

KLASIFIKASI JUDUL SKRIPSI BERDASARKAN PROGRAM STUDI MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*

Mochamad Husni¹⁾, Nuriadin Hidayat²⁾

¹⁾Sistem Informasi, STMIK Pradnya Paramita
husni_stmik@yahoo.co.id

²⁾ Sistem Informasi, STMIK Pradnya Paramita
adhinhidayat22@gmail.com

Abstract

As a university based on the STMIK Pradnya paramita informatics study program, it has implemented an Academic Information System in serving students, and the system has been integrated, but the academic section is still experiencing obstacles, in this case the Study Program, one of which is to classify the thesis titles submitted by students, as for the unsuitability of the thesis titles of each study program, it is still often found, From these problems, it is necessary to have a solution to minimize and even eliminate inappropriate thesis titles in each study program, namely by designing a thesis title submission application by applying the Naïve Bayes Classifier method, it is hoped that the application of the Naïve Bayes Classifier method can determine the thesis title more precisely and determine the supervisor who is in accordance with their respective fields of expertise.

Keywords: Classification, Skripsi, Naïve Bayes Classifier

1. PENDAHULUAN [Times New Roman 11 bold]

Sekolah Tinggi Pradnya Paramita (STIMATA) dalam kegiatan perkuliahan maupun administrasi telah menerapkan teknologi sebagai media pendukung. Pengisian Kartu Rencana Studi (KRS) dan pencetakan Kartu Hasil Studi (KHS) yang semua itu dapat diakses pada *website* kampus yang kemudian diarahkan menuju sistem untuk pengisian KRS maupun pencetakan KHS. Sistem Informasi yang telah disediakan oleh pihak kampus untuk saat ini masih berjalan dengan baik dan bermanfaat bagi mahasiswa maupun dari pihak administrasi kampus.

Sistem yang telah terintegrasi, bagian akademik STIMATA masih mengalami kendala, salah satunya adalah pengklasifikasian judul skripsi oleh mahasiswa. Dalam kegiatan pengklasifikasian ketua program studi baru bisa melakukan pengklasifikasian judul apabila proposal sudah terkumpul. Berdasarkan penelitian Ahmad Fathan (2016:3) pengklasifikasian judul skripsi secara otomatis dengan menggunakan metode *naïve bayes* sangat membantu untuk melakukan klasifikasi setiap judul skripsi dengan cepat dapat membantu memberikan gambaran kepada mahasiswa untuk mencari judul skripsi yang

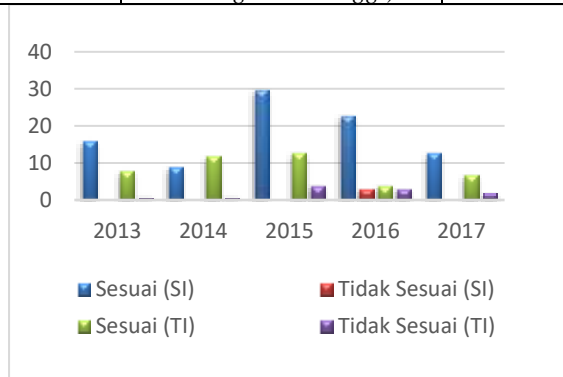
relevan dengan konsentrasi mahasiswa tersebut. Daftar skripsi mahasiswa dari tahun 2013-2017, dapat dilihat pada tabel 1.1 sebagai berikut :

Tabel 1.1 Judul Skripsi Mahasiswa STIMATA tahun 2013 – 2017

Tahun	Sistem Informasi		Teknik Informatika	
	Sesuai	Tidak Sesuai	Sesuai	Tidak Sesuai
2013	16	0	8	1
2014	9	0	12	1
2015	30	0	13	4
2016	23	3	4	3
2017	13	0	7	2

Berdasarkan Tabel 1.1 dapat di ambil kesimpulan bahwa adanya ketidaksesuaian judul skripsi dengan masing-masing program studi. Hal ini dapat ditunjukkan pada Gambar grafik 1.1 sebagai berikut.

