

Penerapan Metode Naïve Bayes Classifier Untuk Penentuan Topik Tugas Akhir Pada Website Perpustakaan STIKOM Binaniaga

Kepin Sihotang^{1*}, Rajib Ghaniy²

¹ STIKOM Binaniaga/Teknik Informatika

Email: sihotang-kepin94@gmail.com

² STIKOM Binaniaga/Sistem Informasi

Email: rajob@stikombinaniaga.ac.id

ABSTRACT

STIKOM Binaniaga is one of the private universities in the city of Bogor, which organizes study programs in the field of Information Systems and Informatics Engineering. The library is an institution that manages collections of written works, printed works, and / or professional record works with a standard system to meet the needs of education, research, preservation, information, and recreation for visitors. This library system is used by all students of TI Stikom Binaniaga Bogor and by librarians, with the aim that students can see, know the reference to the title and topic of thesis thesis. Students majoring in IT can see references to the titles that have been set on the stikom library website system, and research applies the naive bayes classifier method for classification where the existing titles will be classified with related topic statements, thus the title data obtained previously is only students majoring in IT, because the data used for learning trials is only the title of alumni of the IT department in 2015-2017. With the data of IT titles, the Naive Bayes Classifier method can prove it. The results of the questionnaire distributed to lecturers and system experts showed very decent results for both respondents.

Keywords: naive bayes classifier, thesis title, thesis topic.

ABSTRAK

STIKOM Binaniaga adalah salah satu perguruan tinggi swasta di kota Bogor, yang menyelenggarakan program studi dibidang Sistem Informasi dan Teknik Informatika. Perpustakaan adalah institusi pengelola koleksi karya tulis, karya cetak, dan atau karya rekam secara profesional dengan sistem yang baku guna memenuhi kebutuhan pendidikan, penelitian, pelestarian, informasi, dan rekreasi para pemustaka. Sistem perpustakaan ini digunakan oleh seluruh mahasiswa TI Stikom Binaniaga Bogor dan oleh pustakawan, dengan tujuan agar mahasiswa dapat melihat, mengetahui referensi judul dan topik skripsi tugas akhir. Mahasiswa jurusan TI dapat melihat referensi judul-judul yang telah di tetapkan pada sistem website perpustakaan stikom, dan penelitian menerapkan metode naive bayes classifier untuk pengklasifikasian dimana judul yang ada akan terklasifikasi dengan pernyataan topik yang berhubungan, demikian data judul yang didapat sebelumnya hanya mahasiswa jurusan TI, karena data yang dijadikan untuk uji coba pembelajaran adalah hanya judul alumni mahasiswa jurusan TI stikom pada tahun 2015-2017. Dengan data judul-judul TI tersebut metode naive bayes classifier sudah dapat membuktikannya. Hasil dari kuesioner yang disebarkan kepada dosen dan ahli sistem menunjukkan hasil sangat layak pada kedua responden.

Kata Kunci: naive bayes classifier, judul skripsi, topik skripsi

PENDAHULUAN

A. Latar Belakang

Information Retrieval merupakan konsep pencarian informasi dari sumber yang tidak terstruktur. Menurut Manning (2009), *Information Retrieval* adalah menemukan dokumen

dari domain yang tidak terstruktur (bentuk teks) berupa kebutuhan informasi dari jumlah koleksi yang sangat banyak. Perkembangan jaringan dan selalu bertambahnya jumlah data yang tersedia, mengharuskan penerapan *Information Retrieval* dalam segala aspek. Salah satunya adalah perpustakaan. Komponen-komponen dalam *Information Retrieval* adalah *indexing*, pembobotan, dan perhitungan kemiripan *kosinus*. *Vector Space Model* (VSM) adalah sebagai metode yang mempunyai kelebihan mengukur kemiripan antara suatu dokumen dengan suatu *query user* dengan menggunakan *cosinus* dari sudut antar vektor yang dibentuk oleh dokumen dengan vektor dari kata kunci yang diinputkan oleh user. Salah satu kelemahan dari VSM adalah metode ini menganggap bahwa setiap term pada dokumen bersifat independen, yaitu metode ini tidak melihat hubungan makna dengan term lain. Sebagai contoh, apabila user melakukan pencarian dengan kata kunci “*programming*” maka hasil pencariannya adalah semua dokumen yang hanya memiliki kata “*programming*” saja, padahal masih banyak dokumen-dokumen yang masih berhubungan makna dengan kata “*programming*” seperti “*PHP*”, “*Java*”, dan lain-lain.

