

**PENERAPAN DATA MINING UNTUK
MENGHASILKAN POLA KELULUSAN SISWA
MENGUNAKAN ALGORITMA APRIORI
(Studi Kasus: SMKN 1 Bukit Kemuning)**

SKRIPSI



Disusun Oleh:

Nurul Hidayati

1711050096

**PRODI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
INSTITUT INFORMATIKA DAN BISNIS DARMAJAYA
BANDAR LAMPUNG**

2021

**PENERAPAN DATA MINING UNTUK
MENGHASILKAN POLA KELULUSAN SISWA
MENGUNAKAN ALGORITMA APRIORI
(Studi Kasus: SMKN I Bukit Kemuning)**

SKRIPSI

Diajukan Sebagai Salah Satu Syarat Untuk Mencapai Gelar
SARJANA KOMPUTER
Pada Program Studi Sistem Informasi
IIB Darmajaya Bandar Lampung



Disusun Oleh:

Nurul Hidayati
1711050096

**PRODI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
INSTITUT INFORMATIKA DAN BISNIS DARMAJAYA
BANDAR LAMPUNG
2021**

PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi yang saya ajukan ini adalah hasil karya saya sendiri, tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi atau karya yang pernah ditulis atau diterbitkan orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka. Karya ini adalah milik saya dan pertanggungjawaban sepenuhnya berada di pundak saya.

Bandar Lampung, 12 Desember 2021



Penulis

HALAMAN PERSETUJUAN

Judul Skripsi : Penerapan Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Menggunakan Algoritma Apriori (Studi Kasus: SMKN 1 Bukit Kemuning)

Nama Mahasiswa : Nurul Hidayati

NPM : 1711050096

Program Studi : Sistem Informasi

Disetujui Oleh:

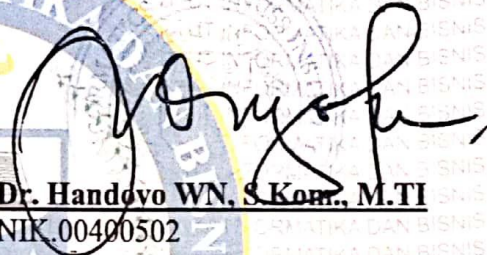
Penguji II

Ketua Program Studi



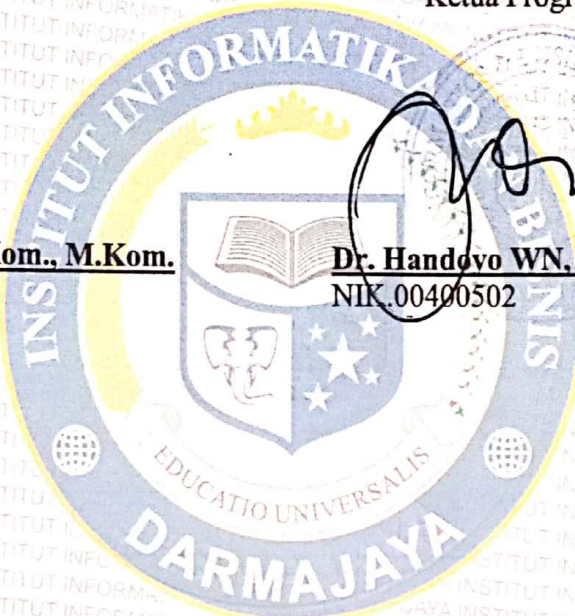
Sri Karnila, S.Kom., M.Kom.

NIK. 01251005



Dr. Handoyo WN, S.Kom., M.TI

NIK. 00400502



HALAMAN PENGESAHAN

Telah diuji dan dipertahankan didepan Tim Penguji Tugas Akhir Program Studi Sistem Informasi IIB Darmajaya dan dinyatakan diterima untuk memenuhi syarat guna memperoleh gelar Sarjana Komputer

Mengesahkan:

1. Tim Penguji

Tanda Tangan


Anggota 1 : Nurjoko, S.Kom., M.TI



Anggota 2 : Sri Karnila, S.Kom., M.Kom.



2. Dekan Fakultas Ilmu Komputer


Zaidir Jamal, S.T., M.Eng
NIK.00590203

Tanggal Lulus Ujian Skripsi: 10 September 2021

PERSEMBAHAN

Untuk yang terkasih.

Bapak Katimin dan Aa Wawan, Lelaki Pejuang dan terhebat dalam hidupku.

Ibu Sarinah dan Tete Meli, kedua perempuan tangguhku.

Lelaki misterius yang masih dalam lindungan Tuhan ;(

MOTTO

Jangan Mundur, Lari Itu Sifat Pengecut.

Jadilah Perempuan Yang Berani Menentukan Pilihan, Seperti Kartini.

Be Brave But Dont Be Stupid.

Sesungguhnya Bersama Kesulitan Itu Ada Kemudahan.

Yakin Usaha Sampai.

RIWAYAT HIDUP



BIODATA

Nama : Nurul Hidayati
Nama Panggilan : Nuri
Tempat Tanggal Lahir : Sukasari, 10 Maret 1999
Jenis Kelamin : Perempuan
Agama : Islam
Alamat : Jl. Bumi Manti Surya Kencana IV, Kampung
Baru, Bandar Lampung
No. HP : 085382057353
E-mail : nurulhidayati100340@gmail.com
Blog : nuripark99.blogspot.com
Hobi : Membaca dan Menulis

RIWAYAT PENDIDIKAN

SDN 02 Sukasari 2005-2011
SMPN 4 Tanjung Raja 2011-2014
SMKN 2 Kotabumi 2014-2017

ABSTRAK

PENERAPAN DATA MINING UNTUK MENGHASILKAN POLA KELULUSAN SISWA MENGGUNAKAN ALGORITMA APRIORI (Studi Kasus: SMKN 1 Bukit Kemuning)

Oleh:

Nurul Hidayati

1711050096

nurulhidayati100340@gmail.com

Informasi kini menjadi elemen penting yang harus kita dapatkan dengan cepat, akurat dan relevan. Kebutuhan akan informasi tidak hanya sekedar tentang kecelakaan lalu lintas, bencana alam, diskon akhir bulan, inflasi uang dan harga saham, tetapi dibutuhkan informasi tentang lokasi strategis pemasaran barang, prediksi kecelakaan lalu lintas, dan sekolah terbaik. Salahsatu indikator sekolah terbaik bisa dilihat dari kualitas alumni yang unggul dan dapat bersaing dengan alumni sekolah lainnya. Maka dibutuhkan sebuah strategi yang tepat untuk memajukan kualitas siswa, dimana strategi tersebut diambil bukan berdasarkan perasaan suka atau tidak suka, melainkan berdasarkan data yang ada. Pertumbuhan data yang pesat telah menciptakan tumpukan-tumpukan data tetapi miskin informasi. Maka dari itu, dibutuhkan suatu teknik untuk mengekstrak informasi dari tumpukan data tersebut. Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dari dalam *database*. Penemuan aturan kombinasi dari setiap item dalam data disebut dengan asosiasi. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien, yaitu analisis pola frekuensi tinggi. Algoritma apriori dapat digunakan untuk membentuk kandidat kombinasi *item* untuk diterapkan aturan asosiatifnya yang mempunyai nilai keseringan tertentu. Hasil analisis menunjukkan bahwa data mining dapat digunakan untuk mencari informasi. Dihasilkan asosiasi yang menerangkan bahwa asal siswa mempengaruhi status kelulusan siswa.

Kata Kunci: Informasi, Data Mining, Apriori.

ABSTRACT

APPLICATION OF DATA MINING TO GENERATING STUDENT GRADUATION PATTERNS USING APRIORI ALGORITHM

(Case Study: SMKN 1 Bukit Kemuning)

By:

Nurul Hidayati

1711050096

nurulhidayati100340@gmail.com

Information is an important element that we must obtain quickly, accurately and relevantly. The need for information is not only about traffic accidents, natural disasters, month-end discounts, money inflation and stock prices, but also information about strategic locations for marketing goods, predictions of traffic accidents, and the best schools. One of the best school indicators can be seen from the quality of alumni who are superior and can compete with other school alumni. So we need an appropriate strategy to advance the quality of students, where the strategy is taken not based on feelings of liking or disliking, but based on existing data. The rapid growth of data has created piles of data but poor information. Therefore, we need a technique to extract information from the pile of data. Data mining is a term used to describe the discovery of knowledge from within the database. Finding the combination rule of each item in the data is called association. In particular, one of the stages of association analysis that has attracted the attention of many researchers to produce efficient algorithms is the analysis of high-frequency patterns. A priori algorithm can be used to form candidate item combinations to apply associative rules that have a certain frequency value. The results of the analysis show that data mining can be used to find information. The resulting association explains that the origin of the student affects the student's graduation status.

Keywords: Information, Data Mining, A priori.

KATA PENGANTAR

Saya bersyukur kepada Tuhan Yang Maha Esa, Tuhan yang ada sebelum saya ada, Tuhan yang ada saat saya ada, Tuhan yang selalu ada saat saya tiada. Tuhan yang dengan kasih sayangNya, selalu memberikan skenario terbaik yang atas ijin-Nya penelitian ini bisa saya selesaikan.

Hari demi hari telah saya lewati, pergantian tahun juga saya alami ketika mengerjakan skripsi ini. Berbagai rasa dan emosi telah mewarnai perjalanan saya, kadang saya semangat menulis satu persatu kata yang ada dipikiran saya, namun tak jarang saya merasa sedih, malas, dan badmood karena beberapa hal. Namun satu hal yang tak pernah saya lupa, apapun yang terjadi saya mesti berusaha untuk menyelesaikan apa yang sudah saya mulai. Hingga akhirnya saya sampai dititik ketika pembimbing memberikan ijin untuk sidang skripsi.

Skripsi ini saya buat dengan jerih payah saya sendiri, yang dengan tidak menyerah meskipun sulit, selalu berusaha menyemangati diri sendiri dan mencoba satu kali lagi. Dengan selesainya skripsi ini, saya mengucapkan terimakasih kepada:

1. Bapak Dr.(Can) Ir. Firmansyah YA, MBA., M.Sc. selaku Rektor IIB Darmajaya.
2. Bapak Dr. RZ. Abdul Aziz, S.T., MT. selaku Wakil Rektor I IIB Darmajaya.
3. Bapak Rony Nazar, S.E., MT. selaku Wakil Rektor II IIB Darmajaya.
4. Bapak Muprihan Thaib, S.Sos., MM. Selaku Wakil Rektor III IIB Darmajaya.
5. Bapak Prof. Dr. Ir. Raden Achmad Bustomi Rosadi, M.S. selaku Wakil Rektor IV IIB Darmajaya.
6. Bapak Dr. Handoyo Widi Nugroho, S.Kom., M.T.I selaku Ketua Jurusan Sistem Informasi sekaligus dosen pembimbing yang telah membimbing saya dengan penuh kesabaran.
7. Bapak Nurjoko, S.Kom., M.T.I dan Ibu Sri Karnila, S.Kom., M.T.I selaku Dosen Penguji.
8. Ibu Halimah, S.Kom., M.T.I selaku Dosen Pembimbing Akademik, yang selalu mendengarkan dan mengarahkan saya selama kuliah di IIB Darmajaya.

9. Kedua orang tua, Aa dan Tete yang selalu menyemangati dan mendukung saya dalam keadaan apapun.
10. Ferdiansyah Roza sebagai teman berpikir yang sudah mengorbankan waktu tidurnya untuk membantu saya membuat tabel tabular secara manual, yang kemudian saya temukan cara cepatnya. Rodi Putra Setiawan sebagai teman berpikir dan bercerita yang tidak pernah lelah mendengarkan keluh kisah saya, yang selalu menemani saya merevisi naskah dan selalu memberikan dukungan. Aprilia Hari Purnama sosok sahabat yang rela membagi pakai laptopnya sehingga saya bisa menyelesaikan penelitian ini.
11. Teman-teman HMI Komisariat Darmajaya, Nufus, Linda, Andika, Fitria, Riski, Fahrul, Reza, dan semuanya yang tidak bisa saya sebutkan satu persatu. Terimakasih telah memberikan dukungan secara moril dan materiil sehingga saya dapat menyelesaikan penelitian ini
12. Teman seperjuangan saya, Bela Savira, Sindi Sintia, Tasha Aninda, Agatha Chiara.
13. Teman-teman Angkatan 2017 IIB Darmajaya, Asrul, Afri, Median, Tiwi, Fanny, Fuzy, Doni, Deni, Ajib, Kustia, dan lainnya yang selalu saling menyemangati yang tidak bisa saya sebutkan satu persatu.

Saya menyadari bahwa penelitian ini memiliki banyak kekurangan dan kesalahan. Saya sangat menerima jika ada kritik dan saran untuk penelitian kedepannya yang lebih baik.

Bandar Lampung, 12 Desember 2021

Penulis

DAFTAR ISI

PERNYATAAN.....	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN.....	v
PERSEMBAHAN.....	vi
MOTTO	vii
RIWAYAT HIDUP	viii
ABSTRAK	ix
<i>ABSTRACT</i>	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR TABEL	xv
DAFTAR GAMBAR.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Ruang Lingkup Masalah	4
1.3 Rumusan Masalah	4
1.4 Tujuan.....	5
1.5 Manfaat.....	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Perkembangan Data Mining.....	7
2.2 Pengelompokan Data Mining.....	9
2.3 KDD (<i>Knowledge Discovery in Database</i>)	12
2.4 <i>Association Rule</i>	14
2.5 Algoritma Apriori	14
2.6 Rapidminer	17
2.7 Penelitian Sebelumnya Tentang Algoritma Apriori dan Pencarian Pola Kelulusan Siswa	18
BAB III METODOLOGI PENELITIAN	22

3.1 Tahapan Penelitian	22
3.1.1 Studi Awal	23
3.1.2 Pengumpulan Data	23
3.1.3 Pre-processing Data	23
3.1.4 Proses Mining Menggunakan Algoritma Apriori	25
3.1.5 Interpretation and Evaluation	25
3.2 Alat dan Bahan	25
3.2.1 Alat	25
3.2.2 Bahan	26
3.3 Metodologi Pengumpulan Data	29
BAB IV HASIL DAN PEMBAHASAN	30
4.1 Analisis Menggunakan Algoritma Apriori	30
4.2 Analisis menggunakan Rapidminer	43
BAB V SIMPULAN DAN SARAN	53
5.1 Kesimpulan	53
5.2 Saran	54
DAFTAR PUSTAKA	55

DAFTAR TABEL

Tabel 3. 1 Format Data Tabular	24
Tabel 4. 1 Dataset.....	30
Tabel 4. 2 <i>Itemset</i> yang memenuhi <i>minimum support</i> 20%	31
Tabel 4. 3 Kombinasi 2 <i>Itemset</i> yang memenuhi <i>minimum support</i> 20%	33
Tabel 4. 4 Kombinasi 3 <i>Itemset</i> yang memenuhi <i>minimum support</i> 20%	35
Tabel 4. 5 Kombinasi 2 <i>itemset</i> yang memenuhi <i>minimum confidence</i> 60%.....	39
Tabel 4. 6 Kombinasi 3 <i>itemset</i> yang memenuhi <i>minimum confidence</i> 60%.....	41

DAFTAR GAMBAR

Gambar 2. 1 <i>Fishbone Diagram</i>	20
Gambar 3. 1 Tahapan <i>Knowledge Discovery in Database</i>	12
Gambar 3. 2 Tahapan Penelitian	22
Gambar 4. 1 <i>Import Data</i>	43
Gambar 4. 2 <i>Store Data</i>	44
Gambar 4. 3 Data Pada Lembar Kerja	44
Gambar 4. 4 <i>Select Attributes</i>	45
Gambar 4. 5 Proses <i>Select Attributes</i>	45
Gambar 4. 6 <i>Replace Missing Values</i>	46
Gambar 4. 7 Pemilihan atribut yang akan dihilangkan	46
Gambar 4. 8 Bagian <i>Parameters Replace Missing Values</i>	47
Gambar 4. 9 Statistik data yang sudah dibersihkan	47
Gambar 4. 10 Proses <i>Write CSV</i>	48
Gambar 4. 11 Proses <i>Import Data</i>	48
Gambar 4. 12 Jendela <i>Specify Your Data Format</i>	49
Gambar 4. 13 Jendela Format Kolom	49
Gambar 4. 14 Proses Pengimporan Data.....	50
Gambar 4. 15 Jendela Proses Pengolahan Data	50
Gambar 4. 16 Jendela Proses Pengolahan Data dengan Operator dan Parameter W- predictive Apriori.....	51
Gambar 4. 17 Pola yang Dihasilkan Proses Analisis	51

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang semakin pesat di era globalisasi saat ini tidak bisa dihindari lagi pengaruhnya terhadap kehidupan sehari-hari. Informasi kini menjadi elemen penting yang harus kita dapatkan dengan cepat, akurat dan relevan. Kebutuhan akan informasi tidak hanya sekedar tentang kecelakaan lalu lintas, bencana alam, diskon akhir bulan, inflasi uang dan harga saham. Tetapi kini informasi tentang lokasi strategis pemasaran barang, prediksi kecelakaan lalu lintas, prediksi terjadinya bencana alam, dan prediksi kelulusan siswa menjadi informasi berharga yang dapat menunjang proses bisnis dalam pengambilan keputusan. Informasi tersebut bisa didapatkan dengan menggunakan proses penambahan data.

