

# PENERAPAN MODEL NAIVE BAYES UNTUK MEMPREDIKSI POTENSI PENDAFTARAN SISWA DI SMK TAMAN SISWA TELUK BETUNG BERBASIS WEB

<sup>1</sup>Sherlyn Eka Yuliana Putri, <sup>2</sup>Saniati, S.ST., M.T, <sup>3</sup>Ade Surahman, S.Kom., M.Kom

S1 Informatika, Fakultas Teknik Dan Ilmu Komputer, Universitas Teknokrat Indonesia

Jl. ZA. Pagar Alam No.9 -11, Labuhan Ratu, Kec. Kedaton, Kota Bandar Lampung, Lampung 35132

Email: [sherlyneka06@gmail.com](mailto:sherlyneka06@gmail.com), [saniati@teknokrat.ac.id](mailto:saniati@teknokrat.ac.id), [adesurahman@teknokrat.ac.id](mailto:adesurahman@teknokrat.ac.id),

**Abstrak** – SMK Taman Siswa setiap awal tahun ajaran baru akan menyelenggarakan penerimaan calon peserta didik baru. Persaingan dalam menarik minat calon siswa baru untuk masuk SMK Taman Siswa dapat terjadi. Dengan adanya persaingan antar SMK, jumlah pendaftaran siswa baru setiap tahunnya akan mengalami peningkatan dan tidak menutup kemungkinan akan mengalami penurunan calon peserta didik baru, sehingga jumlah siswa baru di SMK tersebut tidak stabil. Permasalahannya adalah ketidakstabilan pendaftaran calon peserta didik baru di SMK Taman Siswa. Berdasarkan masalah tersebut, maka penulis merancang sebuah sistem Penerapan Model Naive Bayes Untuk Memprediksi Potensi Pendaftaran Siswa Di SMK Taman Siswa Teluk Betung Berbasis WEB. Sistem yang dapat melakukan prediksi potensi pendaftaran siswa baru yang akan masuk ke SMK Taman Siswa.

Hasil akhir yang diperoleh adalah sebuah sistem informasi prediksi potensi pendaftaran siswa dengan menggunakan model naive bayes yang memiliki kemampuan untuk memprediksi potensi banyaknya siswa yang mendaftar dan tidak mendaftar di SMK Taman Siswa. Aplikasi ini diuji menggunakan perhitungan akurasi dimana hasilnya didapatkan tingkat akurasi sebesar 86% untuk prediksi potensi pendaftaran siswa.

**Kata kunci:** Penerapan, Naive Bayes, Pendaftaran Siswa, Website.

## 1 Pendahuluan

### 1.1 Latar Belakang

Persaingan antar Sekolah Menengah Atas (SMA) atau Sekolah Menengah Kejuruan (SMK) Swasta semakin meningkat belakangan ini. Perkembangan teknologi juga ikut andil dalam mempengaruhi perkembangan pendidikan di Indonesia, khususnya SMK Swasta. Secara tidak langsung menimbulkan persaingan dalam menarik minat calon siswa baru untuk masuk SMK Swasta tersebut. Dengan adanya persaingan antar SMK jumlah pendaftaran siswa baru setiap tahunnya akan mengalami peningkatan dan tidak menutup kemungkinan akan mengalami penurunan calon peserta didik baru, sehingga jumlah siswa baru di SMK tersebut tidak stabil. Pada setiap awal tahun ajaran baru pihak SMK Taman Siswa akan menyelenggarakan penerimaan calon peserta didik baru. Ketidakstabilan dalam pendaftaran calon peserta didik baru di SMK Taman Siswa juga bisa terjadi. Untuk hal itu akan lebih baik jika pihak penyelenggara mampu melakukan antisipasi untuk meminimalisir kelemahan yang ada dengan melakukan prediksi potensi pendaftaran siswa baru yang akan masuk ke SMK Taman Siswa.

