

PENERAPAN ALGORITMA APRIORI DALAM SELEKSI PENJURUSAN CALON PESERTA DIDIK BARU DI SMAK DIPONEGORO BLITAR

Refani Puspa Karina, Sri Lestanti, Filda Febrinita

Program Studi Teknik Informatika S1, Fakultas Teknologi Informasi
Universitas Islam Balitar, Jalan Majapahit No. 2-4 Kota Blitar, Indonesia
rfnpuspa@gmail.com

ABSTRAK

Keputusan merupakan suatu hal yang sangat berpengaruh dalam proses menghadapi alternatif yang dipilih. Berbagai kendala dalam memilih penjurusan yang sesuai dengan kriteria memang cukup membingungkan. Tujuan penelitian ini untuk menentukan konsentrasi penjurusan yang akan diambil oleh calon peserta didik baru dengan menggunakan algoritma Apriori. Algoritma Apriori digunakan untuk menentukan seleksi penjurusan calon peserta didik baru berdasarkan nilai rata-rata mata pelajaran pilihan penjurusan. Dari 79 data calon peserta didik baru, dengan nilai minimum support sebesar 0,45 dan minimum confidence sebesar 0.85 dikarenakan pengkaji sudah melakukan pergeseran nilai sebelumnya, dan dengan minimum support dan minimum confidence sebelumnya didapatkan frequent itemset yang selanjutnya diuji keakuratannya. Didapatkan hasil pengujian menggunakan lift ratio dengan nilai 1.09, setelah dihitung nilai dari pengujian lift tersebut dapat dinyatakan akurat untuk melakukan rekomendasi penjurusan. Nilai lift ratio 1,09 menghasilkan korelasi positif dimana dari 79 pendaftar, 46% didalamnya direkomendasikan masuk ke dalam penjurusan IPA dan IPS.

Kata Kunci : algoritma apriori, peserta didik baru, seleksi penjurusan.

1. PENDAHULUAN

SMA Katolik Diponegoro Blitar merupakan salah satu sekolah swasta tertua di Kota Blitar yang merupakan sekolah unggul dengan nilai akreditasi A. SMA Katolik Diponegoro Blitar memiliki 3 jurusan akademik, yaitu Sains (IPA), Sosial (IPS) dan Bahasa. Pembagian penjurusan penerimaan siswa baru tidak dapat dilihat dari satu faktor saja, tetapi harus dilihat dari banyak faktor yang mempengaruhi pembagian penjurusan, karena setiap minat siswa berbeda. Banyak siswa yang memilih jurusan tidak sesuai minat serta kemampuannya, dan cenderung mengikuti pilihan jurusan yang dipilih oleh teman-temannya sehingga berujung salah jurusan dan berakibat putus sekolah di tengah jalan. Hal tersebut yang perlu mendapat perhatian lebih dari pihak sekolah. Perlu diperhatikan juga faktor internal yaitu kondisi yang ada pada siswa seperti kesehatannya [1].

Berdasarkan observasi yang dilakukan di SMAK Diponegoro Blitar, peneliti menemukan berbagai permasalahan diantaranya tidak sedikit siswa yang memilih jurusan tidak sesuai minat dan kemampuannya dan cenderung mengikuti pilihan teman-temannya, sehingga berujung salah jurusan dan berakibat putus sekolah di tengah jalan.

Dalam mengetahui penjurusan yang diambil secara bersamaan, dapat digunakan *algoritma apriori* untuk menemukan aturan asosiatif suatu kombinasi *item*. Pencarian asosiasi berawal dari pengolahan rata-rata beberapa nilai mata pelajaran yang dipilih dari raport pendidikan terakhir sekolah menengah pertama (SMP) atau sederajat yang diolah dan menghasilkan rekomendasi jurusan sesuai nilai siswa.

Dengan judul Penerapan Algoritma Apriori Dalam Seleksi Penjurusan Calon Peserta Didik Baru

Di SMAK Diponegoro Blitar diharapkan dapat mempermudah siswa dalam memilih penjurusan sesuai nilai dan kemampuan siswa.

2. TINJAUAN PUSTAKA

2.1. Data Calon Peserta Didik Baru

Data calon peserta didik baru diperoleh dari formulir pendaftaran yang diisi saat murid mendaftar ke sekolah. Data ini terdiri dari beberapa atribut, yaitu: nama, tempat/tanggal lahir, agama, jenis kelamin, nama orang tua, alamat tempat tinggal, asal sekolah, nomor telepon dan jurusan yang dipilih. Data ini akan diperbarui bila terjadi perubahan yang dilaporkan oleh murid atau wali murid. Beberapa data yang bersifat privasi hanya dimanfaatkan oleh pihak sekolah [2].

2.2. Data Mining

Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait dari berbagai database besar [3].

Data mining berisi pencarian pola yang diinginkan di dalam basis data yang besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali agar dapat memberikan suatu analisa data yang berguna untuk dapat dipelajari dengan lebih teliti [4].

Data mining merupakan salah satu dari langkah *Knowledge Discovery from Data* atau lebih dikenal dengan singkatan KDD. Berikut langkah-langkah dari KDD [5]:

1. *Data cleaning*
Menghilangkan noise dan data yang tidak konsisten
2. *Data integration*
Sumber-sumber data diintegrasikan
3. *Data selection*
Data yang sesuai tugas analisis diambil dari database
4. *Data transformation*
Data ditransformasikan dan dikonsolidasikan kedalam bentuk yang sesuai dengan melakukan operasi ringkasan dan pengumpulan.
5. *Data mining*
Proses penting dimana metode cerdas diterapkan untuk menemukan pola data
6. *Pattern evaluation*
Mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan ukuran ketertarikan
7. *Knowledge presentation*
Visualisasi dan teknik representasi pengetahuan digunakan untuk menyajikan pengetahuan yang digalu kepada pengguna.

2.3. Algoritma Apriori

Algoritma apriori termasuk kedalam jenis *association rule* pada *data mining* yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market basket analysis*. *Association rule* adalah teknik *data mining* yang digunakan untuk menemukan aturan suatu kombinasi *item* serta berguna untuk menemukan data tersembunyi yang menarik di dataset. Konsep *association rule* berasal dari *frequent pattern mining* yang mencari pola barang yang sering muncul secara bersamaan pada sebuah himpunan data.

