

Algoritma *Naive Bayes* untuk Memprediksi Waktu Pengerjaan Uji Kompetensi Keahlian (UKK) Siswa Sekolah Menengah Kejuruan

Muhammad Zaenal Mutaqin ^{1,*}, Muhamad Malik Mutoffar ²

* Korespondensi: e-mail: zaenal@ibm.ac.id

¹ Teknik Informatika; Institut Bisnis Muhammadiyah Bekasi; Jl. Sersan Aswan No.16, Margahayu, Bekasi Timur, Kota Bekasi, Telp/fax 021-82693450; e-mail: zaenal@ibm.ac.id

² Teknik Informatika, Sekolah Tinggi Teknologi Bandung; Jl. Soekarno Hatta No. 378 Bandung, Jawa Barat, 022-5224000; e-mail: malik@sttbandung.ac.id

Submitted : **6 April 2023**
Revised : **20 April 2023**
Accepted : **11 Mei 2023**
Published : **30 Mei 2023**

Abstract

This research is motivated by the need for a method or method that helps teachers and schools to predict the speed of time for students UKK work so that schools are more effective in preparing students to face UKK with faster processing time where the current problem is that schools are still using manual prediction methods . The hypothesis of the researchers is that by implementing the Naive Bayes algorithms to predict the length of time the UKK Student can work, it can produce more perfect predictions so that school management is more efficient in providing solutions for students who are predicted to work slowly on SMK Bhakti Persada Bekasi . UKK is the final assessment in order to determine the achievement of competencies for vocational students. The use of Data Mining with artificial classification and intelligence models that will predict the length of time spent on UKK in terms of student completion time quickly, normally or slowly. The Algorithm method used is Naive Bayes with the prediction accuracy of 99.11%.

Keywords: Competency Test Expertise (UKK), Data Mining, Naive Bayes, Prediction

Abstrak

Penelitian dilatarbelakangi oleh kebutuhan akan cara atau metode yang membantu guru dan pihak sekolah untuk meprediksi kecepatan waktu pengerjaan UKK siswa sehingga sekolah lebih efektif dalam mempersiapkan siswa dalam menghadapi UKK dengan waktu pengerjaan lebih cepat dimana masalah saat ini sekolah masih menggunakan metode prediksi manual. Hipotesa peneliti bahwa dengan implementasi algoritma *Naive Bayes* untuk memprediksi lama waktu pengerjaan UKK Siswa, dapat menghasilkan prediksi yang lebih sempurna sehingga manajemen sekolah lebih efisien dalam memberikan solusi bagi siswa-siswi yang diprediksi akan mengerjakan UKK dengan lamban pada SMK Bhakti Persada Kota Bekasi. UKK adalah Penilaian akhir dalam rangka menentukan capaian kompetensi bagi siswa SMK. Penggunaan *Data Mining* dengan model klasifikasi dan kecerdasan buatan yang akan memprediksi lama waktu pengerjaan UKK dalam hal waktu penyelesaian siswa dengan cepat, normal atau lambat. Metode Algoritma yang digunakan adalah *Naive Bayes* dengan hasil akurasi prediksi sebesar 99,11%.

Kata kunci: *Data Mining*, *Naive Bayes*, Prediksi, Uji Kompetensi Keahlian (UKK), SMK

1. Pendahuluan

Saat akan memulai penelitian ini peneliti menelusuri dengan melakukan riset awal tentang masalah apa yang sebenarnya yang menjadi kendala dalam meningkatkan kualitas hasil Uji Kompetensi Keahlian (UKK) pada SMK Bhakti Persada Kota Bekasi, guru atau *management* sekolah masih menggunakan metode manual atau tidak melakukan analisa atau prediksi keberhasilan UKK saat mempersiapkan UKK Siswa.

Menurut KBBI Algoritma adalah urutan logis pengambilan keputusan untuk pemecahan masalah, urutan langkah-langkah logis penyelesaian masalah yang disusun secara sistematis (Rinaldi, 2011). *Naive Bayes* salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk machine learning dan Data Mining. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes. Teorema tersebut dikombinasikan dengan naive dimana diasumsikan kondisi antar atribut saling bebas. Algoritma *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Penelitian terkait implementasi algoritma *Naive Bayes* digunakan untuk menentukan kesiapan siswa dalam menghadapi ujian nasional (Muchammad, 2016).

Pendidikan kejuruan suatu program yang berada di bawah pendidikan tinggi yang diorganisasi menyiapkan peserta didik untuk memasuki dunia kerja tertentu atau meningkatkan pekerjaan dalam dunia kerja. Kompetensi adalah kemampuan (*ability*) atau kapasitas seseorang untuk mengerjakan berbagai tugas dalam suatu pekerjaan, dimana kemampuan ini ditentukan oleh 2 (dua) faktor yaitu kemampuan intelektual dan kemampuan fisik (Robbins, 2006). Salah satu penelitian terkait uji kompetensi dilakukan Utomo pada 2015 tentang prediksi nilai ujian nasional produktif sekolah menengah kejuruan menggunakan metode *Neural Network* (Utomo, 2015).

Uji kompetensi ini dinyatakan dalam PP 19 Tahun 2005 tentang Standar Nasional pendidikan pasal 89 ayat 1 bahwa pencapaian kompetensi akhir peserta didik dinyatakan dalam dokumen ijazah dan/atau sertifikat kompetensi Selanjutnya pasal 89 ayat 5 juga dinyatakan bahwa sertifikat kompetensi dimaksud pada ayat 1 diterbitkan oleh satuan pendidikan yang terakreditasi atau lembaga sertifikasi mandiri yang dibentuk oleh organisasi profesi yang diakui Pemerintah sebagai tanda bahwa peserta didik yang bersangkutan telah lulus uji kompetensi. (Surjono, 2008).

Penilaian dilakukan untuk menentukan apakah peserta didik sudah kompeten atau belum kompeten. Penentuan dilakukan dengan membandingkan bukti-bukti hasil belajar yang diperoleh oleh peserta didik dengan kriteria kinerja yang ditetapkan pada standar kompetensi. Proses pengumpulan bukti-bukti tersebut dapat dilakukan melalui berbagai metode, yang disesuaikan dengan karakteristik keahlian dan konteks pengujian (Budiyono, 2011).

Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan

berdasarkan pengalaman di masa lalu sehingga dikenal sebagai Teorema Bayes. *Naive Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari master tabel keputusan (Olson & Delen, 2008).

