TEKNIK DATA MINING MENGGUNAKAN METODE BAYES CLASSIFIER UNTUK OPTIMALISASI PENCARIAN PADA APLIKASI PERPUSTAKAAN

(STUDI KASUS : PERPUSTAKAAN UNIVERSITAS PASUNDAN – BANDUNG)

Oleh :

Selvia Lorena Br Ginting, Reggy Pasya Trinanda

Abstrak

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Salah satu metode data mining adalah klasifikasi yaitu proses pencarian model klasifikasi yang dapat membedakan objek label kelasnya. Naïve Bayes Classifier adalah salah satu teknik yang dapat dipakai untuk membangun model klasifikasi. Pada penelitian ini metode Naïve Bayes Classifier diterapkan sebagai teknik yang membangun model klasifikasi (pengelompokan) dari dokumen-dokumen yang ada di sebuah perpustakaan. Perpustakaan merupakan tempat dimana pengunjung memperoleh akses terhadap informasi dan pengetahuan. Mengingat banyaknya data buku yang dimiliki perpustakaan, pengelompokan dokumen sangat perlu dilakukan. Oleh karena itu sangat diperlukan sebuah perangkat lunak yang handal untuk mengatasi pengelompokan dokumen tersebut, demi terciptanya kenyamanan pencarian data buku yang dilakukan oleh pengunjung. Cara kerja dari metoda lunak Naïve Bayes Classifier menghitung peluang dari satu kelas dari masing-masing kelompok atribut yang ada dan menentukan kelas mana yang paling optimal, artinya pengelompokan dapat dilakukan berdasarkan kategori yang pengguna masukkan pada perangkat lunak. Hasil pengujian diperoleh Naïve Bayes Classifier mengklasifikasikan beberapa judul dan ketegori yang terdapat pada database perpustakaan kemudian pencarian akan dilanjutkan lebih mendalam dengan melibatkan deskripsi dari setiap buku, sehingga akan menampilkan lebih banyak referensi sebagai hasil pencarian. Tentunya referensi tersebut berkaitan dengan kata yang di masukkan oleh pengunjung pada mesin pencarian di aplikasi perpustakaan tersebut. Dengan demikian diharapkan dapat membantu pengunjung dengan memperoleh peluang yang lebih besar dalam pencarian buku yang diinginkan.

1. Pendahuluan

Saat ini, kemajuan teknologi di bidang basis data masa kini semakin meningkat. Kemajuan teknologi ini memungkinan suatu perusahaan atau organisasi mengumpulkan data dari berbagai sumber dengan mudah dan cepat, sehingga membuat volume data menjadi semakin besar dan terus bertambah. Semakin besarnya *volume* data yang ada menimbulkan masalah dalam pengklasifikasian atau pengelompokannya. Data yang tersebar tanpa dikelompokkan dengan aturan tertentu tentunya akan memperlambat proses pencarian. Untuk data skala kecil tentunya tidak ada masalah yang signifikan

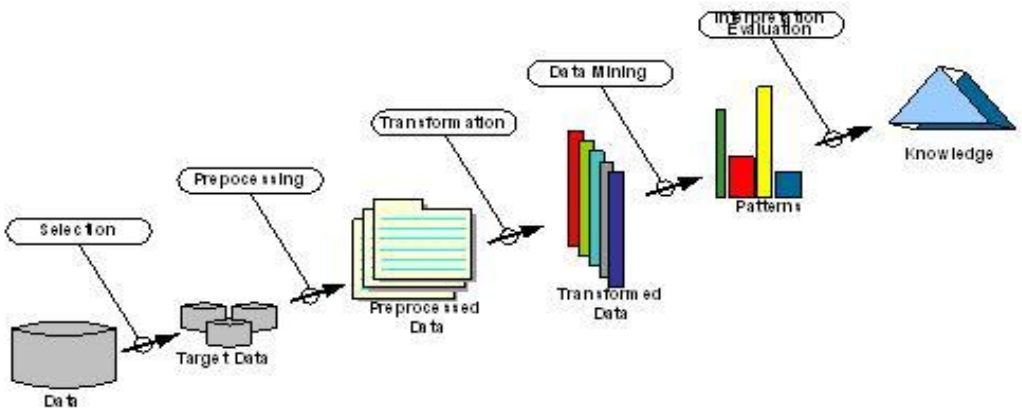
dalam proses pencarian, namun untuk data skala besar sangat dibutuhkan kecepatan dalam proses pencarian data, sehingga dibutuhkan pengelompokan data terlebih dahulu. *Naïve Bayes Classifier* merupakan merupakan salah satu metoda di dalam data mining untuk mengklasifikasikan data. Cara kerja dari metode *Naïve Bayes Classfier* menggunakan perhitungan probabilitas. Konsep dasar yang digunakan oleh *Naïve bayes* adalah Teorema *Bayes*, yaitu teorema yang digunakan dalam statistika untuk menghitung suatu peluang, *Bayes Optimal Classifier* menghitung peluang dari satu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal. Proses pengelompokan atau klasifikasi dibagi menjadi dua fase yaitu *learning/training* dan *testing/classify*. Pada fase *learning*, sebagian data yang telah diketahui kelas, datanya diumpankan untuk membentuk model perkiraan. Kemudian pada fase *testing,* model yang sudah terbentuk diuji dengan sebagian data.