Salah satu tujuan dari proses pembelajaran di Sekolah yaitu memiliki lulusan yang unggul dan kompeten dibidangnya. Setiap guru harus berusaha membimbing siswa-siswi dalam setiap pembelajaran yang dilakukan. Selain itu tentu ada faktor-faktor lain yang mempengaruhi kualitas siswa, baik faktor internal maupun faktor eksternal. Sekolah dituntut untuk memiliki lulusan yang unggul dan dapat bersaing serta memiliki kualitas yang baik. Oleh karena itu, dibutuhkan sebuah strategi yang tepat untuk memajukan kualitas siswa, dimana strategi tersebut diambil bukan berdasarkan perasaan suka atau tidak suka, melainkan berdasarkan data yang ada.

Seiring dengan kemajuan teknologi penyimpanan, yang semakin memungkinkan untuk menyimpan data dalam jumlah besar dengan biaya yang relatif rendah, telah muncul kesadaran yang berkembang bahwa terkubur informasi didalam data tersebut. Pertumbuhan data yang pesat telah menciptakan tumpukan-tumpukan data tetapi miskin informasi yang bermanfaat atau dengan tepat dikatakan bahwa

dunia menjadi kaya data tetapi miskin pengetahuan menjadikan pentingnya tumpukan data tersebut diolah sedemikian rupa sehingga menghasilkan informasi ataupun pengetahuan yang bermanfaat bagi institusi, perusahaan ataupun organisasi (Wardani, 2020) .

Knowledge Discovery in Database (KDD) merupakan proses analisa terstruktur untuk memperoleh informasi yang benar, baru, bermanfaat dan menemukan pola dari data yang besar dan kompleks. Data mining menjadi inti dari proses KDD, yakni dengan menggunakan algoritma tertentu untuk mengeksplorasi data, membangun model dan menemukan pola yang belum diketahui. Model digunakan untuk memahami fenomena data, analisis maupun prediksi (Zanuardi & Suprayitno, 2018).

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dari dalam *database*. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola. (Nofriansyah, 2014)

Data mining dibagi menjadi beberapa bagian berdasarkan tugas yang dilakukan, yaitu : Deskripsi, Estimasi, Prediksi, Klasifikasi, Pengclusteran, dan Asosiasi. Di Pada penelitian ini akan dilakukan proses asosiasi. Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Tugas ini, alam dunia marketing lebih dikenal dengan analisis keranjang pasar (*market based analysis*) (Nofriansyah, 2014).

Analisis Asosiasi atau *Association rule mining* adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi *item*. Analisis asosiasi dikenal

juga sebagai salah satu teknik data mining yang menjadi dasar dari salah satu teknik data mining lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien, yaitu analisis pola frekuensi tinggi (*frequent pattern mining*) (Fauzy, Saleh W, & Asror, 2016) .

Algoritma apriori dapat digunakan untuk membentuk kandidat kombinasi *item* untuk diterapkan aturan asosiatifnya yang mempunyai nilai keseringan tertentu. Pentingnya aturan asosiasi dapat diketahui dari dua parameter, yaitu *minimum support* yang merupakan presentase kombinasi *item* diseluruh transaksi dan *minimum confidence* yaitu kuatnya hubungan antar *item* dalam aturan asosiatif (Rizky & Sadikin, 2019) .

Rapidminer merupakan *software* untuk pengolahan data. Dengan menggunakan prinsip dan algoritma data mining, *rapidminer* mengekstrak pola-pola dari dataset yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan *database* (Novianti, 2019).

Penelitian sebelumnya adalah penentuan pola kelulusan siswa dengan metode *naive bayes* yang menghasilkan keakuratan sistem sekitar 99,82% dan memiliki nilai error 0,18% (Meilani & Susanti, 2015). Pada penelitian tersebut digunakan fungsi *clustering* data mining dengan metode *naive bayes*. Namun dipenelitian tersebut belum ada hasil penelitian berupa pola yang dihasilkan dan tidak dicantumkan sumber data yang digunakan dalam penelitian tersebut.

Berangkat dari kesenjangan tersebut, pada penelitian ini akan dilakukan analisis menggunakan fungsi asosiasi data mining dengan algoritma *apriori*. Data yang akan digunakan dalam penelitian ini adalah data siswa SMKN 1 Bukit Kemuning. Pendekatan yang dilakukan adalah mencari kesamaan antar *item* pada dataset, kemudian dibentuk aturan asosiatif berdasarkan kesamaan-kesamaan tersebut. Aturan asosiatif dijadikan acuan dalam penentuan pola pada penelitian ini.

Penggunaan teknik data mining diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi dalam tumpukan data sehingga akan menjadi informasi yang berharga.

Analisis ini akan membantu pihak sekolah menemukan pola kelulusan siswa-siswi dengan memanfaatkan data siswa yang sudah lulus. Dari pola tersebut diharapkan bisa diketahui faktor-faktor yang sangat berpengaruh pada tingkat kelulusan, sehingga akan membantu pihak sekolah untuk menyusun strategi yang tepat dalam meningkatkan kualitas sekolah dan menjadikan sekolah memiliki daya saing yang tinggi.

1.2 Ruang Lingkup Masalah

Dari latar belakang di atas, ruang lingkup masalah penelitian ini adalah:

1. Berfokus pada penerapan teknik data mining dengan algoritma apriori untuk menemukan pola kelulusan siswa.
2. Data yang digunakan adalah data Siswa SMKN 1 Bukit Kemuning angkatan tahun 2020.

1.3 Rumusan Masalah

Berdasarkan latar belakang dan ruang lingkup masalah di atas, maka rumusan masalah pada penelitian ini adalah:

1. Bagaimana menerapkan teknik data mining dengan algoritma apriori untuk menemukan pola kelulusan siswa?
2. Bagaimana mencari faktor-faktor yang sangat berpengaruh pada tingkat kelulusan sehingga akan ditemukan informasi yang berguna untuk membantu dalam pengambilan keputusan?

1.4 Tujuan

Adapun tujuan dari penelitian ini adalah:

1. Menerapkan teknik data mining dengan Algoritma Apriori dan menyajikan pola kelulusan siswa.
2. Melakukan analisis data kelulusan siswa angkatan 2020 agar dapat diketahui faktor-faktor yang sangat berpengaruh pada pola kelulusan.

1.5 Manfaat

Adapun manfaat yang akan didapat dengan dilakukannya penelitian ini adalah:

1. Dapat mempermudah pihak sekolah dalam menentukan keputusan yang akan diambil. Keputusan berupa strategi apa yang akan diambil untuk meningkatkan kualitas lulusan/alumni.
2. Dapat mengefektifkan keputusan yang diambil, karena analisis dilakukan berdasarkan data, bukan berdasarkan insting atau perasaan.

1.6 Sistematika Penulisan

Dalam penulisan laporan ini terdapat lima (5) BAB dengan sistematika penulisan sebagai berikut:

BAB 1 PENDAHULUAN

Dalam pendahuluan tercantum antara lain latar belakang, ruang lingkup masalah, rumusan masalah, tujuan, manfaat, dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini berisi teori-teori yang berkaitan dan mendukung penelitian serta penulisan skripsi yang dilakukan oleh penulis.

BAB III METODOLOGI PENELITIAN

Pada bab ini dijelaskan metodologi penelitian penyelesaian masalah yang dijelaskan dirumusan masalah yang meliputi metode pengumpulan data, metode pengembangan sistem, serta alat dan bahan pendukung.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas tentang hasil dari penelitian berupa pola kelulusan siswa beserta proses pengolahan datanya.

BAB V SIMPULAN DAN SARAN

Pada bab ini berisi simpulan dari seluruh pembahasan dan saran yang diperlukan untuk perbaikan dimasa yang akan datang.

LAMPIRAN

Berisi lampiran yang diperlukan.

BAB II

TINJAUAN PUSTAKA

2.1 Perkembangan Data Mining

Istilah data mining mulai populer di komunitas pengguna basis data pada tahun 1990-an. Namun teori dan metode dasar dari data mining telah lahir jauh sebelum era 90. Data mining berasal dari berbagai disiplin ilmu, dua ilmu yang paling mendasar adalah statistika dan *machine learning*. Teori-teori statistika yang berakar dari teori matematika yang menitikberatkan pada pembentukan model. Model merupakan asumsi atau pendekatan struktur yang mendekati data sesungguhnya. Sementara itu, *machine learning* lebih mementingkan pengembangan algoritma.

Perkembangan awal data mining dimulai pada tahun 1763 ketika Thomas Bayes mempublikasikan Teorema Bayes. Teori ini sangat penting dalam data mining, karena memungkinkan estimasi suatu kejadian berdasarkan kejadian yang telah berlangsung. Pada tahun 1805 mulai berkembang teori regresi yang mempelajari hubungan antar variabel. Regresi menjadi salah satu alat penting dalam data mining.

Penggunaan komputer untuk mengolah data dalam jumlah besar dimulai ketika Alan Turing memperkenalkan ide mesin pengolah data yang bersifat universal pada tahun 1936. Tahun 1943 Warren McCulloch dan Walter Pitts menciptakan konsep dasar jaringan syaraf tiruan. Konsep jaringan syaraf tiruan berusaha meniru cara kerja otak manusia dalam mengingat pola. Pengembangan sistem basis data yang pesat mulai tahun 1970 memungkinkan manusia untuk menyimpan dan mengelola data berukuran besar. Perkembangan itu diikuti pula oleh perkembangan berbagai algoritma untuk pengolahan data misalnya algoritma genetika pada tahun 1975 dan *Support Vector Machines*(SVM) pada tahun 1992. (Adinugroho & Sari, 2018).

Kalau kita membahas tentang data mining, tentulah kita harus mengetahui terlebih dahulu definisi dari data mining. Secara umum data mining terbagi atas dua kata, yaitu:

1. Data yaitu kumpulan fakta yang terekam atau sebuah entitas yang tidak memiliki arti dan selama ini terabaikan.
2. *Mining* yaitu proses penambangan.

Sehingga data mining itu dapat diartikan sebagai proses penambangan data yang menghasilkan sebuah *output* berupa pengetahuan. (Nofriansyah & Nurcahyo, 2015)

Data Mining adalah suatu istilah yang dapat digunakan untuk menguraikan penemuan pengetahuan dalam *database*. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Data mining didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau sering semi otomatis. Pola yang ditemukan harus berarti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. (Susilo, 2018)

Data mining mampu membantu perusahaan untuk mendapatkan pola dari data yang tersimpan didalam basis data perusahaan. Pengetahuan yang diperoleh tersebut akan menjadi pedoman dalam mengambil tindakan-tindakan bisnis sebagai upaya pemeliharaan dan peningkatan fungsi-fungsi berikut: deskripsi, estimasi, prediksi, klasifikasi, *clustering*, dan asosiasi. (Modyta, 2019).

Istilah data mining dan *Knowledge Discovery in Database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Salah satu tahapan dalam proses KDD adalah data mining.

Berikut adalah tahapan dalam data mining.

1. Pembersihan data (*data cleansing*)
Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau tidak relevan.
2. Integrasi Data (*data integration*)
Integrasi data merupakan penggabungan data dari berbagai *database* kedalam satu *database* baru.
3. Seleksi Data (*data selection*)
Data yang ada dalam *database* seringkali semuanya tidak dipakai, oleh karena itu hanya data yang sesuai untuk proses analisis yang akan diambil dari *database*.
4. Transformasi Data (*data transformation*)
Dalam tahap ini data akan diubah atau digabung kedalam format yang sesuai untuk proses data mining.
5. Proses Mining
Tahapan ini merupakan proses utama dimana metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. Evaluasi Pola (*pattern evaluation*)
Pola-pola yang dihasilkan akan dievaluasi, yaitu dipilih pola-pola yang menarik.
7. Presentasi pengetahuan (*knowledge presentation*)
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.2 Pengelompokan Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari data untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menentukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali *variable* target estimasi lebih kearah numerik dari pada kearah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang. Contoh prediksi bisnis dan penelitian:

- a. Prediksi harga beras dalam tiga bulan yang akan datang.
- b. Prediksi presentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Contoh:

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau tidak.
- b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- c. Mendiagnosis penyakit seorang pasien untuk mendapatkan kategori penyakit.

5. *Clustering*

Clustering merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam *cluster* lain. *Clustering* berbeda dengan klasifikasi, yaitu tidak adanya variabel target dalam *clustering*. *Clustering* tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Algoritma *clustering* mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan, yang mana kemiripan dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal. Contoh:

- a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku *financial* yang baik dan yang mencurigakan.
- c. Melakukan *clustering* terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.

6. Asosiasi

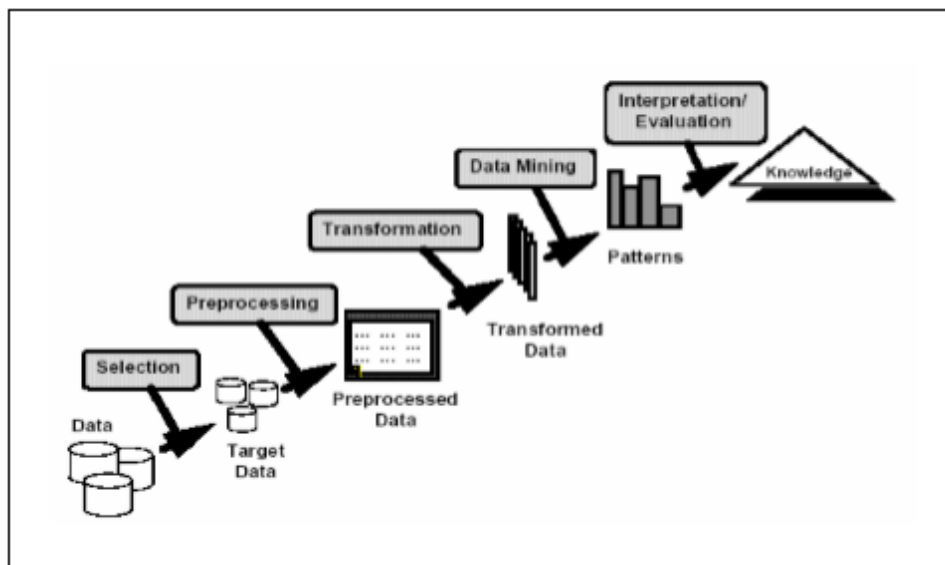
Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh:

- a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran *upgrade* layanan yang diberikan.
- b. Menentukan barang dalam supermarket yang dibeli secara bersamaan dan yang tidak pernah dibeli secara bersamaan.

2.3 KDD (*Knowledge Discovery in Database*)

Metode yang digunakan adalah *Knowledge Discovery in Database* (KDD). *Knowledge Discovery in Database* adalah proses ekstraksi informasi yang potensial, implisit, dan tidak dikenal dalam *database*. Proses *Knowledge Discovery* melibatkan proses data mining didalamnya, kemudian mengubah hasilnya menjadi informasi yang disederhanakan sehingga mudah dipahami. KDD diartikan sebagai rangkaian proses untuk mencari dan mengidentifikasi pola (*pattern*) dalam data. Pola yang ditemukan bersifat sah, baru, dan dapat dimengerti.

Proses KDD (*Knowledge Discovery in Database*) secara garis besar dapat dijelaskan sebagai berikut (Kusrini, 2009).



Gambar 3. 1 Tahapan *Knowledge Discovery in Database*

1) Pemilihan Data (*Data Selection*)

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

2) Pembersihan data (*data cleaning*)

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleansing pada data yang mkenjadi fokus *Knowledge Discovery in Database*. Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak juga dilakukan proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD.

3) *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih sehingga data tersebut sesuai untuk proses data mining.

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 tergantung pada tujuan dan proses KDD secara keseluruhan.

5) *Interpretation* atau *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*.

2.4 Association Rule

Association dalam data mining merupakan pekerjaan untuk menentukan atribut yang akan didapatkan bersamaan. Tugas dari *association* adalah mencari aturan yang tidak mengcover untuk mengukur hubungan antara dua atau lebih atribut. *Association Rule* adalah bentuk jika kejadian sebelumnya kemudian konsekuensinya (*If antecedent, then consequent*). Bersamaan dengan perhitungan aturan *support* dan *confidence*. *Association rule mining* adalah teknik data mining untuk menemukan asosiasi antara suatu kombinasi item. (Maharani, et al., 2017)

Association Rule (aturan asosiasi) atau *affinity analysis* (analisis afinita) berkenaan dengan studi tentang “apa bersama apa”. Asosiasi juga sering dinamakan *market basket analysis* (analisis keranjang belanja), aturan asosiasi ingin memberikan informasi dalam bentuk hubungan “*if-then*” atau “jika-maka”. (Susilo, 2018)

Analisis asosiasi dikenal sebagai salah satu teknik data mining yang menjadi dasar dari salah satu teknik data mining lainnya. Salah satu tahap analisis asosiasi yang menarik banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (*frequent pattern mining*). (Fauzy, Saleh W, & Asror, 2016)

Penting tidaknya suatu asosiasi dapat diketahui dengan dua tolok ukur, yaitu *support* dan *confidence*. *Support* (nilai penunjang) adalah presentase kombinasi item dalam *database*. Sedangkan *confidence* (nilai kepastian) merupakan kuatnya hubungan antar item dalam aturan asosiasi.