Tahapan *association rule* yang menarik banyak perhatian peneliti dalam menghasilkan algoritma yang efisien adalah *frequent pattern mining*. Suatu asosiasi dapat diketahui penting tidaknya melalui dua tolok ukur, yaitu: *support* (nilai penunjang) yang merupakan presentase kombinasi *item* dalam *database* dan *confidence* (nilai kepastian) yang merupakan kuatnya hubungan antara *item-item* dalam *association rule* [6]. Langkah langkah perhitungan *algoritma apriori* adalah sebagai berikut:

1. *Frequent pattern mining* (pola frekuensi tinggi)
Pada tahap dimana pembentukan kombinasi antar *item* yang telah memenuhi syarat minimum *support*. Nilai *support* pada satu *itemset* dapat diperoleh dengan menggunakan sebuah rumus sebagai berikut:

$$Support(A) = \frac{\text{jumlah transaksi mengandung } A}{\text{nilai transaksi}} \quad (1)$$

Rumus diatas digunakan untuk menentukan nilai *support* pada satu *itemset* jumlah transaksi yang mengandung *item A* dibagi dngan jumlah transaksi yang ada pada data.

Rumus nilai *support 2 item* adalah sebagai berikut:

$$Support(A, B) = \frac{\text{Support}(A \cap B) = \frac{\text{jumlah transaksi mengandung } A \text{ dan } B}{\text{nilai transaksi}} \times 100\% \quad (2)$$

Rumus diatas digunakan untuk menentukan nilai *support* pada dua *itemset* atau lebih dibagi dengan jumlah transaksi yang mengandung *item A* dan *B* yang terjadi pada data dan dikalikan seratus persen.

2. Pembentukan *Association rule*
Langkah yang dilakukan setelah menentukan nilai suport pada *itemset* berfrekuensi tinggi lalu dibentuk *Association rule* yang menyatakan kuatnya hubungan *itemset* pada transaksi. Untuk menentukan *Association rule* yang terbentuk berlaku hukum asosiatif $A \rightarrow B$ tidak berlaku $B \rightarrow A$. Untuk menentukan aturan $A \rightarrow B$ digunakan rumus:

$$Support(A|B) = \frac{\text{jumlah transaksi mengandung } A \text{ dan } B}{\text{jumlah transaksi mengandung } A} \times 100\% \quad (3)$$

3. Perhitungan nilai *confidence*
Langkah yang dilakukan setelah terbentuknya rule adalah menghitung nilai *confidence*. Setelah semua pola frekuensi tinggi ditemukan, kemudian dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung aturan asosiatif $A \rightarrow B$. nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dari rumus berikut:

$$Confidence = P(B|A) = \frac{\sum \text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi mengandung } A} \quad (4)$$

2.4. Pengujian Lift Ratio

Proses mining apriori ditandai dengan terbentuknya kekuatan hubungan kombinasi *itemset* dengan alat ukur asosiasi final. Namun untuk mengukur valid atau tidaknya asosiasi final tersebut maka dapat menggunakan lift ratio [7]. Lift ratio adalah alat ukur penting dalam aturan asosiasi. Fungsinya adalah mengukur ketepatan dan kecermatan suatu alat ukur (*support* dan *confidence*) agar dapat dipercaya sepenuhnya. Pada akhirnya sebuah kombinasi *itemset* dinyatakan valid dan kuat jika nilai lift ratio > 1. Rumus perhitungan lift ratio adalah sebagai berikut:

$$Lift(A, B) = \frac{P(A \cup B)}{P(A) P(B)} \quad (5)$$

- Lift (A,B): korelasi antara A dan B
- P (A ∪ B): jumlah kemunculan antara A dan B dibagi dengan total transaksi
- P(A) P(B): jumlah kemunculan A dikali jumlah jkemunculan B pada total transaksi

Apabila dari perhitungan tersebut menghasilkan nilai dibawah 1 maka terdapat korelasi negatif. Untuk perhitungan yang menghasilkan nilai diatas 1 maka terdapat korelasi positif. Namun apabila menghasilkan

nilai sama dengan 1 maka tidak ada korelasi antara X dan Y.

2.5. Kajian Terkait

Penelitian tentang pembagian jurusan penerimaan siswa baru di SMK Negeri 1 Siantar. Penelitian tersebut bertujuan untuk menerapkan algoritma apriori dalam pemilihan penjurusan siswa yang seringkali dipilih dan memiliki nilai kepercayaan tinggi. Hasil penelitian ini terdiri dari data yang mengelompokkan calon mahasiswa baru berdasarkan jurusan yang diinginkan dengan dukungan minimal 50% dan keyakinan minimal 50%, menghilangkan 20 aturan. Salah satu aturan yang ditetapkan adalah jika seorang mahasiswa memilih jurusan fashion (A4), maka jurusan yang lebih tepat untuk mahasiswa tersebut adalah rambut dan kulit (A2) dengan *support* score 0,5 atau sama dengan 50 dan nilai *confidence* 100 sampai 0,5 atau sama dengan 50 [6].

Penelitian tentang penggunaan algoritma apriori untuk mendukung strategi promosi perguruan tinggi. Bahwa penentuan strategi promosi yang tepat akan dapat menekan biaya promosi dan mencapai target promosi yang tepat. Salah satu cara yang mungkin untuk menentukan strategi promosi adalah dengan menggunakan teknik *data mining*. Teknik *data mining* yang digunakan dalam hal ini menggunakan *algoritma apriori*. *Algoritma apriori* adalah salah satu algoritma *data mining* klasik. *Algoritma apriori* digunakan agar komputer dapat mempelajari aturan asosiasi, mencari pola hubungan antara satu atau lebih *item* dalam kumpulan data. Penelitian tersebut dilakukan dengan mengamati beberapa variabel penelitian yang sering menjadi pertimbangan perguruan tinggi untuk menentukan tujuan kenaikan pangkat, yaitu sekolah, wilayah, dan jurusan. Hasil dari penelitian tersebut berupa pola-pola dengan presentase *support* 11-29% dan presentase *confidence* 92-100% [8].