Berdasarkan uraian masalah diatas, peneliti mengusulkan pembuatan sistem yang dapat memberikan informasi prediksi potensi pendaftaran siswa baru dengan menggunakan metode Naive Bayes dan parameter yaitu usia, jenis kelamin, alamat, asal pendidikan, alamat pendidikan, jurusan, agama, prediksi potensi pendaftaran. Peneliti berharap dengan dilakukan penelitian ini dapat memberikan kebutuhan data yang sesuai dalam menginformasikan prediksi potensi pendaftaran mahasiswa baru.

Tujuan penelitian ini adalah untuk membangun sistem prediksi potensi pendaftaran siswa baru di SMK

Taman Siswa dan hasil prediksi potensi pendaftaran mahasiswa baru di SMK Taman Siswa memiliki akurasi yang baik.

### 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan di atas, maka penulis dapat merumuskan permasalahan yang akan dijadikan penyusunan dalam laporan skripsi sebagai berikut:

1. Bagaimana menerapkan model Naive Bayes untuk memprediksi potensi pendaftaran siswa di SMK Taman Siswa?
2. Apakah hasil prediksi potensi pendaftaran siswa baru di SMK Taman Siswa memiliki akurasi yang baik?

### 1.3 Tujuan Penelitian

Adapun tujuan dari penelitian yang ingin dicapai adalah sebagai berikut:

1. Membangun sistem prediksi potensi pendaftaran siswa baru di SMK Taman Siswa.
2. Hasil prediksi potensi pendaftaran mahasiswa baru di SMK Taman Siswa memiliki akurasi yang baik.

### 1.4 Batasan Penelitian

Berdasarkan penelitian ini, maka penulis memberikan batasan-batasan sebagai berikut:

1. Perancangan sistem berisi variabel yaitu usia, jenis kelamin, alamat, asal pendidikan, jurusan, agama, pekerjaan orang tua, prediksi potensi pendaftaran.
2. Hanya memakai 2 Class yaitu Daftar dan Tidak Daftar.

### 1.4. Manfaat Penelitian

Terlaksananya penelitian perancangan sistem diharapkan mampu memberikan manfaat, yaitu:

1. Bagi pihak SMK Taman Siswa  
Dapat menginformasikan potensi pendaftaran calon siswa baru yang akan mendaftar di SMK Taman Siswa.
2. Bagi Peneliti  
Peneliti dapat mengimplementasikan hasil belajar selama perkuliahan di Universitas Teknokrat Indonesia, menambah wawasan bagi peneliti mengenai metode Naive Bayes.

## 2 Tinjauan Pustaka

### 2.1 Data Mining

Data Mining merupakan proses pengekstraksian informasi dari sekumpulan data yang sangat besar Saleh (2015). Pengelolaan Data Mining melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data (Taruna, 2008). Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkan data menjadi informasi-informasi penting yang dapat meningkatkan kualitas data yang akan dipakai. Dapat ditarik kesimpulan bahwa Data Mining merupakan proses mengumpulkan data berukuran besar yang akan diekstraksi kemudian menjadikan data tersebut menjadi informasi – informasi data yang nantinya dapat digunakan.

### 2.2 Naive Bayes

Naive Bayes Classifier merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan (Saleh, 2015). Teorema bayes dikombinasikan dengan “Naive” yang berarti setiap atribut/variabel bersifat bebas (independent). Naive Bayes Classifier dapat dilatih dengan efisien dalam pembelajaran terawasi (supervised learning) (Gunawan, 2018). Untuk menjelaskan metode *Naive Bayes*, proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang sesuai bagi data yang dianalisis.

Karena itu, metode *Naive Bayes* di atas dapat disesuaikan sebagai berikut:

$$P(x_1, x_2, K, x_n | c_j) = P(c_j) \prod_i P(x_i | c_j)$$

Dimana :

- $P(c_j)$  = Peluang hipotesis j (*Prior Probability*)
- $P(x_1, x_2, K, x_n | c_j)$  = Peluang atribut-atribut (inputan) jika diketahui hipotesis  $c_j$  (*Likelihood*)
- $P(x_i | c_j)$  = *Posterior Probability*

Kesimpulan dari rumus diatas adalah :