2.5 Algoritma Apriori

Association rule mempunyai beberapa algoritma yang digunakan untuk menemukan aturan kombinasi dari dua item atau lebih dalam *database*. Dalam penelitian ini, penulis menggunakan algoritma apriori. Algoritma apriori

merupakan salah satu algoritma yang bisa dipakai untuk menemukan aturan asosiasi dalam data mining.

Algoritma apriori adalah suatu algoritma dasar yang diusulkan oleh Agrawal & Srikant pada tahun 1994 untuk menentukan *frequent itemsets* untuk aturan asosiasi *boolean*. Algoritma apriori termasuk jenis aturan asosiasi pada data mining. Bentuk algoritma apriori dapat dituliskan sebagai berikut :

```

L1 = {frequent itemset with one element}
for (k = 2; Lk-1 ≠ ∅; k++)
{
Ck = apriori - gen(Lk-1);
for all transactions t
{
C't = subset(Ckt);
for all candidates c ∈ C't do c.count++;
}
Lk = {c ∈ C't | c.count ≥ minsup}
}
return ∪k Lk;

```

Keterangan:

L : Himpunan *frequent itemset minimum support*

C : Himpunan kandidat *itemset*

c : Kandidat *itemset*

t : Transaksi

Algoritma apriori banyak digunakan untuk menganalisis data transaksi atau bisa disebut *market basket*, misalnya sebuah swalayan memiliki *market basket*, dengan adanya algoritma apriori, pemilik swalayan dapat mengetahui pola pembelian seorang konsumen, jika seorang konsumen membeli item A, B, punya kemungkinan 50% dia akan membeli item C. Pola ini sangat signifikan dengan adanya data transaksi selama ini. (Modyta, 2019).

Prinsip dari algoritma apriori adalah kumpulkan jumlah item tunggal, dapatkan item besar. Dapatkan *candidate pairs*, hitung => *large triplets* dari item-item dan seterusnya. Sebagai petunjuk: Setiap subset dari sebuah *frequent itemset* harus menjadi *frequent*. (Febrianti & Suryadi, 2018).

Kemudian dalam analisa pola frekuensi tinggi dengan algoritma apriori dalam tahapan ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam basis data. Nilai *support* sebuah item diperoleh dengan menggunakan Persamaan (2.1).

$$Support(A) = \frac{\text{Jumlah transaksi mengandung } A}{\text{Total transaksi}} \quad (2.1)$$

Sementara, nilai *support* dari 2 item diperoleh dengan menggunakan Persamaan (2.2).

$$Support(A, B) = P(A \cap B) = \frac{\Sigma \text{Transaksi mengandung } A \text{ dan } B}{\Sigma \text{Transaksi}} \quad (2.2)$$

Support menunjukkan frekuensi kemunculan suatu *item* dalam *dataset*. *Support* dari satu *itemset* disebut juga analisis pola frekuensi tinggi (*frequent pattern mining*). *Itemset-itemset* yang memenuhi nilai *minimum support* yang telah ditentukan selanjutnya akan dilakukan kombinasi dua *item* dengan *itemset* yang lain untuk dicari frekuensi kemunculannya dalam *dataset*. Kombinasi dua item yang memenuhi *minimum support* akan dilakukan kombinasi 3 *item* dengan *itemset* yang lain. Kombinasi akan terus dilakukan sampai tidak ada lagi kombinasi yang memenuhi *minimum support*.

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat *minimum* untuk *confidence* dengan menghitung *confidence* aturan asosiatif A & B. Nilai *confidence* diperoleh dengan Persamaan (2.3).

$$Confidence = P(B|A) = \frac{\Sigma \text{Transaksi mengandung } A \text{ dan } B}{\Sigma \text{Transaksi mengandung } A} \quad (2.3)$$

Untuk menentukan aturan asosiasi yang akan dipilih, maka harus diurutkan berdasarkan *Support x Confidence*. Aturan diambil sebanyak n aturan yang memiliki hasil terbesar. (Yanto & Khoiriah, 2015)

Confidence merupakan nilai kepastian, yaitu nilai yang menunjukkan kuat atau tidaknya suatu aturan dalam *dataset*. Kombinasi yang memenuhi *minimum Confidence* yang telah ditentukan akan digunakan untuk membuat sebuah aturan asosiasi.

Rules terdiri dari *antecedent* dan *consequent*. *Antecedent* adalah bentuk kondisi dari pada *rules* atau merupakan variabel yang menjadi penyebab suatu *rules*. Sedangkan *consequent* adalah bentuk pernyataan dari pada *rules* yang merupakan akibat dari penyebab atau *antecedent*. Sedangkan *lift* menunjukkan tingkat kekuatan *rules* kejadian acak dari *antecedent* dan *consequent* berdasarkan support masing-masing. (Sikumbang, 2018)

2.6 Rapidminer

Rapidminer adalah sebuah *tools* yang digunakan dalam teknik yang berada dilingkungan *machine learning*, data mining, *text mining*, dan *predictive analytics*.

Rapidminer merupakan perangkat lunak yang digunakan dalam pengolahan data. Dengan menggunakan prinsip dan algoritma data mining. Rapidminer mengekstrak pola-pola dari *dataset* yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan, dan *database*. Rapidminer memudahkan penggunaanya dalam melakukan perhitungan data yang sangat banyak menggunakan operator-operator. Operator ini berfungsi untuk memodifikasi data. Data dihubungkan dengan *node-node* pada operator kemudian kita hanya tinggal melihat hasilnya. Hasil yang diperlihatkan dapat ditampilkan secara visual menggunakan grafik. Hal ini yang menjadikan rapidminer sebagai *software*

pilihan untuk melakukan ekstraksi data dengan metode-metode data mining. (Novianti, 2019).

2.7 Penelitian Sebelumnya Tentang Algoritma Apriori dan Pencarian Pola Kelulusan Siswa

Penelitian sebelumnya dilakukan oleh Robi Yanto dan Riri Khoiriyah yaitu mencari pola pembelian obat menggunakan metode algoritma apriori. Penelitian tersebut menghasilkan dua kombinasi *itemset* tentang kecenderungan obat yang sering dibeli konsumen. Pengetahuan tersebut dapat membantu pihak karyawan dalam mengatur tata letak obat yang terdiri dari 2 item secara berdekatan. (Yanto & Khoiriah, 2015)

Penelitian lainnya juga dilakukan oleh Maharani dan teman-temannya tentang pencarian pola untuk pengaturan *layout* produk di minimarket dengan menerapkan *association rule*. Penelitian tersebut menghasilkan 8 *rule* yang telah memenuhi *support* dan *confidence*. Kombinasi tersebut dapat digunakan untuk membantu pengaturan *layout* produk di minimarket tersebut. (Maharani, et al., 2017)

Penelitian lainnya juga dilakukan oleh Andri Anto Tri Susilo tentang pengolahan data transaksi minimarket. Penelitian ini menghasilkan 6 *rule* yang memenuhi *minimum support* dan *confidence*. *Rule* ini dapat membantu dalam pembuatan katalog produk yang akan dijual, karena aturan asosiasi menggambarkan hubungan antar *item* produk dalam sebuah kombinasi *itemset*. (Susilo, 2018)

Penelitian lainnya dilakukan oleh Dicky Nofriansyah dan teman-temannya tentang analisa data penjualan untuk mendapatkan pola rekomendasi penjualan menggunakan algoritma apriori. Penelitian ini menghasilkan 7 aturan asosiasi yang dapat digunakan untuk memberikan rekomendasi penjualan. (Nofriansyah, Yetri, Erwansyah, & Suharsil, 2019)

Penelitian sebelumnya juga dilakukan oleh Eka Lia Febrianti dan Agus Suryadi tentang analisis pola pembelian konsumen. Pada penelitian tersebut, penerapan algoritma apriori pada teknik data mining sangat efisien dan dapat mempercepat proses pembentukan pola kombinasi *itemset*. Pola kombinasi *itemset* ini sangat berguna untuk mempersiapkan stok jenis barang yang diperlukan. Analisis ini sangat membantu perusahaan menemukan informasi yang sangat penting dari gudang data yang semakin menumpuk. (Febrianti & Suryadi, 2018)

Penelitian sebelumnya juga dilakukan oleh Erma Delima Sikumbang tentang penerapan data mining dalam penjualan sepatu menggunakan algoritma apriori. Setelah dilakukan analisis, didapatkan informasi berupa jenis sepatu yang paling diminati oleh konsumen. Informasi ini dapat membantu pengelola toko untuk menyiapkan stok jenis sepatu apa saja yang diperlukan kemudian hari. (Sikumbang, 2018)

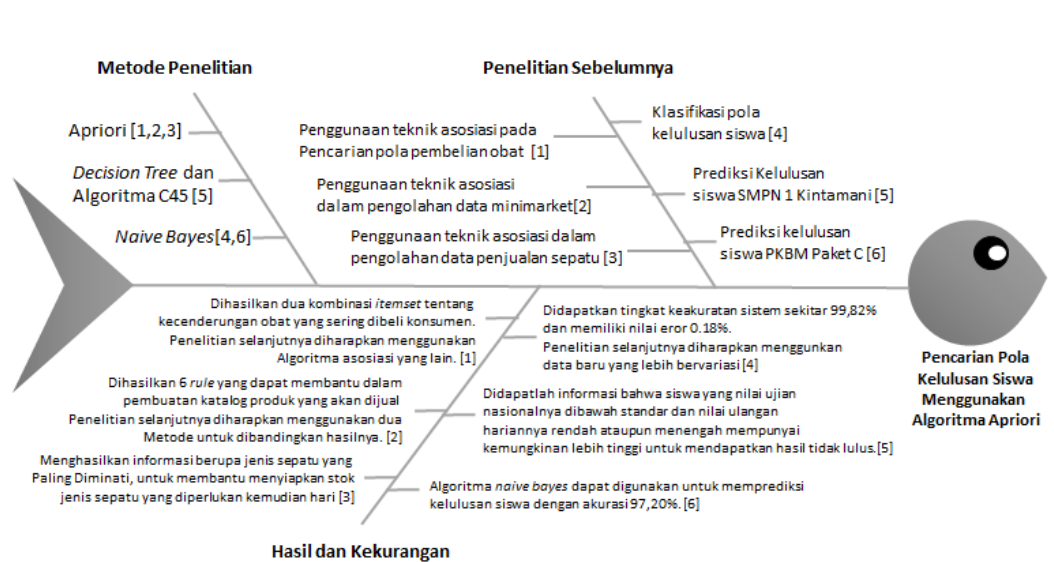
Penelitian lainnya dilakukan oleh Mohamad Fauzy dan teman-temannya tentang penggunaan *association rule* untuk prediksi hujan di Wilayah Kota Bandung. Penelitian ini menggunakan algoritma apriori dan bertujuan mencari hubungan kesamaan antar *item* dalam prediksi hujan. Hasilnya algoritma apriori dapat digunakan untuk prediksi hujan. Semua *rule* yang dihasilkan dalam penelitian ini memiliki nilai *lift ratio* lebih dari 1.00 sehingga dapat digunakan sebagai acuan untuk memprediksi hujan. (Fauzy, Saleh W, & Asror, 2016)

Penelitian sebelumnya dilakukan oleh Budanis Dwi Meilani dan Nofi Susanti tentang pencarian pola kelulusan siswa. Penelitian ini menerapkan teknik klasifikasi dengan algoritma *naive bayes*. Hasil penelitian ini didapatkan tingkat keakuratan sistem sekitar 99,82% dan memiliki nilai eror 0.18%. (Meilani & Susanti, 2015)

Penelitian lainnya dilakukan oleh Putu dan teman-temannya tentang prediksi tingkat kelulusan siswa menggunakan metode *decision tree* di SMPN 1

Kintamani. Pada penelitian ini, digunakan kriteria nilai ujian harian, nilai UTS, nilai UAS, dan nilai UN. Penelitian ini menghasilkan sebuah pohon keputusan beserta tingkat kelulusan masing-masing siswa. Dari hasil pengolahan data, didapatlah informasi bahwa siswa yang nilai ujian nasionalnya dibawah standar dan nilai ulangan hariannya rendah ataupun menengah mempunyai kemungkinan lebih tinggi untuk mendapatkan hasil tidak lulus. (Nugraha, Aribawa, Priyana, & Indrawan, 2016)

Penelitian sebelumnya juga dilakukan oleh Jaya Sumpena dan Nina Kurnia tentang analisis prediksi kelulusan siswa PKBM paket C dengan algoritma *naive bayes*. Pada penelitian menerapkan teknik prediksi data mining dengan menggunakan data nilai sebagai bahan untuk dilakukan analisa. Hasilnya, algoritma *naive bayes* dapat digunakan untuk memprediksi kelulusan siswa dengan akurasi 97,20%. (Sumpena & Kurnia, 2019)



Gambar 2.1 Fishbone Diagram

Keterangan:

[1] Implementasi Data Mining dengan Metode Algoritma Apriori dalam menentukan Pola Pembelian Obat oleh Robi Yanto dan Riri Khoiriyah.

[2] Penerapan Algoritma Apriori pada Pengolahan Data Transaksi Penjualan di Minimarket Priyo Kota Lubuklinggau oleh Andri Anto Tri Susilo.

[3] Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori oleh Erma Delima Sikumbang.

[4] Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode Naive Bayes oleh Budanis Dwi Meilani dan Nofi Susanti.

[5] Penerapan Metode Decision Tree (Data Mining) Untuk Memprediksi Tingkat Kelulusan Siswa SMPN 1 Kintamani oleh Putu Gede Surya Cipta Negara dan teman-temannya.

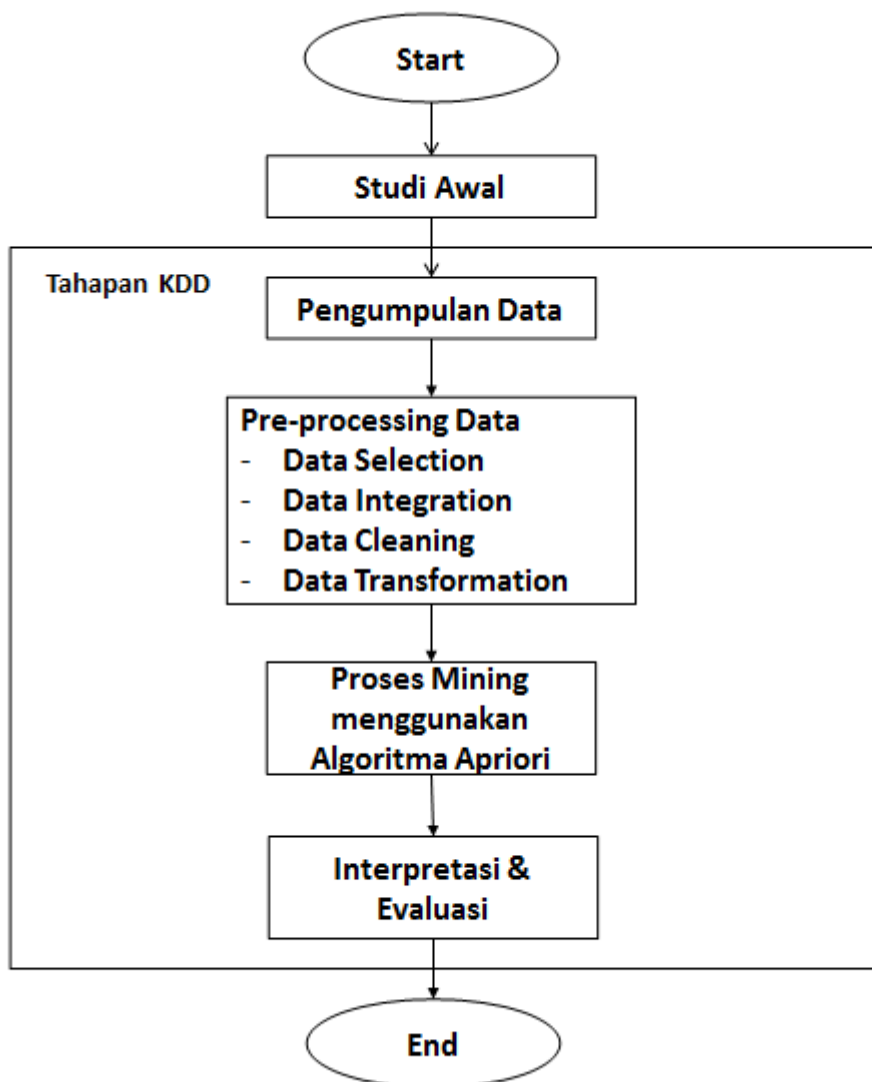
[6] Analisis Prediksi Kelulusan Siswa PKBM Paket C Dengan Metoda Algoritma Naive Bayes oleh Jaya Sumpena dan Nina Kurnia.

Berdasarkan penelitian sebelumnya di atas, dan berdasarkan pemetaan masalah dari beberapa penelitian menggunakan *fishbone diagram*, topik yang menarik untuk dibahas yaitu kombinasi dari fungsi asosiasi data mining dengan algoritma apriori untuk mencari pola kelulusan siswa. Sehingga diharapkan dapat diketahui hubungan asosiasi antara *item-item* dalam data yang dianalisis.

