

**KLASIFIKASI PENJURUSAN SISWA SEKOLAH MENENGAH ATAS DENGAN
MENGUNAKAN ALGORITMA NAÏVE BAYES
(Studi Kasus : SMA Muhammadiyah 3 Jember)**

¹*Siti Nurasyiroh Farianah*

²*Agung Nilogiri*

Email: salumsta@gmail.com, agungnilogiri@unmuhjember.ac.id

ABSTRAK

Penentuan jurusan pada siswa Sekolah Menengah Atas tentunya tidak terlepas dari penguasaan siswa terhadap mata pelajaran tersebut. SMA Muhammadiyah 3 Jember membuka 3 jurusan untuk kelangsungan proses belajar siswanya. Yaitu IPA, IPS dan Bahasa. Penjurusan di SMA Muhammadiyah 3 Jember baik IPA, IPS dan Bahasa dipertimbangkan oleh beberapa faktor yaitu. Nilai rapor SMP (IPA, IPS, Matematika, Bahasa Indonesia, dan Bahasa Inggris) semester ganjil selama 3 tahun. Dalam proses penjurusan sebelumnya masih tercatat dan belum menggunakan metode khusus. Pada penelitian ini algoritma *Naive Baye* diusulkan untuk digunakan sebagai alternatif metode untuk melakukan penjurusan. Penelitian ini menggunakan total dataset sebanyak 200 data. Data diuji menggunakan *k-fold cross validation*. Dari hasil penelitian ini model terbaik kelas IPA ada pada uji coba dengan $K=8$ pada uji ke 2 dengan nilai akurasi 88% dan presisi 71%, untuk kelas IPS ada pada uji coba dengan $K=8$ pada uji ke 6 dengan nilai akurasi 80% dan presisi 100%, dan kelas Bahasa ada pada uji coba dengan $K=10$ pada uji ke 6 dengan nilai akurasi 85% dan presisi 100%.

Kata kunci : Naive Bayes, Data Mining, Klasifikasi, Penjurusan.

**KLASIFIKASI PENJURUSAN SISWA SEKOLAH MENENGAH ATAS
DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES
(Studi Kasus : SMA Muhammadiyah 3 Jember)**

¹*Siti Nurasyiroh Farianah*

²*Agung Nilogiri*

Email: ¹salumsta@gmail.com, ²agungnilogiri@unmuhjember.ac.id

ABSTRACT

The determination of majors in high school students certainly cannot be separated from students' mastery of these subjects. Muhammadiyah 3 Jember High School opened 3 majors for the continuity of the student learning process. Namely Science, IPS and Language. Majors at Muhammadiyah 3 Jember High School in Natural Sciences, Social Sciences and Language are considered by several factors, i.e. Middle school report cards (Science, Social Sciences, Mathematics, Indonesian, and English) odd semester for 3 years. In the previous majors process was recorded and did not use special methods. In this study the Naive Bayes algorithm is proposed to be used as an alternative method for conducting majors. This study uses a total dataset of 200 data. Data were tested using k-fold cross validation. From the results of this study the best model of natural science class is on a trial with $K = 8$ in the second test with an accuracy value of 88% and 71% precision, for the IPS class is in a trial with $K = 8$ in the sixth test with an accuracy value of 80% and 100% precision, and the Language class is in the trial with $K = 10$ in the 6th test with 85% accuracy and 100% precision.

Keywords: Naive Bayes, Data Mining, Classification, Majors.

PENDAHULUAN

1. Latar Belakang

Salah satu perubahan yang dapat dirasakan dengan penerapan kurikulum 2013 adalah proses penjurusan siswa Sekolah Menengah Atas atau SMA. Dalam penerapan kurikulum 2013 proses penjurusan Siswa Sekolah Menengah Atas dilakukan langsung pada awal masuk Sekolah Menengah Atas atau pada kelas X. Hal tersebut berbeda dengan kurikulum sebelumnya di mana pada proses penjurusan siswa Sekolah Menengah Atas dilakukan pada kelas XI.

Dalam penelitian ini mengambil objek penelitian pada SMA Muhammadiyah 3

Jember yang sedang melaksanakan kurikulum 2013. Sekolah Menengah Atas Muhammadiyah 3 Jember membuka 3 jurusan untuk kelangsungan proses belajar siswanya, yaitu jurusan Ilmu Pengetahuan Alam (IPA), Ilmu Pengetahuan Sosial (IPS), dan Bahasa. Penjurusan di SMA Muhammadiyah 3 Jember dilakukan agar para siswa bisa menyalurkan bakat dan minat serta kemampuan di jurusan yang ada.

