

LAPORAN PENELITIAN



ALGORITMA APRIORI UNTUK MENGUKUR KORELASI JURUSAN SEKOLAH DAN NILAI MATA KULIAH KONSENTRASI TERHADAP TINGKAT KELULUSAN MAHASISWA

Oleh :

Mariana Windarti, M.T (NIDN 0606058702)

Hendro Joko Prasetyo, M.Kom (NIDN 0618077101)

Rizka Safitri Lutfiyani, M.Eng (NIDN 0627048802)

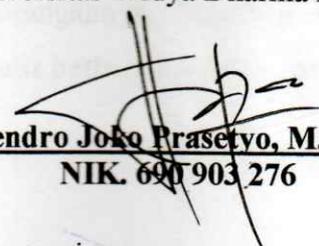
Penelitian ini dilaksanakan atas biaya
Lembaga Penelitian dan Pengabdian Kepada Masyarakat
Unwidha Klaten dengan Nomor Kontrak :
046/F.02.46/PUSLIT/III/2021

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS WIDYA DHARMA KLATEN
2021**

HALAMAN PENGESAHAN

1. Judul
ALGORITMA APRIORI UNTUK MENGUKUR KORELASI JURUSAN SEKOLAH DAN NILAI MATA KULIAH KONSENTRASI TERHADAP TINGKAT KELULUSAN MAHASISWA
2. Ketua Peneliti :
 - a) Nama : Mariana Windarti M.T
 - b) Jenis Kelamin : Perempuan
 - c) NIK/NIDN : 690 116 375 / 0606058702
 - d) Pangkat/Golongan : Penata Muda/ IIIb
 - e) Fakultas / Prodi : Ilmu Komputer / Manajemen Informatika
 - f) No Hp / Email : 081328551511 / marianawindarti@gmail.com
3. Jangka Waktu Penelitian : 10 bulan
4. Jumlah Anggota Peneliti : 2
 - a) Nama Lengkap : Hendro Joko Prasetyo, M.Kom
NIK/NIDN : 690 903 276 / 0618077101
 - b) Nama Lengkap : Rizka Safitri Lutfiyani, M.Eng
NIK/NIDN : 690 116 362 / 0627048802
5. Pembiayaan : Universitas Widya Dharma Klaten
6. Total Biaya : Rp. 3.000.000,00

Menyetujui,
Kaprosdi Manajemen Informatika
Universitas Widya Dharma Klaten

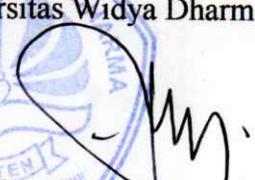

(Hendro Joko Prasetyo, M.Kom)
NIK. 690 903 276

Klaten, 16 Desember 2021
Ketua Pelaksana,


(Mariana Windarti, M.T)
NIK. 690 116 375

Mengetahui,
Dekan Fasilkom
Universitas Widya Dharma Klaten,

(Dr. Th. Kriswianti N, M.Si)
NIP. 19590929 198803 2 005

Mengesahkan,
Ketua LPPM
Universitas Widya Dharma Klaten,

(Arif Julianto S.N, S.E., M.Si)
NIK. 690 301 250

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa atas berkat dan rahmatNya penyusunan penelitian yang berjudul **“ALGORITMA APRIORI UNTUK MENGUKUR KORELASI JURUSAN SEKOLAH DAN NILAI MATA KULIAH KONSENTRASI TERHADAP TINGKAT KELULUSAN MAHASISWA”** telah selesai dilakukan. Pada kesempatan yang baik ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Prof. Dr. Triyono, M.Pd selaku Rektor Universitas Widya Dharma Klaten
2. Arif Julianto S N, S.E., M.Si selaku ketua lembaga penelitian Universitas Widya Dharma Klaten
3. Dr. Th. Kriswianti N, M.Si selaku dekan Fakultas Ilmu Komputer Universitas Widya Klaten
4. Hendro Joko Prasetyo, M.Kom selaku Kaprodi Manajemen Informatika Universitas Widya Dharma Klaten
5. Semua pihak yang tidak bisa penulis sebutkan satu persatu, atas kerjasama dalam pelaksanaan penelitian ini.

Penulis menyadari bahwa penelitian ini masih jauh dari sempurna dan banyak kekurangannya. Masukan dari semua pihak sangat penulis harapkan. Akhirnya penulis berharap semoga karya sederhana ini dapat bermanfaat bagi kita semua.

Klaten, 17 Desember 2021

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN PENGESAHAN	Error! Bookmark not defined.
KATA PENGANTAR	iii
DAFTAR ISI	iv
DAFTAR TABEL	vi
DAFTAR GAMBAR	vii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.5.1 Manfaat Teoritis	4
1.5.2 Manfaat Praktis	4
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI	5
2.1 Tinjauan Pustaka.....	5
2.2 Landasan Teori	8
2.2.1 Data Mining.....	8
2.2.2. RapidMiner	8
2.2.3 Assosiation	9
2.2.4 Algoritma Apriori.....	9
BAB III METODOLOGI PENELITIAN	11
3.1 Alat dan Bahan Penelitian	11
3.1.1 Alat Penelitian.....	11

3.1.2 Bahan Penelitian	11
3.2 Metodologi Penelitian	11
3.2.1 Proses Pengumpulan Data	11
3.2.2 Pengolahan Awal Data	12
3.2.3 Tahapan Penelitian	13
BAB IV HASIL DAN PEMBAHASAN.....	15
4.1. Pengolahan Data	15
4.2 Hasil Implementasi <i>Tool</i> RapidMiner	17
BAB V PENUTUP.....	19
5.1 Kesimpulan	19
5.2 Saran.....	19

DAFTAR TABEL

Tabel 4.1. Transformasi Nilai Mata Kuliah	15
Tabel 4.2. Transformasi Jurusan Sekolah	15
Tabel 4.3. Transformasi Masa Studi	16
Tabel 4.3. Sampel Dataset Konsentrasi Sistem Informasi Sistem Informasi.....	16
Tabel 4.3. Sampel Dataset Konsentrasi Sistem Informasi Sistem Cerdas.....	17