2. Metode Penelitian

Pengumpulan data tidak hanya mengambil data, tetapi juga harus mampu mendeskripsikan data dan memberikan kontribusi pada pengetahuan. Data harus jelas, memiliki relasi, dapat diukur, diprediksi, memiliki generalisasi, dan terkait dengan teori. (Dawson, 2019). Penelitian menggunakan pendekatan kuantitatif. Kegiatan penelitian melalui beberapa tahap dalam pengembangannya yakni pertama dilakukan pengumpulan materi-materi kepustakaan yang berhubungan dengan pengambilan judul.

Pembahasan dalam penelitian difokuskan pada prediksi kelulusan siswa di lingkungan SMK Bhakti Persada Kota Bekasi yang terletak di kota Bekasi dengan alamat Jl. Antara Dalam Kelurahan Jatimakmur Kecamatan Pondok Gede Bekasi, dengan rencana waktu penelitian selama tiga bulan dimulai pada bulan Oktober sampai Desember 2018.

Objek penelitian adalah Data Siswa-siswa jurusan Teknik Komputer dan Jaringan (TKJ) dan Akuntansi (AK) SMK Bhakti Persada Kota Bekasi. Data yang digunakan dalam penelitian merupakan data primer yang diambil dari data dapodik SMK Bhakti Persada per tahun ajaran 2018/2019 (tahun lulus 2017, 2018 dan akan lulus tahun 2019) dan data hasil UKK dan *Try out* UKK per Oktober 2018.

Kegiatan penelitian melalui beberapa tahap dalam pengembangannya yakni pertama dilakukan pengumpulan materi-materi kepustakaan yang berhubungan dengan pengambilan judul. Kemudian langkah selanjutnya yaitu survei langsung ke tempat penelitian yaitu SMK Bhakti Persada Kota Bekasi. Dari penelitian tersebut akan di dapat beberapa kriteria yang digunakan untuk penelitian lebih lanjut. Setelah dilakukan survey maka didapatkan data yang akan digunakan dalam penelitian ini, yang terdiri dari beberapa kriteria. Selanjutnya data diolah dahulu karena kualitas data masukan akan menjadi kurang baik jika data tersebut kurang lengkap, tidak konsisten dan tidak rapi (Vercellis, 2011), untuk menghasilkan laporan yang dijadikan *alternative* keputusan.

Data mining proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Liang et al., 2005). Terdapat beberapa istilah lain yang memiliki makna sama dengan *Data mining*, yaitu *Knowledge discovery in databases (KDD)*, ekstraksi pengetahuan (*knowledge extraction*), Analisa data/pola (*data/pattern analysis*), kecerdasan bisnis (*business intelligence*) dan *data archaeology* dan *data dredging* (Larose, 2005).

Algoritma *Naïve Bayes* metode klasifikasi dengan menggunakan metode probabilitas dan juga statistik (Siregar et al., 2017).

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (1)$$

Dengan $X =$ is an unknown class, $H =$ data hypothesis X , $P(H|X) =$ probability of hypothesis H based on condition x , $P(H) =$ probability of hypothesis H , $P(X) =$ probability of hypothesis X . (Yunianto et al., 2022).

Tingkat akurasi adalah salah satu aspek yang menjadi parameter kehandalan dari algoritma klasifikasi. Sistem akan melakukan klasifikasi yang diharapkan dapat mengklasifikasikan semua set dengan benar, tetapi tidak semua sistem bekerja dengan baik atau tidak bisa 100% akurat (Prasetyo, 2012). Rumus 2 adalah rumus formula untuk menghitung tingkat akurasi (Febriya et al., 2022).

$$\text{Akurasi} = \frac{\text{Jumlah data yang di prediksi secara benar}}{\text{Jumlah prediksi yang dilakukan}} \quad (2)$$

Untuk mengestimasi tingkat kesalahan pada sebuah model algoritma, terdapat beberapa teknik salah satunya adalah *k-fold cross validation*. Cara kerjanya adalah melakukan pengelompokan antara data latih dan data uji, kemudian dilakukan proses pengujian yang diulang sebanyak k kali. Hasil pengujian itu kemudian dirata-ratakan untuk menghasilkan sebuah nilai. Pengaruh nilai k yang terlalu kecil dapat menghasilkan akurasi yang rendah. Saat nilai k sedikit (2-5) maka ukuran sampel (*training*) menjadi lebih kecil, maka ada perbedaan karena ketidakstabilan set pelatihan itu sendiri, yang akan membawa peningkatan varian. *K-fold cross-validation* dengan nilai k sedang (10-20) dapat mengurangi varians ketika meningkatkan bias. (Kohavi & others, 1995).

Proses dari *the Naive Bayesian Classifier* atau *Simple Bayesian Classifier* melibatkan langkah-langkah berikut: (a) Mengumpulkan data pelatihan; (b) Preprocessing data; (c) Menghitung probabilitas dasar; (d) Menghitung probabilitas bersyarat; (e) Menghitung probabilitas posterior; (f) Mengklasifikasikan contoh baru; dan (f) Evaluasi dan pengujian (Han et al., 2012).

Pengujian menggunakan *RapidMiner* terdapat beberapa tahap, diantaranya adalah *import* data dari yang telah disediakan, memilih *file* dalam format *excel*, hasil dari visualisasi data yang telah diimport, penggabungan data dengan operasi yang dibutuhkan dalam penggunaan algoritma yang diinginkan hasil *running* algoritma (Darmawan et al., 2022).

Populasi dalam penelitian ini adalah siswa-siswa jurusan Teknik Komputer dan Jaringan dan Akuntansi SMK Bhakti Persada Kota Bekasi yang beralamat di Jl. Antara Dalam Kelurahan Jatimakmur Kecamatan Pondok Gede.

Data yang digunakan dalam penelitian ini merupakan data primer yang diambil dari data dapodik SMK Bhakti Persada dan telah di tambahkan atribut nilai test penjurusan, nilai raport (rata-rata nilai produktif), nilai UKK & *Try out* (2019) dan lama waktu pengerjaan UKK serta tahun

lulus. Atribut yang diambil dalam penelitian ini antara lain penghasilan orang tua, nilai test penjurusan, nilai rata-rata sample produktif semester I dan II, Nilai UKK & *Try out* (2019), dan waktu pengerjaan UKK yaitu: Cepat, Normal, dan Lambat untuk tahun akademik 2016/2017, 2017/2018 dan 2018/2019, seperti pada Tabel 1.

