

Implementasi *Data Warehouse* dan *Data Mining* Untuk Pengembangan Sistem Rekomendasi Pemilihan SMA

Yusuf Sulistyio Nugroho^{1*}, Triana Dewi Salma¹, Sigid Rokhanudin¹

¹Jurusan Informatika Fakultas Komunikasi dan Informatika

Universitas Muhammadiyah Surakarta

*yusuf.nugroho@ums.ac.id

ABSTRAK

Jumlah penyelenggara pendidikan di Sragen telah mengalami peningkatan dalam kurun beberapa tahun terakhir. Hal ini juga salah satu akibat dari semakin banyaknya jumlah siswa di wilayah tersebut. Namun peningkatan jumlah ini ternyata tidak diimbangi dengan pengelolaan data sekolah yang baik misalnya menggunakan teknologi informasi. Tidak adanya pusat data yang terintegrasi secara baik menyebabkan pihak sekolah kesulitan menyampaikan informasi akademik dan masyarakat sulit mendapatkan informasi yang valid sehingga banyak calon siswa yang ragu untuk menentukan sekolah menengah tingkat atas yang akan dituju sesuai dengan kriterianya. Berdasarkan hal tersebut, sebuah data warehouse diperlukan sebagai pengelola data sekolah-sekolah secara terintegrasi dengan baik dan dikembangkan sistem rekomendasi dengan menerapkan teknik data mining. Dengan demikian, data sekolah dapat digali untuk menghasilkan suatu informasi yang dapat dimanfaatkan sebagai rekomendasi bagi calon siswa sekolah menengah atas untuk menentukan pilihan sekolah yang dituju. Sebuah diagram snowflake dirancang sebagai langkah awal dalam pengembangan data warehouse. Data yang dikirim ke data warehouse terlebih dahulu diproses melalui tahap ekstraksi, transformasi dan load. Sementara itu, sistem rekomendasi dibangun dengan menerapkan metode naïve bayes dengan cara menghitung probabilitas masing-masing kriteria yang diajukan berdasarkan data pelatihan dalam data warehouse. Adapun kriteria yang digunakan yaitu biaya sekolah, jarak sekolah, nilai akreditasi, tingkat kelulusan, dan nilai rerata ujian akhir nasional. Hasil penelitian menunjukkan bahwa sebuah data warehouse telah berhasil dibangun sebagai pengelola data sekolah di Sragen yang terintegrasi dan terhubung dengan sistem rekomendasi untuk membantu calon siswa memilih sekolah yang sesuai kriteria-kriteria yang diajukan. Rekomendasi yang dihasilkan oleh sistem berdasarkan nilai probabilitas tertinggi dari setiap variabel masing-masing sekolah, tingkat prioritas variabel, serta urutan ranking sekolah.

Kata kunci: *data mining, data warehouse, naive bayes, sistem rekomendasi sekolah*

1. Pendahuluan

Pendidikan di berbagai wilayah di Indonesia dalam beberapa tahun terakhir telah menunjukkan adanya kemajuan dan perkembangan. Peningkatan pendidikan ini juga turut dirasakan di Kota Sragen berdasarkan

jumlah sekolah, guru dan siswa yang semakin meningkat. Berdasarkan data sekolah dari Dinas Pendidikan dan Kebudayaan Sragen tahun 2015 [1], jumlah instansi pendidikan dan siswa di Sragen telah mengalami perkembangan meskipun tidak begitu signifikan seperti yang ditunjukkan pada tabel 1.

Tabel 1. Perkembangan Sekolah Menengah Atas di Sragen 2014/2015

Tahun	Jumlah Sekolah	Jumlah Guru	Jumlah Murid
2011/2012	71	2573	14850
2012/2013	69	2590	15355
2013/2014	74	2642	16334
2014/2015	74	2700	16366

Sumber : Dinas Pendidikan & Kebudayaan Kota Sragen, 2015

Di sisi lain, dengan adanya peningkatan jumlah sekolah ataupun siswa ini ternyata tidak diimbangi dengan pengelolaan data sekolah yang terintegrasi dengan baik menggunakan perangkat teknologi informasi, seperti komputer. Hal ini dapat menyebabkan pihak sekolah mengalami kesulitan dalam menyampaikan informasi akademik termasuk informasi pendaftaran siswa baru dan masyarakat juga sulit mendapatkan informasi yang valid sehingga banyak calon siswa yang mengalami kebingungan untuk menentukan sekolah menengah tingkat atas yang akan dituju. Terlebih lagi, calon siswa biasanya perlu mendatangi secara langsung ke sekolah untuk mendapatkan informasi valid terkait persyaratan-persyaratan pendaftaran siswa baru yang diminta oleh pihak sekolah. Untuk mengatasi hal ini, maka sebuah pusat data yang baik sebagai pengelola data dan terintegrasi dengan sebuah sistem informasi diperlukan, yaitu dengan dibangunnya sebuah *data warehouse* [2]. Hal ini diperlukan karena data akademik sekolah akan terus bertambah dari tahun ke tahun terutama data siswa.