Penelitian tentang penggunaan algoritma apriori untuk menentukan strategi promosi perguruan tinggi. Penelitian dilakukan sesuai dengan tahapan pada metode CRISP-DM mulai dari *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Hasil penelitian menunjukkan bahwa diperoleh 6 (enam) aturan asosiatif untuk menunjukkan asosiatif pemilihan program studi terhadap daerah asal dan jenis sekolah dengan nilai *confidence* tertinggi 70%. Mahasiswa yang memilih program studi Teknik Informatika lebih banyak berasal dari jenis sekolah SMK di daerah Badung dan Denpasar, sedangkan mahasiswa yang memilih program studi Sistem Informasi lebih banyak berasal dari jenis sekolah SMA di daerah Badung dan SMK di daerah Denpasar, dan mahasiswa yang memilih program studi Sistem Informasi Akuntansi lebih banyak berasal dari SMA di daerah Badung dan SMK di daerah Denpasar. Selanjutnya hasil penelitian dapat digunakan oleh manajemen STMIK Primakara untuk dapat dijadikan acuan/pedoman dalam menentukan kebijakan dan

pengambilan keputusan dalam menentukan strategi promosi perguruan tinggi [9].

Penelitian tentang penggunaan algoritma apriori pada analisis pola belanja konsumen. Penerapan *algoritma apriori* dalam menganalisa pola belanja konsumen toko buku Gramedia Bintaro. Dengan memanfaatkan data transaksi yang ada di database, pihak manajemen dapat mengetahui buku-buku yang sering dibeli. Pada penelitian tersebut penerapan *Algoritma apriori* digunakan untuk mencari kombinasi *item* terbanyak berdasarkan data transaksi dan kemudian membentuk pola asosiasi dari kombinasi *item* tersebut. Pola asosiasi yang terbentuk dengan nilai minimum *support* 5% dan nilai minimum *confidence* 15% menghasilkan 7 aturan asosiasi. Dan *strong rules* yang didapatkan adalah *schoolbooks indonesia curriculum children's book* dengan nilai *support* 11,23% dan nilai *confidence* 30,66%. Analisis pola yang dihasilkan aplikasi ini dapat digunakan oleh pihak Gramedia untuk pengaturan tata letak buku, rekomendasi pencarian buku pada Gramedia Online dan lain-lainnya, yang mana setiap bulannya pola yang dihasilkan bisa berbeda-beda sesuai data transaksi yang dianalisis. Semakin besar jumlah data yang di proses maka semakin lama waktu yang dibutuhkan. Semakin besar nilai *support* dan nilai *confidence* yang diatur maka akan semakin singkat waktu pemrosesan algoritma [10].

Penelitian tentang penerapan algoritma apriori untuk meningkatkan promosi produk. Dengan algoritma tersebut dilakukan pencarian frekuensi dan *item* barang yang paling sering muncul. Hasil dari penelitian ini menunjukkan bahwa *algoritma apriori* dapat digunakan untuk menganalisis data transaksi sehingga diketahui mana produk yang harus dipromosikan. Perhitungan metode apriori menghasilkan suatu pola pembelian yang terjadi di PD. XYZ. Dengan menganalisis pola tersebut dihasilkan kesimpulan bahwa produk yang akan dipromosikan yaitu cat tembok ekonomis dan peralatan cat berupa kuas tangan dengan nilai *support* 11% dan *confidence* 75% [11].

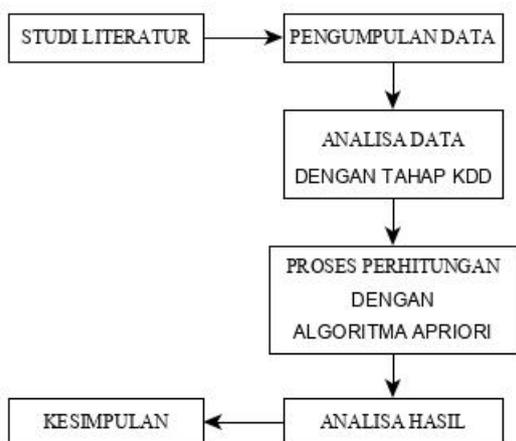
Penelitian tentang penggunaan algoritma apriori untuk rekomendasi sistem. Sistem rekomendasi yang menganalisis pola dan kecenderungan pribadi pelanggan dengan menggunakan analisis aturan asosiasi dan filter kolaboratif untuk mengumpulkan data pelanggan saat mengunjungi perusahaan pelanggan dengan NFC (*Near Field Communication*). Algoritma rekomendasi yang digunakan dalam sistem yang diusulkan menggunakan hasil analisis data dan data jarak dari GPS (*Global Positioning System*) untuk merekomendasikan bisnis lokal yang paling sering dikunjungi orang [12].

Penelitian tentang analisis data lulusan menggunakan algoritma apriori untuk mendukung strategi promosi Universitas Lancang Kuning. Setiap perusahaan maupun organisasi yang ingin tetap bertahan perlu untuk menentukan strategi promosi yang tepat. Penentuan strategi promosi yang tepat akan

dapat mengurangi biaya promosi dan mencapai sasaran promosi yang tepat. Salah satu cara yang dapat dilakukan untuk penentuan strategi promosi adalah dengan menggunakan teknik *data mining*. Teknik *data mining* yang digunakan dalam hal ini adalah dengan menggunakan algoritma *Clustering K-Means*. *Clustering* merupakan pengelompokan record, observasi, atau kasus ke dalam kelas-kelas objek yang mirip. K-Means adalah metode klaster data non-hirarkis yang mencoba untuk membagi data ke dalam satu atau lebih klaster. Penelitian dilakukan dengan mengamati beberapa variabel penelitian yang sering dipertimbangkan oleh perguruan tinggi dalam menentukan sasaran promosinya yaitu asal sekolah, daerah, dan jurusan. Hasil penelitian ini adalah berupa pola menarik hasil *data mining* yang merupakan informasi penting untuk mendukung strategi promosi yang tepat dalam mendapatkan calon mahasiswa baru [13].

