

**PENGUNAAN ALGORITMA NAÏVE BAYES PADA KLASIFIKASI
JUDUL PROYEK AKHIR BERDASARKAN KELOMPOK
BIDANG KOMPETENSI (KBK)**

Dini Nurmalasari¹, Heri R Yuliantoro², Saleha Indri Yanti³

¹ Program Studi Teknik Rekayasa Komputer, Politeknik Caltex Riau

² Program Studi Akuntansi Perpajakan, Politeknik Caltex Riau

³ Program Studi Teknik Informatika, Politeknik Caltex Riau,

Email: dini@pcr.ac.id

ABSTRAK

Proyek Akhir (PA) merupakan salah satu syarat kelulusan yang harus dipenuhi oleh mahasiswa Politeknik Caltex Riau (PCR). Saat ini proses PA dilakukan melalui sistem akademik, mulai dari pengajuan pengajuan judul, penentuan pembimbing dan penguji, log bimbingan, persetujuan seminar dan sidang, sampai dengan proses penilaian. Pada saat proses pengajuan judul, mahasiswa diharuskan menginputkan judul, deskripsi, referensi serta memilih judul tersebut masuk dalam kategori bidang kompetensi (KBK) apa. Setelah judul masuk, maka dosen pada KBK tersebut akan melakukan review kelayakan dari judul tersebut. Terdapat masalah ketika KBK yang dipilih mahasiswa pada saat pengajuan judul ternyata salah, sehingga proses review akan terhambat karena tidak sesuai dengan kompetensi dosen pada KBK tersebut. Dengan adanya teknologi pengolah teks seperti teks mining, hal tersebut dapat diatasi, dengan cara membuat aplikasi yang dapat secara otomatis melakukan pengelompokan judul PA dan memasukan ke dalam KBK yang seharusnya.

Penelitian ini mengimplementasikan penggunaan Algoritma Naive Bayes untuk melakukan klasifikasi judul PA secara otomatis berdasarkan KBK yang sesuai. Sistem juga dapat mendeteksi persentase kemiripan judul yang diinput dengan judul yang sudah ada pada database menggunakan *Cosine Similarity*, sehingga dapat meminimalisir adanya penelitian yang sama. Hasil pengujian menunjukkan sistem klasifikasi menggunakan algoritma Naïve Bayes dapat memberikan informasi terkait KBK dari judul yang di input dengan nilai akurasi sebesar 90.88%.

Kata kunci : Judul Proyek Akhir, Text Mining, Naive Bayes, Cosine Similarity.

ABSTRACT

The Final Project (PA) is one of the graduation requirements that must be met by Caltex Riau Polytechnic (PCR) students. Currently the PA process is carried out through the academic system, starting from submitting submissions for titles, determining supervisors and examiners, guidance logs, approval of seminars and hearings, up to the assessment process. During the title submission process, students are required to input a title, description, reference and choose which category the title falls into. After the title is entered, the lecturer at the KBK will review the feasibility of the title. There is a problem when the CBC that students choose when submitting the title turns out to be wrong, so the review process will be hampered because it does not match the competence of the lecturer in the KBK. With the existence of text processing technology such as text mining, this can

be overcome, by creating an application that can automatically group PA titles and enter them into the proper KBK.

This study implements the use of the Naive Bayes Algorithm to classify PA titles automatically based on the appropriate CBC. The system can also detect the percentage of similarity of the title entered with the existing title in the database using Cosine Similarity, so that it can minimize the existence of the same research. The test results show that the classification system using the Naïve Bayes algorithm can provide information related to the KBK from the title entered with an accuracy value of 90.88%.

Keywords: Final Project Title, Student, Text Mining, Naive Bayes, Cosine Similarity.

1. PENDAHULUAN

Mahasiswa Jurusan Teknologi Informasi (JTI) Politeknik Caltex Riau diwajibkan untuk mengumpulkan judul proyek akhir sebagai salah satu persyaratan kelulusan mahasiswa. Mahasiswa JTI terdiri dari tiga Program Studi yaitu Teknik Informatika, Teknologi Rekayasa Komputer, dan Sistem Informasi. Setiap tahunnya jumlah mahasiswa JTI di Politeknik Caltex Riau semakin bertambah, sehingga mengakibatkan judul proyek akhir semakin banyak dan beragam. Judul proyek akhir yang diajukan mahasiswa akan diperiksa oleh koordinator proyek akhir masing-masing Kelompok Bidang Kompetensi (KBK), untuk direview apakah judul proyek akhir diterima atau tidak. Saat pengumpulan judul proyek akhir, mahasiswa akan diberikan pilihan untuk memilih KBK dari judul proyek akhir yang akan diajukan. Pilihan KBK ini dipilih sendiri oleh mahasiswa tersebut, sehingga muncul permasalahan dimana judul proyek akhir tidak sesuai dengan KBK atau salah KBK. Hal ini menyebabkan judul proyek akhir membutuhkan proses *review* yang lebih lama.