Namun seiring berjalannya waktu jika data yang terus bertambah dan tersimpan di dalam *data warehouse* tersebut hanya dibiarkan menumpuk, maka akan menjadi tidak bermanfaat. Padahal data yang tersimpan tersebut bisa digunakan untuk menghasilkan sebuah informasi yang bermanfaat. Oleh karena itu, sebuah teknik untuk menggali data yang berlimpah tersebut diperlukan untuk menghasilkan suatu informasi yang dapat dimanfaatkan bagi pihak-pihak yang membutuhkan. Salah satu teknik yang dapat diterapkan untuk menggali informasi dari data yang berlimpah adalah *data mining*. *Data mining* dapat digunakan untuk mengelola data pendidikan yang berlimpah secara lebih baik agar proses pembelajaran bisa dilaksanakan secara efektif [3].

Sementara itu, setiap awal tahun akademik pelajaran selalu diiringi dengan aktifitas pendaftaran siswa baru mulai dari tingkat taman kanak-kanak sampai tingkat pendidikan tinggi, termasuk sekolah menengah tingkat atas. Namun tentunya hal ini juga terdapat kendala-kendala yang dirasakan oleh para calon siswa dan orang tua, yaitu informasi yang disampaikan tidak jelas dan bahkan mengalami kebingungan dalam memilih sekolah [4]. Persyaratan-persyaratan yang diminta oleh pihak sekolah terkadang sulit dipenuhi misalnya biaya pendaftaran dan uang gedung yang dirasa mahal, atau bahkan faktor-faktor lain yang terkadang juga menjadi kendala bagi calon siswa untuk menentukan sekolah, misalnya terkait dengan jarak sekolah dimana calon siswa biasanya akan memilih sekolah yang jaraknya lebih dekat.

Sebagai dasar, pada sebuah penelitian menunjukkan bahwa sistem rekomendasi sangat berguna dalam membantu permasalahan dalam bidang transportasi umum [5]. Sistem rekomendasi dalam penelitian ini dikembangkan dengan tujuan untuk membantu sopir taksi dalam rangka mencari penumpang dengan cepat, dan membantu calon penumpang untuk mendapatkan taksi secara lebih mudah. Data yang digunakan dalam pengembangan sistem adalah data trayek taksi berbasis

GPS dalam jumlah yang besar. Informasi rekomendasi yang diberikan kepada sopir taksi diambil berdasarkan pola mobilitas penumpang, sedangkan informasi yang diberikan kepada calon penumpang diambil berdasarkan pola lokasi penjemputan dan penurunan penumpang dari data trayek taksi berbasis GPS. Metode penentuan pola trayek taksi menggunakan teknik *density-based clustering* (OPTICS). Hasil penelitian menunjukkan bahwa sistem rekomendasi telah berhasil membantu sopir taksi mencari calon penumpang dengan lebih mudah, dan membantu calon penumpang mencari taksi secara lebih cepat. Sehingga hal ini dapat menghemat waktu bagi kedua pihak, mengurangi tingkat kepadatan lalu lintas terutama di kota-kota besar, serta dapat mengurangi tingkat polusi yang dihasilkan oleh taksi ketika harus berkeliling mencari penumpang.

Sedangkan pada penelitian lain menyatakan bahwa sistem rekomendasi dapat diterapkan dalam rangka membantu para pelancong untuk mendapatkan tempat penginapan (hotel) yang baik dan sesuai dengan kebutuhan [6]. Sistem rekomendasi ini dikembangkan dengan menggunakan data review yang ditulis oleh para wisatawan lain yang pernah menginap di sebuah hotel dengan menerapkan teknik *text-mining* menggunakan algoritma *unsupervised clustering*. Secara prinsip, review tertulis dapat digunakan untuk menggantikan informasi lisan terkait dengan penilaian sebuah objek. Review yang ditulis baik bersifat positif maupun negatif secara khusus didefinisikan menjadi 3 kategori informasi. Pertama, tujuan kunjungan yang dibedakan menjadi 5 jenis yaitu kunjungan bisnis, liburan perorangan, keluarga, kelompok atau organisasi, dan pasangan. Kedua adalah kewarganegaraan. Ketiga adalah kriteria hotel yang diinginkan oleh pengunjung. Hasil mining dari review kriteria hotel ini ditemukan informasi yang paling sering dibicarakan yaitu lokasi, tingkat pelayanan, makanan, kamar, harga, dan fasilitas hotel. Hasil penelitian menunjukkan bahwa informasi dapat ditemukan dari *preprocessing* sampel teks dalam jumlah yang besar. Teks review dapat digali untuk mengidentifikasi jenis-jenis hotel dan fasilitasnya. Selain itu, hasil yang menarik dari eksperimen menunjukkan bahwa tidak ada korelasi yang kuat antara teks yang ditulis dalam review dengan rating yang diberikan oleh penulisnya.

