

# KLASIFIKASI DAN PENCARIAN BUKU REFERENSI AKADEMIK MENGUNAKAN METODE *NAÏVE BAYES CLASSIFIER* (NBC) (STUDI KASUS: PERPUSTAKAAN DAERAH PROVINSI KALIMANTAN TIMUR)

Agus Setiawan<sup>1)</sup>, Indah Fitri Astuti<sup>2)</sup>, Awang Harsa Kridalaksana<sup>3)</sup>

<sup>1,2,3)</sup>Program Studi Ilmu Komputer, FMIPA, Universitas Mulawarman  
Email: agust.setiawan@gmail.com<sup>1)</sup>, indahfitriastuti@yahoo.com<sup>2)</sup>, awangkid@gmail.com<sup>3)</sup>

## ABSTRAK

Buku merupakan salah satu elemen penunjang dalam proses belajar mengajar, baik itu di sekolah maupun perguruan tinggi. Salah satu tempat yang menyediakan akses untuk mendapatkan buku tersebut adalah Perpustakaan Daerah Provinsi Kalimantan Timur. Kebanyakan pengunjung sulit mendapatkan buku referensi yang tepat sesuai dengan kebutuhan mereka. Hal itu dikarenakan kurangnya fasilitas yang dapat mengklasifikasi buku-buku tersebut sesuai dengan kategorinya. Klasifikasi buku merupakan proses pengelompokan buku sesuai dengan kategori yang dimilikinya. Sebuah buku dapat dikelompokkan ke dalam kategori tertentu berdasarkan kata-kata yang terkandung pada judul buku tersebut. Penelitian ini bertujuan untuk membuat aplikasi klasifikasi dan pencarian buku yang dapat mengatasi masalah-masalah yang berkaitan dengan pengkategorian dan pencarian buku dengan lebih mudah dan akurat. Aplikasi ini dibangun menggunakan bahasa pemrograman PHP, *database MySQL*, dan menggunakan metode *Naïve Bayes Classifier* (NBC). Kategori buku yang digunakan ada lima, yaitu: pemrograman komputer, jaringan komputer, database komputer, multimedia, dan sistem operasi. Dengan mengimplementasikan metode NBC dan kelima kategori tersebut, maka dihasilkan sebuah aplikasi klasifikasi dan pencarian di Perpustakaan Daerah Provinsi Kalimantan Timur.

**Kata kunci :** Klasifikasi, Buku, *Naïve Bayes Classifier* (NBC), Perpustakaan.

## PENDAHULUAN

Buku merupakan jendela dunia. Dengan membaca buku, banyak ilmu yang dapat diperoleh. Salah satu manfaat buku adalah menunjang proses belajar mengajar, baik itu di sekolah maupun perguruan tinggi. Disamping itu, buku juga dapat dijadikan referensi dari penulisan akademik. Bagi kalangan akademisi, khususnya mahasiswa buku menjadi sangat penting karena dijadikan sebagai sumber pustaka dari penulisan akademik.

Kebanyakan mahasiswa sulit mendapatkan buku referensi yang sesuai dengan kebutuhan mereka di perpustakaan. Hal tersebut dikarenakan kurangnya sistem yang dapat mempermudah proses pencarian buku yang sesuai dengan kebutuhan mereka. Oleh sebab itu diperlukan suatu sistem yang dapat mempermudah dan mempercepat pencarian buku referensi tersebut. Salah satunya dengan pengklasifikasian buku sesuai dengan kategorinya sehingga proses pencariannya akan lebih cepat, mudah, dan akurat.

Klasifikasi buku merupakan proses pengelompokan buku sesuai dengan kategori yang dimilikinya. Sebuah buku dapat dikelompokkan ke dalam kategori tertentu berdasarkan kata-kata pada judul buku tersebut. Kata atau kalimat yang terdapat di dalam sebuah buku memiliki makna tertentu dan dapat digunakan sebagai dasar untuk menentukan kategori buku tersebut. Salah satu algoritma klasifikasi teks yang paling populer yaitu naive bayesian yang sering disebut dengan *Naïve Bayes Classifier* (NBC). Algoritma NBC ini memanfaatkan

teorema probabilitas yaitu teorema bayes dan fungsionalitas data mining. Kelebihan naive bayesian diantaranya adalah sederhana tapi memiliki tingkat akurasi yang tinggi dan error rate yang minimum. Algoritma NBC tersebut telah diterapkan pada dokumen teks berbahasa Indonesia dan hasilnya sangat efektif, di mana tingkat akurasinya tinggi, yaitu sebesar 87.63% (Samodra, 2009).

Berdasarkan latar belakang masalah di atas penulis berniat untuk merancang dan membuat suatu sistem komputerisasi yang dapat mengklasifikasi buku sesuai dengan kategorinya serta membuat suatu sistem pencarian buku dari hasil klasifikasi buku tersebut.

## Batasan Masalah

Dalam penelitian ini batasan masalah dibuat agar tidak menyimpang dari tujuan dibuatnya penelitian, adapun batasan-batasan masalah tersebut adalah:

1. Data buku yang diuji berasal dari Badan Perpustakaan Daerah Provinsi Kalimantan Timur.
2. Pengklasifikasian buku berdasarkan 5 kategori, yaitu Pemrograman Komputer, Jaringan Komputer, Database, Multimedia, dan Sistem Operasi.
3. Algoritma yang akan diimplementasikan pada bahasan mengenai proses pengklasifikasian buku adalah *Naïve Bayes Classifier* (NBC).

