Analisis Hubungan Fasilitas Sekolah Dengan Perolehan Nilai Ujian Nasional Menggunakan Metode Apriori Association Rule

Ibnu Agiel's Althur[1], Imelda Atastina[2]

[1,2]Fakultas Informatika, Universitas Telkom, Bandung
[1]althuragiel@students.telkomuniversity.ac.id, [2]imelda@telkomuniversity.ac.id

Abstrak

Salah satu dari kebijakan Kemdikbud adalah pelaksanaan Ujian Nasional (UN). Tujuan dari UN adalah untuk mengukur tingkat keberhasilan siswa dalam menyerap pendidikan selama menempuh pendidikan sesuai tingkat pendidikannya. Meskipun UN sudah tidak menjadi syarat kelulusan siswa-siswi sekarang, tidak dapat dipungkiri UN satu-satunya ujian yang memiliki standar dan dilaksanakan secara nasional. Oleh karena itu banyak usaha yang dilakukan sekolah untuk mendongkrak nilai siswa-siswi mereka. Namun dengan ketidaksetaraan fasilitas antar sekolah di Indonesia memungkinkan adanya perbedaan kualitas siswa-siswi antar sekolah. Dengan menggunakan metode Algoritma Apriori penulis ingin menunjukan adanya keterkaitan antara nilai UN dengan sarana dan prasarana sekolah. Berdasarkan hasil dari dua percobaan terdapat hubungan antara nilai UN dengan fasilitas sekolah secara parsial.

Kata kunci : UN, Algoritma Apriori, Association Rules, Data Mining, Fasilitas

Abstract

One of Kemdikbud (Ministry of Education and Culture) is to arrange National Exam (UN). The objective of UN is to evaluate students during their education according to their level of education. Although UN now is not being a condition for graduation anymore, it cannot be denied that UN is the only examination that has a national standard and implemented nationally. Because of that there was a lot of effort for schools to raise their students scores. However, with facilities inequality between schools in Indonesia make it possible to have a different quality of students between schools. With using Apriori Algorithm method writer like to show there are linkages between UN scores with school facilities and infrastructures. Based on two experiments there are relations between UN scores with school facilities partially.

Keywords: UN, Apriori Algorithm, Association Rules, Data Mining, Facilities

1. Pendahuluan

Latar Belakang

Salah satu dari kebijakan Kemdikbud adalah pelaksanaan Ujian Nasional (UN). Tujuan dari UN adalah untuk mengukur tingkat keberhasilan siswa dalam menyerap pendidikan selama menempuh pendidikan sesuai tingkat pendidikannya. Meskipun UN sudah tidak menjadi syarat kelulusan siswa-siswi sekarang, tidak dapat dipungkiri UN satu-satunya ujian yang memiliki standar dan dilaksanakan secara nasional. Oleh karena itu banyak usaha yang dilakukan sekolah untuk mendongkrak nilai siswa-siswi mereka. Namun dengan ketidaksetaraan fasilitas antar sekolah di Indonesia memungkinkan adanya perbedaan kualitas siswa-siswi antar sekolah.

Kemdikbud memiliki data center untuk menampung data sekolah, siswa, dan guru seluruh Indonesia yang bernama Data Pokok Pendidikan (Dapodik). Dapodik menyimpan data pokok seluruh sekolah di Indonesia. Data yang tersimpan dalam Dapodik mencakup detail sekolah seperti jumlah gedung, ruangan kelas, perpustakaan, toilet, perpustakaan, kantin, dan fasilitas pendamping lainnya seperti ruangan / bangunan ibadah, akses listrik, dan internet. Selain itu Dapodik juga menyimpan data guru dan siswa serta nilai rapor siswa.

Dengan menggabungkan Dapodik dengan data hasil UN penulis ingin menunjukkan apakah ada hubungannya antara kualitas fasilitas sekolah, dengan hasil UN yang diperoleh siswa. Penulis menggunakan Apriori association rule Rules. Penelitian ini dilakukan untuk mencari hubungan antara fasilitas sekolah dengan nilai UN.

Penelitian ini diharapkan bermanfaat bagi peneliti lain dalam melakukan penelitian pada data yang dimiliki lembaga pemerintahan lainnya. Penulis juga berharap jika hasil penelitian ini dapat digunakan sebagai acuan untuk peneliti lain ke depannya. Selain itu penelitian ini diharapkan dapat menjadi pemicu bagi seluruh lembaga pemerintahan khususnya Kemdikbud untuk mulai memanfaatkan dan mengembangkan teknologi data mining dalam menentukan kebijakan ke depannya.

Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

1) Dataset yang tersedia sebanyak 1700 data fasilitas dan non fasilitas sekolah beserta nilai UN mereka pada tahun 2017 – 2018.
2) Dalam satu baris data sekolah terdiri dari 22 atribut.
3) Lingkup sekolah yang dijadikan dataset adalah data sekolah SMP seluruh Indonesia yang didapatkan pada situs SekolahKita[1] dengan pengecualian sekolah Indonesia di luar negeri.

Tujuan Penelitian

Tujuan dari penelitian ini setidaknya ada tiga point:

1) Mencari hubungan antara jumlah dan kondisi fasilitas yang diberikan sekolah terhadap hasil UN para siswa.
2) Mencari tahu apakah dengan algoritma Apriori dapat menyimpulkan keterkaitan antara fasilitas sekolah dengan nilai UN.
3) Mencari tahu apakah dengan preprocessing data dapat membantu menghasilkan rules keterkaitan antara fasilitas sekolah dengan nilai UN.

Organisasi Tulisan

Beberapa poin yang akan dijelaskan pada jurnal ini adalah sebagai berikut. Poin pertama menjelaskan latar belakang, batasan, tujuan dan organisasi dari jurnal ini. Kemudian poin kedua terdapat studi literatur terkait dengan penelitian yang sudah dilakukan sebelumnya dan penelitian yang sedang dilakukan serta beberapa tinjauan pustaka yang terkait dengan penelitian. Kemudian poin ketiga akan menjelaskan proses percobaan yang akan dijelaskan pada jurnal ini. Kemudian poin keempat, akan menjelaskan hasil pengujian dan analisis yang telah dilakukan. Dan yang terakhir poin kelima akan dijelaskan mengenai kesimpulan dan saran yang dihasilkan dari penelitian ini.

2. Studi Literatur

2.1 Data Mining

Menurut Fayyad et al [1] data mining merupakan knowledge discovery in database atau yang disingkat KDD. Pengetahuannya bisa berupa pola atau relasi data yang valid dan tidak diketahui sebelumnya [2]. Sementara beberapa peneliti lain beranggapan bahwa data mining merupakan gabungan dari beberapa disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari sekumpulan data yang sangat besar, yang metode-metode di dalamnya adalah irisan dari artificial intelligence, machine learning, statistics, dan database systems [2] [3] [4]. Berdasarkan dari teori-teori di atas dapat disimpulkan bahwa data mining merupakan proses pengolahan data untuk menemukan sesuatu yang baru berupa pola-pola yang dapat dijadikan patokkan sebagai pengetahuan yang baru dari data tersebut.

Berdasarkan fungsionalitasnya, data mining terbagi menjadi enam kelompok [1] [2]:

Classification: mengelompokkan struktur yang diketahui untuk diaplikasikan pada data baru. Contoh klasifikasi penerima beasiswa, dan klasifikasi makhluk hidup berdasarkan data ciri-ciri fisiknya.

Clustering: mengelompokkan data yang tidak diketahui labelnya ke dalam beberapa kelompok tertentu dengan mencari kemiripan antar datanya.

Regression: memodelkan data dengan suatu fungsi dengan meminimalkan kesalahan prediksi sekecil mungkin.

Anomaly detection: mendeteksi data yang tidak lazim, data tersebut dapat berupa outlier (pencilan), perubahan atau deviasi yang mungkin sangat penting dan perlu dipelajari lebih lanjut.

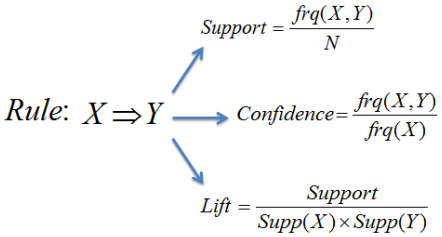
Association rule learning: bisa juga disebut pemodelan ketergantungan (dependency modelling); mencari relasi antar variabel.

Summarization: merepresentasikan data dalam bentuk ringkas, dan sederhana, dapat berupa visualisasi dan pembuatan laporan.

2.2 Association Rules

Secara garis besar *association rule* merupakan proses dari *data mining* untuk mencari aturan asosiatif antara suatu kombinasi item [5]. *Association rules* mencari semua *itemset* yang memiliki *support* lebih besar dari besaran minimum yang telah ditentukan, lalu membuat *rules* yang memiliki *confidence* lebih besar dari minimum yang telah ditetapkan. *Lift* merupakan rasio dari *support* yang telah diobservasi untuk mengetahui apakah *rules* yang telah dibuat valid atau tidak [6].