Metode untuk mengetahui kategori pada sebuah judul tugas akhir secara otomatis tanpa harus dikategorikan dibaca satu persatu maka perlu dilakukan pengukuran kemiripan dokumen terkait dengan menggunakan metode *Naive Bayes Classifier* (NBC) yaitu proses pengenalan teks dan dokumen. Seperti salah satu kasus terdahulu menggunakan metode tersebut untuk penelitian klasifikasi pengaduan melalui web. NBC adalah klasifikasi statistik yang bisa memprediksi probabilitas sebuah kelas, dan kelebihan dari metode ini adalah tingkat akurasi yang tinggi juga waktu komputasi yang lebih cepat. Peminatan adalah suatu keputusan yang dilakukan mahasiswa/i akademik untuk memilih topik atau tema sesuai minat, bakat, dan kemampuan selama mengikuti pembelajaran di kuliah.

Pemilihan peminatan dilakukan atas dasar kebutuhan untuk melanjutkan tugas akhir dan di sesuaikan dengan judul skripsi. Pada penelitian yang dilakukan pada tugas akhir ini, akan dibuat suatu instrument yang membantu dalam penelitian penyusun yaitu suatu program pencarian dokumen atau data teks melalui *query* tertentu. Program ini akan menerapkan salah satu metode yaitu metode text mining dan metode *naive bayes classifier*. *Naive Bayes Classifier* (NBC) merupakan sebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan (*independent*) yang tinggi. Perpustakaan adalah mencakup suatu ruangan, bagian dari gedung / bangunan atau gedung tersendiri yang berisi buku-buku koleksi, yang diatur dan disusun demikian rupa, sehingga mudah untuk dicari dan dipergunakan apabila sewaktu-waktu diperlukan oleh pembaca.

Metode *text mining* merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar. Langkah-langkah yang dilakukan dalam *text mining* adalah, *Text Preprocessing* tindakan yang dilakukan pada tahap ini adalah *toLowerCase*, yaitu mengubah semua karakter huruf menjadi huruf kecil, dan *Tokenizing* yaitu proses penguraian deskripsi yang semula berupa kalimat – kalimat menjadi kata-kata dan menghilangkan *delimiter-delimiter* seperti tanda titik(.), koma(,), spasi dan karakter angka yang ada pada kata tersebut. Ada pun aplikasi pencarian buku dan lain-lain yang berbasis web, seperti contoh yang ada pada kampus STIKOM BINANIAGA. Ada beberapa penelitian sebelumnya yang menerapkan metode text mining dan *naive bayes classifier* salah satunya adalah “KLASIFIKASI ARTIKEL BERITA MENGGUNAKAN METODE TEXT MINING DAN NAIVE BAYES CLASSIFIER” (Ira Anggraeni Setiawan, Tacbir Hendro , Dian Nursantika, 2017). Penelitian tersebut menerapkan metode *text mining* dan *naive bayes classifier* telah menghasilkan sistem klasifikasi artikel berita terhadap sembilan kategori atau kanal yaitu *news*, *finance*, *sport*, *otomotif*, *entertain*, *healty*, *food*, *travel* dan tekno. Sistem klasifikasi artikel berita yang dilakukan dengan menggunakan algoritma *naive bayes classifier* menerima data masukkan berupa data *text* artikel yang diproses dengan *text mining* yaitu proses *casefolding*, *tokenizing* dan *filtering*. Setelah didapat kata kunci dari proses *text mining* akan dilakukan perhitungan dengan *naive bayes classifier* yang menghasilkan keluaran berupa artikel yang sudah terkategori.