DAFTAR GAMBAR

Gambar 3.1. Tahapan Penelitian	10
Gambar 4.1. Tampilan Implementasi Apriori pada RapidMiner.....	17
Gambar 4.2. Tampilan Table View Association Rules.....	18

RINGKASAN

Jumlah data mahasiswa yang bertambah setiap tahun tentu mengakibatkan penumpukan data di perguruan tinggi, khususnya di Universitas Widya Dharma Klaten Program Studi Teknik Informatika. Dibutuhkan suatu teknik pengolahan data agar data yang menumpuk tidak sulit untuk dianalisa. Penelitian ini dilakukan untuk menganalisis hubungan antara data jurusan sekolah dan nilai mata kuliah konsentrasi mahasiswa dengan tingkat kelulusan mahasiswa. Program Studi (PS) Teknik Informatika memiliki 2 (dua) konsentrasi atau penjurusan yaitu Sistem Informasi (SI) dan Sistem Cerdas (SC). Teknik pengolahan data saat ini sudah bervariasi dan perlu disesuaikan dengan kebutuhan analisis data. Metode yang digunakan dalam penelitian ini adalah algoritma Apriori, yaitu algoritma Asosiasi yang menggunakan pengetahuan frekuensi atribut yang telah diketahui sebelumnya untuk memproses informasi selanjutnya. Penelitian ini dilakukan dengan menggunakan data jurusan sekolah sebelum masuk perguruan tinggi dan data akademik berupa nilai mata kuliah konsentrasi dan data kelulusan mahasiswa. Data tingkat kelulusan yang digunakan adalah lama studi yang diperoleh melalui tanggal kelulusan dan IPK (Indeks Prestasi Kumulatif). Setelah semua data terkumpul kemudian mencari persentase hubungan atau korelasi antara jurusan sekolah, nilai mata kuliah konsentrasi mahasiswa terhadap tingkat kelulusan menggunakan teknik data mining. Informasi yang ditampilkan berupa nilai support (Nilai Penunjang) dan confidence (Nilai Kepastian).

Kata kunci : *data mining*, asosiasi, korelasi, Support dan Confidence

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan tinggi di Indonesia saat ini sedang berkembang. Hal ini ditunjukkan dengan semakin banyak muncul perguruan-perguruan tinggi baik negeri maupun swasta. Jumlah pendaftar untuk masuk ke sebuah perguruan tinggi juga semakin meningkat setiap tahunnya, sehingga menyebabkan adanya persaingan diantara perguruan tinggi di Indonesia. Untuk dapat bersaing dengan perguruan tinggi lain, tentunya sebuah perguruan tinggi harus memiliki kualitas dan mutu pendidikan yang bagus. Berbagai upaya telah dilakukan oleh sebuah universitas guna meningkatkan mutu dan kualitas.

Kualitas perguruan tinggi ditunjukkan dengan prestasi akademik yang dicapai oleh para mahasiswa. Prestasi akademik pada dunia pendidikan merupakan salah satu tolak ukur dari keberhasilan proses belajar mengajar. Pada pendidikan tinggi prestasi akademik dapat dilihat dari nilai Indeks Prestasi Kumulatif (IPK) dan lamanya studi atau masa studi dalam menempuh pendidikan di perguruan tinggi. Menurut Wikipedia Indeks Prestasi adalah salah satu alat ukur prestasi di bidang akademik. Meskipun bernama "indeks", IP sebenarnya bukanlah indeks dalam pengertian sebenarnya, melainkan merupakan rerata terboboti. Sedangkan Indeks Prestasi Kumulatif merupakan perhitungan IP dengan menggabungkan semua matakuliah yang telah ditempuh sampai suatu semester tertentu . Masa studi merupakan lama studi terjadwal yang harus ditempuh oleh mahasiswa sesuai dengan rentang waktu yang dipersyaratkan. Semakin cepat masa studi mahasiswa maka kinerja mahasiswa semakin baik, begitu juga sebaliknya.

Sebelum memasuki perguruan tinggi, seorang mahasiswa pasti telah menempuh pendidikan Sekolah Menengah (SMA/SMK) atau sederajat. Jurusan sekolah saat SMA/SMK juga berbeda-beda seperti SMA memiliki jurusan IPA, IPS dan Bahasa. Tingkat SMK memiliki jurusan Teknik Komputer & Jaringan, Akuntansi, Manajemen Bisnis, Multimedia, dll. Beberapa penelitian telah dilakukan untuk menganalisa tentang kinerja mahasiswa yaitu variabel yang

mempengaruhi kinerja mahasiswa baik IPK maupun lama studi, faktor lain yang mempengaruhi kinerja seperti faktor internal dan eksternal. Selain itu juga terdapat penelitian untuk menganalisa korelasi antara atribut yang berkaitan dengan kinerja mahasiswa seperti nilai UN yang mempengaruhi IPK yang diperoleh selama menempuh studi atau IPK yang mempengaruhi masa studi, dan lain-lain. Atribut untuk memprediksi kelulusan mahasiswa dapat dilakukan berdasarkan usia, jenis kelamin, Indeks Prestasi semester 1 sampai semester 4 (Romadhona, Suprapedi, & Himawan, 2017)

Untuk dapat melakukan proses analisa dengan prediksi ataupun mengukur tingkat korelasi menggunakan sejumlah data mahasiswa yang tersimpan dalam basis data. Perolehan pengetahuan dalam basis data (sejumlah data yang besar) biasa disebut dengan *data mining*. Istilah *data mining* atau penambangan data merupakan proses untuk mendapatkan informasi yang berguna dari basis data yang besar (Jiawei & Kamber, 2001). Data mining adalah metode penting untuk meningkatkan efisiensi dalam menemukan informasi baru atau tersembunyi yang berguna, valid dan mudah dipahami dari basis data yang sangat besar. Data Mining adalah proses analisis data dari berbagai sudut pandang dan meringkas data tersebut ke dalam format informasi identik yang bermanfaat dan dapat digunakan untuk memprediksi tren di masa depan.