[1] http://sekolah.data.kemdikbud.go.id/index.php/



Gambar 1 Rumus Umum *Association Rules*
Sumber (http://www.saedsayad.com/association_rules.htm)

$$Support(X \rightarrow Y) = \frac{frq(X \cup Y)}{N}$$

- $frq(X,Y)$: frekuensi munculnya *item* X dengan *item* Y
- N: Jumlah data

Support merupakan presentasi kombinasi *item* dalam *itemset* yang mempresentasikan presentasi $X \cup Y$ dari total *records* dalam *database*. asumsikan *support* dari sebuah *item* sebesar 0.1%, berarti hanya 0.1 persen dari semua transaksi yang mengandung pembelian *item* tersebut [7].

$$Confidence(X \rightarrow Y) = \frac{frq(X \cup Y)}{frq(X)}$$

- $frq(X,Y)$: frekuensi munculnya *item* X dengan *item* Y
- $frq(X)$: frekuensi munculnya item X dalam *itemset*.

Confidence merupakan presentasi dari data yang memiliki $X \cup Y$ terhadap jumlah data yang memiliki *X*. *Confidence* digunakan untuk mengukur kekuatan dari *association rules* yang telah dibuat. Asumsi sebuah *rule* $X \rightarrow Y$ memiliki nilai *confidence* 80%, maka dalam keseluruhan data yang mengandung *item X* ada 80% data memiliki *item X* dan *Y* secara bersamaan [7].

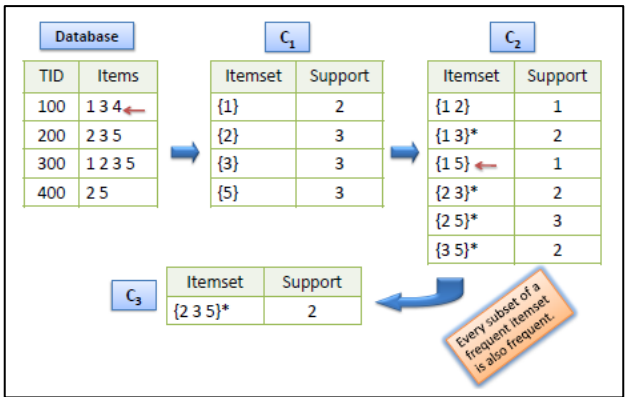
$$Lift(X \rightarrow Y) = \frac{Support(X \rightarrow Y)}{Support(X)\ x\ Support(Y)} = \frac{Confidence(X \rightarrow Y)}{Support(Y)}$$

- $Support(X \rightarrow Y)$: nilai *support itemset* X bersama Y
- $Support(X) = frq(X)$.
- $Support(Y) = frq(Y)$.
- $Confidence(X \rightarrow Y)$ = nilai *confidence rule* antara *itemset* X terhadap Y.

Lift merupakan rasio dari *support* yang diobservasi jika menganggap X dan Y merupakan *item*/*itemset* yang terpisah. Nilai *lift* dapat disebut sebagai nilai kepentingan dari sebuah *rule*. Rentang nilai *lift* mulai dari nol (0) hingga tak terbatas. Jika nilai *lift* lebih rendah dari satu (1) maka *rule* yang dihasilkan kurang relevan atau hubungan antara *itemset* X dan Y tidak terlalu sering muncul atau sangat jarang ditemukan bersama. Jika nilai *lift* pada suatu *rule* mendekati satu (1), maka dapat diprediksi bahwa hubungan *itemset* X dengan Y akan mungkin sering muncul. Jika nilai *lift* lebih besar dari satu (1), maka dapat diprediksi jika *itemset* X akan lebih sering muncul diiringi dengan *item/itemset* Y dalam sebuah *rule* [8].

2.3 Algoritma Apriori

Apriori adalah algoritma seminal yang diusulkan oleh R. Agrawal dan R. Srikant pada tahun 1994 untuk mining frequent itemset untuk aturan asosiasi Boolean [9]. Hal yang membedakan Apriori dengan metode asosiasi yang lain adalah algoritma ini mengambil fakta jika subset dari *itemset* yang sering muncul dapat dianggap sebagai *itemset* yang sering muncul juga. Sehingga pada setiap iterasi jika ada subset yang jarang muncul atau tidak memenuhi kriteria *support* dan *confidence* minimum pada proses pembuatan *rules* dapat dihapus dan tidak digunakan pada iterasi selanjutnya. Akibatnya proses pembuatan *rules* menggunakan algoritma Apriori dapat menghemat waktu karena terpotongnya dataset pada setiap iterasi [6].



Gambar 2 Ilustrasi dari algoritma Apriori
Sumber (http://www.saedsayad.com/association_rules.htm)

2.4 Data Pokok Pendidikan (Dapodik)

Dapodik atau Data Pokok Pendidikan adalah sistem pendataan skala nasional yang terpadu, dan merupakan sumber data utama pendidikan nasional, yang merupakan bagian dari Program perancanaan pendidikan nasional dalam mewujudkan insan Indonesia yang Cerdas dan Kompetitif. Karena tanpa perencanaan pendidikan yang matang, maka seluruh program yang terbentuk dari perencanaan tersebut akan jauh dari tujuan yang diharapkan.

Untuk melaksanakan perencanaan pendidikan, maupun untuk melaksanakan program-program pendidikan secara tepat sasaran, dibutuhkan data yang cepat, lengkap, valid, akuntabel dan terus *up to date*.

Dengan ketersediaan data yang cepat, lengkap, valid, akuntabel dan up to date tersebut, maka proses perencanaan, pelaksanaan, pelaporan dan evaluasi kinerja program-program pendidikan nasional dapat dilaksanakan dengan lebih terukur, tepat sasaran, efektif, efisien dan berkelanjutan. Sehubungan dengan hal tersebut, Departemen Pendidikan Nasional telah mengembangkan suatu sistem pendataan skala nasional yang terpadu yang disebut dengan Data Pokok Pendidikan (Dapodik) [10].

2.5 Fasilitas Sekolah

Fasilitas Sekolah menurut Dimyati dan Mudjiono [11] adalah sarana dan prasarana pembelajaran. Prasarana meliputi gedung sekolah, ruang belajar, lapangan olahraga, ruang ibadah, ruang kesenian dan peralatan olah raga. Sarana pembelajaran meliputi buku pelajaran, buku bacaan, alat dan fasilitas laboraturium sekolah dan berbagai media pembelajaran yang lain. Sedangkan menurut Keputusan Menteri P dan K No. 079/1975 [12], fasilitas belajar terdiri dari 3 kelompok besar yaitu:

1. Bangunan dan perabot sekolah

Bangunan di sekolah pada dasarnya harus sesuai dengan kebutuhan pendidikan dan harus layak untuk ditempati siswa pada proses kegiatan belajar mengajar di sekolah. Bangunan sekolah terdiri atas berbagai macam ruangan. Secara umum jenis ruangan ditinjau dari fungsinya dapat dikelompokkan dalam ruang pendidikan untuk menampung proses kegiatan belajar mengajar baik teori maupun praktek, ruang administrasi untuk proses administrasi sekolah dan berbagai kegiatan kantor, dan ruang penunjang untuk kegiatan yang mendukung proses belajar mengajar. Sedangkan perabot sekolah yang pada umumnya terdiri dari berbagai jenis mebel, harus dapat mendukung semua semua kegiatan yang berlangsung di sekolah, baik kegiatan belajar mengajar maupun kegiatan administrasi sekolah.

2. Alat pelajaran

Alat pelajaran yang dimaksudkan disini adalah alat peraga dan buku-buku bahan ajar. Alat peraga berfungsi untuk memperlancar dan memperjelas komunikasi dalam proses belajar mengajar antara guru dan siswa. Buku-buku pelajaran yang digunakan dalam kegiatan belajar mengajar, biasanya terdiri dari buku pegangan, buku pelengkap, dan buku bacaan.

3. Media pendidikan

Media pengajaran merupakan sarana non personal yang digunakan atau disediakan oleh tenaga pengajar yang memegang peranan dalam proses belajar untuk mencapai tujuan instruksional. Media pengajaran dapat dikategorikan dalam media visual yang menggunakan proyeksi, media auditif, dan media kombinasi.