Sementara itu, pada penelitian tentang sistem rekomendasi dalam bidang penjualan menyatakan bahwa penggabungan beberapa produk menjadi satu item dapat memenuhi kebutuhan konsumen sekaligus, sementara keuntungan bagi pihak perusahaan adalah dapat meningkatkan pelanggan dan secara otomatis pendapatan perusahaan [7]. Tujuan pengembangan sistem rekomendasi penggabungan produk ini adalah untuk memperluas pasar yang mencakup produk-produk baru yang kemungkinan besar tidak akan dibeli oleh konsumen jika bukan bagian dari paket produk. Sehingga dengan penggabungan menjadi satu paket dengan produk lain dapat meningkatkan pendapatan dan keuntungan perusahaan. Metode yang digunakan untuk

menggabungkan paket produk adalah dengan pendekatan *collaborative filtering* (CF) yang merupakan metode paling populer dan efektif untuk mengembangkan sistem rekomendasi [8]. Metode ini secara iteratif mengkluster konsumen berdasarkan perspektifnya terhadap pola rating dan hubungan tingkat kepercayaan sosial. Hasil penelitian menunjukkan bahwa sistem telah dikembangkan dengan memanfaatkan metode rekomendasi item tunggal, dan menggabungkan dengan rekomendasi paket, sehingga dapat meningkatkan lingkup pembeli dan pendapatan perusahaan. Dengan demikian meningkatkan nilai yang diberikan oleh sistem rekomendasi.

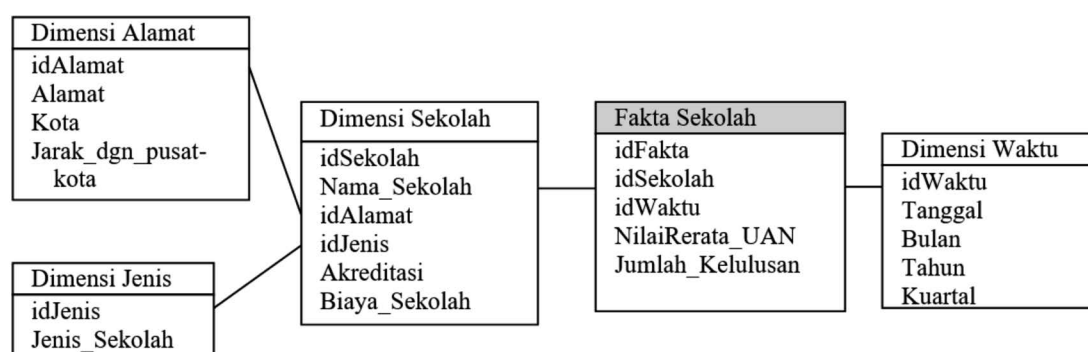
Berdasarkan latar belakang yang dialami oleh calon siswa di Sragen serta referensi beberapa penelitian tersebut, maka sebuah sistem diperlukan untuk memberikan informasi terkait prosedur pendaftaran yang diintegrasikan dengan sebuah sistem rekomendasi untuk membantu dalam pemilihan sekolah sesuai yang diharapkan oleh calon siswa. Sistem ini dikembangkan dengan memanfaatkan *data warehouse* sebagai pusat datanya dan menerapkan teknik *data mining* untuk menghasilkan

informasi rekomendasi yang dapat dijadikan sebagai saran bagi calon siswa untuk memilih sekolah menengah atas sesuai dengan kriteria dan persyaratan yang diajukan.

2. Metodologi

A. Snowflake Diagram

Data dalam *data warehouse* disimpan pada sebuah tabel fakta dan beberapa tabel dimensi. Data transaksi yang terkait akan diinputkan ke dalam tabel fakta, sedangkan data objek yang dapat diukur dan dianalisis akan dimasukkan ke dalam tabel dimensi dan dapat direlasikan ke tabel fakta jika diperlukan [9]. *Data warehouse* ini dikembangkan dengan menganut pola *snowflake diagram* karena penerapan struktur *snowflake* ini dapat menghemat waktu untuk akses data. Namun di sisi lain penyimpanan data akan memerlukan lebih banyak kapasitas *storage* karena jumlah data yang bisa mencapai ribuan bahkan jutaan transaksi akibat sifat *data warehouse* yang tidak diijinkan untuk melakukan *data manipulation language* (DML) kecuali untuk aktifitas input data saja. Diagram *snowflake* dirancang sebagai dasar pengembangan *data warehouse* seperti pada gambar 1.