Nama Peneliti	Judul	Akurasi
Liu Z, dkk (2015)	<i>A Classification Method for Complex Power Quality Disturbances Using EEMD and Rank Wavelet SVM</i>	92.8%
Shaleh A (2015)	<i>Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga</i>	78.3%
Tang B, dkk (2016)	<i>Toward Optimal Feature Selection in Naïve Bayes For Text Categorization</i>	80.0%
Muin, A. (2016)	<i>Metode Naïve Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)</i>	93.3%



Gambar 1.1 Grafik Kesesuaian Judul Skripsi

Mahasiswa yang mengajukan judul skripsi diharuskan sesuai dengan konsentrasi masing-masing Program Studi. Apabila judul skripsi yang diambil oleh mahasiswa tidak sesuai dengan konsentrasi Program Studinya akan menyebabkan waktu pengerjaan skripsi menjadi tidak tepat waktu.

Permasalahan yang terjadi pada Program Studi adalah tidak adanya sistem yang menyeleksi proposal skripsi, sehingga tidak semua proposal penelitian mahasiswa sesuai dengan konsentrasinya.

Dari kelemahan tersebut, program studi perlu melakukan klasifikasi judul skripsi agar sesuai konsentrasi Program Studi. Adapun cara yang lebih relevan untuk mengetahui judul yang diajukan oleh masing-masing mahasiswa adalah melakukan analisis pengklasifikasian setiap judul yang didaftarkan. Pengklasifikasian secara manual akan membutuhkan waktu dan tenaga untuk proses seleksi. Berdasarkan penelitian Sharma N. dan Singh M. (2016:1) menggunakan metode pengklasifikasian dapat

meningkatkan kinerja serta data statistik yang lebih akurat dan cepat.

Berdasarkan permasalahan tersebut, maka solusi yang ditawarkan adalah dengan melakukan pengklasifikasian judul skripsi menggunakan *datamining* dengan metode *naïve bayes classifier*. Didukung oleh penelitian terdahulu yang ditunjukkan pada tabel 1.2 berikut ini:

Tabel 1 Hasil penelitian terdahulu

2. KAJIAN LITERATUR

2.1. Naïve Bayes Classifier

Pernyataan dari Gamallo P. Garcia M. (2014:10) bahwa *Naïve Bayes Classifier* adalah model probabilistik sederhana berdasarkan aturan Bayes dengan asumsi independensi yang kuat. Model *Naïve Bayes* melibatkan penyederhanaan asumsi independen bersyarat. Dari pernyataan tadi diberikan kelas (positif atau negatif), kata-kata yang bersyarat independen satu sama lain. Asumsi ini tidak mempengaruhi akurasi klasifikasi teks dengan berapa banyak tapi membuat algoritma klasifikasi sangat cepat berlaku untuk masalah tersebut.

Terakhir menurut Kusriani, (2009:3) *Bayesian classification* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

2.2. Konsep Naïve Bayes Classifier

Naive bayes adalah tehnik yang di terapkan untuk menentukan kelas dari tiap masalah, yang sudah di bagi berdasarkan tiap-tiap masalah. perhitungan numerik berdasarkan pada pendekatan grup. *Naive bayes* memiliki beberapa manfaat seperti sederhana, cepat, memiliki tingkat akurasi yang tinggi.

Menurut Jurafsky, (2015:3) Intuisi *classifier* yang ditunjukkan pada gambar 2.1. Representasi sebuah dokumen teks seolah-olah itu *bag-of-words* (tas-dari-kata), yaitu, satu set kata-kata yang tidak di perintahkan dengan posisi mereka diabaikan, hanya menjaga

frekuensi mereka dalam dokumen. Dalam contoh pada gambar 2.1, bukan mewakili urutan kata dalam semua frasa seperti "Saya suka film ini" dan "Saya akan merekomendasikan hal", kita hanya dicatat bahwa kata saya terjadi 5 kali di seluruh kutipan, kata itu 6 kali, kata *word love, recommend, and movie once*, dan sebagainya. Menurut Dhande dan Patnaik, (2014:14) K. Ming Leung, menjelaskan rumus *Bayes* adalah:

$$y(\alpha|\beta) = y + \frac{y(\alpha)*y(\beta|\alpha)}{y(\beta)} \quad (2.1)$$

Dimana:

α : *Specific class.*

β : *Document wants to classify.*

$\gamma(\alpha)$ and $\gamma(\beta)$: *Prior probabilities.*

$\gamma(\alpha|\beta)$ and $\gamma(\beta|\alpha)$: *Posterior probabilities.*

Penilaian dari kelas α bisa berupa positif atau negatif. Dokumen adalah ulasan dari film tertentu. Model angka dari *Naive Bayes* merekam informasi tentang frekuensi kata pada dokumen. *Maximum Likelihood Estimate* (MLE) adalah frekuensi relatif dan sesuai dengan kemungkinan nilai masing-masing parameter yang diberikan data pelatihan. Untuk probabilitas sebelum perkiraan ini ditunjukkan pada Persamaan 1.

$$y(\alpha) = \frac{N_c}{N} \quad (2.2)$$

Dimana:

N_c : nomor dokumen pada kelas α .