Berdasarkan permasalahan di atas, maka akan dibangun sebuah sistem yang akan mengelompokkan judul Proyek Akhir berdasarkan KBK secara otomatis dengan menggunakan deskripsi judul proyek akhir. Hal ini dikarenakan pada bagian deskripsi proyek akhir terdapat banyak informasi yang disampaikan berhubungan dengan proyek akhir yang akan dibuat. Pengelompokkan judul proyek akhir dilakukan untuk mempermudah dalam mengetahui jenis KBK dari setiap judul proyek akhir. Berhubung data judul proyek akhir mahasiswa volumenya terus bertambah, ilmu yang sesuai untuk digunakan adalah *text mining*. Penerapan *text mining* dapat digunakan untuk membantu dalam menganalisa data yang diperoleh dari deskripsi singkat PA mahasiswa. Untuk menerapkan *text mining*, dibutuhkan deskripsi PA mahasiswa yang sebelumnya sehingga dapat menggali pola-pola yang dapat dijadikan pengetahuan baru. Selain itu, mahasiswa juga dapat mengetahui persentasi kemiripan judul yang akan diajukan dengan judul-judul yang sudah pernah ada sebelumnya. Sehingga mencegah terjadinya penginputan judul yang sama.

Text mining dapat didefinisikan sebagai suatu proses menggali informasi dimana seorang *user* berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis yang merupakan komponen-komponen dalam *data mining* yang salah satunya adalah kategorisasi (Feldman & Sanger, 2007). Dalam proyek akhir ini, teknik yang digunakan untuk mengelompokkan judul adalah menggunakan teknik klasifikasi. Dengan menggunakan klasifikasi, judul mahasiswa nantinya akan diklasifikasikan kedalam KBK dari masing-masing program studi JTI yang ada di PCR.

Klasifikasi adalah bentuk analisis data yang dapat digunakan untuk menggambarkan kelas data penting atau untuk memprediksi tren data masa depan (Han & Kamber, 2006). Teknik klasifikasi yang akan digunakan pada sistem ini adalah *Naive*

Bayes. Tujuan dari klasifikasi adalah untuk memprediksi secara akurat kelas dari setiap kasus data.

Selain klasifikasi berdasarkan KBK, sistem juga mampu mengecek kemiripan judul yang diajukan Mahasiswa JTI dengan judul yang sebelumnya pernah ada pada *database* sistem. Teknik yang digunakan adalah *Cosine Similarity*. *Cosine Similarity* adalah salah satu metode dalam menentukan nilai kemiripan antar dua objek. *Cosine Similarity* dapat diterapkan dalam menentukan nilai kemiripan pada dua berkas dokumen teks. Parameter yang digunakan adalah jumlah kata-kata pada dua dokumen teks yang dibandingkan. Teknik ini akan menghasilkan *output* berupa persentase kemiripan antara 2 judul proyek akhir atau lebih yang ada pada *database* sistem.

2. METODOLOGI PENELITIAN

Pada penelitian ini dilakukan metodologi penelitian dengan tahapan sebagai berikut :

a. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data judul proyek akhir mahasiswa Politeknik Caltex Riau Jurusan Teknologi Informasi dari tahun 2015 sampai dengan tahun 2019 dengan format .xls yang diperoleh dari BSTI. Data ini digunakan sebagai data training dan data testing pada penelitian penulis.

b. Analisis dan Perancangan

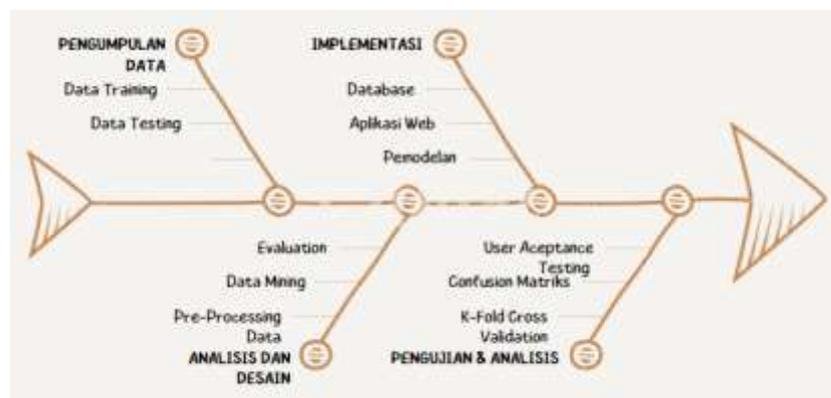
Pada tahapan ini dilakukan analisis kebutuhan pengguna serta perancangan data dan perancangan proses. Pada tahapan ini juga dilakukan analisis perhitungan Naïve Bayes secara manual, serta dilakukan pembuatan algoritma naïve bayes berdasarkan kebutuhan data. Pada tahap ini juga dilakukan proses KDD (*Knowledge Data Discovery*), dengan tahapan : *Selection, Pre-Processing Data, Transformation, Data Mining (Klasifikasi)* dan *Evaluation*.

