

# Sistem Rekomendasi Pemilihan Sekolah Menengah Tingkat Atas Menggunakan Metode Naive Bayes

Triana Dewi Salma<sup>1\*</sup>, Yusuf Sulistyio Nugroho<sup>1</sup>

<sup>1</sup>Program Studi Informatika, Fakultas Komunikasi dan Informatika  
Universitas Muhammadiyah Surakarta

\*tdsalma@gmail.com, yusuf.nugroho@ums.ac.id

## ABSTRAK

Setiap tahun pendidikan di Indonesia semakin maju dan berkembang seiring dengan semakin bertambahnya jumlah sekolah atau lembaga pendidikan. Sekolah-sekolah ini tidak jarang didirikan dalam jarak yang berdekatan dalam satu wilayah seperti halnya dengan sekolah menengah atas dan sederajat di kota Sragen. Namun, permasalahan yang sering dihadapi yaitu banyak faktor yang akhirnya melatarbelakangi siswa untuk memilih suatu calon sekolah baik faktor internal maupun eksternal sekolah. Oleh karena itu, siswa dituntut cerdas dalam menentukan calon sekolah terbaik dan sesuai dengan kriterianya. Penelitian ini dilakukan dalam rangka membantu siswa untuk memilih calon sekolah menengah tingkat atas terbaik sesuai dengan kriteria-kriteria yang diinginkan. Penerapan metode *Naive Bayes* dalam aplikasi berbasis web berbasis PHP, HTML dan *JavaScript* digunakan untuk merekomendasikan calon sekolah terbaik bagi siswa dengan memanfaatkan data sekolah menengah atas di Kota Sragen. Metode *Naive Bayes* diterapkan untuk mencari probabilitas terbesar dari semua *instance* pada atribut target seperti jarak, biaya SPP, akreditasi, tingkat kelulusan, dan nilai rata-rata minimal UAN. Hasil dari penelitian ini yaitu sebuah sistem rekomendasi yang dapat membantu menyajikan informasi sekolah-sekolah menengah atas yang sesuai dengan kriteria dengan mempertimbangkan nilai confidence terbesar, pemenuhan variabel, prioritas variabel dan rangking favorit sekolah. Hal tersebut dapat menjadi pertimbangan bagi siswa untuk memilih calon sekolah menengah atas yang terbaik sesuai dengan kriteria yang diajukan.

**Kata kunci :** *data mining, Naive Bayes, pemilihan sekolah, sekolah menengah tingkat atas, sistem rekomendasi.*

## 1. Pendahuluan

Setiap tahun pendidikan di Indonesia semakin maju dan berkembang seiring dengan semakin berkembangnya

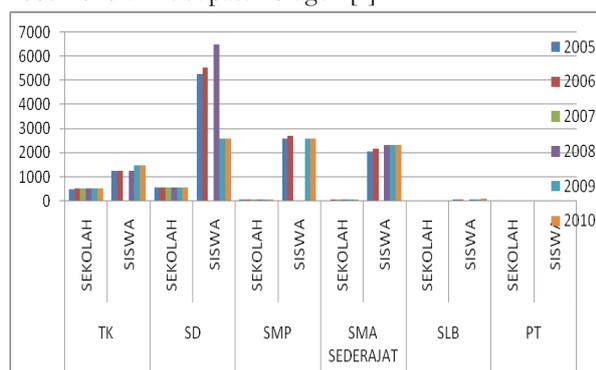
sekolah atau lembaga pendidikan. Tabel 1 menunjukkan perkembangan sekolah atau lembaga pendidikan di Indonesia menurut data dari Kemendikbud dalam ikhtisar data pendidikan tahun 2011/2012 [1].

Tabel 1. Data Perkembangan Sekolah Menurut Jenjang Pendidikan Tahun 2007/2008-2011/2012

No	Jenjang Pendidikan	Tahun				
		2007/2008	2008/2009	2009/2010	2010/2011	2011/2012
1	TK	63,444	63,624	67,55	69,326	70,917
	a. Negeri	692	804	1,616	1,804	2,083
	b. Swasta	62,752	62,82	65,934	67,522	68,834
2	SLB	1,455	1,686	1,803	1,783	1,924
	a. Negeri	343	412	437	330	496
	b. Swasta	1,112	1,274	1,366	1,453	1,428
3	SD	144,567	144,228	143,252	146,804	146,826
	a. Negeri	132,513	131,490	130,563	133,406	133,597
	b. Swasta	12,054	12,738	12,689	13,398	13,229
4	SMP	26,277	28,777	29,866	30,290	33,668
	a. Negeri	15,024	16,898	17,714	17,930	20,594
	b. Swasta	11,253	11,879	12,152	12,360	13,074

No	Jenjang Pendidikan	Tahun				
		2007/2008	2008/2009	2009/2010	2010/2011	2011/2012
5	SMA	10,239	10,762	11,036	11,306	11,654
	a. Negeri	4,493	4,797	5,034	5,322	5,57
	b. Swasta	5,746	5,965	6,002	5,984	6,084
6	SMK	6,746	7,592	8,399	9,164	10,256
	a. Negeri	1,748	2,003	2,218	2,459	2,697
	b. Swasta	4,998	5,589	6,181	6,705	7,559
7	PT	2,68	2,975	3,011	3,185	3,17
	a. Negeri	82	83	83	88,000	92
	b. Swasta	2,598	2,892	2,928	3,097	3,078

Sementara itu pemerintah telah mencanangkan untuk penerapan wajib sekolah 9 tahun. Persaingan dunia kerja juga menuntut masyarakat untuk memiliki pendidikan yang tinggi. Seiring banyaknya masyarakat yang mulai menyadari arti pentingnya pendidikan, maka banyak sekolah baru berdiri dan bersaing dalam memberikan mutu pendidikan yang baik. Sedangkan data jumlah sekolah bertambah banyak diimbangi dengan jumlah bertambahnya siswa tiap tahunnya. Gambar 1 menunjukkan perkembangan sekolah atau lembaga pendidikan dan perkembangan jumlah siswa dari tahun 2005-2010 di Kabupaten Sragen [2].



Gambar 1. Data Perkembangan Sekolah Menurut Jenjang Pendidikan dan siswa Tahun 2005/2006 - 2010/2011 di Kabupaten Sragen.