4. Data Pelatihan bersifat statis yang berasal dari katalog buku online di internet.
5. Proses klasifikasi buku berdasarkan judul dari buku tersebut.
6. Proses pencarian buku berdasarkan judul, penulis, dan kategori buku.
7. Buku-buku yang digunakan hanya buku yang ditulis dalam bahasa Indonesia.

### Tujuan Penelitian

Dengan adanya penelitian ini, tujuan yang ingin dicapai ialah membangun sistem klasifikasi dan pencarian buku dengan metode *Naïve Bayes Classifier* (NBC).

### Manfaat Penelitian

Adapun manfaat yang ingin diharapkan dari penelitian ini adalah:

1. Membantu administrator menentukan kategori yang tepat untuk setiap buku.
2. Mempermudah dan mempercepat proses pencarian buku yang sesuai dengan kebutuhan.

### TINJAUAN PUSTAKA

#### Kategorisasi Teks

Klasifikasi atau kategorisasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui (Rachli, 2007). Pengkategorisasian teks menjadi suatu hal yang penting dan kebutuhannya akan semakin meningkat seiring dengan berjalannya waktu, karena data semakin lama akan semakin bertambah dan media penyimpanannya juga akan semakin besar. Sehingga perlu dilakukan pengkajian metode untuk kategorisasi teks dan uji coba terhadap hal tersebut melalui melakukan eksperimen terhadap beberapa metode-metode kategorisasi.

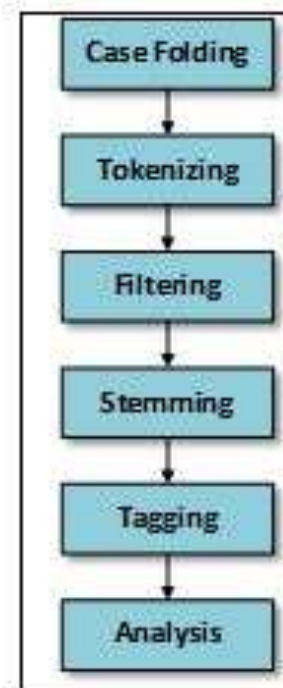
Pada kategorisasi teks, diberikan sekumpulan kategori (label) dan koleksi data teks yang berfungsi sebagai data latih, yaitu data yang digunakan untuk membangun model, dan kemudian dilakukan proses untuk menemukan kategori yang tepat untuk data teks, yaitu data yang digunakan untuk menentukan akurasi dari model. Misalkan ada sebuah dokumen  $x$  sebagai input, maka output yang dihasilkan oleh model tersebut adalah kelas atau kategori  $y$  dari beberapa kategori tertentu yang telah didefinisikan sebelumnya ( $y_1, \dots, y_k$ ). Adapun contoh dari pemanfaatan kategorisasi teks adalah pengkategorisasian berita ke dalam beberapa kategori seperti bisnis, teknologi, kesehatan dan lain sebagainya; pengkategorisasian e-mail sebagai spam atau bukan; pengkategorisasian kilasan film sebagai film favorit, netral atau tidak favorit dan pengkategorisasian paper yang menarik dan tidak menarik.

Berdasarkan uraian di atas, dapat disimpulkan bahwa kategorisasi teks adalah sebuah proses untuk mengkategorisasi sebuah teks sesuai dengan

kategori yang telah ditentukan yang bertujuan untuk mempermudah dalam mengorganisir teks dalam jumlah besar.

#### Tahap-Tahap Kategorisasi Teks

Inti dari suatu sistem klasifikasi adalah tahap penemuan pola (pattern discovery) namun secara lengkap proses kategorisasi teks dibagi menjadi 3 tahap utama, yaitu proses awal terhadap teks (text pre-processing), transformasi teks ke dalam bentuk antara (text transformation/ feature generation), dan penemuan pola (pattern discovery). Masukan awal dari proses ini adalah suatu data teks dan menghasilkan keluaran berupa pola sebagai hasil interpretasi (Even dan Zohar, 2002). Adapun ilustrasi tahapan di dalam kategorisasi teks dapat dilihat pada Gambar 1.



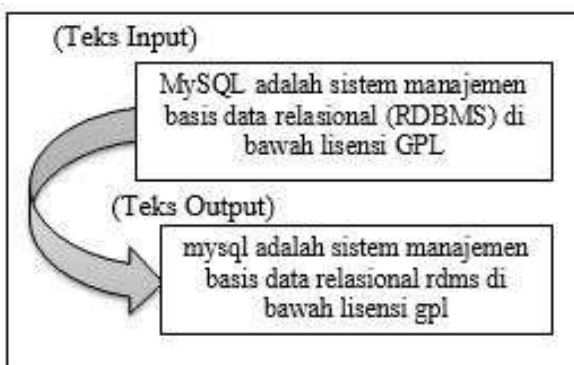
Gambar 1. Tahapan Klasifikasi Teks (Suhartono, 2013)