*Posterior = prior x likelihood*

### 2.3 MySQL

MySQL merupakan program pengakses Databases yang bersifat jaringan sehingga dapat digunakan untuk aplikasi multi user. Kelebihan lain dari MySQL adalah MySQL menggunakan bahasa query standar yang dimiliki SQL (Nugroho,2004). Kepopuleran MySQL antara lain karena MySQL menggunakan Sql sebagai bahasa dasar untuk mengakses Database-nya sehingga mudah untuk digunakan. MySQL juga bersifat open source dan free pada berbagai platform kecuali pada windows yang bersifat shareware. MySQL didistribusikan dengan lisensi open source GPL (General Public License) mulai versi 3.23, pada bulan Juni 2000 (Suhartanto, 2012).

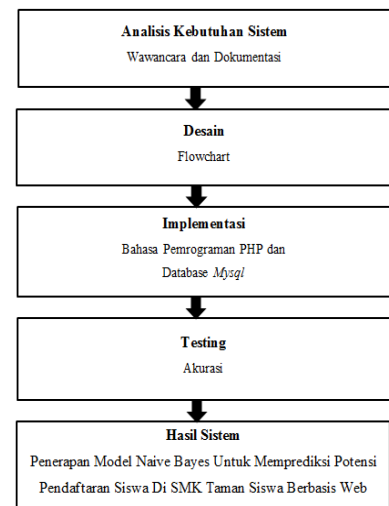
### 2.4 PHP

Hypertext preprocessor (PHP) adalah salah satu bahasa pemrograman yang dirancang untuk membangun aplikasi berbasis Web. Aplikasi Web adalah aplikasi yang disimpan dan dieksekusi (oleh php engine) di lingkungan Web server. Setiap permintaan yang dilakukan oleh user melalui aplikasi klien (Web browser) akan direspon oleh aplikasi Web dan hasilnya akan dikembalikan lagi ke hadapan user (Raharjo, 2011).

## 3 Metode Penelitian

### 3.1 Tahapan Penelitian

Tahapan penelitian pada dasarnya adalah hubungan antara konsep-konsep yang ingin diamati atau diukur melalui penelitian yang akan dilakukan. Berdasarkan kerangka penelitian yang ada, kerangka penelitian dapat dilihat pada gambar 3.1.



### 3.2 Analisis Kebutuhan Sistem

Analisis permasalahan yang terjadi pada penelitian ini adalah suatu hal yang akan menjadi objek penelitian untuk mencoba memberikan solusi berupa hasil dari perancangan sistem prediksi potensi pendaftar, adapun permasalahan yang dihadapi adalah :

1. Ketidakstabilan pendaftar untuk mendaftar di SMK TamanSiswa Teluk Betung.
2. Belum adanya sistem untuk prediksi potensi pendaftaran siswa baru.

Dalam penelitian ini, peneliti melakukan analisis kebutuhan sistem dengan cara pengumpulan data dengan cara :

1. Wawancara

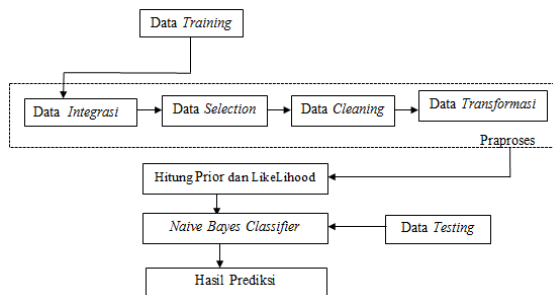
Wawancara merupakan teknik pengumpulan data yang dilakukan dengan cara menanyakan langsung kepada pihak SMK Taman Siswa Teluk Betung.

2. Dokumentasi

Dokumentasi yaitu melakukan pengumpulan data yang dibutuhkan untuk penelitian. Dalam penelitian ini, data-data penelitian didapatkan dari data siswa dengan Tahun ajaran 2017/2018 sebanyak 66 dataset Tahun ajaran 2018/2019 sebanyak 81 dataset, Jumlah data pendaftaran keseluruhan 147.