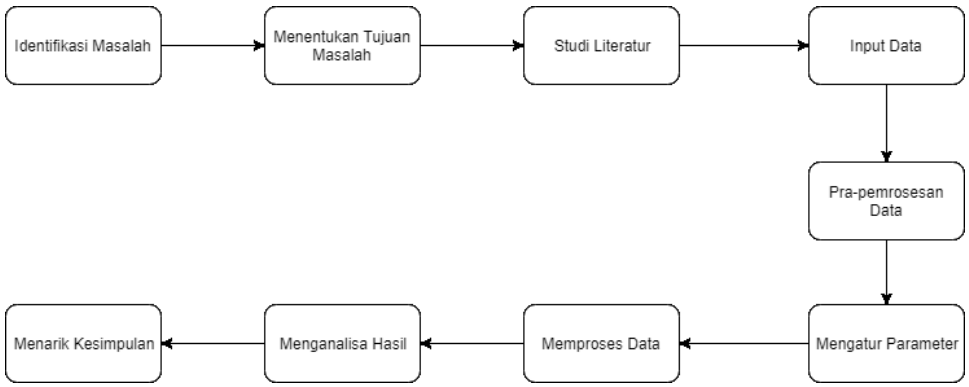
2.6 Penelitian Terkait

Penelitian yang terkait dengan penelitian ini adalah penelitian dari Dewi Setianingsih, RB Fajriya Hakim tahun 2015 tentang Penerapan Data Mining Dalam Analisis Kejadian Tanah Longsor Di Indonesia Dengan Menggunakan Association Rule Algoritma Apriori, dan penelitian Mohamad Fauzy, Kemas Rahmat Saleh W, Ibnu Asror tahun 2016 tentang Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada

Simulasi Prediksi Hujan Wilayah Kota Bandung. Meskipun dengan kasus yang berbeda, kedua penelitian tersebut dapat dijadikan panduan dalam penelitian ini.

3 Penelitian yang Dilakukan

Tahapan penelitian adalah tahapan yang dilakukan penulis dalam melakukan penelitian ini. Berikut adalah tahapan penelitian "*Analisis Hubungan Fasilitas Sekolah Dengan Perolehan Nilai UN Menggunakan Association rule Mining*".



Gambar 3 Tahapan Penelitian

Identifikasi masalah dimulai dengan mencari masalah yang ingin diteliti. Masalah yang ditemukan adalah ketidakmerataan fasilitas sekolah di Indonesia, apakah ketidakmerataan tersebut berpengaruh kepada nilai UN yang dihasilkan pada sekolah tersebut. Berdasarkan masalah yang ditemukan, maka perlu dicari apa tujuan penelitian yang akan dilakukan. Tujuan penelitian harus berhubungan dengan masalah yang telah ditemukan.

Studi literatur dilakukan untuk mencari tahu penelitian-penelitian apa yang telah dilakukan sebelumnya yang berhubungan dengan penelitian yang akan dilakukan. Studi literatur juga dilakukan untuk mempelajari metode-metode yang akan digunakan pada penelitian ini. Input data dilakukan dengan mengumpulkan data yang tersedia dari sumber data. Dari sumber data yang ditemukan, data diinput secara manual dengan menyalin data satu per satu dari sumber data. Prapemrosesan data dilakukan untuk menjadikan data yang telah dikumpulkan menjadi data yang bisa diproses. Prapemrosesan data dilakukan karena data yang diinput tidak semuanya bisa digunakan untuk dilakukan proses Apriori.

Parameter diatur sesuai dengan parameter yang ada di algoritma Apriori. Parameter yang diatur adalah nilai minimum *support*, minimum *confidence*, minlen, dan maxlen. Data yang sudah dipraproses akhirnya diproses menggunakan algoritma Apriori dengan parameter yang sudah ditentukan. Pemrosesan Apriori ini dilakukan menggunakan R Studio dengan plugin Arules. Hasil yang didapatkan dari proses Apriori sebelumnya dianalisa untuk mencari hubungan antara fasilitas sekolah dengan nilai UN. Dari hasil Analisa maka akan ditentukan apakah diperlukan percobaan selanjutnya atau langsung menarik kesimpulan. Setelah melakukan percobaan dengan segala hasil yang didapatkan, maka dapat dilakukan penarikan kesimpulan. Penarikan kesimpulan harus bersifat objektif terhadap hasil percobaan.

4 Pembahasan

4.1 Sistematika Percobaan

Proses percobaan yang dilakukan dimulai dengan input data. Hal ini dilakukan secara manual dengan menyalin data secara manual di situs Sekolah Kita. Data sekolah dicari berdasarkan kecamatan yang terdaftar di situs tersebut, dengan mengecualikan daerah luar negeri dikarenakan penulis ingin fokus terhadap kondisi sekolah di Indonesia.

4.2 Perangkat yang Digunakan

Berikut adalah perangkat-perangkat yang digunakan dalam penelitian ini:
Perangkat Keras:
Laptop ASUS GL502VS
- Processor: Intel Core i7 6700HQ
- RAM: 16 GB
- Storage: HDD 1 TB, SSD NVME 512 GB

Perangkat Lunak:
- Operating System: Windows 10 Home 64-bit build 17134

- Text Editor: Visual Code Studio
- Spreadsheet Editor: Microsoft Office Excel 365, Google Spreadsheet
- IDE: JetBrains PyCharm 2019.1.1
- RStudio 1.1.463

4.3 Karakteristik Data

Dataset diperoleh dengan input manual dari website Sekolah Kita. Dataset yang digunakan pada percobaan ini berupa data 1700 sekolah yang terdiri dari 50 sekolah per provinsi. Berikut adalah karakteristik dari data yang diambil dari Sekolah Kita:

4.3.1 Data Fasilitas Sekolah

Berdasarkan definisi fasilitas sekolah sebelumnya, yang dimaksud fasilitas sekolah adalah bangunan dan perabot sekolah, alat belajar dan media pembelajaran. Akan tetapi data yang diambil pada situs Sekolah Kita terbatas pada data berikut.

Tabel 1 Data fasilitas sekolah yang didapatkan melalui situs Sekolah Kita

No.	Nama Kategori	Penjelasan	Atribut Data
1	Akses Internet	Ada atau tidaknya akses internet pada sekolah tersebut.	Nominal
2	Sumber Listrik	Ada atau tidaknya sumber listrik pada sekolah tersebut.	Nominal
3	Daya Listrik	Besar daya listrik pada sekolah.	Ratio
4	Luas Tanah	Besar luas tanah yang dimiliki sekolah.	Ratio
5	Ruang Kelas	Jumlah ruangan kelas (kondisi layak dan rusak ringan) yang dimiliki sekolah.	Ratio
6	Laboratorium	Jumlah laboratorium yang dimiliki sekolah.	Ratio
7	Perpustakaan	Jumlah perpustakaan yang dimiliki sekolah.	Ratio
8	Sanitasi Siswa	Jumlah sanitasi siswa / toilet siswa yang dimiliki sekolah.	Ratio
9	Persentasi Kelas Layak	Jumlah persentasi (%) ruangan kelas yang layak. Persentasi ini sama dengan jumlah data ruang kelas – ruang kelas yang rusak.	Interval

4.3.2 Data Non-fasilitas Sekolah

Selain data fasilitas sekolah, diperlukan juga data non-fasilitas sekolah yang diperkirakan memiliki hubungan dengan nilai UN.

Tabel 2 Data non-fasilitas sekolah yang didapatkan dari situs Sekolah Kita.

No.	Nama Kategori	Penjelasan	Atribut Data
1	ID Sekolah	NPSN (Nomor Pokok Sekolah Nasional).	Nominal
2	Guru	Jumlah guru dalam satu sekolah.	Ratio
3	Rombongan Belajar	Jumlah rombongan belajar dalam satu sekolah (total dari kelas VII, VIII, dan IX).	Ratio
4	Kurikulum	Jenis kurikulum yang digunakan sekolah (K-13 atau KTSP).	Nominal
5	Penyelenggaraan	Waktu penyelenggaraan kegiatan belajar mengajar.	Nominal
6	Manajemen Berbasis Sekolah	Ada atau tidaknya aplikasi manajemen berbasis sekolah pada sekolah tersebut.	Nominal
7	Persentasi Guru Kualifikasi	Jumlah persentasi (%) guru sekolah yang memiliki kualifikasi sebagai guru.	Interval
8	Persentasi Guru Sertifikasi	Jumlah persentasi (%) guru sekolah yang telah memiliki sertifikasi.	Interval
9	Persentasi Guru PNS	Jumlah persentasi (%) guru yang berstatus PNS.	Interval
10	Provinsi	Lokasi provinsi dimana sekolah tersebut berada.	Nominal
11	Status	Status sekolah negeri atau swasta.	Nominal
12	Akreditasi	Status akreditasi sekolah pada waktu data diinput.	Nominal

4.3.3 Data Nilai UN

Terakhir adalah data nilai UN yang menjadi variabel tetap dari penelitian ini.