#### Text Preprocessing

Tahapan awal dari klasifikasi teks ialah *Text Preprocessing* yang bertujuan untuk mempersiapkan teks menjadi data yang akan mengalami pengolahan pada tahapan berikutnya. Beberapa contoh tindakan yang dapat dilakukan pada tahap ini, seperti proses *Tokenizing*, yaitu memecah suatu kalimat menjadi sekumpulan kata. Selain itu pada tahapan ini biasanya juga dilakukan *Case Folding*, yaitu perubahan karakter huruf menjadi huruf kecil. Adapun penjelasan rinci mengenai *Case Folding* dan *Tokenizing* adalah:

##### 1. Case Folding

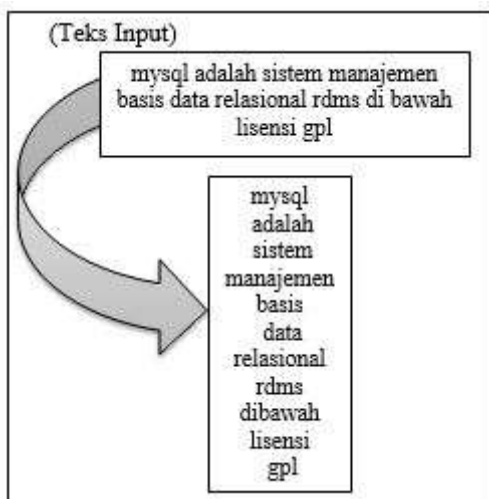
*Case Folding* yaitu mengubah semua huruf dalam teks menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter. Contoh dari tahap ini dapat dilihat pada Gambar 2.



Gambar 2. Proses Case Folding

2. Tokenizing

Tokenizing adalah sebuah proses untuk memilah isi teks sehingga menjadi satuan kata-kata. Proses ini cukup rumit untuk sebuah program komputer karena beberapa karakter dapat dijadikan sebagai pembatas (*delimiter*) dari token-token itu sendiri. Contoh dari tahap ini dapat dilihat pada Gambar 3.



Gambar 3. Proses Tokenizing

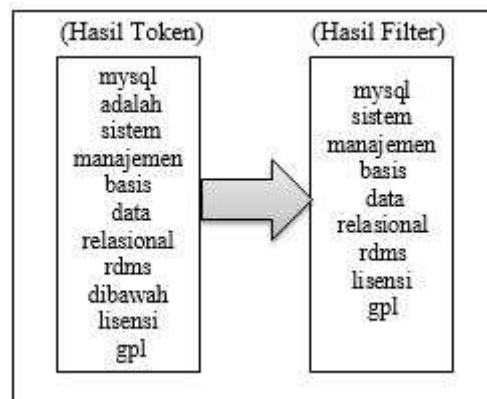
Text Transformation

Proses transformasi ini dilakukan dengan mengurangi jumlah kata-kata yang ada dengan penghilangan *stopword* dan juga dengan mengubah kata-kata dalam bentuk dasarnya (*stemming*). *Stopword* adalah kata-kata yang bukan merupakan ciri (kata unik) dari suatu teks seperti kata sambung, kata kepunyaan. Proses penghilangan *stopword* digunakan karena proses ini akan sangat mengurangi beban kerja sistem.

1. Filtering

Tahapan *filtering* merupakan tahap mengambil kata-kata penting dari hasil *token*. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist / stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh *stopwords* adalah “yang”, “dan”, “di”, “dari” dan seterusnya. Dengan menggunakan daftar *stoplist*, maka setiap kata dalam koleksi akan dicocokkan dengan kata-kata yang ada

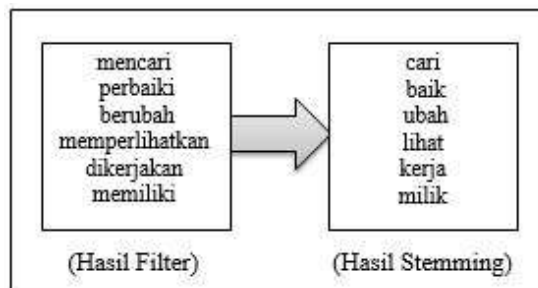
dalam *stoplist*. Apabila kedua kata itu sama, maka kata itu akan dibuang dari koleksi.



Gambar 4. Proses Filtering

2. Stemming

*Stemming* merupakan suatu proses untuk mereduksi kata ke bentuk dasarnya. *Stemming* dengan kata lain merupakan suatu proses yang menyediakan suatu pemetaan antara berbagai kata dengan morfologi yang berbeda menjadi satu bentuk dasar (*stem*). Tahap *stemming* merupakan tahap mencari akar (*root*) kata dari tiap kata hasil *filtering*. Contoh dari tahapan ini pada teks berbahasa Indonesia dapat dilihat pada Gambar 5.



Gambar 5. Proses Stemming

Naïve Bayes Classifier (NBC)

*Naïve Bayes Classifier* (NBC) merupakan salah satu metode yang digunakan untuk mengklasifikasikan sekumpulan teks. 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 (Ismail, 2008). Performa pengklasifikasi biasanya diukur dengan ketepatan (atau tingkat galat). Teori bayes sebagai dasar pada klasifikasi di sini adalah B adalah sampel dengan kelas (label) yang tidak diketahui, di mana A merupakan hipotesa bahwa B adalah data dengan klas (label). P(B) adalah peluang hipotesa B, sedangkan P(A) adalah peluang data sampel yang diamati. P(B|A) adalah peluang data sampel B, bila diasumsikan bahwa hipotesa benar. Kaitannya dengan masalah klasifikasi, yang dihitung adalah P(A|B), yaitu peluang bahwa hipotesa benar untuk data sampel B yang diamati (lihat persamaan 2.2). *Naïve Bayes Classifier* adalah metode *classifier* yang berdasarkan probabilitas dan teorema bayesian dengan asumsi bahwa setiap variabel B bersifat

independen. Dengan kata lain, *Naïve Bayes Classifier* mengasumsikan bahwa keberadaan sebuah atribut (variabel) tidak ada kaitannya dengan keberadaan atribut (variabel) yang lain.