Setiap perguruan tinggi dituntut untuk meningkatkan kualitasnya, hal ini tentu saja harus disertai dengan sumber daya manusia yang berkualitas, serta sarana dan prasarana yang menunjang mahasiswa dapat lulus tepat waktu. Data yang ada di pendidikan tinggi dapat digunakan oleh pihak perguruan tinggi untuk dapat meningkatkan mutu dan kualitas dengan melakukan upaya atau sejumlah strategi tertentu. Di Universitas Widya Dharma Klaten (Unwidha) khususnya pada PS Teknik Informatika, tingkat kelulusan mahasiswa terbilang masih rendah dengan nilai IPK cukup bahkan terbilang kecil. Masa studi yang ditempuh mahasiswa juga terbilang lama, sebagian mahasiswa menyelesaikan masa studi kurang dari 5 tahun, sebagian lagi lebih dari 5 tahun, dimana lama studi yang dijadwalkan adalah 4 tahun.

Oleh karena itu dengan memanfaatkan data akademik dan data kelulusan mahasiswa, penulis mengimplementasikan data mining untuk membuat suatu

aplikasi, yaitu teknik asosiasi dengan algoritma apriori. Diharapkan dengan adanya aplikasi ini program studi Teknik Informatika baik sekretariat, dosen maupun mahasiswa dapat mengambil solusi atau kebijakan yang lebih baik dalam proses evaluasi pembelajaran, sehingga dapat meningkatkan tingkat kelulusan mahasiswa dan kualitas pada program studi Teknik Informatika .

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang diatas dapat dirumuskan masalah sebagai berikut:

1. Bagaimana menerapkan algoritma apriori untuk mengukur korelasi jurusan sekolah dan nilai mata kuliah konsentrasi dengan tingkat kelulusan mahasiswa.
2. Bagaimana melalui hasil korelasi dan analisa dapat digunakan pihak terkait khususnya PS Teknik Informatika sebagai bahan pertimbangan dan kebijakan dalam proses evaluasi pembelajaran guna meningkatkan tingkat kelulusan mahasiswa.

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini agar lebih mudah dipahami dalam penulisannya adalah sebagai berikut:

1. Studi Kasus dilakukan di Fakultas Ilmu Komputer UNWIDHA Klaten yang khususnya program studi Teknik Informatika.
2. Data yang digunakan adalah data alumni S1 Teknik Informatika dengan tahun lulus 2018-2020.
3. Atribut yang digunakan antara lain jurusan sekolah, tahun masuk, nilai mata kuliah konsentrasi, tanggal kelulusan dan IPK.
4. Untuk mengukur tingkat korelasi menggunakan metode apriori yang menggunakan perangkat lunak atau *tools* Rapid Miner.

1.4 Tujuan Penelitian

Tujuan penelitian yang ingin dicapai oleh peneliti antara lain:

1. Bagaimana menerapkan algoritma apriori untuk mengukur korelasi jurusan sekolah dan nilai mata kuliah konsentrasi dengan tingkat kelulusan mahasiswa.
2. Bagaimana hasil korelasi dan analisa dapat digunakan pihak terkait khususnya program studi Teknik Informatika sebagai bahan pertimbangan dan kebijakan dalam proses evaluasi pembelajaran guna meningkatkan kualitas lulusan mahasiswa.

1.5 Manfaat Penelitian

Dengan adanya penelitian ini diharapkan mampu memberikan manfaat sebagai berikut:

1.5.1 Manfaat Teoritis

Hasil penelitian ini diharapkan mampu memberikan sumbangan pemikiran dan memperkaya wawasan dalam melakukan analisis dan mengukur tingkat korelasi menggunakan metode apriori.

1.5.2 Manfaat Praktis

a. Bagi Peneliti

Sebagai sarana belajar untuk dapat mengintegrasikan pengetahuan dan hasil analisis yang diperoleh. Selain itu penelitian ini juga mampu memberikan gambaran bagaimana proses untuk mengukur korelasi jurusan sekolah, data akademik berupa nilai mata kuliah konsentrasi dan tingkat kelulusan (lama studi dan IPK).

b. Bagi Pihak Universitas

Hasil pengukuran tingkat korelasi mampu memberikan gambaran atau pertimbangan bagi pihak program studi untuk melakukan antisipasi terhadap mahasiswa yang masa studinya melebihi dari target yang diharapkan.

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Penelitian yang dilakukan dibidang data mining telah banyak dilakukan, antara lain penerapan data mining untuk melakukan prediksi berdasarkan keadaan tertentu atau atribut yang dimiliki. Selain itu juga dapat mengukur tingkat korelasi antara variabel yang mempengaruhi variabel lain. Terdapat teknik atau metode yang digunakan untuk menerapkan proses didalam data mining seperti *Decicion Tree* (DT), K-NN, *Bayesian Network* (BN), *Neural Network*, *Asosiation Rules* (Prasetyo, 2012).

Asosiasi atau Association Rule merupakan suatu aturan yang memberikan informasi dalam bentuk hubungan “if-then” atau “jika-maka” yang dihitung berdasarkan database dan sifatnya probabilistik. Ide dari aturan ini adalah untuk mencari hubungan antara data dan mencari semua kemungkinan yang paling mungkin (most likely). Misalnya suatu cara yang dikembangkan oleh Wal-Mart, yaitu dengan menganalisis data belanja di gudang data, dan menemukan bahwa pembeli laki-laki yang membeli popok cenderung membeli bir, berdasarkan data tersebut Wal-Mart kemudian menempatkan popok dan bir secara berdekatan dan ternyata penjualan bir meningkat. Dari contoh dapat dilihat aturan asosiasi dengan istilah antecedent yaitu popok mewakili bagian “jika” dan consequent yaitu bir mewakili bagian “maka”.