N : jumlah nomor pada dokumen.

Dalam model *Multinomial*, mengasumsikan nilai atribut yang bebas satu sama lain yang diberikan untuk kelas tertentu $\gamma = (\alpha | \beta) = \gamma(\omega_1 \dots \omega_n | \alpha)$. Dalam model *multinomial*, dokumen memerintahkan urutan peristiwa kata, ditarik dari kosakata V . Asumsikan bahwa panjang dokumen *independen* dari kelas. dengan demikian, masing-masing $i\beta$ dokumen diambil dari pembagian *multinomial* kata-kata dengan banyak percobaan *independen* sebagai panjang $i\beta$. ini menghasilkan hal yang umum yaitu seperti kantong yang berisi banyak kata yang merepresentasikan dokumen dokumen. Model *BOW* umumnya digunakan dalam metode klasifikasi dokumen, di mana (frekuensi) terjadinya setiap kata digunakan sebagai fitur untuk pelatihan pengklasifikasian. dari dokumen yang diberikan, daftar standar dari kata-kata yang muncul dalam dokumen (kamus) dan menghitung vektor frekuensi dari kata-kata yang muncul dalam dokumen. Sudut antara dua vektor adalah ukuran yang banyak digunakan pada kedekatan antara dokumen.

Misalkan W kamus-himpunan semua hal (kata) yang terjadi setidaknya sekali dalam koleksi dokumen D . *BOW* representasi dn dokumen adalah *vektor* yang mempengaruhi $(\omega_1 n \dots \omega_n | W | n)$. Dalam kasus yang paling sederhana, pengaruh ω in $\in 1, -1$ menyatakan ada atau tidak adanya istilah tertentu dalam dokumen. Lebih umum, s mewakili frekuensi istilah i dalam dokumen n , sehingga menghasilkan frekuensi representasi. Transformasi set dokumen D ke dalam representasi *BOW* memungkinkan set diubah untuk dilihat sebagai matriks, dimana baris mewakili *vektor* dokumen, dan kolom adalah kata. Sebuah fitur *unigram* menandai kehadiran atau tidak adanya satu kata dalam teks. Memperkirakan kemungkinan $\gamma = (\omega | \alpha)$ sebagai *cy* frekuensi relatif merupakan istilah dari ω termasuk dalam dokumen kelas α dan juga beberapa istilah dalam dokumen.

$$y(\omega + \alpha)^2 = \frac{\text{count}(\omega, \alpha) + 1}{\text{count}(\alpha) + |V|} \quad (2.3)$$

Dimana:

$\text{Count}(\omega, \alpha)$: *Number of occurrences of ω in training documents from class α*

$\text{Count}(\alpha)$: *Number of words in that class*

$|V|$: *Number of terms in the vocabulary*

Masalah dengan estimasi *MLE* adalah bahwa hal itu adalah nol untuk kombinasi *termclass* yang tidak terjadi dalam data pelatihan. Data pelatihan tidak akan cukup besar untuk mewakili frekuensi dari kejadian yang cukup langka. Untuk menghilangkan masalah probabilitas nol, menggunakan *add-one* atau *Laplace smoothing*, ini hanya menambahkan satu untuk setiap hitungan. Tambahkan satu *smoothing* dapat diartikan sebagai *uniform prior* (setiap istilah terjadi sekali untuk setiap kelas) yang kemudian diperbarui sebagai bukti dari data pelatihan yang masuk. Kemudian, probabilitas dokumen yang diberikan kelasnya secara sederhana distribusi *multinomial* di representasikan pada persamaan 2. Akhirnya mengklasifikasikan dokumen baru menggunakan probabilitas posteriori. $\alpha | NB$ adalah probabilitas posterior, α_j adalah salah satu kelas dari α kelas dan β_i adalah dokumen.