c. Implementasi

Pada tahap ini dilakukan implementasi algoritma Naïve Bayes untuk pengelompokan judul PA dengan kelompok KBK yang terdiri dari : *Software Engineering (SE), Business Analyst (BA), Data Engineering (DE), Soft Computing (SC), Multimedia (Mu), dan Operating System Computer Network (OSCN)*. Pada tahapan ini dibangun aplikasi berbasis web dengan menggunakan bahasa pemrograman PHP, *Framework CodeIgniter* dan MySQL sebagai *database*.

d. Pengujian dan Analisis

Pengujian yang dilakukan terdiri dari pengujian kemiripan dokumen dengan *K-Fold Cross Validation, Confusion Matriks* untuk menghitung akurasi dan laju error, serta pengujian *User Acceptance Testing (UAT)*.



Gambar 1. Fishbone Diagram Metodologi Penelitian

3. TINJAUAN PUSTAKA

3.1. Studi Literatur

Penelitian pertama adalah penelitian yang dilakukan oleh (Firmansyah, 2015) yang berjudul implementasi text mining klasifikasi objek wisata dengan metode Naive Bayes Classifier di dinas pariwisata dan kebudayaan Jawa Barat. Penelitian ini mengelompokkan berdasarkan 8 kategori objek wisata yaitu Karakter Hutan, Karakter Pantai, Karakter Religius, Karakter Tradisi dan Komunikasi Adat, Karakter Kepurbakalaan, Karakter Gelar Seni & Ragam Festival, Karakter Heritage, dan Karakter Wisata Belanja. Pengklasifikasian saat ini dilakukan melalui analisa terhadap kemunculan kata-kata yang dianggap mewakili dari masing-masing karakter berdasarkan asumsi pribadi orang yang mengklasifikasikan atau deskripsi dari objek wisata tersebut.

Penelitian kedua adalah penelitian yang dilakukan oleh (Frisilya, 2016) yang berjudul klasifikasi kompetensi tugas akhir secara otomatis berdasarkan deskripsi singkat menggunakan perbandingan algoritma K-NN dan Naive Bayes. Penelitian ini mengelompokkan judul PA berdasarkan KBK. Sistem yang dibangun akan mengelompokkan judul PA berdasarkan deskripsi singkat dengan menerapkan text mining dan perbandingan dua algoritma klasifikasi. Algoritma yang dibandingkan yaitu K-Nearest Neighbor (K-NN) dan Naive Bayes. Kedua algoritma ini akan dibandingkan dengan menggunakan k-Fold Cross Validation dan Confusion Matrix serta pengujian nilai k pada k-NN untuk menentukan jenis klasifikasi terbaik yang akan diimplementasikan pada sistem. Berdasarkan hasil penelitian didapatkan akurasi tertinggi diperoleh oleh algoritma Naive Bayes yaitu 65,4 persen dan k-NN diperoleh rata-rata hasil akurasi 51,14 persen dengan menggunakan minimum support 0,024.

Penelitian ketiga adalah penelitian yang dilakukan oleh (Widyastuti, 2015) yang berjudul Sistem Rekomendasi Publikasi Ilmiah Berdasarkan Keterkaitan Konteks Menggunakan Metode Frquent Itemset Mining. Penelitian ini menggunakan metode Frequent Itemset Mining untuk menemukan kata yang frequent pada publikasi ilmiah. Penelitian ini juga menggunakan algoritma Cosine Similarity untuk menampilkan rekomendasi publikasi ilmiah berdasarkan frekuensi kumpulan kata yang relevansinya dihitung dalam persentasi. Dari pengujian yang telah dilakukan dapat diperoleh bahwa FIM berhasil mereduksi dimensi fitur kata dengan rata-rata sebesar 80,17% menggunakan minimum support 0,1.