Pada algoritma Apriori cara menentukan kandidat yang mungkin muncul yaitu dengan mencari kombinasi antar item yang memenuhi nilai minimum support dan nilai minimum confidence. Support adalah nilai penunjang atau persentase kombinasi item dalam database. Nilai Support dapat diperoleh dengan Rumus :

$$Support (A) = \frac{Jumah\ Transaksi\ Mengandung\ A}{Total\ Transaksi}$$

Sedangkan confidence merupakan nilai kepastian: kuatnya hubungan antar-item dalam aturan asosiasi.. Nilai Confidence dapat ditemukan, setelah pola

frekuensi tinggi sebuah item ditemukan. Rumus untuk menghitung nilai confidence adalah sebagai berikut:

“Jika A maka B” = (A B) maka:

$$\text{Confidence } P(B|A) = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi mengandung A}}$$

Penerapan data mining juga dilakukan di berbagai bidang seperti bidang kesehatan untuk memprediksi penyakit tertentu, kemudian untuk prediksi curah hujan, dibidang pendidikan untuk prediksi kinerja mahasiswa, dan lain-lain. Penerapan data mining di bidang pendidikan disebut dengan *Educational Data Mining* (EDM). EDM sangat berguna dan berpengaruh terhadap mahasiswa, pengajar atau pendidik, dan institusi dalam hal ini adalah perguruan tinggi untuk meningkatkan kinerja mahasiswa yang berdampak pada peningkatan mutu dan kualitas perguruan tinggi (Mohamed, Husain, & Rashid, 2015).

Selain itu penelitian lain mengenai *Data Mining* Korelasi Pekerjaan Orang Tua Terhadap Indeks Prestasi Kumulatif Lulusan Studi Kasus STMIK Kaputama Binjai menggunakan metode *data mining* berbasis rule atau aturan. Penelitian ini membahas mengenai faktor-faktor yang mempengaruhi prestasi akademik antara lain motivasi belajar, peranan orang tua dan perekonomian keluarga. Penelitian ini dilakukan karena mahasiswa di STMIK Kaputama Binjai memiliki IPK lulusan masih dominan dibawah 3.0 sehingga diperlukan upaya atau strategi untuk menaikkan IPK. Dengan mengimplementasikan data mining pada data lulusan mahasiswa dengan algoritma apriori, dapat diketahui berapa besar confidence dan support pekerjaan orang tua dengan IPK lulusan. Pekerjaan orang tuanya PNS memiliki kemungkinan 60% akan lulus dengan IPK cukup memuaskan, dan bagi mahasiswa yang pekerjaan orang tuanya wiraswasta kemungkinan 54% lulus dengan IPK memuaskan, sedangkan yang pekerjaannya petani kemungkinan 43% lulus dengan IPK memuaskan (Novriyenni & Sihombing, 2016).

Penelitian serupa di bidang pendidikan tinggi dengan menerapkan metode backpropagation yaitu untuk mengetahui korelasi antara nilai rata-rata rapor dan IPK akhir mahasiswa. Setelah dilakukan pelatihan dan pengujian dapat disimpulkan bahwa nilai rata-rata rapor siswa tidak cukup menjamin untuk

mendapatkan IPK akhir yang tinggi di perguruan tinggi. Sehingga tidak ada korelasi antara nilai rata-rata rapor dan IPK akhir Mahasiswa. Hal ini dibuktikan dengan hasil pengujian, dimana semua data uji termasuk ke dalam kelas 1 ($IPK \geq 3.0$) sedangkan target sebenarnya pada 5 data termasuk kedalam kelas 0 ($IPK < 3.0$) (Utami, 2016).

Penelitian lain dengan menggunakan metode yang berbeda yaitu selain metode backpropagation dilakukan pada tahun 2014. Penelitian ini membahas mengenai korelasi jurusan siswa tingkat SMA/ sederajat dengan jurusan di perguruan tinggi terhadap IPK dengan metode Apriori. Berdasarkan hasil penelitian tersebut dapat diketahui confidence dan support perolehan IPK yakni mahasiswa dari jurusan IPA yang memilih prodi selain Sistem Informasi kemungkinan 66% lulus dengan IPK memuaskan, sedang mahasiswa dari jurusan IPS yang memilih program studi apapun memiliki tingkat kelulusan dengan IPK kurang memuaskan sebesar 49% (Buaton, Manoor, & Maulita, 2014).

Algoritma Apriori juga digunakan dalam penelitian tahun 2017 untuk menampilkan korelasi nilai akademik dengan kelulusan mahasiswa. Hasil penelitian menunjukkan adanya hubungan antara kategori kelulusan dengan nilai mata kuliah berdasarkan proses asosiasi dan kombinasi data. Kemudian berdasarkan training dan pengujian jumlah data training dapat mempengaruhi persentase kecocokan atau keakurasian data mining. Pada pengujian aplikasi dapat dilihat bahwa jika mengubah nilai threshold, akan dihasilkan kombinasi yang bervariasi. Nilai threshold yang besar belum tentu menjadi nilai threshold terbaik, dengan tingkat keberhasilan yang tinggi dan demikian juga sebaliknya. Nilai threshold yang terbaik dapat dipengaruhi oleh jumlah data dan jumlah kombinasi data yang digunakan (Br Ginting, Purba, & Sumitra, 2017).

Metode Data Mining yaitu Association dengan Algoritma Apriori digunakan untuk mencari hubungan antara data murid dengan nilai sekolah. Data murid merupakan data awal masuk murid yang terdiri dari tempat dan tahun lahir, agama, dan tahun masuk. Nilai minimal support yang digunakan sebesar 20% dan nilai confidence merupakan parameter yang menentukan seberapa besar hubungan antara data murid dengan nilai sekolah. Nilai confidence terbesar diperoleh dari hubungan antara agama Kristen dengan nilai sekolah $\uparrow 77-82$ sebesar 54,5%.

Penelitian ini dilakukan pada SD Kristen Kalam Kudus IV Alam Raya (Grand, 2018).

Berdasarkan pada penelitian-penelitian yang dilakukan sebelumnya belum ada penelitian yang membahas tentang keterkaitan atau korelasi antara atribut nilai UAN SMA/Sederajat, IPK tahun pertama terhadap masa studi mahasiswa. Sehingga penulis bermaksud untuk melakukan penelitian untuk menganalisa dan mengukur tingkat korelasi nilai UAN, IPK tahun pertama terhadap masa studi.