BAB III METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Tahapan penelitian ini digambarkan dalam sebuah diagram alir seperti berikut ini:



Gambar 3. 2 Tahapan Penelitian

3.1.1 Studi Awal

Pada tahap studi awal penulis mempelajari masalah yang akan dijadikan topik penelitian. Setelah didapatkan topik penelitian, penulis menentukan ruang lingkup masalah, latar belakang masalah, dan mempelajari beberapa literatur yang berkaitan dengan penelitian untuk memecahkan masalah tersebut.

3.1.2 Pengumpulan Data

Pengumpulan data dilakukan dengan cara wawancara dan studi pustaka. Penulis melakukan tanya jawab untuk menggali informasi mengenai asal data, kemudian penulis meminta data siswa kepada bagian administrasi SMKN 1 Bukit Kemuning. Data yang dikumpulkan adalah data siswa yang lulus pada tahun 2020. Penulis diberi dua file dalam format Microsoft Excel. File pertama adalah file data peserta didik, dan file kedua berisi nilai ujian sekolah peserta didik SMKN 1 Bukit Kemuning lulusan tahun 2020

3.1.3 Pre-processing Data

Dalam data mining, tahap *pre-processing* merupakan proses yang paling banyak menghabiskan waktu yaitu sebanyak 60% dari semua proses. Artinya tahap *pre-processing* memegang peranan penting dalam proses data mining.

Tahap pertama, akan dilakukan proses pemilihan data. Penulis memilih atribut yang akan digunakan, dari 9 atribut data diantaranya Nama, NISN, Kelas, Kecamatan, Pekerjaan Ayah, Kuliah, Bekerja, Menganggur, dan Rata-rata hanya dipilih 6 atribut yang akan digunakan yaitu atribut Kecamatan, Pekerjaan Ayah, Kuliah, Bekerja, Menganggur, dan Rata-rata.

Setelah didapatkan atribut yang terpilih, selanjutnya dilakukan proses integrasi data. Dalam penelitian ini, penulis melakukan penggabungan data menggunakan Microsoft Excel dikarenakan adanya ketidaksamaan urutan nama siswa dalam kedua file data yang akan digunakan. Hal ini akan jadi masalah tentunya jika data

yang diolah sangat banyak. Setelah itu penulis memberikan kategori pada atribut rata-rata. Kategori yang dimaksud yaitu sangat memuaskan, memuaskan, dan tidak memuaskan. Kemudian atribut bekerja, kuliah dan menganggur dijadikan satu atribut status setelah kuliah.

Setelah diintegrasikan, maka selanjutnya dilakukan pembersihan data. Pada proses ini dilakukan pembersihan data berupa menghilangkan data yang tidak lengkap, data yang redundan, dan data yang kosong. Penulis membersihkan data dengan operator Replace missing value di Rapidminer. Operator ini melengkapi data yang kosong dengan rata-rata data yang ada dalam satu atribut.

Kemudian penulis mentransformasikan data kedalam format tabular, yaitu format yang bisa diolah oleh Rapidminer. Contoh data dalam bentuk tabular bisa dilihat di tabel 3.1

Tabel 3. 1 Format Data Tabular

NO	Abung Tinggi	Abung Barat	Banjit	Baradatu	Bekasi Barat	Bukit Kemuning	Bukit Kerman	Gunung Labuhan
1	1	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1
5	0	0	0	0	0	1	0	0
6	0	0	0	0	0	0	0	1
7	0	0	0	0	0	0	0	1
8	0	0	0	0	0	0	0	1
9	0	0	0	0	0	1	0	0
10	0	0	0	0	0	0	0	1
11	1	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	1
13	0	0	0	0	0	0	0	1
14	1	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	1
16	0	0	0	0	0	1	0	0
17	0	0	0	0	0	1	0	0
18	0	0	0	0	0	0	0	1
19	0	0	0	0	0	0	0	1
20	1	0	0	0	0	0	0	0

3.1.4 Proses Mining Menggunakan Algoritma Apriori

Pada tahapan ini dilakukan proses mining, yaitu dengan menerapkan model yang akan dipakai, dimana dalam penelitian ini digunakan algoritma apriori untuk menganalisis hubungan asosiasi dari beberapa variabel. Analisis ini dilakukan menggunakan tool rapidminer. Setelah model diaplikasikan, maka dihasilkan suatu aturan asosiasi.

3.1.5 Interpretation and Evaluation

Setelah dianalisis, didapatkan hasil berupa *rules* yang menunjukkan hubungan asosiasi dalam data. *Rules* tersebut lalu diinterpretasikan atau dievaluasi. *Rules* diinterpretasikan dengan cara dipresentasikan dengan bahasa yang mudah dipahami. Pada tahap ini dilakukan Presentasi dari pola yang dihasilkan.

3.2 Alat dan Bahan

3.2.1 Alat

Alat yang digunakan dalam penelitian ini dibagi menjadi dua bagian, yaitu *hardware* dan *software*.

Adapun *hardware* yang digunakan dalam penelitian ini adalah satu buah laptop dengan spesifikasi sebagai berikut:

- Processor 2.00 Ghz
- RAM 4 GB

Software yang digunakan dalam penelitian ini yaitu:

- Windows 10 64 Bit
- Rapidminer 9.8.1
- Microsoft Excel 2010

3.2.2 Bahan

Bahan yang digunakan dalam penelitian ini adalah data kelulusan siswa tahun 2020 yang didapat dari bagian administrasi SMKN 1 Bukit Kemuning.

3.2.2.1 Data Mentah

Data mentah berupa daftar peserta didik lulusan tahun 2020 yang berisi data identitas siswa, nilai siswa, dan status siswa setelah lulus. Data yang dicatat berupa Nama, NISN, Asal, Pekerjaan Orang Tua, dan Status setelah lulus. Atribut dapat dilihat dalam Tabel 3.1 berikut ini.

NO	Atribut	Keterangan
1	Nama	Merupakan nama lengkap siswa bersangkutan.
2	NISN	Nomor Induk Siswa Nasional merupakan kode unik yang dimiliki siswa sebagai nomor identitas nasional
3	Kelas	Merupakan Kelas siswa yang bersangkutan
4	Kecamatan	Merupakan nama Kecamatan wilayah tempat tinggal siswa yang bersangkutan
5	Pekerjaan Ayah	Merupakan pekerjaan ayah siswa yang bersangkutan.
6	Kuliah	Merupakan perguruan tinggi tempat melanjutkan studi siswa yang bersangkutan.
7	Bekerja	Merupakan tempat bekerja siswa yang bersangkutan setelah lulus dari SMKN 1 Bukit Kemuning.
8	Menganggur	Merupakan status siswa yang bersangkutan setelah lulus dari SMKN 1 Bukit kemuning.
9	Rata-rata	Merupakan nilai rata-rata dari jumlah keseluruhan yang diperoleh siswa yang

NO	Atribut	Keterangan
		bersangkutan.

3.2.2.2 Atribut Yang Akan Digunakan

Dalam penelitian ini akan digunakan beberapa atribut dari data-data tersebut. Tidak semua atribut dapat menunjang kegiatan analisis, maka dilakukan pemilihan dan pembersihan data untuk menghilangkan atribut yang tidak digunakan dalam data mentah. Atribut yang digunakan diantaranya adalah sebagai berikut:

1. Kecamatan, yang merupakan asal siswa yang bersangkutan.
2. Pekerjaan Ayah, yaitu atribut yang menunjukkan pekerjaan orang tua siswa yaitu Ayah siswa.
3. Rata-rata, yaitu atribut yang menunjukkan nilai rata-rata ujian sekolah yang didapatkan siswa yang bersangkutan. Atribut ini akan diklasifikasikan menjadi memuaskan dan sangat memuaskan.
4. Bekerja, merupakan atribut yang menjelaskan status siswa yang bekerja setelah dinyatakan lulus.
5. Kuliah, merupakan atribut yang menjelaskan status siswa yang kuliah setelah dinyatakan lulus.
6. Menganggur, yaitu atribut yang menjelaskan status siswa yang menganggur setelah dinyatakan lulus.

3.2.2.3 Transformasi Data

Transformasi data merupakan proses perubahan data kedalam format yang dibutuhkan untuk diproses dalam data mining tanpa mengubah nilai data. Transformasi dilakukan karena data yang akan digunakan mempunyai format yang berbeda, sehingga tidak bisa langsung dilakukan analisis. Atribut yang akan ditransformasikan yaitu atribut Rata-rata, Bekerja, Kuliah, dan Menganggur.

Atribut rata-rata akan ditransformasikan kedalam bentuk nominal dengan tiga kategori, yaitu kategori tidak memuaskan, cukup memuaskan, memuaskan, dan sangat memuaskan. Penentuan kategori menggunakan Kriteria Kelulusan Minimal (KKM) yang dimiliki Sekolah SMKN 1 Bukit Kemuning. Dalam penelitian ini penulis mentransformasikan data menggunakan Microsoft Excel dengan rumus IF bersarang yaitu :

$=IF(AND(BG2 \geq 90, BG2 \leq 100), "Sangat Memuaskan", IF(AND(BG2 \geq 80, BG2 < 90), "Memuaskan", IF(AND(BG2 \geq 70, BG2 < 80), "Cukup Memuaskan", "Tidak Memuaskan"))))$.

Transformasi data dapat dilihat pada Tabel 3.2 dibawah ini.

Rata-rata	Kategori
≥ 90	Sangat Memuaskan
80 - <90	Memuaskan
70 - <80	Cukup Memuaskan
0- <70	Tidak Memuaskan

Atribut Bekerja, Kuliah, dan Menganggur akan ditransformasikan menjadi satu atribut yaitu Status Setelah Lulus. Hal ini dikarenakan alumni hanya bisa memilih salah satu diantara ketiga atribut di atas sehingga tidak ada alumni yang memilih sedang bekerja dan sedang kuliah dalam waktu bersamaan. Atribut Bekerja, Kuliah, dan Menganggur akan ditransformasikan menggunakan Microsoft Excel. Transformasi data dapat dilihat di Tabel 3.3 berikut ini.

Nama Atribut	Status Setelah Lulus
Bekerja	Bekerja
Kuliah	Kuliah
Menganggur	Menganggur

3.3 Metodologi Pengumpulan Data

Dalam penelitian ini dibutuhkan data yang relevan sebagai bahan yang diteliti. Maka dari itu, dilakukan pengumpulan data untuk mendapatkan informasi yang diperlukan. Adapun metode pengumpulan data dalam penelitian ini dapat dijelaskan sebagai berikut:

a. Wawancara

Data dikumpulkan dengan cara mengadakan tanya jawab langsung kepada pihak sekolah. Tujuan dari wawancara adalah untuk memperoleh informasi yang lebih akurat dan lengkap. Dalam penelitian ini dilakukan wawancara dengan petugas bagian admin di SMKN 1 Bukit Kemuning.

b. Studi Pustaka

Informasi didapat dengan cara mengumpulkan dan mengkaji data kelulusan siswa SMKN 1 Bukit Kemuning lulusan tahun 2020.

BAB IV HASIL DAN PEMBAHASAN

4.1 Analisis Menggunakan Algoritma Apriori

Dalam penghitungan manual ini digunakan 20 *dataset* sebagai sampel uji coba, dengan *minimum support* 20% dan *minimum confidence* 60%. *Dataset* dapat dilihat pada tabel 4.1.

Tabel 4. 1 Dataset

No	Kecamatan	Pekerjaan Orang Tua	Kategori Kelulusan	Status Setelah Lulus
1	Abung Tinggi	Petani	Cukup Memuaskan	Menganggur
2	Abung Barat	Petani	Cukup Memuaskan	Bekerja
3	Abung Tinggi	Petani	Memuaskan	Bekerja
4	Gunung Labuan	Petani	Memuaskan	Bekerja
5	Bukit Kemuning	Petani	Memuaskan	Menganggur
6	Gunung Labuan	Buruh	Cukup Memuaskan	Menganggur
7	Gunung Labuan	Petani	Cukup Memuaskan	Menganggur
8	Gunung Labuan	Wiraswasta	Cukup Memuaskan	Menganggur
9	Bukit Kemuning	Petani	Memuaskan	Kuliah
10	Gunung Labuan	Petani	Cukup Memuaskan	Kuliah
11	Abung Tinggi	Petani	Memuaskan	Menganggur
12	Gunung Labuan	Petani	Cukup Memuaskan	Menganggur
13	Gunung Labuan	Petani	Cukup Memuaskan	Kuliah
14	Abung Tinggi	Petani	Cukup Memuaskan	Menganggur
15	Gunung Labuan	Petani	Memuaskan	Menganggur
16	Bukit Kemuning	Petani	Memuaskan	Menganggur
17	Bukit Kemuning	Petani	Memuaskan	Kuliah
18	Gunung Labuan	Wiraswasta	Cukup Memuaskan	Kuliah
19	Gunung Labuan	Petani	Cukup Memuaskan	Menganggur
20	Abung Tinggi	Petani	Cukup Memuaskan	Menganggur

Proses selanjutnya akan dihitung *Support* dari setiap *itemset* menggunakan persamaan 2.1 yaitu: $Support (A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ transaksi}$

Berikut ini adalah perhitungan *support* dari setiap *itemset* menggunakan persamaan 2.1.

$$Support (Abung Tinggi) = \frac{5}{20} = 0,25 = 25\%$$

$$Support (Abung Barat) = \frac{1}{20} = 0,05 = 5\%$$

$$Support (Gunung Labuan) = \frac{10}{20} = 0,5 = 50\%$$

$$Support (Bukit Kemuning) = \frac{4}{20} = 0,2 = 20\%$$

$$Support (Petani) = \frac{17}{20} = 0,85 = 85\%$$

$$Support (Buruh) = \frac{1}{20} = 0,05 = 5\%$$

$$Support (Wiraswasta) = \frac{2}{20} = 0,1 = 10\%$$

$$Support (Cukup Memuaskan) = \frac{12}{20} = 0,6 = 60\%$$

$$Support (Memuaskan) = \frac{8}{20} = 0,4 = 40\%$$

$$Support (Menganggur) = \frac{12}{20} = 0,6 = 60\%$$

$$Support (Bekerja) = \frac{3}{20} = 0,15 = 15\%$$

$$Support (Kuliah) = \frac{5}{20} = 0,25 = 25\%$$

Setelah semua *itemset* dihitung *support* dari setiap *itemset*, hasil perhitungan dari *itemset* yang memenuhi *minimum support* disajikan didalam Tabel 4.2 untuk memudahkan proses perhitungan selanjutnya.

Tabel 4. 2 Itemset yang memenuhi *minimum support* 20%

Itemset	Support
Abung Tinggi	25%
Gunung Labuan	50%
Bukit Kemuning	20%
Petani	85%
Cukup Memuaskan	60%

Memuaskan	40%
Menganggur	60%
Kuliah	25%

Berdasarkan Tabel 4.2 , didapatkan *itemset* yang memenuhi *minimum support* sebesar 20%. *Itemset* yang memenuhi *minimum support* diantaranya Abung Tinggi, Gunung Labuan, Bukit Kemuning, Petani, Cukup Memuaskan, Memuaskan, Menganggur, dan Kuliah. *Itemset* Petani menjadi item dengan nilai *support* tertinggi yaitu sebesar 85 %.

Selanjutnya dilakukan kombinasi 2 *itemset* dari *item-item* yang memenuhi *minimum support* menggunakan persamaan 2.2

$$Support (A, B) = P(A \cap B) = \frac{\Sigma Transaksi \text{ mengandung } A \text{ dan } B}{\Sigma Transaksi}$$

Berikut ini adalah perhitungan *support* dari 2 *itemset* menggunakan persamaan 2.2

$$Support (Abung Tinggi, Gunung Labuan) = \frac{0}{20} = 0,00 = 0\%$$

$$Support (Abung Tinggi, Bukit Kemuning) = \frac{0}{20} = 0,00 = 0\%$$

$$Support (Abung Tinggi, Petani) = \frac{5}{20} = 0,25 = 25\%$$

$$Support (Abung Tinggi, Cukup Memuaskan) = \frac{3}{20} = 0,15 = 15\%$$

$$Support (Abung Tinggi, Memuaskan) = \frac{2}{20} = 0,1 = 10\%$$

$$Support (Abung Tinggi, Menganggur) = \frac{4}{20} = 0,2 = 20\%$$

$$Support (Abung Tinggi, Kuliah) = \frac{0}{20} = 0,00 = 0\%$$

$$Support (Gunung Labuan, Bukit Kemuning) = \frac{0}{20} = 0,00 = 0\%$$

$$Support (Gunung Labuan, Petani) = \frac{7}{20} = 0,35 = 35\%$$

$$Support (Gunung Labuan, Cukup Memuaskan) = \frac{8}{20} = 0,4 = 40\%$$

$$Support (Gunung Labuan, Memuaskan) = \frac{2}{20} = 0,1 = 10\%$$

$$Support (Gunung Labuan, Menganggur) = \frac{6}{20} = 0,3 = 30\%$$

$$Support (Gunung Labuan, Kuliah) = \frac{3}{20} = 0,15 = 15\%$$

$$\text{Support (Bukit Kemuning, Petani)} = \frac{4}{20} = 0,2 = 20\%$$

$$\text{Support (Bukit Kemuning, Cukup Memuaskan)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Bukit Kemuning, Memuaskan)} = \frac{4}{20} = 0,2 = 20\%$$

$$\text{Support (Bukit Kemuning, Menganggur)} = \frac{2}{20} = 0,1 = 10\%$$

$$\text{Support (Bukit Kemuning, Kuliah)} = \frac{2}{20} = 0,1 = 10\%$$

$$\text{Support (Petani, Cukup Memuaskan)} = \frac{9}{20} = 0,45 = 45\%$$

$$\text{Support (Petani, Memuaskan)} = \frac{8}{20} = 0,40 = 40\%$$

$$\text{Support (Petani, Menganggur)} = \frac{10}{20} = 0,5 = 50\%$$

$$\text{Support (Petani, Kuliah)} = \frac{3}{20} = 0,15 = 15\%$$

$$\text{Support (Cukup Memuaskan, Memuaskan)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Cukup Memuaskan, Menganggur)} = \frac{8}{20} = 0,4 = 40\%$$

$$\text{Support (Cukup Memuaskan, Kuliah)} = \frac{3}{20} = 0,15 = 15\%$$

$$\text{Support (Memuaskan, Menganggur)} = \frac{4}{20} = 0,2 = 20\%$$

$$\text{Support (Memuaskan, Kuliah)} = \frac{2}{20} = 0,1 = 10\%$$

$$\text{Support (Menganggur, Kuliah)} = \frac{0}{20} = 0,00 = 0\%$$

Setelah dilakukan perhitungan *support* dari kombinasi 2 itemset, hasil perhitungan kombinasi itemset yang memenuhi *minimum support* sebesar 20% disajikan dalam Tabel 4.3 untuk memudahkan proses selanjutnya.