Berdasarkan referensi di atas, maka dari itu metode *text mining* dan *naive bayes classifier* akan dicoba untuk di terapkan sebagai penelitian dengan judul “PENERAPAN METODE NAIVE BAYES CLASSIFIER UNTUK PENENTUAN TOPIK TUGAS AKHIR PADA WEBSITE PERPUSTAKAAN STIKOM BINANIAGA” untuk referensi pencarian judul penelitian tugas akhir berdasarkan topik menggunakan *framework codeigniter* dan di harapkan mahasiswa dapat mengetahui dengan kecocokan judul tugas akhir berdasarkan topik.

B. Permasalahan

Berdasarkan latar belakang yang telah diuraikan, maka identifikasi masalah dalam penelitian ini adalah :

1. Hasil pencarian judul skripsi di website e-library Stikom Binaniaga belum menampilkan topik yang sesuai.
2. Hasil pencarian judul belum dikelompokkan berdasarkan topik penelitian atau belum terklasifikasi dengan judul penelitian.

C. Landasan Teori

1. Naïve Bayes Classifier (NBC)

Salah satu penerapan theorem bayes dalam klasifikasi adalah Naïve Bayes.

$$V_{map} = \operatorname{argmax}_{v_j \in V} P(V_j | a_1, a_2, a_3, \dots, a_n)$$

$$V_{map} = \operatorname{argmax}_{v_j \in V} \frac{P(a_1, a_2, a_3, \dots, a_n | V_j) P(V_j)}{P(a_1, a_2, a_3, \dots, a_n)}$$

Naïve bayes berdasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output, atau dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu atau $(a_1, a_2, a_3, \dots, a_n | v_j) = \pi_i p(a_i | v_j)$. Memasukan persamaan ini ke $n_{NB} = \operatorname{argmax}_{v_j \in V} P(V_j) \prod P(a_i | v_j)$ maka akan didapat pendekatan yang dipakai dalam naïve bayes classifier dimana n_{NB} adalah nilai output hasil klasifikasi.

2. Text Mining

Text Mining merupakan penerapan konsep dari teknik data mining untuk mencari pola dalam teks, bertujuan untuk mencari informasi yang bermanfaat dengan tujuan tertentu. Berdasarkan ketidakteraturan struktur data teks, maka proses text mining memerlukan beberapa tahap awal yang pada intinya mempersiapkan agar teks dapat diubah menjadi lebih terstruktur (Budi Susanto, Teknik Informatika UKDW Yogyakarta). Proses text mining dibagi menjadi 3 tahap utama, yaitu proses awal terhadap teks (text preprocessing), transformasi teks (text transformation), dan penemuan pola (pattern discovery)(Even, Yahir Zohar, 2002).

a. Text Preprocessing

Pada text mining, struktur data yang baik dapat mempermudah proses komputerisasi secara otomatis. Maka dari itu, diperlukan beberapa tahapan untuk pengubahan dari informasi yang strukturnya sembarang menjadi lebih terstruktur sesuai dengan kebutuhan. Tahapan awal dari text mining adalah text preprocessing yang bertujuan untuk mempersiapkan teks menjadi data yang terstruktur dan dapat diproses pada tahapan berikutnya.

b. Case Folding

Tahap Case Folding adalah Mengubah seluruh huruf dari “a” sampai dengan “z” dalam dokumen menjadi huruf kecil. Tidak semua dokumen konsisten dengan penggunaan huruf kapital. Maka dari itu Case folding Mengkonversi keseluruhan teks dalam dokumen menjadi Huruf kecil.