Penelitian tentang penggunaan algoritma apriori dalam pemilihan program studi calon mahasiswa baru di Universitas Widyatama Bandung. Belum pernah diketahui rule pemilihan program studi yang dilakukan oleh calon mahasiswa. Sehingga strategi khusus yang dilakukan dalam kegiatan adalah promosi. Salah satu alternatif solusi dari masalah tersebut menerapkan teknik *data mining*, sehingga dapat dilakukan penelusuran pada data historis untuk mengidentifikasi pola data yang dimiliki terhadap subyek yang diteliti yang didasarkan pada sifat-sifat data yang teridentifikasi sebelumnya. Dalam penelitian ini akan menggunakan *algoritma apriori* untuk mengetahui asosiasi rule pemilihan program studi calon mahasiswa. Hasil penelitian tersebut menghasilkan pola yang relative tinggi dengan presentase support 10-28% dan presentase confidence 87-94% [14].

3. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

3.1. Studi Literatur

Studi literatur merupakan cara untuk menyelesaikan persoalan atau masalah dengan menelusuri sumber-sumber tulisan seperti buku-buku

karya pengarang terpercaya, jurnal ilmiah, dan hasil penelitian mahasiswa dalam bentuk skripsi.

3.2. Pengumpulan data

Pada penelitian ini pengumpulan data dilakukan dengan melakukan observasi dan wawancara langsung ke tempat penelitian guna mendapatkan data yang diperlukan oleh peneliti.

3.3. Analisis data dengan tahap KDD

Analisa data dengan tahap KDD merupakan proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi penting. Pada penelitian ini analisa data dengan tahap KDD dilakukan untuk melakukan seleksi data dan pembersihan data dimana proses tersebut dilakukan untuk membuang duplikasi data dan memperbaiki kesalahan pada data seperti kesalahan cetak.

3.4. Proses perhitungan dengan algoritma apriori

Proses perhitungan dengan *algoritma apriori* merupakan proses yang dilakukan dengan menghitung *itemsets* yang diperlukan dan menentukan *support* dan *confidence* untuk mendapatkan kombinasi *item* tertinggi. *Algoritma apriori* digunakan agar komputer dapat mempelajari aturan asosiasi, mencari pola hubungan antar satu atau lebih *item* dalam suatu dataset. *Algoritma apriori* adalah algoritma yang paling terkenal untuk menentukan pola frekuensi tinggi. Pola frekuensi tinggi merupakan pola *item*-iem didalam suatu *database* yang memiliki frekuensi atau *support* diatas ambang batas tertentu yang disebut istilah *minimum support*. Pola frekuensi ini digunakan untuk menyusun aturan asosiatif dan juga beberapa *data mining* lainnya. Proses perhitungan menggunakan *algoritma apriori* pada penelitian ini menggunakan rumus:

Nilai *support* pada satu *itemset* dapat diperoleh dengan menggunakan sebuah rumus sebagai berikut:

$$Support(A) = \frac{jumlah\ transaksi\ mengandung\ A}{nilai\ transaksi} \quad (6)$$

Rumus *support* pada dua *itemset* dapat menggunakan rumus:

$$Support(A, B) = Support(A \cap B) = \frac{jumlah\ transaksi\ mengandung\ A\ dan\ B}{nilai\ transaksi} \times 100\% \quad (7)$$

Rumus diatas digunakan untuk menentukan nilai *support* pada dua *itemset* atau lebih dibagi dengan jumlah transaksi yang mengandung *item* A dan B yang terjadi pada data dan dikalikan seratus persen.

Saat seluruh pola frekuensi tinggi didapatkan, maka selanjutnya dicari aturan relasi yang memenuhi kriteria persyaratan minimal untuk confidence dengan menghitung confidence aturan asosiatif $A \rightarrow B$ menggunakan rumus:

$$Support(A, B) = P \left(\frac{A}{B} \right) \frac{\Sigma \text{transaksi mengandung } A \text{ dan } B}{\Sigma \text{jumlah transaksi mengandung } A} \quad (8)$$

Rumus diatas digunakan untuk mendapatkan nilai confidence dengan melakukan pembagian jumlah transaksi yang memuat *item* A dan B dengan jumlah transaksi yang memuat *item* A

Dengan data yang dihitung berjumlah 8 dataset dengan kondisi dimana bila ingin memasuki jurusan IPA, maka nilai yang masuk kedalam mata pelajaran IPA harus meemnuhi syarat masuk jurusan IPA. Rata-rata masuk jurusan IPA adalah 80, begitu pula masuk IPS harus dengan rata-rata raport 80, dan penjurusan Bahasa dengan rata-rata nilai 78. Pada penelitian ini, pengkaji menggunakan 3 frequents pattern item, dengan minimum *support* 45% dan minimum *confidence* 60%.

3.5. Pengujian dan analisa hasil

a. Pengujian

Pengujian merupakan hal yang harus dilakukan dalam penelitian yang bertujuan untuk mendapatkan informasi mengenai ketepatan dan keakuratan dari kualitas data transaksi yang sedang diuji (*under test*). Pengujian pada penelitian ini menggunakan pengujian *lift rasio*. *Lift* berupa angka rasio yang mengidentifikasi banyaknya kemungkinana dalam menemukan semua atribut yang muncul bersama dibandingkan dengan semua atribut yang memenuhi syarat. *Lift* mengekspresikan level kekuatan *rule* atas kejadian acak dari *antecedent* dan *consequent* berdasarkan pada *supportnya*. Perhitungan *lift* dapat dilihat pada rumus berikut:

$$LiftRasio = \frac{confidence}{benchmark\ confidence} \quad (9)$$

Sedangkan perhitungan dari *benchmark confidence* dapat dilihat pada rumus:

$$ConfidenceBenchmark = \frac{\Sigma \text{transaksidenganitemdalamconsequent}}{\Sigma \text{transaksidalamdatabase}} \quad (10)$$

Lift rasio yang bernilai lebih besar dari satu mengidentifikasikan aturan tersebut memiliki manfaat. Semakin besar nilai *lift* rasio maka makin besar pula kekuatan asosiasi.

b. Analisa hasil

Untuk mendapatkan analisa hasil yang akurat dalam penelitian ini perlu melakukan pengujian, peroses pengujian ini dilakukan dengan menganalisa hasil dari proses pehitungan metode tersebut. *Item* yang menghasilkan aturan asosiasi terbaik akan dijadikan acuan dalam proses seleksi penjurusan berdasarkan nilai *support* dan *confidence*.