3.3 Desain

Tahapan ini dilakukan untuk pengembangan yang menentukan alur sistem secara keseluruhan. Dalam penelitian ini peneliti membuat bagan flowchart untuk menentukan alur sistem, dapat dilihat pada Gambar 3.2.



3.3 Data Training

Sumber data training adalah data pendaftaran siswa SMK Taman Siswa tahun ajaran 2017/2018 dan 2018/2019. Jumlah data pendaftaran keseluruhan 147 data dari:

- a. Tahun ajaran 2017/2018 sebanyak 66 dataset
- b. Tahun ajaran 2018/2019 sebanyak 81 dataset

3.4 Implementasi

Tahapan ini implementasi sistem terdapat beberapa menu yaitu login, home, data siswa asli, data siswa asli (praproses), data testing, kalkulasi naive bayes, persentase prior, likelihood. Pembuatan sistem prediksi peneliti menggunakan:

- a. Bahasa pemrograman PHP
- b. Database *Mysql*

3.5 Testing

Data testing berasal dari data training yang sudah di praproses. Pengujian data testing yaitu sebanyak 100% yang akan nantinya diteliti dengan metode Naive Bayes. Adapun data testing yang digunakan berdasarkan data siswa SMK Taman Siswa tahun ajaran 2017/2018 dan 2018/2019 tersimpan pada file Ms. Excel data telah di praproses terlebih dahulu sebelum di uji. Diketahui jumlah data testing sebanyak 147 dataset dimana sebanyak 8 atribut dan Class sebanyak 2.

3.6 Alat Penelitian

Dalam penelitian ini membutuhkan beberapa alat yang terdiri dari perangkat keras (Hardware) dan perangkat lunak (Software). Alat-alat penelitian dijelaskan sebagai berikut :

a. Perangkat Keras (Hardware)

Spesifikasi perangkat keras yang digunakan oleh peneliti untuk membangun sistem adalah sebagai berikut :

- 1. Laptop dengan spesifikasi Processor Intel(R) Core(TM)2 Duo CPU P7570 @2.26Hz (2CPUs), ~2.3GHz
- 2. RAM 4GB
- 3. Harddisk 300GB
- 4. VGA 1759MB
- 5. Monitor LCD dengan resolusi layar 1366 x 768 Pixel

b. Perangkat Lunak (Software)

Spesifikasi perangkat lunak yang digunakan oleh peneliti untuk membangun sistem adalah sebagai berikut:

- 1. Sistem Operasi Microsoft Windows 10 x64bit
- 2. Bahasa Pemrograman PHP (Personal Home Page)
- 3. Koneksi Database XAMPP
- 4. Database *MySQL*

4 Analisis dan Perancangan Sistem

Analisis kebutuhan fungsional dilakukan untuk mengetahui proses-proses apa saja yang dilakukan oleh sistem. Berikut ini merupakan fungsionalitas apa saja yang dibutuhkan oleh :

- 1. Sistem harus dapat melakukan pengolahan prediksi potensi pendaftaran siswa SMK Taman Siswa.
  - a. Admin dapat masuk kedalam sistem
  - b. Admin dapat memasukkan data siswa asli dan data siswa praproses
  - c. Sistem dapat menghitung nilai Prior
  - d. Sistem dapat menghitung nilai Likelihood
  - e. Admin dapat menyimpan, mengubah, menghapus data yang ada pada sistem
  - f. Sistem dapat menampilkan informasi data siswa
  - g. Sistem dapat menampilkan data siswa berdasarkan atribut yang diinginkan

4.1 Analisis Kebutuhan Non Fungsional

4.1.1 Analisis Kebutuhan Non Fungsional

Analisis kebutuhan non fungsional dilakukan untuk mengetahui perilaku yang dimiliki oleh sistem. Spesifikasi kebutuhan non fungsional melibatkan analisis perangkat keras (*hardware*), analisis perangkat lunak (*software*) dan analisis keamanan (*security*).