Klasifikasi *Naïve Bayes Classifier* dilakukan dengan cara mencari probabilitas  $P(V = v_j | D = d_i)$ , yaitu probabilitas kategori  $v_j$  jika diketahui teks  $d_i$ . Pada saat klasifikasi, pendekatan Bayes akan menghasilkan label kategori yang paling tinggi probabilitasnya ( $V_{MAP}$ ) dengan masukan atribut ( $a_1, a_2, a_3, \dots a_n$ ) dimana  $a_1$  adalah kata pertama,  $a_2$  dan seterusnya. Selanjutnya klasifikasi teks mencari nilai maksimum dari kategori teks (McCallum and Nigam, 1998):

$$V_{MAP} = \underset{v_j \in V}{\operatorname{arg\,max}} P(v_j | a_1, a_2, \dots, a_n) \quad (2.1)$$

Teorema Bayes menyatakan probabilitas bersyarat adalah:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.2)$$

Dengan menggunakan Teorema Bayes (Persamaan 2.2), maka Persamaan (2.1) ini dapat ditulis:

$$V_{MAP} = \underset{v_j \in V}{\operatorname{arg\,max}} \frac{P(a_1, a_2, \dots, a_n | v_j) P(v_j)}{P(a_1, a_2, \dots, a_n)} \quad (2.3)$$

Karena  $P(a_1, a_2, \dots, a_n)$  nilainya konstan untuk semua  $v_j$  sehingga Persamaan 2.3 dapat ditulis menjadi Persamaan 2.4.

$$V_{MAP} = \underset{v_j \in V}{\operatorname{arg\,max}} P(a_1, a_2, \dots, a_n | v_j) P(v_j) \quad (2.4)$$

Tingkat kesulitan menghitung  $P(a_1, a_2, \dots, a_n | v_j)$  menjadi tinggi karena jumlah  $P(a_1, a_2, \dots, a_n | v_j)$  bisa jadi akan sangat besar. Ini disebabkan jumlah *term* tersebut sama dengan jumlah semua kombinasi posisi kata dikali dengan jumlah kategori. *Naïve Bayes Classifier* menyederhanakan hal ini dengan mengasumsikan bahwa di dalam setiap kategori, setiap kata independen satu sama lain. Dengan kata lain Persamaan 2.4 dapat ditulis menjadi Persamaan 2.5.

$$P(a_1, a_2, a_3, \dots, a_n | v_j) = \prod_i P(a_i | v_j) \quad (2.5)$$

Substitusi Persamaan 2.5 dengan Persamaan 2.4 akan menghasilkan Persamaan 2.6.

$$V_{MAP} = \underset{v_j \in V}{\operatorname{arg\,max}} P(v_j) \prod_i P(w_k | v_j) \quad (2.6)$$

$V_{MAP}$  merupakan nilai probabilitas hasil perhitungan *Naïve Bayes Classifier* untuk nilai fungsi target yang bersangkutan. Frekuensi kemunculan kata menjadi dasar perhitungan nilai dari  $P(v_j)$  dan  $P(w_k | v_j)$ . Sehingga pelatihan teks untuk algoritma naive bayes dapat didefinisikan menjadi Persamaan 2.7 dan Persamaan 2.8.

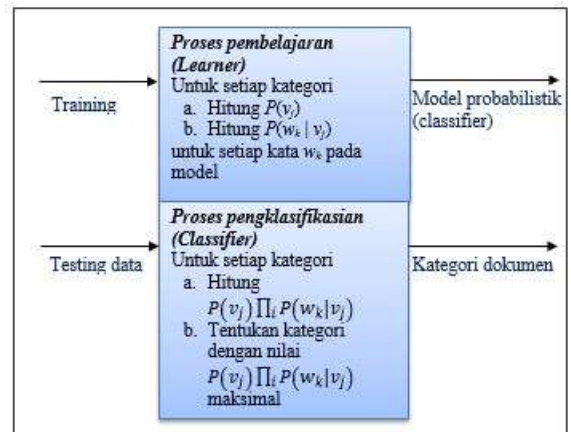
$$P(v_j) = \frac{|docs_j|}{|dokpel|} \quad (2.7)$$

$$P(w_k | v_j) = \frac{n_k + 1}{n + |kosakata|} \quad (2.8)$$

$P(v_j)$  dan probabilitas kata  $w_k$  untuk setiap kategori  $P(w_k | v_j)$  dihitung pada saat pelatihan. Dimana  $|docs_j|$  adalah jumlah data pada kategori  $j$  dan  $|dokpel|$  adalah jumlah data yang digunakan dalam pelatihan. Sedangkan  $n_k$  adalah jumlah kemunculan kata  $w_k$  pada kategori  $v_j$ ,  $n$

adalah jumlah semua kata pada kategori  $v_j$  dan  $|kosakata|$  adalah jumlah kata pada semua data latih.