Sekolah-sekolah ini tidak jarang didirikan dalam jarak yang berdekatan seperti halnya Sekolah Menengah Atas dan Sederajat di Kota Sragen. Mutu sekolah memang menjadi pertimbangan dalam pemilihan sekolah. Namun banyak faktor yang melatarbelakangi para siswa untuk memilih suatu sekolah. Dewasa ini, sekolah-sekolah tidak hanya bersaing dalam hal akademik melalui prestasi sekolah maupun akreditasi namun juga bersaing dalam mutu pendidikan yang baik untuk menarik minat calon siswa. Tidak hanya itu, banyak faktor yang akhirnya melatarbelakangi siswa untuk memilih sebuah calon sekolah baik faktor eksternal sekolah maupun internal dari siswanya sendiri.

Oleh karena itu, dibutuhkan suatu sistem yang membantu siswa untuk merekomendasikan calon sekolah terbaik dari berbagai faktor internal maupun eksternal. Sistem tersebut diharapkan dapat membantu siswa dalam

memilih calon sekolah dan merekomendasikan sekolah-sekolah yang sesuai dengan kriteria yang diinginkan. Metode yang sering digunakan untuk memprediksikan sesuatu dalam data mining sangatlah banyak. Dalam hal ini, *Naive Bayes* digunakan untuk merekomendasikan calon sekolah terbaik bagi siswa dengan memanfaatkan data Sekolah Menengah Atas dan sederajat di Kota Sragen.

Penelitian mengenai metode *Naive Bayes* telah banyak digunakan dalam prediksi berbagai hal. Penelitian mengenai sistem rekomendasi pemilihan sekolah juga banyak dilakukan. Purnamasari dalam [3] menggunakan metode *Naive Bayes*. Latar belakang penelitian ini adalah Yogyakarta sebagai kota pelajar dimanfaatkan sebagai lahan bisnis terutama tempat kos, maka sistem ini berguna untuk informasi mahasiswa dalam mencari kos. Penelitian Purnamasari ini menghasilkan penerapan metode *Naive Bayes* pada kasus tersebut yaitu rekomendasi tempat kos berdasarkan penghitungan probabilitas yang dipengaruhi masukan data awal.

Uyun dan Madikhaturon meneliti model rekomendasi berbasis *fuzzy* untuk pemilihan sekolah lanjutan tingkat atas [4]. Model rekomendasi yang digunakan adalah *fuzzy model* tahani. Secara umum, logika *fuzzy* dapat menangani faktor ketidakpastian dengan baik sehingga dapat diimplementasikan dalam rekomendasi pemilihan sekolah. Penelitian ini mengambil data sekolah lanjutan tingkat atas di kota Yogyakarta. Secara keseluruhan, sistem yang dibangun mampu memberikan rekomendasi kepada calon siswa dalam mengambil keputusan memilih sekolah.

Fuadillah menjelaskan penelitiannya mengenai sistem aplikasi penunjang keputusan pemilihan sekolah menengah atas dan sederajat menggunakan metode *Fuzzy Multiple Criteria Decision Making* (FMCDM) [5]. Penelitian ini bertujuan membantu calon siswa-siswi SMA dan sederajat mendapatkan informasi sekolah-sekolah yang ada di Banjarbaru dan membantu mereka dalam penentuan keputusan dalam pemilihan sekolah. Sampel penelitian ini adalah SMA dan sederajat yang ada di Banjarbaru yang berjumlah 30 Sekolah. Kesimpulan penelitian bahwa siswa lebih mudah mendapatkan informasi mengenai sekolah menengah atas yang ada di Banjarbaru dan siswa

juga lebih mudah dalam menentukan pemilihan sekolah menengah atas dengan menggunakan aplikasi sistem.

Nugroho, dkk mengangkat masalah seleksi sekolah menengah lanjutan sebagai penelitiannya [6]. Metode yang digunakan yaitu AHP (*Analytical Hierarchy Process*) dengan tujuan penelitian membuat aplikasi sistem pendukung keputusan untuk membantu orang tua dalam memilih sekolah yang sesuai bagi anak. Hasil dari penelitian ini yaitu sistem pendukung keputusan untuk membantu orang tua calon siswa menentukan sekolah sesuai dengan kriteria yang diinginkan berdasarkan nilai bobot kemungkinan diterima atau tidaknya bagi anak di sekolah tersebut.

Perbedaan mendasar dari penelitian sebelumnya yaitu tentang pembuatan rekomendasi pemilihan sekolah menengah atas dan sederajat menggunakan metode *Naive Bayes*. HMAP untuk mencari probabilitas terbesar dari semua instance pada atribut target atau semua kemungkinan keputusan. Sampel yang digunakan yaitu Sekolah Menengah Atas dan sederajat di Kota Sragen. Tujuan penelitian ini adalah untuk menghasilkan sebuah sistem berbasis web sebagai rekomendasi siswa dalam memilih calon sekolah menengah tingkat atas sesuai dengan kriteria-kriteria yang diinginkan menggunakan metode data mining yaitu *Naive Bayes*. Dengan demikian, dapat membantu menyajikan informasi sekolah-sekolah menengah atas yang sesuai dengan kriteria dan dapat menjadi pertimbangan bagi siswa untuk memilih calon sekolah menengah atas yang terbaik sesuai dengan kriteria yang diinginkan.

## 2. Metode

### A. Peralatan Pendukung

Peralatan yang digunakan selama penelitian terdiri dari *software* dan *hardware*. *Software* yang digunakan antara lain Sublime Text 2, XAMPP, Google Chrome, DB Designer 4, dan Rapid Miner 5. Sedangkan *hardware* yang digunakan adalah laptop dengan spesifikasi prosesor Intel Core i3, RAM 2 GB, *Harddisk* 500 GB dan sistem operasi windows 7. Pembuatan aplikasi berbasis web ini memanfaatkan bahasa pemrograman PHP (*Hypertext Processor*), HTML (*Hyper Text Markup Language*), dan *JavaScript*.