1. Analisis Perangkat Keras/Hardware

Kebutuhan perangkat keras (*hardware*) yang diperlukan untuk mengimplementasikan sistem ini untuk penginputan data. Spesifikasi perangkat keras yang dibutuhkan yaitu Laptop dengan spesifikasi :

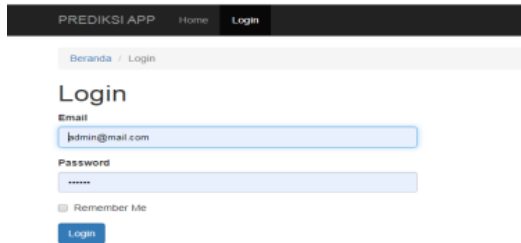
- Processor Intel(R) Core(TM)2 Duo CPU P7570 @2.26Hz (2CPUs), ~2.3GHz,
- RAM 4GB,
- Harddisk 300GB, VGA 1759MB,

- Monitor LCD dengan resolusi layar 1366 x 768 Pixel.
2. **Analisis Perangkat Lunak / Software**  
Perangkat lunak (*software*) yang dibutuhkan dalam implementasi sistem aplikasi ini yaitu :
- Sistem Operasi *Microsoft Windows 10*,
  - PHP,
  - *Google Chrome*,
  - *MySql*
3. **Analisis Keamanan/Security**  
Kebutuhan keamanan berisi pernyataan tentang mekanisme pengamanan aplikasi data yang akan diimplementasikan pada sistem. Keamanan yang dibutuhkan oleh sistem ini yaitu dilengkapi *password* untuk masuk kedalam sistemnya.

5 Hasil Penelitian dan Pembahasan

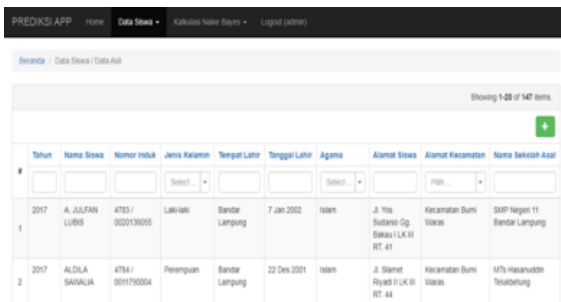
5.1 Tampilan Halaman Utama Sistem

Tampilan saat pertama kali aplikasi digunakan, pada form ini terdapat list fitur-fitur aplikasi yang dapat digunakan setelah user melakukan login pada area yang telah disediakan.



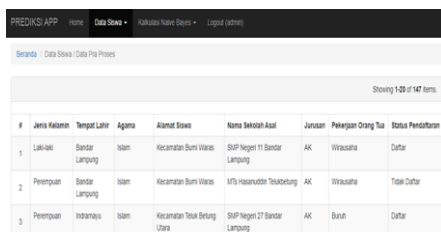
5.2 Tampilan Data Asli Siswa

Menampilkan data siswa asli yang didapatkan dari SMK Taman Siswa. Tampilan data asli diatas juga menampilkan 150 atau 100 % data siswa yang telah diinputkan melalui sistem



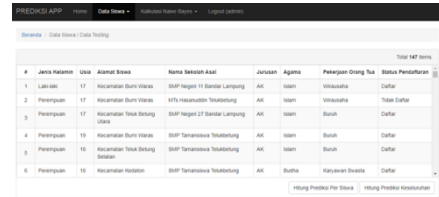
5.3 Tampilan Data Siswa Praproses

Tampilan pada data siswa praproses diatas menampilkan data yang di praproses dari data asli untuk kesesuaian yang dibutuhkan perhitungan naive bayes.



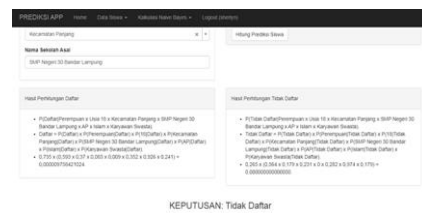
5.4 Tampilan Data Testing

Tampilan pada data testing diatas memiliki fungsi yaitu menampilkan data testing siswa yang akan dihitung prediksi potensi pendaftarannya menggunakan metode naive bayes dalam sistem.