- c. *Tokenizing*
Tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Dalam tokenizing juga dapat membuang beberapa karakter yang dianggap sebagai tanda baca.
- d. *Filtering*
Tahap filtering adalah tahap mengambil kata-kata penting dari hasil tokenizing menggunakan algoritma stopword removal (membuang kata yang kurang penting). Stopword adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-word.
- e. *Stemming*
Tahap stemming adalah tahap mencari root kata dari tiap kata hasil filtering. Pada tahap ini dilakukan proses pengembalian berbagai bentuk kata kedalam suatu representasi yang sama.
- f. *TF(Term Frequency)*
Term frequency merupakan salah satu metode untuk menghitung bobot tiap term dalam teks. Dalam metode ini, tiap term diasumsikan memiliki nilai kepentingan yang sebanding dengan jumlah kemunculan term tersebut pada teks (Mark Hall & Lloyd Smith, 1999).
- g. *DF(Document Frequency)*
Document frequency adalah jumlah dokumen yang mengandung suatu *term* tertentu. Tiap *term* akan dihitung nilai *document frequency*-nya (DF). Lalu *term* tersebut diseleksi berdasarkan jumlah nilai DF. Jika nilai DF berada dibawah *threshold* yang telah ditentukan, maka *term* akan dibuang. Asumsi awalnya adalah bahwa *term* yang lebih jarang muncul tidak memiliki pengaruh besar dalam proses pengelompokan dokumen. Pembuangan *term* yang jarang ini dapat mengurangi dimensi fitur yang besar pada *text mining*.
- h. *IDF(Invers Document Frequency)*
Invers document frequency (IDF) yaitu pengurangan dominasi *term* yang sering muncul di berbagai dokumen. Hal ini diperlukan karena *term* yang banyak muncul di berbagai dokumen, dapat dianggap sebagai *term* umum (*common term*) sehingga tidak penting nilainya. Sebaliknya faktor kejarang muncul kata (*term scarcity*) dalam koleksi dokumen harus diperhatikan dalam pemberian bobot.

METODE PENGEMBANGAN

Metode penelitian merupakan cara yang dilakukan dalam penelitian untuk mendapatkan data dengan sebuah hasil yang diharapkan. Dalam penelitian ini metode yang digunakan adalah eksperimen, artinya bahwa penelitian yang dilakukan untuk melakukan uji coba terhadap permasalahan tertentu dengan penggunaan teori tertentu sehingga didapatkan hasil pengujian yang tepat antara permasalahan yang diambil dengan teori yang digunakan.

Menurut Pressman (2012:50), dalam melakukan perancangan sistem yang akan dikembangkan dapat menggunakan metode prototype. Metode ini cocok digunakan untuk mengembangkan sebuah perangkat yang akan dikembangkan kembali. Metode ini dimulai dengan pengumpulan kebutuhan pengguna, dalam hal ini pengguna dari perangkat yang dikembangkan adalah peserta didik. Kemudian membuat sebuah rancangan kilat yang selanjutnya akan dievaluasi kembali sebelum dikembangkan secara benar. Prototype bukanlah merupakan sesuatu yang lengkap, tetapi sesuatu yang harus dievaluasi dan dimodifikasi kembali. Segala perubahan dapat terjadi pada saat prototype dibuat untuk memenuhi kebutuhan pengguna dan pada saat yang sama memungkinkan pengembang untuk lebih memahami kebutuhan pengguna secara lebih baik. Kelebihan dari pembuatan prototype yaitu :

1. Pengguna berpartisipasi aktif dalam pengembangan sistem, sehingga hasil produk pengembangan akan semakin mudah disesuaikan dengan keinginan dan kebutuhan pengguna.
2. Penentuan kebutuhan lebih mudah diwujudkan

3. Mempersingkat waktu pengembangan produk perangkat lunak.
4. Adanya komunikasi yang baik antara pengembang dan pengguna.
5. Pengembang dapat bekerja lebih baik dalam menentukan kebutuhan pengguna.
6. Lebih menghemat waktu dalam pengembangan sistem.
7. Penerapan menjadi lebih mudah karena pengguna mengetahui apa yang diharapkannya.

