

OPTIMASI ALGORITMA NAIVE BAYES MENGGUNAKAN METODE CROSS VALIDATION UNTUK MENINGKATKAN AKURASI PREDIKSI TINGKAT KELULUSAN TEPAT WAKTU

Yohakim Benedictus Samponu¹⁾, Kusri²⁾

^{1,2)} Magister Teknik Informatika Universitas Amikom Yogyakarta
e-mail: mikahoy045@gmail.com¹⁾, kusri@amikom.ac.id²⁾

ABSTRACT

Education at this time is an important requirement in facing the demands of an increasingly advanced era in technology. To compensate this, the existing educational standards in universities must also be improved, this is a bit much affect the pattern of teaching from universities that produce qualified graduates who can compete in the world of work later and indirectly give a positive impact on the university itself. Qualified graduates are of course not only depending on the role of a university but also majors and quality of education as long as students are still in high school / vocational school also plays an important role. Results of the on-time graduation rate prediction research can be used as an information to improve the quality and optimization of the education system but it requires a maximum degree of accuracy. This research predicts on time graduation rates by conducting analysis using data mining classification techniques. Naïve Bayes algorithm that are used for this research will be discussed as a reference in conducting research. The author performs a series of different experimental scenarios / cross validation to perform comparisons that can give a difference in the level of accuracy gained from this research. The results of this research indicate that with the addition of Cross Validation testing scenario there is an increase of 2% accuracy of the test.

Keywords: data mining, Cross Validation, Naïve Bayes.

ABSTRAK

Pendidikan pada saat ini sudah merupakan suatu kebutuhan penting dalam menghadapi tuntutan jaman yang semakin maju dalam teknologi. Untuk mengimbangi hal tersebut standar pendidikan yang ada di perguruan tinggi pun harus ditingkatkan, hal ini sedikit banyak mempengaruhi pola pengajaran dari perguruan tinggi yang menghasilkan lulusan yang berkualitas yang mampu bersaing dalam dunia kerja nantinya dan secara tidak langsung memberikan dampak yang positif terhadap kampus tersebut. Lulusan yang berkualitas tentu saja tidak hanya bergantung dari peran perguruan tinggi semata, jurusan dan kualitas pendidikan selama mahasiswa masih berada di SMA / SMK juga memegang peranan penting. Hasil penelitian prediksi tingkat kelulusan tepat waktu dapat menjadi masukan untuk meningkatkan kualitas dan optimasi sistem pendidikan tetapi tentu saja penelitian ini membutuhkan tingkat akurasi yang maksimal. Penelitian ini melakukan prediksi tingkat kelulusan tepat waktu dengan melakukan analisis menggunakan teknik klasifikasi data mining. Algoritma Naïve Bayes yang digunakan untuk penelitian ini akan dibahas juga sebagai acuan dalam melakukan penelitian. Penulis melakukan serangkaian skenario percobaan yang berbeda / cross validation untuk melakukan perbandingan yang dapat memberikan perbedaan dalam tingkat akurasi yang dihasilkan dari penelitian ini. Hasil penelitian ini menunjukkan dengan penambahan skenario pengujian Cross Validation terdapat peningkatan akurasi sebesar 2% dari pengujian yang dilakukan.

Kata Kunci: data mining, Cross Validation, Naïve Bayes.

I. PENDAHULUAN

Universitas Sam Ratulangi merupakan salah satu Perguruan Tinggi yang banyak diminati di daerah Sulawesi Utara. Ada 11 Fakultas di Perguruan Tinggi ini, dan Fakultas Teknik merupakan salah satu fakultas yang menarik banyak minat mahasiswa baru. Masa studi di salah satu Program Studi Fakultas Teknik khususnya Program Studi Informatika dapat ditempuh dalam waktu 4 tahun (8 semester). Tidak seluruh mahasiswa dapat lulus tepat waktu karena berbagai macam faktor

sedangkan tingkat kelulusan tepat waktu ini berpengaruh tidak hanya untuk mahasiswa itu sendiri saja, tapi dapat mempengaruhi akreditasi dari perguruan tinggi tersebut.