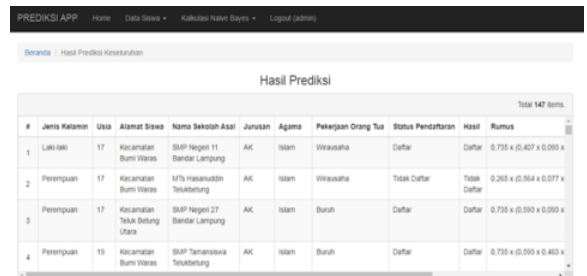
5.5 Tampilan Hitung Prediksi Persiswa

Tampilan pada hitung prediksi persiswa diatas memiliki fungsi yaitu meng-inputkan data yang akan diuji secara manual sesuai degan keinginan admin, hitung prediksi persiswa menampilkan prediksi satu dataset uji



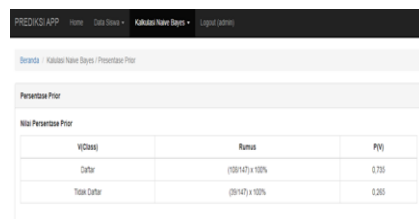
5.6 Tampilan Hitung Prediksi Keseluruhan

Tampilan pada hitung prediksi keseluruhan diatas memiliki fungsi yaitu menampilkan seluruh perhitungan prediksi potensi pendaftaran siswa serta perbandingan hasil prediksi sistem dan data asli siswa yang sudah diinputkan pada sistem. Pada tampilan hitung prediksi keseluruhan juga menampilkan rumus dari perhitungan naive bayes.



5.7 Tampilan Persentase Prior

Tampilan diatas merupakan tampilan yang menampilkan hitungan presentase prior yang dihitung secara otomatis dari data siswa yang diinputkan. Pada tampilan persentase prior terdapat prior daftar dan prior tidak daftar serta rumus yang digunakan.



### 5.8 Tampilan Likelihood

Tampilan diatas merupakan tampilan yang menampilkan hitungan likelihood yang dihitung secara otomatis dari data siswa yang diinputkan. Pada tampilan likelihood terdapat sebanyak 104 kriteria dari daftar dan tidak daftar

No	P(A B)	Nilai
1	P(A B)Kuliah	0,407
2	P(A B)Kuliah	0,500
3	P(A B)Kuliah dan Buku (Kelas)	0,200
4	P(A B)Kuliah dan Tidak Belajar (Kelas)	0,200
5	P(A B)Kuliah dan Tidak Belajar (Siswa)	0,200
6	P(A B)Kuliah dan Habis	0,500
7	P(A B)Kuliah dan Tidak Belajar (Tugas)	0,500

### 5.9 Pengujian

Penilaian akurasi sistem menggunakan rumus :

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Jumlah\ Total\ Data\ Testing} \times 100\% = \frac{128}{147} \times 100\% = 0,870\%$$

Hasil Uji

Jumlah seluruh data : 147  
 Jumlah data sesuai : 128  
 Jumlah data tidak sesuai : 19

## 6 Kesimpulan dan Saran

### 6.1 Kesimpulan

Berdasarkan pembahasan dan evaluasi dari bab-bab sebelumnya, maka dapat diperoleh beberapa kesimpulan sebagai berikut:

1. Menerapkan model *naive bayes* dengan cara mengolah data asli siswa SMK Taman Siswa Teluk Betung menjadi data praproses yang akan dibutuhkan untuk perhitungan *naive bayes*, menentukan Prior dan Likelihood dengan pembuatan sistem menggunakan bahasa pemrograman PHP, dan semua data *entry* merupakan hasil perhitungan menggunakan metode *naive bayes clasifier*.
2. Iya, menghasilkan akurasi yang baik berdasarkan hasil pengujian sistem yang dibandingkan dengan data yang sudah ada atau data asli dari SMK Taman Siswa didapatkan tingkat akurasi sebesar 87%.