Tabel 1. Data Dapodik Tahun 2019

No	NIS	Nama	Gel. Masuk SMK	Asal SMP	Penghasilan Orangtua	Nilai Test Jurusan	Nilai Rata2 Produktif Semester I	Nilai Rata2 Produktif Semester II	Nilai UKK & Try out (2019)	Lama Waktu Pengerjaan	Tahun Lulus
1	141510003	ADHITYA DWI YANSYAH	1	BEKASI	1 juta - 2 juta/bulan	55	85	85.1	79	Lambat	2017
2	141510007	ALFAN ANDREANSAH	2	BEKASI	1 juta - 2 juta/bulan	75.5	89.5	90	98	Cepat	2017
3	141510012	ANDI BAHTIAR	1	BEKASI	1 juta - 2 juta/bulan	55	85	85.1	79	Lambat	2017
...
164	151610235	TIARA ANGGRAINI	2	JAKARTA	1 juta - 4 juta/bulan	63.5	86.5	86.5	91	NORMAL	2018
...
221	141510226	WULAN TRIDARA AMANDA	2	JAKARTA	1 juta - 4 juta/bulan	63.5	86.5	86.5	91	NORMAL	2017
222	141510229	YULIA AFTIANU NINGSIH	2	JAKARTA	1 juta - 4 juta/bulan	63.5	86.5	86.5	91	NORMAL	2017
...
279	161710125	YUYUN YUNENGSIH	2	KARAWANG	1 juta - 2 juta/bulan	75	89	89.5	87	?	2019

Sumber : Hasil Penelitian (2023)

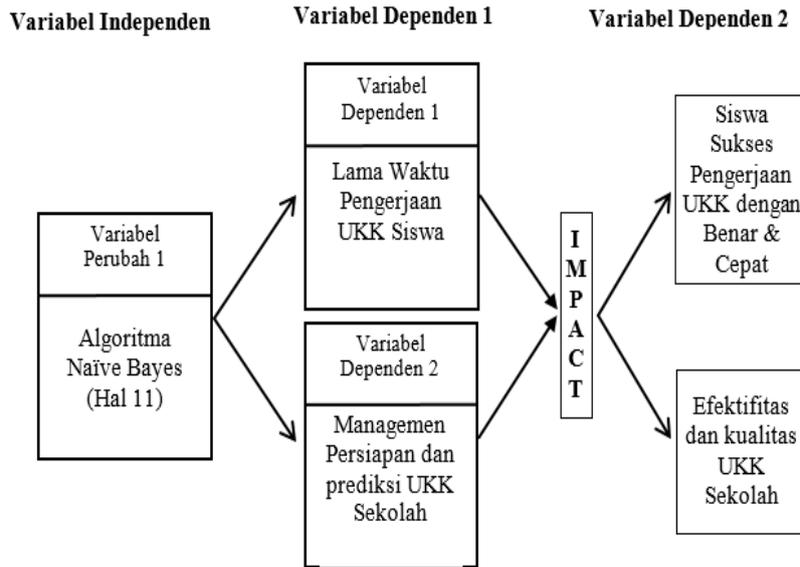
Adapun data yang diambil dalam penelitian ini adalah data Siswa lulusan tahun 2017 sebanyak 114 siswa dan 2018 sebanyak 108 siswa dengan jumlah data sebanyak 222 siswa (untuk pemodelan prediksi) dan 57 siswa yang akan lulus di tahun 2019 untuk prediksi sebenarnya, sehingga total keseluruhan sebanya 279 siswa. Untuk pemilihan sampel disesuaikan dengan aturan dari penggunaan metode algoritma *Naive Bayes*.

a. Data *Training* dan data Testing

Data dalam perhitungan manual, pada *dataset* untuk pemodelan sebanyak 222 *record*, dibagi menjadi 10 subset/split dan tiap subset akan dibagi 2 bagian dimana sebagian pertama digunakan sebagai data *training* dan dan sebagian sisanya sebagai data testing, dimana prosentase bagian pada 10 subset tersebut akan di lakukan otomatis oleh *Rapidminer*, dan akan dijadikan iterasi pemodelan dari masing-masing algoritma prediksi. *Dataset* ini akan digunakan juga untuk proses mining dan pengujian dimana atribut yang digunakan adalah Penghasilan Orang Tua, Nilai Test Penjurusan, Nilai Rata-rata Mapel Produktif Semester I, Nilai Rata-rata Mapel Produktif Semester II, Nilai UKK & *Try out* (2019), dan Lama Waktu Pengerjaan UKK.

b. Data Target

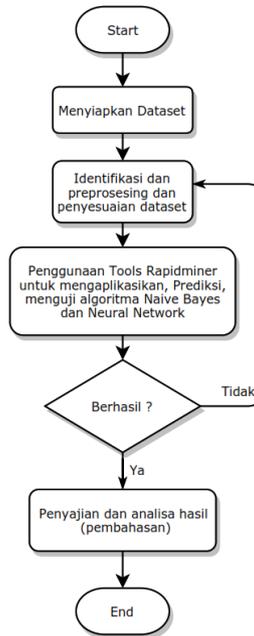
Data ini berupa sampel yang akan diujikan atau diprediksi kecepatan waktu pengerjaan test nya yang telah diambil modelnya dari proses *Data mining* kelulusan. Data ini adalah Siswa lulusan tahun 2018 (*test target*) dan 2019 (target sebenarnya). *Data mining* adalah proses yang melibatkan tahap interpretasi dan evaluasi untuk menghasilkan pengetahuan baru. Tujuan utamanya adalah memberikan kontribusi yang lebih baik dalam pemahaman dan penggunaan data yang ada (Fayyad et al., 1996).



Sumber : Hasil Penelitian (2023)

Gambar 1. Kerangka Konseptual

Pada Gambar 1 menggambarkan perihal variabel-variabel penelitian secara konseptual yang menjadi landasan atau dasar acuan penelitian. Dalam menyelesaikan penelitian dijelaskan pada Gambar 2 *flowchart* proses penyelesaian penelitian.

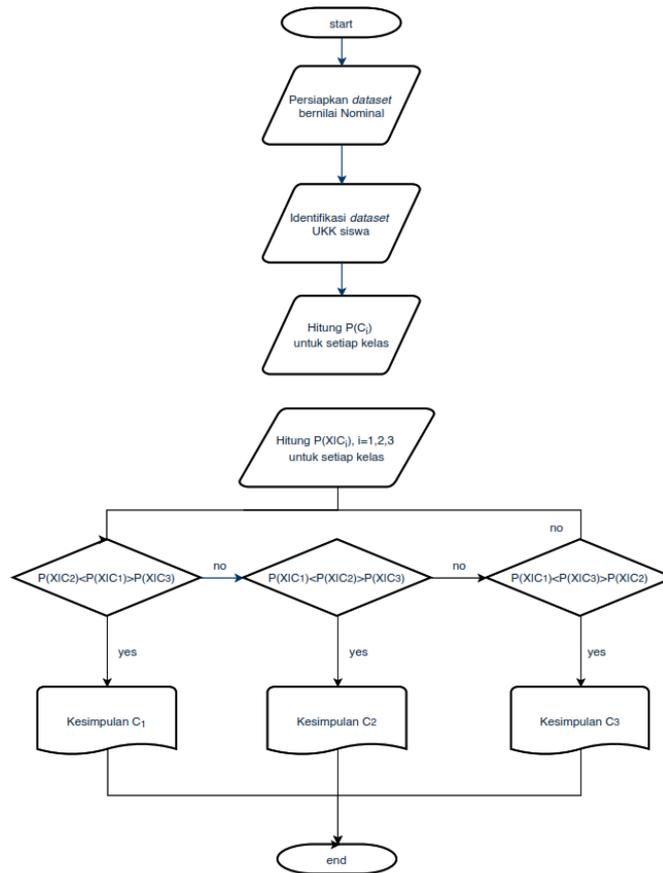


Sumber : Hasil Penelitian (2023)