Berdasarkan data dari jurusan Teknik Elektro yang merupakan salah satu jurusan yang berada di bawah Fakultas Teknik Universitas Sam Ratulangi (<http://fatek.unsrat.ac.id/elektro/index.php/data-jurusan/alumni>) bisa diketahui bahwa lama studi dari mahasiswa sebagian besar berada di atas 4 Tahun. Data lama studi mahasiswa dapat diperhatikan pada Tabel 1.

TABEL 1
LAMA STUDI ALUMNI JURUSAN ELEKTRO FAKULTAS TEKNIK UNSRAT

Lama Studi	Presentasi Kelulusan
9 Tahun	2%
8 Tahun	13%
7 Tahun	17%
6 Tahun	28,5%
5 Tahun	28,5%
4 Tahun	11%

Dari Tabel 1 dapat dilihat dengan jelas bahwa presentasi tingkat kelulusan tidak tepat waktu lebih tinggi dari pada yang lulus tepat waktu yang selanjutnya bisa dijadikan *target variable*, maka untuk melakukan prediksi tingkat kelulusan teknik data mining yang digunakan adalah Klasifikasi. Pada dasarnya klasifikasi adalah teknik mempelajari data sehingga menghasilkan aturan yang dapat memprediksi data baru. Dengan semakin majunya teknologi prediksi bisa dilakukan dengan berbagai algoritma sesuai dengan kebutuhan. Adapun algoritma yang dapat digunakan adalah Naïve Bayes.

Walaupun sudah banyak penelitian yang dilakukan dengan metode Naïve Bayes, tapi belum ada penelitian yang mengulas tentang prediksi tingkat kelulusan mahasiswa menggunakan metode Naïve Bayes dan *Cross Validation*. Penelitian terdahulu yang menggunakan algoritma Naïve Bayes antara lain: Mengetahui minat beli pelanggan terhadap kartu internet XL dengan metode Naïve Bayes [9], Prediksi besaran listrik rumah tangga menggunakan metode Naïve Bayes [10], Klasifikasi Status Gizi menggunakan metode Naïve Bayes [6]

Penelitian di atas menggunakan algoritma Naïve Bayes untuk melakukan prediksi. Algoritma Naïve Bayes sangat menarik untuk dibahas karena dari penelitian sebelumnya algoritma ini mampu memanfaatkan data training untuk menghasilkan probabilitas setiap kriteria untuk kelas yang berbeda seperti melakukan prediksi minat beli pelanggan terhadap kartu internet yang baru akan diluncurkan dengan melakukan prediksi berdasarkan kartu internet yang sudah beredar sebelumnya.

Penelitian ini memilih algoritma Naïve Bayes karena berdasarkan kesimpulan beberapa penelitian terdahulu algoritma ini memiliki tingkat akurasi yang tinggi yang kemudian bisa ditingkatkan lagi tingkat akurasinya dengan menambahkan metode *cross validation*.

II. RUMUSAN MASALAH

Sebagaimana sudah dikemukakan sebelumnya pada latar belakang, dengan teknologi sekarang maka memungkinkan untuk mengatasi / melakukan prediksi terhadap tingkat kelulusan mahasiswa tepat waktu yang tentunya secara tidak langsung dapat menjadi masukan untuk mengoptimalkan sistem pengajaran dari perguruan tinggi khususnya Fakultas Teknik Universitas Sam Ratulangi Manado.

Masalah yang akan diselesaikan melalui penelitian ini yaitu Seberapa besar peningkatan akurasi prediksi tingkat kelulusan tepat waktu antara data pengujian menggunakan Algoritma Naïve Bayes yang ditambahkan Metode *Cross Validation*?

III. BATASAN MASALAH

Batasan masalah dalam melakukan prediksi kelulusan mahasiswa tepat waktu dalam penelitian ini hanya terbatas pada :

- Prediksi tingkat kelulusan tepat waktu menggunakan Algoritma Naïve Bayes dengan menambahkan metode *cross validation*.
- Data training yang digunakan dalam penelitian ini menggunakan data simulasi dengan ketentuan total data lulusan tahun 2012 dengan 50% data testing dari total data.

- c. Untuk algoritma dilakukan testing terlebih dahulu. Pengetesan dilakukan dengan beberapa skenario yang berbeda menggunakan metode *Cross Validation*
- d. Tools yang digunakan untuk melakukan pengujian / penelitian ini menggunakan bahasa pemrograman yang sudah ada, yaitu bahasa pemrograman R.