Tabel 3 Data Nilai UN yang didapatkan dari situs Sekolah Kita

No.	Nama Kategori	Penjelasan	Atribut Data
1	Bahasa Indonesia	Nilai rata-rata UN Bahasa Indonesia pada satu tahun terakhir.	Interval
2	Bahasa Inggris	Nilai rata-rata UN Bahasa Inggris pada satu tahun terakhir.	Interval
3	Matematika	Nilai rata-rata UN Matematika pada satu tahun terakhir.	Interval
4	IPA	Nilai rata-rata UN IPA pada satu tahun terakhir.	Interval

Selain itu dataset yang dikumpulkan masih ada beberapa *missing value* dan beberapa data yang tidak valid. *Missing value* adalah data "kosong" yang biasanya ada dalam sebuah kumpulan data. *Missing value* biasanya bernilai nol atau kosong tanpa nilai, dan dapat mempengaruhi hasil dari pengolahan data. Sementara data yang tidak valid dapat ditentukan dengan menganalisa hubungan nilai suatu data dengan data lain atau menganalisa hubungan nilai suatu data dengan kejadian nyata. Dalam dataset yang telah dikumpulkan, ditemukan beberapa *missing value* dan data yang tidak valid. Berikut adalah detail dari data tersebut.

1) *Missing value*

Tabel 4 Jumlah *missing value* pada dataset yang digunakan

Data	Jumlah *Missing Value*	Presentase
Jenis Penyelenggaraan	28	1.647%
Kurikulum	1	0.058%

Dalam pencarian *missing value* dalam dataset ada beberapa pengecualian dimana data dapat bernilai nol (0). Hal ini dikarenakan selain data yang dimaksud pada Tabel 4, semua data yang ada dalam dataset memiliki arti jika suatu sekolah tidak memiliki fasilitas tersebut. Misalnya sebuah sekolah mungkin saja tidak memiliki laboratorium, santisasi siswa, perpustakaan, bahkan listrik.

2) Data tidak valid

Tabel 5 Jumlah data tidak valid pada dataset yang digunakan

Data	Jumlah Data Tidak Valid	Presentase
Daya Listrik dan Sumber Listrik	101	5.941%
LuasTanah	80	4.706%

Ada alasan kenapa data pada 5 dianggap tidak valid. Untuk data daya listrik dan sumber listrik, alasannya adalah keterkaitan antara dua data tersebut. Dalam beberapa baris data dalam dataset yang telah dikumpulkan, ditemukan data yang menyebutkan bahwa sebuah sekolah memiliki sumber listrik tetapi nilai data daya listriknya nol (0), atau berisi nilai yang tidak lazim seperti daya listrik bernilai 1, 2, 1234, dan sebagainya. Sementara untuk data luas tanah, hal yang menyebabkan data tersebut tidak valid adalah ketidaklaziman nilai pada data tersebut. Ketidaklaziman yang dimaksud adalah jika dikaitkan dengan kejadian nyata. Beberapa dari data luas tanah tidak valid berisi nol (0), tetapi ada juga yang berisi nilai-nilai yang tidak mungkin untuk luas tanah sekolah seperti 60, 19, 1, dan sebagainya.

4.4 Siklus Percobaan

4.4.1 Percobaan Pertama

Pada *pra-processing data* pertama dilakukan analisis secara statistik dari data yang telah diinput. Hal ini dilakukan untuk mengetahui karakteristik dari dataset, selain itu dapat membantu dalam konversi data untuk data yang berjenis *ordinal*, hingga data *ratio*. Data *ordinal* dan data *ratio* yang berbentuk numerik kurang cocok untuk dilakukan analisis menggunakan algoritma Apriori. Hal ini dikarenakan perbedaan satu digit pada data tersebut dapat dianggap sebagai rule yang berbeda sehingga perlu adanya pengkategorian untuk data yang bersifat tersebut. Berikut adalah hasil statistik dari data yang telah dilakukan terhadap dataset.

Tabel 5 Hasil Statistik Deskriptif dari dataset

Data	Maximum	Minimum	Mean	Modus	Standar Deviasi
Guru	86	0	20.45	8	15.1563
Rombongan Belajar	47	1	10.7929	3	8.6861
Kurikulum	-	-	-	K-13	-
Penyelenggaraan	-	-	-	Pagi/6h	-
Akses Internet	1	0	0.5082	1	0.5
Sumber Listrik	1	0	0.97	1	0.1706
Daya Listrik	560000	0	6863.48	900	22108.94
Luas Tanah	1000000	0	9534.88	10000	26540.71
Ruang Kelas	66	0	10.2158	3	8.4347
Laboratorium	7	0	0.8788	1	0.8216
Perpustakaan	3	0	0.7588	1	0.4773
Sanitasi Siswa	25	0	1.8229	2	2.2419
Presentase Guru Kualifikasi	100	0	93.13	100	11.1187
Presentase Guru Sertifikasi	100	0	38.7255	0	27.2419
Presentase Guru PNS	100	0	49.1853	0	31.6044
Presentase Kelas Layak	100	0	89.08	100	24.3018
Bahasa Indonesia	89.87	31.43	61.0685	60	10.004
Bahasa Inggris	90.06	29.25	47.9501	37.33	12.049
Matematika	96.84	19.9	47.1471	32.5	14.907
IPA	88.56	26.09	48.6415	37.5	11.9732
Status	-	-	-	Negeri	-
Akreditasi	-	-	-	B	-

Lepas dari hasil statistik di atas, diperlukan juga analisis dari dataset per kategori / kolom untuk membantu pembuatan metode konversi data. Dari analisis per kategori, ternyata ada tiga kategori yang tidak perlu dipakai dalam dataset Apriori. Berikut tiga kategori yang dikeluarkan dari dataset:

- ID Sekolah: ID Sekolah hanya digunakan sebagai penanda bahwa setiap baris data itu unik, sehingga tidak perlu digunakan dalam dataset.
- Manajemen Berbasis Sekolah: Manajemen Berbasis Sekolah awalnya dimasukkan ke dalam dataset dengan harapan jika ada sistem informasi aplikasi Manajemen Berbasis Sekolah dapat mempengaruhi nilai Ujian Nasional. Akan tetapi dikarenakan sebanyak 1700 data yang diinput, tidak ada satupun sekolah yang memiliki aplikasi Manajemen Berbasis Sekolah sehingga tidak perlu dimasukan ke dalam dataset.
- Provinsi: Data Provinsi hanya digunakan sebagai penanda bahwa data tersebut diambil dari daerah provinsi tertentu dan juga sebagai penanda bahwa ada lima puluh data per provinsi. Target lingkup penelitian ini adalah data seluruh Indonesia secara umum tanpa memandang regional. Maka dari itu data provinsi tidak perlu dimasukkan dalam dataset.

Setelah mengeluarkan tiga kategori dari dataset, maka dilakukan analisis mendalam per kategori. Analisis ini dilakukan dengan menghitung jumlah data per kategori, lalu membuat parameter untuk mengelompokkan data tersebut.

Berdasarkan tabel-tabel di atas, maka dibuat aturan konversi data seperti di bawah ini.

1. Guru:
   a. Guru_A: jumlah guru > 30
   b. Guru_B: 30 ≥ jumlah guru ≥ 16
   c. Guru_C: 15 ≥ jumlah guru ≥ 10
   d. Guru_D: 10 > jumlah guru > 0
   e. Guru_0: jumlah guru = 0
2. Rombel:

   a. Rombel_A: jumlah rombel > 15
   b. Rombel_B: 15 ≥ jumlah rombel ≥ 11
   c. Rombel_C: 10 ≥ jumlah rombel ≥ 5
   d. Rombel_D: 5 > jumlah rombel > 0
3. Kurikulum, tidak ada perubahan, karena cuma ada dua entri data yaitu K-13 dan KTSP.
4. Penyelenggaraan, tidak ada perubahan selain yang tidak ada data dianggap "Penyelenggaraan_Null"
5. Internet:
   a. No_Int: Tidak ada internet
   b. Int: Ada internet
6. Sumber Listrik:
   a. List: Ada sumber listrik
   b. No_List: Tidak ada sumber listrik
7. Daya Listrik:
   a. DL_A: Daya listrik > 5000 Watt
   b. DL_B: 5000 Watt ≥ Daya listrik ≥ 2001 Watt
   c. DL_C: 2000 Watt ≥ Daya listrik ≥ 901 Watt
   d. DL_D: 900 Watt > Daya listrik ≥ 0 Watt
   e. DL_0: Daya listrik = 0 Watt
8. Luas Tanah:
   a. LT_A: Luas tanah > 12000 M^2
   b. LT_B: 12000M^2 ≥ Luas tanah ≥ 6001 M^2
   c. LT_C: 6000 M^2 ≥ Luas tanah ≥ 2001 M^2
   d. LT_D: 0 M^2 > Luas tanah ≥ 2000 M^2
   e. LT_0: Luas tanah = 0 M^2
9. Ruang Kelas:
   a. RK_A: Ruang kelas > 15
   b. RK_B: 15 ≥ Ruang kelas ≥ 11
   c. RK_C: 10 ≥ Ruang kelas ≥ 5
   d. RK_D: 5 > Ruang kelas > 0
   e. RK_0: Ruang kelas = 0
10. Laboratorium:
   a. Lab_0: Laboratorium = 0
   b. Lab_1: Laboratorium = 1
   c. Lab_2: Laboratorium = 2
   d. Lab_3: Laboratorium = 3
   e. Lab_4: Laboratorium = 4
   f. Lab_5: Laboratorium = 5
   g. Lab_6: Laboratorium = 6
   h. Lab_7: Laboratorium = 7
   i. Lab_x: Laboratorium > 7
11. Perpustakaan:
   a. Pus_0: Perpustakaan = 0
   b. Pus_1: Perpustakaan = 1
   c. Pus_2: Perpustakaan = 2
   d. Pus_3: Perpustakaan = 3
   e. Pus_x: Perpustakaan > 3
12. Sanitasi Siswa:
   a. San_A: Sanitasi siswa > 20
   b. San_B: 20 ≥ Sanitasi siswa ≥ 16
   c. San_C: 15 ≥ Sanitasi siswa ≥ 11
   d. San_D: 10 ≥ Sanitasi siswa ≥ 6
   e. San_E: 5 ≥ Sanitasi siswa ≥ 1
   f. San_0: Sanitasi siswa = 0
13. Guru Kualifikasi:
   a. Kualifikasi_A: 100 ≥ Guru kualifikasi > 80
   b. Kualifikasi_B: 80 ≥ Guru kualifikasi > 60
   c. Kualifikasi_C: 60 ≥ Guru kualifikasi > 40
   d. Kualifikasi_D: 40 ≥ Guru kualifikasi > 20
   e. Kualifikasi_E: 20 ≥ Guru kualifikasi > 0
   f. Kualifikasi_0: Guru kualifikasi = 0

14. Guru Sertifikasi:
    a. Sertifikasi_A: 100 ≥ Guru sertifikasi > 80
    b. Sertifikasi_B: 80 ≥ Guru sertifikasi > 60
    c. Sertifikasi_C: 60 ≥ Guru sertifikasi > 40
    d. Sertifikasi_D: 40 ≥ Guru sertifikasi > 20
    e. Sertifikasi_E: 20 ≥ Guru sertifikasi > 0
    f. Sertifikasi_0: Guru sertifikasi = 0
15. Guru PNS:
    a. PNS_A: 100 ≥ Guru PNS > 80
    b. PNS_B: 80 ≥ Guru PNS > 60
    c. PNS_C: 60 ≥ Guru PNS > 40
    d. PNS_D: 40 ≥ Guru PNS > 20
    e. PNS_E: 20 ≥ Guru PNS > 0
    f. PNS_0: Guru PNS = 0
16. Kelas Layak:
    a. KL_A: 100 ≥ Kelas layak > 80
    b. KL_B: 80 ≥ Kelas layak > 60
    c. KL_C: 60 ≥ Kelas layak > 40
    d. KL_D: 40 ≥ Kelas layak > 20
    e. KL_E: 20 ≥ Kelas layak > 0
    f. KL_0: Kelas layak = 0
17. Bahasa Indonesia:
    a. BIndo_A: 100 ≥ Bahasa Indonesia > 80
    b. BIndo_B: 80 ≥ Bahasa Indonesia > 60
    c. BIndo_C: 60 ≥ Bahasa Indonesia > 40
    d. BIndo_D: 40 ≥ Bahasa Indonesia > 20
    e. BIndo_E: 20 ≥ Bahasa Indonesia > 0
    f. BIndo_0: Bahasa Indonesia = 0
18. Bahasa Inggris:
    a. BInggris_A: 100 ≥ Bahasa Inggris > 80
    b. BInggris_B: 80 ≥ Bahasa Inggris > 60
    c. BInggris_C: 60 ≥ Bahasa Inggris > 40
    d. BInggris_D: 40 ≥ Bahasa Inggris > 20
    e. BInggris_E: 20 ≥ Bahasa Inggris > 0
    f. BInggris_0: Bahasa Inggris = 0
19. Matematika:
    a. MTK_A: 100 ≥ Matematika > 80
    b. MTK_B: 80 ≥ Matematika > 60
    c. MTK_C: 60 ≥ Matematika > 40
    d. MTK_D: 40 ≥ Matematika > 20
    e. MTK_E: 20 ≥ Matematika > 0
    f. MTK_0: Matematika = 0
20. IPA:
    a. IPA_A: 100 ≥ IPA > 80
    b. IPA_B: 80 ≥ IPA > 60
    c. IPA_C: 60 ≥ IPA > 40
    d. IPA_D: 40 ≥ IPA > 20
    e. IPA_E: 20 ≥ IPA > 0
    f. IPA_0: IPA = 0
21. Status: Tidak perlu ada perubahan karena sudah terbagi menjadi dua kategori.
22. Akreditasi:
    a. Ak_A: Akreditasi A
    b. Ak_B: Akreditasi B
    c. Ak_C: Akreditasi C
    d. Ak_Tidak/Belum: Tidak/Belum Terakreditasi

Aturan konversi data tersebut sebagian besar masih dibuat berdasarkan perkiraan mentah semata dan statistik dasar tanpa memperhatikan hubungan antar data. Sementara untuk data *missing value* dan data tidak valid tidak dilakukan tindakan perubahan dikarenakan presentase data tersebut tergolong rendah dibawah 10 persen.

Parameter tuning dalam percobaan ini adalah pengaturan *hyperparameter* yang akan digunakan pada algoritma Apriori. Dalam kasus ini ada empat *hyperparameter* yang diatur dalam percobaan ini.

a. Minlen: *Minimum length*, atau panjang aturan minimum dari aturan yang akan dibuat oleh algoritma Apriori. Secara *default*, minlen bernilai 0.
b. Maxlen: *Maximum length*, atau Panjang aturan maximum dari aturan yang akan dibuat oleh algoritma Apriori. Secara *default*, maxlen bernilai 8.
c. Support: Nilai minimum *support* yang dibutuhkan yang dibutuhkan dalam membuat aturan.
d. Confidence: Nilai minimum *confidence* yang dibutuhkan dalam membuat aturan.

Pada percobaan pertama dilakukan empat pengaturan. Berikut adalah hasil pengaturan parameter yang digunakan pada percobaan pertama.

1. support=0.1, confidence=0.1, minlen=4, maxlen=22
2. support=0.05, confidence=0.05, minlen=4, maxlen=22
3. support=0.03, confidence=0.03, minlen=4, maxlen=22
4. support=0.02, confidence=0.02, minlen=4, maxlen=22

Alasan mengapa minlen ditentukan dengan nilai tetap adalah untuk mencari tahu aturan pada mata pelajaran yang diujiankan dalam UN yaitu Bahasa Indonesia, Bahasa Inggris, Matematika dan IPA. Total jumlah mata pelajaran yang diujikan adalah empat mata pelajaran, maka nilai minlen adalah empat. Hal ini dilakukan karena diharapkan setidaknya akan menghasilkan aturan yang ada didalamnya semua mata pelajaran yang diuji tersebut. Sementara alasan maxlen kenapa nilainya selalu sama karena jumlah *item* dalam satu *itemset* adalah 22 *item*, sehingga diharapkan akan muncul aturan yang menampilkan semua *item-item* tersebut. Untuk nilai support dan confidence yang dibuat pada percobaan pertama masih ditentukan secara sembarang.

Dari parameter yang telah ditentukan pada proses sebelumnya, maka dilakukan pemrosesan data menggunakan RStudio dengan *plugin* Arules. Hasil dari pemrosesan data adalah sebagai berikut.

Tabel 6 Hasil pemrosesan data percobaan pertama

Rule	Support	Confidence	Minlen	Maxlen	Total Rules
1	0.1	0.1	4	22	123323
2	0.05	0.05	4	22	1120309
3	0.03	0.03	4	22	4784115
4	0.02	0.02	4	22	13824346

Berdasarkan hasil yang didiapatkan dari pemrosesan data, aturan yang dihasilkan cukup banyak sehingga diperlukan penyaringan hasil aturan. Untuk menyaring hasil aturan dilakukan pembuatan *subset* dari aturan yang telah dibuat. Pembuatan subset dilakukan dengan mengambil aturan yang memiliki mata pelajaran yang diujikan dalam UN. Dikarenakan *item* untuk mata pelajaran UN telah dikelompokkan berdasarkan kategori nilainya jadi pembuatan subset dilakukan dengan melakukan kombinasi subset mata pelajaran UN satu per satu. Berikut hasil dari aturan yang telah didapatkan berdasarkan mata pelajaran UN.

