bayes* sebagai algoritma yang digunakan pada perancangan sistem informasi penerimaan mahasiswa baru di ITK.

1.5. Kerangka Pemikiran Penelitian

Kerangka pemikiran penelitian menjelaskan alur berpikir dalam pengerjaan penelitian mulai dari masalah hingga hasil yang didapat. Adapun kerangka pemikiran yang dimaksud dapat dilihat pada Gambar 1.1.



Gambar 1.1 Bagan kerangka pemikiran penelitian

BAB 2

TINJAUAN PUSTAKA

Pada bab ini akan dijelaskan mengenai teori-teori terkait *Knowledge Discovery in Databases* (KDD) yang bersumber dari buku, jurnal, ataupun artikel yang berfungsi sebagai dasar dalam mengerjakan tugas akhir.

2.1. Seleksi Nasional Masuk Perguruan Tinggi Negeri

Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) merupakan seleksi berdasarkan penelusuran prestasi akademik calon Mahasiswa dilakukan oleh masing-masing Perguruan Tinggi Negeri (PTN) di bawah koordinasi panitia pusat (Karsidi, 2017). SNMPTN diikuti seluruh PTN yang sudah ditetapkan oleh Majelis Rektor Perguruan Tinggi Negeri Indonesia (MRPTNI), diselenggarakan dalam suatu sistem yang terpadu dan serentak. Biaya pelaksanaan SNMPTN ditanggung oleh Pemerintah, sehingga peserta tidak dipungut biaya seleksi.

2.1.1. Tujuan SNMPTN

Adapun tujuan dari SNMPTN adalah sebagai berikut:

1. Memberikan kesempatan kepada siswa Sekolah Menengah Atas (SMA), Sekolah Menengah Kejuruan (SMK), Madrasah Aliyah (MA), atau yang sederajat di dalam dan luar negeri (Sekolah Republik Indonesia/SRI) yang memiliki prestasi unggul untuk menempuh pendidikan tinggi di PTN.
2. Memberikan peluang kepada PTN untuk mendapatkan calon mahasiswa baru yang mempunyai prestasi akademik tinggi.

2.1.2. Ketentuan Umum dan Persyaratan SNMPTN

A. Ketentuan Umum

1. SNMPTN merupakan pola seleksi nasional berdasarkan hasil penelusuran prestasi akademik dengan menggunakan rapor semester 1 (satu) sampai dengan semester 5 (lima) bagi SMA/SMK/MA atau sederajat dengan masa belajar 3

- (tiga) tahun atau semester 1 (satu) sampai dengan semester 7 (tujuh) bagi SMK dengan masa belajar 4 (empat) tahun, serta portofolio akademik.
2. Pangkalan Data Sekolah dan Siswa (PDSS) merupakan basis data yang berisikan rekam jejak kinerja sekolah dan prestasi akademik siswa.
 3. Sekolah yang siswanya mengikuti SNMPTN harus mempunyai Nomor Pokok Sekolah Nasional (NPSN) dan mengisikan data prestasi siswa di PDSS.
 4. Siswa yang berhak mengikuti seleksi adalah siswa yang memiliki Nomor Induk Siswa Nasional (NISN), memiliki prestasi unggul dan rekam jejak prestasi akademik di PDSS.
 5. Siswa yang akan mendaftar SNMPTN wajib membaca informasi pada laman PTN yang dipilih tentang ketentuan yang terkait dengan penerimaan mahasiswa baru.

B. Ketentuan Khusus

1. Persyaratan Sekolah

Sekolah yang siswanya berhak mengikuti SNMPTN adalah SMA/SMK/MA atau sederajat (termasuk SRI di luar negeri) yang mempunyai NPSN dan telah mengisi PDSS dengan lengkap dan benar.

2. Persyaratan Siswa Pendaftar

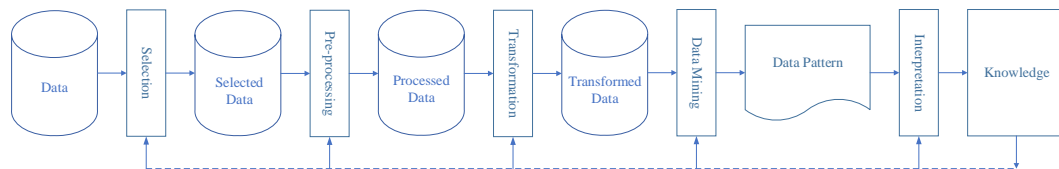
Siswa SMA/SMK/MA atau sederajat (termasuk SRI di luar negeri) kelas terakhir pada tahun 2017 yang memenuhi persyaratan sebagai berikut:

- a. Memiliki prestasi unggul yaitu calon peserta masuk peringkat terbaik di sekolah, dengan ketentuan berdasarkan akreditasi sekolah sebagai berikut:
 - 1) akreditasi A, 50% terbaik di sekolahnya;
 - 2) akreditasi B, 30% terbaik di sekolahnya;
 - 3) akreditasi C, 10% terbaik di sekolahnya;
 - 4) belum terakreditasi, 5% terbaik di sekolahnya.Pemeringkatan dilakukan oleh Panitia Pusat.
- b. Memiliki NISN dan terdaftar pada PDSS.
- c. Memiliki nilai rapor semester 1 sampai semester 5 (bagi siswa SMA/SMK/MA atau sederajat tiga tahun) atau nilai rapor semester 1 sampai semester 7 (bagi SMK empat tahun) yang telah diisikan pada PDSS.

- d. Memenuhi persyaratan lain yang ditentukan oleh masing masing PTN (dapat dilihat pada laman PTN bersangkutan).
3. Penerimaan di PTN
Peserta diterima di PTN, jika:
 - a. lulus satuan pendidikan;
 - b. lulus SNMPTN 2017; dan
 - c. lulus verifikasi data dan memenuhi persyaratan lain yang ditentukan oleh masing-masing PTN penerima.

2.2. Knowledge Discovery in Databases

Knowledge Discovery in Databases (KDD) adalah sekumpulan proses untuk mengidentifikasi pola data yang valid, baru, berpotensi berguna, dan dapat dipahami (Rikhi, 2015). Istilah *data mining* dan KDD sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar (Kusrini dan Luthfi, 2009). Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Gambar 2.1 menunjukkan tahap penemuan *knowledge* pada KDD.



Gambar 2.1 Tahap penemuan knowledge pada KDD (Saputra dkk, 2016)

Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

1. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing/Cleaning*

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki

kesalahan pada data. Selanjutnya dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation/Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya sehingga menghasilkan sebuah *knowledge* baru.