Data yang digunakan di sini adalah data perpustakaan. Perpustakaan merupakan tempat yang cukup sering dikunjungi baik hanya sekedar membaca ataupun untuk mencari referensi. Fasilitas dan kenyamanan bagi pengunjung merupakan hal yang senantiasa perlu ditingkatkan diantaranya mempermudah pengunjung dalam hal pencarian buku dengan memanfaatkan *software* yang ada sehingga dapat membantu para pengunjung lebih cepat mengetahui daftar buku serta tempat penyimpanan buku yang ada pada perpustakaan tersebut dengan keakuratan pengklasifikasian dokumen yang baik. Untuk itu dibutuhkan metode *Naïve Bayes Classfier* untuk klasifikasi dokumen (di sini dokumen berupa data buku yang ada di perpustakaan) yang akan diterapkan dalam membangun perangkat lunak pencarian pada Aplikasi Perpustakaan. Fungsinya untuk mempercepat proses pencarian data buku pada Aplikasi Perpustakaan sehingga membantu meningkatkan pelayanan pada perpustakaan.

2. Dasar Teori

2.1 Data Mining

Knowledge Discovery in Database (KDD) merupakan proses pencarian pengetahuan yang bermanfaat dari kumpulan data. Proses KDD bersifat interaktif dan iteratif, meliputi sejumlah langkah dengan melibatkan pengguna dalam membuat keputusan dan dapat dilakukan pengulangan di antara dua buah langkah. *Data mining* merupakan salah satu proses inti yang terdapat dalam *Knowledge Data Discovery* (KDD). Banyak orang memperlakukan data mining sebagai sinonim dari KDD, karena sebagian besar pekerjaan dalam KDD difokuskan pada data mining. Namun, langkah-langkah ini merupakan proses yang penting yang menjamin kesuksesan dari aplikasi KDD.



Gambar 1 Ilustrasi Langkah-Langkah dalam Proses KDD

2.2 Naïve Bayes Classifier (NBC)

Salah satu tugas *Data Mining* adalah klasifikasi data, yaitu memetakan (mengklasifikasikan) data ke dalam satu atau beberapa kelas yaang sudah didefinisikan sebelumnya. Salah satu metoda dalam klasifikasi data adalah *Naïve Bayes Classifier* (NBC). *Naïve Bayes Classifier* merupakan salah satu metoda *machine learning* yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

Dasar dari *Naïve Bayes* yang dipakai dalam pemrograman adalah rumus Bayes:

$P(A|B) = (P(B|A) * P(A))/P(B)$ ..... (1)

Peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B. Pada pengaplikasiannya nanti rumus ini berubah menjadi :

$P(Ci|D) = (P(D|Ci)*P(Ci)) / P(D)$ ..... (2)

Naïve Bayes Classifier atau bisa disebut sebagai *Multinomial Naïve Bayes* merupakan model penyederhanaan dari Metoda *Bayes* yang cocok dalam pengklasifikasian teks atau dokumen. Persamaannya adalah:

$V_{MAP} = \arg\max P(V_j | a_1, a_2, ...a_n)$ ..... (3)

Menurut persamaan (3), maka persamaan (1) dapat ditulis:

$$V_{MAP} = \underset{v_j \in V}{\arg max} \frac{P(a_1, a_2, \ldots . a_n | v_j) P(v_j}{P(a_1, a_2, \ldots . a_n)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

P(a1, a2,.....an) konstan, sehingga dapat dihilangkan menjadi

$$V_{MAP} = \underset{v_j \in V}{\arg max} P(a_1, a_2, \ldots . a_n | v_j) P(v_j) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . (5)$$

Karena P(a1 , a2,….. an | vj) sulit untuk dihitung, maka akan diasumsikan bahwa setiap kata pada dokumen tidak mempunyai keterkaitan.

$$V_{MAP} = \underset{v_j \in V}{\arg max} P(v_j) \prod_i P(a_i | v_j) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . . (6)$$

Keterangan :

$$P(v_j) = \frac{|docs_j|}{|Contoh|} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

$$P(w_k | v_j) = \frac{n_k + 1}{n + |kosakata|} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . \ldots\ldots\ldots (8)$$

Di mana untuk :

- P(vj) : Probabilitas setiap dokumen terhadap sekumpulan dokumen.
- P(wk|vj) : Probabilitas kemunculan kata wk pada suatu dokumen dengan kategori class vj.
- | docs | : frekuensi dokumen pada setiap kategori.
- | Contoh | : jumlah dokumen yang ada.
- Nk : frekuensi kata ke-k pada setiap kategori.
- kosakata : jumlah kata pada dokumen test.