Menurut Pressman (2010), seringkali pelanggan mendefinisikan satu set tujuan umum untuk perangkat lunak, tetapi tidak mengidentifikasi persyaratan rinci untuk fungsi dan fitur. Di lain kasus, pengembang mungkin tidak yakin dari efisiensi dari sebuah algoritma, adaptasi dari sistem operasi, atau bentuk yang interaksi manusia-mesin harus ambil dan situasi lain. Dalam hal ini, paradigma prototype mungkin menawarkan pendekatan yang terbaik. Model prototype ini memungkinkan adanya interaksi antara pengembang sistem dengan pengguna sistem nantinya, sehingga dapat mengatasi ketidak serasian antar pengembang dan pengguna. Bagian mengenai prototype model dapat dilihat pada gambar 1.



Gambar 1. Model Prototype

HASIL DAN PEMBAHASAN

Hasil

A. Penerapan Metode Naïve Bayes Classifier

Metode Naive Bayes Classifier ini di terapkan pada website perpustakaan STIKOM, untuk penentuan topik tugas akhir dan pengklasifikasian judul dan topik yang berhubungan. Pada proses pembelajaran, data judul skripsi jurusan TI yang sudah dimasukkan akan dipisahkan berdasarkan kategori yang telah ditentukan. Kemudian data yang sudah dipisahkan masuk ketahap preprocessing (casefolding, tokenizing dan filtering). Setelah itu dihitung nilai probabilitas kata dan probabilitas kategori pada setiap data yang dijadikan data pembelajaran, kemudian disimpan ke dalam database pembelajaran yang berisi kata – kata penting pada setiap kategori.

Sedangkan pada proses pengujian yang menjadi data masukkan yaitu data judul skripsi jurusan TI baru yang belum diketahui kategorinya. Pada tahap preprocessing yang dilakukan sama seperti yang dilakukan pada proses preprocessing di proses pembelajaran, yang membedakannya yaitu pada saat perhitungan probabilitas kata. Setelah melakukan tahap preprocessing, maka dokumen baru tersebut akan melalui proses persamaan kata. Kata – kata yang ada di data baru dengan kata – kata yang ada di database pembelajaran. Gunanya untuk menghitung probabilitas kata yang sama pada database. Sehingga menghasilkan probabilitas pada setiap kategori yang ada.

B. Rancangan Database

Sistem yang dibuat untuk penelitian ini menggunakan beberapa tabel namun yang di tambahkan pada penelitian ini ada 10 tabel, karena table keseluruhan yang ada pada website perpustakaan itu memiliki fungsi yang berbeda, namun pada penelitian ini ada beberapa penambahan tabel yang di terapkan pada website tersebut, hanya berfungsi untuk perancangan dimana penambahan tabel tersebut didalamnya adalah fungsi perhitungan metode naive bayes classifier.

Pembahasan

A. Hasil Perhitungan Naïve Bayes Classifier

1. Langkah Pertama

Dibuatkan rincian mengenai judul skripsi yang sudah masuk kedalam database perpustakaan dan sesuaikan topik dari judul tersebut. Dimana judul yang ada akan dipisahkan dengan pengabilan kata per kata, kecuali kata tersebut harus yang bukan kata penyambung kalimat.

2. Langkah Kedua

Setelah penseleksian kata maka, langkah berikutnya adalah memberi nilai setiap kata dengan menggunakan rumus berikut :

$$P(W_k|V_j) = (nk+1) / (Jml\ Frekuensi+Jml\ judul).$$

Dimana :

$P(W_k | V_j)$: Probabilitas bobot kata sesuai Topik

nk : Nilai kemunculan frekuensi kata

NETWORKING

$$P(\text{ simple network management protocol (SNMP)}|\text{Networking}) = (1+1)/(11+19) = 0,06667$$

$$P(\text{ monitoring}|\text{Networking}) = (2+1)/(11+19) = 0,1$$

$$P(\text{ IP}|\text{Networking}) = (1+1)/(11+19) = 0,06667$$

$$P(\text{ link}|\text{Networking}) = (1+1)/(11+19) = 0,06667$$

$$P(\text{ bandwidth}|\text{Networking}) = (1+1)/(11+19) = 0,06667$$

$$P(\text{ traffic shaping}|\text{Networking}) = (1+1)/(11+19) = 0,06667$$

$$P(\text{ QoS (Quality Of Service)}|\text{Networking}) = (1+1)/(11+19) = 0,06667$$