2. Rumusan Masalah

- Berapa tingkat presisi Algoritma *Naive Bayes* dalam menentukan Penjurusan di SMA Muhammadiyah 3 Jember
- Berapa akurasi Algoritma *Naive Bayes* dalam menentukan penjurusan di SMA Muhammadiyah 3 Jember?

3. Batasan Masalah

- Data yang digunakan untuk penelitian adalah data nilai siswa baru atau siswa kelas X di SMA Muhammadiyah 3 Jember berupa data rata-rata raport (IPA, IPS, Matematika, Bahasa Indonesia, Bahasa Inggris) semester ganjil Sekolah Menengah Pertama (SMP) selama 3 tahun.
- Data yang digunakan adalah data siswa baru tahun 2018-2019 sebanyak 200 data.

4. Tujuan Penelitian

- Mengukur tingkat Presisi dalam menentukan penjurusan di SMA Muhammadiyah 3 Jember.
- Mengukur tingkat Akurasi menggunakan algoritma klasifikasi *Naive Bayes*.

5. Manfaat Penelitian

- Membantu dan memudahkan pihak SMA Muhammadiyah 3 Jember khususnya guru Bimbingan Karier (BK) dalam melakukan Keputusan penjurusan siswa secara tepat dan akurat.
- Sebagai alternatif metode penjurusan di SMA Muhammadiyah 3 Jember.

METODE PENELITIAN

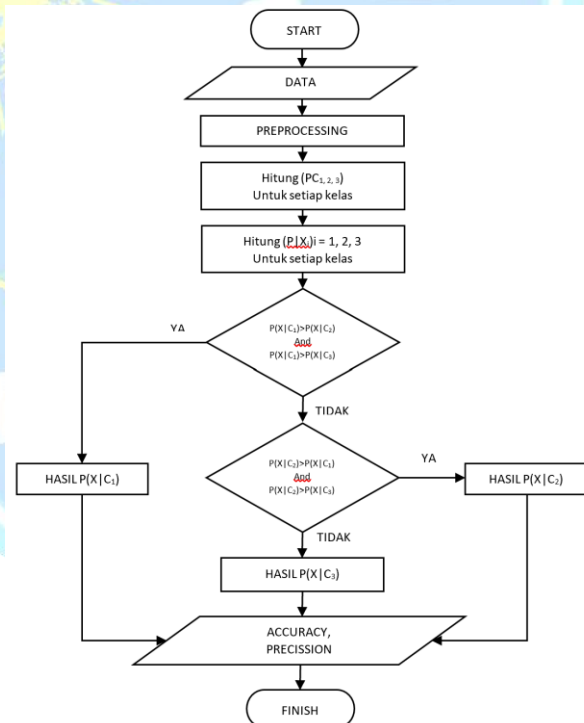
1. Pengumpulan Data

Data yang digunakan untuk penelitian ini merupakan data dari SMA Muhammadiyah 3 Jember yang terdiri dari 200 data. Pengumpulan data diawali dengan datang ke SMA Muhammadiyah 3 Jember dan melakukan wawancara untuk mengetahui bagaimana sistem pengambilan jurusan yang dilakukan di SMA Muhammadiyah 3 Jember.

2. Data Set

Data set untuk penelitian ini adalah data nilai siswa baru yang berupa nilai raport (IPA, IPS, Matematika, Bahasa Indonesia, Bahasa Inggris) semester ganjil selama tiga tahun di Sekolah Menengah Pertama (SMP).