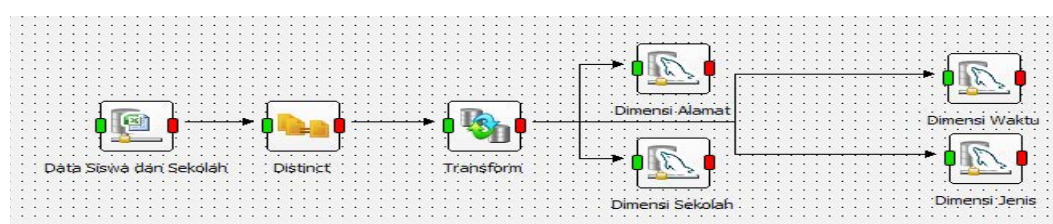


Gambar 1. Rancangan *Snowflake Diagram* dalam *Data Warehouse*

B. Proses ETL (Extract, Transform and Load)

Proses ETL memiliki tujuan untuk mengirim data transaksional yang dimiliki menuju ke suatu *data warehouse*. Proses ini diperlukan karena data dari database

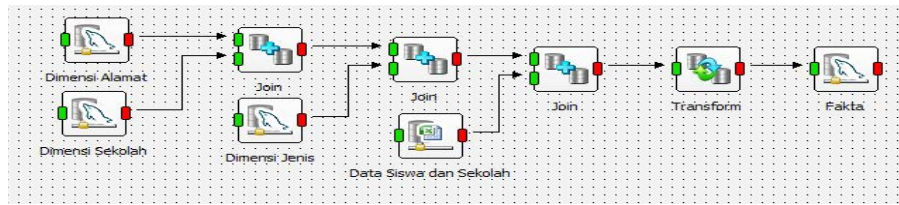
transaksional masih mengandung kesalahan-kesalahan yang tidak sesuai dengan format *data warehouse*. Diagram proses ETL yang dilakukan untuk menghasilkan 4 buah dimensi secara keseluruhan dapat dilihat pada gambar 2.



Gambar 2. Diagram Proses ETL untuk 4 Dimensi

Data sumber diambil dari data sekolah yang dimiliki oleh pihak Dinas Pendidikan Kota Sragen yang masih tersimpan dalam bentuk spreadsheet (*Microsoft Excel*). Data tersebut berisi tentang data semua siswa yang terdaftar di sekolah-sekolah menengah tingkat atas di Kota Sragen. Berdasarkan data tersebut, maka dapat dibentuk menjadi 4 (empat) buah dimensi yaitu dimensi sekolah, dimensi alamat, dimensi waktu, dan dimensi jenis sekolah.

Sementara itu, data transaksi yang berisi data siswa dari tahun ke tahun ditransformasikan menjadi sebuah tabel fakta dengan dihubungkan dengan keempat dimensi yang sudah dibentuk pada proses ETL sebelumnya. Proses ETL untuk pembuatan tabel fakta dapat dilihat pada gambar 3.



Gambar 3. Diagram Proses ETL untuk Tabel Fakta

C. Pengolahan Data

Sistem rekomendasi dikembangkan menggunakan data yang terdapat dalam *data warehouse* yang telah dibangun berdasarkan rancangan *snomflake*. Data tersebut dihubungkan ke dalam sistem dan kemudian akan diolah menggunakan algoritma *naïve bayes* untuk menghasilkan suatu informasi strategis yang berupa rekomendasi kepada calon siswa sekolah menengah berdasarkan kriteria-kriteria yang diajukan. Berikut tahap-tahap yang dilakukan dalam pengembangan sistem rekomendasi.

1) Penentuan Variabel Data Pelatihan

Data pelatihan adalah data yang sudah terjadi dan tersimpan dalam database sesuai dengan kriteria-kriteria yang disediakan. Tabel 2 merupakan rincian variabel yang digunakan dalam data pelatihan untuk melakukan perhitungan *naïve bayes* dengan tujuan menghasilkan nilai probabilitas kesesuaian sekolah terhadap kriteria yang diajukan.

Tabel 2. Variabel dalam Data Pelatihan

Variabel	Atribut	Keterangan
Y	Nama Sekolah	Nama Sekolah yang akan diprediksikan. Meliputi 10 SMA sederajat di Kota Sragen
X1	Biaya SPP	Merupakan biaya yang harus dibayarkan setiap semester di sekolah.
X2	Jarak Sekolah	Merupakan jarak sekolah yang diukur dari pusat kota yaitu dari titik Kantor Bupati Kabupaten Sragen.
X3	Akreditasi Sekolah	Merupakan nilai akreditasi sekolah yang bersangkutan.
X4	Tingkat kelulusan siswa	Merupakan tingkat kelulusan dari siswa sekolah yang bersangkutan.
X5	Nilai Rata-rata UAN	Nilai minimum rata-rata UAN yang diterima oleh sekolah yang bersangkutan. Dalam hal ini diambil dari nilai standar minimal tertinggi dari data 5 tahun.

2) Penentuan Variabel Data Uji

Data uji merupakan data yang belum terjadi dan dapat digunakan untuk melakukan percobaan atau prediksi terhadap aturan-aturan yang diperoleh berdasarkan analisis data pelatihan [10]. Data pengujian dalam penelitian ini adalah data siswa yang sudah lulus SMP sederajat ataupun siswa yang masih berstatus aktif sebagai siswa kelas 3 SMP dan akan melanjutkan ke jenjang SMA atau sederajat. Data siswa ini diambil secara random dari beberapa siswa yang berada di Sragen. Atribut yang digunakan

dalam data uji ini ada atribut X1 sampai dengan X5 dalam tabel 1.