3.2. Klasifikasi

Klasifikasi adalah pengelompokan yang sistematis mengenai objek, gagasan, buku atau benda-benda lain ke dalam kelas atau golongan tertentu berdasarkan ciri-ciri yang sama. Di dalam klasifikasi bahan pustaka, dipergunakan penggolongan berdasarkan beberapa ciri tertentu seperti berdasarkan bentuk fisik, berdasarkan penggunaan bahan pustaka, dan berdasarkan isi atau subyek buku (Ha (Feldman, 2007).

Dalam pengklasifikasian data terdapat dua proses yang dilakukan yaitu: (Gurusamy, 2014)

a. Proses *training*

Pada proses *training* digunakan *training set* yang telah diketahui label-labelnya untuk membangun model atau fungsi.

b. Proses *testing*

Untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses *training*, maka digunakan data yang disebut dengan *testing set* untuk memprediksi label-labelnya.

3.3. Kelompok Bidang Kompetensi (KBK)

Kelompok Bidang Kompetensi (KBK) pada Jurusan Teknologi Informasi (JTI) di Politeknik Caltex Riau yaitu sebanyak 6 KBK, yaitu:

a. *Software Engineering (SE)*

Berisi fokus keilmuan terkait dengan berbagai metode pada tahapan-tahapan pengembangan / rekayasa perangkat lunak (SDLC): *Software Requirement, Software Design, Software Testing, Software Quality, Refactoring, Software Re-engineering, Mobile/Web/Desktop Application, Human Computer Interaction.*

b. *Business Analyst (BA)*

Berisi fokus keilmuan pemanfaatan IT/software untuk pengembangan bisnis proses pada suatu organisasi. Beberapa cakupan pada BA: *Business Process Management, ERP, CRM, SCM, Manajemen Stok, Manajemen Proyek.*

c. *Data Engineering (DE)*

Berisi fokus keilmuan pengolahan data dari level transaksional hingga level analisa menggunakan berbagai metode dalam mendukung pengambilan keputusan bisnis. Beberapa cakupan pada DE: *Data Mining, Data Warehousing, Business Intelligence Data Integration, Pre Processing, Big Data, Text Mining.*

d. *Soft Computing (SC)*

Berisi fokus keilmuan pengembangan / pemanfaatan Algoritma untuk membantu dalam pengambilan keputusan bisnis. Perbedaan dengan DE adalah bahwa pada SC, fokus keilmuannya pada pemanfaatan Algoritma untuk menghasilkan output keputusan. Sedangkan pada DE, output keputusan dihasilkan dari pola data yang diolah. Beberapa cakupan pada SC: *Sistem Pakar, Fuzzy Logic, Semantic Web, Sistem Pendukung/Pengambil Keputusan, Artificial Intelligence, Context Aware, Knowledge Management.*

e. *Multimedia (Mu)*

Berisi fokus keilmuan terkait multimedia dalam bidang industri, meliputi: *Video Processing, Image Processing, Animasi, Game.*

f. *Operating System Computer Network (OSCN)*

Berisi fokus keilmuan terkait infrastruktur IT, meliputi: *Network Infrastructure, Cloud Computing, Operating System, Embedded System, Micro, Networking Security*

3.4. Naïve Bayes Classifier

(Indranandita, 2008) *Naïve Bayes Classifier* merupakan salah satu metode machine learning yang digunakan untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. *Naive Bayes Classifier* populer untuk keperluan *text classification* karena pemrosesan memiliki waktu yang cepat, mudah diimplementasikan dengan strukturnya yang cukup sederhana dan untuk memiliki efektifitas yang tinggi (Taheri dan Mammadov, 2013). Dasar dari teorema *naïve bayes* yang dipakai dalam pemrograman adalah rumus *Bayes*.

$$P(c_j | w_i) = \frac{P(c_j) P(w_i | c_j)}{P(w_i)} \dots \dots (1)$$

Keterangan:

$P(c_j | w_i)$: Peluang kategori j ketika terdapat kemunculan kata i

$P(c_j)$: Peluang kemunculan sebuah kategori j

$P(w_i | c_j)$: Peluang sebuah kata i masuk ke dalam kategori j

$P(w_i)$: Peluang kemunculan sebuah kata

Naïve Bayes Classifier merupakan model penyederhanaan dari *Metode Bayes* yang cocok dalam pengklasifikasian teks atau dokumen [1]. Persamaannya adalah:

$$V_{nb} = \operatorname{argmax}_{v_{w \in v}} P(c_j) \prod P(w_i | c_j) \dots \dots \dots (2)$$