### B. Pengumpulan Data

Data yang digunakan dalam penelitian yaitu data dari Sekolah Menengah Atas dan sederajat di Kota Sragen sebagai data training. Data tersebut diperoleh dari Dinas Pendidikan dan Kebudayaan Kota Sragen. Selain itu, juga menggunakan data testing yang didapatkan dari sampel siswa.

#### 1) Data Training

*Data training* merupakan data yang digunakan untuk pengenalan pola. *Data training* yang digunakan adalah data Sekolah Menengah Atas dan sederajat di Kota Sragen. Data yang diambil dalam sampel

adalah 10 Sekolah Menengah Atas dan sederajat di Kota Sragen dari tahun ajaran 2010/2011 sampai 2014/2015. Atribut yang digunakan meliputi nama sekolah, biaya SPP, jarak sekolah yang diukur dari Kantor Bupati Kab Sragen (Jl. Raya Sukowati no 255, Sragen, Jawa tengah), akreditasi sekolah, tingkat kelulusan, dan nilai rata-rata minimal UAN.

#### 2) Data Testing

*Data testing* merupakan data yang sedang atau akan terjadi dan digunakan untuk menguji pola yang telah didapatkan dari data training. *Data testing* yang digunakan adalah data siswa lulusan SMP dan sederajat maupun yang masih aktif di kelas IX dan ingin melanjutkan ke Sekolah Menengah Atas dan sederajat. Data Siswa ini diambil sampel secara acak. Atribut yang digunakan yaitu biaya SPP, jarak sekolah yang diukur dari Kantor Bupati Kab Sragen (Jl. Raya Sukowati no 255, Sragen, Jawa tengah), akreditasi sekolah, tingkat kelulusan, dan nilai rata-rata minimal UAN.

### C. Kebutuhan Data

Penentuan kebutuhan merupakan suatu hal yang diperlukan dalam membantu pembangunan data mining ini. Dari atribut data sekolah dan data siswa dibagi dalam bermacam-macam variabel, yaitu :

- 1) Variabel X1 biaya SPP.
- 2) Variabel X2 jarak sekolah yang diukur dari pusat kota.
- 3) Variabel X3 prestasi Sekolah yang didasarkan pada akreditasi sekolah.
- 4) Variabel X4 tingkat lulusan yang diperoleh dari perhitungan menggunakan persamaan (1)

$$\text{Tingkat Kelulusan} = \frac{\sum \text{siswa yang lulus}}{\sum \text{siswa kelas III}} \times 100\% \quad (1)$$

### D. Data Cleaning

Al-Janabi dan Janicki menjelaskan bahwa *data cleaning* atau pembersihan data merupakan proses yang paling penting dalam pengolahan data agar data benar – benar relevan dengan kebutuhan, konsisten, dan tidak mengalami redudansi [7]. Dengan demikian, dari atribut data sekolah dan data siswa tidak terjadi *noise* atau ketidakkonsistenan data dalam pengujian sistem. Pembersihan data dilakukan dengan menyelaraskan seluruh *class* dari setiap variabelnya.

### E. Penggunaan Algoritma *Naive Bayes*

*Naive Bayes Classifier* (NBC) menurut Prasetyo merupakan teknik prediksi berbasis probabilistik sederhana yang merujuk pada penerapan teorema Bayes dengan asumsi independensi yang kuat. Metode *Bayes* menggunakan probabilitas bersyarat sebagai dasarnya [8]. Yang, dkk. dalam [9] mengemukakan bahwa algoritma *Naive Bayes* didasarkan pada fungsi probabilitas untuk

setiap *instance* dan menjelaskan pemetaan antara atribut dan klasifikasi atribut dalam sistem yang memiliki efisiensi klasifikasi stabil dan kompleksitas rendah. *Naive Bayes Classifier* mengasumsikan bahwa nilai atribut saling independen. Metode HMAP atau *Naive Bayes* menurut [5] ditunjukkan seperti persamaan (2).

$$P(S|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2)$$

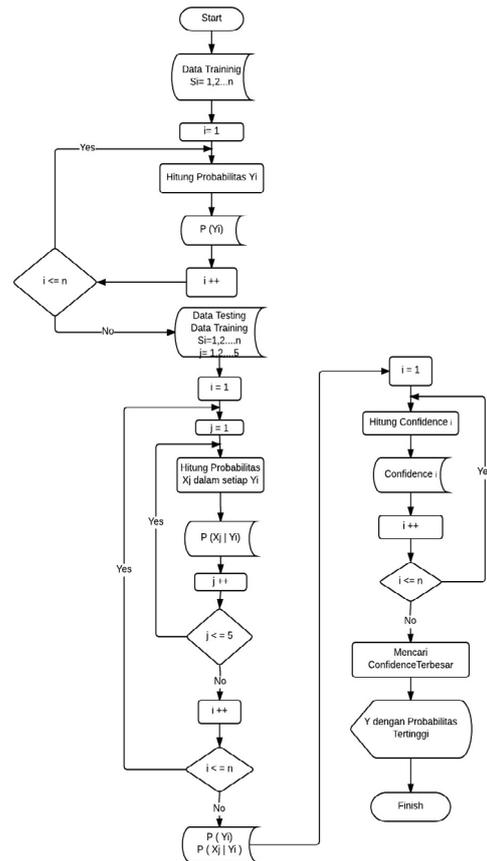
Dimana algoritma HMAP (*Hypothesis Maximum Apriori Probability*) yang menyatakan hipotesa yang diambil berdasarkan nilai probabilitas berdasarkan kondisi prior yang diketahui. HMAP inilah yang digunakan di dalam *machine learning* sebagai metode untuk mendapatkan hipotesis untuk suatu keputusan.

Penerapan algoritma *Naive bayes* dalam penelitian ini yaitu untuk mencari probabilitas variabel pada setiap kemungkinan sekolah. *Data training* yang digunakan sebanyak 10 sekolah, oleh karena itu perulangan pencarian *confidence* sebanyak jumlah data tersebut. Pada tabel 2 merupakan algoritma *Naive bayes* yang diterapkan untuk mencari rekomendasi sekolah.