Gambar 2. *Flowchart* Proses Penyelesaian Penelitian

Pada Gambar 3 tentang *flowchart* algoritma *Neural Network* yaitu, peneliti mulai dengan menyiapkan *dataset*, indentifikasi dan preprosesing *dataset*, membaca input, bobot, jumlah *layer*, *target error* dan *training cycle* yang diinginkan, hitung hasil bobot tiap fitur (*input*), nilai yang terbesar merupakan kesimpulan (kelas) dan selesai.

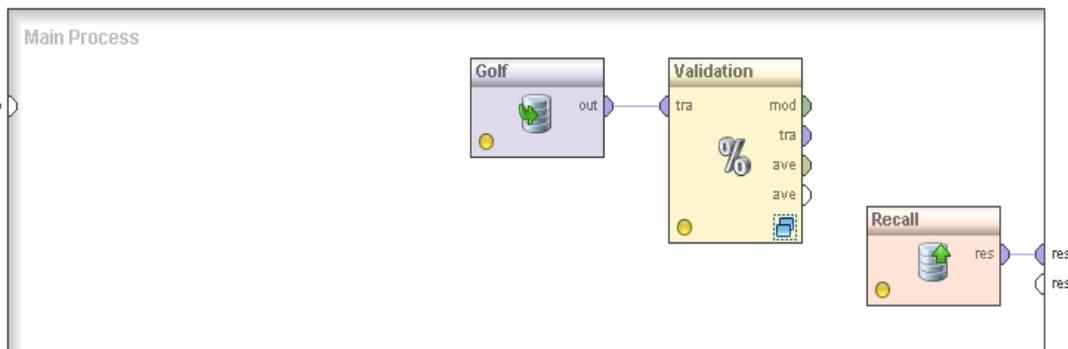
Algoritma Naïve Bayes untuk Memprediksi Waktu Pengerjaan Uji Kompetensi Keahlian (UKK) Siswa Sekolah Menengah Kejuruan



Sumber : Hasil Penelitian (2023)

Gambar 3. Flowchart Algoritma Naïve Bayes

Gambar 4 merupakan tampilan dari area kerja (*workspace*) *Rapidminer Studio*, Desainer visual alur kerja yang memberikan ilmu data dan mesin pembelajaran ke seluruh tim analitik, dari analis hingga pakar (Akthar et al., 2012).



Sumber : Hasil Penelitian (2023)

Gambar 4. Workspace Main Process *Rapidminer*

Untuk menyelesaikan penelitian digunakan *Rapidminer*, sebanyak 222 data siswa, dilakukan proses *cleaning* agar tidak terjadi kesalahan saat dilakukan pembelajaran oleh *Neural*

Network. Nilai pada atribut diubah dalam bentuk tipe data integer, seperti pada Tabel 3 konversi nilai atribut dan tipe data dengan hasil pada Tabel 4 Data siswa hasil konversi.

Tabel 3. Konversi Nilai Atribut dan Tipe Data

No	Atribut	Nilai	Konversi Angka	Tipe Data
1	Penghasilan Orang Tua	1 Juta - 2 Juta/bulan	1	Integer
		1 Juta - 3 Juta/bulan	2	Integer
		1 Juta - 4 Juta/bulan	3	Integer
		1 Juta - 5 Juta/bulan	4	Integer
2	Nilai Test Penjurusan	Real	Real	Real
3	Nilai Rata-rata Mapel Produktif Semester I	Real	Real	Real
4	Nilai Rata-rata Mapel Produktif Semester II	Real	Real	Real
5	Nilai UKK dan <i>Try out</i> (2019)	Real	Real	Real
6	Target (Lama Waktu Pengerjaan)	Cepat	Cepat	Polynomial
		Normal	Normal	
		Lambat	Lambat	

Sumber : Hasil Penelitian (2023)

Tabel 4. Data Siswa Hasil Konversi

No	NIS	Penghasilan Orangtua	Nilai Test Penjurusan	Nilai Rata2 Mapel Produktif Semester I	Nilai Rata2 Mapel Produktif Semester II	Nilai UKK & TryOut (2019)	Lama Waktu Pengerjaan
1	141510003	1	55	85	85.1	79	LAMBAT
2	141510007	1	75.5	89.5	90	98	CEPAT
3	141510012	1	55	85	85.1	79	LAMBAT
4	141510023	1	59.5	85.5	86	80	NORMAL
5	141510030	1	55	85	85.1	79	LAMBAT
6	141510031	1	59	85.5	85.5	79	NORMAL
7	141510033	1	59	85.5	85.5	79	NORMAL

Sumber : Hasil Penelitian (2023)

3. Hasil dan Pembahasan

Untuk mendapatkan hasil dari implementasi algoritma *Naïve Bayes* dan *Neural Network* dalam memprediksi lama waktu pengerjaan UKK siswa, maka dilakukan proses implementasi menggunakan *Rapidminer. Software* tersebut menjadi pilihan peneliti dikarenakan kemudahan dalam penggunaan dan *open source* digunakan sebatas penggunaan sampai 1000 *record dataset*.

Implementasi Algoritma *Naïve Bayes* untuk Pemodelan Prediksi.

Tahap-tahap proses algoritma *Naïve Bayes* dalam menghitung atau mencari nilai prediksi yang ingin diketahui.

- a. Tahap 1 menghitung jumlah *class* dari data *training* dan *testing*

$$P = (Y=Cepat) = 36/222 = 0,162$$

jumlah target lama waktu pengerjaan CEPAT dibagi dengan jumlah siswa atau baris atau *record* pada tabel.

$$P = (Y=Normal) = 140/222 = 0,631$$

jumlah target lama waktu pengerjaan NORMAL dibagi dengan jumlah siswa atau baris atau *record* pada tabel

$$P = (Y=Lambat) = 46/222 = 0,207$$

jumlah target lama waktu pengerjaan LAMBAT dibagi dengan jumlah siswa atau baris atau *record* pada tabel.