2.2 Landasan Teori

2.2.1 Data Mining

Data Mining (penambangan data) merupakan proses untuk menemukan pola yang menarik dan pengetahuan dari sejumlah besar data (Jiawei & Kamber, 2001). Sumber data dapat berupa basis data, data warehouse, web, gudang informasi lain dan data yang berasal dari sistem yang dinamis. Proses penambangan data dapat dilakukan melalui algoritma. Kinerja sebuah algoritma memiliki kriteria antara lain (Widodo, Handayanto, & Herlawati, 2013):

- 1) Akurasi prediksi, menentukan seberapa akurat suatu algoritma dalam memprediksi keluaran,
- 2) Kecepatan, menunjukkan seberapa cepat dalam memproses data masukan,
- 3) *Robustness*, menggambarkan kemampuan melakukan prediksi yang akurat walau dalam kondisi ekstrim dan banyak gangguan.
- 4) Skalabilitas, kemampuan memproses data baik dalam ukuran yang lebih besar maupun dari bidang lain yang berbeda.
- 5) Interpretability, menggambarkan kemudahan untuk dipahami dan diinterpretasikan.
- 6) Kesederhanaan, merupakan sifat yang cenderung dipilih untuk menyelesaikan permasalahan.

2.2.2. RapidMiner

RapidMiner merupakan perangkat lunak yang bersifat open source. RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. RapidMiner menyediakan GUI (*Graphic User Interface*) untuk merancang sebuah pipeline analitis. GUI ini akan menghasilkan file XML (*Extensible*

Markup Language) yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke data (Aprilla C, Aji Baskoro, Ambarwati, & Wicaksana, 2013).

2.2.3 Assosiation

Association merupakan aturan yang mempelajari *item* atau atribut yang selalu muncul secara bersamaan (Ye, 2015). Konsep aturan *association* berasal dari *frequent pattern mining* yang mencari pola barang yang sering muncul secara bersamaan pada sebuah himpunan data. *Frequent pattern mining* mengarah pada penemuan asosiasi dan korelasi antar item dalam kumpulan data transaksional atau relasional yang besar. Nilai *support* dan *confidence* merupakan ukuran dari aturan ketertarikan (Jiawei & Kamber, 2001). Contoh: nilai *support* sebesar 2% berarti bahwa 2% dari semua transaksi menunjukkan barang A dan B dibeli secara bersamaan. Dan nilai *confidence* sebesar 50% berarti bahwa 50% pembeli barang A juga membeli barang B. Aturan *association* dianggap menarik jika memenuhi ambang batas *minimum support* dan ambang batas *minimum confidence*. Ambang batas ini bisa diatur oleh pengguna atau pakar domain. Secara umum, aturan *association mining* terdiri dari 2 langkah, yaitu (Jiawei & Kamber, 2001):

1. Temukan semua *itemset* yang sering muncul. Setiap *itemset* memenuhi *minimum support*.
2. Buat sebuah aturan asosiasi yang kuat dari *itemset* yang sering muncul. Peraturan harus memenuhi dukungan minimum *support* dan *confidence*.

2.2.4 Algoritma Apriori

Algoritma Apriori adalah sebuah algoritma untuk menemukan semua kumpulan item (*itemset*) yang memiliki dukungan tidak kurang dari *minimum support* (Wu & Kumar, 2009). Algoritma Apriori diusulkan oleh Agrawal dan Srikant pada tahun 1994 untuk *mining itemset* yang sering muncul untuk aturan asosiasi Boolean (Jiawei & Kamber, 2001). Pada tahap pertama, menghitung semua *itemset* yang ditemukan di *database* transaksi yang dukungannya memiliki *minimum support*. *Itemset* ini disebut *frequent itemset*, meski hanya memiliki beberapa persen minimum ambang batas *support*. Pada fase kedua, minimum ambang batas *confidence* mulai berlaku: Algoritma

menghasilkan semua aturan dari *frequent itemset* yang dikonfirmasi melebihi ambang batas (Maimon & Rokach, 2010).

BAB III

METODOLOGI PENELITIAN

3.1 Alat dan Bahan Penelitian

3.1.1 Alat Penelitian

Alat-alat yang digunakan dalam penelitian adalah sebagai berikut :

- a. Sebuah laptop
- b. Software
 - 1) Microsoft Excel
 - 2) RapidMiner 5

3.1.2 Bahan Penelitian

Bahan yang digunakan dalam penelitian ini adalah data alumni Fakultas Ilmu Komputer Program Studi Teknik Informatika UNWIDHA tahun lulus 2018 sampai dengan 2020 yang terdiri dari atribut jurusan sekolah, tahun masuk, nilai mata kuliah konsentrasi dan tingkat kelulusan (masa studi dan IPK). Data belum semuanya tersimpan dalam sistem komputer karena masih ada sebagian data yang disimpan atau direkap secara manual.

3.2 Metodologi Penelitian

3.2.1 Proses Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data alumni program studi Teknik Informatika Fakultas Ilmu Komputer UNWIDHA Klaten dengan tahun lulus 2018-2020, yang digunakan sebagai data pelatihan dan pengujian. Data tersebut antara lain jurusan sekolah, tahun masuk, nilai mata kuliah konsentrasi dan tanggal kelulusan untuk mengetahui lama studi mahasiswa. Pengumpulan data dilakukan secara langsung yaitu melalui wawancara dengan bagian akademik Fakultas Ilmu Komputer UNWIDHA. Hasil wawancara tersebut diperoleh data alumni seperti nama mahasiswa, NIM, tahun lulus, mata kuliah penunjang konsentrasi (Sistem Informasi dan Sistem Cerdas). Selain dengan wawancara pengumpulan data juga dilakukan melalui studi pustaka

dengan mencari referensi terkait penelitian seperti sumber dari buku, jurnal, internet, dan lain-lain.