Tabel 7 Pola aturan subset yang berhasil diekstrak pada percobaan pertama

No	Pola Aturan Subset	Kode Pola	Subset Rule
1	BIndo_B, BInggris_B, MTK_B, IPA_B	BBBB	2, 3, 4
2	BIndo_B, BInggris_B, MTK_B, IPA_C	BBBC	4
3	BIndo_B, BInggris_C, MTK_B, IPA_B	BCBB	4
4	BIndo_B, BInggris_C, MTK_C, IPA_C	BCCC	1, 2, 3, 4
5	BIndo_B, BInggris_C, MTK_D, IPA_C	BCDC	2, 3, 4
6	BIndo_B, BInggris_D, MTK_C, IPA_C	BDCC	4
7	BIndo_B, BInggris_D, MTK_D, IPA_C	BDDC	3, 4
8	BIndo_C, BInggris_C, MTK_C, IPA_C	CCCC	3, 4
9	BIndo_C, BInggris_C, MTK_D, IPA_C	CCDC	4
10	BIndo_C, BInggris_C, MTK_D, IPA_D	CCDD	4
11	BIndo_C, BInggris_D, MTK_D, IPA_C	CDDC	3, 4
12	BIndo_C, BInggris_D, MTK_D, IPA_D	CDDD	1, 2, 3, 4

Dari pola aturan subset-subset di atas, ditemukan bahwa tidak semua subset ditemukan pada satu *rule.* Hal ini disebabkan oleh parameter *support* yang berbeda pada setiap *rule*. Semakin rendah nilai minimum *support*, maka semakin banyak *rules* yang dihasilkan oleh algoritma ini. Dengan *rule* 4 sebagai *rule* yang memiliki nilai minimum *support* terkecil menghasilkan semua subset yang telah ditemukan. Karena *rule* 4 mencakup semua *subset* yang dicari, maka pengaturan *support* yang rendah akan digunakan pada percobaan berikutnya. Berikut adalah contoh salah satu *subset* pada pola BBBB yang telah diurutkan berdasarkan 10 nilai *confidence* tertinggi.

	lhs		rhs	support	confidence	lift	count
[1]	{BIndo_B,BInggris_B,IPA_B,Lab_1,MTK_B}	=>	{List}	0.02176471	1	1.030928	37
[2]	{BIndo_B,BInggris_B,Int,IPA_B,MTK_B}	=>	{List}	0.02529412	1	1.030928	43
[3]	{BIndo_B,BInggris_B,IPA_B,LT_A,MTK_B,Negeri}	=>	{K-13}	0.02176471	1	1.061836	37
[4]	{BIndo_B,BInggris_B,IPA_B,Kualifikasi_A,Lab_1,MTK_B}	=>	{List}	0.02000000	1	1.030928	34
[5]	{BIndo_B,BInggris_B,IPA_B,K-13,Lab_1,MTK_B}	=>	{List}	0.02000000	1	1.030928	34
[6]	{BIndo_B,BInggris_B,Int,IPA_B,MTK_B,Negeri}	=>	{List}	0.02235294	1	1.030928	38
[7]	{BIndo_B,BInggris_B,Int,IPA_B,Kualifikasi_A,MTK_B}	=>	{List}	0.02352941	1	1.030928	40
[8]	{BIndo_B,BInggris_B,Int,IPA_B,K-13,MTK_B}	=>	{List}	0.02352941	1	1.030928	40
[9]	{BIndo_B,BInggris_B,IPA_B,MTK_B,Pus_1,San_E}	=>	{List}	0.02764706	1	1.030928	47
[10]	{BIndo_B,BInggris_B,IPA_B,KL_A,MTK_B,San_E}	=>	{List}	0.03176471	1	1.030928	54

Gambar 4 Daftar 10 hasil aturan pola BBBB berdasarkan confidence tertinggi

Berdasarkan hasil *filter* di atas dapat diketahui bahwa *support* tertinggi dari semua *subset* yang diteliti hanya sebesar 0.20 atau sekitar 20%. Selain itu nilai *support* hanya merepresentasikan perbandingan *rule* terhadap total *item* dalam *itemset* keseluruhan. Dengan begitu *rule* yang dapat dianggap bagus bergantung pada nilai *confidence* dan *lift*. Dalam percobaan pertama, penulis menentukan kriteria *rule* yang dapat diterima:

$$1 > confidence \geq 0.7$$

Alasan dalam penentuan aturan tersebut disebabkan karena batas *confidence* dalam menerima sebuah *rule* tidak ada aturan pasti, maka setidaknya batas *confidence* minimum harus cukup besar setidaknya 0.7 atau 70%. Akan tetapi algoritma Apriori dijelaskan sebagai kemunculan Bersama *itemset* X dengan *item* Y secara bersamaan dengan dilambangkan sebagai X→Y, dan dalam Apriori nilai *confidence* adalah nilai persentase kemunculan *item* Y terhadap total kemunculan *itemset* X. Hal ini berarti jika kemunculan X dalam sebuah *rule* sebanyak 100 kemunculan, dengan nilai *confidence* sebesar 0.7 maka kemunculan Y dalam X→Y sebanyak 70 kemunculan. Dengan begitu meskipun nilai *support* dalam sebuah *rule* rendah, jika nilai *confidence* cukup tinggi maka *rule* tersebut termasuk bagus. Selain itu dalam hasil yang telah ditampilkan pada gambar di atas [lhs] = X dan [rhs] = Y.

Berdasarkan aturan yang telah ditentukan maka bisa disimpulkan bahwa semua *rules* berdasarkan 10 *rules* terbaik berdasarkan *confidence* merupakan *rules* yang dapat diterima dengan nilai *confidence* terendah yang ditemukan adalah sebesar 0.89. Akan tetapi selain dari parameter *confidence* yang telah ditentukan, perlu adanya tim ahli dalam bidang pendidikan nasional untuk menentukan *rules* yang benar-benar dapat diterima atau tidak. Dari beberapa hasil *rules* di atas, ada beberapa *subsets* yang menjadi perhatian, yaitu *subsets* yang menghasilkan *rules* dengan *confidence* sama dengan 1. Berikut adalah *subsets* tersebut.

Tabel 8 Daftar *subset* yang memiliki *confidence* = 1 pada percobaan pertama

***Subset* yang memiliki *confidence* = 1**
BBBB
BCBB
BCCC
BCDC
BDCC
BDDC
CCCC
CCDC
CCDD
CDDC
CDDD

Secara khusus, semua *item* kecuali atribut guru, dan akreditasi dalam *itemset* berpengaruh pada pembuatan *rules* pada percobaan pertama. Hal ini dibuktikan dengan munculnya semua *item* pada semua *subsets*, meskipun tidak muncul secara serempak dalam satu *rule* tetapi dengan menyebar secara parsial ke semua *subset*. Hal ini membuktikan adanya hubungan data fasilitas pada sekolah terhadap nilai UN.

Tabel 9 Daftar kemunculan atribut dalam 10 *rules* terbaik berdasarkan *confidence* pada percobaan pertama

Atribut	**Frekuensi Kemunculan dalam 10 *rules* terbaik**	**Paling Banyak Muncul Pada Subset**
Guru	0	-
Rombongan Belajar	2	CDDD

Kurikulum	35	BCBB
Penyelenggaraan	5	CDDC
Akses Internet	5	BBBB
Sumber Listrik	81	BDDC, CCCC, CCDC, CDDC
Daya Listrik	2	CDDD
Luas Tanah	1	BBBB
Ruang Kelas	4	CDDD
Laboratorium	13	CCCC
Perpustakaan	7	CCCC, CDDC
Sanitasi Siswa	8	CDDC
Presentase Guru Kualifikasi	32	BCBB
Presentase Guru Sertifikasi	9	BCDC
Presentase Guru PNS	7	BCDC
Presentase Kelas Layak	10	BCBB
Status	25	BCBB, CCDC, CCDD
Akreditasi	0	-

Berdasarkan 9 atribut sumber listrik memiliki frekuensi kemunculan paling banyak dalam 10 *rules* terbaik berdasarkan *confidence*. Jika berdasarkan frekuensi kemunculan maka dapat disimpulkan bahwa sumber listrik menjadi atribut yang paling berpengaruh dalam 10 *rules* terbaik berdasarkan *confidence*. Selain itu atribut jumlah guru dan akreditasi sekolah menjadi atribut yang paling tidak berpengaruh karena tidak muncul dalam 10 *rules* terbaik berdasarkan *confidence*.

4.5 Percobaan Kedua

Pada percobaan kedua praproses data yang dilakukan adalah dengan mengaitkan hubungan antara data Rombel dengan data Guru, dan data Rombel dengan data Ruang Kelas. Hubungan antara data Rombel dengan data Guru adalah dengan mencari selisih dari data Guru dengan data Rombel. Alasan dari tindakan ini adalah untuk mencari tahu keterkaitan antara jumlah guru dengan jumlah rombongan belajar. Selain itu hal ini dilakukan karena pada percobaan pertama atribut jumlah guru tidak muncul dalam 10 *rules* terbaik berdasarkan *confidence*. Hal ini juga berlaku juga dengan hubungan antara data Rombel dengan data Ruang Kelas.