Pada persamaan (8) terdapat suatu penambahan 1 pada pembilang, hal ini dilakukan untuk mengantisipasi jika terdapat suatu kata pada dokumen uji yang tidak ada pada setiap dokumen data training.

2.3 Cara Kerja Naïve Bayes Classifier

Cara kerja *Naïve Bayes Classifier* melalui dua tahapan, yaitu :

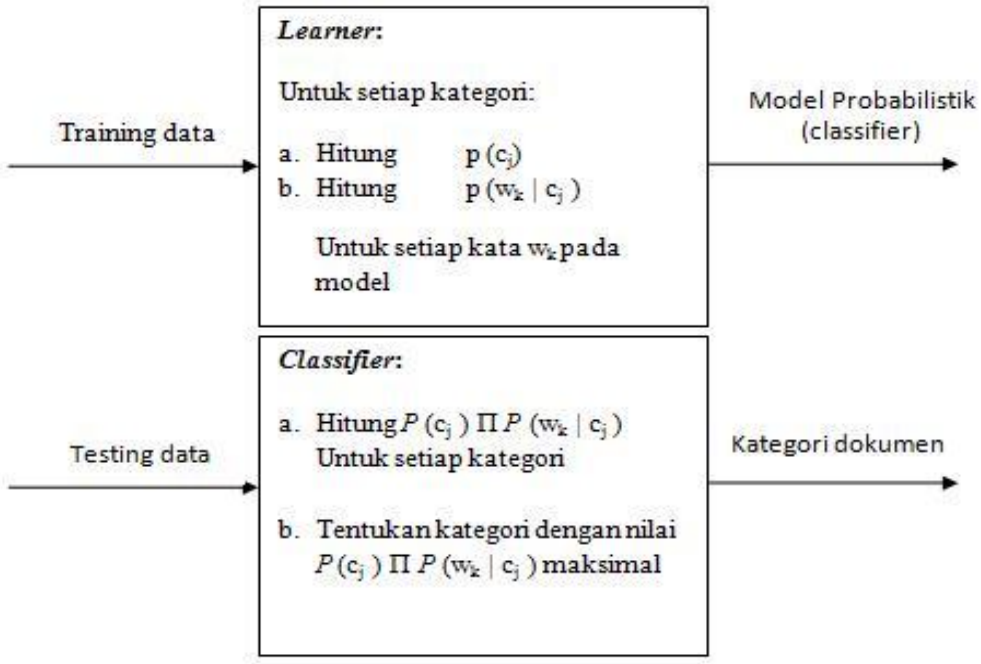
➔ *Learning* (Pembelajaran)
Naïve Bayes adalah suatu metoda yang termasuk ke dalam supervised learning, maka akan dibutuhkan pengetahuan awal untuk dapat mengambil keputusan. Langkah-langkah :

- *Step* 1 : Bentuk *vocabulary* pada setiap dokumen *data training*

- *Step* 2 : Hitung probabilitas pada setiao kategori P(vj).
- *Step 2* : Tentukan frekuensi setiap kata wk pada setiap kategori P(wk|vj)

➔ *Classify* (Pengklasifikasian). Langkah-langkahnya adalah :

- *Step* 1 : Hitung P(vj)Π P(wk | vj) untuk setiap kategori.
- *Step* 2 : Tentukan kategori dengan nilai P(vj)Π P(wk | vj) maksimal.



Gambar 2 Tahapan Proses Klasifikasi Dokumen *Naïve Bayes Classifier*

3. Analisis, Perancangan, Implementasi dan Pengujian Sistem

Pencarian data buku pada Aplikasi Perpustakaan dapat dilakukan berdasarkan judul, pengarang, maupun kata kunci (*keyword).* Pada umumnya hasil pencarian hanya berupa buku yang memiliki judul, pengarang maupun kata kunci (*keyword)* yang sesuai dengan kata atau kalimat yang di ketik oleh *user* saja, buku-buku yang memiliki makna sejenis dengan kata yang dicari, namun tidak mengandung kata tersebut pada *keyword* yang diketikkan oleh user, tidak akan ditampilkan pada hasil pencarian.