2.3. **Data mining**

Data mining merupakan suatu metode menemukan suatu pengetahuan dalam suatu database yang cukup besar (Efendi, 2014). *Data mining* sendiri 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. Jananto menyebutkan alasan utama mengapa *data mining* diperlukan adalah karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan *knowledge* yang berguna (Jananto, 2013). Informasi tersebut dapat digunakan pada banyak bidang, mulai dari bisnis, kontrol produksi, kesehatan, dan lain-lain. *Data mining* dibagi menjadi

beberapa metode yaitu deskripsi, estimasi, prediksi, klasifikasi, *clustering*, dan asosiasi (Gunadi dan Sensuse, 2012).

2.4. Metode Klasifikasi

Metode klasifikasi adalah metode yang mengklasifikasikan data yang tidak diketahui ke beberapa data yang dikenal, yang telah ditentukan kategori sesuai dengan konsep dari berbagai kategori, dan menggunakan deskripsi ini untuk membangun kategori model (biasanya ditandai dengan peraturan atau modus pohon pengambilan keputusan) (Maghfirah dkk, 2015). Kategori deskripsi dan model klasifikasi dapat diperoleh dengan memanfaatkan bank dataset melalui algoritma tertentu. Klasifikasi dapat juga disebut perkiraan atribut, yaitu, mengklasifikasikan data ke dalam kategori yang paling tepat. Algoritma utama dari metode klasifikasi adalah pohon keputusan, statistik Bayesian, *nearest neighbor*, jaringan saraf pengambilan keputusan, algoritma genetika, teori himpunan kasar, logika fuzzy dan sebagainya.

2.5. Algoritma Klasifikasi

Terdapat banyak algoritma dalam metode klasifikasi antara lain C4.5, K-*Nearest Neighbor*, Naïve Bayes, dan lain-lain. Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama dari algoritma yang lainnya (Kamagi dan Hansun, 2014). Kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima dan dapat menangani atribut bertipe diskret dan numerik. Dalam mengkonstruksi pohon, algoritma C4.5 membaca seluruh sampel data training dari storage dan memuatnya ke memori. Hal yang menjadi salah satu kelemahan algoritma C4.5 dalam kategori “skalabilitas” adalah algoritma ini hanya dapat digunakan jika *data training* dapat disimpan secara keseluruhan dan pada waktu yang bersamaan di memori.

Algoritma K-*Nearest Neighbor* adalah sebuah algoritma untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Nisa, 2017). Tujuan dari metode K-*Nearest Neighbor* adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan training

dimana hasil dari uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori. Kelebihan dari algoritma ini adalah dapat menghasilkan data yang lebih akurat dan lebih efektif pada *data training* yang besar. Akan tetapi kelemahan pada algoritma ini adalah memiliki nilai komputasi yang tinggi dan perlu menentukan parameter 'k' yang optimal untuk menyatakan jumlah tetangga yang terdekat.

Algoritma Naïve Bayes merupakan sebuah algoritma pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan (Saleh, 2015). Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Kelebihan dari algoritma ini adalah hanya membutuhkan *data training* yang kecil sehingga komputasi yang dilakukan juga relatif rendah. Kelebihan lain pada algoritma ini yaitu dapat dipersonalisasi sedemikian rupa sehingga dapat disesuaikan dengan sifat dan kebutuhan masing-masing data (Natalius, 2010). Namun terdapat kelemahan pada algoritma ini yaitu ukuran dari vektor fitur yang dihasilkan cukup besar dan butuh teknik untuk memperkecil ukuran vektor tersebut (Darujati dan Gumelar, 2012).

Kesimpulan yang dapat diambil adalah bahwa pada praktiknya, algoritma yang telah disebutkan tidak ada yang sempurna karena masing-masing memiliki kelebihan dan kekurangan masing-masing. Hal ini disebabkan dari karakteristik algoritma itu sendiri dalam pengklasifikasian yang ingin dicapai. Dari pertimbangan yang sudah dilakukan, untuk studi kasus pada penelitian ini yang mana menggunakan data penerimaan mahasiswa baru dengan sifat dan kebutuhan data yang berbeda, maka algoritma yang akan digunakan adalah algoritma Naïve Bayes karena mampu menyesuaikan kondisi data tersebut.

2.6. Algoritma Naïve Bayes

Algoritma Naïve Bayes atau dikenal dengan Bayesian *Classification* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Algoritma Naïve Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* (Kusrini dan Luthfi, 2009).

Teorema Bayes memiliki bentuk umum sebagai berikut:

$$P(H | X) = \frac{P(X | H) P(H)}{P(X)} \quad 2.1$$

Dimana:

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$ = Probabilitas hipotesis H (prior prob.)

$P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut

$P(X)$ = Probabilitas dari X

Ide dasar dari Algoritma Naïve Bayes adalah bahwa hasil hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan pada beberapa bukti (X) yang diamati. Ada beberapa hal penting dari aturan Bayes yaitu (Prasetyo, 2012):

1. Sebuah probabilitas awal / *prior* H atau $P(H)$ adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir / *posteriori* H atau $P(H|X)$ adalah probabilitas dari suatu hipotesis setelah bukti diamati.

Klasifikasi dengan Naïve Bayes bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik Naïve Bayes sebagai berikut (Prasetyo, 2012):

1. Metode Naïve Bayes bekerja teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outliner*). Naïve Bayes juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.
2. Tangguh menghadapi atribut yang tidak relevan.
3. Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi Naïve Bayes karena asumsi independensi atribut tersebut sudah tidak ada.

2.7. Pengukuran Kinerja Klasifikasi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya (Arfiana, 2014). Umumnya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Ridwan dkk menggunakan matriks konfusi pada penelitiannya dengan topik penerapan *data mining* untuk evaluasi kinerja akademik mahasiswa menggunakan algoritma Naïve Bayes, bertujuan untuk mengetahui unjuk kerja dari algoritma Naïve Bayes dalam mengklasifikasikan data ke dalam kelas yang telah ditentukan (Ridwan dkk, 2013). Efendi juga menggunakan matriks konfusi pada penelitiannya dengan topik implementasi data mining menggunakan algoritma Naïve Bayes dalam menentukan pengunduran diri calon mahasiswa pada Universitas Dian Nuswantoro Semarang, bertujuan untuk mengetahui akurasi algoritma Naïve Bayes dalam memprediksi registrasi calon mahasiswa baru pada Universitas Dian Nuswantoro Semarang (Efendi, 2014).

Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi (Prasetyo, 2012). Kuantitas matriks konfusi dapat diringkas menjadi dua nilai, yaitu akurasi dan laju *error*. Dengan mengetahui jumlah data yang diklasifikasikan secara benar, dapat diketahui akurasi hasil prediksi dan dengan mengetahui jumlah data yang diklasifikasikan secara salah, dapat diketahui laju *error* dari prediksi yang dilakukan. Dua kuantitas ini digunakan sebagai matrik kinerja klasifikasi. Formula 2.2 digunakan untuk menghitung akurasi, sedangkan formula 2.3 digunakan untuk menghitung laju *error* (kesalahan prediksi).

$$\text{akurasi} = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{Jumlah prediksi yang dilakukan}} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad 2.2$$

$$\text{laju error} = \frac{\text{Jumlah data yang diprediksi secara salah}}{\text{Jumlah prediksi yang dilakukan}} = \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad 2.3$$

Dimana:

f_{00} = Data negatif yang diprediksi negatif

f_{01} = Data negatif yang diprediksi positif

- f_{10} = Data positif yang diprediksi negatif
 f_{11} = Data positif yang diprediksi positif

Semua algoritma klasifikasi berusaha membentuk model yang mempunyai akurasi tinggi atau laju error yang rendah. Umumnya, model yang dibangun memprediksi dengan benar pada semua data yang menjadi data latihnya, tetapi ketika model berhadapan dengan data uji, barulah kinerja model dari sebuah algoritma klasifikasi ditentukan. Seperti pada penelitian yang dilakukan oleh Ridwan dkk, hasil akurasi dari algoritma Naïve Bayes yang didapat adalah sebesar 70%, sedangkan pada penelitian yang dilakukan oleh Efendi, hasil akurasi yang didapat dengan algoritma yang sama adalah sebesar 78%. Sehingga dapat ditarik kesimpulan bahwa masing-masing model memiliki kinerja yang berbeda, tergantung dengan bagaimana model tersebut dibentuk.

2.8. Penelitian Terdahulu

Belum terdapat banyak penelitian terkait penerapan KDD menggunakan algoritma Naïve Bayes pada data penerimaan mahasiswa baru melalui jalur SNMPTN. Penelitian yang dilakukan oleh Nugroho pada tahun 2014, membahas bagaimana mengklasifikasi kelulusan mahasiswa Universitas Dian Nuswantoro (UDINUS) angkatan 2009 dan mendapatkan akurasi yang tepat untuk melakukan klasifikasi tersebut dengan menggunakan Naïve Bayes (Nugroho, 2014). Hasil yang didapat adalah nilai akurasi terhadap klasifikasi kelulusan sebesar 82.08%. Hasil tersebut menunjukkan bahwa algoritma Naïve Bayes sukses dalam mengklasifikasikan status kelulusan.

Lalu pada tahun 2015, penelitian yang dilakukan oleh Setiawan dkk membahas bagaimana membangun sistem pendukung keputusan berbasis web dengan tujuan mempermudah panitia SNMPTN dalam menyeleksi calon mahasiswa baru dan mempercepat waktu seleksi (Setiawan dkk, 2015). Dari tujuan penelitian tersebut dapat diketahui bahwa waktu seleksi yang dibutuhkan menjadi masalah utama sehingga harus diefisienkan.

Selanjutnya, penelitian yang dilakukan oleh Hernawati dkk pada tahun yang sama adalah analisis pola hubungan dalam data indeks prestasi kumulatif dan data dasar SNMPTN dengan algoritma Apriori (Hernawati dkk, 2015). Atribut dari data

SNMPTN yang digunakan pada penelitian tersebut meliputi tingkat pendidikan orang tua, pekerjaan orang tua, penghasilan orang tua, asal SMA, dan asal kabupaten (SMA asal). Hal tersebut akan menjadi referensi dalam menentukan atribut yang digunakan pada penelitian ini.

Tabel 2.1 Penelitian terdahulu

No	Nama dan Tahun Publikasi	Hasil
1	Nugroho, 2014	Metode: Data Mining Menggunakan Algoritma Naïve Bayes untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro Hasil: Nilai akurasi terhadap klasifikasi kelulusan adalah sebesar 82.08 %.
2	Setiawan dkk, 2015	Metode: Implementasi Metode Electre pada Sistem Pendukung Keputusan SNMPTN Jalur Undangan Hasil: Aplikasi yang telah dibuat mengefisiensi proses seleksi calon mahasiswa pada SNMPTN jalur undangan di Perguruan Tinggi Negeri.
3	Hernawati dkk, 2015	Metode: Analisis Pola Hubungan dalam Data Indeks Prestasi Kumulatif dan Data Dasar SNMPTN dengan Algoritma Apriori Hasil: Data dasar SNMPTN tidak memengaruhi perolehan Indeks Prestasi Kumulatif (IPK)
4	Nasution dkk, 2015	Metode: Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes (Studi Kasus: Fasilkom Unilak) Hasil: Atribut yang memiliki pengaruh yang paling signifikan untuk meningkatkan nilai akurasi adalah Asal Sekolah, Indeks Prestasi (IP) Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, IP Semester 5 dan Indeks Prestasi Kumulatif (IPK)

Kemudian pada tahun yang sama juga, penelitian yang dilakukan oleh Nasution dkk adalah evaluasi kinerja akademik mahasiswa menggunakan algoritma Naïve Bayes di Universitas Lancang Kuning (Nasution dkk, 2015). Pada penelitian tersebut, Nasution dkk melakukan pemetaan antar atribut untuk mengetahui atribut mana yang memiliki peranan yang signifikan atau yang berpengaruh dalam menentukan mahasiswa lulus tepat waktu atau lulus tidak tepat waktu. Hal tersebut juga akan menjadi referensi dalam menentukan atribut yang berpengaruh terhadap status kelulusan calon mahasiswa baru pada jalur SNMPTN. Tabel 2.1 menunjukkan rangkuman hasil penelitian terdahulu yang memiliki keterkaitan dengan penelitian yang akan dilakukan.