Selanjutnya, adalah perhitungan yang dilakukan pada *record* pertama saja dari 222 *record* pada *dataset* Data Training UKK.xlsx.

- b. Tahap 2 menghitung jumlah fitur atau kasus yang sama dengan *class* yang sama, dengan penjelasan:

$$P = (Y = \text{Nilai Test Penjurusan 55} \mid Y = \text{Cepat}) = 0/36$$

jumlah data Nilai Test Penjurusan 55 dengan lama waktu pengerjaan CEPAT dibagi dengan jumlah data dengan Lama Waktu Pengerjaan CEPAT.

$$P = (Y = \text{Nilai Test Penjurusan 55} \mid \text{Normal}) = 0/140$$

jumlah data Nilai Test Penjurusan 55 dengan lama waktu pengerjaan NORMAL dibagi dengan jumlah data dengan lama waktu pengerjaan NORMAL

$$P = (Y= \text{Nilai Test Penjurusan 55} \mid Y = \text{Lambat}) = 5/46$$

jumlah data Nilai Test Penjurusan 5 dengan lama waktu pengerjaan LAMBAT dibagi dengan jumlah data dengan lama waktu pengerjaan LAMBAT

Dilanjutkan dengan fitur-fitur kasus berikutnya yang di persingkat:

$$P = (Y = \text{Penghasilan Orang Tua 1 juta – 2 juta/bulan} \mid Y = \text{Cepat}) = 14/36$$

$$P = (Y = \text{Penghasilan Orang Tua 1 juta – 2 juta/bulan} \mid Y = \text{Normal}) = 56/140$$

$$P = (Y = \text{Penghasilan Orang Tua 1 juta – 2 juta/bulan} \mid Y = \text{Lambat}) = 26/46$$

$$P = (Y = \text{Nilai Rata-rata Mapel Produktif Semester I 85} \mid Y = \text{Cepat}) = 0/36$$

$$P = (Y = \text{Nilai Rata-rata Mapel Produktif Semester I 85} \mid Y = \text{Normal}) = 0/140$$

$$P = (Y = \text{Nilai Rata-rata Mapel Produktif Semester I 85} \mid Y = \text{Lambat}) = 46/46$$

$$P = (Y = \text{Nilai Rata-rata Mapel Produktif Semester II 85,1} \mid Y = \text{Cepat}) = 0/36$$

$$P = (Y = \text{Nilai Rata-rata Mapel Produktif Semester II 85,1} \mid Y = \text{Normal}) = 0/140$$

$$P = (Y = \text{Nilai Rata-rata Mapel Produktif Semester II 85,1} \mid Y = \text{Lambat}) = 8/46$$

$$P = (Y = \text{Nilai UKK \& Try out (2019) 79} \mid Y = \text{Cepat}) = 0/36$$

$$P = (Y = \text{Nilai UKK \& Try out (2019) 79} \mid Y = \text{Normal}) = 8/140$$

$$P = (Y = \text{Nilai UKK \& Try out (2019) 79} \mid Y = \text{Lambat}) = 3/46$$

- c. Tahap 3 kalikan semua hasil variabel dengan target CEPAT, NORMAL, dan LAMBAT

$$= 0/36 * 14/36 * 0/36 * 0/36 * 0/36 * 36/222$$

$$= 0 * 0,38 * 0 * 0 * 0 * 0,16 = 0$$

$$P = (P(\text{Nilai Test Penjurusan 55} \mid Y=Normal) * P(Y = \text{Penghasilan Orang Tua 1 juta – 2 juta/bulan} \mid Y = \text{Normal}) * P(\text{Nilai Rata-rata Mapel Produktif Semester I 85} \mid Y = \text{Normal}) *$$

$$P(\text{Nilai Rata-rata Mapel Produktif Semester II } 85,1 \mid Y = \text{Normal}) * P(Y = \text{Nilai UKK \& Try out (2019) } 79 \mid Y = \text{Normal}) * P(Y=\text{Normal})$$

$$= 0/140 * 56/140 * 0/140 * 0/140 * 8/140 * 140/222$$

$$= 0 * 0.40 * 0 * 0 * 0.05 * 0.16 = 0$$

$$P = (P(\text{Nilai Test Penjurusan } 55 \mid Y=\text{Lambat}) * P(Y = \text{Penghasilan Orang Tua } 1 \text{ juta} - 2 \text{ juta/bulan} \mid Y = \text{Lambat}) * P(\text{Nilai Rata-rata Mapel Produktif Semester I } 85 \mid Y = \text{Lambat}) * P(\text{Nilai Rata-rata Mapel Produktif Semester II } 85,1 \mid Y = \text{Lambat}) * P(Y = \text{Nilai UKK \& Try out (2019) } 79 \mid Y = \text{Lambat}) * P(Y=\text{Lambat})$$

$$= 5/46 * 26/46 * 46/46 * 8/46 * 3/46 * 46/222$$

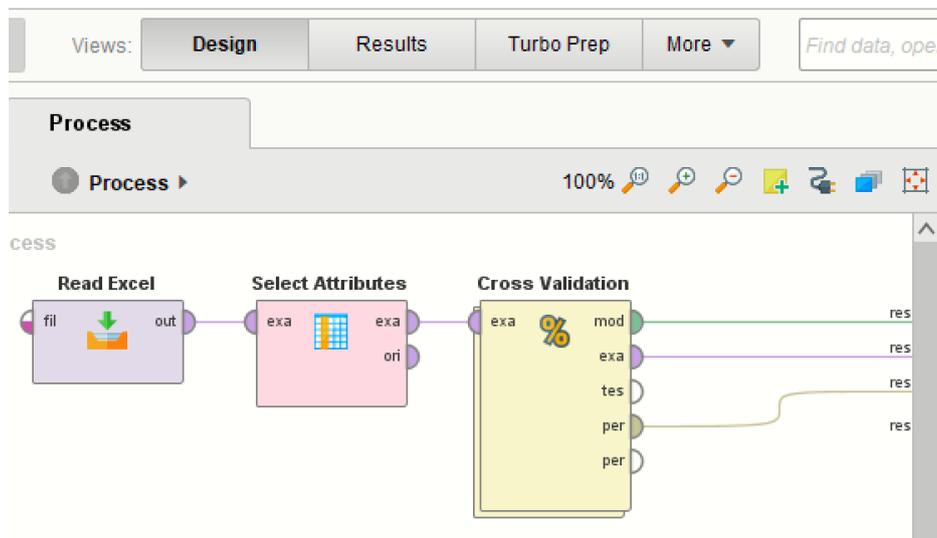
$$= 0.10 * 0.56 * 1 * 0.17 * 0.06 * 0.20 = 0.0001$$