3. Flowchart Perhitungan Naive Bayes



4. Perhitungan Klasifikasi Algoritma Naive Bayes

Nama	KATEGORI					KEP
	MAT	IPA	IPS	B.ING	B.IND	
ADELA IBRAHIM	SEDANG	TINGGI	TINGGI	SEDANG	SEDANG	IPS
AKBAR YUDHA PRASETYO	TINGGI	TINGGI	TINGGI	SEDANG	SEDANG	IPA
DEVA OCTA VINONA	RENDAH	RENDAH	RENDAH	RENDAH	RENDAH	IPS
DEVIRA DWI PUSPITASARI	RENDAH	RENDAH	TINGGI	SEDANG	RENDAH	IPS
GALUH SILA PUTRI ADI	RENDAH	SEDANG	SEDANG	SEDANG	TINGGI	BHS
GHINAR ZAHIRANI AISYAHAKIM	TINGGI	SEDANG	TINGGI	TINGGI	TINGGI	BHS
JIHAN ROSYIDAH MAGHFIROH	TINGGI	TINGGI	TINGGI	SEDANG	TINGGI	IPA
JONAVITO KURNIA SEJATI	SEDANG	SEDANG	SEDANG	SEDANG	SEDANG	IPA
KEVIN AKBAR MAULANA	RENDAH	RENDAH	RENDAH	RENDAH	RENDAH	IPS
MAYA ANGRAINI	RENDAH	SEDANG	RENDAH	RENDAH	SEDANG	BHS
MUHAMMAD ROMADHONI	SEDANG	TINGGI	TINGGI	TINGGI	TINGGI	IPA
PERDANA NANDA HARI PUTRA	SEDANG	TINGGI	TINGGI	TINGGI	TINGGI	BHS
PUTRI NUR FITRIYANI	RENDAH	SEDANG	RENDAH	RENDAH	RENDAH	IPA
PUTRI NUR SYIFA'	SEDANG	TINGGI	TINGGI	SEDANG	TINGGI	BHS
PUTRI SEPTIAN A.W	SEDANG	RENDAH	SEDANG	RENDAH	RENDAH	IPS
PYVICO FADERA GARDY	RENDAH	SEDANG	SEDANG	TINGGI	RENDAH	IPS
RATRI SEPTIANA DEWI	TINGGI	SEDANG	TINGGI	SEDANG	TINGGI	IPS
RICHO ANDRO ARDIANSYAH W	RENDAH	RENDAH	RENDAH	SEDANG	RENDAH	IPA
WAFDA MARAYA RIZAL	RENDAH	RENDAH	RENDAH	RENDAH	RENDAH	BHS
YOGA LAKSONO	SEDANG	TINGGI	SEDANG	RENDAH	SEDANG	IPA

Hitung jumlah class/label

Attribute	Jumlah Kasus	IPA	IPS	BHS
		C1	C2	C3
TOTAL	20	7	7	6

Terdapat 3 class dari data training tersebut, yaitu :

- C1 (Class 1) → Prediksi IPA = 7 record
- C2 (Class 2) → Prediksi IPS = 7 record
- C3 (Class 3) → Prediksi Bahasa = 6 Record
- Total = 20 record

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

Maka :

- $P(C1) = 7/20 = 0,35$
- $P(C2) = 7/20 = 0,35$
- $P(C3) = 6/20 = 0,3$

5. Pengujian

Teknik pengujian dalam penelitian ini yaitu dengan menggunakan pengujian *K-fold Cross Validation*. *K-fold Cross Validation* merupakan teknik untuk menghasilkan

sebuah akurasi dengan cara membagi data set ke dalam data testing dan data training.

IMPLEMENTASI DAN PENGUJIAN

1. Gambaran Data

Data dalam penelitian ini adalah siswa baru dari SMA Muhammadiyah 3 Jember tahun 2018-2019 sebanyak 200 data. Dari 200 data siswa tersebut 70 siswa terklasifikasi dalam jurusan Bahasa sebanyak 70 siswa, 64 siswa terklasifikasi jurusan IPA dan 66 siswa terklasifikasi jurusan IPS.

2. Hasil Klasifikasi Pengujian K-vold Cross Validation

Berikut adalah hasil pengujian K-Vold Cross Validation dengan K = 2.

Skenario 2.1

Skenario 2.1	1-100	101-200
--------------	-------	---------

Kriteria IPA

		Keputusan	
		+ IPA	- Selain IPA
Klasifikasi	+ IPA	TP = 7	FP = 6
	-Selain IPA	FN = 21	TN = 66

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{7 + 66}{100} \times 100\% \\
 &= 73\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Presisi} &= \frac{TP}{TP+FP} \times 100\% \\
 &= \frac{7}{7+6} \times 100\% \\
 &= \frac{7}{13} \times 100\% \\
 &= 54\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Presisi} &= \frac{TP}{TP+FP} \times 100\% \\
 &= \frac{14}{14+19} \times 100\% \\
 &= \frac{14}{33} \times 100\% \\
 &= 42\%
 \end{aligned}$$

Kriteria IPS

		Keputusan	
		+ IPS	-Selain IPS
Klasifikasi	+ IPS	TP = 23	FP = 31
	-Selain IPS	FN = 11	TN = 35

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{23+35}{23+35+31+11} \times 100\% \\
 &= 58\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Presisi} &= \frac{TP}{TP+FP} \times 100\% \\
 &= \frac{23}{23+31} \times 100\% \\
 &= \frac{23}{54} \times 100\% \\
 &= 43\%
 \end{aligned}$$

Kriteria Bahasa

		Keputusan	
		+ Bahasa	-Selain Bahasa
Klasifikasi	+ Bahasa	TP = 14	FP = 19
	-Selain Bahasa	FN = 24	TN = 43

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{14+43}{14+43+19+24} \times 100\% \\
 &= 57\%
 \end{aligned}$$