3.6. Kesimpulan

Kesimpulan merupakan tahap penyusunan hasil penelitian yang diuraikan kedalam bentuk teks yang berorientasi penemuan.

4. HASIL DAN PEMBAHASAN

Berdasarkan kegiatan penelitian yang dilakukan, data yang digunakan merupakan private dataset berupa data calon peserta didik baru SMAK Diponegoro Blitar yang diperoleh melalui proses observasi dan wawancara. Data yang diperoleh merupakan data nilai siswa dan rata-rata nilai untuk mengetahui penjurusan yang dapat dimasuki. Selanjutnya data diproses dengan tahap *pre-processing* dimana tahap ini digunakan agar pengolahan data mentah menjadi data yang siap digunakan. Pada penelitian ini pengkaji menggunakan tahapan proses data cleaning dimana data berupa nilai dan nomor penndaftar diproses dan dihapus yang tidak diperlukan. Setelah melewati tahapan data cleaning, data yang telah diproses kemudian ditransformasi sebagai berikut

4.1. Transformasi Data

Tabel 1. Transformasi Data

No.	ΣA1	ΣS1	ΣB1	No.	ΣA1	ΣS1	ΣB1
1	0	0	1	41	1	1	1
2	1	1	1	42	0	0	1
3	1	0	1	43	0	1	0
4	1	1	1	44	0	1	1
5	0	1	1	45	1	1	1
6	1	1	1	46	0	1	1
7	1	1	1	47	1	1	1
8	1	0	1	48	1	1	1
9	1	1	1	49	1	1	1
10	0	0	1	50	0	1	1
11	0	0	1	51	0	0	1
12	1	1	1	52	1	1	1
13	0	1	1	53	0	1	1
14	1	1	1	54	0	1	0
15	1	1	1	55	0	0	1
16	1	1	1	56	0	1	0
17	0	1	1	57	1	1	1
18	1	0	1	58	1	1	1
19	1	0	1	59	1	1	1
20	1	1	1	60	1	1	1
21	0	1	1	61	0	1	1
22	1	1	1	62	0	1	1
23	1	0	1	63	0	1	1
24	1	1	1	64	0	0	1
25	1	1	1	65	0	1	0
26	0	1	1	66	0	1	1
27	1	1	1	67	0	1	1
28	1	1	1	68	1	1	1
29	0	1	1	69	0	1	1
30	0	0	1	70	0	1	1
31	1	0	1	71	1	0	1
32	1	0	0	72	1	1	1
33	0	0	1	73	1	1	1
34	1	1	1	74	1	1	1
35	1	1	1	75	0	1	0
36	1	1	1	76	1	1	1
37	0	1	0	77	0	1	1
38	1	1	1	78	0	0	1
39	1	1	1	79	1	1	1
40	1	1	1				

Keterangan:

- a. $\sum A1$: Rata-rata nilai mata pelajaran IPA
- b. $\sum S1$: Rata-rata nilai mata pelajaran IPS
- c. $\sum B1$: Rata-rata nilai mata pelajaran Bahasa
- d. 0 : nilai kurang yang tidak memenuhi syarat masuk penjurusan
- e. 1 : nilai yang memenuhi syarat masuk penjurusan

Data diatas merupakan hasil data seleksi dimana nilai yang memenuhi syarat masuk penjurusan IPA dan IPS dengan nilai ≥ 80 dan nilai yang memenuhi syarat masuk penjurusan Bahasa ≥ 78 yang dicatat dengan angka 1, dan nilai yang tidak memenuhi syarat masuk penjurusan IPA dan IPS < 80 dan penjurusan Bahasa < 78 dicatat dengan angka 0 yang selanjutnya akan diproses untuk mencari berapa nilai support dan confidence lalu dilakukan pengujian menggunakan lift ratio.

Setelah data berhasil ditransformasi selanjutnya dilakukan proses perhitungan menggunakan algoritma apriori dengan menentukan minimum support sebesar 0,45 dan minimum confidence sebesar 0,85 dengan cara mencari rule penentuan 1 sampai 3 *frequent itemset* dan dilanjutkan pencarian hasil nilai *confidence*.

4.2. Pembentukan 1-Itemset

Tabel 2. Pembentukan 1-Itemset

1-itemsets	support	ket
A1 < 80	0,4303797	
A1 \geq 80	0,5696203	lolos
S1 < 80	0,2405063	
S1 \geq 80	0,7721519	lolos
B1 < 78	0,0886076	
B1 \geq 78	0,9113924	lolos

Keterangan:

- a. A1 < 80 : nilai mata pelajaran yang rata-ratanya tidak memenuhi syarat masuk penjurusan IPA
- b. A1 \geq 80 : nilai mata pelajaran yang rata-ratanya memenuhi syarat masuk penjurusan IPA
- c. S1 < 80 : nilai mata pelajaran yang rata-ratanya tidak memenuhi syarat masuk penjurusan IPS
- d. S1 \geq 80 : nilai mata pelajaran yang rata-ratanya memenuhi syarat masuk penjurusan IPS
- e. B1 < 80 : nilai mata pelajaran yang rata-ratanya tidak memenuhi syarat masuk penjurusan Bahasa
- f. B1 \geq 80 : nilai mata pelajaran yang rata-ratanya memenuhi syarat masuk penjurusan Bahasa

Pada tahap diatas nilai yang memenuhi syarat masuk penjurusan IPA di hitung dan dibagi dengan data keseluruhan dimana nilai yang memenuhi syarat masuk IPA sebesar $45/79 * 100\% = 0,5696$ dan nilai yang tidak memenuhi syarat masuk sebesar $34/79 * 100\% = 0,4303$

Pada tahap diatas nilai yang memenuhi syarat masuk penjurusan IPS di hitung dan dibagi dengan data keseluruhan dimana nilai yang memenuhi syarat masuk IPS sebesar $60/79 * 100\% = 0,7721$ dan nilai

yang tidak memenuhi syarat masuk sebesar $19/79 * 100\% = 0,2405$

Pada tahap diatas nilai yang memenuhi syarat masuk penjurusan Bahasa di hitung dan dibagi dengan data keseluruhan dimana nilai yang memenuhi syarat masuk Bahasa sebesar $72/79 * 100\% = 0,9113$ dan nilai yang tidak memenuhi syarat masuk sebesar $7/79 * 100\% = 0,0886$