$$P = (P(\text{Nilai Test Penjurusan } 55 \mid Y=\text{Cepat}) * P(Y = \text{Penghasilan Orang Tua } 1 \text{ juta} - 2 \text{ juta/bulan} \mid Y = \text{Cepat}) * P(\text{Nilai Rata-rata Mapel Produktif Semester I } 85 \mid Y = \text{Cepat}) * P(\text{Nilai Rata-rata Mapel Produktif Semester II } 85,1 \mid Y = \text{Cepat}) * P(Y = \text{Nilai UKK \& Try out (2019) } 79 \mid Y = \text{Cepat}) * P(Y=\text{Cepat})$$

d. Tahap 4 bandingkan kelas CEPAT, NORMAL dan LAMBAT

Hasil $(P|\text{Cepat}) = 0$; $(P|\text{Normal}) = 0$; $(P|\text{Lambat}) = 0,0001$, maka karena hasil kelas pada lama waktu pengerjaan Lambat lebih besar dibandingkan dengan Cepat dan Normal untuk prediksi lama waktu pada (contoh *record* pertama) *dataset* kelulusan UKK Siswa 2018 dengan Nilai Test Penjurusan = 55, Nilai Rata-rata Mapel Produktif Semester I = 85, Penghasilan orang tua 1 juta – 2 juta/bulan dan Nilai UKK & Try out (2019) = 79 diprediksikan lulus dengan Lama Waktu Pengerjaan Lambat.

Rangkaian proses selanjutnya adalah melakukan implementasi algoritma *Naïve Bayes* tersebut menggunakan *Rapidminer*.

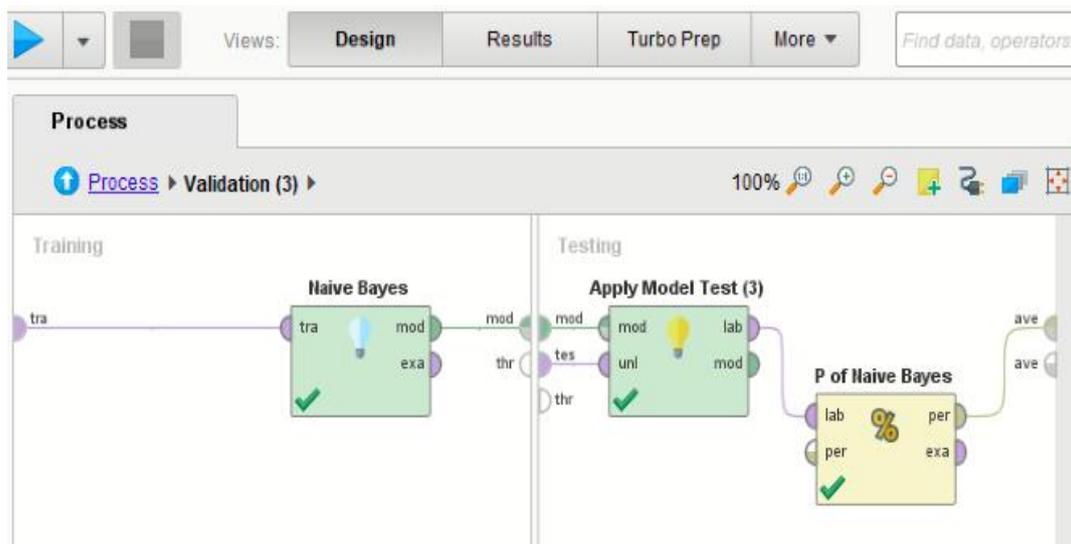


Sumber : Hasil Penelitian (2023)

Gambar 5. *Process Naïve Bayes* pada *Rapidminer*

Gambar 5 *design process Naïve Bayes* pada *Rapidminer* dengan *Dataset Data Training UKK Update.xlsx*, merupakan tampilan *workspace* pada tahap pertama (proses) dari *Rapidminer*

untuk melakukan proses prediksi dari file *Dataset Data Training UKK Update.xlsx* menggunakan algoritma *Naïve Bayes*.



Sumber : Hasil Penelitian (2023)

Gambar 6. *Design Validation Naïve Bayes* pada *Rapidminer*

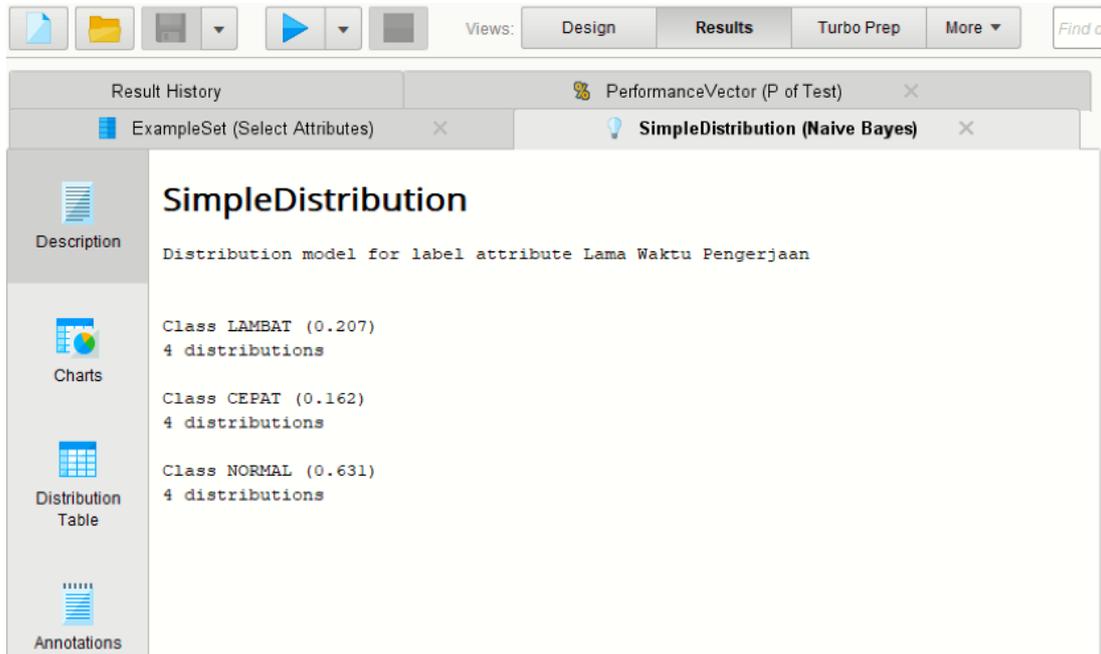
Gambar 6 design validation *Naïve Bayes* pada *Rapidminer* dengan *Dataset Data Training UKK Update.xlsx*, merupakan tampilan *workspace* pada tahap berikutnya (*validation*) dari *Rapidminer* untuk melakukan proses prediksi file *Dataset Data Training UKK Update.xlsx* menggunakan algoritma *Naïve Bayes*.