Pada perhitungan klasifikasi terdapat prior yang digunakan untuk menghitung peluang kemunculan kategori pada semua dokumen, perhitungan prior dapat dilihat pada persamaan berikut (Prananda Antinasari, 2017):

$$P(c_j) = \frac{Nc}{N} \dots \dots \dots (3)$$

Keterangan:

$P(c_j)$ = peluang kelas c

Nc = jumlah kata pada kelas c , jumlah kata pada kategori c_j di data training

N = total data training yang digunakan

Perhitungan *conditional probability* dilakukan untuk menghitung peluang kemunculan suatu kata (w_i) dalam setiap kelas (c_j). Pada persamaan *conditional probability* untuk menghindari nilai 0 digunakan *Laplace Smoothing*, dengan menambahkan angka 1 untuk tiap perhitungan kata uji dengan persamaan berikut (Antinasari et al., 2017):

$$P(w_i | c_j) = \frac{\operatorname{count}(w_i, c_j) + 1}{(\sum_{w \in V} \operatorname{count}(w, c_j)) + |V|} \dots \dots \dots (4)$$

Keterangan:

$P(w_i | c_j)$: peluang kemunculan suatu kata dalam setiap kategori j

$\operatorname{count}(w_i, c_j)$: merupakan jumlah dari suatu kata *query* yang muncul dalam suatu kelas atau kategori

$\sum_{w \in V} \operatorname{count}(w, c_j)$: merupakan jumlah dari seluruh kata yang ada pada kategori j

$|V|$ = jumlah seluruh kata yang tidak duplikat (kata unik) pada seluruh kategori

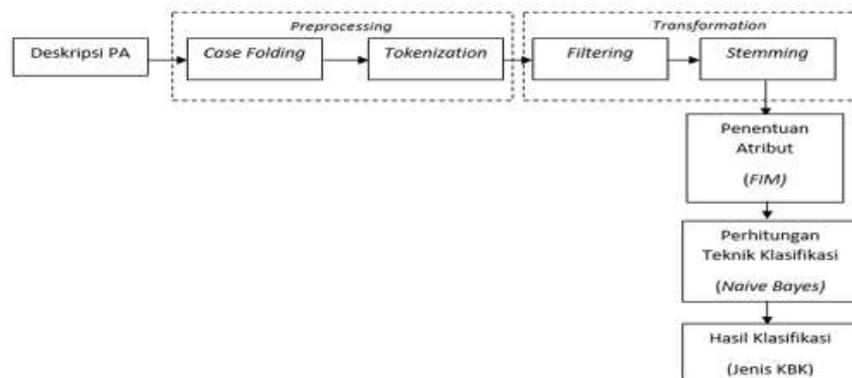
Perhitungan posterior merupakan perhitungan yang dilakukan dengan mengalikan prior dengan total *conditional probability*. Rumus perhitungan dapat dilihat pada persamaan:

$$P(c_j | w_i) = P(c_j) \times P(w_i | c_j) \times \dots \times P(w_n | c_j) \dots \dots \dots (5)$$

4. HASIL

4.1. Proses Klasifikasi

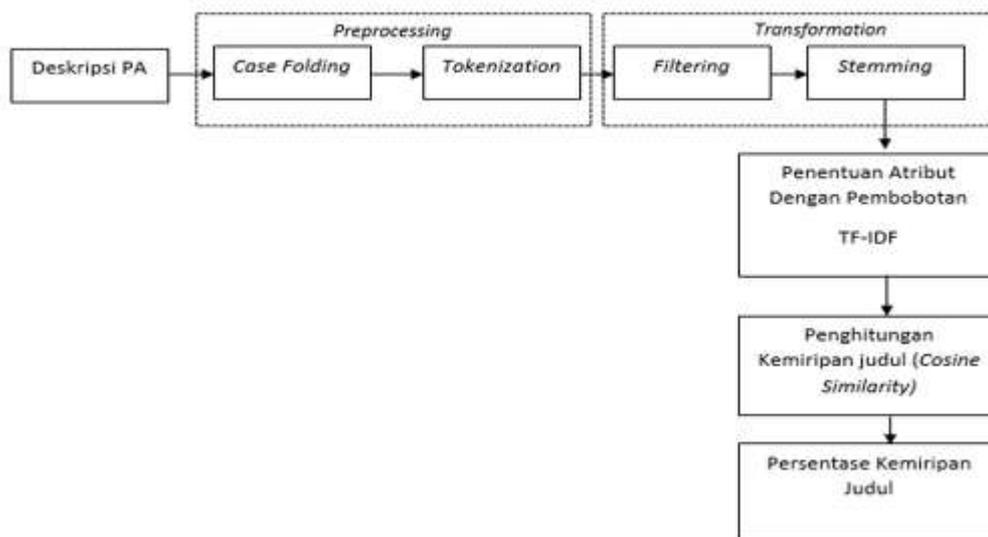
Secara umum sistem ini dibagi menjadi beberapa tahapan proses, pada Gambar 2 ini merupakan alur kerja proses *text mining* klasifikasi judul PA.