Tabel 2. Algoritma Naive Bayes

Algoritma Naive Bayes	
Data Training	$S_i = (X_1, X_2, X_3, X_4, X_5, Y)$ , $i = 1, 2, \dots, n$ ;
For	$i = 1$ sampai $n$
	Mencari Probabilitas Setiap $Y_i$
	End For
	Mendapatkan Fakta $P(Y)$ untuk setiap $Y_i$
Data Testing	$(X_1, X_2, X_3, X_4, X_5)$
For	$i = 1$ sampai $n$
	Mencari probabilitas $X_1, X_2, X_3, X_4, X_5$ dalam setiap $Y_i$
	End For
	Mendapatkan $P(X Y)$ untuk setiap $X$ dalam $Y_i$
Confidence	$Confidence_1, Confidence_2, \dots, Confidence_n$
For	$i = 1$ sampai $n$
	Menghitung Confidence setiap atribut target = $P(Y_i) \cdot P(X_{1,2,\dots,5}   Y_i)$
	End for
	Membandingkan dan mencari nilai confidence yang tertinggi dari setiap atribut target

Nilai *confidence* tertinggi diperoleh dengan mencari probabilitas untuk setiap  $y$  yang ada dalam *data training*. Kemudian mencari *probabilitas* variabel  $X_1, X_2, X_3, X_4, X_5$  dari data testing untuk setiap  $y$ . Nilai *probabilitas*  $y$  dan  $x|y$  digunakan untuk menghitung *confidence* pada setiap atribut target. Gambar 2 merupakan *flowchart* yang menjelaskan langkah-langkah algoritma *Naive Bayes* untuk menghasilkan nilai *confidence* tertinggi pada atribut target sebagai pertimbangan sistem dalam memberikan rekomendasi sekolah bagi siswa sesuai dengan kriteria yang diajukannya.



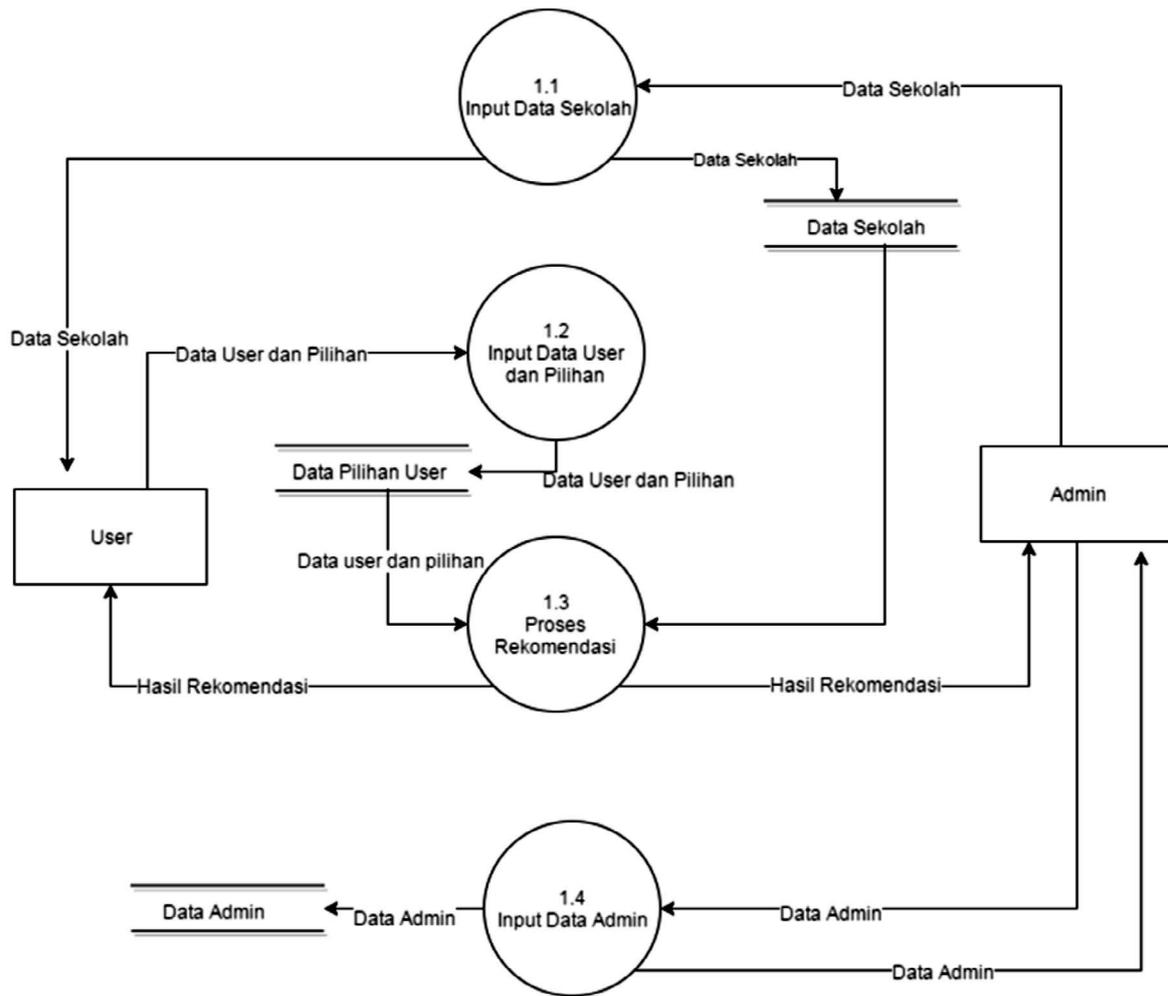
Gambar 2. Flowchart Naive Bayes dalam system

Pada Gambar 3 merupakan DFD level 0 atau diagram alir sistem secara umum. Sedangkan pada gambar 4 merupakan DFD level 1 yang menjelaskan alir data pada setiap proses. Pada gambar 5 merupakan DFD Level 2 yang menjelaskan alir data lebih terperinci pada proses rekomendasi sebagai inti sistem.