D. Perhitungan Naïve Bayes

Naïve Bayes adalah salah satu algoritma klasifikasi yang paling efektif dan efisien. Sebuah *classifier* yang memiliki label *class*, dibangun dari sekumpulan data pelatihan dengan label *class* [11]. Algoritma *Naïve Bayes* dapat dihitung dengan menggunakan persamaan 1 [12] untuk mencari nilai probabilitas sekolah sesuai dengan variabel yang diajukan oleh pengguna.

$$P(S | X) = \frac{P(X | H) \cdot P(H)}{P(X)} \dots\dots\dots(1)$$

E. Pengembangan Sistem Rekomendasi

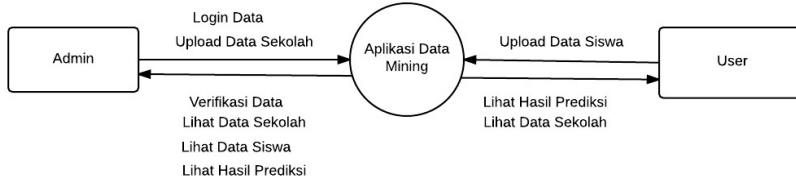
Sistem rekomendasi merupakan sistem penyaring informasi yang bertujuan untuk memprediksi nilai rating terhadap pengguna dan items, terutama yang berasal dari data besar untuk merekomendasikan tingkat kesukaan atau pilihan pengguna sistem [13]. Sementara itu, sistem rekomendasi dalam penelitian ini dikembangkan dengan berbasis web untuk memudahkan para calon siswa pencari sekolah menengah lebih mudah mendapatkan

informasi masing-masing sekolah dan dapat menentukan sekolah sesuai dengan kriteria yang diinginkan. Sistem ini dibangun dengan menerapkan algoritma *naïve bayes* (persamaan 1) untuk menghasilkan informasi yang bermanfaat bagi calon siswa sebagai rekomendasi. Dengan adanya sistem ini, calon siswa akan lebih mudah dalam menentukan sekolah sesuai dengan kriteria dan persyaratan yang diajukan. Adapun sistem rekomendasi ini dikembangkan dengan tahapan sebagai berikut:

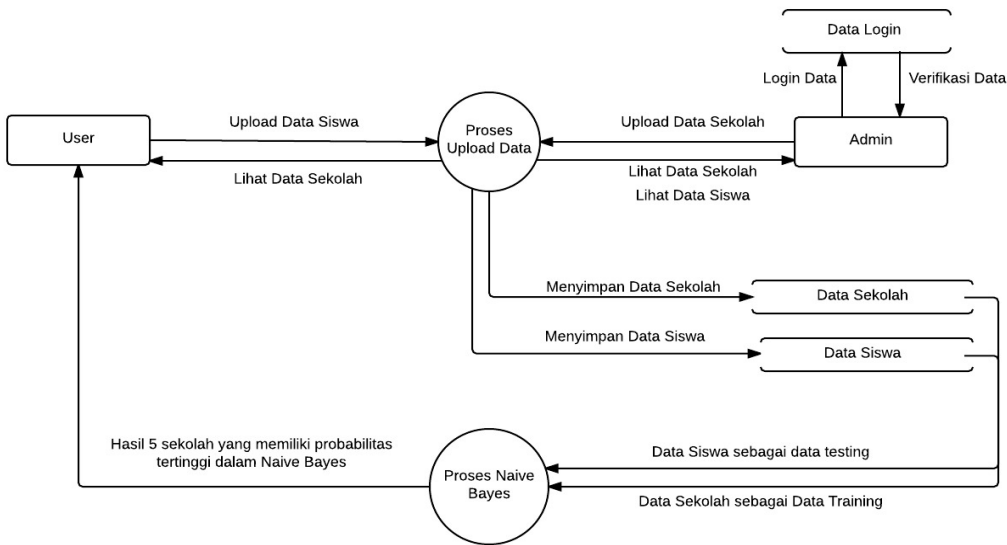
1) **Perancangan Diagram Alir Data**

Sistem rekomendasi dikembangkan untuk digunakan oleh administrator sebagai user utama dan calon siswa sebagai user umum. Sehingga

sistem perlu disesuaikan alur datanya. Gambar 4 menunjukkan diagram alir data pada level 0, sedangkan gambar 5 memperlihatkan diagram alir data untuk level 1.