Data yang digunakan diperoleh dari KHS dan transkrip nilai mahasiswa. Sampel data yang digunakan terdiri dari 42 data. Hal ini dikarenakan terdapat beberapa data yang memiliki nilai atribut yang tidak lengkap (missing value) sehingga data yang benar-benar dapat digunakan adalah data yang telah melewati tahapan *preprocessing*.

Data yang diperlukan adalah data mahasiswa dengan variabel yang digunakan antara lain:

1. Jurusan sekolah
2. NIM
3. Konsentrasi atau peminatan
4. Nilai mata kuliah penunjang SI,
5. Nilai mata kuliah penunjang SC,

Data awal atau data mentah berupa dokumen atau file Ms.Excel. Sedangkan perangkat lunak (tools) yang digunakan dalam melakukan pengelompokan adalah RapidMiner. RapidMiner merupakan perangkat lunak berbasis open source atau berlisensi gratis yang digunakan sebagai alat bantu dalam data mining dan knowledge discovery. Rapidminer dilisensikan oleh GNU Affero General Public License versi 3 dan saat ini tersedia dalam versi 5.3. Awalnya dikembangkan mulai tahun 2001 untuk kecerdasan buatan Universitas Dortmund. Metode yang digunakan dalam memproses data mining menggunakan algoritma Apriori.

3.2.2 Pengolahan Awal Data

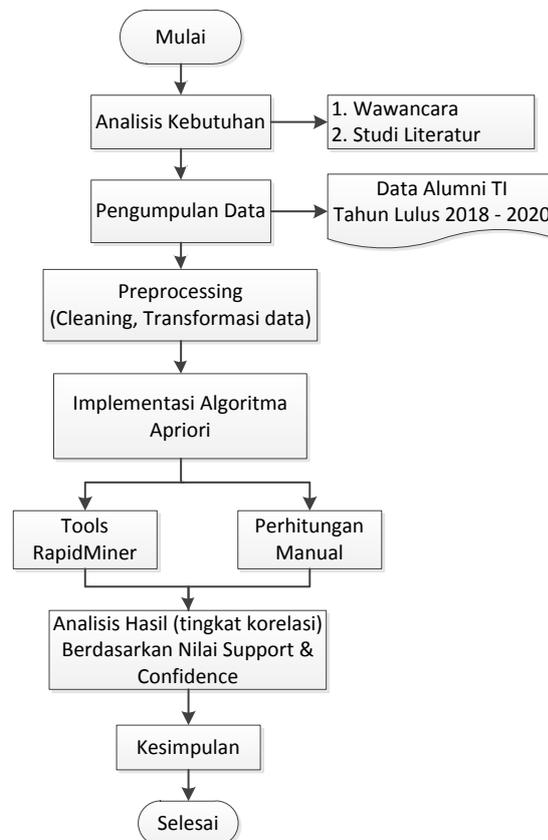
Pada data mentah sering ditemukan nilai yang hilang, tidak tersimpannya nilai, sampling tidak cukup bagus dan sebagainya. Maka perlu ditingkatkan kualitasnya dengan melakukan penyiapan data. Kurang baiknya kualitas data mentah dikarenakan adanya kesalahan dalam penyimpanan data. Data awal atau data mentah didapat dalam format .xls atau .xlsx (Microsoft Excel). Kemudian masing-masing nilai dari variabel atau atribut akan dikategorikan ke dalam

kriteria yang sudah ditentukan. Didalam data mining pengolahan data di awal atau penyiapan data termasuk kedalam tahapan *preprocessing* antara lain:

1. Transformasi data, untuk meningkatkan performa dengan melakukan normalisasi data.
2. Data reduction yaitu memperoleh data yang hanya mempunyai atribut yang berhubungan dengan penelitian, atribut yang tidak diperlukan akan dikurangi.
3. Pembersihan data yaitu data yang memiliki nilai atribut yang tidak lengkap (missing value) atau kosong akan dihilangkan.

3.2.3 Tahapan Penelitian

Tahapan yang dilakukan dalam penelitian ini dapat dilihat pada gambar 3.1 berikut ini.



Gambar 3.1 Tahapan Penelitian

Berdasarkan gambar diatas dapat dijelaskan bahwa:

1. Melakukan analisis kebutuhan dengan cara wawancara kepada pihak terkait atau narasumber yaitu Kepala Seksi atau Bagian Akademik Fakultas Ilmu Komputer. Kemudian melakukan studi pustaka yaitu dengan mengumpulkan data atau informasi dengan membaca dan mempelajari buku, jurnal, penelitian maupun referensi lainnya.
2. Pengumpulan data yang dilakukan secara langsung yaitu data alumni program studi Teknik Informatika.
3. Melakukan tahapan preprocessing dengan mereduksi data yang kosong atau yang memiliki nilai atribut tidak lengkap.
4. Implementasi metode apriori menggunakan tools RapidMiner dengan mengukur tingkat korelasi berdasarkan nilai Support dan Confidence yang diperoleh.
5. Menganalisa hasil yang diperoleh dari pengukuran tingkat korelasi menggunakan atribut yang telah ditentukan.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Pengolahan Data

Dataset yang digunakan sebanyak 42 record data yang merupakan data alumni program studi Teknik Informatika Unwidha Klaten tahun lulus 2018-2020. Dari berbagai informasi yang diperoleh berupa data beserta atributnya didapat suatu pola dan pengetahuan yang berharga terkait dengan pengelompokan mahasiswa berdasarkan kemampuan akademis melalui teknik klustering pada data mining.

Variabel yang digunakan terdiri dari data jurusan sekolah sebelum masuk perguruan tinggi, NIM (Nomor Induk Mahasiswa), data tahun lulus dan data nilai mata kuliah penunjang konsentrasi yaitu Basis Data, Praktikum Pemrograman I, Praktikum Pemrograman II, Praktikum Pemrograman III, Desain Berorientasi Objek (DBO), Praktikum DBO, Kecerdasan Buatan, Pengolahan Citra Digital (PCD), Metode Statistik dan Matematika Lanjut. Setelah semua data terkumpul kemudian menghitung persentase hubungan atau korelasi antara jurusan sekolah, nilai mata kuliah konsentrasi SI & SC terhadap tingkat kelulusan menggunakan teknik asosiasi dalam data mining yaitu algoritma apriori.