Tabel 4.3 Kombinasi 2 Itemset yang memenuhi *minimum support* 20%

Itemset	Support
Abung Tinggi, Petani	25%
Abung Tinggi, Menganggur	20%
Gunung Labuan, Petani	35%
Gunung Labuan, Cukup Memuaskan	40%
Gunung Labuan, Menganggur	30%
Bukit Kemuning, Petani	20%

Bukit Kemuning, Memuaskan	20%
Petani, Cukup Memuaskan	45%
Petani, Memuaskan	40%
Petani, Menganggur	50%
Cukup Memuaskan, Menganggur	40%
Memuaskan, Menganggur	20%

Berdasarkan Tabel 4.3 didapatkan 12 kombinasi yang memenuhi *minimum support* sebesar 20%. Kombinasi *item* Petani dan Menganggur mendapat nilai *support* paling tinggi sebesar 50 %.

Selanjutnya *item-item* yang memenuhi *minimum support* akan dilakukan kombinasi 3 *itemset* menggunakan persamaan 2.2. Berikut ini adalah perhitungan *support* dari kombinasi 3 *itemset*.

$$Support (\text{Abung Tinggi, Petani, Menganggur}) = \frac{4}{20} = 0,2 = 20\%$$

$$Support (\text{Abung Tinggi, Petani, Gunung Labuan}) = \frac{0}{20} = 0,00 = 0\%$$

$$Support (\text{Abung Tinggi, Petani, Cukup Memuaskan}) = \frac{3}{20} = 0,15 = 15\%$$

$$Support (\text{Abung Tinggi, Petani, Bukit Kemuning}) = \frac{0}{20} = 0,00 = 0\%$$

$$Support (\text{Abung Tinggi, Petani, Memuaskan}) = \frac{2}{20} = 0,1 = 10\%$$

$$Support (\text{Abung Tinggi, Menganggur, Gunung Labuan}) = \frac{0}{20} = 0,00 = 0\%$$

$$Support (\text{Abung Tinggi, Menganggur, Cukup Memuaskan}) = \frac{3}{20} = 0,15 = 15\%$$

$$Support (\text{Abung Tinggi, Menganggur, Bukit Kemuning}) = \frac{0}{20} = 0,00 = 0\%$$

$$Support (\text{Abung Tinggi, Menganggur, Memuaskan}) = \frac{1}{20} = 0,05 = 5\%$$

$$Support (\text{Gunung Labuan, Petani, Menganggur}) = \frac{4}{20} = 0,2 = 20\%$$

$$Support (\text{Gunung Labuan, Petani, Cukup Memuaskan}) = \frac{4}{20} = 0,2 = 20\%$$

$$Support (\text{Gunung Labuan, Petani, Bukit Kemuning}) = \frac{0}{20} = 0,00 = 0\%$$

$$Support (\text{Gunung Labuan, Petani, Memuaskan}) = \frac{2}{20} = 0,1 = 10\%$$

$$\text{Support (Gunung Labuan, Cukup Memuaskan, Abung Tinggi)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Gunung Labuan, Cukup Memuaskan, Menganggur)} = \frac{5}{20} = 0,25 = 25\%$$

$$\text{Support (Gunung Labuan, Cukup Memuaskan, Bukit Kemuning)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Gunung Labuan, Cukup Memuaskan, Memuaskan)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Gunung Labuan, Menganggur, Bukit Kemuning)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Gunung Labuan, Menganggur, Memuaskan)} = \frac{1}{20} = 0,05 = 5\%$$

$$\text{Support (Bukit Kemuning, Petani, Cukup Memuaskan)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Bukit Kemuning, Petani, Menganggur)} = \frac{2}{20} = 0,1 = 10\%$$

$$\text{Support (Bukit Kemuning, Petani, Memuaskan)} = \frac{4}{20} = 0,2 = 20\%$$

$$\text{Support (Bukit Kemuning, Memuaskan, Menganggur)} = \frac{2}{20} = 0,1 = 10\%$$

$$\text{Support (Bukit Kemuning, Memuaskan, Cukup Memuaskan)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Bukit Kemuning, Memuaskan, Gunung Labuan)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Petani, Cukup Memuaskan, Menganggur)} = \frac{6}{20} = 0,3 = 30\%$$

$$\text{Support (Petani, Cukup Memuaskan, Memuaskan)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Petani, Memuaskan, Menganggur)} = \frac{4}{20} = 0,2 = 20\%$$

$$\text{Support (Cukup Memuaskan, Menganggur, Memuaskan)} = \frac{0}{20} = 0,00 = 0\%$$

Setelah dilakukan perhitungan *support* dari kombinasi 3 *itemset*, hasil perhitungan yang memenuhi *minimum support* sebesar 20 % disajikan dalam Tabel 4.4 untuk memudahkan proses selanjutnya.

Tabel 4. 4 Kombinasi 3 *Itemset* yang memenuhi *minimum support* 20%

Itemset	Support
Abung Tinggi, Petani, Menganggur	20%
Gunung Labuan, Petani, Menganggur	20%
Gunung Labuan, Petani, Cukup Memuaskan	20%
Gunung Labuan, Cukup Memuaskan, Menganggur	25%

Bukit Kemuning, Petani, Memuaskan	20%
Petani, Cukup Memuaskan, Menganggur	30%
Petani, Memuaskan, Menganggur	20%

Berdasarkan Tabel 4.4 didapatkan 7 kombinasi yang memenuhi *minimum support* sebesar 20%. Kombinasi *item* Petani, Cukup Memuaskan, dan Menganggur mendapat nilai *support* paling tinggi sebesar 30 %.

Selanjutnya *item-item* yang memenuhi *minimum support* akan dilakukan kombinasi 4 *itemset* menggunakan persamaan 2.2. Berikut ini adalah perhitungan *support* dari kombinasi 4 *itemset*.

$$\text{Support (Abung Tinggi, Petani, Menganggur, Gunung Labuan)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Abung Tinggi, Petani, Menganggur, Cukup Memuaskan)} = \frac{3}{20} = 0,15 = 15\%$$

$$\text{Support (Abung Tinggi, Petani, Menganggur, Bukit Kemuning)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Abung Tinggi, Petani, Menganggur, Memuaskan)} = \frac{2}{20} = 0,1 = 10\%$$

$$\text{Support (Gunung Labuan, Petani, Menganggur, Cukup Memuaskan)} = \frac{3}{20} = 0,15 = 15\%$$

$$\text{Support (Gunung Labuan, Petani, Menganggur, Memuaskan)} = \frac{1}{20} = 0,05 = 5\%$$

$$\text{Support (Gunung Labuan, Petani, Menganggur, Bukit Kemuning)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Gunung Labuan, Petani, Cukup Memuaskan, Memuaskan)} = \frac{0}{20} = 0,00 = 0\%$$

$$\text{Support (Gunung Labuan, Petani, Cukup Memuaskan, Abung Tinggi)} = \frac{0}{20} = 0,00 = 0\%$$

$Support$ (Gunung Labuan, Petani, Cukup Memuaskan, Bukit Kemuning) =

$$\frac{0}{20} = 0,00 = 0\%$$

$Support$ (Gunung Labuan, Cukup Memuaskan, Menganggur, Abung Tinggi) =

$$\frac{0}{20} = 0,00 = 0\%$$

$Support$ (Gunung Labuan, Cukup Memuaskan, Menganggur, Bukit Kemuning) =

$$\frac{0}{20} = 0,00 = 0\%$$

$Support$ (Gunung Labuan, Cukup Memuaskan, Menganggur, Memuaskan) =

$$\frac{0}{20} = 0,00 = 0\%$$

$Support$ (Bukit Kemuning, Petani, Memuaskan, Menganggur) = $\frac{2}{20} = 0,1 = 10\%$

$Support$ (Bukit Kemuning, Petani, Memuaskan, Cukup Memuaskan) = $\frac{0}{20} =$

$$0,00 = 0\%$$

$Support$ (Bukit Kemuning, Petani, Memuaskan, Abung Tinggi) = $\frac{0}{20} = 0,00 =$

$$0\%$$

$Support$ (Bukit Kemuning, Petani, Memuaskan, Gunung Labuan) = $\frac{0}{20} = 0,00 =$

$$0\%$$

$Support$ (Petani, Cukup Memuaskan, Menganggur, Bukit Kemuning) = $\frac{0}{20} =$

$$0,00 = 0\%$$

$Support$ (Petani, Cukup Memuaskan, Menganggur, Memuaskan) = $\frac{0}{20} = 0,00 =$

$$0\%$$

Dari perhitungan di atas tidak ditemukan kombinasi 4 *itemset* yang memenuhi *minimum support* sebesar 20%, maka perhitungan *support* berhenti di kombinasi 3 *itemset*.

Setelah semua pola *item* yang memiliki frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat *minimum confidence* sebesar 60%. Nilai *confidence* dihitung menggunakan persamaan 2.3

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

Berikut ini adalah perhitungan *confidence* 2 itemset.

$$\text{Confidence (Abung Tinggi, Gunung Labuan)} = \frac{0}{5} = 0,00 = 0\%$$

$$\text{Confidence (Abung Tinggi, Bukit Kemuning)} = \frac{0}{5} = 0,00 = 0\%$$

$$\text{Confidence (Abung Tinggi, Petani)} = \frac{5}{5} = 1 = 100\%$$

$$\text{Confidence (Abung Tinggi, Menganggur)} = \frac{4}{5} = 0,8 = 80\%$$

$$\text{Confidence (Abung Tinggi, Cukup Memuaskan)} = \frac{3}{5} = 0,6 = 60\%$$

$$\text{Confidence (Abung Tinggi, Memuaskan)} = \frac{2}{5} = 0,4 = 40\%$$

$$\text{Confidence (Gunung Labuan, Bukit Kemuning)} = \frac{0}{10} = 0,00 = 0\%$$

$$\text{Confidence (Gunung Labuan, Petani)} = \frac{7}{10} = 0,7 = 70\%$$

$$\text{Confidence (Gunung Labuan, Menganggur)} = \frac{6}{10} = 0,6 = 60\%$$

$$\text{Confidence (Gunung Labuan, Cukup Memuaskan)} = \frac{8}{10} = 0,8 = 80\%$$

$$\text{Confidence (Gunung Labuan, Memuaskan)} = \frac{2}{10} = 0,2 = 20\%$$

$$\text{Confidence (Bukit Kemuning, Petani)} = \frac{4}{4} = 1 = 100\%$$

$$\text{Confidence (Bukit Kemuning, Menganggur)} = \frac{2}{4} = 0,5 = 50\%$$

$$\text{Confidence (Bukit Kemuning, Cukup Memuaskan)} = \frac{0}{4} = 0,00 = 0\%$$

$$\text{Confidence (Bukit Kemuning, Memuaskan)} = \frac{4}{4} = 1 = 100\%$$

$$\text{Confidence (Petani, Menganggur)} = \frac{10}{17} = 0,588 = 58,8\%$$

$$\text{Confidence (Petani, Cukup Memuaskan)} = \frac{9}{17} = 0,529 = 52,9\%$$

$$\text{Confidence (Petani, Memuaskan)} = \frac{8}{17} = 0,475 = 47,5\%$$

$$\text{Confidence (Menganggur, Cukup Memuaskan)} = \frac{8}{12} = 0,666 = 66,6\%$$

$$\text{Confidence (Menganggur, Memuaskan)} = \frac{4}{12} = 0,333 = 33,3\%$$

Setelah dilakukan perhitungan *confidence* dari 2 *itemset*, hasil perhitungan yang memenuhi *minimum confidence* sebesar 60 % disajikan dalam Tabel 4.5 untuk memudahkan proses selanjutnya.

Tabel 4. 5 Kombinasi 2 *itemset* yang memenuhi *minimum confidence* 60%

Itemset	Support	Confidence
Abung Tinggi, Petani	25%	100%
Abung Tinggi, Menganggur	20%	80%
Abung Tinggi, Cukup Memuaskan	15%	60%
Gunung Labuan, Petani	35%	70%
Gunung Labuan, Menganggur	30%	60%
Gunung Labuan, Cukup Memuaskan	40%	80%
Bukit Kemuning, Petani	20%	100%
Bukit Kemuning, Memuaskan	20%	100%
Menganggur, Cukup Memuaskan	40%	66%

Berdasarkan Tabel 4.5 didapatkan 9 kombinasi yang memenuhi *minimum confidence* sebesar 60%. Selanjutnya *item-item* yang memenuhi *minimum confidence* akan dilakukan perhitungan *confidence* kombinasi 3 *itemset* menggunakan persamaan 2.3. Berikut ini adalah perhitungan *support* dari kombinasi 3 *itemset*.

$$\text{Confidence (Abung Tinggi, Petani, Menganggur)} = \frac{4}{5} = 0,8 = 80\%$$

$$\text{Confidence (Abung Tinggi, Petani, Cukup Memuaskan)} = \frac{3}{5} = 0,6 = 60\%$$

$$\text{Confidence (Abung Tinggi, Petani, Memuaskan)} = \frac{2}{5} = 0,4 = 40\%$$

$$\text{Confidence (Abung Tinggi, Petani, Gunung Labuan)} = \frac{0}{5} = 0,00 = 0\%$$

$$\text{Confidence (Abung Tinggi, Petani, Bukit Kemuning)} = \frac{0}{5} = 0,00 = 0\%$$

$$\text{Confidence (Abung Tinggi, Menganggur, Cukup Memuaskan)} = \frac{3}{4} = 0,75 = 75\%$$

$$\text{Confidence (Abung Tinggi, Menganggur, Memuaskan)} = \frac{1}{4} = 0,25 = 25\%$$

$$\text{Confidence (Abung Tinggi, Menganggur, Gunung Labuhan)} = \frac{0}{4} = 0,00 = 0\%$$

$$\text{Confidence (Abung Tinggi, Menganggur, Bukit Kemuning)} = \frac{0}{4} = 0,00 = 0\%$$

$$\text{Confidence (Abung Tinggi, Cukup Memuaskan, Memuaskan)} = \frac{0}{3} = 0,00 = 0\%$$

$$\text{Confidence (Abung Tinggi, Cukup Memuaskan, Gunung Labuhan)} = \frac{0}{3} = 0,00 = 0\%$$

$$\text{Confidence (Abung Tinggi, Cukup Memuaskan, Bukit Kemuning)} = \frac{0}{3} = 0,00 = 0\%$$

$$\text{Confidence (Gunung Labuhan, Petani, Menganggur)} = \frac{0}{7} = 0,00 = 0\%$$

$$\text{Confidence (Gunung Labuhan, Petani, Cukup Memuaskan)} = \frac{5}{7} = 0,714 = 71,4\%$$

$$\text{Confidence (Gunung Labuhan, Petani, Memuaskan)} = \frac{2}{7} = 0,285 = 28,5\%$$

$$\text{Confidence (Gunung Labuhan, Petani, Abung Tinggi)} = \frac{0}{7} = 0,00 = 0\%$$

$$\text{Confidence (Gunung Labuhan, Petani, Bukit Kemuning)} = \frac{0}{7} = 0,00 = 0\%$$

$$\text{Confidence (Gunung Labuhan, Menganggur, Abung Tinggi)} = \frac{0}{6} = 0,00 = 0\%$$

$$\text{Confidence (Gunung Labuhan, Menganggur, Bukit Kemuning)} = \frac{0}{6} = 0,00 = 0\%$$

$$\text{Confidence (Gunung Labuhan, Menganggur, Cukup Memuaskan)} = \frac{5}{6} = 0,833 = 83,3\%$$

$$\text{Confidence (Gunung Labuhan, Menganggur, Memuaskan)} = \frac{1}{6} = 0,166 = 16,6\%$$

$$\text{Confidence (Gunung Labuhan, Cukup Memuaskan, Abung Tinggi)} = \frac{0}{8} = 0,00 = 0\%$$

$$\text{Confidence (Gunung Labuhan, Cukup Memuaskan, Bukit Kemuning)} = \frac{0}{8} = 0,00 = 0\%$$

$$\text{Confidence (Gunung Labuhan, Cukup Memuaskan, Memuaskan)} = \frac{0}{8} = 0,00 = 0\%$$

$$\text{Confidence (Bukit Kemuning, Petani, Abung Tinggi)} = \frac{0}{4} = 0,00 = 0\%$$

$$\text{Confidence (Bukit Kemuning, Petani, Gunung Labuan)} = \frac{0}{4} = 0,00 = 0\%$$

$$\text{Confidence (Bukit Kemuning, Petani, Menganggur)} = \frac{2}{4} = 0,5 = 50\%$$

$$\text{Confidence (Bukit Kemuning, Petani, Cukup Memuaskan)} = \frac{0}{4} = 0,00 = 0\%$$

$$\text{Confidence (Bukit Kemuning, Petani, Memuaskan)} = \frac{1}{4} = 1 = 100\%$$

$$\text{Confidence (Bukit Kemuning, Memuaskan, Abung Tinggi)} = \frac{0}{4} = 0,00 = 0\%$$

$$\text{Confidence (Bukit Kemuning, Memuaskan, Gunung Labuan)} = \frac{0}{4} = 0,00 = 0\%$$

$$\text{Confidence (Bukit Kemuning, Memuaskan, Menganggur)} = \frac{2}{4} = 0,5 = 50\%$$

$$\text{Confidence (Bukit Kemuning, Memuaskan, Cukup Memuaskan)} = \frac{0}{4} = 0,00 = 0\%$$

$$\text{Confidence (Menganggur, Cukup Memuaskan, Abung Tinggi)} = \frac{3}{8} = 0,375 = 0\%$$

$$\text{Confidence (Menganggur, Cukup Memuaskan, Bukit Kemuning)} = \frac{2}{8} = 0,375 = 37,5\%$$

$$\text{Confidence (Menganggur, Cukup Memuaskan, Gunung Labuan)} = \frac{5}{8} = 0,625 = 62,5\%$$

$$\text{Confidence (Menganggur, Cukup Memuaskan, Petani)} = \frac{6}{8} = 0,75 = 75\%$$

$$\text{Confidence (Menganggur, Cukup Memuaskan, Memuaskan)} = \frac{0}{8} = 0,00 = 0\%$$

Setelah dilakukan perhitungan *confidence* dari 3 *itemset*, hasil perhitungan yang memenuhi *minimum confidence* sebesar 60 % disajikan dalam Tabel 4.6 untuk memudahkan proses selanjutnya.