4.3. pembentukan 2-itemsets

Tabel 3. Pembentukan 2-Itemsets

2-itemsets	support	ket
A1, S1	0,4683544	lolos
A1, B1	0,556962	lolos
S1, A1	0,4556962	lolos
S1, B1	0,6962025	lolos
B1, A1	0,5443038	lolos
B1, S1	0,6962025	lolos

Keterangan:

- a. A1: penjurusan IPA
- b. S1: penjurusan IPS
- c. B1: penjurusan Bahasa
- d. Support: nilai dukungan
- e. Ket: keterangan

Pada tahap diatas, merupakan tahap dimana nilai yang sebelumnya memenuhi minimal support di hitung lagi dengan 2 itemsets sebagaimana:

Nilai yang memenuhi syarat masuk IPA dan IPS ditotal sebagaimana A1, S1 : $37/79 * 100\% = 0,4683$ sehingga didapat nilai support sebesar 0,4683

Nilai yang memenuhi syarat masuk IPA dan Bahasa ditotal sebagaimana A1, B1 : $44/79 * 100\% = 0,5569$ sehingga didapat nilai support sebesar 0,5569

Nilai yang memenuhi syarat masuk IPS dan IPA ditotal sebagaimana S1, A1 : $36/79 * 100\% = 0,4556$ sehingga disapat nilai support sebesar 0,4556

Nilai yang memenuhi syarat masuk IPS dan Bahasa ditotal sebagaimana S1, B1 : $55/79 * 100\% = 0,6962$ sehingga didapat nilai support sebesar 0,6962

Nilai yang memenuhi syarat masuk Bahasa dan IPA ditotal sebagaimana B1, A1 : $43/79 * 100\% = 0,5443$ sehingga didapat nilai support sebesar 0,5443

Nilai yang memenuhi syarat masuk Bahasa dan IPS ditotal sebagaimana B1, S1 : $55/79 * 100\% = 0,6962$ sehingga didapat nilai support sebesar 0,6962

4.4. Pembentukan 3-Itemset

Tabel 4 Pembentukan 3-Itemset

3-itemsets	support	ket
A1,S1,B1	0,46835	lolos

Keterangan:

- a. A1,S1,B1 : nilai yang memenuhi minimal support pada tabel sebelumnya
- b. Support: nilai dukungan
- c. Ket: keterangan

Pada tabel diatas, didapatkan nilai yang memenuhi syarat 3 penjurusan dengan cara nilai yang

memenuhi syarat masuk penjurusan IPA, IPS, dan Bahasa dicari minimal supportnya dengan cara total data yang mendukung nilai masuk jurusan ketiganya/total data keseluruhan dikalikan 100% sehingga didapatkan hasil $37/79 \times 100\% = 0,4683$ dimana 0,4683 telah memenuhi nilai minimal support yang telah ditentukan sebelumnya.

4.5. Hasil Nilai Confidence

Tabel 5. Hasil Nilai Confidence

A	B	SUP of A	SUP of B	SUP of A∩B	CON
A1	B1,S1	0,5696	0,6962	0,4683	0,8
S1	A1, B1	0,7722	0,5443	0,4683	0,59016
B1	A1, S1	0,9114	0,4557	0,4683	0,5
A1, B1	S1	0,5570	0,7595	0,4683	0,81818
S1, B1	A1	0,6962	0,5696	0,4683	0,65455
A1, S1	B1	0,4684	0,9114	0,4683	0,97297

Keterangan:

- A: antesedens/paruh prtama/if
- B: consequent/paruh kedua/then
- SUP of A: nilai support A
- SUP of B: nilai support B
- SUP of A∩B: nilai support yang mengandung item A dan B
- CON: nilai kepercayaan

Pada tabel diatas merupakan hasil nilai confidence yang didapatkan dari hasil pemecahan tabel 3 frequent itemsets. Setelah ditentukan minimal confidence 0,85 didapatkan nilai confidence yang memenuhi kriteria minimum yaitu 0,9729 dengan kombinasi itemset A1,S1→B1.. Hasil didapatkan dengan cara support A∩B/support of A dan didapatkan nilai confidence sebagaimana tabel diatas 0,9729 yang selanjutnya akan diuji menggunakan Lift Ratio sebagai pengujian akurasi algoritma.

4.6. Pengujian

Berdasarkan uji coba dataset yang telah dilakukan didapatkan nilai minimum support yaitu 0.45 dan minimum confidence yaitu 0.85. dan didapatkan hasil sebagaimana tabel diatas dimana dari item diatas nilai confidence yang lolos yaitu 0.972 dengan item A1,S1→B1.

Pengujian selanjutnya yaitu lift ratio yang bertujuan untuk mengetahui kuat tidaknya aturan asosiasi yang telah terbentuk. Pada akhirnya sebuah kombinasi itemset dinyatakan valid dan kuat jika nilai lift ratio > 1.

Pada penelitian ini, pengkaji menggunakan lift ratio sebagai pengujian akurasi algoritma, setelah didapatkan hasil nilai confidence sebagaimana tabel 4.9, maka dilakukan pengujian akurasi algoritma apriori dengan rumus berikut:

$$LiftRasio = \frac{confidence}{benchmark\ confidence} \quad (11)$$

$$ConfidenceBenchmark = \frac{\sum transaksidenganitemdalamconsequent}{\sum transaksidalamdatabase} \quad (12)$$

Berikut ini adalah tabel hasil pengujian lift ratio terhadap hasil dari analisis data calon peserta didik baru di SMAK Diponegoro Blitar menggunakan algoritma apriori:

Tabel 6. Hasil Pengujian Korelasi Lift

ant	cons	ant supp	cons supp	supp	conf	lift ratio	korelasi
A1	B1	0,569620	0,9113924	0,46	0,97	1,07283951	positif
S1	B1	0,772152	0,9113924	0,69	0,90	0,98929872	negatif
A1,S1	B1	0,468354	0,9113924	0,46	1,000	1,0972	positif

Keterangan:

- Ant: pendahuluan/if
- Cons: akibat/then
- Ant supp: nilai pendukung pendahuluan
- Cons supp: nilai pendukung akibat
- Supp: pendukung
- Conf: kepercayaan atau penunjang
- Lift ratio: nilai yang didapatkan
- Korelasi: keterangan hasil pengujian

Adapun nilai dari pengujian lift ratio dapat dilihat pada tabel 4.10 diatas Aini. berdasarkan pengujian lift ratio dapat diambil analisis bahwa nilai support dan nilai confidence akan mempengaruhi besarnya lift ratio. Semakin rendah nilai support dan semakin tinggi nilai confidence maka nilai lift ratio akan semakin besar. Begitu pula dengan semakin tinggi support dan semakin rendah nilai confidence maka nilai lift ratio akan semakin kecil. Besar kecilnya nilai lift ratio menunjukkan kuatnya hubungan dalam aturan asosiasi tersebut. hal ini ditunjukkan dengan jika nilai lift ratio lebih besar dari 1 berarti hubungan antar item pada pola yang dihasilkan semakin kuat.