Pada aplikasi ini pencarian akan dilakukan lebih mendalam, sampai kepada deskripsi dari buku-buku tersebut dan akan dilakukan pengklasifikasian dengan menggunakan metoda *Naïve Bayes Clasiffier* (NBC) sehingga akan mendapatkan hasil yang lebih optimal dikarenakan *user* mendapatkan referensi buku yang lebih banyak.

3.1 Analisis Sistem

Data yang digunakan adalah data buku dari perpustakaan Universitas Pasundan yang meliputi pencarian berdasarkan judul, kategori, pengarang, serta deskripsi dari buku-buku tersebut. Jurmlah data ada 560 judul buku, sebagian data dapat dilihat pada tabel berikut :

pk_buku	judul	kategori	pengarang	deskripsi	lokasi
5	Auditing: pende...	AKUNTANSI	ARENS, ALVIN A.	Pembahasan me...	1C2
9	Accounting princ...	AKUNTANSI	DONALD	merupakan sala...	1C2
10	Intermediate ac...	AKUNTANSI LAN...	DONALD	Karena munculn...	1C3
37	Dasar-dasar An...	EKONOMI MIKRO	Budi Rahardjo	Buku ini menjela...	2D2
37	Teori Akuntansi	AKUNTANSI	Sofyan Syafri H...	Teori Akuntansi ...	1C2
38	Pesona Bisnis Ba...	BISNIS	Eva Sativa Nilawati	Peluang bisnis b...	4F1
83	Teknologi Otomo...	TEKNIK MESIN	Nyoman Sutantr...	Perkembangan T...	3E5
94	Panduan Sistem ...	TEKNIK ARSITEK...	Jimmy S. Juwana	Bangunan tinggi ...	3E7
95	10 Strategi Man...	EKONOMI MANA...	Michael LeBoeuf	Kunci untuk mem...	2D3
96	Analisis Ekonom...	EKONOMI MAKRO	Wing Wahyu Wi...	Sudah banyak b...	2D4
117	Membuat aneka ...	TEKNOLOGI TEP...	FACHRUDDIN, L...	Teknologi tepat ...	1B1
168	Fisika dasar	FISIKA	SUTRISNO	Seri fisika	1A1
1082	Cara sehat inver...	INVESTASI-EKO...	WIDOATMODJO...	Membuat uang b...	2D5
1585	Menjadikan bua...	PETERNAKAN	REDAKSI TRUBUS	Seri Agrikiat	1G1
1761	Pengelolaan har...	KEKAYAAN NEG...	AAK	satu tujuan yuridis	1H1
6014	ALGEBRA AND T...	MATEMATIKA	MARCUS	ilmu matematika	1A2
6097	ILMU NEGARA U...	JATI DIRI BANGSA	KRANENBURG MR	ilmu negara	1H2
6582	CONTOH BENTU...	HUKUM ACARA ...	SOEROSO	PRAKTIK HUKUM...	1H3
7064	BISIKAN DARI L...	NOVEL	LUFITYANTO GA...	karya sastra fiks...	1J1
7221	BIOLA TAK BER...	NOVEL	AJIDARMA SEN...	karya sastra fiks...	1J1
7398	PETUNJUK PRAK...	KEWIRAUSAHAAN	SUTAMTO	PERUSAHAAN K...	1J2
1	Afeksi Islam	DINAMIKA ISLAM	TIM DOSEN PAI	Materi pokok pe...	1K1
4	MANAJEMEN ST...	MANAJEMEN EK...	THOMAS L. WHE...	Buku ini lebih me...	1D1
14	STUDI KELAYAK...	MANAJEMEN PR...	HUSNAN, SUAD	Informasi Bibliog...	1D2
15	Manajemen sum...	MANAJEMEN SU...	ROBERT L MALT...	Buku ini berorien...	1D3
21	Menggambar me...	DINAMIKA PERM...	SATO, G. TAKESHI	Menyajikan dasa...	1B2

1 of 560

Gambar 3 Cuplikan Data

Pada proses pencarian dengan menggunakan metoda *Naïve Bayes Classifier* akan melalui dua tahap, yaitu proses *learning* dan proses *classifier*. Dimana proses *learning* akan membentuk *vocabulary* pada setiap dokumen data *training*, yaitu berupa kamus kata dasar yang nantinya akan menjadi perbandingan antara satu kata dengan kata yang lainnya. Kemudian proses *learning* akan menghitung probabilitas pada setiap kategori dan menentukan frekuensi dari setiap kata yang muncul pada setiap kategori tersebut untuk nantinya diklasifikasikan.