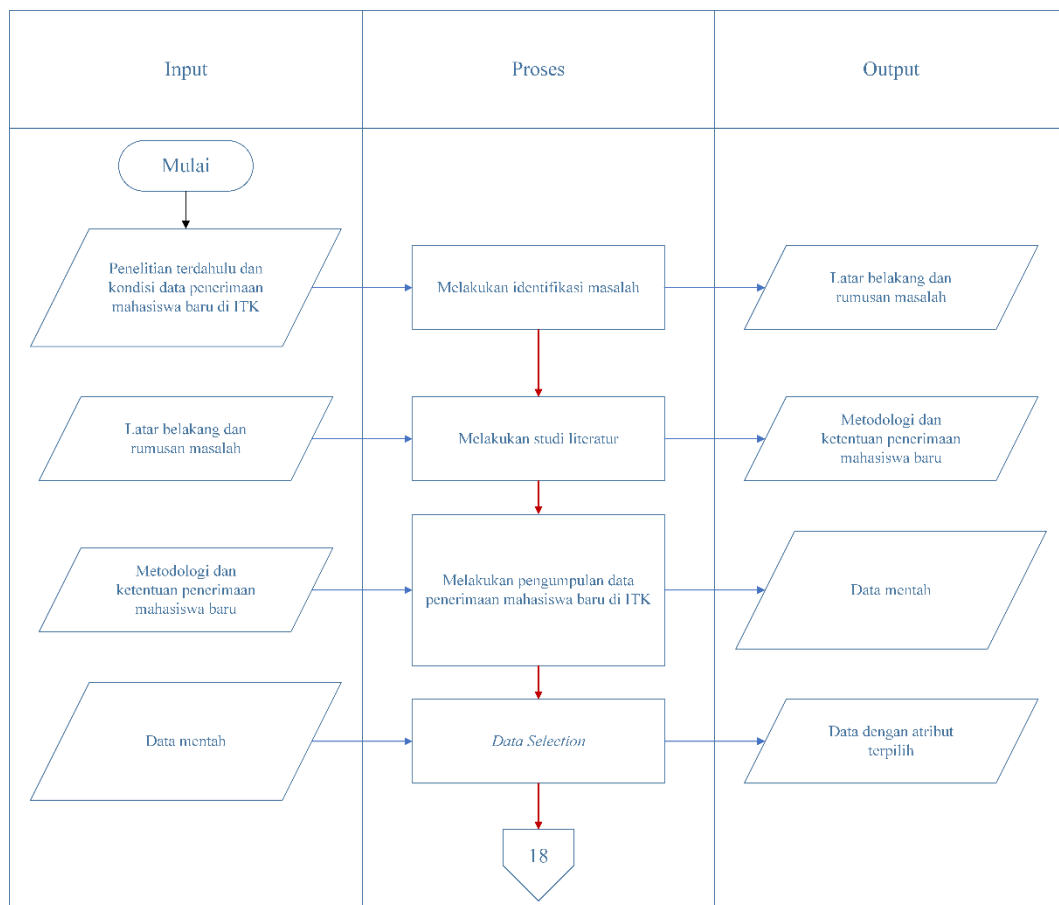
BAB 3

METODE PENELITIAN

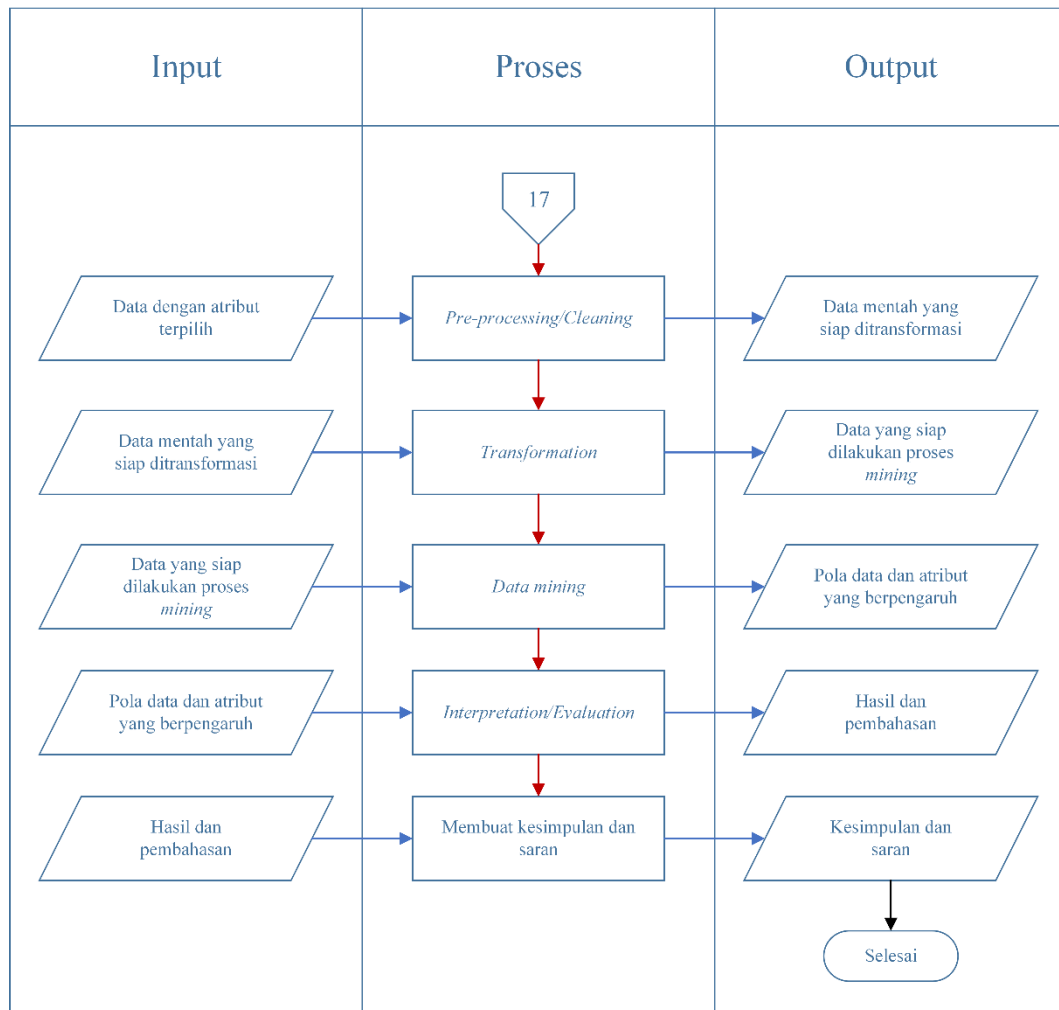
Pada bab ini dijelaskan mengenai metodologi yang digunakan dalam penelitian, termasuk tahapan yang dilakukan dalam analisis data penerimaan mahasiswa baru menggunakan teknik *Knowledge Discovery in Databases* (KDD).

3.1. Metodologi

Bagian ini membahas mengenai metodologi yang digunakan dalam melakukan penelitian berdasarkan bagan alur yang ditunjukkan pada Gambar 3.1 dan Gambar 3.2. Dari bagan tersebut dapat dilihat input, proses, dan *output* yang dibutuhkan dan dilakukan dalam penelitian.



Gambar 3.1 Bagan metodologi bagian 1



Gambar 3.2 Bagan metodologi bagian 2

Adapun penjelasan lebih dalam mengenai metodologi yang digunakan dalam penelitian adalah sebagai berikut:

1. Identifikasi Masalah

Tahap identifikasi masalah merupakan langkah awal dalam penelitian ini. Identifikasi permasalahan dilakukan untuk mendapatkan latar belakang dan rumusan permasalahan penelitian sehingga didapatkan penjelasan yang lebih rinci mengenai penelitian ini.

Identifikasi masalah pada penelitian ini menggunakan metode wawancara terhadap ketua pelaksana panitia penerimaan mahasiswa baru mengenai data penerimaan mahasiswa baru di ITK. Hasil dari identifikasi masalah tersebut adalah latar belakang dan perumusan masalah penelitian yaitu pada setiap tahunnya, jumlah mahasiswa baru yang diterima ITK semakin meningkat. Namun proses

seleksi dalam menentukan kelulusan calon mahasiswa baru masih dilakukan secara manual sehingga membutuhkan waktu yang cukup lama.