Pola yang memiliki nilai lift ratio tertinggi pada itemset A1,S1→B1 dengan nilai lift 1,09 dan berkorelasi Positif.

4.7. Pengujian Sistem

Dalam pengujian sistem, penguji menggunakan Google Colab untuk menguji keakuratan perhitungan algoritma dengan hasil sebagai berikut:

```
def APRIORI MY(data, min support=0.45, max
_length = 7):
    support = {}
    L = set(data.columns)

    for i in range(1, max length+1):
        c = list(combinations(L,i))
        L = set()

        for j in list(c):
            sup = data.loc[:,j].product(ax
is=1).sum()/len(data.index)
            if sup >= min support:
                support[j] = sup

    L = set(set(L) | set(j))

    result = pd.DataFrame(list(support.ite
ms()), columns = ['items', 'support'])
    return(result)
```

Source code diatas merupakan source code untuk mencari nilai support 1 itemset, 2 itemset dan 3 itemset dan didapatkan hasil sebagai berikut:

	items	support
0	(A.)	0.569620
1	(B.)	0.911392
2	(S.)	0.772152
3	(A, B)	0.556962
4	(A, S)	0.468354
5	(B, S)	0.696203
6	(A, B, S)	0.468354

Gambar 2. Hasil frequent itemset

```
def ASSOCIATION_RULE(df, min_confidence=0.85):
    support = pd.Series(df.support.values,
        index=df['items'].values).to_dict()
    data = []
    L = df['items'].values

    p = list(permutations(L,2))

    for i in p:
        if set(i[0]).issubset(i[1]):
            conf = support[i[1]]/support[i[0]]
            if round(conf, 2) >= min confidence:
                j = i[1][not i[1].index(i[0][0])]
                lift = support[i[1]]/(support[i[0]]* support [(j,))
                #lift = conf/(support[(j,))
                data.append([i[0], (j,), support[i[0]], support[(j,)], support[i[1]], conf, lift])
                result = pd.DataFrame(data, columns=['antecedents', 'consequents', 'antecedent support', 'consequent support', 'support', 'confidence', 'lift'])
    return(result)
```

Source code diatas merupakan source code untuk mendapatkan nilai confidence dan mendapatkan hasil pengujian, didapatkan nilai sebagai berikut:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	(A.)	(B.)	0.569620	0.911392	0.556962	0.977778	1.072840
1	(S.)	(B.)	0.772152	0.911392	0.696203	0.901639	0.989299
2	(A, S)	(B.)	0.468354	0.911392	0.468354	1.000000	1.097222

Gambar 3. hasil pengujian

```
my_assoc['korelasi'] = np.select([(my_assoc['lift'] == 1), (my_assoc['lift'] > 1), (my_assoc['lift'] < 1)], ['Independen', 'Positif', 'Negatif'])
my_assoc
```

Source code diatas digunakan untuk mendapatkan korelasi positif dan negatif dengan hasil sebagai berikut:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	korelasi
0	(A.)	(B.)	0.569620	0.911392	0.556962	0.977778	1.072840	Positif
1	(S.)	(B.)	0.772152	0.911392	0.696203	0.901639	0.989299	Negatif
2	(A, S)	(B.)	0.468354	0.911392	0.468354	1.000000	1.097222	Positif

Gambar 4. hasil pengujian korelasi lift

Dari hasil pengujian sistem yang telah pengkaji lakukan seperti halnya diatas, dapat disimpulkan bahwa pengujian lift ini akurat dengan hasil pengujian tertinggi 1,09 dan berkorelasi positif dengan rule A,S→B atau hasilnya akurat dan sama dengan pengujian manual.

4.8. Pembahasan

Pada penelitian ini, pengkaji ingin menyampaikan bahwa setelah pengambilan data dari pihak sekolah, data dibersihkan dengan cara data cleaning dimana data yang tidak dielukan dihapus. Setelah melalui tahap data celaning, data selanjutnya ditransformasi kedalam angka biner 0 dan 1 dimana 0 merupakan data yang tidak memenuhi target dan 1 merupakan data yang memenuhi target. Tahap selanjutnya adalah penerapan algoritma apriori pada total 79 data tersebut ditentukan minimum support sebesar 0,46 dan minimum confidence sebesar 0,85 dan dihasilkan rule terkuat yaitu A1,S1→B1 dengan confidence 0.97297. Tahap selanjutnya adalah pengujian manual menggunakan pengujian akurasi algoritma apriori yaitu Lift Ratio, yang apabila dari perhitungan tersebut menghasilkan nilai dibawah 1 maka terdapat korelasi negatif. Untuk perhitungan yang menghasilkan nilai diatas 1 atau >1 maka terdapat korelasi positif. dan perhitungan dan pengujian akurasi menggunakan Microsoft Excel sebagai perhitungan manual dan Google Colab sebagai perhitungan sistem, dimana itemset dengan nilai confidence tertinggi yang diuji menggunakan lift ratio oleh pengkaji mendapatkan hasil pengujian tertinggi 1,09 dan berkorelasi positif dengan rule A,S→B atau hasilnya akurat dan sama dengan pengujian manual sebagaimana yang dilampirkan pada Tabel 4.10 Hasil Pengujian Korelasi lift yang mana diuji menggunakan Microsoft Excel dan Gambar 4.3 hasil pengujian korelasi lift yang mana diuji menggunakan Google Colab.