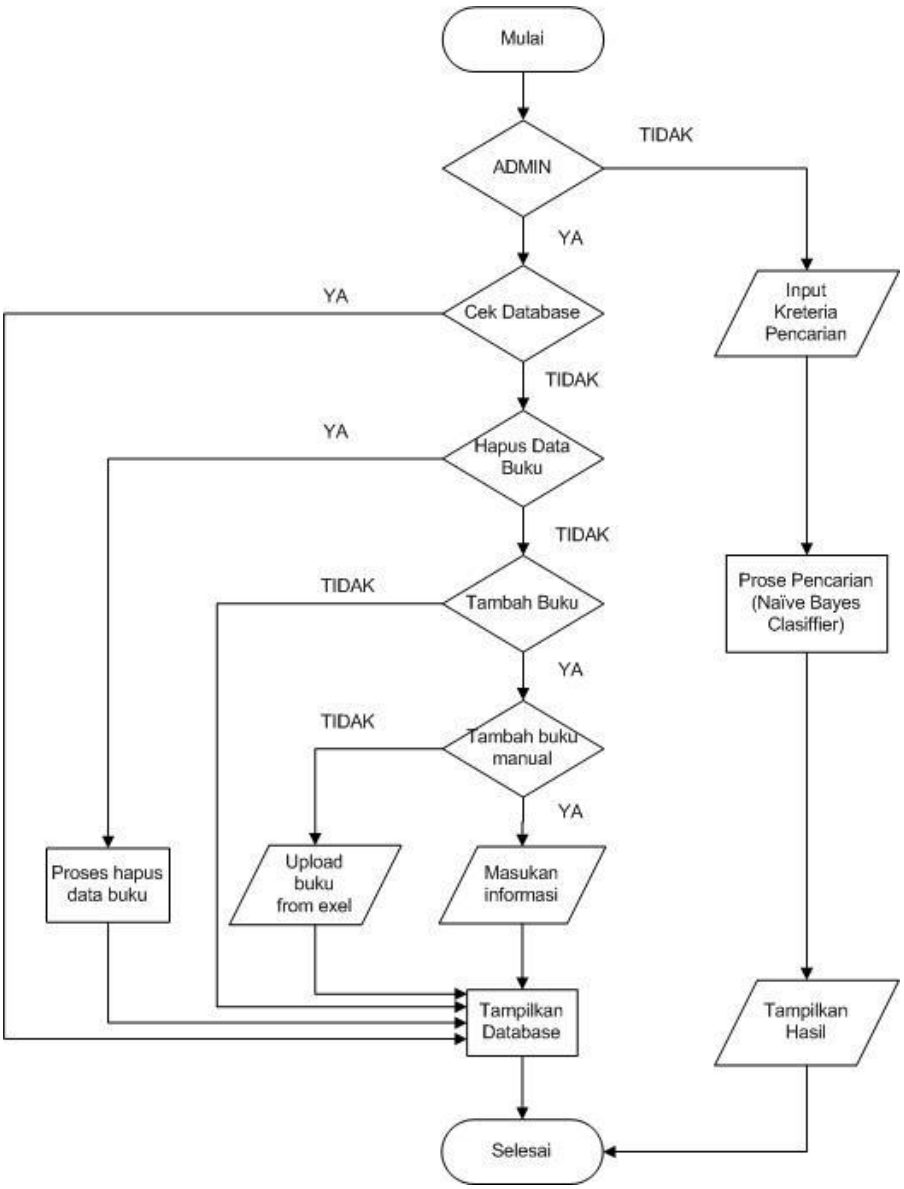
Sedangkan pada proses *classifier*, langkahnya yaitu menghitung probabilitas pada setiap dokumen terhadap sekumpulan dokumen dan kemudian menentukan probabilitas kemunculan kata yang terbesar pada suatu dokumen dengan kategori *class* tersebut. Sehingga kata yang memiliki presentasi terbesarlah yang akan di munculkan pada hasil pencarian.

Data *Output* yang diharapkan yang akan diperoleh oleh *user* dari aplikasi pencarian menggunakan metoda *Naïve Bayes classifier* ini adalah berupa hasil

pencarian yang diinginkan oleh *user* itu sendiri dengan menampilkan kombinasi kolom judul, kategori, pengarang, deskripsi, serta tempat buku tersebut disimpan.