IV. TUJUAN PENELITIAN

Adapun yang tujuan yang ingin dicapai dalam penelitian klasifikasi tingkat kelulusan tepat waktu dengan metode Naïve Bayes agar dapat mengetahui tingkat akurasi algoritma naive bayes dengan penambahan metode cross validation.

V. MANFAAT PENELITIAN

Manfaat dari penelitian diharapkan dapat :

- a. Memberikan kontribusi dalam bidang ilmu pengetahuan khususnya dalam pengembangan bidang data mining terutama dalam mengatasi masalah optimasi sistem pendidikan.
- b. Penelitian ini dapat menjadi masukan dalam penyusunan silabus bagi perguruan tinggi yang dapat membantu proses optimasi sistem pendidikan.
- c. Diharapkan dengan implementasi hasil penelitian ini dapat meningkatkan kuantitas lulusan tepat waktu dari Fakultas Teknik Universitas Sam Ratulangi tanpa mengurangi kualitasnya yang secara tidak langsung dapat meningkatkan akreditasi.

VI. TINJAUAN PUSTAKA

Dalam melakukan penelitian ini, penulis mengumpulkan referensi tentang penelitian terdahulu yang dijadikan patokan dalam menyelesaikan masalah ini.

Teknik klasifikasi data mining bukanlah hal yang baru karena sebelumnya sudah ada beberapa penelitian yang menggunakan algoritma klasifikasi data mining lengkap dengan studi kasus antara lain:

- 1) Penelitian yang menggunakan teknik klasifikasi data Naïve Bayes. Penelitian ini menggunakan teknik klasifikasi data mining menggunakan algoritma Naïve Bayes untuk mengetahui minat beli pelanggan kartu internet XL berdasarkan kartu internet yang sebelumnya sudah beredar di pasaran. Penelitian ini hanya fokus pada 1 algoritma saja sehingga tidak dapat mendapatkan hasil optimal dengan cara membandingkannya dengan algoritma yang lain. [9]
- 2) Penelitian yang menggunakan teknik klasifikasi untuk melakukan prediksi penggunaan listrik rumah tangga masih sama menggunakan algoritma Naïve Bayes. Penelitian ini hanya fokus pada 1 algoritma saja sehingga tidak dapat mendapatkan hasil optimal dengan cara membandingkannya dengan algoritma yang lain. [10]
- 3) Penggunaan algoritma Naïve Bayes untuk klasifikasi bakteri gram negatif yaitu berupa bakteri e-coli yang berperan penting dalam penentuan obat yang akan diberikan pada pasien yang terinfeksi nantinya. Penelitian ini hanya fokus pada 1 algoritma saja sehingga tidak dapat mendapatkan hasil optimal dengan cara membandingkannya dengan algoritma yang lain. [8]

Penelitian dilakukan dengan melakukan pengembangan dari referensi tersebut di atas.

VII. DASAR TEORI

Data mining adalah proses seleksi data berdasarkan data yang sebelumnya sudah ada dengan cara memahaminya lebih dahulu dan melakukan pemrosesan awal untuk meningkatkan kualitas datanya. ada yang akan menjelaskan tentang penelitian terdahulu dan keaslian dari penelitian ini. Data Mining didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan tren baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik Statistik dan Matematika [2].