Dalam hal ini dilakukan pencarian selisih antara data Ruang Kelas terhadap jumlah Rombongan Belajar. Dengan pembuatan dua baru di atas, maka data Guru, Rombel, dan Ruang Kelas dapat dihapus karena kedua data baru tersebut telah merepresentasikan data ketiga data sebelumnya. Kedua hal tersebut dapat dideskripsikan dalam tabel.

Tabel 10 Variabel baru dalam praproses kedua.

Guru – Rombel	Untuk mencari tahu apakah jumlah guru melebihi, sebanding, atau kurang dari jumlah Rombel.
Ruang Kelas - Rombel	Untuk mencari tahu apakah jumlah ruang kelas melebihi, sebanding, atau kurang dari jumlah Rombel.

Berikut adalah aturan konversi untuk dua data baru:

1. Guru – Rombel
   a. Guru_dan_Rombel_sama: Guru - Rombel = 0
   b. Guru_lebih_B: 5 ≥ Guru – Rombel > 0
   c. Guru_lebih_A: 10 ≥ Guru – Rombel ≥ 6
   d. Guru_kebanyakan: Guru – Rombel > 10
   e. Guru_kurang_A: 0 ≥ Guru – Rombel > -5
   f. Guru_kurang_B: -6 ≥ Guru – Rombel > -10
   g. Kekurangan_Banyak_Guru: Guru – Rombel < -10
2. Kelas – Rombel
   a. Ruang_Kelas_dan_Rombel_sama: Kelas – Rombel = 0
   b. Ruang_Kelas_lebih_B: 5 ≥ Kelas – Rombel > 0
   c. Ruang_Kelas_lebih_A: 10 ≥ Kelas – Rombel ≥ 6
   d. Ruang_Kelas_kebanyakan: Kelas – Rombel > 10
   e. Ruang_Kelas_kurang_A: 0 ≥ Kelas – Rombel > -5
   f. Ruang_Kelas_kurang_B: -6 ≥ Kelas – Rombel > -10
   g. Kekurangan_Banyak_Ruang_Kelas: Kelas – Rombel < -10

Untuk aturan konversi Guru – Rombel dan Kelas – Rombel dibuat tanpa analisis tertentu. Konversi dibuat berdasarkan selisih 5 per nilai atribut.

Parameter untuk percobaan kedua mengacu pada hasil dari percobaan pertama. Untuk parameter *support* mengikuti nilai terkecil dari percobaan pertama yaitu 0.02. Hal ini dikarenakan nilai tersebut sudah cukup kecil. Jika dibandingkan dengan jumlah *itemset* yang ada sejumlah 1700 *itemset*, maka 0.02 sekitar 34 *itemset*. Parameter *confidence* yang digunakan pada percobaan kedua adalah menggunakan peraturan yang telah ditentukan sebelumnya yaitu 0.7. Untuk parameter *minlen* masih tetap 4, dan *maxlen* turun menjadi 21 karena jumlah *item* yang berubah pada hasil praproses data. Pada percobaan kedua ini tidak menggunakan parameter set yang lain.

Dari parameter yang telah ditentukan pada proses sebelumnya, maka dilakukan pemrosesan data menggunakan RStudio dengan plugin Arules. Hasil dari pemrosesan data adalah sebagai berikut.

Tabel 11 Hasil pemrosesan data percobaan kedua

Rule	Support	Confidence	Minlen	Maxlen	Total Rules
1	0.02	0.7	4	21	4985066

Berbeda dengan percobaan pertama, percobaan kedua hanya menggunakan satu set parameter. Meskipun nilai minimum *confidence* telah dinaikkan secara drastis, *rules* yang dihasilkan tetap sangat banyak. Maka dari itu pada percobaan kedua tetap dilakukan proses pembuatan *subset* dan penyortiran untuk mengambil sepuluh data terbaik berdasarkan parameter *confidence* untuk setiap *subset*. Berikut hasil dari aturan yang telah didapatkan berdasarkan pembuatan *subset* terhadap mata pelajaran UN pada percobaan kedua.

Tabel 12 Pola aturan subset yang berhasil diekstrak pada percobaan kedua

No	Pola Aturan Subset	Kode Pola
1	BIndo_B, BInggris_B, MTK_B, IPA_B	BBBB
2	BIndo_B, BInggris_B, MTK_B, IPA_C	BBBC
3	BIndo_B, BInggris_C, MTK_B, IPA_B	BCBB
4	BIndo_B, BInggris_C, MTK_C, IPA_C	BCCC
5	BIndo_B, BInggris_C, MTK_D, IPA_C	BCDC
6	BIndo_B, BInggris_D, MTK_C, IPA_C	BDCC
7	BIndo_B, BInggris_D, MTK_D, IPA_C	BDDC
8	BIndo_C, BInggris_C, MTK_C, IPA_C	CCCC
9	BIndo_C, BInggris_C, MTK_D, IPA_C	CCDC
10	BIndo_C, BInggris_C, MTK_D, IPA_D	CCDD
11	BIndo_C, BInggris_D, MTK_D, IPA_C	CDDC
12	BIndo_C, BInggris_D, MTK_D, IPA_D	CDDD

Hasil dari percobaan kedua sedikit berbeda dari percobaan pertama. Perbedaan yang paling jelas adalah perbedaan hasil *rules* yang didapatkan. Secara umum jika dilihat dari parameter *confidence*, *subsets* yang memiliki *rules* dengan nilai *confidence* sama dengan 1 masih sama, tetapi jumlah *rules* yang memiliki *confidence* sama dengan 1 berbeda. Jika dilihat dari nilai *confidence* semua *rules* pada 10 *rules* terbaik berdasarakan *confidence* pada percobaan kedua masuk dalam *rules* yang dapat diterima dengan nilai *confidence* terendah yang ditemukan sebesar 0.89; Berikut adalah daftar *subsets* yang memiliki *confidence* 1 dan *subsets* yang memiliki *lift* yang memiliki setidaknya 1.5 ke atas pada percobaan kedua.

Tabel 13 Daftar *subset* yang memiliki *confidence* = 1 pada percobaan bercobaan kedua

Subsets yang memiliki *confidence* = 1
BBBB
BCBB
BCCC
BCDC
BDCC
BDDC
CCCC
CCDC
CCDD
CDDC

Subsets yang memiliki *confidence* = 1
CDDD

Jika diteliti hasil *rules* pada percobaan kedua, berdasarkan perubahan data yang dilakukan pada praproses data percobaan kedua, ada dua *subsets* yang memiliki *rules* yang berhubungan dengan perubahan data tersebut yaitu *subset* BBBB dan BCCC. Akan tetapi hanya data RuangKelas - Rombel yang paling berpengaruh. Alhasil semua data dari *itemset* hasil praproses pada percobaan kedua memiliki hubungan terhadap nilai UN kecuali Guru-Rombel yang tidak temasuk pada daftar 10 *rules* terbaik pada setiap *subset* yang ditemukan.

Secara khusus, semua *item* kecuali atribut guru - rombel, dan akreditasi dalam *itemset* berpengaruh pada pembuatan *rules* pada percobaan kedua. Hal ini dibuktikan dengan munculnya semua *item* pada semua *subsets*, meskipun tidak muncul secara serempak dalam satu *rule* tetapi dengan menyebar secara parsial ke semua *subset*. Hal ini membuktikan adanya hubungan data fasilitas pada sekolah terhadap nilai UN.

Tabel 14 Daftar kemunculan atribut dalam 10 *rules* terbaik berdasarkan *confidence* pada percobaan kedua

Atribut	Frekuensi Kemunculan dalam 10 *rules* terbaik	Paling Banyak Muncul Pada Subset
Guru – Rombel	0	-
Kelas – Rombel	2	BBBB
Kurikulum	35	BCBB
Penyelenggaraan	7	CDDC
Akses Internet	5	BBBB
Sumber Listrik	82	BDDC, CCCC, CCDC, CDDC
Daya Listrik	4	CDDD
Luas Tanah	1	BBBB
Laboratorium	12	CCCC
Perpustakaan	6	CCCC, CDDC
Sanitasi Siswa	8	CDDC
Presentase Guru Kualifikasi	30	BCBB
Presentase Guru Sertifikasi	9	BCDC
Presentase Guru PNS	8	BCDC, CDDD
Presentase Kelas Layak	11	BCBB
Status	25	BCBB, CCDC, CCDD
Akreditasi	0	-

Berdasarkan 14 atribut sumber listrik memiliki frekuensi kemunculan paling banyak dalam 10 *rules* terbaik berdasarkan *confidence*. Jika berdasarkan frekuensi kemunculan maka dapat disimpulkan bahwa sumber listrik menjadi atribut yang paling berpengaruh dalam 10 *rules* terbaik berdasarkan *confidence*. Selain itu atribut jumlah guru dan akreditasi sekolah menjadi atribut yang paling tidak berpengaruh karena tidak muncul dalam 10 *rules* terbaik berdasarkan *confidence*. Selain itu atribut guru – rombel tidak ditemukan pada 10 *rules* terbaik berdasarkan *confidence*. Ini menunjukan bahwa jumlah guru bisa jadi tidak berpengaruh pada hasil nilai UN sekolah. Hal ini berlaku juga untuk atribut akreditasi.