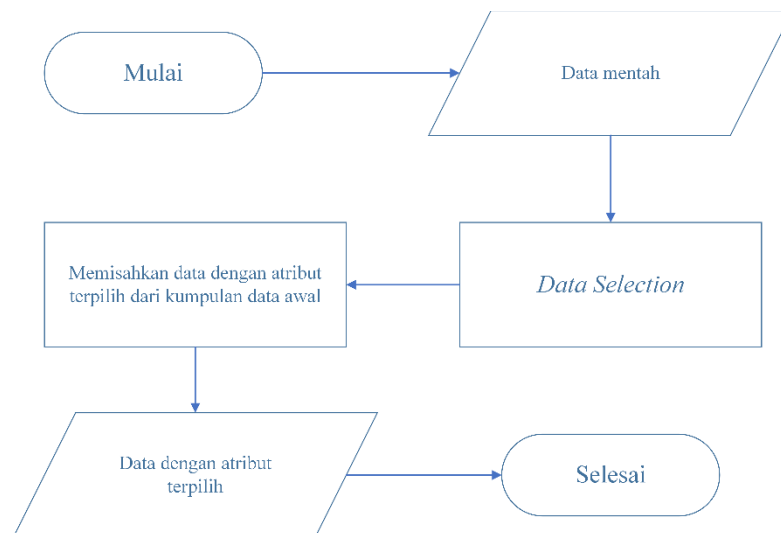
2. Studi Literatur

Studi literatur digunakan untuk mendukung, memperkuat data dan pernyataan, serta sebagai referensi sehingga integritas data dan informasi yang digunakan dapat terjaga. Studi literatur pada penelitian ini dilakukan dari berbagai sumber yaitu berupa penelitian-penelitian yang telah dilakukan sebelumnya, jurnal, skripsi, dan juga buku terkait penerapan KDD, serta dokumen yang memuat informasi terkait penerimaan mahasiswa baru. Tahap ini dilakukan dengan tujuan dapat memahami dasar teori yang berhubungan dengan penelitian dan mempermudah dalam menyelesaikan penelitian. Setelah melakukan studi literatur, ditemukan metodologi yang akan digunakan dan ketentuan penerimaan mahasiswa baru sebagai acuan dalam pengerjaan Tugas Akhir.

3. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data yang dibutuhkan untuk keperluan penelitian. Pengumpulan data dilakukan dengan mengajukan surat permohonan data kepada pihak akademik ITK. Luaran yang didapat pada tahap ini adalah data mentah berupa data penerimaan mahasiswa baru tahun 2016 dan 2017 di ITK yang selanjutnya akan diteliti dalam pengerjaan Tugas Akhir.

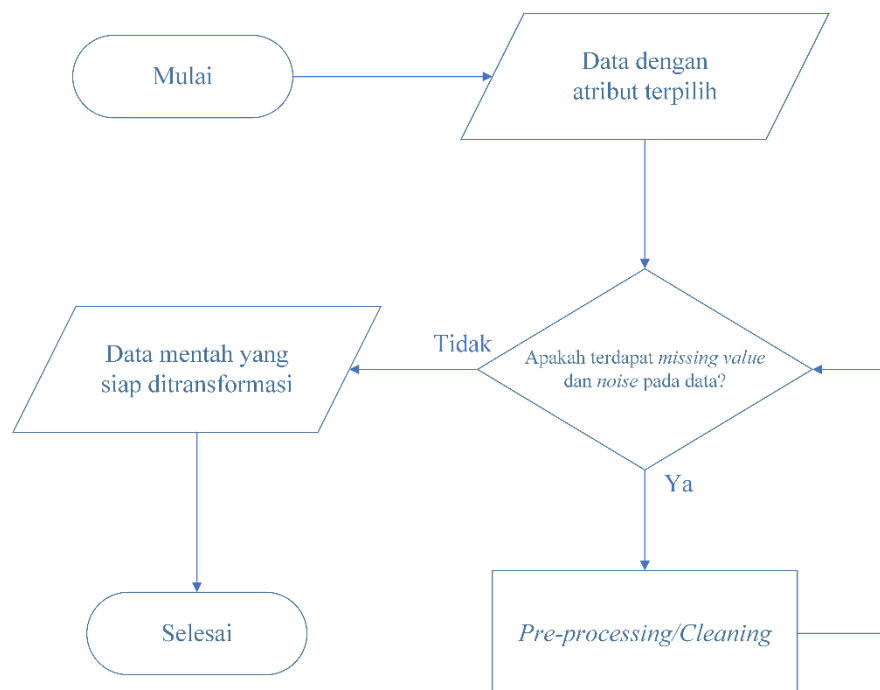
4. *Data Selection*



Gambar 3.3 Diagram alir tahap *data selection*

Data penerimaan mahasiswa baru memiliki banyak atribut di dalamnya. Sehingga pada tahap ini, dilakukan pemilihan atribut (*data selection*) terhadap data mentah yang akan digunakan pada proses *data mining*. *Data selection* dilakukan dengan cara menganalisa setiap atribut yang memiliki kemungkinan berpengaruh dalam pengambilan keputusan terhadap kelulusan calon mahasiswa baru. Selanjutnya data dengan atribut terpilih dipisahkan dari kumpulan data awal untuk digunakan pada tahap selanjutnya. Luaran yang didapat pada tahap ini adalah data penerimaan mahasiswa baru dengan atribut terpilih. Gambar 3.3 menunjukkan diagram alir pada tahap *data selection*.

5. *Pre-processing/Cleaning*



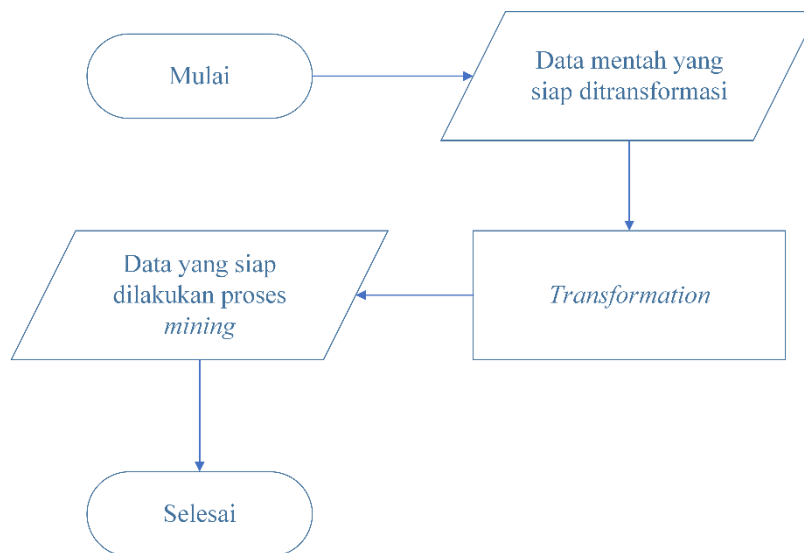
Gambar 3.4 Diagram alir tahap *pre-processing/cleaning*