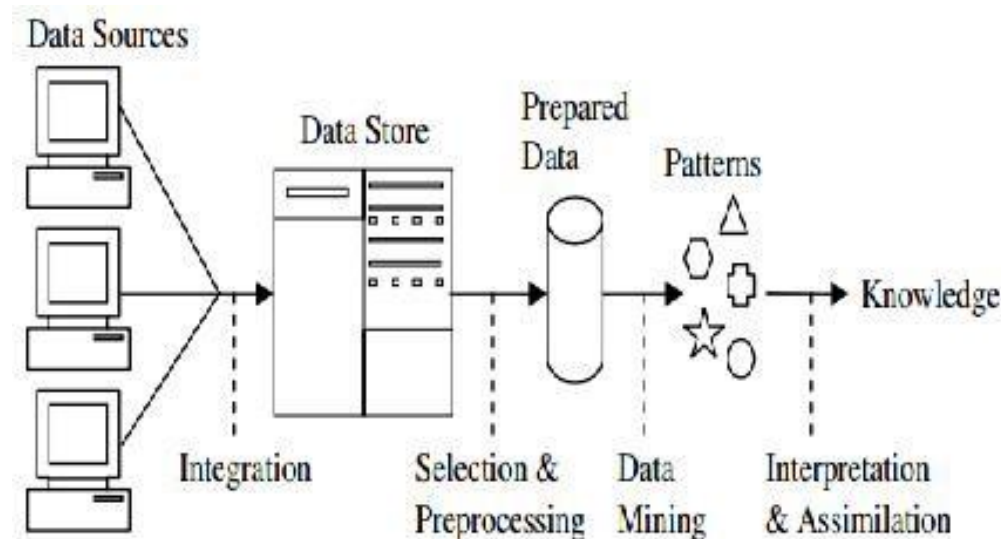
Berdasarkan fungsionalitasnya, tugas-tugas data mining bisa dikelompokkan ke dalam 6 kelompok berikut ini [3] :

- 1) Klasifikasi
- 2) Klasterisasi
- 3) Regresi
- 4) Deteksi Anomali
- 5) Pembelajaran Aturan Asosiasi atau Pemodelan Ketergantungan
- 6) Perangkuman

Teknik klasifikasi data mining merupakan proses seleksi data berdasarkan data yang sebelumnya sudah ada dengan cara memahaminya lebih dahulu dan melakukan pemrosesan awal untuk meningkatkan kualitas datanya. Data mining diperlukan mengingat dengan semakin berkembangnya teknologi saat ini yang menyediakan data yang sangat besar (diakibatkan juga karena biaya storage / ruang penyimpanan data yang semakin murah) maka akan sangat sulit jika pemrosesan suatu data dilakukan secara manual.

Data mining diaplikasikan juga dalam bidang Marketing dan Bisnis, Sains dan Teknik, dan bidang Seni dan Hiburan.

Proses Knowledge Discovery in Database (KDD) dapat digambarkan pada Gambar 2 sebagai berikut [2] :



Gambar 1. Knowledge Discovery in Database

Pada dasarnya klasifikasi adalah teknik mempelajari data sehingga menghasilkan aturan yang dapat memprediksi data baru. Teknik ini tergolong dalam Supervised Learning karena pembelajaran ini telah mengetahui target variabel yang akan diprediksi. Teknik lainnya adalah Unsupervised Learning dimana dengan cara ini lebih mengutamakan menemukan pola yang bisa mengidentifikasi data yang dipilih dalam dataset.

Sebelum melakukan klasifikasi data, dilakukan pembersihan data / data cleaning. Data cleaning dimaksudkan di sini agar teknik data mining bisa bekerja secara optimal hingga memberikan hasil yang paling baik. Sebuah data dikatakan tidak bersih jika mengandung kotoran yang berupa nilai kosong dan / atau derau dan / atau pencilan dan / atau inkonsistensi [3]. Data yang kurang baik ini tentu saja akan memberikan hasil yang kurang baik. Proses pembersihan data dilakukan dengan memperbaiki nilai kosong, menghaluskan data yang berderau (himpunan data berupa kesalahan variasi bersifat acak), membuang pencilan, dan memperbaiki inkonsistensi data yang disebabkan oleh beberapa faktor antara lain kurang bagusnya desain formulir pemasukan data, dll.

Algoritma yang akan digunakan dalam penelitian ini adalah algoritma yang tergolong Klasifikasi, yaitu :

Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah algoritma yang melakukan klasifikasi dengan cara menghitung peluang berdasarkan rumus Bayesian Rule berikut :

$$P(C | X) = (P(X | C) \times P(C)) / (P(X)) \tag{1}$$

Dimana algoritma ini bisa digunakan untuk memecahkan masalah klasifikasi data dengan feature bernilai nominal ataupun numerik. Jenis Naïve Bayes untuk berbagai tipe data adalah sebagai berikut :

- 1) Gaussian Naïve Bayes.
- 2) Bernaulli Naïve Bayes.
- 3) Multinomial Naïve Bayes

VIII. ANALISIS DATA

Untuk analisis data penelitian ini menggunakan metode CRISP-DM (Cross-Industry Standard Process for Data Mining) dikembangkan tahun 1996 menyediakan standar proses dalam pemecahan masalah untuk data mining melalui 6 Fase [2]:

- 1) Fase Pemahaman Bisnis (Business Understanding Phase)
- 2) Fase Pemahaman Data (Data Understanding Phase)
- 3) Fase Pengolahan Data (Data Preparation Phase)
- 4) Fase Pemodelan (Modelling Phase)
- 5) Fase Evaluasi (Evaluation Phase)
- 6) Fase Penyebaran (Deployment Phase)

Tools yang digunakan adalah bahasa pemrograman R yang juga merupakan environment untuk komputasi statistik. Tool yang tersedia dalam R ada bermacam-macam dari tool statistik linier dan non linier. R juga menyediakan tool grafis memvisualisasikan data yang diolah dalam bentuk grafik.

A. Metode Pengukuran

Algoritma klasifikasi yang digunakan adalah dengan menggunakan fungsi algoritma. Sebagai contoh data training yang disimpan dalam objek data_training (Tabel 2) dan data_testing (Tabel 3) adalah sebagai berikut :

TABEL 2
OBJEK DATA_TRAINING

	Feature 1	Feature 2	Target Variabel
1	1.2	2.1	Class1
2	4.1	2.2	Class2
3	3.2	2.0	Class1

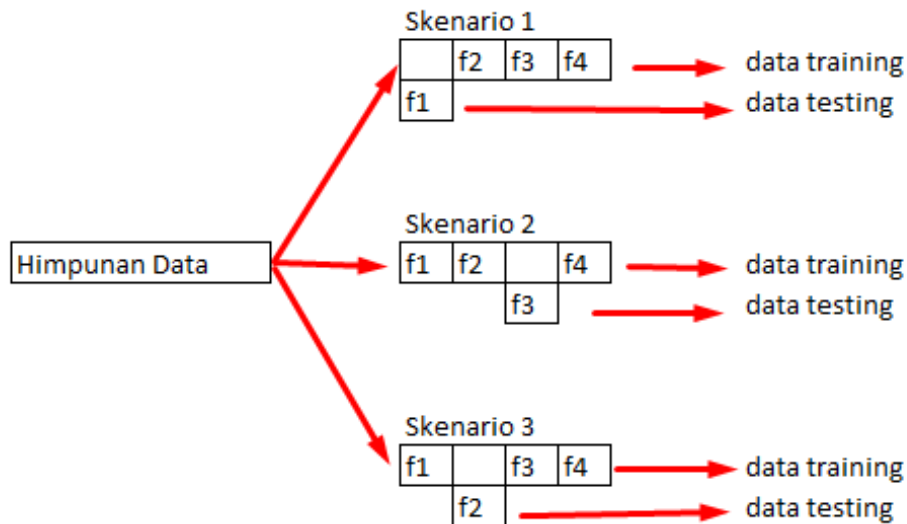
TABEL 3
OBJEK DATA_TESTING

	Feature 1	Feature 2	Target Variabel
1	2.7	3.0	Class1
2	2.2	3.3	Class2

Dengan menggunakan fungsi algoritma klasifikasi `algorithm()` maka contoh pseudocode implementasi menggunakan fungsi algoritma di atas adalah :

$$Prediction = algorithm(data_training, data_testing) \tag{2}$$

Untuk pengembangan dari tool yang sudah ada dan untuk meningkatkan nilai akurasi, prediksi juga akan dilakukan bukan hanya dengan mengandalkan algoritma / library yang ada tetapi juga akan dimaksimalkan dengan menggunakan Cross Validation, dimana untuk parameternya akan diuji lebih dari sekali, sehingga untuk setiap himpunan data bagian fold yang diuji akan berbeda-beda. Untuk lebih jelasnya dapat dilihat pada Gambar 4.