Beberapa penelitian menunjukkan bahwa meskipun kemerdekaan asumsi antara kata-kata dalam sebuah dokumen tidak sepenuhnya terpenuhi, jelas bahwa asumsi kemerdekaan bersyarat jarang terjadi di sebagian besar aplikasi dunia nyata. Pendekatan langsung untuk mengatasi keterbatasan *Naive Bayes*

adalah untuk menggabungkan dengan *Neural Network* untuk mewakili eksplisit ketergantungan antar atribut.

2.3. CodeIgniter

Menurut Hakim (2010:8) *CodeIgniter* adalah sebuah *framework* PHP yang dapat membantu mempercepat *developer* dalam pengembangan aplikasi web berbasis PHP dibanding jika menulis semua kode program dari awal.

CodeIgniter pertama kali dibuat oleh Rick Ellis, CEO Ellislab, Inc. sebuah perusahaan yang memproduksi CMS (*Content Management System*) yang cukup handal, yaitu *Expression Engine*. Saat ini, *CodeIgniter* dikembangkan dan dimaintain oleh *Expression Engine Development Team*

3. METODE PENELITIAN

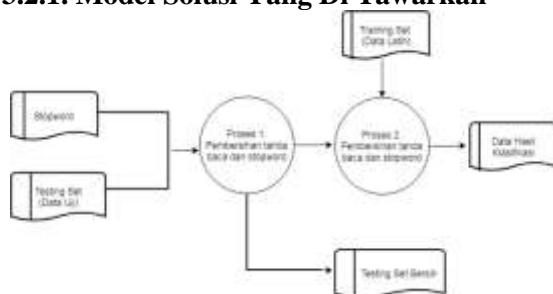
3.1. Deskripsi Permasalahan

Permasalahan yang akan diungkapkan dalam penelitian ini yaitu mengenai ketidaksesuaian judul skripsi dengan program studi yang salah satunya menyebabkan penelitian yang di kerjakan oleh mahasiswa tidak selesai tepat waktu.

Oleh karena itu perlu dilakukannya klasifikasi terhadap judul skripsi dengan menggunakan metode *naive bayes classifier* agar permasalahan tersebut dapat di minimalisir.

3.2. Konsep Solusi Permasalahan

3.2.1. Model Solusi Yang Di Tawarkan



Gambar 2 Model yang ditawarkan

Konsep implementasi yang diusulkan diberikan pada Gambar 2, dengan penjelasan sebagai berikut:

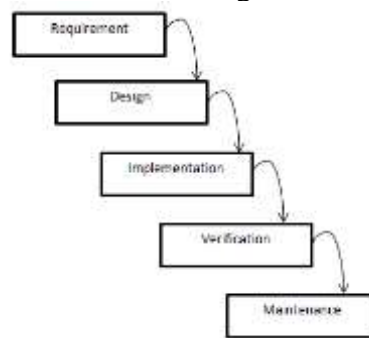
1. Sebagai masukan (*input*), Aplikasi membutuhkan:
 - a. Dokumen data latih berisi data judul skripsi yang sudah diklasifikasi secara manual oleh admin yang nantinya dijadikan acuan pembandingan dari data

yang akan dianalisis. Klasifikasi yang dimaksud sesuai atau tidak sesuai.

- b. Dokumen data uji, yaitu data target yang akan diklasifikasikan. Kata kata yang terkandung di kedua dokumen masukan dibersihkan dari kata-kata yang tidak berguna (atau *stop words*).
2. Proses ke-2 data uji bersih dibandingkan dengan data data latih, menggunakan *Naive Bayes Classifier*. Dari proses ke-2 ditampilkan data hasil klasifikasi judul yang berasal dari klasifikasi yang dilakukan oleh aplikasi. Kemudian dari hasil klasifikasi yang dilakukan aplikasi, dihitung prosentase data masing-masing kelas. Data klasifikasi yang dilakukan aplikasi akan disimpan pada dokumen apabila hasil klasifikasi sesuai.