Berdasarkan uraian persamaan di atas, maka ilustrasi tahapan proses *training* dan klasifikasi dapat dilihat pada Gambar 6.



Gambar 6. Tahapan klasifikasi Dokumen Algoritma *Naïve Bayes*

### HASIL PENELITIAN DAN PEMBAHASAN

#### Deskripsi Sistem

Sistem klasifikasi dan pencarian buku referensi akademik menggunakan metode *Naïve Bayes Classifier* (NBC) ini merupakan aplikasi yang membantu pihak perpustakaan dalam menentukan kategori yang tepat pada setiap buku serta memudahkan user dalam mencari buku yang sesuai dengan keinginan mereka.

Kategori-kategori yang digunakan dalam pengklasifikasian buku ini ada lima, yaitu: pemrograman komputer, jaringan komputer, database, multimedia, dan sistem operasi.

#### Perancangan Sistem

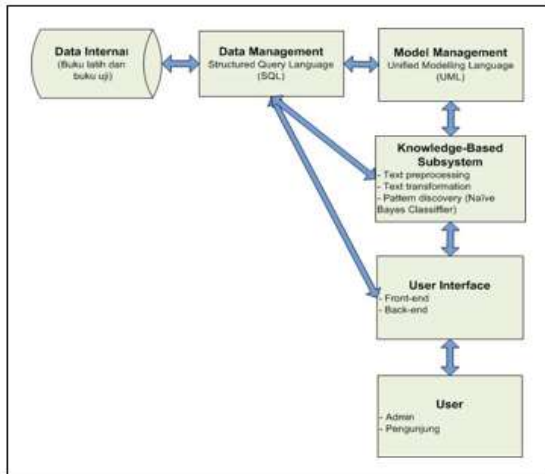
Perancangan sistem adalah rencana sistematis dalam proses pengembangan sistem setelah memiliki data yang cukup untuk menunjang kegiatan pengembangan sistem. Dalam perancangan sistem ini digunakan dua diagram UML yaitu use case diagram dan activity diagram.

Pada sistem ini terdapat dua level hak pengguna yaitu admin dan pengunjung sebagai user. Level admin berfungsi manajemen buku-buku pada aplikasi ini, seperti menambah dan menghapus buku, melakukan klasifikasi buku, serta menambah dan menghapus admin. Disamping itu admin juga dapat melihat statistik web dan daftar buku hasil pelatihan.

Level user digunakan oleh pengunjung perpustakaan untuk melihat daftar buku terbaru, melakukan pencarian buku, dan melihat daftar seluruh buku hasil dari klasifikasi.

#### Perancangan Arsitektur Sistem

Aplikasi ini dibuat dengan rancangan arsitektur dari sistem klasifikasi dan pencarian buku yang dapat dilihat pada gambar 7.



Gambar 7. Arsitektur Sistem Klasifikasi dan Pencarian Buku.

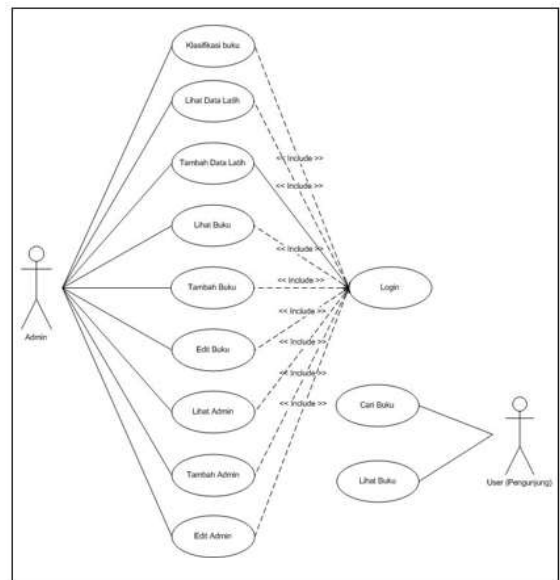
Pada arsitektur klasifikasi dan pencarian buku ini ada enam komponen :

1. Data internal, data yang diperlukan pada aplikasi ini adalah buku-buku yang berasal dari Badan Perpustakaan Daerah Kalimantan Timur dan katalog buku *online* yang terdapat pada internet.
2. Data *Management*, pada sistem ini untuk mengolah data menggunakan MySQL.
3. Model *Management*, dalam mengembangkan sistem ini menggunakan *Unified Modelling Language* (UML).
4. *Knowledge-Based Subsystem*, dalam proses klasifikasi ini terdapat tiga tahapan yaitu: *text preprocessing*, *text transformation*, dan *pattern discovery* dengan menggunakan metode *Naive Bayes Classifier* (NBC).
5. *User Interface*, sistem ini mempunyai beberapa halaman, yaitu halaman *front-end* bagi pengunjung dan halaman *back-end* untuk admin.
6. *User*, pengguna akhir dari sistem ini adalah admin dan pengunjung perpustakaan.

**Use Case Diagram**

Perancangan *use case diagram* merupakan tahap awal dan utama dalam proses pengembangan sistem, dimana dalam tahapan ini dijelaskan dan didefinisikan fungsi-fungsi serta fitur-fitur apa saja yang dapat disediakan oleh sistem. Pada *use case diagram* mempunyai dua aktor, yaitu *user* (pengunjung) dan *admin* seperti digambarkan pada gambar 8.