Gambar 2. Skenario Pengujian Cross Validation

Setelah didapat hasil prediksi algoritma klasifikasi kemudian dibandingkan dengan data aktual. Untuk lebih jelasnya ditunjukkan pada Tabel 4

TABEL 4
PERBANDINGAN NILAI PREDIKSI DAN DATA AKTUAL

	Prediksi	Data_aktual
1	Class1	Class1
2	Class1	Class2

Dari Tabel 4 terdapat 1 kesalahan prediksi, maka dengan rumus sederhana dapat dilakukan pengukuran kinerja algoritma adalah sebagai berikut:

$$kinerja = \frac{\text{jumlah instance yang diprediksi benar}}{\text{jumlah instance}} \tag{3}$$

Untuk melakukan perbandingan pengukuran dari metode yang menggunakan seleksi forward selection dan yang tidak adalah menghitung selisih akurasi dari kinerja dengan menggunakan seleksi forward selection dan yang tidak menggunakan seleksi forward selection sebagai berikut :

$$\text{Besar pengaruh} = \text{kinerja}(fs) - \text{kinerja} \tag{4}$$

B. Implementasi Pengukuran

Data yang diperoleh sebanyak 100 data dengan target variable “tepat_waktu” dengan view pada R Studio sebagai berikut :

	Kelamin	Sekolah	lpk	Lulus	Angkatan	Tepat
1	L	SMA Swasta	< 3.5	2017	2012	TIDAK
2	P	SMA Swasta	< 3.5	2016	2012	YA
3	L	SMA Swasta	> 3.5	2017	2012	TIDAK
4	P	SMA Negeri	< 3.5	2016	2012	YA
5	L	SMA Negeri	< 3.5	2016	2012	YA
6	L	SMA Negeri	< 3.5	2017	2012	TIDAK
7	P	SMA Negeri	< 3.5	2016	2012	YA
8	L	SMA Swasta	> 3.5	2016	2012	YA
9	L	SMA Swasta	> 3.5	2017	2012	TIDAK
10	L	SMA Swasta	< 3.5	2017	2012	TIDAK

Showing 1 to 10 of 100 entries

Gambar 3. Dataset

Dari data tersebut didapatkan summary sebagai berikut :

```
> summary(is)
  kelamin          sekolah      Ipk      Lulus      Angkatan      Tepat
L:66   Pondok Pesantren: 2   < 3.5:89   Min.   :2016   Min.   :2012   TIDAK:52
P:34   SMA Negeri      :45   > 3.5:11   1st Qu.:2016   1st Qu.:2012   YA   :48
      SMA Swasta      :48
      SMK              : 5
      Mean   :2017
      3rd Qu.:2017
      Max.   :2017
      Mean   :2012
      3rd Qu.:2012
      Max.   :2012
```

Gambar 4. Summary Dataset

Berdasarkan summary tersebut diatas maka dilakukan pembagian data dengan menggunakan metode k-Fold Cross Validation, dalam hal ini 2-Fold Cross Validation. Adapun penentuan datasetnya adalah sebagai berikut :

Dataset1	Dataset2	Tepat
24	24	YA
26	26	TIDAK

Gambar 5. fold Cross Validation

Teknik cross validation akan melakukan proses pembelajaran dan pengujian sebanyak 2 kali. Pertama, proses pembelajaran akan menggunakan data kotak abu-abu. Kemudian melakukan proses pengujian menggunakan data kotak hitam. Kedua, proses pembelajaran akan menggunakan data kotak hitam. Kemudian melakukan proses pengujian menggunakan data kotak abu-abu. Kinerja metode klasifikasi di atas adalah rata-rata kinerja dari dua proses pembelajaran dan pengujian.

Dari penjelasan singkat di atas, dapat dilihat semua data digunakan sebagai data training dan testing. Sehingga algoritma klasifikasi yang digunakan lebih teruji dan nilai kinerja yang didapat lebih valid.

	Kelamin	Sekolah	Ipk	Lulus	Angkatan	Tepat
1	L	SMA Swasta	< 3.5	2016	2012	YA
2	L	SMK	< 3.5	2016	2012	YA
3	P	SMA Negeri	< 3.5	2016	2012	YA
4	L	SMA Swasta	< 3.5	2016	2012	YA
5	P	SMA Negeri	< 3.5	2016	2012	YA
6	L	SMA Swasta	< 3.5	2016	2012	YA
7	P	SMA Swasta	< 3.5	2016	2012	YA
8	P	SMA Negeri	< 3.5	2016	2012	YA

Showing 1 to 9 of 50 entries

Gambar 6. Sample data testing 2-fold cross validation (kolom *lulus* dan *tepat* tidak diikuti untuk testing)