3.2.2. Model Tahapan Pengembangan Sistem

Tahapan pengembangan sistem pada penelitian ini adalah sebagai berikut:



Gambar 3 Tahap Pengembangan Sistem
Sumber: Massey and satao , *Evolving a New Software Development Life Cycle Model (SDLC) incorporated with release Management*, (2012:1536)

3.2.3. Use Case

Pemodelan *use case diagram* pada digambarkan pada gambar 4:



Gambar 4 Use Case Diagram

4. HASIL DAN PEMBAHASAN

4.1. Hasil Pengujian Sistem

Setiap halaman pada sistem telah diuji, dengan cara memasukkan data pada setiap form input yang ada, baik berbeda karakter maupun nilai. Hasil yang keluar akan dicocokkan dengan hasil yang diharapkan. Berdasarkan pengujian sistem yang telah dilakukan, berikut adalah tampilan dari hasil pengujian.

a. Halaman Login

Halaman login digunakan untuk masuk ke sistem, maka Admin dan User harus login terlebih dahulu dengan melengkapi seluruh form yang ditampilkan.

Gambar 5. Halaman Login

Halaman Form Pengajuan

Menu proposal di bagian mahasiswa yang berfungsi sebagai halaman untuk pengajuan proposal. Mahasiswa akan mengisi form tersebut dan kemudian proposal tersebut akan otomatis tersimpan apabila form diisi dengan benar.

Berikut ini adalah form pendaftaran yang sudah diisi, dan bisa dikirim.

Gambar 6 Form pendaftaran

Halaman Perhitungan Naïve Bayes

Pada form ini, hasil perhitungan dari judul yang dimasukkan bisa dilihat bagaimana rekomendasi dari sistem terhadap judul tersebut, apakah sesuai dengan program studi atau tidak.

Gambar 7. Form Perhitungan Naïve Bayes

Gambar selanjutnya menampilkan bagaimana hasil perhitungan terhadap judul yang di masukkan.

Gambar 8 Form Hasil Perhitungan Naïve Bayes

Hasil Pengujian Akurasi dan Kecepatan

Berdasarkan hasil pengujian yang telah dilakukan berikut ini merupakan hasil

pengujian tingkat akurasi dan waktu eksekusi program

Tabel 2 Hasil Pengujian Naïve Bayes Terhadap Judul Skripsi Sistem Informasi

No	Judul skripsi	Akurasi
1	SISTEM PENUNJANG KEPUTUSAN DENGAN METODE AHP	51.77%
2	ANALISIS PERBANDINGAN METODE PREWITT DAN SOBEL PADA PROSES SEGMENTASI CITRA DIGITAL	51.18%
3	AUDIT DIGITAL LIBRARY MENGGUNAKAN FRAMEWORK COBIT 5	50.53%
4	CAREER DEVELOPMENT CENTRE ONLINE	50.46%
5	SISTEM INFORMASI GEOGRAFIS PARWISATA BERBASIS ANDROID	52.46%

Tabel 3 Hasil Pengujian Naïve Bayes Terhadap Judul Skripsi Teknik Informatika

No	Judul skripsi	Akurasi
1	APLIKASI SISTEM PAKAR DIAGNOSIS PENYAKIT TANAMAN JERUK METODE FORWARD CHAINING	52.64%
2	IMPLEMENTASI PRIVAT CLOUD STORAGE CIVITAS AKADEMIK	50.28%
3	PENERAPAN ALGORITMA ROUGH SET DALAM MENGANALISA PRESTASI BELAJAR MAHASISWA	50.48%
4	SISTEM PAKAR UNTUK MENDIAGNOSA PENYAKIT TANAMAN ADENIUM	51.05%
5	SECURITY ALARM RUMAHAN MENGGUNAKAN PERANGKAT RASBERRY	50.16%

Berdasarkan hasil pengujian yang telah dilakukan berikut ini merupakan hasil pengujian kecepatan waktu eksekusi program dibandingkan dengan menggunakan manual.

Tabel 4. Hasil Pengujian Perbandingan Waktu Eksekusi

No	Jenis Data Pengujian	Waktu (Menit) Mahasiswa	Waktu (Menit) Kaprodi
1	Pendaftaran (Menggunakan Sistem)	00:04:20	00:01:08
2	Pendaftaran (Konvensional)	00:60:00	00:60:00

Pengujian perbandingan yang telah dilakukan menghasilkan seperti pada tabel 4, dimana lama waktu pendaftaran dan pengklasifikasian dipengaruhi oleh sistem yang digunakan. Apabila menggunakan sistem konvensional tahap yang harus dilalui lumayan panjang yaitu mahasiswa harus mengambil formulir terlebih dahulu kemudian harus mengumpulkannya lagi dengan semua dokumen yang diperlukan. Dari kelemahan tersebut juga berdampak terhadap kinerja kaprodi dalam melakukan pengklasifikasian judul.

