

# ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI KEPRIBADIAN SISWA SEKOLAH VOKASI BERBASIS PEMROGRAMAN PHP

*Faizal Riza*

*Program Studi Teknik Informatika, Institut Teknologi Budi Utomo Jakarta  
faizalriza@itbu.ac.id*

## **Abstrak**

Sekolah adalah salah satu komponen penting di masyarakat yang mengakomodasi siswa dengan beragam latar belakang kepribadian. Beragamnya latar belakang kepribadian siswa akan menyulitkan guru untuk menerapkan metode pembelajaran yang sama. Kepribadian siswa yang dipetakan sejak awal masa Pendidikan akan memudahkan guru melakukan transfer ilmu dan nilai kehidupan. Keberhasilan pendidikan di sekolah akan meningkatkan kepercayaan masyarakat terhadap kualitas sekolah. Beberapa hal tersebut menjadi latar belakang diperlukannya cara untuk memprediksi kepribadian siswa dalam rangka memfasilitasi proses pendidikan, baik dalam Pendidikan kecerdasan maupun karakter. Prediksi yang digunakan dalam penelitian ini adalah Naive Bayes. Aplikasi yang dibuat menggunakan 100 dataset latih dan 5 data uji. Pada aplikasi yang dibuat, para siswa mengisi tanggapan terhadap serangkaian pernyataan. Respon siswa dicatat dan dihitung menggunakan algoritma Naive Bayes untuk memprediksi kepribadian utama siswa. Akurasi yang diperoleh sebesar 60% untuk memprediksi kepribadian siswa.

Kata kunci: *naive bayes*, algoritma, vokasi, kepribadian

## **1. PENDAHULUAN**

Tingkat pengangguran terbuka (TPT) di Indonesia pada Februari 2018 adalah 6,87 juta orang atau 5,13%. Angka ini turun sekitar 2% dibandingkan Februari 2017 yang berjumlah 7,01 juta orang atau 5,33%. Pada bulan Februari 2016, tingkat pengangguran terbuka tertinggi berada di level SMK 9,84%. Angka ini meningkat 0,79% dibandingkan dengan pengangguran pada Februari 2015 (BPS-Statistics Indonesia, 2017).

Lulusan kejuruan saat ini tidak mampu menjawab masalah ketenagakerjaan yang dibutuhkan oleh tenaga kerja. Karena pengetahuan dan keterampilan dibidang vokasi yang mereka miliki masih rendah. Hal ini memunculkan kesenjangan antara kualifikasi lulusan dan kebutuhan industri, Menteri Koordinator Bidang Perekonomian Darmin Nasution mengatakan bahwa perubahan revolusioner diperlukan untuk meningkatkan kualitas pendidikan kejuruan seperti sekolah kejuruan, politeknik dan pusat pelatihan kejuruan (BLK). Saat ini, setidaknya ada 144 kompetensi keahlian dalam pendidikan kejuruan di Indonesia dengan 5 jurusan dengan siswa terbanyak termasuk Departemen Teknik Komputer dan Jaringan

berjumlah 578.000 siswa, Teknik Kendaraan Ringan 574.000 siswa, Akuntansi 430.000 siswa, Administrasi Kantor 428.000 siswa, dan Sepeda Rekayasa 270.000 sepeda motor siswa. Lima jurusan berkontribusi 54,9% dari total siswa kejuruan di Indonesia (Muamar Zain, 2015).

Tes statistik tidak dapat menunjukkan solusi tepat untuk mengembangkan soft skill siswa. Hal tersebut dikarenakan hasil statistik tidak dapat membaca informasi secara semantik. Menurut Faizal Riza et. al. (2020), dinyatakan bahwa *"the data of matching method will count the number of keywords that appear in the document then return the order of documents with the highest number of occurrences to the user. This can result in the performance of information retrieval becoming less good because they do not pay attention to the information semantically"*.

Tes psikologi adalah salah satu tes yang digunakan untuk mendidik siswa. Pengetahuan tentang kepribadian seseorang dianggap penting karena memengaruhi perilaku orang-orang yang membuat keputusan yang memiliki dampak positif atau negatif. Selain itu, kepribadian seseorang dapat menjadi salah satu faktor yang digunakan untuk memberinya

pekerjaan (Sarwono, 2011). Setelah kurikulum SMK ditetapkan dan diselaraskan dengan kebutuhan Industri, perlu untuk memperkuat hasil belajar siswa agar dapat menghasilkan lulusan yang memiliki kualifikasi industri yang direncanakan dengan baik (by design) bukan karena kebetulan saja (by accident). Hal terpenting dalam pendidikan vokasional adalah ketercapaian standar pengetahuan dan keterampilan terhadap bidang vokasi yang ditempuhnya. Keberhasilan pencapaian standar pengetahuan dan keterampilan sangat bergantung pada keberhasilan proses pembelajaran. (Furnham, Chamorro-Premuzic, & McDougall, 2003). Proses pembelajaran yang dilakukan guru akan berhasil apabila metode pembelajaran bersesuaian dengan kepribadian siswa. Penelitian ini dibuat untuk mengatasi masalah identifikasi kepribadian siswa yang dihadapi oleh guru.