Tabel 4. 6 Kombinasi 3 *itemset* yang memenuhi *minimum confidence* 60%

Itemset	Support	Confidence
Abung Tinggi, Petani, Menganggur	20%	80%
Abung Tinggi, Petani, Cukup Memuaskan	15%	60%
Abung Tinggi, Menganggur, Cukup Memuaskan	15%	75%

Gunung Labuan, Petani, Cukup Memuaskan	20%	71%
Gunung Labuan, Menganggur, Cukup Memuaskan	25%	83%
Bukit Kemuning, Petani, Memuaskan	20%	100%
Menganggur, Cukup Memuaskan, Petani	30%	75%

Berdasarkan Tabel 4.6 didapatkan 7 kombinasi yang memenuhi *minimum confidence* sebesar 60%. Untuk menentukan *rules*, akan dipilih kombinasi yang memenuhi *minimum support* dan *minimum confidence*. Maka dari Tabel 4.6 di atas, hanya ada 5 kombinasi yang memenuhi *minimum support* dan *minimum confidence*, yaitu:

1. Abung Tinggi, Petani, Menganggur
2. Gunung Labuan, Petani, Cukup Memuaskan
3. Gunung Labuan, Menganggur, Cukup Memuaskan
4. Bukit Kemuning, Petani, Memuaskan
5. Menganggur, Cukup Memuaskan, Petani

Rules tersebut dideskripsikan sebagai berikut:

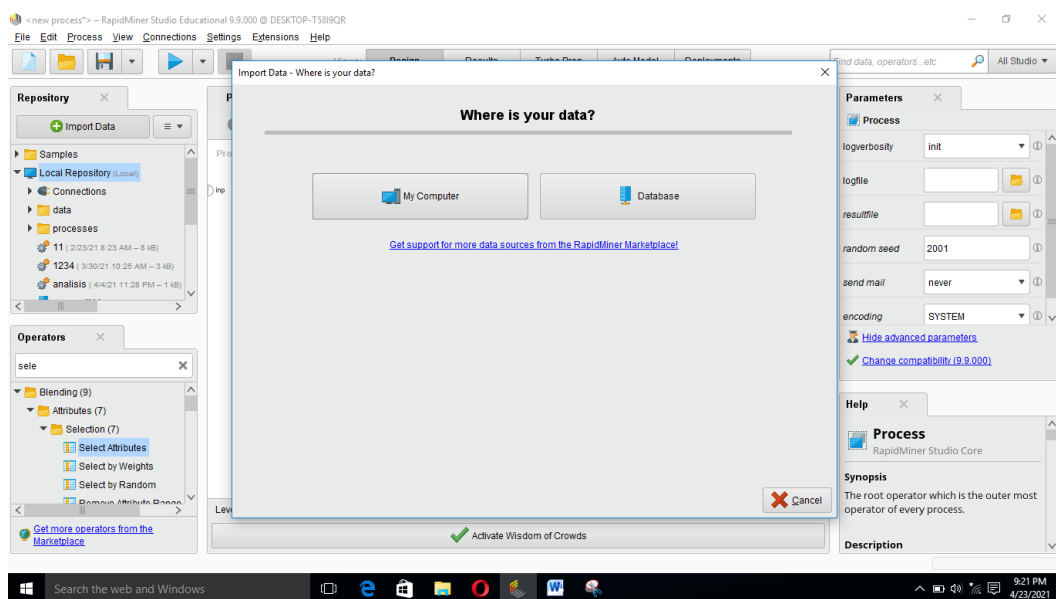
1. Jika berasal dari Abung Tinggi, Pekerjaan orangtuanya Petani, maka status setelah lulus menganggur dengan didukung 20 % kemunculan dalam data dan nilai kepastian 80%.
2. Jika berasal dari Gunung Labuan dan pekerjaan orangtuanya Petani, maka kategori kelulusannya Cukup Memuaskan, didukung 20% kemunculan dalam data dan nilai kepastian 71%.
3. Jika berasal dari Gunung Labuan dan status setelah lulusnya Menganggur, maka kategori kelulusannya Cukup Memuaskan, didukung 25% kemunculan dalam data dan nilai kepastian 83%.
4. Jika berasal dari Bukit Kemuning dan pekerjaan orangtuanya Petani, maka kategori kelulusannya Memuaskan, didukung 20% kemunculan dalam data dan nilai kepastiannya 100%.

5. Jika status setelah lulusnya Mengganggu dan kategori kelulusannya Cukup Memuaskan, maka pekerjaan orangtuanya Petani, didukung 30% kemunculan dalam data dan nilai kepastiannya 75%.

4.2 Analisis menggunakan Rapidminer

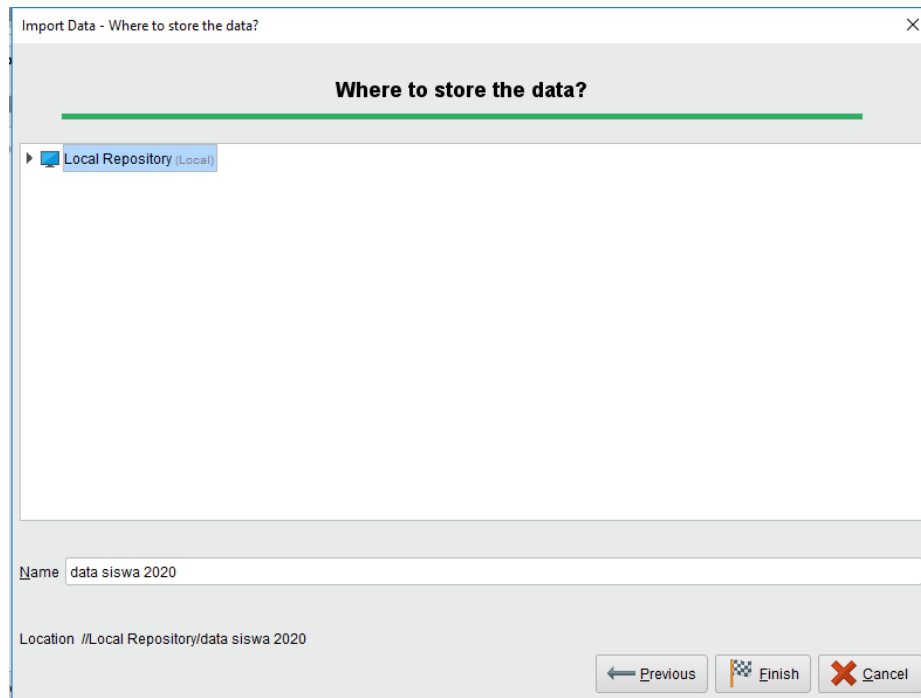
Proses analisis menggunakan *tools* Rapidminer dilakukan dengan dua proses, yaitu proses *preprocessing data* dan proses analisis data.

Pada proses *preprocessing data*, pertama-tama dilakukan proses seleksi data, dan pembersihan data menggunakan *tools* rapidminer. Langkah-langkahnya yaitu Buka rapidminer kemudian pilih *new process*.



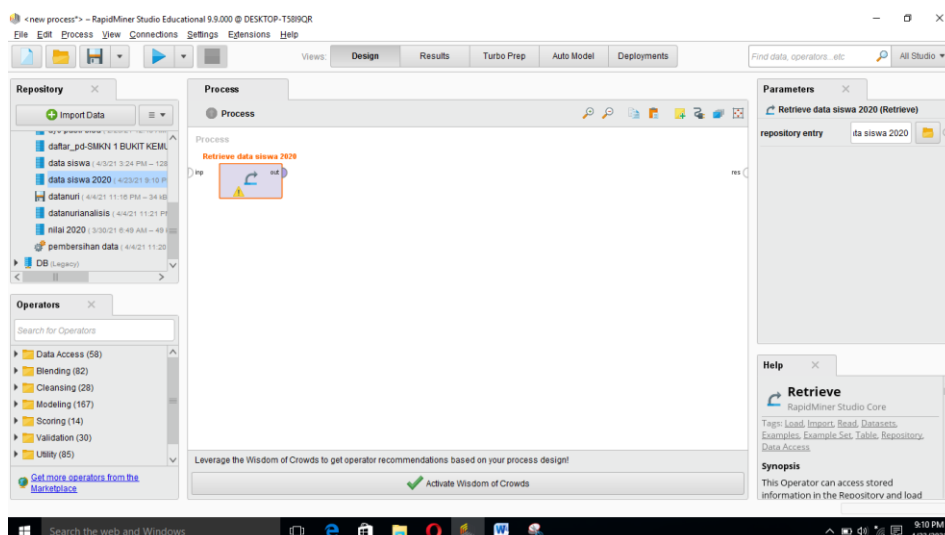
Gambar 4. 1 *Import Data*

Pilih *Import data* kemudian pilih *my computer* karena data yang akan diolah berada di Laptop. Kemudian pilih file yang akan *diimport*.



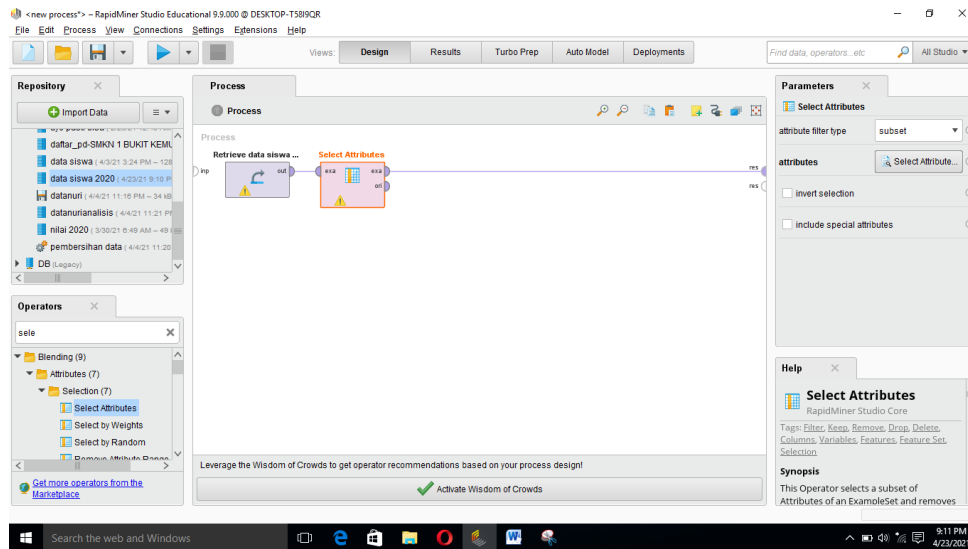
Gambar 4. 2 Store Data

Pilih lokasi penyimpanan di *local repository*, beri nama sesuai dengan nama filenya, pada Gambar 4.2 penulis memberi nama **data siswa 2020**, kemudian pilih *finish*. Selanjutnya tarik data yang sudah diimport di *repository* ke lembar proses, seperti pada Gambar 4.3.



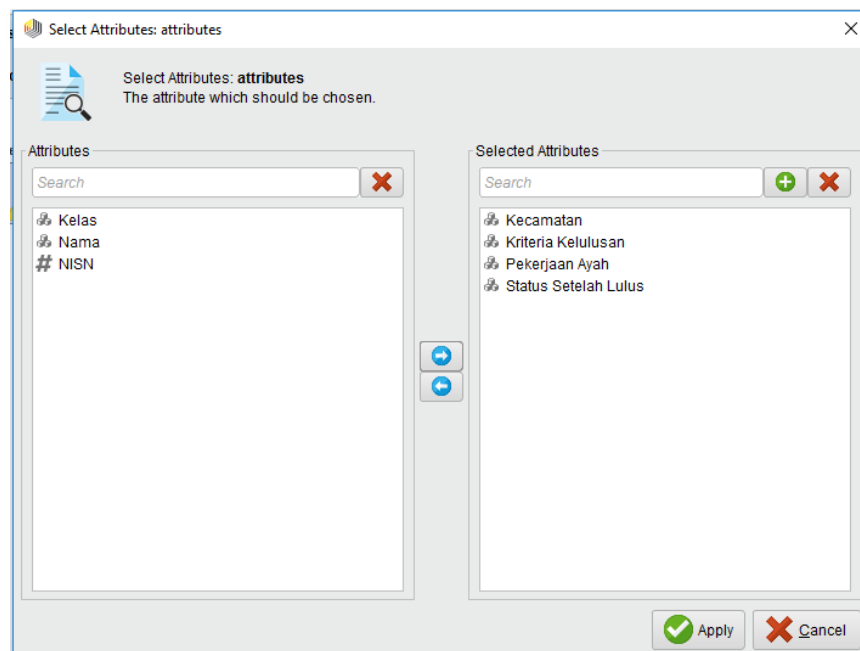
Gambar 4. 3 Data Pada Lembar Kerja

Selanjutnya pilih *Select attributes* dibagian *operators*, dan dibagian *parameters* pada *Attribute filter type* pilih *subset*. *Select attributes* digunakan untuk menyeleksi data yang akan digunakan. Proses dapat dilihat pada Gambar 4.4.



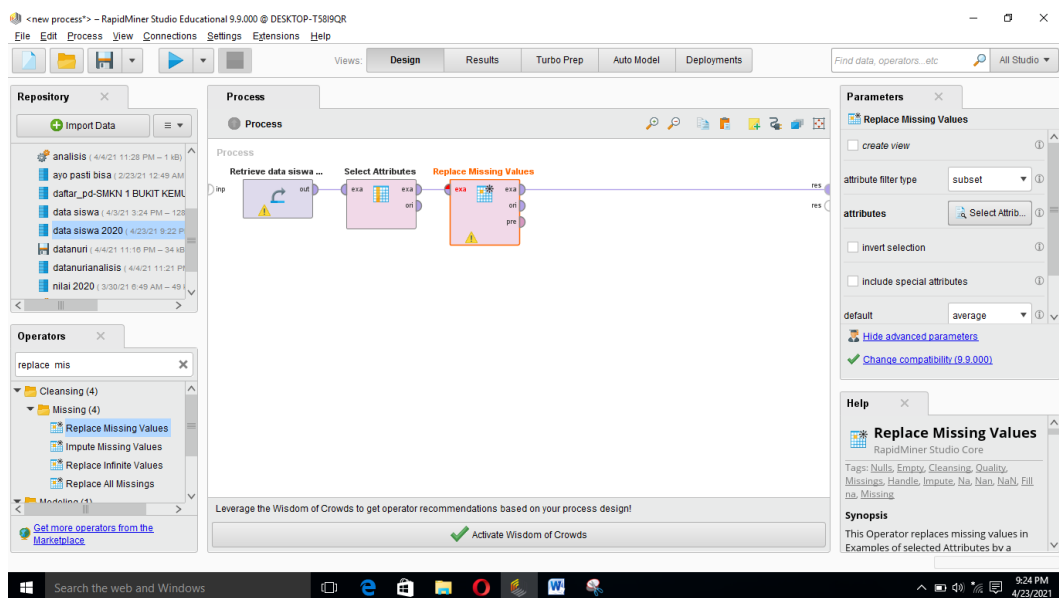
Gambar 4.4 *Select Attributes*

Pada bagian atribut klik *select attributes* untuk memilih atribut apa saja yang akan diambil. Dalam proses ini penulis menghilangkan atribut nama, nik, dan kelas, maka selected attributes-nya adalah Kecamatan, Kriteria Kelulusan, Pekerjaan Ayah, dan Status Setelah Lulus. Proses dapat dilihat pada Gambar 4.5.