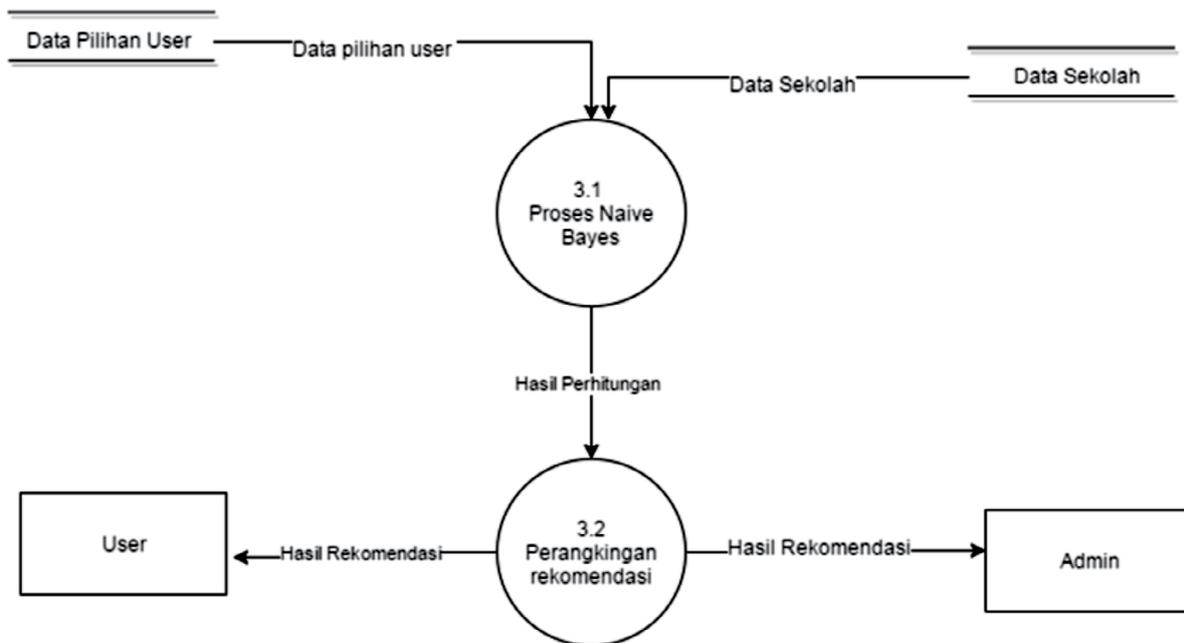


DFD Level 0

Gambar 3. DFD Level 0



Gambar 4. DFD Level 1



Gambar 5. DFD Level

### 3. Hasil dan diskusi

#### A. Hasil

Pengumpulan data selama penelitian menghasilkan dua data yaitu *data training* dan *data testing* dengan atribut diantaranya biaya SPP, jarak sekolah yang diukur dari Kantor Bupati Kab Sragen (Jl. Raya Sukowati no 255,

Sragen, Jawa tengah), akreditasi sekolah, tingkat kelulusan siswa, dan nilai rata-rata minimal UAN sebagai data pendukung dalam menghasilkan rekomendasi tersebut. *Data training* pada tabel 3 merupakan data yang digunakan dalam penelitian ini dan *data testing* pada tabel 4 sebagai data uji.

Tabel 3. Data training

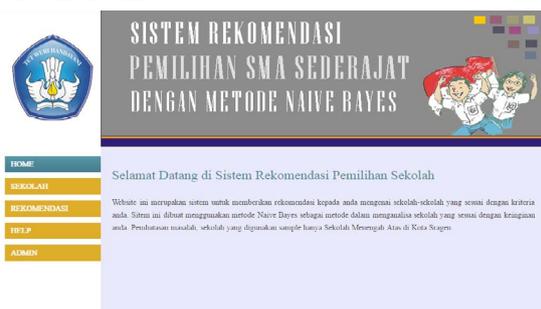
Nama Sekolah	Biaya Spp (Rupiah)	Jarak (Km)	Akreditasi	Tingkat Kelulusan (%)	Nilai UAN
SMA Negeri 1 Sragen	200.000	2,1	A	100	32
SMA Negeri 2 Sragen	150.000	2,1	A	100	30,25
SMA Negeri 3 Sragen	150.000	1,9	A	89	32,85
MAN 1 Sragen	85.000	2,4	A	100	16,9
SMK Negeri 1 Sragen	85.000	1,5	A	95	34,85
SMK Negeri 2 Sragen	150.000	2,2	A	99	16,52
SMA Muhammadiyah 1 Sragen	250.000	4,2	A	95	16,02
SMK Muhammadiyah 1 Sragen	150.000	2,4	B	99	13,8
SMK Muhammadiyah 2 Sragen	160.000	2,4	A	96	17,4
SMK Muhammadiyah 4 Sragen	270.000	2,2	B	89	12,15

Tabel 4. Data Testing

No	Biaya_SPP	Jarak	Akreditasi	Tingkat Kelulusan	Nilai Ujian	Nama Sekolah
1	200.001-300.000	1-2	B	90 - 98%	31-40	?
2	0 - 100.000	3,1-4	A	80-89 %	21-30	?
3	100.001-200.000	2,1 -3	C	99-100%	21-30	?

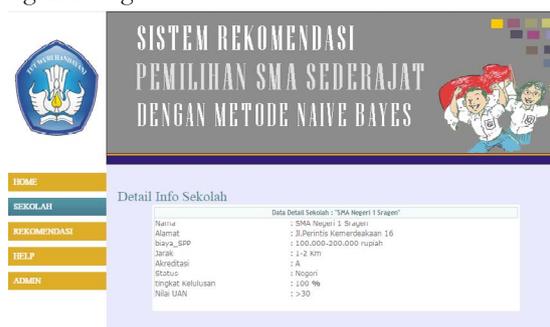
Perhitungan menggunakan algoritma *Naive Bayes* dan menghasilkan nilai *confidence* pada setiap variabel Y untuk setiap atribut. Sekolah yang nantinya direkomendasikan akan ditampilkan diurutkan berdasarkan nilai *confidence* terbesar. Selain itu, perekomendasi juga memperhatikan pemenuhan variabel dan prioritas variabel yang dapat di-custom-isasi oleh user serta ranking favorit sekolah.

Tampilan Program halaman utama merupakan tampilan awal ketika user membuka sistem berupa penjelasan singkat mengenai sistem. Terdapat beberapa menu yang tersedia untuk user diantaranya menu Home, Sekolah, Rekomendasi dan Help. Sedangkan menu login adalah menu dimana admin dapat login dengan memasukkan username dan password yang telah terdaftar. Gambar 6 menunjukkan tampilan halaman awal dari sistem.