### 6.2 Saran

Agar Sistem ini dapat digunakan dan berjalan dengan baik seperti yang diharapkan, maka disarankan hal-hal sebagai berikut :

1. Dapat menggunakan lebih banyak data *training* agar hasil hitung uji semakin baik dan valid. Dapat dikembangkan menjadi aplikasi yang lebih luas dari pengetahuan yang dimiliki.
2. Penelitian ini diharapkan dapat dikembangkan menggunakan metode yang berbeda untuk membandingkan nilai keakuratannya.

### Daftar Pustaka

Andini T.I., Witanti W., dan Renaldi F. 2016. Prediksi Potensi Pemasaran Produk Baru dengan Metode

Naïve Bayes Classifier dan Regresi Linear, Seminar Nasional Aplikasi Teknologi Informasi (SNATi)

Anhar. 2010. PHP dan MySQL Secara Otodidak, Jurnal Ilmiah Ilmu Komputer. PT Trans Media Jakarta

Dhanashree S.M., Mayur P.B., dan Shruti D.D. 2012. Prediction System For Heart Disease Using Naive Bayes, International journal of Advanced Computer and Mathematical Sciences

Firman A., Wowor H., dan Najooan X. 2014. Sistem Informasi Perpustakaan Berbasis Web Application, Jurnal Sistem Informasi Bisnis

Fithri D. L., dan Darmanto E. 2014. Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes, Vol.319-324.

Gunawan I. 2018. Sistem Pakar Diagnosenyakit Kulit Pada Kucing Menggunakan Naive Bayes Berbasis Web, Lampung: Universitas Teknokrat

Herkutanto C. 2013. Pengembangan Dan Analisis Kualitas Aplikasi Pencarian Gambar Berdasar Histogram Warna Berbasis Web, Yogyakarta: Fakultas Teknik Universitas Negeri Yogyakarta

Jogiyanto. 2009. Analisis dan Desain Sistem Informasi, Yogyakarta: Andi Offset

Kusrini dan Luthfi, E. T. 2009. Algoritma Data Mining, Yogyakarta

Kustiyaningsih. 2011. Pemrograman Basis Data Berbasis Web Menggunakan PHP dan MySQL, Yogyakarta: Graha Ilmu

Mujiasih S. 2011. Pemanfaatan Data Mining Untuk Prakiraan Cuaca Utilization Of Data Mining For Weather Forecasting, Jurnal Meteorologi Dan Geofisika,

Saleh A. 2015. Implementasi Metode Klasifikasi Naïve Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga, Citec Journal 2(3), V2I3.49

Shyara. dan Taruna R, M. S. H. 2008. Enhanced Naïve Bayes Algorithm for Intrusion Detection in Data Mining, Journal of Computer Science and Information Technologies, Vol. 4 (6)

Sommerville, Ian , 2009, Software Engineering–Eight Edition, AddisonWesley, Massachussets.

Suhartanto, M. 2012. Pembuatan Website Sekolah Menengah Pertama Negeri 3 Delanggu Dengan Menggunakan Php Dan MySQL

- Syarli, S. and Muin, A. A. 2016. Metode Naive Bayes Untuk Prediksi Kelulusan ( Studi Kasus : Data Mahasiswa Baru Perguruan Tinggi, Jurnal Ilmiah Ilmu Komputer
- Tampubolon, K., Saragih H., Reza B. 2013. Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan
- Wahyudi, D., Mirza, A. H. dan Merrieayu, P. H. 2013. 'Implementasi Data Mining Dengan Naive Bayes Classifier Untuk Mendukung Strategi Promosi, Palembang: Studi kasus Universitas Bina Darma Palembang
- Wardhana, A. 2018. Strategi Promosi Penerimaan Siswa Baru, Yogyakarta: (Studi Kasus Smk Kesehatan Cipta Bhakti Husada Yogyakarta)
- H Sulistiani, K Muludi and A Syarif (2021). Implementation of various artificial intelligence approach for prediction and recommendation of personality disorder patient. *Journal of Physics: Conference Series*, [online] 1751, p.012040. Available at: <https://doi.org/10.1088/1742-6596/1751/1/012040>.