Setelah melakukan tahap *data selection*, tahap yang akan dilakukan selanjutnya adalah tahap *pre-processing/cleaning*. Pada tahap ini, data dengan atribut terpilih akan diperiksa apakah masih terdapat data yang hilang (*missing value*) atau data yang tidak benar (*noise*) yang meliputi duplikasi data, data yang inkonsisten, dan kesalahan pada data. Jika terdapat *missing value* dan *noise*, maka data tersebut akan dilakukan *cleaning* dengan membuang duplikasi data, dan memperbaiki data yang salah maupun inkonsisten. Jika sudah tidak terdapat *missing value* dan *noise* pada kumpulan data, maka tahap ini menghasilkan luaran berupa data mentah yang siap

ditransformasi (data yang terbebas dari *missing value* dan *noise*). Gambar 3.4 menunjukkan diagram alir pada tahap *pre-processing/cleaning*.

6. Transformation

Pada tahap ini, data yang sudah terbebas dari *noise* dan *missing value* akan ditransformasikan (*transformation*) atau diubah ke dalam bentuk atau format yang sesuai untuk dilakukan tahap *data mining*. Pada studi kasus ini, kumpulan data yang atributnya bertipe numerik akan diubah menjadi kategorik sesuai dengan kebutuhan algoritma klasifikasi yang akan digunakan pada tahap *data mining*. Luaran dari tahap ini adalah kumpulan data yang siap dilakukan proses *mining*. Gambar 3.5 menunjukkan diagram alir pada tahap *transformation*.



Gambar 3.5 Diagram alir tahap *transformation*

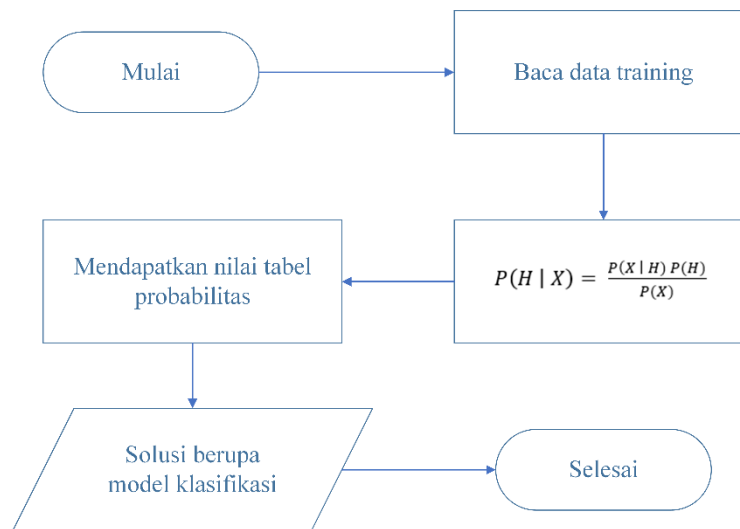
7. Data Mining

Pada tahap ini, dilakukan proses *data mining* terhadap data yang dihasilkan dari tahap sebelumnya yaitu tahap *transformation*. Metode yang digunakan pada tahap ini adalah metode klasifikasi dengan algoritma Naïve Bayes. Tujuan dari tahap ini adalah untuk menemukan pola atau informasi menarik dalam kumpulan data yang dilakukan proses *mining*.

Kumpulan data akan dibagi menjadi dua set data yaitu data *training* (x,y) dan data *testing* (x,y), dimana 'x' merupakan kumpulan atribut prediktor dan 'y' merupakan atribut respons. Data *training* digunakan untuk membangun model klasifikasi yang akan dipakai untuk memprediksi data *testing* (Arfiana, 2014).

Dalam membangun model klasifikasi diperlukan algoritma klasifikasi yaitu algoritma Naïve Bayes. Gambar 3.6 menunjukkan diagram alir algoritma Naïve Bayes. Adapun langkah-langkah pada Gambar 3.6 dapat diuraikan sebagai berikut (Ridwan dkk, 2013):

- a. Membaca data *training*.
- b. Mencari nilai probabilitas dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
- c. Mendapatkan nilai tabel probabilitas.
- d. Menghasilkan solusi berupa model klasifikasi.



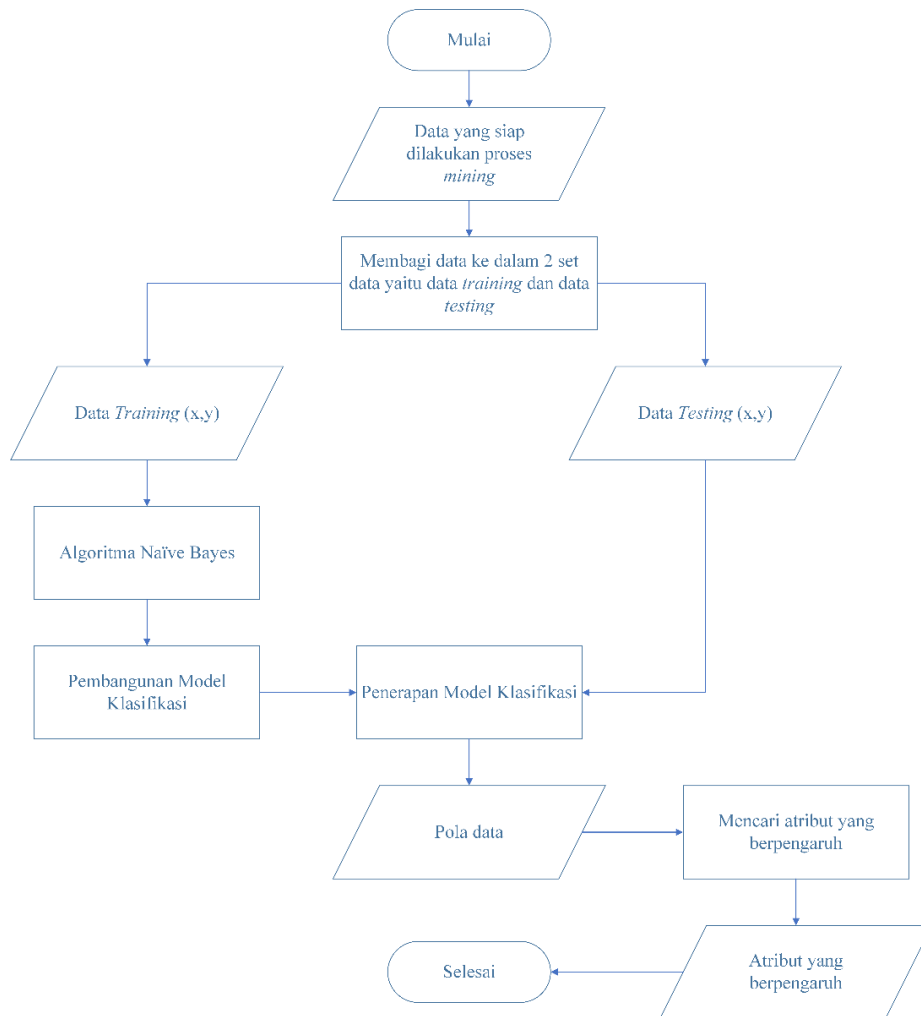
Gambar 3.6 Diagram alir algoritma Naïve Bayes