Gambar 2. Alur Kerja Proses *Text Mining* Klasifikasi Judul Proyek Akhir

Pada alur kerja proses *text mining* pertama berupa deskripsi PA yang akan melakukan proses *preprocessing* (*case folding*, *tokenization*) dan *transformation*, (*filtering*, *stemming*). Setelah melakukan tahap *preprocessing* dan *transformation* selanjutnya penentuan atribut dengan melakukan pengecekan terhadap setiap kemunculan kata. Setelah penentuan atribut selanjutnya akan dilakukan perhitungan teknik klasifikasi yang merupakan tahap penting untuk menemukan pola atau pengetahuan dari keseluruhan teks dengan menggunakan *data mining*. Teknik klasifikasi yang akan diterapkan pada sistem ini adalah *Naive Bayes*.

4.2. Proses Pengecekan Kemiripan Judul

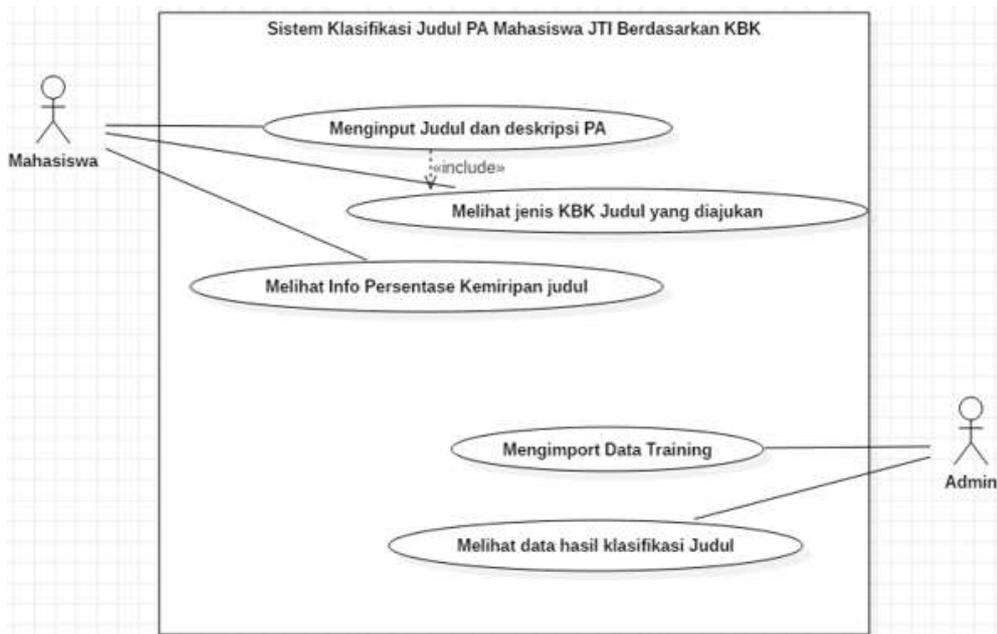


Gambar 3. Proses Pengecekan Kemiripan Judul

Gambar 3 merupakan Proses pengecekan kemiripan judul menggunakan proses penentuan atribut dengan pembobotan TF-IDF. Sebelum ke tahap pembobotan kata dengan TF-IDF, terlebih dahulu lakukan proses *preprocessing* (*case folding*, *tokenizing*) dan *transformation* (*filtering*, *stemming*) dari data deskripsi PA. Proses tersebut digunakan untuk menghasilkan kumpulan kata yang dapat digunakan untuk proses pencarian kata yang sering muncul pada judul PA. Proses selanjutnya adalah mencari kata yang sering muncul yang akan digunakan sebagai fitur kata menggunakan metode TF-IDF. Setelah fitur kata ditemukan selanjutnya dilakukan perhitungan menggunakan algoritma *Cosine Sjmlarity* untuk menentukan persentase kemiripan dari judul PA yang baru dengan judul PA terdahulu. Terakhir adalah menampilkan hasil persentasi kemiripan judul yang baru diinput dengan yang berada di arsip data judul sistem.

