PENGGUNAAN METODE NAÏVE BAYES CLASSIFIER PADA APLIKASI PERPUSTAKAAN

Selvia Lorena Br Ginting,S.Si., MT[1], Reggy Pasya Trinanda[2]
[1,2] Jurusan Teknik Komputer Unikom, Bandung
[1]selvialorena@yahoo.com, [2]reggypasya@gmail.com

ABSTRAK

Perpustakaan tempat yang sering di kunjungi. Hampir setiap instansi mempunyai perpustakaan, maka dari itu diperlukan peningkatan pelayanan untuk kenyamanan pengunjung. Pencarian buku pada aplikasi di setiap perpustakaan harus dapat mencakup kebutuhan perpustakaan tersebut, dengan kemudahan melakukan pencarian buku maka akan menambah kenyamanan pengunjung dalam mencari buku yang diinginkan. Dengan metode pengklasifikasian dokumen atau teks maka hal itu dapat dilakukan. Pada karya ilmiah ini digunakan metode Naïve Bayes Classifier untuk mengklasifikasikan beberapa judul dan ketegori yang terdapat pada database perpustakaan tersebut. Dengan menggunakan metode ini maka pencarian akan dilakukan lebih mendalam dengan melibatkan deskripsi dari setiap buku, sehingga ada lebih banyak referensi yang di tampilkan dalam 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 dalam melakukan pencarian buku yang diinginkan.

Kata kunci : metode pencarian, perpustakaan, Naïve Bayes Classifier.

1. PENDAHULUAN

1.1 Latar Belakang

Perpustakaan merupakan tempat yang cukup sering dikunjungi baik hanya sekedar membaca ataupun untuk mencari referensi. Dapat menggunakan sebuah software untuk pengklasifikasian buku dimana diharapkan dapat membantu para pengunjung mengetahui list berikut tempat penyimpanan buku yang ada pada perpustakaan tersebut dengan keakuratan pengklasifikasian dokumen yang baik.

Naïve Bayes merupakan salah satu metode machine learning yang menggunkan 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 klasifikasi biasanya 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 lainnya untuk mengetahui akurasi dari model tersebut.

1.2 Tujuan

Adapun tujuan yang ingin dicapai adalah :

1. Mempelajari metode *Naïve Bayes Classifier* untuk mengklasifikasikan dokumen.
2. Merancang sebuah perangkat lunak pencarian buku perpustakaan.
3. Mengimplementasikan algoritma *Naïve Bayes Classifier*.

1.3 Manfaat

Sedangkan manfaat nya adalah :

1. Mempercepat proses pencarian buku pada perpustakaan dengan menggunakan perangkat lunak dan metode *Naïve Bayes Classifier* ini.
2. Membantu peningkatan pelayanan pada perpustakaan untuk mempermudah proses pencarian buku.

2. DASAR TEORI

2.1 Data Mining

Data Mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data atau bisa disebut dengan KDD (*Knowledge Discovery in Database*). Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data.

KDD (*Knowledge Discovery in Database*) adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti.

2.2 Algoritma Naïve Bayes Classifier (NBC)
Naïve Bayes merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas. Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.
Dasar dari teorema 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 algoritma bayes yang cocok dalam pengklasifikasian text atau dokumen. Persamaannya adalah:

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

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

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

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

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

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

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

Keterangan:

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

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

Dimana 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 Algoritma Naïve Bayes Classifier (NBC)
Cara kerja dari *Naïve Bayes Classifier* yaitu melalui dua tahapan yaitu:
➔ Learning (Pembelajaran)
Naïve Bayes adalah algoritma 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 setiap kategori P(vj).
• Step 3 : 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.

3. PERANCANGAN DAN IMPLEMENTASI

aplikasi ini pencarian akan dilakukan lebih mendalam dengan mendreskipsikan buku-buku tersebut dan akan dilakukan pengklasifikasian

dengan menggunakan metoda naive bayes clasiffier (NBC) sehingga akan mendapatkan hasil yang lebih optimal dikarenakan user mendapatkan referensi buku yang lebih banyak.