$$P(\text{ Internet}|\text{Networking}) = (1+1)/(11+19) = 0,06667$$

$$P(\text{ network}|\text{Networking}) = (2+1)/(11+19) = 0,1$$

$$P(\text{ avaiability }|\text{Networking}) = (1+1)/(11+19) = 0,06667$$

Wireless

$$P(\text{ captive portal }|\text{Wireless}) = (1+1)/(4+19) = 0,08696$$

$$P(\text{ radius }|\text{Wireless}) = (1+1)/(4+19) = 0,08696$$

$$P(\text{ authentication }|\text{Wireless}) = (1+1)/(4+19) = 0,08696$$

$$P(\text{ wireless }|\text{Wireless}) = (1+1)/(4+19) = 0,08696$$

Cryptography

$$P(\text{ File }|\text{Cryptography}) = (1+1)/(6+19) = 0,0800$$

$$P(\text{ Drawing }|\text{Cryptography}) = (1+1)/(6+19) = 0,0800$$

$$P(\text{ algoritma AES (Advanced Encryption Standard) }|\text{Cryptography}) = (1+1)/(6+19) = 0,0800$$

$$P(\text{ Kompresi }|\text{Cryptography}) = (1+1)/(6+19) = 0,0800$$

$$P(\text{ Algoritma Vigenere Chiper-Huffman }|\text{Cryptography}) = (1+1)/(6+19) = 0,0800$$

$$P(\text{ Enkripsi }|\text{Cryptography}) = (1+1)/(6+19) = 0,0800$$

Social Computing

$$P(\text{ Near Field Communication (NFC)}|\text{Social Computing}) = (1+1)/(4 +19) = 0,08696$$

$$P(\text{ JSON}|\text{Social Computing}) = (1+1)/(4 +19) = 0,08696$$

$$P(\text{ protokol }|\text{Social Computing}) = (1+1)/(4 +19) = 0,08696$$

$$P(\text{ web service}|\text{Social Computing}) = (1+1)/(4 +19) = 0,08696$$

Data Mining

$$P(\text{ naive bayes classifier}|\text{Data Mning}) = (1+1)/(1+19) = 0,1$$

Selanjutnya setiap kata yang telah bernilai, maka seluruh kata tersebut akan diperhitungkan nilai probabilitasnya dengan menggunakan rumus :

$$P(V_j) = \text{Jumlah dokumensetiap kategori} / \text{Total dokumen}$$

Diketahui : Topik Networking = 2
 Topik Cryptography = 1
 Topik Social Computing = 1
 Topik Data Mining = 1

Jadi Probabilitas adalah :

$P(\text{Networking}) = 4/8 = 0,25$
 $P(\text{Cryptography}) = 1/8 = 0,125$
 $P(\text{Social Computing}) = 1/8 = 0,125$
 $P(\text{Data Mining}) = 1/8 = 0,125$

B. Proses Pengujian

Contoh judul Skripsi baru yang belum diketahui kategori atau Topiknya. Contoh judul Skripsi : “PENERAPAN METODE NAIVE BAYES CLASSIFIER UNTUK PENENTUAN TOPIK TUGAS AKHIR PADA WEBSITE PERPUSTAKAAN STIKOM BINANIAGA“. Setelah melalui proses text mining yaitu preprocessing, maka kata – kata yang didapat yaitu : “ NAIVE BAYES CLASSIFIER”. Seperti pada Tabel 1 berikut

Tabel 1. Frekwensi Kata

Kata	Frekwensi
Naïve Bayes Classifier	Jumlah kata 2 yaitu Naïve bayes dan Classifier