3.2 Perancangan Sistem

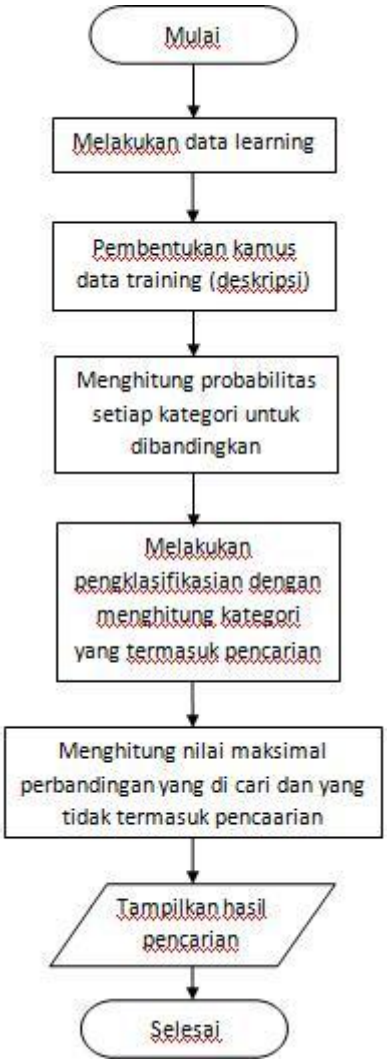
Flowchart perangkat lunak pencarian data buku menggunakan *Naïve Bayes classifier* adalah sebagai berikut :



Gambar 4 *Flowchart* Perangkat Lunak Pencarian Data Buku

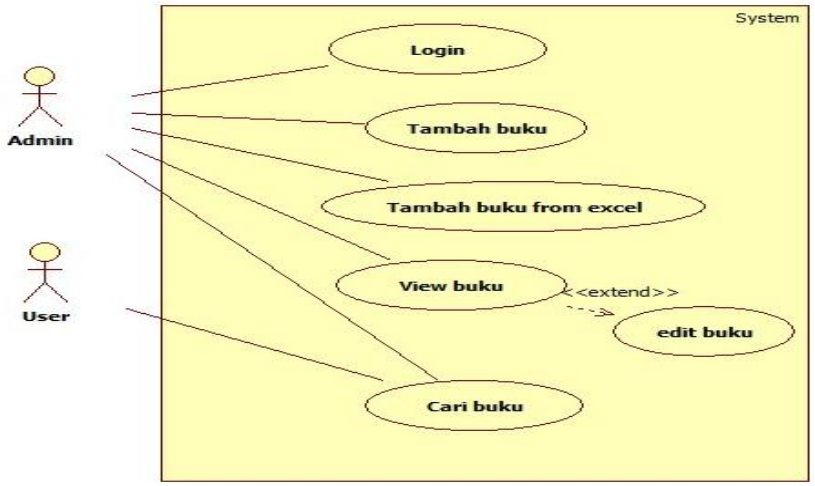
Berdasarkan *flowchart* tersebut, apabila yang mengakses perangkat lunak adalah seorang admin maka akses yang dapat dilakukan adalah cek *database*, edit *database* menambah buku baik secara manual maupun melalui *excel* serta dapat menampilkan *database*. Namun apabila yang mengakses adalah *user* perpustakaan, maka hanya bisa mencari buku yang ingin dicari dengan cara memasukkan kriteria pada kolom yang telah disediakan dan akan memperoleh hasilnya.

Berikut *flowchart* perangkat lunak pencarian mengunakan *Naïve Bayes Classifier :*



Gambar 5 *Flowchart* Perangkat Lunak Pencarian Menggunakan *Naïve Bayes Classifier*

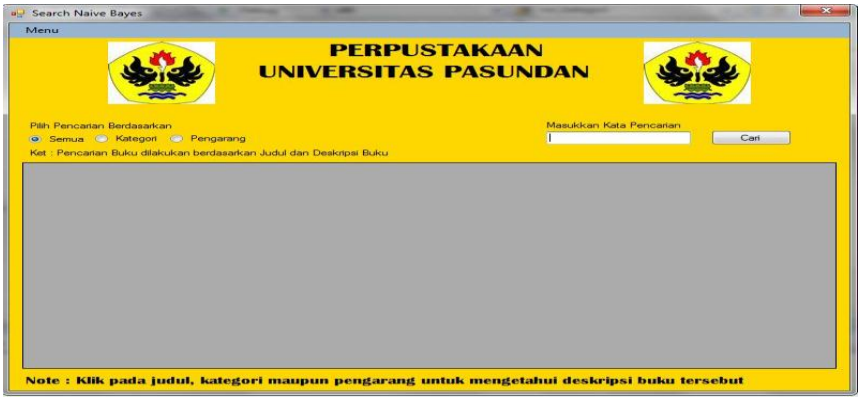
Use case diagram digunakan untuk memodelkan dan menyatakan unit fungsi/layanan yang disediakan oleh sistem (atau bagian sistem : subsistem atau *class*) ke pemakai.



Gambar 6 *Use Case Diagram* Perangkat Lunak Pencarian

3.3 Implementasi Sistem

a. Berikut hasil perancangan antarmuka sistem (menu awal) yang telah dibangun. Menu utama ini merupakan menu yang dapat di akses *user* dan *user* sudah dapat melakukan pencarian data buku yang diinginkan berdasarkan judul, kategori maupun deskripsinya. Sedangkan admin dapat melakukan login untuk mengakses menu utama.