Row No.	Lama Wakt...	prediction(L...	confidence(L...	confidence(...	confidence(...	Penghasian...	Nilai Test Pe...	Nilai Rata2 ...	Nilai...
1	CEPAT	CEPAT	1	0	0	1	75.500	89.500	90
2	LAMBAT	LAMBAT	0	0.000	1.000	1	55	85	85
3	NORMAL	NORMAL	0	1	0	1	59	85.500	85.5
4	NORMAL	NORMAL	0	1	0	4	59	85.500	85.5
5	NORMAL	NORMAL	0	1	0	1	59	85.500	85.5
6	LAMBAT	LAMBAT	0	0.000	1.000	1	54.500	85	85
7	LAMBAT	LAMBAT	0	0.000	1.000	1	54.500	85	85
8	CEPAT	CEPAT	1	0	0	1	75.500	89.500	90
9	NORMAL	NORMAL	0.000	1.000	0	1	63.500	86.500	86.5
10	NORMAL	NORMAL	0.000	1.000	0	1	85	87	87
11	LAMBAT	LAMBAT	0	0.000	1.000	1	57.500	85	85.5
12	NORMAL	NORMAL	0.000	1.000	0	1	85	87	87

Sumber : Hasil Penelitian (2023)

Gambar 7. Tampilan *Tab Results*

Gambar 7 merupakan tampilan *workspace* pada tahap *Results* dari *Rapidminer* setelah dilakukan proses *running* (tombol *Play*) untuk melakukan proses prediksi dari file *Dataset Data Training UKK Update.xlsx* menggunakan algoritma *Naïve Bayes*.

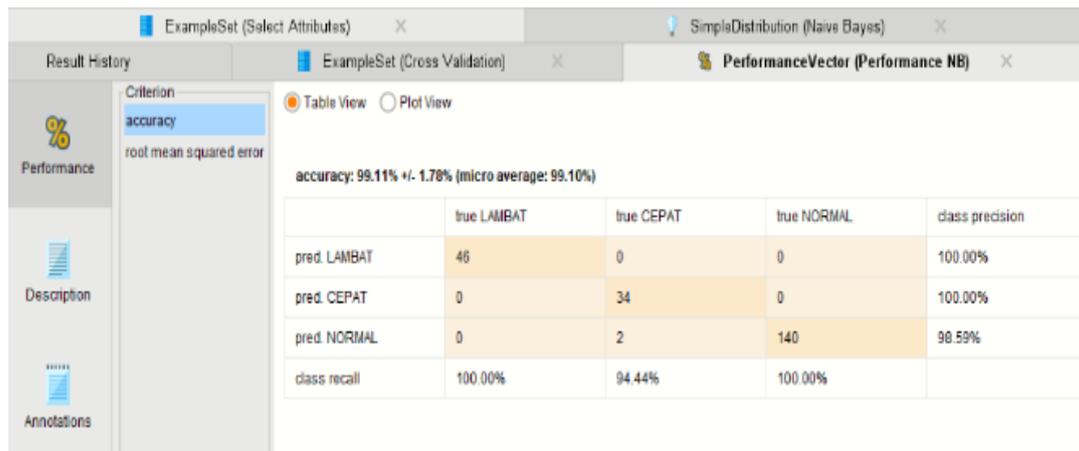


Sumber : Hasil Penelitian (2023)

Gambar 8. Tampilan *SimpleDistribution (Naïve Bayes) sub Description*

Gambar 8 merupakan tampilan *workspace* pada *Tab Result (SimpleDistribution)* dari *Rapidminer* setelah dilakukan proses *running* (tombol *Play*) untuk melakukan proses prediksi dari file *Dataset Data Training UKK Update.xlsx* menggunakan algoritma *Naïve Bayes*.

Distribusi 3 kelas yang dijadikan target atau label yaitu : Lambat = 0,2017 ; Cepat : 0,163 ; Normal : 0,631.

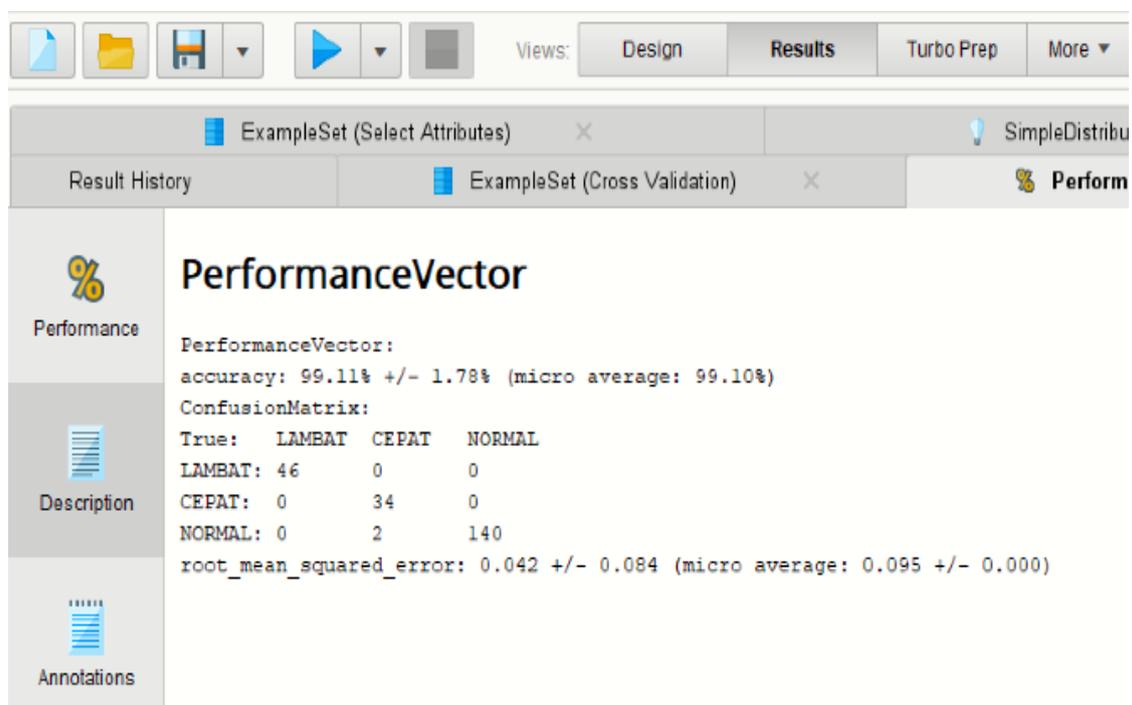


Sumber : Hasil Penelitian (2023)

Gambar 9 Tampilan *PerformanceVector sub Accuracy*

Gambar 9 merupakan tampilan *workspace* pada *Tab Result (PerformanceVector)* dari *Rapidminer* setelah dilakukan proses *running* (tombol *Play*) untuk melakukan proses prediksi dari file *Dataset Data Training UKK Update.xlsx* menggunakan algoritma *Naïve Bayes*.