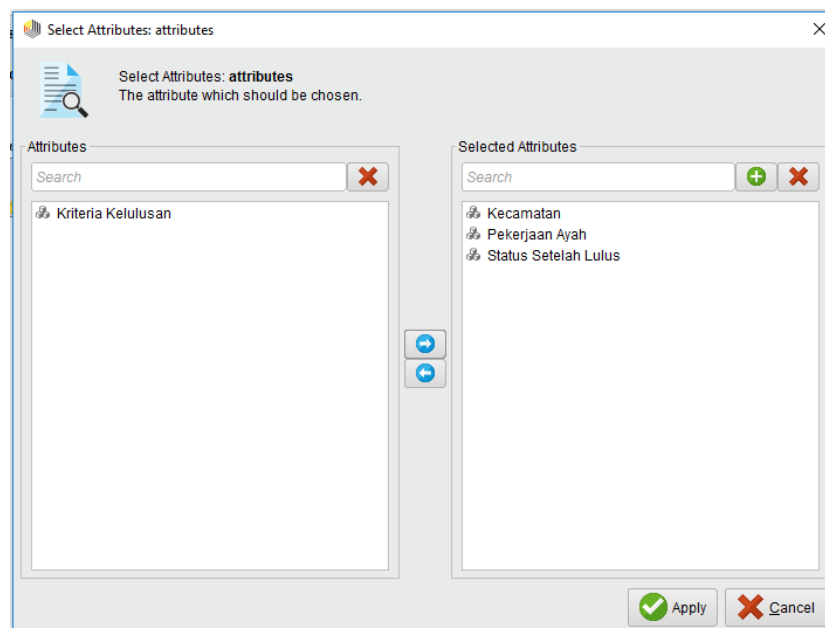
Gambar 4.5 Proses *Select Attributes*

Setelah atribut dipilih, klik *apply*. Selanjutnya pilih *Replace missing values* untuk melengkapi data yang kosong. Pilih *subset* pada *attribute filter type* dibagian *parameters* lalu klik *select atribut* untuk memilih atribut apa saja yang akan dilengkapi. Proses dapat dilihat pada Gambar 4.6.



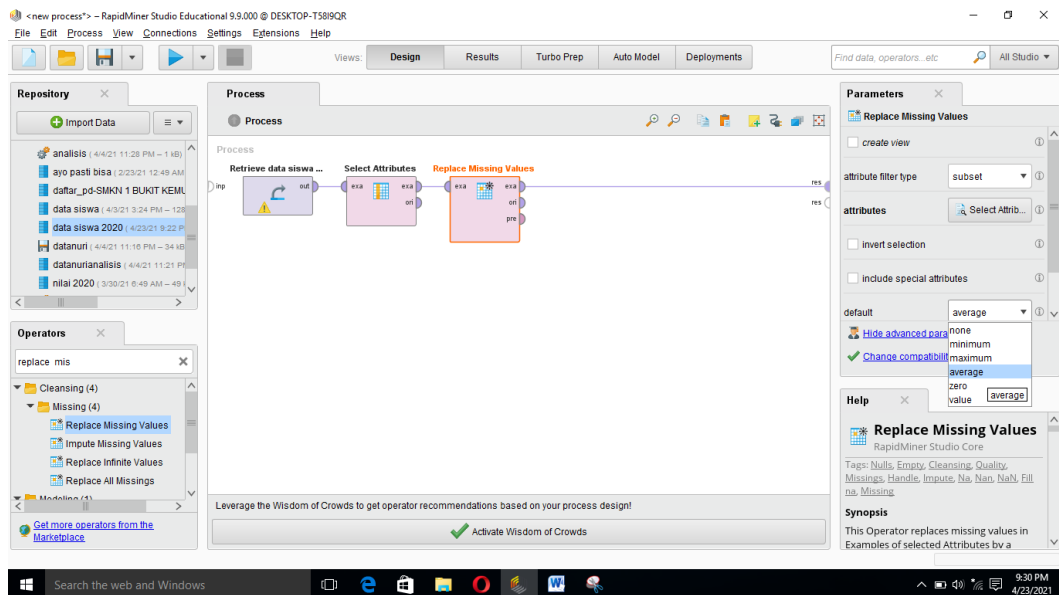
Gambar 4. 6 Replace Missing Values

Pilih atribut yang akan di-*replace missing value*, disini saya memilih atribut Kecamatan, Pekerjaan Ayah, dan Status Setelah Lulus. Setelah atribut terpilih, klik *apply*. Proses dapat dilihat pada Gambar 4.7.



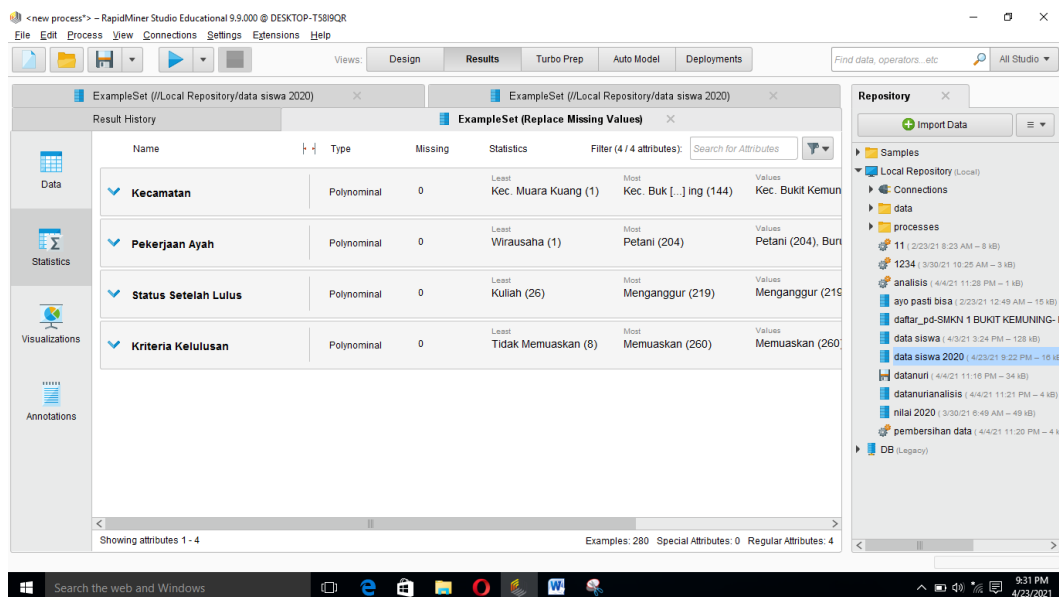
Gambar 4. 7 Pemilihan atribut yang akan dihilangkan

Dibagian *Default*, pilih *average* karena data yang kosong akan diisi dengan rata-rata data. Selain *average*, data yang kosong bisa juga diisi dengan nilai minimal, maksimal, atau *zero*, tetapi penulis memilih *average* untuk melengkapi data yang kosong tersebut. Proses dapat dilihat pada Gambar 4.8.



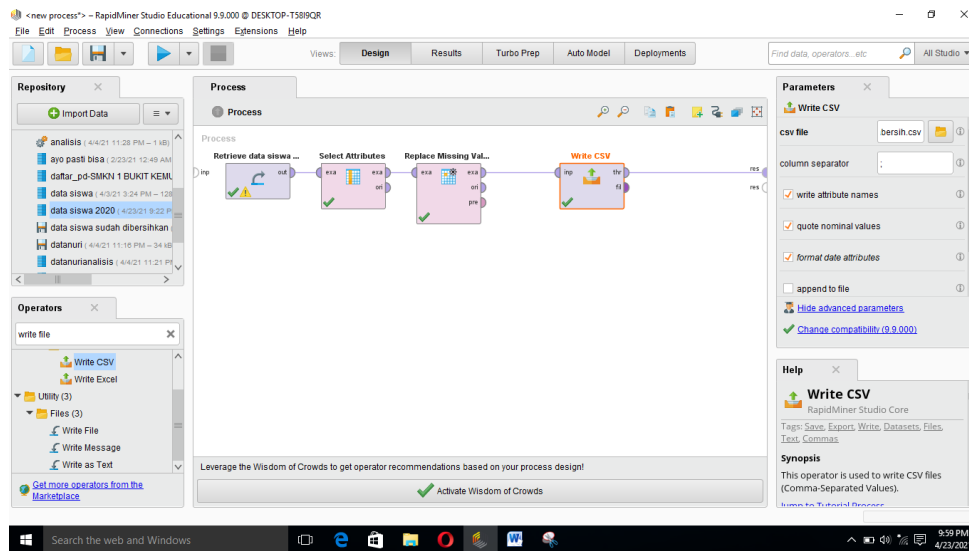
Gambar 4. 8 Bagian *Parameters Replace Missing Values*

Setelah itu klik *Run* untuk menjalankan proses. Dapat dilihat pada bagian *statistic* sudah tidak ada *missing value* yang artinya data sudah bersih. Proses dapat dilihat pada Gambar 4.9 berikut ini.



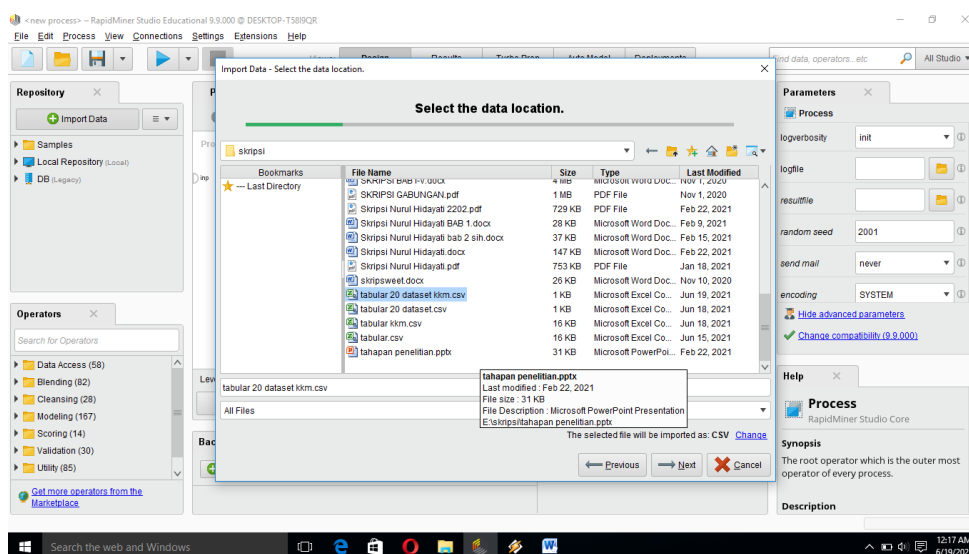
Gambar 4. 9 Statistik data yang sudah dibersihkan

Selanjutnya kita simpan data yang sudah dibersihkan dalam bentuk csv. Dengan memilih *write csv* pada bagian *operators*, lalu pilih lokasi penyimpanan dibagian *parameters*, lalu hubungkan dengan *operator* yang lain dan *run* prosesnya maka akan ada di lokasi penyimpanan yang dipilih seperti pada Gambar 4.10.



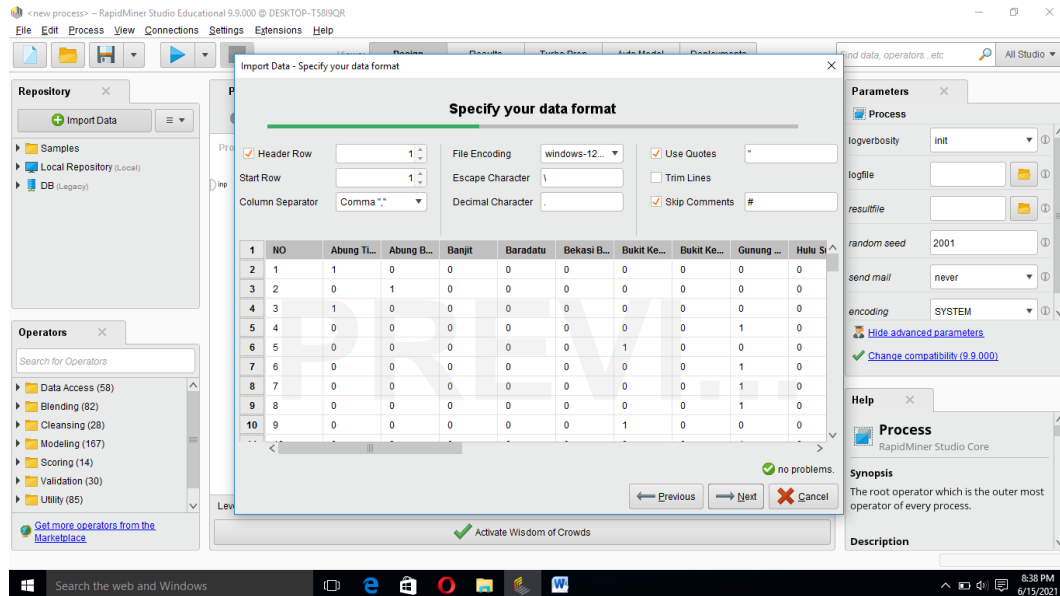
Gambar 4. 10 Proses Write CSV

Data sudah tersimpan di lokasi penyimpanan yang dipilih. Selanjutnya akan dilakukan pengolahan data menggunakan Algoritma Apriori. Langkah pertama *import* data yang sudah dibersihkan dan dijadikan format tabular. Proses dapat dilihat pada Gambar 4.11.



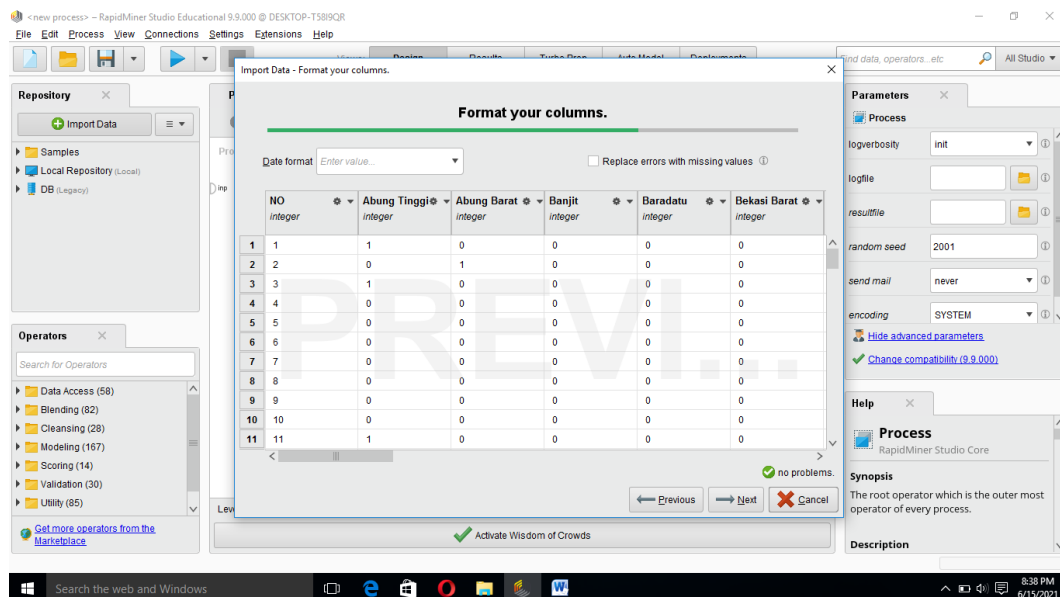
Gambar 4. 11 Proses Import Data

Selanjutnya akan muncul jendela yang menunjukkan format dari data yang kita *import*, pilih *next* untuk melanjutkan. Proses dapat dilihat pada Gambar 4.12.