Gambar 4. Diagram Alir Data Level 0

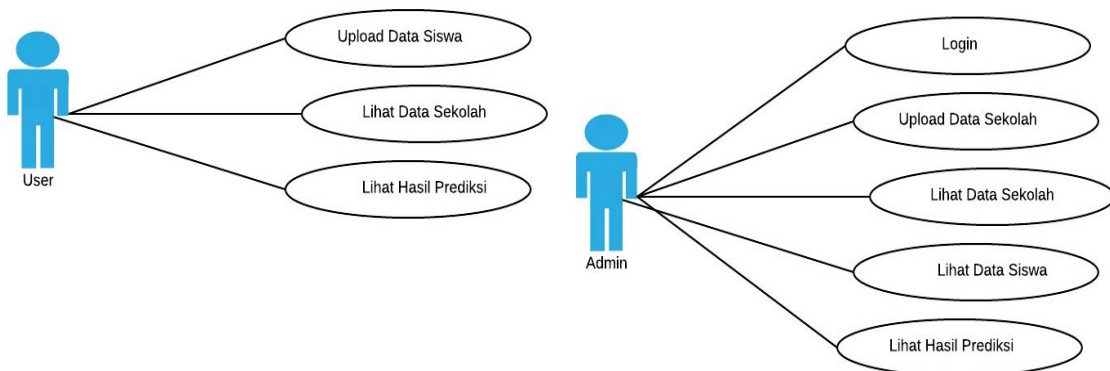


Gambar 5. Diagram Alir Data Level 1

2) **Perancangan Diagram Use Case**

Sesuai dengan kriteria user yang dibedakan menjadi 2 kategori maka diagram *use case* dirancang

dengan 2 kategori, yaitu *use case diagram* untuk administrator (gambar 6.a) dan *use case diagram* untuk user umum (gambar 6.b).



Gambar 6. *Use Case Diagram* User Umum dan Administrator

3. Hasil dan Pembahasan

A. Data Pelatihan dan Data Pengujian

Data yang digunakan dalam sistem rekomendasi ini sepenuhnya diambil dari tabel fakta yang telah dibangun

dalam *data warehouse*. Data tersebut digunakan sebagai data pelatihan (*training data*). Tabel 3 menunjukkan potongan data sekolah menengah atas berdasarkan variabel yang telah ditentukan dalam data pelatihan.

Tabel 3. Potongan Data Pelatihan

Nama Sekolah	Biaya SPP (Rupiah)	Jarak (Km)	Akreditasi	Persentase Kelulusan	Nilai UAN
MAN 1 Sragen	85.000	2,4	A	100%	16,9
SMA Negeri 1 Sragen	200.000	2,1	A	100%	32
SMA Muhammadiyah 1 Sragen	250.000	4,2	A	95%	16,02
SMK Muhammadiyah 4 Sragen	270.000	2,2	B	89%	12,15
SMK Negeri 1 Sragen	85.000	1,5	A	95%	34,85

Sedangkan tabel 4 menunjukkan contoh data uji yang akan digunakan untuk menguji sistem rekomendasi. Data uji ini diperoleh dari salah satu calon siswa sekolah menengah atas sebagai user yang menggunakan sistem sebagai percobaan untuk menghasilkan rekomendasi terhadap sekolah yang dipilih. Data yang diinputkan oleh

user meliputi 5 variabel bebas, yaitu biaya SPP, jarak, akreditasi, tingkat kelulusan dan nilai ujian. Sedangkan nama sekolah adalah variabel yang akan dicari nilainya berdasarkan perhitungan *naïve bayes*. Informasi inilah yang akan disampaikan kepada user sebagai rekomendasi sekolah yang akan dituju.

Tabel 4. Contoh Data Uji Sistem Rekomendasi

Biaya SPP (Rupiah)	Jarak (Km)	Akreditasi	Tingkat Kelulusan	Nilai UAN	Nama Sekolah
X_1	X_2	X_3	X_4	X_5	Y
200.001-300.000	1-2	B	90-98%	31-40	?

B. Perhitungan *Naïve Bayes*

Perhitungan algoritma *Naïve Bayes* dengan menggunakan data sekolah sebagai data pelatihan yang diambil dari tabel 2 untuk mencari nilai probabilitasnya dan data kriteria dari tabel 3 yang diajukan oleh user sebagai data pengujian yang dipilih sebagai sampel. Berdasarkan hasil perhitungan maka dapat ditemukan sekolah-sekolah yang memiliki nilai probabilitas berdasarkan tingkat *confidence*-nya. Tabel 4 merupakan data sekolah yang terpenuhi kriterianya berdasarkan hasil perhitungan *naïve bayes*. Nilai 1 pada tabel 4 menunjukkan variabel yang

terpenuhi berdasarkan kriteria yang diajukan. Sedangkan prioritas variabel berdasarkan urutan dari yang tertinggi yaitu akreditasi, biaya SPP, nilai UAN minimal, tingkat kelulusan dan jarak sebagai prioritas terendah.

Sebagai contoh berdasarkan tabel 5, variabel SPP (X_1) pada SMA Muhammadiyah 1 Sragen memiliki nilai 1 karena variabel X_1 yang diajukan oleh calon siswa berdasarkan tabel 3 ($X_1 = 200.001-300.000$) dapat dipenuhi nilainya jika dilihat dari kriteria SPP dalam tabel 2 untuk SMA tersebut (X_1 SMA Muhammadiyah 1 Sragen = 250.000).