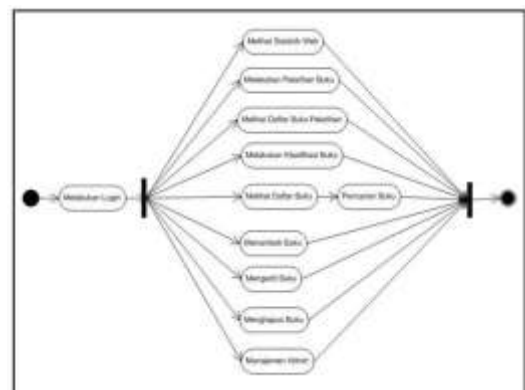
Seperti terlihat pada gambar 8, *user* berperan dalam melakukan pencarian buku dan melihat daftar buku yang ada. Sedangkan admin berperan dalam melakukan klasifikasi buku, menambah data latih, melihat daftar buku hasil pelatihan, menambah buku, melihat daftar buku dari hasil penambahan buku, manajemen buku yang meliputi menghapus dan mengedit data buku, melihat daftar admin, menambah admin, manajemen admin yang meliputi menghapus dan mengedit data admin.



Gambar 8. Use Case Diagram Pencarian dan Klasifikasi Buku

**Activity Diagram Sistem**

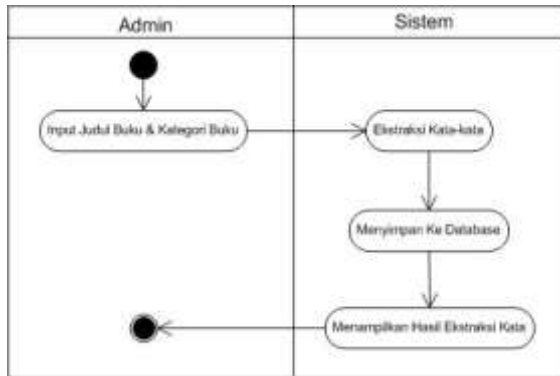
Gambar 9 adalah *activity diagram* untuk admin, dimana admin melakukan login terlebih dahulu untuk bisa masuk ke dalam sistem. Kemudian admin dapat melihat statistik web yang meliputi jumlah data latih, jumlah data buku, dan jumlah admin. Admin juga dapat melakukan pelatihan buku serta melihat daftar buku-buku hasil pelatihan tersebut. Selain itu admin dapat melakukan klasifikasi buku untuk menentukan kategori buku yang tepat. Selanjutnya admin dapat melakukan penambahan buku dengan menginputkan rincian buku yang meliputi: judul buku, kode buku, penulis buku, sinopsis buku, dan penerbit buku. Pada menu tambah buku ini, sistem secara otomatis akan menentukan kategori buku yang berhasil diinputkan. Penambahan Buku tersebut pada nantinya akan ditampilkan juga pada halaman pengunjung. Fungsi admin selanjutnya adalah melihat dan melakukan pencarian daftar buku yang berhasil ditambah. Admin pun dapat memanajemen buku yang meliputi pengeditan dan penghapusan buku. Fungsi yang terakhir ialah manajemen admin, yang meliputi penambahan admin, edit data admin, dan menghapus data admin.



Gambar 9. Activity Diagram Admin

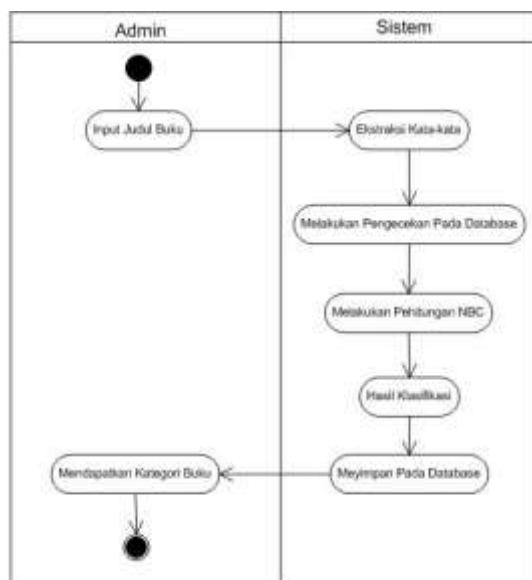


Di bawah ini gambar 10 adalah *activity diagram* dalam aktifitas menambah data latihan. Pertama admin melakukan penginputan judul buku beserta kategori buku tersebut, kemudian sistem akan mengekstrak kata-kata pada judul tersebut (*text preprocessing* dan *text transformation*) lalu menyimpan hasil dari ekstraksi tersebut dalam database setelah itu sistem akan menampilkan hasil ekstraksi judul tersebut kepada admin.



Gambar 10. Activity Diagram Admin Tambah Data Latihan

Gambar 11 merupakan *activity diagram* dalam proses klasifikasi buku. Pertama admin menginputkan judul buku, setelah itu sistem akan mengekstrak kata-kata pada judul tersebut (*text preprocessing* dan *text transformation*), selanjutnya system akan mengecek kata-kata tersebut pada database, jika kata-kata tersebut ada dalam database maka akan dilakukan perhitungan NBC, setelah nilai perhitungan didapatkan sistem akan menampilkan hasil klasifikasi berupa kategori buku tersebut serta menyimpannya dalam *database*.