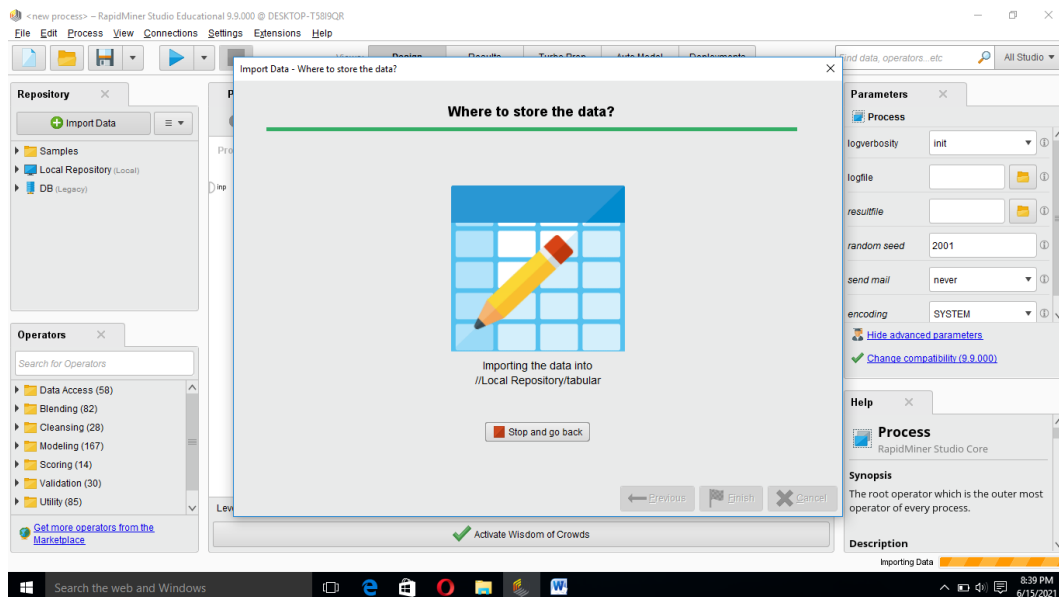
Gambar 4. 12 Jendela *Specify Your Data Format*

Lalu akan muncul tampilan *preview* data yang akan diimpor, pilih *next* untuk melanjutkan. Proses dapat dilihat pada Gambar 4.13.



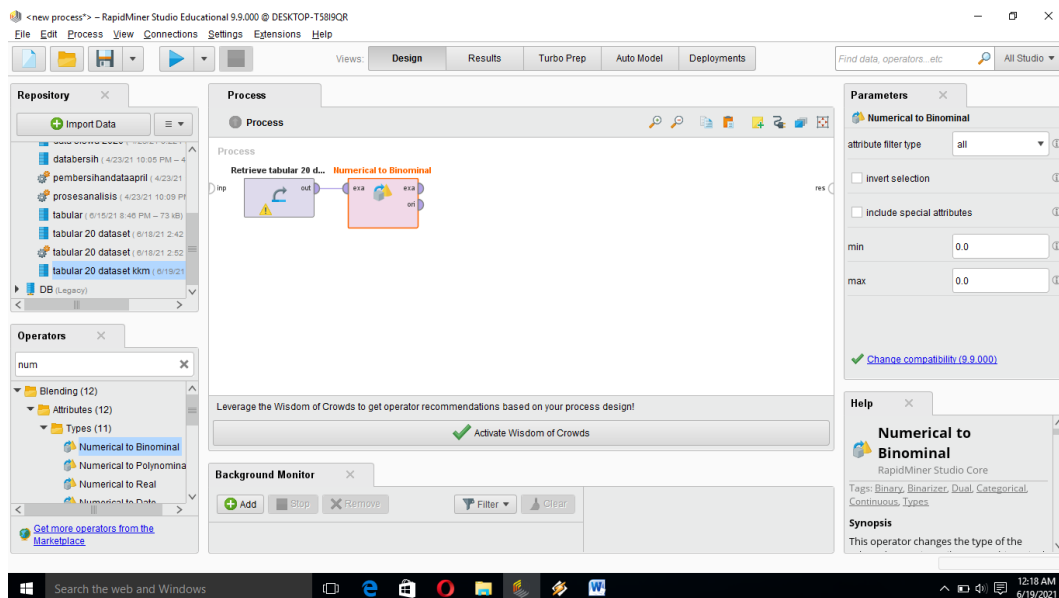
Gambar 4. 13 Jendela *Format Kolom*

Selanjutnya data akan diimpor kedalam *tools*. Proses dapat dilihat pada Gambar 4.14.



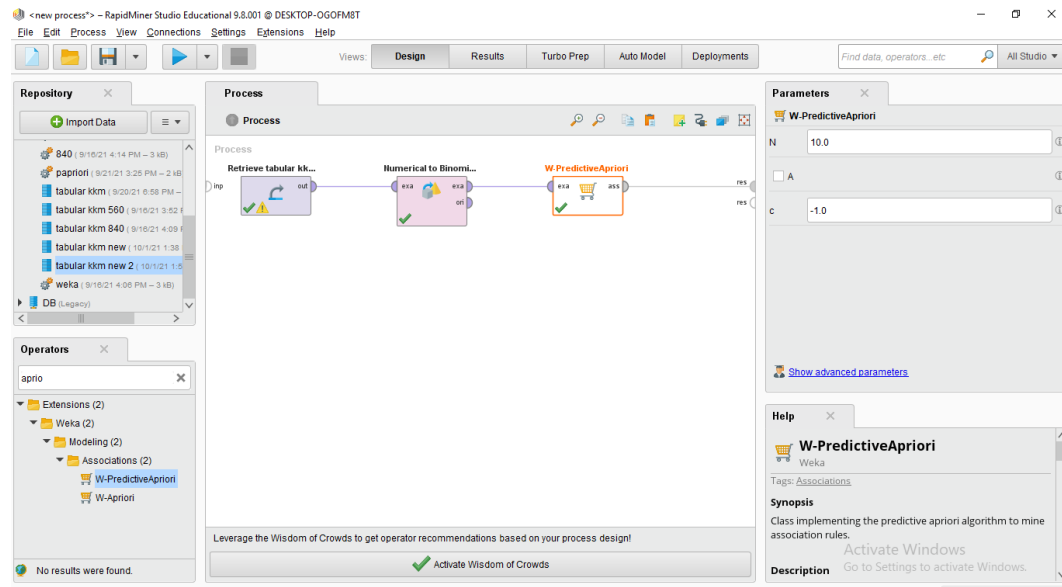
Gambar 4. 14 Proses Pengimporan Data

Buka jendela proses dan tarik data yang sudah diimpor. Setelah itu tambahkan operator *numerical to binominal*. Operator ini berfungsi untuk mengubah format data menjadi format yang bisa dibaca *tools* rapidminer. Proses dapat dilihat pada Gambar 4.15.



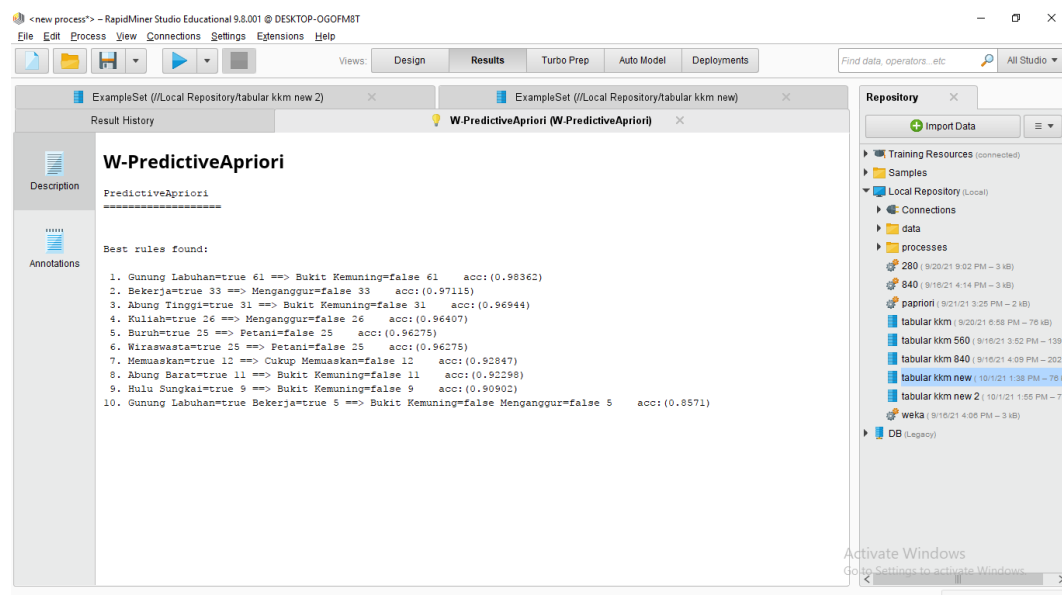
Gambar 4. 15 Jendela Proses Pengolahan Data

Tambahkan operator *W-predictive Apriori* dan ubah parameternya, yaitu jumlah *rule* yang ingin ditampilkan. Penulis ingin menampilkan 10 *rules*. Proses dapat dilihat pada Gambar 4.16.



Gambar 4. 16 Jendela Proses Pengolahan Data dengan Operator dan Parameter *W-predictive Apriori*

Selanjutnya *running proses* tersebut untuk melihat pola yang dihasilkan. Pola yang dihasilkan dapat dilihat pada Gambar 4.17.



Gambar 4. 17 Pola yang Dihasilkan Proses Analisis

Didapatkan 10 rule yaitu:

1. Jika berasal dari Gunung Labuan, maka siswa tidak mungkin berasal dari Bukit Kemuning.
2. Jika Status setelah lulusnya bekerja, maka tidak mungkin menganggur dalam waktu bersamaan.
3. Jika berasal dari Abung Tinggi maka siswa tidak mungkin berasal dari Bukit Kemuning.
4. Jika status setelah lulusnya kuliah, maka siswa tidak mungkin menganggur dalam waktu bersamaan.
5. Jika pekerjaan orang tuanya Buruh, maka tidak mungkin pekerjaan orang tuanya petani.
6. Jika pekerjaan orang tuanya wiraswasta, maka tidak mungkin petani dalam waktu bersamaan
7. Jika hasil ujiannya memuaskan, maka tidak mungkin tidak memuaskan dalam waktu bersamaan
8. Jika siswa berasal dari Abung Barat, maka tidak mungkin berasal juga dari Bukit Kemuning.
9. Jika berasal dari Hulu Sungkai, maka tidak mungkin berasal dari Bukit Kemuning.
10. Jika berasal dari Gunung Labuan, maka stataus setelah lulusnya bekerja. Dan jika berasal dari Bukit Kemuning, maka status setelah lulusnya menganggur.

BAB V

SIMPULAN DAN SARAN

5.1 Kesimpulan

Dari uraian di atas dapat ditarik kesimpulan sebagai berikut:

1. Teknik data mining dengan algoritma apriori dapat diterapkan untuk mencari pola kelulusan siswa. Proses dilakukan mengikuti tahapan-tahapan metode KDD sehingga menghasilkan pola-pola kelulusan siswa. Selain mengikuti tahapan metode KDD, proses persiapan data juga dilakukan berdasarkan tahapan data mining, yaitu seleksi data, integrasi data, pembersihan data, dan transformasi data. Selanjutnya data diolah menggunakan algoritma apriori dengan melakukan pencarian nilai *support* dan *confidence* hingga mendapatkan *rule* yang memenuhi *minimum support* dan *minimum confidence*. *Rule* yang memenuhi *minimum support* dan *minimum confidence* itu yang kemudian menjadi pola kelulusan siswa.
2. Faktor-faktor yang sangat menentukan kelulusan siswa yaitu pekerjaan orang tua dan asal siswa mempengaruhi status setelah siswa lulus dan kategori kelulusan siswa. Berikut ini *rule* yang dihasilkan dari proses analisis:
 - 1) Jika berasal dari Gunung Labuan, maka siswa tidak mungkin berasal dari Bukit Kemuning.
 - 2) Jika Status setelah lulusnya bekerja, maka tidak mungkin menganggur dalam waktu bersamaan.
 - 3) Jika berasal dari Abung Tinggi maka siswa tidak mungkin berasal dari Bukit Kemuning.
 - 4) Jika status setelah lulusnya kuliah, maka siswa tidak mungkin menganggur dalam waktu bersamaan.
 - 5) Jika pekerjaan orang tuanya Buruh, maka tidak mungkin pekerjaan orang tuanya petani.
 - 6) Jika pekerjaan orang tuanya wiraswasta, maka tidak mungkin petani dalam waktu bersamaan.

- 7) Jika hasil ujiannya memuaskan, maka tidak mungkin tidak memuaskan dalam waktu bersamaan
- 8) Jika siswa berasal dari Abung Barat, maka tidak mungkin berasal juga dari Bukit Kemuning.
- 9) Jika berasal dari Hulu Sungkai, maka tidak mungkin berasal dari Bukit Kemuning.
- 10) Jika berasal dari Gunung Labuan, maka status setelah lulusnya bekerja. Dan jika berasal dari Bukit Kemuning, maka status setelah lulusnya menganggur.

5.2 Saran

Berdasarkan pembahasan di atas, saran penulis sebagai berikut:

1. Perlu dilakukan penelitian lain menggunakan data yang sama dengan algoritma yang berbeda, misalnya algoritma FP-Growth. Hal ini dimaksudkan untuk menguji kebenaran pola yang dihasilkan.
2. Perlu dilakukan penelitian lain dengan data yang sama menggunakan metode data mining yang lain seperti klasifikasi dan *Clustering*. Hal ini dimaksudkan untuk menguji kebenaran pola yang dihasilkan. Penelitian menggunakan asosiasi dengan data sekolah dirasa kurang bervariasi karena semua dataset memiliki 4 item yang menyebabkan tidak banyak pola-pola yang dihasilkan.

DAFTAR PUSTAKA

- Adinugroho, S., & Sari, Y. A. (2018). *Implementasi Data Mining Menggunakan WEKA*. Malang: UB Press.
- Fauzy, M., Saleh W, K. R., & Asror, I. (2016). Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Kota Bandung. *Jurnal Ilmiah Teknologi Informasi Terapan Volume II, No.2*, 221.
- Febrianti, E. L., & Suryadi, A. (2018). PENERAPAN DATA MINING DENGAN ALGORITMA APRIORI UNTUK ANALISIS POLA PEMBELIAN KONSUMEN . *Seminar Nasional Royal (SENAR)*, 173-176.
- Maharani, Hasibuan, N. A., Silalahi, N., Nasution, S. D., Mersan, Suginam, et al. (2017). IMPLEMENTASI DATA MINING UNTUK PENGATURAN LAYOUT MINIMARKET DENGAN MENERAPKAN ASSOCIATION RULE . *Jurnal Riset Komputer*, 6-11.
- Meilani, B. D., & Susanti, N. (2015). Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode Naive Bayes. *Jurnal Ilmiah NERO Vol.1*, 182.
- Modyta, A. (2019). Penerapan Data Mining Untuk Menganalisa Pola Peminjaman Buku Pada Perpustakaan IIB Darmajaya Menggunakan Algoritma Apriori. *Skripsi*.
- Nofriansyah, D. (2014). *Konsep Data Mining VS Sistem Pendukung Keputusan*. Yogyakarta: Deepublish.
- Nofriansyah, D., & Nurcahyo, G. W. (2015). *Algoritma Data Mining Dan Pengujian*. Sleman: Penerbit Deepublish.
- Nofriansyah, D., Yetri, M., Erwansyah, K., & Suharsil. (2019). Penerapan Data Mining Dalam Menganalisa Data Penjualan Untuk Mendapatkan Pola Rekomendasi Penjualan Menggunakan Algoritma Apriori Pada K3 Mart . *Sains dan Komputer*, 176-182.
- Novianti, D. (2019). Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner . *Jurnal Komputer dan Informatika* , 49.
- Nugraha, P. G., Aribawa, I. W., Priyana, I. P., & Indrawan, G. (2016). Penerapan Metode Decision Tree (Data Mining) Untuk Memprediksi Tingkat Kelulusan Siswa SMPN 1 Kintamani. *Seminar Nasional Vokasi dan Teknologi*, (hal. 35-44). Denpasar-Bali.

- Rizky, A. R., & Sadikin, M. (2019). Data Mining Menggunakan Algoritma Apriori Untuk Rekomendasi Produk Bagi Pelanggan. *Jurnal Teknologi dan Sistem Komputer*, 103.
- Sikumbang, E. D. (2018). Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori . *Jurnal Teknik Komputer*, 156-161.
- Sumpena, J., & Kurnia, N. (2019). ANALISIS PREDIKSI KELULUSAN SISWA PKBM PAKET C . *TEDC*, 127-133.
- Susilo, A. A. (2018). Penerapan Algoritma Apriori Pada Pengolahan Data Transaksi Penjualan di Minimarket Priyo Kota Lubuklinggau. *JTKSI*, 39-46.
- Wardani, N. W. (2020). *Penerapan Data Mining Dalam Analytic CRM*. Medan: Yayasan Kita Menulis.
- Yanto, R., & Khoiriah, R. (2015). Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat. *Citec Journal*, 102 - 113.
- Zanuardi, A., & Suprayitno, H. (2018). Analisa Karakteristik Kecelakaan Lalu Lintas di Jalan Ahmad Yani Surabaya melalui Pendekatan Knowledge Discovery in Database . *Jurnal Manajemen Aset Infrastruktur & Fasilitas – Vol. 2, No. 1*, 47.
- Tampubolon Kennedy, Saragih Hoga, Reza Bobby. *Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan*. Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI). Vol-1 No.1, Oktober 2013.
- Kusrini dan Luthfi. E. Taufiq. 2009. *Algoritma Data Mining*. Yogyakarta. Andi.
- Modyta Anggraini. 2019. *Penerapan Data Mining Untuk Menganalisa Pola Peminjaman Buku Pada Perpustakaan IIB Darmajaya Menggunakan Algoritma Apriori*. Skripsi.
- Susilo Andri Tri Anto. *Penerapan Algoritma Apriori Pada Pengolahan Data Transaksi Penjualan di Minimarket Priyo Kota Lubuklinggau*. JTKSI. Vol.01 No.3 September 2018.