Tabel 5. Data Sekolah yang Terpenuhi dari Hasil Perhitungan *Naïve Bayes*

No	Y	X_1	X_2	X_3	X_4	X_5
1	MAN 1 Sragen	-	-	-	-	-
2	SMA N 1 Sragen	-	-	-	-	1
3	SMA N 2 Sragen	-	-	-	-	-
4	SMA N 3 Sragen	-	1	-	-	1
5	SMA Muhammadiyah 1 Sragen	1	-	-	1	-
6	SMK Muhammadiyah 1 Sragen	-	-	1	-	-
7	SMK Muhammadiyah 2 Sragen	-	-	-	1	-
8	SMK Muhammadiyah 4 Sragen	1	-	-	1	-
9	SMK N 1 Sragen	-	1	-	1	1
10	SMK N 2 Sragen	-	-	-	-	-

Data sekolah dalam tabel 5 dideskripsikan berdasarkan nilai variabel yang terpenuhi dari hasil perhitungan *naive bayes* sebagai berikut:

1. SMK N 1 Sragen, variabel yang terpenuhi = nilai UAN, Kelulusan, Jarak
2. SMK Muhammadiyah 4 Sragen, variabel yang terpenuhi = akreditasi, SPP
3. SMA Muhammadiyah 1 Sragen, variabel yang terpenuhi = SPP, Kelulusan
4. SMA Negeri 3 Sragen, variabel yang terpenuhi = nilai UAN, jarak
5. SMK Muhammadiyah 1 Sragen, variabel yang terpenuhi = akreditasi
6. SMA Negeri 1 Sragen, variabel yang terpenuhi = nilai UAN
7. SMK Muhammadiyah 2 Sragen, variabel yang terpenuhi = kelulusan

Sedangkan urutan sekolah berdasarkan tingkat favoritnya pada variabel Y yang memiliki variabel yang sama sebagai berikut:

1. SMKN 1 Sragen = rangking favorit 5
2. SMK Muhammadiyah 4 Sragen = rangking favorit 7
3. SMA Muhammadiyah 1 Sragen = rangking favorit 8
4. SMA Negeri 3 Sragen = rangking favorit 3
5. SMK Muhammadiyah 1 Sragen = rangking favorit 10
6. SMA Negeri 1 Sragen = rangking favorit 1
7. SMK Muhammadiyah 2 Sragen = rangking favorit 9

C. Sistem Rekomendasi

Sistem rekomendasi telah berhasil dikembangkan dengan berbasis *data warehouse*. Sistem ini secara umum dapat digunakan untuk menyampaikan berbagai informasi terkait akademik sekolah yang ada di Sragen. Namun secara khusus, sistem ini digunakan untuk memberikan informasi strategis yang dapat berupa rekomendasi pilihan sekolah yang dituju sesuai dengan kriteria-kriteria yang diajukan oleh calon siswa sekolah menengah atas.

Adapun sistem rekomendasi ini terdapat 5 menu utama, yaitu halaman muka (*Home*), Sekolah, Rekomendasi, Bantuan dan Admin. Menu *Home* digunakan untuk menampilkan informasi-informasi umum mengenai akademik sekolah ataupun informasi lain. Menu Sekolah digunakan untuk menampilkan data tentang sekolah-sekolah yang ada di Sragen. Sementara itu, menu Rekomendasi digunakan oleh user umum untuk menginputkan data kriteria untuk memperoleh rekomendasi dari sistem terkait sekolah-sekolah yang sesuai dengan data yang diinputkan berdasarkan probabilitas *naive bayes*. Tampilan halaman form input kriteria yang diajukan oleh calon siswa ditunjukkan pada gambar 7.

Gambar 7. Halaman Form Input Kriteria

Semua data kriteria yang diinputkan oleh user ke dalam form input rekomendasi tersebut kemudian diolah oleh sistem dengan menggunakan algoritma *naive bayes*. Hasil perhitungan metode ini kemudian ditampilkan pada halaman hasil rekomendasi seperti yang ditunjukkan pada gambar 8. Hasil ini berupa daftar sekolah yang ditampilkan dalam bentuk tabel yang diurutkan berdasarkan probabilitas tertinggi dari hasil perhitungan dan prioritas favorit masing-masing sekolah. Hasil yang ditampilkan meliputi nomor urut sekolah, nama sekolah, variabel berdasarkan kriteria yang terpenuhi sesuai yang diajukan oleh calon siswa, dan rangking favorit sekolah yang diambil berdasarkan urutan rangking dibandingkan dengan sekolah menengah atas yang lain dalam data pelatihan.

Hasil Rekomendasi Sekolah untuk Anda			
No	Sekolah	Variable	Rangking Favorit
1	SMK Negeri 1 Sragen	UAN KELULUSAN JARAK	5
2	SMK Muhammadiyah 4 Sragen	AKREDITASI SPP	7
3	SMA Muhammadiyah 1 Sragen	SPP KELULUSAN	8
4	SMA Negeri 3 Sragen	UAN JARAK	3
5	SMK Muhammadiyah 1 Sragen	AKREDITASI	10
6	SMA Negeri 1 Sragen	UAN	12
7	SMK Muhammadiyah 2 Sragen	KELULUSAN	9

Gambar 8. Hasil Rekomendasi Sekolah

Hasil yang ditunjukkan dalam gambar 8 memiliki makna bahwa SMK Negeri 1 Sragen merupakan sekolah dengan tingkat probabilitas yang tinggi dibandingkan dengan sekolah yang lain karena menempati urutan pertama. Hal ini diperoleh berdasarkan kriteria yang diajukan oleh calon siswa dapat dipenuhi sebanyak tiga variabel meliputi nilai UAN, tingkat kelulusan, serta jarak sekolah meskipun secara umum SMK Negeri 1 Sragen memiliki peringkat nomor 5 se-Sragen.