3.1 Analisis Proses

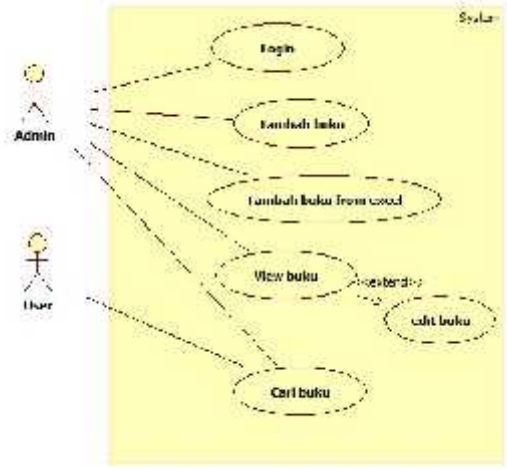
Pada proses pencarian dengan menggunakan metoda *Naïve Bayes Clasiffier* akan melalui dua tahap, yaitu proses learning dan proses clasiffier. Dimana proses learning akan membentuk vocabulary pada setiap dokumen data training, yaitu berupa kamus kata dasar yang nantinya akan menjadi perbandingan antara 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 di klasifikasikan.

Sedangkan pada proses clasiiffier 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.

3.2 Perancangan Software

3.2.1 Use case Diagram

Perancangan software dapat dilihat dari use case diagram di bawah ini:



Gambar 1. *Usecase* Diagram Sistem Perpustakaan

Terdapat dua actor yaitu:

1. Admin

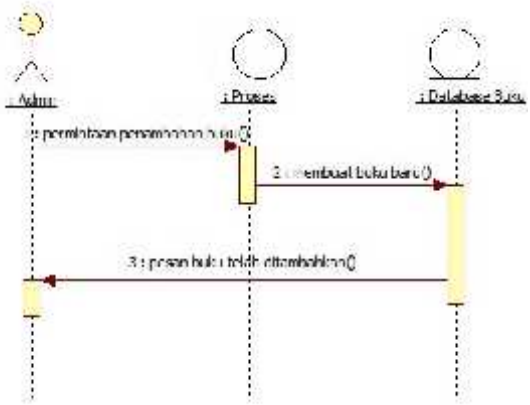
Dapat menjalankan system pada bagian login dan masuk ke menu utama untuk melakukan tambah buku, dan edit buku.

2. User

Dapat melakukan pencarian buku dengan memasukkan keyword yang diinginkan dengan mengacu pada judul, pengarang atau deskripsi dari buku yang dicari.

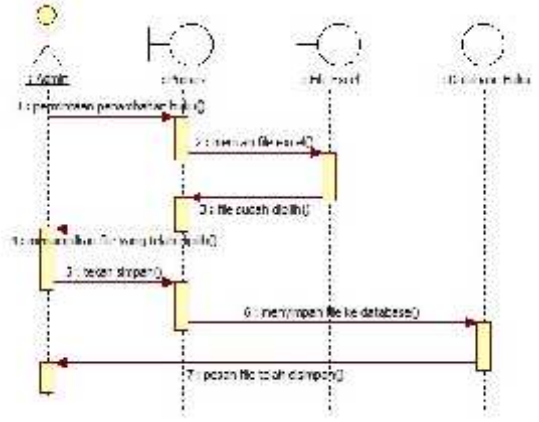
3.2.2 Sequence Diagram

Penjelasan antar proses akan digambarkan dengan sequence diagram berikut:



Gambar 2. *Sequence* Diagram Untuk Tambah Buku Secara Manual

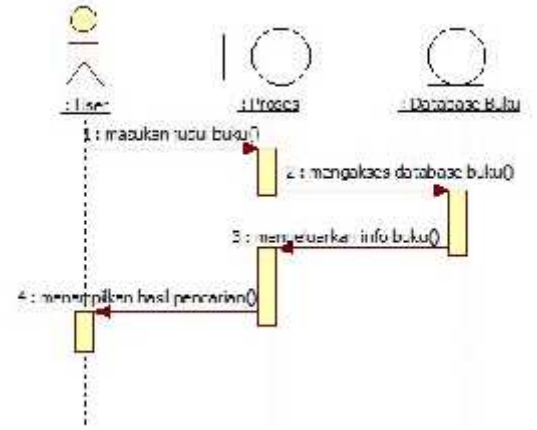
Penambahan buku dapat dilakukan secara manual maupun dengan mengungguhnya dari file excel atau .xls.