Hasil Perhitungan untuk data testing pertama data yang benar diprediksi sebanyak 28 data sehingga dapat dikatakan tingkat akurasi untuk data testing pertama sebesar 56% dengan keterangan pada gambar 7 :

```
Naive Bayes Classifier for Discrete Predictors
Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
TIDAK YA
0.52 0.48

Conditional probabilities:
kelamin
Y L P
TIDAK 0.7692308 0.2307692
YA 0.5833333 0.4166667

sekolah
Y Pondok Pesantren SMA Negeri SMA Swasta SMK
TIDAK 0.0000000 0.38461538 0.57692308 0.03846154
YA 0.04166667 0.58333333 0.37500000 0.00000000

Ipk
Y < 3.5 > 3.5
TIDAK 0.80769231 0.19230769
YA 0.91666667 0.08333333
```

Gambar 7. Hasil data testing dataset1

Hasil Perhitungan untuk data testing kedua data yang benar diprediksi sebanyak 29 data sehingga dapat dikatakan tingkat akurasi untuk data testing pertama sebesar 58% dengan keterangan pada gambar 8 :

```

A-priori probabilities:
Y
TIDAK    YA
0.52    0.48

Conditional probabilities:
kelamin
Y          L          P
TIDAK 0.6538462 0.3461538
YA     0.6250000 0.3750000

Sekolah
Y          Pondok Pesantren SMA Negeri SMA Swasta SMK
TIDAK 0.03846154 0.50000000 0.42307692 0.03846154
YA     0.00000000 0.33333333 0.54166667 0.12500000

Ipk
Y          < 3.5    > 3.5
TIDAK 0.88461538 0.11538462
YA     0.95833333 0.04166667
    
```

Gambar 8. Hasil data testing dataset2

IX. KESIMPULAN

Dari hasil penelitian di atas maka dapat diambil kesimpulan terjadi peningkatan akurasi dengan menambahkan metode pengujian Cross Validation sebesar 2% dari data pengujian yang dilakukan berdasarkan data yang ada.

DAFTAR PUSTAKA

- [1] Kusrini; Taufiq, Luthfi Emha, Algoritma Data Mining, Penerbit Andi, Yogyakarta, 2009
- [2] Larose, Daniel, T, Discovering Knowledge in Data an Introduction to Data Mining, Wiley-Interscience, United States of America, 2005
- [3] Suyanto, Data Mining Untuk Klasifikasi dan Klasterisasi Data, Penerbit Informatika, Bandung, 2017
- [4] Ishtiaq Ahmed; Donghai Guan; Tae Choong Chung, 2014, SMS Classification Based on Naïve Bayes Classifier and Apriori Algorithm Frequent Itemset, International Journal of Machine Learning and Computing, Vol. 4, No. 2, April 2014
- [5] Kumar Krishna Venkata S; Kiruthika P: 2015, An Overview of Classification Algorithm in Data mining, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 12, December 2015
- [6] Kusumadewi Sri, 2009, Klasifikasi Status Gizi Menggunakan Naïve Bayesian Classification, CommIT, Vol. 3 No. 1 Mei 2009, hlm. 6 – 11
- [7] Manjusha K. K., Sankaranarayanan, K., Seena P., 2014, Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification, International Journal of Advanced Research in Computer Science and Software Engineering, Vol 4, No 1, Hal 864868.
- [8] Priyanti, Evi; 2017, Penerapan Algoritma Naïve Bayes untuk klasifikasi Bakteri Gram Negatif, Jurnal Teknik Komputer Vol 3 No. 2 Agustus 2017
- [9] Nofriansyah Dicky; Erwansyah Kamil; Ramadhan Mukhlis, 2016, Penerapan Data Mining dengan Algoritma Naïve Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi), Jurnal SAINTIKOM Vol.15, No. 2, Mei 2016
- [10] Saleh Alfa, 2015, Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga, Citec Journal, Vol. 2, No. 3, Mei 2015 – Juli 2015
- [11] Faisal, M Reza, Seri Belajar Data Science Klasifikasi Dengan Bahasa Pemrograman R, <https://bukudatasienceklasifikasir.codeplex.com/>, September 2017
- [12] Han, J., Kamber, M., & Pei, J. , Data Mining: Concepts and Techniques Third. Elsevier, <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>, 2012