Berdasarkan hasil penelitian ini, sistem rekomendasi dapat digunakan oleh calon siswa untuk menentukan pilihan sekolah menengah atas sesuai dengan kriteria yang diajukan berdasarkan nilai probabilitas variabel yang diambil dari persyaratan pendaftaran siswa baru. Nilai probabilitas ini diperoleh dari perhitungan menggunakan algoritma naïve bayes berdasarkan nilai *confidence*-nya. Hal ini berbeda dengan sistem rekomendasi yang pernah dilakukan dalam penelitian lainnya [5][6][7]. Sistem rekomendasi ini selain ditujukan dalam bidang pendidikan yang jelas berbeda dengan ketiga penelitian tersebut, algoritma yang digunakan dalam penyelesaian permasalahan juga berbeda, yaitu menggunakan algoritma naïve bayes.

4. Kesimpulan

Kesimpulan yang dapat diambil dari hasil penelitian adalah sebagai berikut:

1. Telah berhasil dibangun sebuah *data warehouse* sebagai pusat data sekolah menengah atas dan sederajat di Sragen secara terintegrasi.
2. Sistem rekomendasi dengan menerapkan *Naive Bayes* berhasil dikembangkan untuk mempermudah calon siswa dalam menentukan sekolah yang sesuai dengan variabel-variabel yang ada.
3. Rekomendasi yang dihasilkan dengan memperhitungkan tingkat probabilitas tertinggi masing-masing atribut pada tiap sekolah, peringkat prioritas atribut, dan juga peringkat favorit dari sebuah sekolah.

5. Daftar Pustaka

- [1] Dinas Pendidikan dan Kebudayaan Sragen, "Number of Senior High Schools, Teachers and Students in Environmental Education Office in Sragen 2014/2015," 2015. [Online]. Available: <http://www.sragenkab.go.id/home.php?menu=50>. [Accessed: 17-Dec-2015].
- [2] R. Jindal and S. Taneja, "Comparative Study of Data Warehouse Design Approaches: a Survey," *Int. J. Database Manag. Syst.*, vol. 4, no. 1, pp. 33–45, 2012.
- [3] Y. S. Nugroho and Setyawan, "Klasifikasi Masa Studi Mahasiswa Fakultas Komunikasi dan Informatika," *KomuniTi*, vol. VI, no. 1, pp. 84–91, 2014.
- [4] D. B. Rahayu and P. Erwin Gunadhi, "Perancangan Sistem Informasi Pendaftaran Peserta Didik Baru Berbasis Web Studi Kasus Di SMA Negeri 14 Garut," *J. Algoritm.*, vol. 9, no. 27, pp. 112–120, 2012.
- [5] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-Finder: A Recommender System for Finding Passengers and Vacant Taxis," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 10, pp. 2390–2403, Oct. 2013.
- [6] A. Levi, O. Mokryn, C. Diot, and N. Taft, "Finding a needle in a haystack of reviews," in *Proceedings of the sixth ACM conference on Recommender systems - RecSys '12*, 2012, p. 115.
- [7] M. Beladev, L. Rokach, and B. Shapira, "Recommender systems for product bundling," *Knowledge-Based Syst.*, vol. 111, pp. 193–206, 2016.
- [8] R. Burke and Robin, "Hybrid Recommender Systems: Survey and Experiments," *User Model. User-adapt. Interact.*, vol. 12, no. 4, pp. 331–370, 2002.
- [9] P. Ponniah, *Data Warehousing Fundamentals for IT Professionals, Second Edition*, Second. New York, United States: John Wiley & Sons, Inc., 2010.
- [10] Y.-W. Chang, C.-J. Hsieh, K.-W. Chang, M. Ringgaard, and C.-J. Lin, "Training and Testing Low-degree Polynomial Data Mappings via Linear SVM," *J. Mach. Learn. Res.*, vol. 11, pp. 1471–1490, 2010.
- [11] H. Zhang and S. Shengli Sheng, "Learning Weighted Naive Bayes with Accurate Ranking," in *Fourth IEEE International Conference on Data Mining (ICDM'04)*, 2004, pp. 567–570.
- [12] Y. S. Nugroho and S. N. Haryati, "Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali," *Khazanah Inform.*, vol. 1, no. 1, 2015.
- [13] R. Katarya and O. P. Verma, "An effective collaborative movie recommender system with cuckoo search," *Egypt. Informatics J.*, 2016.