Untuk tingkat akurasi prediksi sebagaimana hasil perhitungan menggunakan rumus 1 dan rumus 2 sebesar 99,11% dengan rincian (a) Kebenaran Lambat = 46, di Prediksi Lambat = 46, di Prediksi Cepat = 0, di Prediksi Normal = 0, nilai prosentase akurasi pada kelas Lambat adalah 100,00%; (b) Kebenaran Cepat = 36, di Prediksi Lambat = 0, di Prediksi Cepat = 34, di Prediksi Normal = 0, nilai prosentase akurasi pada kelas Cepat adalah 94,44%; dan (c) Kebenaran Normal = 140, di Prediksi Lambat = 0, di Prediksi Cepat = 0, di Prediksi Normal = 140, nilai prosentase akurasi pada kelas Normal adalah 100,00% .



Sumber : Hasil Penelitian (2023)

Gambar.10 Tampilan PerformanceVector sub Description

Gambar 10 merupakan tampilan *workspace* pada *Tab Result PerformanceVector (sub Description)* dari *Rapidminer* setelah dilakukan proses *running* (tombol *Play*) untuk melakukan proses prediksi dari file *Dataset Data Training UKK Update.xlsx* menggunakan algoritma *Naïve Bayes*.

Dari tampilan hasil pemrosesan aplikasi, hasilnya dapat di baca pada *Description (PerformanceVector)*, yaitu :

accuracy: 99.11% +/- 1.78% (micro average :99.10%)

root_mean_squared_error: 0.042 +/- 0.084 (micro average :0.095.10 +/- 0.000)

Hasil akurasi tersebut hampir mendekati 100%, yaitu dengan nilai 99,11%, dengan tingkat kesalahan atau *Root Mean Squarred Error (RMSE)* yang didapat sebesar 0,042

5758ccc77fb7929df47a9

- Febriya, F., Citra, R. S., Sumarni, T., & Mutoffar, M. M. (2022). Klasifikasi Ujaran Kebencian Dengan Naive Bayes Classification pada Page Meme Comic Indonesia. *Seminar Nasional CORIS 2022*. <https://corisindo.stikom-bali.ac.id/penelitian/index.php/semnas/article/download/80/61>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques Third Edition. University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University*.
https://www.academia.edu/download/43034828/Data_Mining_Concepts_And_Techniques_3rd_Edition.pdf
- Kohavi, R., & others. (1995). A Study of cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Ijcai*, 14(2), 1137–1145. https://www.researchgate.net/profile/Ron-Kohavi/publication/2352264_A_Study_of_Cross-Validation_and_Bootstrap_for_Accuracy_Estimation_and_Model_Selection/links/02e7e51bcc14c5e91c000000/A-Study-of-Cross-Validation-and-Bootstrap-for-Accuracy-Estimation-and-
- Larose, D. T. (2005). *Discovering Knowledge in Data. New Jersey: John Willey & Sons. Inc.*
- Liang, T. P., Turban, E., & Aronson, J. E. (2005). *Decision support systems and intelligent systems. Yogyakarta: Penerbit Andi.*
- Muchammad, A. (2016). *Implementasi Algoritma Naive Bayes untuk Menentukan Kesiapan Siswa Dalam Menghadapi Ujian Nasional*.
http://eprints.dinus.ac.id/18177/2/jurnal_17690.pdf
- Olson, D. L., & Delen, D. (2008). *Advanced Data Mining Techniques*. Springer Science & Business Media. https://www.researchgate.net/profile/David-Olson-21/publication/220695151_Advanced_Data_Mining_Techniques/links/00b4952b07ee813b62000000/Advanced-Data-Mining-Techniques.pdf
- Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab. Yogyakarta: Andi*, 1. @inproceedings%7Bfebriya2022klasifikasi,%0A title=%7BKlasifikasi Ujaran Kebencian Dengan Naive Bayes Classification pada Page Meme Comic Indonesia%7D,%0A author=%7BFebriya, Fadly and Citra, Regina Sukma and Sumarni, Tarsinah and Mutoffar, Muhammad Malik%7D,%0A booktitle=%7BSeminar Nasional CORIS 2022%7D,%0A year=%7B2022%7D%0A%7D
- Rinaldi, M. (2011). *Algoritma dan Pemrograman Dalam Bahasa Pascal dan C. Bandung: Penerbit Informatika.*
- Robbins, S. P. (2006). *Perilaku Organisasi, Edisi Bahasa Indonesia. Jakarta: PT. Indeks Kelompok Gramedia.*
- Siregar, A. M., Puspabhuana, A. (2017). *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group.
- Surjono, H. D. (2008). *Landasan Pemanfaatan Teknologi Informasi Untuk Uji Kompetensi TIK di*

- SMA. *Seminar Nasional Pemanfaatan Teknologi Informasi dan Komunikasi (TIK) Untuk Ujian Online di Sekolah Menengah Atas*.
<http://staffnew.uny.ac.id/upload/131666733/penelitian/13-landasanpemanfaatanti.pdf>
- Utomo, W. (2015). Prediksi Nilai Ujian Nasional Produktif Sekolah Menengah Kejuruan Menggunakan Metode *Neural Network*. *Techno. Com*, 14(1), 33–41.
- Vercellis, C. (2011). *Business intelligence: Data Mining And Optimization For Decision Making*. John Wiley & Sons. <https://www.scinapse.io/papers/599503780>
- Yunianto, I., Kurniawan, A., & Mutoffar, M. M. (2022). Comparison of Decision Tree, KNN and Naive Bayes Methods In Predicting Student Late Graduation In the Informatics Engineering Department, Institute Business XYZ. *Adpebi International Journal of Multidisciplinary Sciences*, 1(1), 374–383.
<https://journal.adpebi.com/index.php/AIJMS/article/download/304/387>