5. Perancangan Sistem

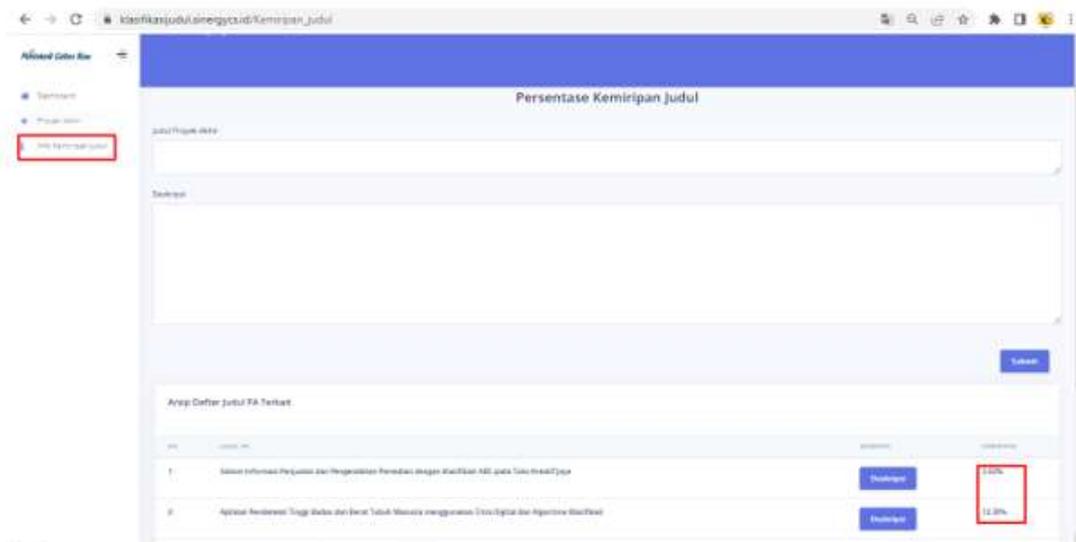
Use Case Diagram merupakan model diagram UML (*Unified Modeling Language*) yang digunakan untuk menggambarkan semua aktor dari sebuah sistem. Aktor (*actor*) yang dimaksud adalah seseorang atau sesuatu yang berinteraksi dengan sistem (pertukaran informasi dalam sistem). Pada sistem ini aktor yang digunakan adalah mahasiswa dan admin. Perancangan *use case* dapat dilihat pada Gambar 4.



Gambar 4. Use Case Diagram

5.1. Hasil Implementasi

Pada sistem klasifikasi judul PA berdasarkan KBK pada JTI hal pertama yang harus dilakukan user adalah login. Terdapat dua user yang terlibat dalam penggunaan sistem ini, yaitu mahasiswa dan admin. Mahasiswa dapat melakukan input judul PA dengan mengisi judul PA, deskripsi serta daftar Pustaka, yang kemudian system akan menampilkan KBK berdasarkan data yang sudah diinputkan. Selain itu system juga akan menampilkan referensi judul serupa beserta persentase kemiripannya.



Gambar 4. Halaman Persentase Kemiripan Judul

5.2. Pengujian dan Analisis

5.2.1. Pengujian *Confusion Matrix*

accuracy: 90.88% +/- 2.66% (micro average: 90.88%)

	true SE	true BA	true DE	true SC	true M	true OSCN	class preci...
pred. SE	19360	568	184	0	0	0	96.28%
pred. BA	154	8250	102	0	0	0	96.99%
pred. DE	6	70	9919	290	0	0	96.44%
pred. SC	0	32	505	16694	3574	155	79.65%
pred. M	0	0	0	58	7691	173	97.08%
pred. OSCN	0	0	0	18	1665	13372	88.82%
class recall	99.18%	92.49%	92.61%	97.85%	59.48%	97.61%	

Gambar 5. Hasil Pengujian *Confusion Matrix*

Dari hasil Gambar 6 dapat dilihat bahwa nilai akurasi (*Confusion Matrix*) dengan menggunakan metode cross validation adalah 90.88%. Sehingga dapat disimpulkan bahwa data training akurat jika dilakukan perhitungan dengan menggunakan metode naïve bayes. Pada penelitian lainnya penggunaan Naïve Bayes pada teks judul PA menghasilkan akurasi sebesar 65.4 % (Frisilya, 2016). Perbedaan akurasi tersebut salah satunya dipengaruhi oleh jumlah data training yang digunakan, jenis data yang dipakai, serta jumlah fitur yang digunakan. Klasifikasi KBK berdasarkan judul PA dengan menggunakan algoritma Naïve Bayes dapat menghasilkan akurasi yang lebih baik dikarenakan kecocokan dengan karakteristik data yang digunakan, diantaranya memiliki fitur yang banyak, memiliki ketidakseimbangan kelas (*skewed class*), serta fitur data berupa data kategorikal.