Setelah model klasifikasi dibangun, model tersebut digunakan untuk memprediksi data *testing* (x,y) sehingga dapat diketahui pola dari data tersebut. Selanjutnya akan dianalisa atribut mana yang memiliki pengaruh signifikan terhadap kelulusan calon mahasiswa jalur SNMPTN dengan membandingkan keterkaitan antar atribut. Luaran dari tahap ini adalah pola dari data yang diteliti serta atribut yang berpengaruh. Gambar 3.7 menunjukkan diagram alir pada tahap *data mining*.

8. Interpretation/Evaluation

Pada tahap ini dilakukan evaluasi terhadap pola yang dihasilkan dari tahap sebelumnya. Evaluasi dilakukan dengan cara memetakan kesesuaian hasil prediksi dengan data yang sebenarnya. Selanjutnya dihitung nilai akurasi untuk mengetahui jumlah data yang diklasifikasikan dengan benar, dan dihitung nilai laju *error* untuk

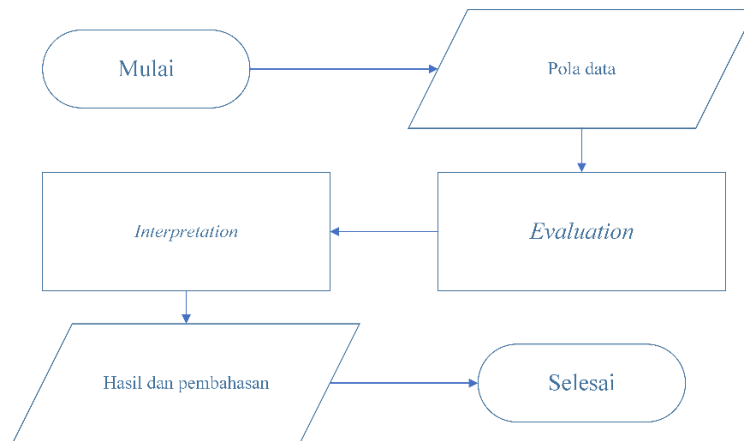
mengetahui jumlah data yang diklasifikasikan secara salah. Kemudian, pola yang dihasilkan juga akan ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Sehingga luaran yang dihasilkan pada tahap ini berupa hasil dan pembahasan terkait seluruh proses KDD yang dilakukan. Gambar 3.8 menunjukkan diagram alir pada tahap *interpretation/evaluation*.



Gambar 3.7 Diagram alir tahap *data mining*

9. Kesimpulan dan Saran

Pada tahap akhir ini setelah melakukan keseluruhan metodologi diberikan kesimpulan mengenai hasil penelitian berdasarkan analisa mengenai data penerimaan mahasiswa baru di ITK, serta rekomendasi yang dapat menjadi pembelajaran untuk penelitian selanjutnya.



Gambar 3.8 Diagram alir tahap interpretation/evaluation

3.2. Rencana Jadwal Penelitian

Tabel 3.1 menunjukkan rencana jadwal penelitian yang dilakukan dalam pengerjaan Tugas Akhir.

Tabel 3.1 Rencana jadwal penelitian

Kegiatan	Minggu Ke-															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Penyusunan Proposal Tugas Akhir																
Identifikasi Masalah																
Studi Literatur																
Pengumpulan Data																
Seminar Proposal Tugas Akhir																
<i>Data Selection</i>																
<i>Pre-processing / Cleaning</i>																
<i>Transformation</i>																
<i>Data Mining</i>																
<i>Interpretation / Evaluation</i>																
Penyusunan Laporan Tugas Akhir																
Sidang Tugas Akhir																

DAFTAR PUSTAKA

- Arfiana, F. (2014), *Klasifikasi Kendaraan Roda Empat Menggunakan Metode Naive Bayes*, Skripsi, Universitas Widyatama, Bandung.
- Darujati, C., dan Gumelar, A. B. (2012), "Pemanfaatan Teknik Supervised untuk Klasifikasi Teks Bahasa Indonesia", *Jurnal LINK*, Vol. 16, No. 1, hal. 52-58.
- Defiyanti, S., dan Jajuli, M. (2015), "Integrasi Metode Klasifikasi dan Clustering dalam Data Mining", *Konferensi Nasional Informatika*, hal. 39-44.
- Efendi, M. (2014), *Implementasi Data Mining Menggunakan Algoritma Naive Bayes dalam Menentukan Pengunduran Diri Calon Mahasiswa pada Universitas Dian Nuswantoro Semarang*, Skripsi, Universitas Dian Nuswantoro, Semarang.
- Gunadi, G., dan Sensuse, D. I. (2012), "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku dengan Menggunakan Algoritma Apriori dan Frequent Pattern Growth (FP-Growth) : Studi Kasus Percetakan PT. Gramedia", *Jurnal TELEMATIKA MKOM*, Vol. 4, No. 1, hal. 118-132.
- Hernawati, K., Insani, N., Sumarno, B., dan Waryanto, N. (2015), "Analisis Pola Hubungan Dalam Data Indeks Prestasi Kumulatif dan Data Dasar SNMPTN dengan Algoritma Apriori", *Jurnal Sains Dasar 2015*, Vol. 4, No. 2, hal. 153-163.
- Jananto, A. (2013), "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa", *Jurnal Teknologi Informasi DINAMIK*, Vol. 18, No. 1, hal. 9-16.
- Kamagi, D. H., dan Hansun, S. (2014), "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa", *ULTIMATICS*, Vol. 6, No. 1, hal. 15-20.
- Karsidi, R. (2017), *Informasi SNMPTN 2017*, Kementerian Riset, Teknologi dan Pendidikan Tinggi, Jakarta.
- Kusrini dan Luthfi, E. T. (2009), *Algoritma Data Mining*, Penerbit ANDI, Yogyakarta.