Gambar 7 Tampilan Menu Awal Untuk Melakukan Pencarian

b. Menu utama hanya dapat diakses oleh admin. Disini admin dapat melakukan tambah buku secara manual maupun dari excel dan juga dapat *view* buku untuk melihat *database* buku.

Gambar 8 Tampilan Menu Utama

c. Admin dapat menambahkan data buku secara manual maupun melalui format excel dengan mengunggah file excel dari komputer lalu menambahkan data buku tersebut ke dalam *database*. Kemudian admin dapat merubah data yang sudah ada dan data baru tersebut disimpan otomatis ke dalam *database*.

Gamber 9 Tampilan *View* Buku

3.4 Pengujian Sistem

Pengujian dilakukan pada pencarian berdasarkan deskripsi, dengan menggunakan metoda *Naïve Bayes Classifier*. Sehingga dapat di tentukan

judul buku yang merupakan klasifikasi dari *keyword* yang di masukkan oleh *user* misalnya “komputer”. Berikut hasil pengujian dengan memasukkan kata ‘komputer’ dalam kolom pencarian.

$$(v_j) = \frac{|docs_j|}{|Contoh|} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

Tabel 1 Hasil Perhitungan *Naïve Bayes Classifier*

No	|docs|	|Contoh|	P(vj)	Hasil
1	1	17	1/17	0.05882353
2	1	17	1/17	0.05882353
3	3	17	3/17	0.17647059
4	1	17	1/17	0.05882353
5	1	17	1/17	0.05882353
6	1	17	1/17	0.05882353
7	1	17	1/17	0.05882353
8	1	17	1/17	0.05882353
9	1	17	1/17	0.05882353
10	1	17	1/17	0.05882353
11	1	17	1/17	0.05882353
12	1	17	1/17	0.05882353
13	1	17	1/17	0.05882353
14	1	17	1/17	0.05882353
15	1	17	1/17	0.05882353
16	1	17	1/17	0.05882353
17	1	17	1/17	0.05882353

Keterangan:

- |docs| : Frekuensi dokumen pada setiap kategori (kemunculan kata "komputer")
- |Contoh| : Jumlah dokumen yang ada.

Dengan melihat nilai P(vj) akan di ambil nilai maksimal atau nilai terbesar untuk nantinya dimasukan ke dalam Tabel *Naïve Bayes* untuk di tampilkan sebagai hasil pencarian pada aplikasi.

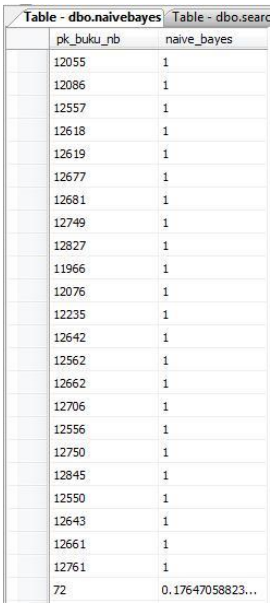
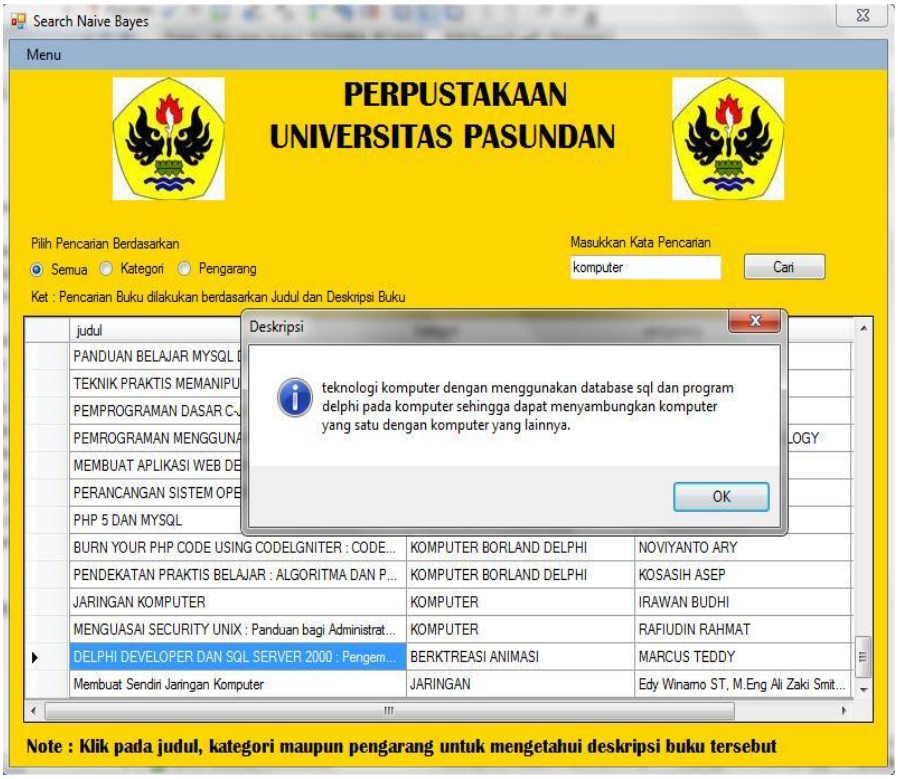