Gambar 11. Activity Diagram Admin Klasifikasi Buku

**Implementasi dan Pengujian Sistem**

Implementasi Sistem merupakan tahapan realisasi yang dilakukan setelah rancangan aplikasi. Pengujian sistem merupakan tahapan dimana semua

fungsi yang terdapat pada sistem diuji dan dievaluasi.

**Implementasi dan Pengujian Sistem User**

Halaman awal dari fungsi *user* dapat dilihat pada gambar 12. Pada halaman awal ini, *user* dapat melihat daftar buku terbaru yang berhasil diunggah oleh *admin*. Selain itu juga terdapat menu pencarian cepat berdasarkan judul buku.



Gambar 12. Halaman awal User

Setelah *User* mengetikkan kata kunci pada kolom pencarian cepat maka sistem akan menampilkan hasil pencarian tersebut. Apabila tidak terdapat buku yang sesuai, maka akan muncul notifikasi bahwa hasil pencarian tidak ditemukan.

*User* juga dapat melihat *detail* atau rincian buku dari hasil pencarian maupun dari daftar buku terbaru yang terdapat pada halaman awal.

Halaman selanjutnya adalah pencarian buku. Pada halaman ini *User* dapat melakukan pencarian berdasarkan tiga hal yaitu judul buku, penulis buku, dan kategori buku. Halaman pencarian buku dapat dilihat pada gambar 13.



Gambar 13. Halaman Pencarian Buku

Hasil pencarian tersebut disertai rincian dari buku tersebut. Adapun rincian yang dimaksud antara lain: kode buku, judul buku, penuli, penerbit, dan kategori. *User* juga dapat melihat rincian yang lebih lengkap seperti sinopsis buku dan cover buku dengan mengklik tombol "Details".

Menu terakhir dari *User* (pengunjung) adalah melihat daftar buku yang berhasil diunggah oleh admin. Halaman ini tampak pada gambar 14. Terlihat pada gambar tersebut *list* atau daftar buku disertai rinciannya berupa: kode buku, judul buku, penulis, penerbit dan kategori. Selain itu *User* juga dapat melihat rincian lebih lengkap dari buku tersebut seperti sinopsis dan *cover* buku dengan mengklik tombol “Details”.



Gambar 14. Halaman Daftar Buku

**Implementasi dan Pengujian Sistem Admin**

Administrator merupakan pengguna sistem dengan hak akses yang lebih tinggi dibandingkan *user* atau pengunjung. Fungsi admin pada intinya adalah melakukan pelatihan buku, melakukan klasifikasi buku, mengunggah serta manajemen buku, dan yang terakhir manajemen data admin. Untuk mengakses fungsi-fungsi tersebut diperlukan *username* dan *password* untuk login ke dalam sistem. Halaman login administrator bisa dilihat pada gambar 15.



Gambar 15. Halaman Login Admin

Setelah berhasil login, admin akan masuk ke halaman *dasboard* dari aplikasi ini. Pada halaman ini admin dapat melihat statistik website yang meliputi jumlah buku yang berhasil diunggah, jumlah data latih, serta jumlah admin. Halaman *dashboard* dapat dilihat pada gambar 16.



Gambar 16. Halaman Dashboard

Halaman pertama yang dapat diakses oleh admin adalah halaman daftar data latih. Pada halaman ini admin dapat melihat daftar buku yang berhasil ditambahkan melalui proses pelatihan. Seperti tampak pada gambar 17, daftar buku hasil pelatihan tersebut disertai dengan rincian berupa judul buku, tanggal unggah, dan kategori. Selain itu terdapat pula *form* yang digunakan untuk melakukan pencarian berdasarkan judul buku.



Gambar 17. Daftar Data Latih

Admin juga dapat menghapus data latih. Ketika akan dihapus akan muncul notifikasi apakah data tersebut akan dihapus apa tidak.

Data latih pada penelitian ini berasal dari katalog online dari berbagai penerbit buku berbahasa Indonesia. Penerbit tersebut diantaranya adalah Andi Publisher, Informatika dan Media Kita. Total buku yang digunakan pada data latih ini berjumlah 250 buku. Dari jumlah tersebut masing-masing kategori terdiri dari 50 buku.

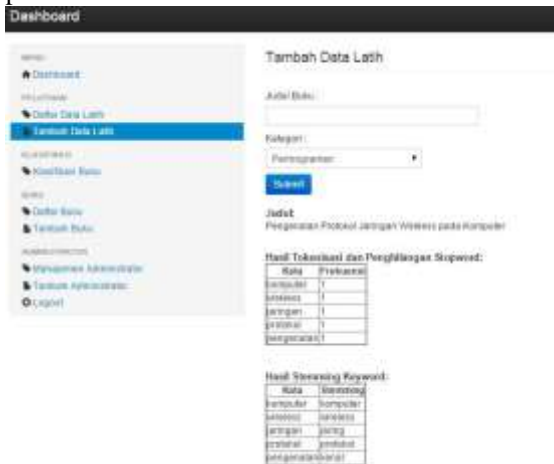
Halaman selanjutnya yang dapat diakses oleh administrator adalah halaman tambah data latih. Pada halaman ini admin dapat menambah data latih dengan menginputkan judul buku lalu memilih kategori yang sesuai dengan judul tersebut. Halaman ini dapat dilihat pada gambar 18.