Gambar 6. Tampilan Halaman Awal

Pada menu sekolah terdapat informasi mengenai sekolah-sekolah yang dikelompokkan menjadi sekolah berstatus negeri dan swasta. User dapat memilih informasi dari sekolah mana yang ingin dilihat. Gambar 7 menunjukkan tampilan informasi detail sekolah SMA Negeri 1 Sragen.



Gambar 7. Tampilan Informasi detail sekolah

Menu Rekomendasi merupakan menu dimana user harus memasukkan data dan menjawab pertanyaan-pertanyaan sebagai input dalam analisis *Naive Bayes*. Gambar 8 menunjukkan pengisian data dan kriteria sekolah yang diinginkan user. User harus memasukkan data pribadi yang akan tersimpan dalam database serta memilih kriteria dalam setiap pertanyaan. Selain itu, juga dapat mengatur limit atau batasan data sekolah yang akan ditampilkan sebagai rekomendasi dan status sekolahnya.

Untuk prioritas kriteria, user dapat mengatur kriteria yang menjadi pertimbangan pertama dalam memilih sekolah hingga prioritas terakhir.

Gambar 8. Tampilan Form pengisian data user dan kriteria sekolah

Berdasarkan data testing pada tabel 4, terdapat tiga data testing. Gambar 9 merupakan hasil pengujian sistem dengan inputan data testing pertama pada tabel 4 dan menampilkan sekolah yang direkomendasikan berdasarkan nilai *confidence* nya, pemenuhan variabel, prioritas variabel dan ranking favorit sekolah sebagai outputnya.

No	Sekolah	Variable	Ranking Favorit
1	SMK Negeri 1 Sragen	UAN KELULUSAN JARAK	5
2	SMA Muhammadiyah 1 Sragen	SPP KELULUSAN	8
3	SMA Negeri 3 Sragen	UAN JARAK	3
4	SMK Muhammadiyah 4 Sragen	SPP	7
5	SMA Negeri 1 Sragen	UAN	12
6	SMK Muhammadiyah 2 Sragen	KELULUSAN	9

Gambar 9. Tampilan Hasil rekomendasi Sekolah berdasarkan data testing pertama

Gambar 10 merupakan hasil pengujian sistem dengan inputan data testing kedua pada tabel 4 dan menampilkan sekolah yang direkomendasikan berdasarkan nilai *confidence* nya, pemenuhan variabel, prioritas variabel dan ranking favorit sekolah sebagai outputnya.

No	Sekolah	Variable	Ranking Favorit
1	SMK Negeri 1 Sragen	SPP	5
2	MAN 1 Sragen	SPP	6
3	SMA Negeri 3 Sragen	KELULUSAN	3
4	SMK Muhammadiyah 4 Sragen	KELULUSAN	7

Gambar 10. Tampilan Hasil rekomendasi Sekolah berdasarkan data testing kedua

Gambar 11 merupakan hasil pengujian sistem dengan inputan data testing data ketiga pada tabel 4 dan menampilkan sekolah yang direkomendasikan berdasarkan berdasarkan nilai *confidence* nya, pemenuhan variabel, prioritas variabel dan ranking favorit sekolah sebagai outputnya.

No	Sekolah	Variable	Ranking Favorit
1	SMK Negeri 2 Sragen	SPP KELULUSAN JARAK	4
2	SMK Muhammadiyah 1 Sragen	SPP KELULUSAN JARAK	10
3	SMA Negeri 2 Sragen	SPP KELULUSAN	2
4	SMA Negeri 1 Sragen	SPP KELULUSAN	12
5	SMK Muhammadiyah 2 Sragen	SPP JARAK	9
6	MAN 1 Sragen	KELULUSAN JARAK	6
7	SMA Negeri 3 Sragen	SPP	3
8	SMK Muhammadiyah 4 Sragen	JARAK	7

Gambar 11. Tampilan Hasil rekomendasi Sekolah berdasarkan data testing ketiga

Admin bisa mengatur *range* setiap variabel yang berpengaruh pada setiap pilihan siswa. *Range* setiap variabel tersebut yang akan menjadi *range* untuk data yang dikelompokkan. Pada gambar 12 merupakan tampilan pada menu rekomendasi yang menjadi menu untuk admin dapat merubah *range* variabel.

Gambar 12. Menu rekomendasi dimana admin bisa mengedit range variabel

## B. Diskusi

Data training yang berupa data sekolah diambil dari Dinas Pendidikan dan kebudayaan Kota Sragen. Namun, pada nyatanya data yang terpusat pada dinas tidak lengkap dan terdapat data yang memang tidak tercatat seperti

data biaya SPP dan Nilai rata-rata UAN. Oleh karena itu, untuk mengatasi keterbatasan data dari Dinas, data training harus diambil dari masing-masing sekolah secara langsung. Jadi, Data training dapat dikumpulkan secara maksimal dan tidak mengganggu dalam proses analisis lebih lanjut.

Berikut perhitungan manual algoritma Naive Bayes dengan menggunakan data sekolah sebagai data training dan data siswa sebagai data testing yang dipilih sebagai sampel.

Perhitungan dilakukan pada *data testing* dengan mengambil data pertama dari tabel 4 yaitu biaya SPP 200.001 - 300.000, memiliki jarak 1-2 Km dari Kantor Bupati Kab Sragen (Jl. Raya Sukowati no 255, Sragen, Jawa tengah), akreditasi sekolah B, tingkat kelulusan 90-98% dan nilai rata-rata minimal UAN 31-40.