## 2. METODOLOGI

Atribut yang digunakan untuk perbandingan metode dalam data mining ditentukan sesuai dengan kebutuhan analisis. Daftar atribut yang digunakan dapat dilihat pada tabel 1. Ada dua jenis variabel dalam proses ini adalah:

- a) Variabel Dependent (Y)  
Variabel dependen adalah variabel yang nilainya tergantung atau terikat berdasarkan nilai-nilai variabel lain
- b) Variabel Independen (X)  
Variabel independen adalah variabel yang nilainya tidak tergantung pada nilai-nilai variabel lain.

Tabel 1.  
Atribut yang Digunakan

Attribut	Variabel
Personality	Y
Gender	X1
Age	X2
School Status	X3
Answer A	X4
Answer B	X5
Answer C	X6
Answer D	X6
Kepribadian Asli	X7

Sumber Data : Hasil Olahan Data Penelitian

Pengumpulan data yang diambil dari data dari respon siswa SMK. Data pelatihan yang digunakan dalam penelitian ini adalah 100 siswa. Untuk pengujian akurasi, penelitian ini menggunakan 5 siswa.

Analisis data menggunakan Algoritma Naive Bayes, algoritma yang mampu memprediksi peluang masa depan berdasarkan pengalaman masa lalu sehingga dikenal sebagai Teorema Bayes. Karakteristik utama dari algoritma Naive Bayes adalah asumsi yang sangat kuat (naif) akan independensi dari setiap kondisi / peristiwa.

Keuntungan menggunakan algoritma ini yaitu membutuhkan hanya sedikit data latih untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi. Karena diasumsikan sebagai variabel independen, hanya varians variabel dalam kelas yang diperlukan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians (Nasution *et al*, 2015).

Tahapan proses algoritma Naive Bayes adalah:

- 1) Hitung jumlah kelas / label.
- 2) Hitung Jumlah Kasus Per Kelas
- 3) Gandakan Semua Variabel Kelas
- 4) Bandingkan Hasil dengan Kelas.

Algoritma hasil elaborasi menyebabkan faktor yang semakin kompleks - kondisi yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisis satu per satu. Akibatnya, perhitungan menjadi sulit dilakukan. Di sini digunakan asumsi independensi (naif) yang sangat tinggi, bahwa masing-masing instruksi (F1, F2, ..., Fn) bebas (independen) satu sama lain.

## 3. HASIL DAN PEMBAHASAN

Data yang diperoleh terbagi menjadi dua jenis yaitu data pelatihan dan data pengujian dengan variabel yaitu nama, jenis kelamin, usia, status sekolah, jawaban A, jawaban B, jawaban C, jawaban D dan kelas asli (kepribadian). Dalam tabel 2 adalah data latih yang merupakan data pelatihan yang digunakan sebagai data pendukung untuk aplikasi ini dan data tabel 3 adalah data pengujian.

No	Nama	Jenis Kelamin	Usia	Teknik	Jawaban A	Jawaban B	Jawaban C	Jawaban D	Risiko Asli
1	Yusuf Alimudin D.P	L	22	Seni	14	6	9	5	3 Sangat
2	Yusuf Alimudin D.P	L	22	Seni	13	6	9	12	3 Sangat
3	Yusuf Alimudin D.P	L	22	Seni	12	7	11	8	3 Sangat
4	Yusuf Alimudin D.P	L	22	Seni	11	8	10	10	3 Sangat
5	Yusuf Alimudin D.P	L	22	Seni	10	9	11	10	3 Sangat

Gambar 2. Tabulasi Data Latih

Sistem melakukan perhitungan menggunakan Naive Bayes, untuk setiap variabel menghasilkan nilai kepercayaan pada variabel Y. Prediksi yang diungkapkan oleh sistem didasarkan pada nilai terbesar. (Nasution *et al*,2015)