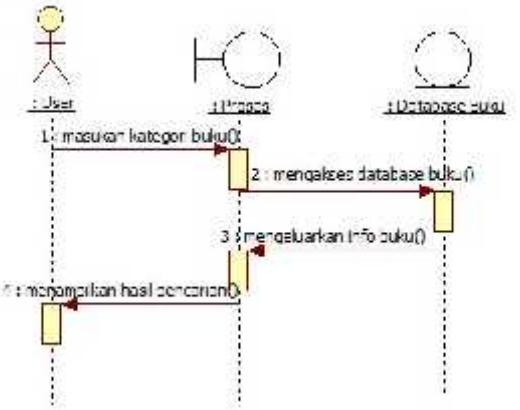
Gambar 3. *Sequence* Diagram Untuk Tambah Buku *From* Excel

Sedangkan untuk melakukan pencarian dapat dilakukan berdasarkan judul, kategori maupun deskripsi dari buku yang ingin di cari. Berikut

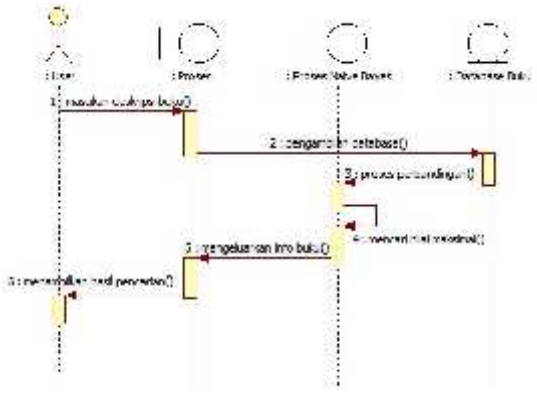
prosesnya yang di gambarkan melalui sequence diagram.



Gambar 4. *Sequence* Diagram Untuk Pencarian Berdasarkan Judul

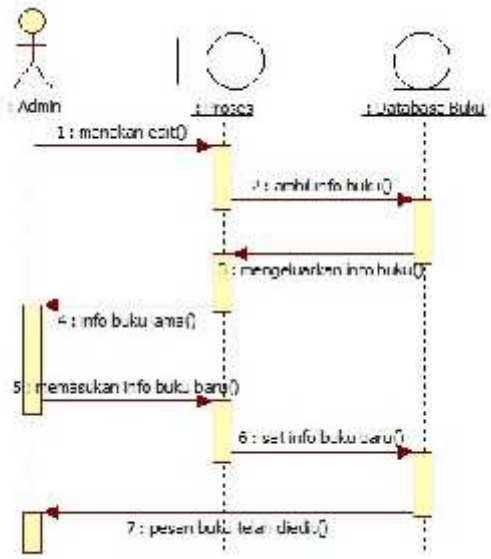


Gambar 5. *Sequence* Diagram Untuk Pencarian Berdasarkan Kategori

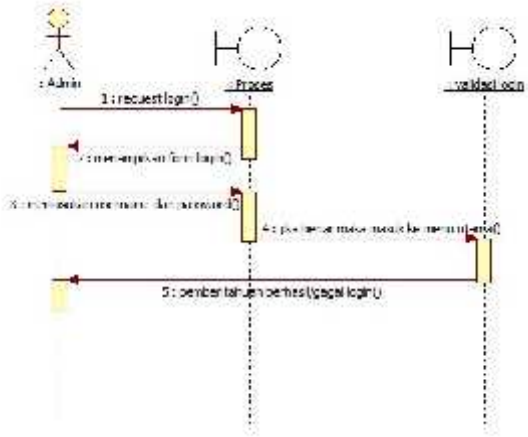


Gambar 6. *Sequence* Diagram Untuk Pencarian Berdasarkan Deskripsi

Dapat pula melakukan edit buku untuk mengedit kesalahan ketika memasukkan buku ke dalam *database*

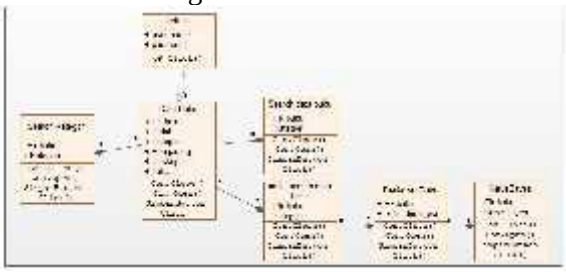


Gambar 7. *Sequence* Diagram Untuk Edit Buku



Gambar 8. *Sequence* Diagram Untuk Login

3.2.3 Class Diagram

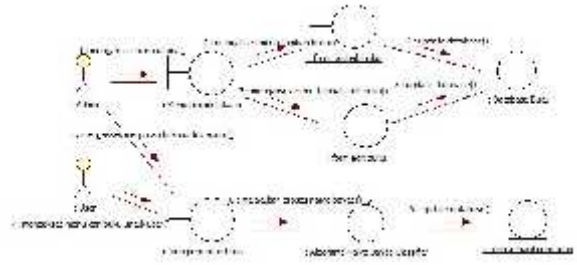


Gambar 9. *Class* Diagram

Dari kelas diagram diatas dapat dilihat bagaimana relasi antar table dan atribut yang terdapat pada table. Data buku menjadi inti dari proses pencarian, yang nantinya akan di kelompokan bedasarkan pencarian judul, kategori maupun deskripsi yang di masukkan oleh user.