Berikut hasil transformasi data nilai yaitu dengan mengubah nilai setiap mata kuliah kedalam kategori atau kelas-kelas tertentu:

Tabel 4.1 Transformasi Nilai Mata Kuliah

No	Nilai	Kategori
1	A	5
2	AB	4
3	B	3
4	BC	2
5	C	1

Tabel 4.2 Transformasi Jurusan Sekolah

No	Jurusan	Kategori
1	SMA/MA	0
2	SMK	1

Tabel 4.3 Transformasi Masa Studi

No	Masa Studi (tahun)	Kategori
1	< 4	1
2	4 – 4,5	2
3	4,5 - 5	3
4	> 5	4

Berikut sampel dataset yang telah melalui tahapan *preprocessing* dan siap untuk dilakukan proses asosiasi dengan apriori.

Tabel 4.4 Sampel Dataset Konsentrasi Sistem Informasi

No	NIM	Jur Sekolah	Konsentrasi (0=SI, 1 = SC)	Mata Kuliah Konsentrasi SI			MasaStudi
				PBO	PEM_Mobile	JST	
1	1471101614	SMK	0	AB	AB	AB	2
2	1471101610	SMK	0	AB	AB	AB	2
3	1471101600	SMK	0	B	A	AB	2
4	1471101609	SMA	0	AB	A	AB	3
5	1471101613	SMA	0	AB	B	AB	4
6	1471101605	SMK	0	AB	AB	A	3
7	1471101672	SMA	0	A	AB	B	2
8	1471101644	SMK	0	AB	AB	B	2
9	1471101649	SMK	0	AB	A	AB	2
10	1471101669	SMA	0	B	A	AB	2
11	1471101662	SMA	0	AB	A	AB	4
12	1471101650	SMK	0	AB	A	AB	2
13	1471101626	SMA	0	AB	A	A	2
14	1471101628	SMA	0	AB	A	AB	2
15	1471101630	SMK	0	A	A	AB	2
16	1471101639	SMA	0	AB	A	AB	2
17	1471101616	SMK	0	AB	B	B	4
18	1571101685	SMA	0	BC	B	B	4
19	1571101693	SMA	0	AB	A	AB	3
20	1571101695	SMK	0	A	AB	B	2
21	1571101697	SMK	0	B	AB	B	4
22	1571101704	SMA	0	AB	A	AB	2
23	1571101711	SMA	0	AB	B	BC	2
24	1571101712	SMK	0	B	B	B	3
25	1571101713	SMK	0	B	AB	B	3
26	1571101718	SMA	0	AB	A	AB	2
31	1671100013	SMK	0	AB	A	AB	3
32	1671100015	SMK	0	A	AB	B	3
36	1671100019	SMA	0	AB	A	AB	3
38	1671100037	SMA	0	AB	A	AB	3

Tabel 4.5 Sampel Dataset Konsentrasi Sistem Cerdas

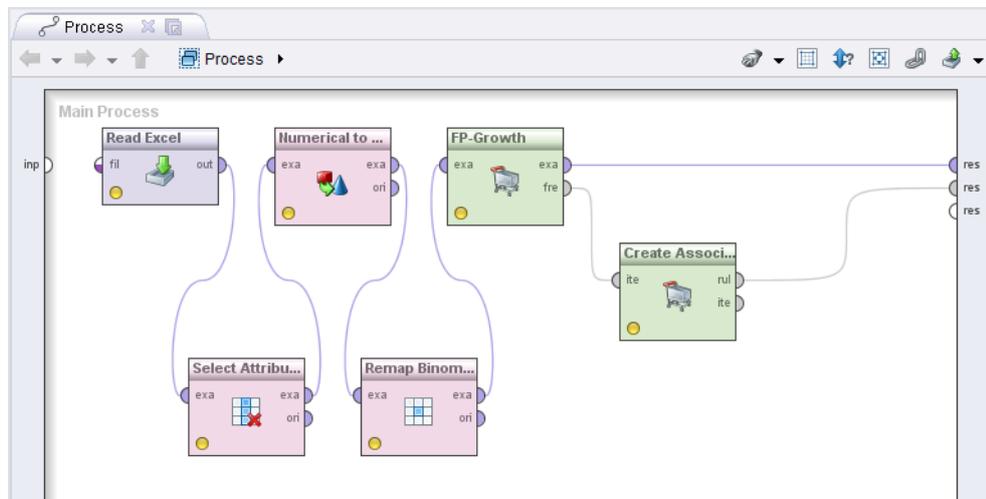
No	NIM	Jur Sekolah	Konsentrasi (0=SI, 1 = SC)	Mata Konsentrasi SC			MasaStudi
				LogikaSamar	DataMining	SistemPakar	
1	1671100002	SMA	1	AB	A	AB	1
2	1671100004	SMA	1	B	AB	B	2
3	1671100009	SMA	1	BC	B	B	2
4	1671100011	SMK	1	B	B	B	3
5	1671100022	SMA	1	AB	B	B	3
6	1671100041	SMA	1	BC	B	BC	3
7	1671100049	SMK	1	BC	B	BC	2
8	1671100030	SMK	1	B	AB	B	2
9	1671100038	SMK	1	A	A	AB	4
10	1671100054	SMK	1	AB	A	AB	3
11	1671100006	SMA	1	BC	AB	B	2
12	1671100061	SMA	1	BC	AB	B	3

4.2 Hasil Implementasi *Tool RapidMiner*

Data yang digunakan dalam proses asosiasi pada RapidMiner berjumlah 42 data. Dimana data dengan kelompok konsentrasi SC terdiri dari 12 data dan konsentrasi SI terdiri dari 30 data. Hasil teknik asosiasi mahasiswa Teknik Informatika Unwidha Klaten menggunakan RapidMiner sebagai berikut:

1. Proses Association

Proses penerapan apriori dapat dilihat pada gambar berikut ini.