Table - dbo.naivebayes | Table - dbo.searc

pk_buku_nb	naive_bayes
12055	1
12086	1
12557	1
12618	1
12619	1
12677	1
12681	1
12749	1
12827	1
11966	1
12076	1
12235	1
12642	1
12562	1
12662	1
12706	1
12556	1
12750	1
12845	1
12550	1
12643	1
12661	1
12761	1
72	0.17647058823...

Gambar 10 Tabel *Naïve Bayes Classifer*



Gambar 11 Hasil Pengujian Algoritma *Naïve Bayes Classifier*

Dari pengujian di atas dapat diketahui bahwa perangkat lunak ini memenuhi beberapa spesifikasi seperti di bawah ini :

1. Kebutuhan
   Dari segi kebutuhan, perangkat lunak ini dibutuhkan untuk pengklasifikasian dokumen yang terdapat pada database.
2. Kegunaan
   Kegunaan dari perangkat lunak untuk mengklasifikasikan dokumen sehingga pencarian buku dapat dilakukan dengan lebih optimal dan akurat, karena perangkat lunak ini melakukan pengklasifikasian dari judul, kategori, pengarang serta deksripsi dari buku yang ingin dicari.
3. Performa
   Hasil pencarian yang dilakukan berdasarkan judul, kategori, pengarang dan deskripsi untuk data buku cukup efisien berdasarkan *loading* pada *database* dan penampilan hasil pencarian dengan cepat.

4. Simpulan dan Saran

Kesimpulan yang diperoleh dari hasil penelitian ini adalah perangkat lunak yang dibangun hanya dapat mencari judul, kategori dan deskripsi dari buku yang dicari,

tidak meliputi isi dari seluruh buku. Metode *Naïve Bayes Classifier* dapat diimplementasikan pada pengklasifikasian buku sehingga memudahkan pencarian. Disamping itu hasil pencarian lebih akurat dikarenakan buku yang ditampilkan bukan saja berdasarkan judul atau kategori atau pengarang saja melainkan berdasarkan deskripsi dari buku. Pengecekan terhadap deskripsi dilihat berapa banyak kata yang sesuai dengan kata kunci yang muncul pada deskripsi tersebut, sehingga hasil yang ditampilkan akurasinya lebih tinggi, dengan demikian *user* memiliki lebih banyak referensi dibandingkan bila hasil pencarian hanya berdasarkan judul, kategori atau pengarang saja. Selain itu deskripsi juga ditampilkan sebagai hasil pencarian sehingga *user* bisa melihat kesesuaian *keyword* dengan buku yang dicari.

Saran yang dapat disampaikan peneliti untuk pengembangan dan penelitian lebih lanjut adalah data buku sebaiknya diperbanyak misalnya puluhan bahkan sampai jutaan *record* sehingga dapat terlihat apakah keakurasian metode ini masih dapat diterima. Selain memperbanyak *data learning*¸ format data buku juga diperbanyak (bervariasi) misalnya dalam bentuk format pdf atau format file lainnya, sehingga dapat dilihat apakah metoda *Naïve Bayes Classifier* masih dapat bekerja dengan baik.

5. Daftar Pustaka

[1]. Oded Z. Maimon. Lior Rokach. *Decomposotion Methodology for Knowledge Discovery and Data Mining* : *Theory and Applications*. World Scientific : 2005.

[2]. Han, Jiawei, Kamber, Micheline. *Data Mining : Concepts and Techniques*. Morgan Kaufmann. 2001.

[3]. Susanto, S. Ph. D., Dedi S., S.T., M.S. *Pengantar Data Mining : Menggali Pengetahuan dari Bongkahan Data.*

[4]. Munawar. *Pemodelan Visual dengan UML*. GrahaIlmu : Yogyakarta. 2005.

[5]. Presman, Roger, Ph.D. *Software Engineering*, *A Practitioner's Approach*. Sixth Edition. USA : Mc. Graw Hill. 2005.

[6]. Pang-Ning, Tan. *Introduction to Data Mining*, USA : Pearson Addison Wesley. 2006.