3.2.4 Kolaborasi Diagram

Kolaborasi diagram adalah suatu diagram yang memperlihatkan pengorganisasian interaksi yang terdapat disekitar objek (seperti halnya *sequence* diagram) dan hubungannya terhadap yang lainnya.



Gambar 10. Kolaborasi Diagram Sistem Perpustakaan

4. IMPLEMENTASI DAN PENGUJIAN

4.1 Implementasi

4.1.1 From menu awal

Menu utama merupakan menu awal keluar setelah program dijalankan, dalam menu utama ini user sudah dapat melakukan pencarian buku yang diinginkan berdasarkan judul, kategori maupun deskripsinya.

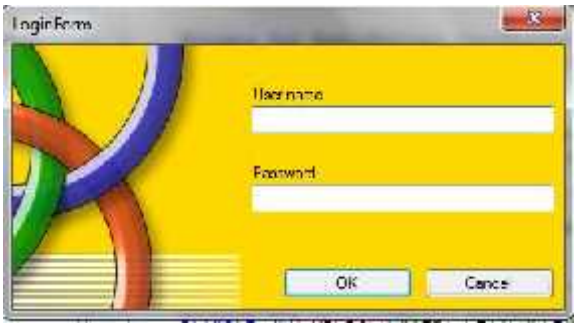
Sedangkan admin dapat melakukan login untuk mengakses menu utama.



Gambar 10. Tampilan Menu Awal Untuk Melakukan Pencarian

1.1.2 Menu Login

Menu login dapat dilakukan oleh admin dengan mengisi *user name* dan *password* untuk memasuki menu utama.



Gambar 11. Tampilan *Form* Login Untuk Admin

1.1.3 Menu Utama

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



Gambar 12. Tampilan Menu Utama



Gambar 13. Tampilan Tambah Buku



Gambar 14. Tampilan View Buku

1.2 Pengujian

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. Berikut hasil pengujian dengan memasukkan kata 'komputer' dalam kolom pencarian.

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

Tabel 1 Hasil Perhitungan Naïve Bayes

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 naivebayes untuk di tampilkan ke hasil pencarian pada aplikasi.



Tabel 15. Tabel Naïve Bayes



Gambar 16. Hasil Pengujian Algoritma *Naïve Bayes Classifier*

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

1. Kebutuhan

Dari segi kebutuhan aplikasi ini dibutuhkan untuk pengklasifikasian dokumen yang terdapat pada database.

2. Kegunaan

Kegunaan dari sistem aplikasi ini dari awal bertujuan untuk mengklasifikasikan dokumen sehingga pencarian buku dapat dilakukan dengan lebih optimal dan akurat, karena aplikasi ini melakukan pengklasifikasian dari judul,

kategori, pengarang serta deksripsi dari buku yang ingin dicari.

3. Performa

Untuk performa pada aplikasi ini terbilang cepat pada pemrosesan pencarian buku, demi kenyamanan pengunjung perpustakaan.

5. SIMPULAN DAN SARAN

Kesimpulan dari hasil pengujian yang dilakukan adalah:

Pada perangkat lunak yang dibangun hanya dapat mencari judul, kategori dan deskripsi dari buku yang dicari, tidak meliputi isi dari seluruh buku. Penggunaan metode *Naïve Bayes Classifier* dapat diimplementasikan pada pencarian di perangkat lunak ini, karena berhasil mencari deskripsi yang berkaitan dengan kata yang dicari oleh *user*. Dengan menggunakan *naïve bayes* maka pengklasifikasian dokumen dapat dilakukan dengan baik, dapat dililat pada pengujian. Deskripsi kata yang di cari oleh *User* akan muncul pada hasil pencarian.

Yang dapat disarankan pada pengujian ini adalah:

a. Dapat digunakan beberapa format buku lain seperti file PDF.

b. Disempurnakan lagi algoritma pencariannya untuk mendapatkan hasil yang lebih maksimal.

c. Melakukan data leraning yang lebih banyak untuk mendapatkan hasil yang lebih baik.

6. 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, Micheline Kamber, *Data Mining : Concepts and Techniques*. Morgan Kaufmann, 2001

[3] Hidayatullah, P. *Visual basic. NET membuat aplikasi database dan pemrograman kreatif*.

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

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