No	Nama	Jenis Kelamin	Usia	Teknik	Jawaban A	Jawaban B	Jawaban C	Jawaban D	Risiko Asli
1	Yusuf Alimudin D.P	L	22	Seni	14	6	9	5	3 Sangat
2	Yusuf Alimudin D.P	L	22	Seni	13	6	9	12	3 Sangat
3	Yusuf Alimudin D.P	L	22	Seni	12	7	11	8	3 Sangat
4	Yusuf Alimudin D.P	L	22	Seni	11	8	10	10	3 Sangat
5	Yusuf Alimudin D.P	L	22	Seni	10	9	11	10	3 Sangat

Gambar 3. Tabulasi Data Uji

Diberikan input oleh user potongan data testing baru seperti pada tabel III , maka dapat dijelaskan melalui perhitungan sebagai berikut :

Jenis Kelamin	Seni	Seni	Seni	Seni
Seni	0.4000	0.5000	0.5000	0.5000
Seni	0.5000	0.4000	0.4000	0.4000
Seni	0.5000	0.5000	0.5000	0.5000
Seni	0.5000	0.4000	0.4000	0.4000

Gambar 4. Tabulasi Data Uji-1

Kemudian aplikasi akan menampilkan nilai masing-masing prediksi terbobot dari empat fitur kriteria kepribadian yang ditetapkan. Hasilnya ditampilkan pada gambar 5.

**PROBABILITAS**

Nilai Senilai = 0.2000 x 0.5000 x 0.4000 x 0.5000 = 0.2500

Nilai Senilai = 0.2000 x 0.4000 x 0.5000 x 0.5000 = 0.2000

Nilai Senilai = 0.2000 x 0.4000 x 0.5000 x 0.5000 = 0.2000

Nilai Senilai = 0.2000 x 0.4000 x 0.5000 x 0.5000 = 0.2000

Hasil Prediksi : Koleris

Gambar 5. Prediksi Naive Bayes pada Data Uji

Analisis perhitungan pada hasil prediksi kepribadian menghasilkan rincian akurasi dan laju error pada data uji. Rincian akurasi dan laju error ditampilkan pada gambar 6.

No	Nama	Jenis Kelamin	Usia	Teknik	Jawaban A	Jawaban B	Jawaban C	Jawaban D	Risiko Asli	Risiko Asli
1	Yusuf Alimudin D.P	L	22	Seni	14	6	9	5	3 Sangat	3 Sangat
2	Yusuf Alimudin D.P	L	22	Seni	13	6	9	12	3 Sangat	3 Sangat
3	Yusuf Alimudin D.P	L	22	Seni	12	7	11	8	3 Sangat	3 Sangat
4	Yusuf Alimudin D.P	L	22	Seni	11	8	10	10	3 Sangat	3 Sangat
5	Yusuf Alimudin D.P	L	22	Seni	10	9	11	10	3 Sangat	3 Sangat

Akurasi = 60 %  
LAJU ERROR = 40 %

Gambar 5. Akurasi dan Laju Error pada Data Uji

Hasil perhitungan menggunakan algoritma Naive Bayes untuk memprediksi kepribadian siswa mendapatkan akurasi sebesar 60 % untuk 5 data testing dari pengolahan 100 data latih

#### 4. KESIMPULAN

Berdasarkan hasil pengujian penggunaan algoritma Naive Bayes untuk memprediksi kepribadian siswa dalam penelitian ini disimpulkan bahwa algoritma Naive Bayes memiliki kinerja yang baik dalam hal akurasi yaitu sebesar 60%. Ini karena data pelatihan siswa masih perlu direproduksi dan diberikan arahan.

#### DAFTAR PUSTAKA

A. Nasution, N Djahara, K Zamsure (2015), "Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes (Studi Kasus: Fasilkom Unilak)," Teknol. Informasi dan Komunilasi Digital. Volume 6 BPS-Statistics Indonesia. (2017). Statistical Yearbook of Indonesia 2017. BPS-Statistics Indonesia. <https://doi.org/10.2307/2541047>

Faizal Riza et al (2020). Information Retrieval Technique for Indonesian PDF Document with Modified Stemming Porter Method Using PHP. J. Phys.: Conf. Ser. 1477 032016

Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2003). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. Learning and Individual Differences. <https://doi.org/10.1016/j.lindif.2003.08.002>

Muamar Zain, D. S. (2015). ESP vs EGP: A Reflection of Implementation of Curriculum 2013 in Vocational Schools. The 62nd TEFLIN International Conference 2015.

Sarwono, S. W. (2011). Psikologi Remaja Edisi Revisi. In Psikologi Remaja. <https://doi.org/10.1108/09513551011032482.Ba stian>