5.2.2. *User Acceptance Testing*

Pengujian User Acceptance Test (UAT) adalah pengujian yang dilakukan oleh pengguna secara langsung untuk menentukan apakah sistem dapat berjalan dengan benar dan sesuai dengan kebutuhan pengguna (Nielsen, 2012). Berdasarkan hasil pengujian yang sudah dilakukan dapat dihitung persentase keberhasilan dari masing-masing aspek pertanyaan dengan rata-rata sebagai berikut:

a. *Usefulness*

$$= (92\% + 87\% + 94\% + 91\%) / 4$$

$$= 91\%$$

Dari hasil perhitungan persentase diatas maka diketahui bahwa aspek *usefulness* berada pada range 91% yang berarti berada pada kriteria Sangat Setuju.

b. *Easy Of Use*

$$= (89\% + 91\% + 89\%) / 3$$

$$= 90\%$$

Dari hasil perhitungan persentase diatas maka diketahui bahwa aspek *Easy Of Use* berada pada range 90% yang berarti berada pada kriteria Sangat Setuju.

c. *Easy Of Learning*

$$= (93\% + 91\% + 91\%) / 3$$

$$= 92\%$$

Dari hasil perhitungan persentase diatas maka diketahui bahwa aspek *Easy Of Learning* berada pada range 92% yang berarti berada pada kriteria Sangat Setuju.

d. *Satisfaction*

$$= (86\% + 94\%) / 2$$

$$= 90 \%$$

Dari hasil perhitungan persentase diatas maka diketahui bahwa aspek *Satisfaction* berada pada range 90% yang berarti berada pada kriteria Sangat Setuju.

Untuk mendapatkan total keseluruhan dari hasil pengujian *usability* yaitu menjumlahkan bagian total nilai dari seluruh aspek di atas lalu dibagi 4. Maka didapatkan hasil sebagai berikut:

$$\text{Total} = (91\% + 90\% + 92\% + 90\%) / 4 = 91 \%$$

6. KESIMPULAN

Dari hasil pengujian yang telah dilakukan, maka dapat diambil beberapa kesimpulan sebagai berikut:

- a. Sistem klasifikasi menggunakan algoritma Naïve Bayes dapat memberikan informasi terkait KBK dari judul yang di input dengan nilai akurasi sebesar 90.88%. Akurasi tersebut dipengaruhi oleh kecocokan dengan karakteristik data yang digunakan, diantaranya memiliki fitur yang banyak, memiliki ketidakseimbangan kelas (*skewed class*), serta fitur data berupa data kategorikal.
- b. Sistem dapat menampilkan persentase kemiripan judul yang diajukan dengan judul-judul yang ada pada *database* sistem.
- c. Seluruh Fitur dan menu pada sistem dapat berjalan dengan baik dan benar dengan total nilai *usability* sebesar 91%, yang bermakna bahwa system yang dibangun mudah untuk digunakan dan di pelajari.

DAFTAR PUSTAKA

- [1] D. R. Firmansyah, "Implementasi Text Mining Klasifikasi Objek Wisata dengan Metode Naive Bayes Classifier di Dinas Pariwisata dan Kebudayaan Jawa Barat," *Skripsi Program Sarjana SI Univeristas Komputer Indonesia*, 2015.
- [2] A. Frisilya, "Klasifikasi Kompetensi Tugas Akhir Secara Otomatis Berdasarkan Deskripsi Singkat Menggunakan Perbandingan Algoritma K-Nndannaive Bayes," *Jurnal Aksara Komputer Terapan*, vol. 5 No. 1, 2016.
- [3] R. Widyastuti, "Sistem Rekomendasi Publikasi Ilmiah Menggunakan TF-IDF Cosine Smilarity dan Reduksi Dimensi Menggunakan Frequent Itemset Mining," *Jurnal Aksara Komputer Terapan*, vol. 4 No. 2, 2015.
- [4] R. d. S. J. Feldman, *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*, New York: Cambridge University Press, 2007.
- [5] A. Indranandita, "Sistem Klasifikasi dan Pencarian Jurnal dengan Menggunakan Metode Niave Bayes dan Vector Space Model.," *Jurnal Universitas Kristen Duta Wacana.*, 2008.
- [6] V. &. K. S. Gurusamy, *Preprocessing Techniques for Text Mining*. RTRICS, 2014.
- [7] J. Nielsen, *Usability 101: Introduction to Usability.*, 2012.
- [8] R. S. P. M. A. F. Prananda Antinasari, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dengan Perbaikan Kata Tidak Baku," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, 2017.