Perhitungan *data training* diasumsikan sebagai berikut :

- Y = Nama Sekolah
- X<sub>1</sub> = Biaya SPP
- X<sub>2</sub> = Jarak Sekolah yang diukur dari Kantor Bupati Kab Sragen (Jl. Raya Sukowati no 255, Sragen, Jawa tengah)

- X<sub>3</sub> = Akreditasi Sekolah
- X<sub>4</sub> = Tingkat Kelulusan
- X<sub>5</sub> = Nilai rata-rata minimal UAN

Fakta menunjukkan :

- P (Y = SMA N 1 Sragen) = 1/10 = 0,1
- P (Y = SMA N 2 Sragen) = 1/10 = 0,1
- P (Y = SMA N 3 Sragen) = 1/10 = 0,1
- P (Y = MAN 1 Sragen) = 1/10 = 0,1
- P (Y = SMK N 1 Sragen) = 1/10 = 0,1
- P (Y = SMK N 2 Sragen) = 1/10 = 0,1
- P (Y = SMA Muhammadiyah 1 Sragen) = 1/10 = 0,1
- P (Y = SMK Muhammadiyah 1 Sragen) = 1/10 = 0,1
- P (Y = SMK Muhammadiyah 2 Sragen) = 1/10 = 0,1
- P (Y = SMK Muhammadiyah 4 Sragen) = 1/10 = 0,1

Perhitungan setiap variabel Y terhadap setiap variabel X dapat dihitung seperti pada tabel 5. Menggunakan Naive bayes untuk mencari probabilitas variabel Y yang ada untuk setiap variabel X.

Tabel 5. Perhitungan Variabel X untuk setiap Variabel Y

Variabel X \ Variabel Y	X <sub>1</sub> Biaya SPP (200.001-300.000)	X <sub>2</sub> Jarak ( 1- 2 Km)	X <sub>3</sub> Akreditasi ( A )	X <sub>4</sub> Tingkat Kelulusan ( 90 – 98 )	X <sub>5</sub> Nilai rata-rata minimal UAN (31-40)
Y <sub>1</sub> = SMAN 1 Sragen	0/1 = 0	0/1 = 0	0/1 = 0	0/1 = 0	1/1 = 1
Y <sub>2</sub> = SMAN 2 Sragen	0/1 = 0	0/1 = 0	0/1 = 0	0/1 = 0	0/1 = 0
Y <sub>3</sub> = SMAN 3 Sragen	0/1 = 0	1/1 = 1	0/1 = 0	0/1 = 0	1/1 = 1
Y <sub>4</sub> = MAN 1 Sragen	0/1 = 0	0/1 = 0	0/1 = 0	0/1 = 0	0/1 = 0
Y <sub>5</sub> = SMKN 1 Sragen	0/1 = 0	1/1 = 1	0/1 = 0	1/1 = 1	1/1 = 1
Y <sub>6</sub> = SMKN 2 Sragen	0/1 = 0	0/1 = 0	0/1 = 0	0/1 = 0	0/1 = 0
Y <sub>7</sub> = SMA Muhammadiyah 1 Sragen	1/1 = 1	0/1 = 0	0/1 = 0	1/1 = 1	0/1 = 0
Y <sub>8</sub> = SMK Muhammadiyah 1 Sragen	0/1 = 0	0/1 = 0	1/1 = 1	0/1 = 0	0/1 = 0
Y <sub>9</sub> = SMK Muhammadiyah 2 Sragen	0/1 = 0	0/1 = 0	0/1 = 0	1/1 = 1	0/1 = 0
Y <sub>10</sub> = SMK Muhammadiyah 4 Sragen	1/1 = 1	0/1 = 0	1/1 = 1	0/1 = 0	0/1 = 0

Menentukan HMAP dari data tersebut dapat dihitung pada persamaan (3) sampai dengan persamaan (12) berikut.

- P (X<sub>1</sub> = 200.001 – 300.000, X<sub>2</sub> = 1-2 Km, X<sub>3</sub> = B, X<sub>4</sub> = 90-98 %, X<sub>5</sub> = 31-40 | Y = SMA N 1 Sragen) = 0,1. 0. 0. 0. 0. 1 = 0 (3)
- P (X<sub>1</sub> = 200.001 – 300.000, X<sub>2</sub> = 1-2 Km, X<sub>3</sub> = B, X<sub>4</sub> = 90-98 %, X<sub>5</sub> = 31-40 | Y = SMA N 2 Sragen) = 0,1. 0. 0. 0. 0. 0 = 0 (4)
- P (X<sub>1</sub> = 200.001 – 300.000, X<sub>2</sub> = 1-2 Km, X<sub>3</sub> = B, X<sub>4</sub> = 90-98 %, X<sub>5</sub> = 31-40 | Y = SMA N 3 Sragen) = 0,1. 0. 1. 0. 0. 1 = 0 (5)
- P (X<sub>1</sub> = 200.001 – 300.000, X<sub>2</sub> = 1-2 Km, X<sub>3</sub> = B, X<sub>4</sub> = 90-98 %, X<sub>5</sub> = 31-40 | Y = MAN 1 Sragen) = 0,1. 0. 0. 0. 0. 0 = 0 (6)

- P (X<sub>1</sub> = 200.001 – 300.000, X<sub>2</sub> = 1-2 Km, X<sub>3</sub> = B, X<sub>4</sub> = 90-98 %, X<sub>5</sub> = 31-40 | Y = SMK N 1 Sragen) = 0,1. 0. 1. 0. 1. 1 = 0 (7)
- P (X<sub>1</sub> = 200.001 – 300.000, X<sub>2</sub> = 1-2 Km, X<sub>3</sub> = B, X<sub>4</sub> = 90-98 %, X<sub>5</sub> = 31-40 | Y = SMK N 2 Sragen) = 0,1. 0. 0. 0. 0. 0 = 0 (8)
- P (X<sub>1</sub> = 200.001– 300.000, X<sub>2</sub> = 1-2 Km, X<sub>3</sub> = B, X<sub>4</sub> = 90-98 %, X<sub>5</sub> = 31-40 | Y = SMA Muhammadiyah 1 Sragen) = 0,1. 1. 0. 0. 1. 0 = 0 (9)
- P (X<sub>1</sub> = 200.001 – 300.000, X<sub>2</sub> = 1-2 Km, X<sub>3</sub> = B, X<sub>4</sub> = 90-98 %, X<sub>5</sub> = 31-40 | Y = SMK Muhammadiyah 1 Sragen) = 0,1. 0. 0. 1. 0. 0 = 0 (10)
- P (X<sub>1</sub> = 200.001 – 300.000, X<sub>2</sub> = 1-2 Km, X<sub>3</sub> = B, X<sub>4</sub> = 90-98 %, X<sub>5</sub> = 31-40 | Y = SMK Muhammadiyah 2 Sragen) = 0,1. 0. 0. 0. 1. 0 = 0 (11)