Gambar 4.1 Tampilan Implementasi Apriori pada RapidMiner

2. Table View

Table view merupakan sheet untuk menampilkan data yang telah diolah secara keseluruhan. Table view dapat lihat pada gambar berikut ini.

No.	Premises	Conclusion	Support	Confidence
50	Jur_Sekolah	MasaStudi, LogikaSamar	0.417	1
51	MasaStudi, Jur_Sekolah	LogikaSamar	0.417	1
52	LogikaSamar, Jur_Sekolah	MasaStudi	0.417	1
53	Jur_Sekolah	MasaStudi, DataMining	0.417	1
54	MasaStudi, Jur_Sekolah	DataMining	0.417	1
55	DataMining, Jur_Sekolah	MasaStudi	0.417	1
56	Jur_Sekolah	LogikaSamar, DataMining	0.417	1
57	LogikaSamar, Jur_Sekolah	DataMining	0.417	1
58	DataMining, Jur_Sekolah	LogikaSamar	0.417	1
59	SistemPakar	MasaStudi, LogikaSamar, DataMi	1	1
60	MasaStudi	SistemPakar, LogikaSamar, Data	1	1

Gambar 4.2 Tampilan Table View Association Rules

Dari gambar 4.2 diatas dapat dilihat pada baris dengan No = 55 yaitu nilai Support = 0.417 dan nilai Confidence = 1 jika premises jurusan sekolah yaitu Sistem Cerdas, mata kuliah adalah data mining dan yang menjadi cloncuSSION adalah masa studi. Hal ini menunjukkan bahwa jurusan sekolah dan nilai mata kuliah data mining memiliki tingkat kepercayaan 41.7 % dan didukung oleh 100% data keseluruhan.

BAB V

PENUTUP

5.1 Kesimpulan

Kesimpulan yang diperoleh setelah melakukan penelitian ini antara lain:

1. Data pelatihan yang digunakan sebanyak 42 *record* data dan menggunakan teknik asosiasi yaitu algoritma apriori.
2. Hasil implementasi menggunakan apriori memiliki nilai Support 0.417 dan nilai Confidence 1, berarti jurusan sekolah dan nilai mata kuliah data mining memiliki tingkat kepercayaan 41.7 % dan didukung oleh 100% data keseluruhan .

5.2 Saran

Saran yang dapat dilakukan untuk dilakukan penelitian selanjutnya antara lain:

1. Dibutuhkan faktor-faktor lain yang mendukung masa studi mahasiswa, tidak hanya faktor jurusan sekolah dan nilai mata kuliah konsentrasi. Selain itu juga perlu memperbanyak data yang digunakan dalam proses pelatihan dan pengujian.
2. Proses pelatihan dan pengujian dapat dilakukan kedalam beberapa model atau bentuk dengan mengubah nilai parameter yang digunakan.

DAFTAR PUSTAKA

- Aprilla C, D., Aji Baskoro, D., Ambarwati, L., & Wicaksana, I. W. S. (2013). *Belajar Data Mining dengan RapidMiner*. (R. Sanjaya, Ed.). Jakarta.
- Br Ginting, S. L., Purba, S. A., & Sumitra, I. D. (2017). Algoritma Apriori untuk Menampilkan Korelasi Nilai Akademik dengan Kelulusan Mahasiswa : Data Mining. *Komputika: Jurnal Sistem Komputer*, 6(2), 59–65. <https://doi.org/10.34010/komputika.v6i2.1706>
- Buaton, R., Manaor, A., & Maulita, Y. (2014). Implementasi Data Mining Untuk Mengetahui Korelasi Jurusan Siswa PadaTingkat SMA/Sederajat Dengan Jurusan di Perguruan Tinggi Terhadap IPK dengan metode A Priori (Studi Kasus STMIK Kaputama). In *Prosiding Seminar Nasional PB3I ITM* (pp. 151–159).
- Grand. (2018). Penerapan Algoritma Apriori Untuk Menemukan Hubungan Data Murid dengan Nilai Sekolah. *IKRAITH-INFORMATIKA*, 2(1), 7–12.
- Jiawei, H., & Kamber, M. (2001). *Data mining: concepts and techniques*. San Francisco, CA, itd: Morgan Kaufmann. [https://doi.org/10.1002/1521-3773\(20010316\)40:6<9823::AID-ANIE9823>3.3.CO;2-C](https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C)
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook (Second)*. (O. M. and L. Rokach, Ed.). Springer.
- Mohamed, A., Husain, W., & Rashid, A. (2015). A Review on Predicting Student ' s Performance using Data Mining Techniques. *Procedia - Procedia Computer Science*, 72, 414–422. <https://doi.org/10.1016/j.procs.2015.12.157>
- Novriyenni, & Sihombing, A. (2016). Data Mining Korelasi Pekerjaan Orang Tua Terhadap Indek Prestasi Kumulatif Lulusan Studi Kasus STMIK Kaputama Binjai. In *Seminar Nasional Ilmu Komputer (SNIKOM)*.
- Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: C.V Andi Offset.
- Romadhona, A., Suprapedi, & Himawan, H. (2017). PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU BERDASARKAN USIA, JENIS KELAMIN DAN INDEKS PRESTASI MENGGUNAKAN ALGORITMA DECISION TREE. *Jurnal Teknologi Informasi*, 13(1), 69–83.
- Utami, D. (2016). *PENERAPAN JARINGAN SARAF TIRUAN UNTUK MENGETAHUI KORELASI ANTARA NILAI RATA-RATA RAPOR DAN IPK AKHIR MAHASISWA DENGAN MENGGUNAKAN METODE BACKPROPAGATION*. Skripsi Jurusan Teknik Informatika UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU.
- Widodo, P. P., Handayanto, R. T., & Herlawati. (2013). *Penerapan Data Mining Dengan Matlab*. Rekayasa Sains. <https://doi.org/10.1017/CBO9781107415324.004>
- Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. CRC Press.
- Ye, N. (2015). *Data Mining Theories, Algorithms, and Examples*. Boca Raton : CRC Press.