Setelah terdapat frekuensi nya lalu di cari nilai terbesar dengan perbandingan masing-masing topik, sehingga penentuan topik akan di ketahui jika setelah di hitung berikut ini untuk mencari nilai terbesar dari masing-masing topik di tentukan dengan nilai tinggi yang didapat, atau hasil dari perhitungan yang dengan cara di berikut ini:

Data mining	Networking
$P(\text{naive bayes classifier} \text{Software Engineering})= 0,1$	$P(\text{naive bayes classifier} \text{Networking})= 0,00001$
Jadi $P(\text{Data mining}) = 0,1$	Jadi $P(\text{Networking})= 0,00001$
$= 0,1$	$=0,00001$
Probabilitas $= P(\text{Data mining})*(P \text{Data mining})$	Probabilitas $= P(\text{Networking})*(P \text{Networking})$
$= 0,125*0,1$	$= 0,25*0,00001$
$= 0,0125$	$= 0,0000025$

Cryptogrhapny
 P(naive bayes classifier|Cryptogrhapny)=0,00001
 Jadi P(|Cryptogrhapny) = 0,00001
 = 0,00001
 Probabilitas = P(Cryptogrhapny)*(P | Cryptogrhapny)
 = 0,125*0,00001
 = 0,00000125

Social Computing
 P(naive bayes classifier|Social Computing)=0,00001
 Jadi P(|Social Computing) =0,00001
 = 0,00001
 Probabilitas = P(Social Computing)*(P | Social Computing)
 = 0,125*0,00001
 = 0,00000125

Jadi kategori dari dokumen artikel berita baru itu termasuk Kategori Topik Data Mining karena memiliki probabilitas paling tinggi yaitu = 0,000385864

C. Hasil Uji Kelayakan

1. Ahli Sistem Informasi

Bobot skor yang digunakan untuk perhitungan pada kuesioner kelayakan sistem untuk ahli sistem ini menggunakan skala likert seperti yang sudah ditunjukkan pada Tabel 2

Tabel 2. Hasil Uji Ahli Sistem Informasi

Aspek Penilaian	Indikator	R1	R2
Komponen Input	User Interface	5	4
	Interaksi Sistem	5	4
Komponen Model	Prosedur Sistem	5	4
	Logika Program	5	4
Komponen Output	Ketepatan Informasi	5	4
	Kekinian Informasi	4	4
Komponen Teknologi	Waktu Respon	4	4
	Keluwesannya Sistem	4	4
Komponen Kontrol	Keamanan Sistem	4	4
Komponen Basis Data	Konten Basis Data	5	4
Jumlah		46	40
Nilai Tertinggi		50	50

$$\text{Persentase Kelayakan (\%)} = \frac{\text{Skor yang diobservasi}}{\text{Skor yang diharapkan}} \times 100\%$$

$$\text{Persentase Kelayakan (\%)} = \frac{46 + 40}{50 + 50} \times 100\%$$

$$\text{Persentase Kelayakan (\%)} = \frac{86}{100} \times 100\%$$

$$\text{Persentase Kelayakan (\%)} = 86\%$$

2. Uji Kelayakan Pengguna

$$\text{Persentase Kelayakan (\%)} = \frac{\text{Skor yang diobservasi}}{\text{Skor yang diharapkan}} \times 100\%$$

$$\text{Persentase Kelayakan (\%)} = \frac{75 + 63 + 78 + 68 + 71 \dots}{100 + 100 + 100 + 100 + \dots} \times 100\%$$

$$\text{Persentase Kelayakan (\%)} = \frac{2.329}{3000} \times 100\%$$

$$\text{Persentase Kelayakan (\%)} = 77\%$$

SIMPULAN DAN SARAN

A. Simpulan

Berdasarkan hasil penelitian yang telah diuraikan, maka dapat ditarik kesimpulan :

1. Metode Naive Classifier dapat digunakan untuk penentuan dan menampilkan topik jurusan TI pada judul yang di ajukan.
2. Ketepatan dan menemukan dalam menentukan topik pada data judul skripsi jurusan TI baru dipengaruhi oleh data pembelajaran atau data latih pada setiap kategori. Data latih ini berisi kata – kata yang sering muncul pada masing – masing kategori atau kata – kata yang dapat mewakili kategori tertentu.