5. KESIMPULAN DAN SARAN

Penelitian ini adalah untuk mengetahui dan menganalisa data nilai calon peserta didik baru untuk melakukan seleksi penjurusan menggunakan algoritma apriori dengan minimum support 0,45 dan minimum confidence 0,85. Berdasarkan hasil pembahasan dapat disimpulkan sebagai berikut: Dengan 79 data calon peserta didik baru serta dengan minimum sopport 0,45 dan minimum confidence 0.85 didapatkan frequent itemset terkuat dengan pola

A1,S1→B1 dengan nilai support 0,4557 dan confidence 0,973 yang berarti dari 79 pendaftar, 46 persennya direkomendasikan masuk kedalam jurusan IPA dan IPS, Dari pengujian yang dilakukan dengan menggunakan perhitungan maupun pengujian dengan sistem, hasil yang didapatkan bernilai sama. Dengan mengambil pola itemsets tertinggi dengan minimum support 0,45 dan minimum confidence 0.85 serta lift ratio dikatakan akurat atau valid ketika nilai hasil pengujian >1. Pengujian menggunakan lift ratio pada penelitian ini dapat dikatakan akurat atau valid dengan hasil 1,09 dan berkorelasi positif dengan rule A1,S1→B1 sehingga pada penelitian ini penggunaan algoritma apriori dalam seleksi jurusan dapat dinyatakan valid maupun akurat.

Untuk menambah referensi penelitian selanjutnya ada beberapa saran yang dikemukakan sebagai berikut: Analisa yang dihasilkan pada skripsi ini merupakan analisa yang mendasar dan perlu dikembangkan lagi. Akan lebih baik jika pengembangan selanjutnya analisa dilakukan lebih spesifik, kemudian disertai proses implementasi dan ujicoba menggunakan aplikasi data mining.

DAFTAR PUSTAKA

- [1] N. G. A. A. L. Dewi, L. E. Tripalupi, and M. Artana, "Pengaruh Pelaksanaan Pembelajaran dan Kebiasaan Belajar Terhadap Hasil Belajar Ekonomi Kelas X SMA Lab Singaraja," *Tesis*, p. 1, 2015, [Online]. Available: <https://media.neliti.com/media/publications/5192-ID-pengaruh-pelaksanaan-pembelajaran-dan-kebiasaan-belajar-terhadap-hasil-belajar-e.pdf>
- [2] Nizarman, "Manajemen Penerimaan Siswa Baru," *Manajer Pendidik.*, vol. 9, no. 2, pp. 224–234, 2015.
- [3] Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika. Jurnal Edik Informatika*, 2.
- [4] J. Halim, "Penerapan Data Mining Untuk Mengukur Tingkat Kepuasan Siswa Terhadap Pelayanan Di Bimbingan Belajar Al-Misbah Dengan Menggunakan Metode K-Means," *J. SAINTIKOM*, vol. 16, no. 1, pp. 1–6, 2017, [Online]. Available: <https://prpm.trigunadharma.ac.id/public/fileJurnal/hplH11Jufri.pdf>
- [5] J. Han and M. Kamber, "Data mining concepts and techniques / Jiawei Han, Micheline Kamber." 2006. [Online]. Available: <http://search.ebscohost.com/login.aspx?direct=true&db=cat02326a&AN=usl.1498894&site=eds-live>
- [6] Saragih, R. A., Damanik, I. S., & Saragih, I. S. (2021). Implementasi Metode Apriori Dalam Pembagian Jurusan Penerimaan Siswa Baru Di SMK Negeri 1 Siantar. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 6(1), 181.
- [7] Fitriati, D. (2016). Implementasi Data Mining untuk Menentukan Kombinasi Media Promosi Barang Berdasarkan Perilaku Pembelian Pelanggan Menggunakan Algoritma Apriori. *Prosiding ANNUAL RESEARCH SEMINAR 2016*, 2(1), 472–480. <https://doi.org/979-587-626-0>
- [8] H. Kusumo, E. Sedyono, and M. Marwata, "Analisis Algoritma Apriori untuk Mendukung Strategi Promosi Perguruan Tinggi," *Walisongo J. Inf. Technol.*, vol. 1, no. 1, p. 49, 2019, doi: 10.21580/wjit.2019.1.1.4000.
- [9] E. H. Pujiarini *et al.*, "Analisis Algoritma Apriori untuk Mendukung Strategi Promosi Perguruan Tinggi," *J. Edik Inform.*, vol. 6, no. 2, p. 49, 2019, doi: 10.15294/sji.v5i2.16199.
- [10] Listriani, D., Setyaningrum, A. H., & Eka, F. (2018). PENERAPAN METODE ASOSIASI MENGGUNAKAN ALGORITMA APRIORI PADA APLIKASI ANALISA POLA BELANJA KONSUMEN (Studi Kasus Toko Buku Gramedia Bintaro). *Jurnal Teknik Informatika*, 9(2), 120–127. <https://doi.org/10.15408/jti.v9i2.5602>
- [11] D. Winarso and A. Karnaidi, "Association Rule Mining Untuk Meningkatkan Promosi Produk (Studi Kasus Pada Pd. Xyz)," *J. Fasilkom*, vol. 7, no. 2, pp. 284–288, 2018, doi: 10.37859/jf.v7i2.789.
- [12] J. Jooa, S. Bangb, and G. Parka, "Implementation of a Recommendation System Using Association Rules and Collaborative Filtering," *Procedia Comput. Sci.*, vol. 91, no. Itqm 2016, pp. 944–952, 2016, doi: 10.1016/j.procs.2016.07.115.
- [13] E. Asril, F. Wiza, and Y. Yunefri, "Analisis Data Lulusan dengan Data Mining untuk Mendukung Strategi Promosi Universitas Lancang Kuning," *J. Teknol. Inf. Komun. Digit. Zo.*, vol. 6, no. 2, pp. 24–32, 2015.
- [14] F. Rumaisa, "Penentuan Association Rule Pada Pemilihan Program Studi Calon Mahasiswa Baru Menggunakan Algoritma Apriori Studi Kasus Pada Universitas Widyatama Bandung," *Semin. Nas. Apl. Teknol. Inf.*, vol. 2012, no. Snati, pp. 15–16, 2012.