3. Hasil

- Dari hasil skenario *k-fold cross validation* dengan nilai $k = 2, 4, 5, 8$ dan 10 didapat bahwa pada kelas IPA, model terbaik ada pada uji coba dengan $k = 8$ pada uji ke dua dengan nilai akurasi 88% dan presisi 71% .
- Dari hasil skenario *k-fold cross validation* dengan nilai $k = 2, 4, 5, 8$ dan 10 didapat bahwa pada kelas IPS, model terbaik ada pada uji coba dengan $k = 8$ pada uji ke enam dengan nilai akurasi 80% dan presisi 100% .
- hasil skenario *k-fold cross validation* dengan nilai $k = 2, 4, 5, 8$ dan 10 didapat bahwa pada kelas Bahasa, model terbaik ada pada uji coba dengan $k = 10$ pada uji ke enam dengan nilai akurasi 85% dan presisi 100% .

PENUTUP

1. Kesimpulan

Dari hasil penelitian metode *Naive Bayes* terhadap klasifikasi penjurusan Siswa Sekolah Menengah Atas dengan total data 200 data dan skenario uji menggunakan *k-fold cross*

validation dalam penelitian ini menggunakan $K = 2, 4, 5, 8, 10$ adalah :

- a. Dari hasil skenario *k-fold cross validation* dengan nilai $k = 2, 4, 5, 8$ dan 10 didapat bahwa pada kelas IPA, model terbaik ada pada uji coba dengan $k = 8$ pada uji ke dua dengan nilai akurasi 88% dan presisi 71% .
- b. Dari hasil skenario *k-fold cross validation* dengan nilai $k = 2, 4, 5, 8$ dan 10 didapat bahwa pada kelas IPS, model terbaik ada pada uji coba dengan $k = 8$ pada uji ke enam dengan nilai akurasi 80% dan presisi 100% .
- c. Dari hasil skenario *k-fold cross validation* dengan nilai $k = 2, 4, 5, 8$ dan 10 didapat bahwa pada kelas Bahasa, model terbaik ada pada uji coba dengan $k = 10$ pada uji ke enam dengan nilai akurasi 85% dan presisi 100% .

2. Saran

Adapun saran untuk penelitian selanjutnya dapat dilakukan dengan membandingkan antara metode Naive Bayes dan metode yang lain.

DAFTAR PUSTAKA

- [1] Bustami. 2014. Penerapan Algoritma Naive Bayes untuk Mengklasifikasi Data Nasabah Asuransi. *jurnal informatika* vol. 8, no. 1. Aceh Indonesia.
- [2] Fithri, D. L., & Darmanto, E. (2014). Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naive Bayes. In *Seminar Nasional Teknologi dan Informatika 2014*. Muria Kudus University.
- [3] Han, J., & Kamber, M. (2006). Classification and prediction. *Data mining: Concepts and techniques*, 347-350.
- [4] Irawan. 2005. Kategorisasi dokumen web secara otomatis Berdasarkan folksonomy menggunakan Multinomial naive bayes classifier (automatic folksonomy categorization Of web documents using multinomial Naive bayes classifier). Tugas Akhir Diajukan Sebagai Syarat Kelulusan Pendidikan Program Sarjana Jurusan Teknik Informatika Sekolah Tinggi Teknologi Telkom. Bandung.
- [5] J. Han and M. Kamber. (2006) *Data Mining : Concept and Techniques Second Edition*. San Francisco: Morgan Kaufmann Publishers.
- [6] Kristanto, O. (2014). Penerapan Algoritma Klasifikasi Data Mining ID3 Untuk Menentukan Penjurusan Siswa SMAN 6 Semarang. *Dian Nuswantoro Fakultas Ilmu Komputer Skripsi*
- [7] Kementerian Pendidikan dan Kebudayaan, Salinan – *Permendikbud Nomor 81A Tahun 2013 Tentang Implementasi Kurikulum*. Jakarta, Indonesia, 2013.
- [8] Nugroho, Y. S. (2014). Data Mining Menggunakan Algoritma Naive Bayes Untuk

Klasifikasi Kelulusan Mahasiswa Universitas
Dian Nuswantoro. *Dian Nuswantoro Fakultas
Ilmu Komputer Skripsi.*

[9] Prastya, F. A. (2014). Penerapan Algoritma
C4. 5 untuk Prediksi Jurusan Siswa SMAN 3
Rembang. *Semarang. Jurusan Teknik
Informatika Fasilkom Udinus.*

[10] Santosa, B. (2007). Data mining teknik
pemanfaatan data untuk keperluan
bisnis. *Yogyakarta: Graha Ilmu, 978(979),
756.*

[11] Wihardi, Y. (2013). K-Folds Cross
Validation. Retrieved April 25, 2018.