B. Saran

Penentuan topik judul skripsi ini hanya di khususkan dengan judul jurusan TI dan topik jurusan TI yang ada, jika tidak tepat dalam penentuan maka kesalahannya adalah data atau pengajuan yang ingin di tentukan topik nya tidak sesuai dengan judul jurusan TI begitu juga tidak tersedia nya data yang bukan jurusan TI, karena di penelitian ini data yang di gunakan hanya khusus judul jurusan TI pada tahun 2015-2017.

Saran untuk kedepan nya data yang digunakan tidak hanya judul skripsi TI berikut juga dengan judul jurusan SI, adapun sistem ini di harapkan dapat dijadikan rekomendasi sistem atau recommended system.

DAFTAR RUJUKAN

- Ardhana, YM Kusuma. 2013. Pemrograman PHP : CodeIgniter Black Box. Jasakom. Jakarta.
- Arikunto, Suharsimi. (2009). Manajemen Penelitian, edisi Revisi. Jakarta : Rineka Cipta.
- Even, Yahir Zohar. Santoso, Budi. 2002. Text mining. Teknik Informatika UKDW Yogyakarta. <https://repository.widyatama.ac.id/xmlui/bitstream/handle/123456789/5867/Bab%202.pdf?sequence=9> 04[Di akses tgl 04 Maret 2018]
- George zipf 1999 IDF(Invers Document Frequency)
- Kurniawan, Bambang. Effendi Syahril. Salim, Opim Sitompul. 2012. Klasifikasi Konten Berita Dengan Metode Text Mining. Universitas Sumatera Utara
- Mas`udia, Putri Elfa. Atmadja, Martono Dwi. Mustafa, Lis Diana. 2017. Information Retrieval Tugas Akhir Dan Perhitungan Kemiripan Dokumen Mengacu Pada Abstrak Menggunakan Vector Space Model. Politeknik Negeri Malang
- Pressman, RS. (2009). Software Engineering A Practitioner's Approach 7th Ed - Roger S. Pressma.
- Pressman, RS. (2012). Rekayasa Perangkat Lunak. Yogyakarta: Andi.
- Pressman, RS. (2016). Manajemen Operasi dan Rantai Pasok, Operations and Supply Chain Management Edisi 14. Salemba Empat.

- Sani, Ramadhan Rakhmat. Zeniarja, Junta. Luthfiarta, Ardytha. 2016. Penerapan Algoritma K-Nearest Neighbor pada Information Retrieval dalam Penentuan Topik Referensi Tugas Akhir. Universitas Dian Nuswantoro
- Salton. G, 1989 IDF(Invers Document Frequency)
- Setiawan, Ira Anggraeni. Hendro, Tachir P. Nursantika Dian. 2017. Klasifikasi Artikel Berita Menggunakan Metode Text Mining Dan Naive Bayes Classifier. Universitas Jenderal Achmad Yani
- Santoso, Budi. 2007. Data mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Graha Ilmu: Yogyakarta
- Sugiyono. (2010). Metode Penelitian Pendidikan, Pendekatan Kuantitatif. BandungWahyudi, Ikhsan Pratama. Wuriyanto, Tutut. Sulistiowati. 2016.
- Rancang Bangun Aplikasi Meningkatkan Tingkat Relevansi Pencarian Tugas Akhir Studi Kasus Pada Perpustakaan Institut Bisnis Dan Informatika Stikom Surabaya. STMIK STIKOM Surabaya
- Witten, 1999 IDF(Invers Document Frequency)
- Yang Yiming & Jan O. Pedersen, 1997 DF(Document Frequency)
- YasirZain, Muhammad. Suswati. 2016. Information Retrieval System Pada Pencarian File Dokumen Berbasis Teks Dengan Metode Vector Space Model Dan Algoritma ECS Stemmer. Universitas Madura