Jika dilihat dari nilai tertinggi yang ditemukan dalam percobaan pertama dan kedua yaitu pola BBBB, maka adanya akses internet cukup berpengaruh dalam menghasilkan nilai yang cukup tinggi. Hal ini dibuktikan dengan adanya 5 *rules* dengan munculnya atribut internet yang hanya ditemukan pada pola BBBB pada kedua percobaan.

Selain itu atribut-atribut lain dapat dijelaskan pada tabel di bawah.

Tabel 15 Daftar atribut yang berpengaruh terhadap nilai UN

Atribut	Frekuensi pada Percobaan Pertama	Frekuensi pada Percobaan Kedua	Paling Banyak Muncul Pada Subset
Rombongan Belajar	2	-	CDDD
Ruang Kelas	4	-	CDDD
Kelas – Rombel	-	2	BBBB
Kurikulum	35	35	BCBB

Penyelenggaraan	5	7	CDDC
Akses Internet	5	5	BBBB
Sumber Listrik	81	82	BDDC, CCCC, CCDC, CDDC
Daya Listrik	2	4	CDDD
Luas Tanah	1	1	BBBB
Laboratorium	13	12	CCCC
Perpustakaan	7	6	CCCC, CDDC
Sanitasi Siswa	8	8	CDDC
Presentase Guru Kualifikasi	32	30	BCBB
Presentase Guru Sertifikasi	9	9	BCDC
Presentase Guru PNS	7	8	BCDC, CDDD
Presentase Kelas Layak	10	11	BCBB
Status	25	25	BCBB, CCDC, CCDD

Berdasarkan tabel di atas adanya sumber listrik yang termasuk fasilitas yang paling banyak muncul dalam 10 *rules* terbaik berdasarkan *confidence*. Hal ini dibuktikan dengan frekuensi sumber listrik yang paling sering muncul. Selain itu pola BBBB yang menjadi pola UN terbaik yang didapatkan pada percobaan ini memiliki atribut yang hanya ditemukan pada pola tersebut yaitu adanya internet. Bisa disimpulkan adanya internet memberi kemungkinan bahwa suatu sekolah dapat menghasilkan rata-rata nilai UN dengan pola BBBB. Sementara untuk nilai UN terburuk yaitu pola CDDD ada beberapa atribut yang sering muncul pada pola tersebut yaitu atribut rombongan belajar dengan nilai Rombel_B atau jumlah rombongan belajar antara 11 sampai 15 rombel. Lalu ada atribut ruang kelas dengan nilai RK_B atau jumlah ruang kelas antara 11 sampai 15. Selanjutnya atribut daya listrik dengan nilai DL_A sampai DL_D. Tapi dengan munculnya nilai yang beragam pada atribut daya listrik menjadikan atribut tersebut tidak dapat dijadikan atribut yang paling berpengaruh karena tidak ada nilai pasti yang berfokus pada atribut tersebut. Selain atribut-atribut tersebut masih banyak atribut lain yang memiliki hubungan khusus terhadap pola nilai UN seperti kurikulum dengan pola BCBB, penyelenggaraan dengan pola CDCC, laboratorium dengan pola CCCC, perpustakaan dengan pola CCCC dan CDDC, sanitasi siswa dengan pola CDDC, persentase guru kualifikasi dengan pola BCBB, persentase gutu sertifikasi dengan BCDC, persentase guru PNS dengan pola BCDC dan CDDD, persentase kelas layak dengan pola BCBB, dan status dengan pola BCBB, CCDC, dan CCDD.

Dari dua percobaan yang telah dilakukan dapat disimpulkan secara umum jika adanya sumber listrik menjadi fasilitas yang paling umum ditemukan pada sekolah di Indonesia. Jika lihat secara statistik, setidaknya sekolah yang memiliki sumber listrik setidaknya akan menghasilkan nilai UN dengan pola BDDC, CCCC, CCDC atau CDDC. Selain itu, atribut guru dan akreditasi menjadi atribut yang paling tidak berpengaruh karena tidak muncul pada 10 *rules* terbaik berdasarkan *confidence* pada dua percobaaan yang telah dilakukan.

5 Kesimpulan

Berdasarkan percobaan-percobaan di atas dapat ditarik kesimpulan sebagai berikut:

1. Ada hubungan antara fasilitas sekolah dan atribut non fasilitas sekolah terhadap hasil nilai UN. Hal ini dibuktikan dengan ditemukannya 12 subsets hubungan nilai UN dengan fasilitas sekolah.
2. Dengan menggunakan algoritma Apriori telah dapat disimpulkan bahwa ada keterkaitan antara fasilitas sekolah dengan nilai UN. Dari total 22 atribut pada percobaan pertama dan 21 atribut pada percobaan kedua, atribut sumber listrik menjadi atribut yang paling berpengaruh pada percobaan ini. Selain itu atribut guru dan akreditasi menjadi atribut yang paling tidak berpengaruh karena tidak muncul di 10 rules terbaik pada setiap subsets yang telah ditemukan.
3. Preprocessing data sangat berpengaruh terhadap rules yang dihasilkan algoritma Apriori. Hal ini dibuktikan dengan adanya perbedaan rules yang dihasilkan pada kedua percobaan yang telah dilakukan. Dengan metode preprocessing data yang tepat maka dapat menghasilkan rules yang lebih akurat.

Adapun saran untuk penelitian lebih lanjut mengenai analisis nilai UN terhadap fasilitas sekolah kedepannya adalah sebagai berikut:

1. Perlunya tambahan data sekolah yang dibutuhkan untuk menghasilkan rules yang lebih akurat.
2. Seiring penambahan data, maka diperlukan juga penambahan kapasitas komputasi yang digunakan dalam melakukan mining rules.
3. Perlunya pendalaman pada metode praprocessing data pada kasus ini. Terlebih dengan tindakan terhadap missing value dan data tidak valid. Hal ini diperlukan untuk menghasilkan rules yang lebih relevan dari percobaan yang telah dilakukan sebelumnya.

4. Perlunya kajian hasil percobaan ini dengan pihak yang terkait dengan bidang pendidikan terutama Kemdikbud dalam menanggapi hasil dari rules yang telah dihasilkan.

Daftar Pustaka

[1] Fayyad et al. 1996. From Data Mining to Knowledge Discovery in Databases. AI Magazine Volume 17 Number 3. http://www.aaai.org/ojs/index.php/aimagazine/article/download/1230/1131/. [2 September 2018].

[2] Suyanto. 2017. Data Mining Untuk Klasifikasi dan Klasterifikasi Data. Penerbit Informatika. Bandung.

[3] ACM. 2006. ACM SIGKDD, Data Mining Curriculum.

[4] Clifton, C. 2010. Encyclopædia Britannica: Definition of Data Mining.

[5] Zhang, Chenqi, dan Shichao Zhang. 2002. Association Rule Mining Models and Algorithms. Springer.

[6] Sayad, Saed. 2010. Association Rules. Diambil dari: http://www.saedsayad.com/association_rules.htm.

[7] Arora, Jyoti, Nidhi Bhalla, Sanjeev Rao. A Review On Association Rule Mining Algorithms. Diambil dari: http://www.rroij.com/open-access/a-review-on-association-rulemining-algorithms.php?aid=43382. (5 November 2018)

[8] IBM. Lift in an Association Rule. Diambil dari: https://www.ibm.com/support/knowledgecenter/en/SSEPGG_9.7.0/com.ibm.im.model.doc/c_lift_in_an_association_rule.html. (5 Agustus 2019)

[9] Han, Jiawei, Micheline Kamber, Jian Pei. 2012. Data Mining Concepts and Techniques: Third Edition. Morgan Kaufmann Publisher, Elsevier. USA.

[10] PT. Nufaza. 2018. Laporan Akhir Pengembangan Aplikasi Pelaporan Bantuan Sarana dan Prasarana SD Tahun 2017. Direktorat Pembinaan Sekolah Dasar, Direktorat Jendral Pendidikan Dasar dan Menengah. Jakarta.

[11] Dimyati dan Mudjiono. 1999. Belajar dan Pembelajaran. Jakarta: Rineka Cipta.

[12] Kementrian Pendidikan dan Kebudayaan Republik Indonesia. 1975. Keputusan Menteri P dan K No. 079/1975.