- Maghfirah, Adji, T. B., dan Setiawan, N. A. (2015), "Menggunakan Data Mining untuk Segmentasi Customer pada Bank untuk Meningkatkan Customer Relationship Management (CRM) dengan Metode Klasifikasi (Algoritma J-48, Zero-R Dan Naive Bayes)", *Prosiding SNST Fakultas Teknik UNHAWAS ke-6 Tahun 2015*, Semarang, hal. 65-70.
- Nasir, M. (2017), *Statuta Institut Teknologi Kalimantan*, Menteri Riset, Teknologi, dan Pendidikan Tinggi, Jakarta.
- Nasution, N., Djahara, K., dan Zamsuri, A. (2015), "Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes (Studi Kasus: Fasilkom Unilak)", *Jurnal Teknologi Informasi & Komunikasi Digital Zone*, Vol. 6, No. 2, hal. 1-11.
- Natalius, S. (2010), *Metoda Naïve Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen*, Institut Teknologi Bandung, Bandung.
- Nisa, S. A. (2017), *Rekomendasi Pemilihan Tipe Mobil Menggunakan Metode K-Nearest Neighbor (KNN)*, Skripsi, Universitas Nusantara PGRI Kediri, Kediri.
- Nugroho, Y. S. (2014), *Data Mining Menggunakan Algoritma Naïve Bayes untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro. (Studi Kasus: Fakultas Ilmu Komputer Angkatan 2009)*, Skripsi, Universitas Dian Nuswantoro, Semarang.
- Prasetyo, E. (2012), *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*, Penerbit ANDI, Yogyakarta.
- Ridwan, M., Suyono, H., dan Sarosa, M. (2013), "Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier", *Jurnal EECCIS*, Vol. 7, No. 1, hal. 59-64.
- Rikhi, N. (2015), "Data Mining and Knowledge Discovery in Database", *International Journal of Engineering Trends and Technology (IJETT)*, Vol. 23, No. 2, hal. 64-70.
- Saleh, A. (2015), "Implementasi Metode Klasifikasi Naïve Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga", *Citec Journal*, Vol. 2, No. 3, hal. 207-217.

- Saputra, N. E., Tania, K. D., dan Heroza, R. I. (2016), "Penerapan Knowledge Management System (KMS) Menggunakan Teknik Knowledge Data Discovery (KDD) pada PT PLN (Persero) WS2JB Rayon Kayu Agung", *Jurnal Sistem Informasi (JSI)*, Vol. 8, No. 2, hal. 1038-1055.
- Setiawan, F., Indriani, F., dan Muliadi (2015), "Implementasi Metode Electre pada Sistem Pendukung Keputusan SNMPTN Jalur Undangan", *Kumpulan jurnaL Ilmu Komputer (KLIK)*, Vol. 2, No. 2, hal. 88-101.
- Tambun, H. R. M. dan Anofrizen (2015), "Rancang Bangun Aplikasi Data Mining untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa dengan Algoritma Apriori", *Jurnal Rekayasa dan Manajemen Sistem Informasi*, Vol. 1, No. 1, hal. 31-38.

LAMPIRAN A

INTERVIEW PROTOCOL

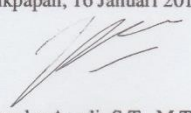
Nama : Jatmoko Awali, S.T., M.T.
 Jabatan : Dosen Prodi Material dan Metalurgi
 Tanggal : 16 Januari 2018
 Waktu : 14.20 – 14.55 WITA
 Tempat : Ruang Dosen Prodi Material dan Metalurgi – Gedung B R304 ITK
 Interviewer : Gardian Bimo Lukito

No.	Pertanyaan	Jawaban
1.	Apa posisi bapak dalam panitia penerimaan mahasiswa baru di Institut Teknologi Kalimantan (ITK)?	Sebagai ketua pelaksana / ketua koordinator.
2.	Bagaimana struktur kepanitiaan pada penerimaan mahasiswa baru jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) di ITK?	Untuk SNMPTN, panitia dibagi menjadi 4 sie yaitu: 1. Sie pengolah data. 2. Sie pelaksana seleksi akademik. 3. Sie penyusun kriteria seleksi akademik. 4. Sie sistem informasi proses seleksi.
3.	Apa saja tanggung jawab dan tugas pokok pada masing-masing elemen panitia?	Sie pengolah data bertanggung jawab mengolah data-data dari peserta SNMPTN dengan atribut antara lain nilai rapor semester 1 sampai 5, prestasi sekolah, kelinearan antara jurusan di sekolah dan program studi yang dituju, dan lain-lain; Sie pelaksana dan penyusun kriteria seleksi akademik bertanggung jawab dalam proses penentuan penerimaan peserta sebagai calon mahasiswa; Sie sistem informasi proses seleksi bertanggung jawab dalam membuat sistem terkait SNMPTN meliputi pengumuman nama yang diterima dan pengumuman daftar ulang, serta keperluan Uang Kuliah Tunggal (UKT) mahasiswa.

No.	Pertanyaan	Jawaban
4.	Sejak kapan kampus ITK terdaftar pada penerimaan mahasiswa jalur SNMPTN?	Sejak tahun 2016.
5.	Bagaimana alur proses penerimaan mahasiswa baru jalur SNMPTN?	Adapun garis besar alur proses penerimaan mahasiswa baru jalur SNMPTN adalah: 1. Pengisian dan verifikasi Pangkalan Data Sekolah dan Siswa (PDSS). 2. Proses Seleksi SNMPTN. 3. Pengumuman SNMPTN. 4. Pengumuman prasyarat verifikasi berkas SNMPTN. 5. <i>Upload</i> data UKT mahasiswa. 6. Pengumuman UKT. 7. Verifikasi / daftar ulang SNMPTN.
6.	Apakah kriteria penentu penerimaan mahasiswa ditentukan dari pusat atau dari ITK?	Dari ITK.
7.	Bagaimana menentukan kriteria tersebut?	Seperti yang telah dijawab pada poin 3, sie pengolah data menganalisis serta mengolah data tersebut yang kemudian akan ditentukan kriteria sebagai penentu kelulusan pendaftar. Proses ini masih dilakukan secara manual sehingga membutuhkan waktu yang cukup lama.
8.	Siapa yang bertanggung jawab terhadap data penerimaan mahasiswa baru setelah masa penerimaan mahasiswa baru telah selesai?	Akademik ITK.
9.	Apa tanggung jawab dari orang tersebut (jawaban dari poin 7)?	Adapun tanggung jawab dari bagian akademik antara lain: 1. Membuat <i>database</i> . 2. Sebagai informasi akademik. 3. Menyimpan dan menjaga data penerimaan mahasiswa baru dengan baik.

No.	Pertanyaan	Jawaban
10.	Apakah terdapat penggunaan lain terkait data penerimaan mahasiswa baru pada jalur SNMPTN? Jika ya, seperti apa penggunaan yang dimaksud?	Penggunaan dari data itu untuk keperluan beasiswa dan keperluan akreditasi institut.
11.	Apakah terdapat kritik dan saran yang ingin bapak berikan terkait data SNMPTN di ITK?	Saran untuk masa <i>upload</i> data UKT perlu diperpanjang.

Balikpapan, 16 Januari 2018


(Jatmoko Awali, S.T., M.T.)