5. KESIMPULAN

Berdasarkan hasil proses yang telah dilakukan terhadap klasifikasi judul skripsi berdasarkan program studi dengan metode *Naïve Bayes* menggunakan aplikasi yang telah di rancang dapat disimpulkan bahwa, aplikasi mampu menganalisis judul yang telah dikumpulkan dalam dua kelas dengan tingkat akurasi antara 50% - 52%, dengan waktu yang dibutuhkan 2.10 – 2.20 detik. Aplikasi ini menyediakan fitur pendaftaran skripsi *online* yang didalamnya dapat menampilkan hasil judul yang direkomendasikan kepada setiap masing-masing program studi. Aplikasi juga memungkinkan data uji diset sesuai keinginan pengguna dalam jumlah data.

6. REFERENSI

Ahmad Fathan Hidayatullah. 2014. *Analisis Sentimen dan klasifikasi kategori tokoh*

- publik pada data twitter menggunakan naïve bayes classifier. Yogyakarta: Universitas Pembangunan Nasional "Veteran".
- AS., Rosa dan Shalahuddin, M. 2013. *Rekayasa Perangkat Lunak Terstruktur Dan Berorientasi Objek*. Bandung: Informatika.
- Ayub, Mewati. 2007. *Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer*. Jurnal Sistem Informasi Vol. 2 No. 1
- Tang Bo, dkk. 2014. *Toward Optimal Feature Selection in Naive Bayes for Text Categorization*. 10.1109/TKDE.2016.2563436, *IEEE Transactions on Knowledge and Data Engineering*, 1-15.
- Dhande, L. L., & Patnaik, G. K. 2014. *Analyzing Sentiment of Movie review Data using Naive Bayes Neural Classifier*. *IJETTCS*, 8.
- Gamallo P, Garcia M, 2014. *Citius: A Naive-Bayes Strategy for Sentiment Analysis on English Tweets*, <http://alt.qcri.org/semeval2014/cdrom/pdf/SemEval2014026.pdf>, diakses tgl 3 Oktober 2016 PM
- Hakim, Lukmanul. 2010. *Trik Rahasia Master PHP Terbongkar Lagi*. Yogyakarta : Lokomedia..
- Jurafsky, dkk "Semantic Role Labeling," [Online]. Available: http://www.deepsky.com/~merovech/voynich/voynich_manchu_reference_materials/PDFs/jurafsky_martin.pdf. [Accessed 4 November 2015].
- Kusrini, luthfi taufiq Emha, 2009. *Algoritma Data Mining*, Penerbit Andi, Yogyakarta. Yogyakarta: Penerbit Andi.
- Hakim, Lukmanul. 2010. *Trik Rahasia Master PHP Terbongkar Lagi*. Yogyakarta : Lokomedia.
- Masumi S . 2015. *Wikipedia-based Semantic Similarity Measurements for Noisy Short Texts Using Extended Naive Bayes*. *IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING, VOL. X, NO. X, XXX* (pp. 1-15). IEEE.
- Muin A. 2016. *Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)*. Jurnal ilmiah ilmu komputer, Vol 2, No 1
- Shaleh A. 2015. *Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga*. *Citec Journal* Vol, 2 No.3
- Sharma N & Singh M. 2016. *Modifying Naive Bayes Classifier for Multinomial Text Classification*. *IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2016)*, (pp. 1-7). Jaipur, India : IEEE.
- Prasetyo, E. . 2014. *Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta : ANDI.
- Kim S dkk. 2006. *Some Effective Techniques for Naive Bayes*. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 18, NO. 11*, (pp. 1-10). IEEE..Prenhalindo
- Liu Z dkk . 2015. *A Classification Method for Complex Power Quality Disturbances Using EEMD and Rank Wavelet SVM*. *IEEE TRANSACTIONS ON SMART GRID, VOL. 6, NO. 4*, (pp. 1-8). IEEE.