$P(X_1 = 200.001 - 300.000, X_2 = 1-2 \text{ Km}, X_3 = B, X_4 = 90-98 \%, X_5 = 31-40 | Y = \text{SMK Muhammadiyah 4 Sragen}) = 0,1. 0. 1. 0. 0. = 0$  (12)

Memperhatikan pemenuhan variabel dengan memperhatikan berapa banyak variabel yang terpenuhi pada setiap Y. Hal tersebut dapat dituangkan seperti pada tabel 6

Tabel 6. Pemenuhan Variabel

No	Y	X <sub>1</sub> SPP	X <sub>2</sub> Jarak	X <sub>3</sub> Akreditasi	X <sub>4</sub> Kelulusan	X <sub>5</sub> Nilai UAN
1	SMA N 1 Sragen	-	-	-	-	1
2	SMA N 2 Sragen	-	-	-	-	-
3	SMA N 3 Sragen	-	1	-	-	1
4	MAN 1 Sragen	-	-	-	-	-
5	SMK N 1 Sragen	-	1	-	1	1
6	SMK N 2 Sragen	-	-	-	-	-
7	SMA Muhammadiyah 1 Sragen	1	-	-	1	-
8	SMK Muhammadiyah 1 Sragen	-	-	1	-	-
9	SMK Muhammadiyah 2 Sragen	-	-	-	1	-
10	SMK Muhammadiyah 4 Sragen	1	-	-	1	-

Pengurutan berdasarkan tabel 6 dan prioritas variabel dengan urutan prioritas tertinggi akreditasi, biaya SPP, nilai UAN minimal, tingkat kelulusan dan jarak sebagai prioritas terendah.

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

Pengurutan Berdasarkan Rangking Favorit Sekolah pada y yang memiliki pemenuhan variabel yang sama.

SMKN 1 Sragen = nilai UAN, Kelulusan, Jarak (rangking Favorit 5)

1. SMK muhammadiyah 4 Sragen = akreditasi, SPP (rangking Favorit 7)
2. SMA Muhammadiyah 1 Sragen = SPP, Kelulusan (rangking Favorit 8)
3. SMA Negeri 3 Sragen = nilai UAN, jarak (rangking Favorit 3)
4. SMK Muhammadiyah 1 Sragen = akreditasi (rangking Favorit 10)
5. SMA Negeri 1 Sragen = nilai UAN (rangking Favorit 1)
6. SMK Muhammadiyah 2 Sragen = kelulusan (rangking Favorit 9)

Berdasarkan perhitungan dan pengurutan, maka sekolah yang ditampilkan pada hasil rekomendasi dengan urutan yang disesuaikan pemenuhan variabel dan rangking favorit sekolah. Data rangking sekolah diambil berdasarkan survei kepada sampel siswa maupun masyarakat umum di Kota Sragen.

#### 4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, kesimpulan yang dapat diambil yaitu aplikasi rekomendasi berbasis web menggunakan metode Naive Bayes yang telah dibuat dapat membantu siswa untuk memilih sekolah terbaik berdasarkan kriteria-kriteria yang telah diajukan.

Cara perekomendasi yaitu dengan memperhatikan tingkat *confidence* tertinggi yang dihasilkan setiap variabel dalam setiap sekolah, pemenuhan dan prioritas variable, maupun rangking favorit sekolah.

#### 5. Daftar Pustaka

- [1] Kemendikbud. 2012. *Iktisar Data Pendidikan Tahun 2011/2012*. Pusat data dan statistik pendidikan, Sekretariat jendral kementerian pendidikan dan kebudayaan.
- [2] Dinas Pendidikan dan Kebudayaan & BPS Sragen, 2010. [online] Available at <http://www.sragenkab.go.id/home.php?menu=50> [Accessed 1 Oktober 2015].
- [3] Purnamasari, H.J. 2011. *Program Bantu Pemilihan Tempat Kost di Yogyakarta Berbasis Web Menggunakan Metode Naive Bayes*. Undergraduate thesis, Duta Wacana Christian University.
- [4] Uyun, Shofwatul. & Madihatun, Yuni. 2011. *Model Rekomendasi Berbasis Fuzzy Untuk Pemilihan Sekolah Lanjutan Tingkat Atas*. Jurnal Informatika. No. 1 / Vol.5 / January 2011. Yogyakarta: Program Studi Teknik Informatika, Universitas Ahmad Dahlan Yogyakarta.
- [5] Fuadillah. 2011. *Sistem Aplikasi Penunjang Keputusan Pemilihan Sekolah Menengah Atas Sederajat Menggunakan Metode Fuzzy Multiple Criteria Decision*

- Making (FMCDM)*. Skripsi. Program Studi Sistem Informasi. Sekolah Tinggi Manajemen Informatika Dan Komputer Banjarbaru.
- [6] Nugroho, Yusuf Sulistyoy. Ulinnuha, Agus. & Aji, Nuruddin Nova Sekti. 2013. *Seleksi Sekolah Lanjutan Menggunakan Analytic Hierarchy Process*. KomuniTi, Vol. V, No. 2 September 2013. Surakarta: Universitas Muhammadiyah Surakarta.
- [7] Al-Janabi, S., & Janicki, R. 2016. *A density-based data cleaning approach for deduplication with data consistency and accuracy*. 2016 SAI Computing Conference (SAI). London, UK.
- [8] Prasetyo, Eko. 2012. *Data Mining konsep dan Aplikasi menggunakan Matlab Data Mining konsep*. Penerbit Andi. Yogyakarta
- [9] Yang, B. Lei, Y. & Yan, B. 2016. *Distributed Multi-Human Location Algorithm Using Naive Bayes Classifier for a Binary Pyroelectric Infrared Sensor Tracking System*. IEEE Sensors J. IEEE Sensors Journal, 16(1), 216-223.
- [10] Nugroho, Yusuf Sulistyoy. & Haryati, Syarifah N. 2015. *Klasifikasi dan Klustering Penjurusan Siswa SMA Negeri 3 Boyolali*. Jurnal Khazanah Informatika Vol. I No. 1 Desember 2015 Available at <http://journals.ums.ac.id/index.php/khif/article/view/1175/1012> [Accessed 14 Maret 2015]