Gambar 18. Halaman Tambah Data Latih

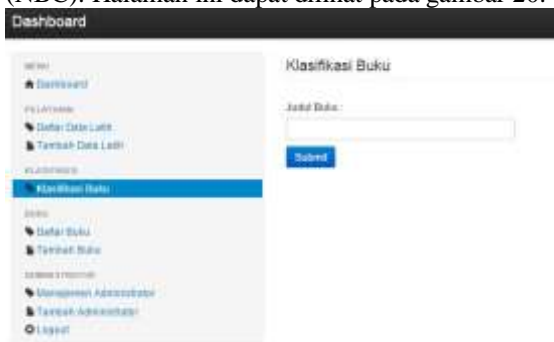
Gambar 19 adalah contoh kasus dari menu Tambah Data Latih. Seperti yang terlihat pada gambar tersebut, data latih yang ingin ditambahkan berjudul “Pengenalan Protokol Jaringan Wireless Pada Komputer”. Setelah judul tersebut diinputkan

dan disubmit oleh admin, maka akan terjadi proses *Preprocessing* dan *Text Transformation*. Dimana proses tersebut meliputi empat tahapan yaitu *Casefolding*, *Tokenizing*, *Filtering*, dan *Stemming*. *Casefolding* adalah proses mengubah semua huruf dalam teks menjadi huruf kecil. *Tokenizing* adalah proses untuk memilah isi teks sehingga menjadi satuan kata-kata. *Filtering* adalah proses penghapusan kata yang dianggap kurang penting, seperti kata sambung serta menghilangkan kata-kata yang bukan merupakan alfabet. Pada contoh kasus dengan judul “Pengenalan Protokol Jaringan Wireless Pada Komputer” kata “pada” dihilangkan karena termasuk dalam *Stoplist*. Proses selanjutnya adalah proses *Stemming*. *Stemming* adalah proses untuk mereduksi kata ke bentuk dasarnya. Seperti contoh kasus pada gambar 19, kata “pengenalan” berubah menjadi kata “kenal”. Hasil dari *Casefolding*, *Tokenizing*, *Filtering*, dan *Stemming* ini dapat dilihat pada tabel pertama dan kedua pada gambar 19. Hasil dari penambahan data latih ini juga akan disimpan pada database sebagai acuan pada proses klasifikasi.



Gambar 19. Contoh Hasil Penambahan Data Latih

Halaman administrator selanjutnya adalah Halaman Klasifikasi Buku. Pada halaman ini admin dapat melakukan klasifikasi atau penentuan kategori buku menggunakan metode *Naïve Bayes Classifier* (NBC). Halaman ini dapat dilihat pada gambar 20.



Gambar 20. Halaman Klasifikasi Buku

Pada contoh kasus kali ini, buku yang digunakan berjudul “Membangun Aplikasi Berbasis PHP dan MySQL”. Seperti halnya proses pelatihan,

proses klasifikasi judul juga melewati tahapan *Preprocessing* dan *Text Transformation*. Setelah melewati dua tahapan tersebut didapatkan hasil seperti gambar 21. Seperti terlihat pada gambar tersebut, hasil *Stemming* menunjukkan bahwa buku dengan judul “Membangun Aplikasi Berbasis PHP dan MySQL” menghasilkan beberapa kata. Kata-kata tersebut adalah “mysql”, “php”, “basis”, “aplikasi”, dan “bangun”. Dari kumpulan kata-kata inilah penentuan kategori akan dilakukan.



Gambar 21. Hasil *Preprocessing* dan *Text Transformation* Klasifikasi Buku

Setelah melewati proses akhir yaitu *Stemming*, proses selanjutnya adalah penentuan kategori buku menggunakan metode *Naïve Bayes Classifier*. Hal pertama yang perlu dilakukan adalah menghitung menghitung nilai  $P(v_j)$ . Nilai  $P(v_j)$  setiap kategori didapat dari jumlah data latih pada kategori tertentu dibagi total keseluruhan data latih yang digunakan pada sistem. Rumus:  $P(v_j) = \frac{|docs_j|}{|dokpel|}$

Hasil dari perhitungan nilai  $P(v_j)$  ini dapat dilihat pada Tabel 1.

Tabel 1. Nilai  $P(v_j)$  Setiap Kategori

Kategori	Nilai $P(v_j)$
Penrograman	$\frac{50}{250} = 0.2$
Jaringan	$\frac{50}{250} = 0.2$
Database	$\frac{50}{250} = 0.2$
Multimedia	$\frac{50}{250} = 0.2$
Sistem Operasi	$\frac{50}{250} = 0.2$

Hasil kumpulan kata-kata yang terdiri dari “mysql”, “php”, “basis”, “aplikasi”, dan “bangun” tadi, selanjutnya akan melalui proses pengecekan setiap kata, di masing-masing kategori pada data latih.

Hasil kumpulan kata-kata yang terdiri dari “mysql”, “php”, “basis”, “aplikasi”, dan “bangun” tadi, selanjutnya akan melalui proses pengecekan



setiap kata, di masing-masing kategori pada data latih. Hal tersebut dapat dilihat pada gambar 22.

Hasil Pengecekan kata pada Data Latih:

Kata	Stemming	Kategori	Frekuensi
mysql	mysql	pemrograman	6
php	php	pemrograman	12
aplikasi	aplikasi	pemrograman	10
membangun	bangun	pemrograman	6
basis	basis	pemrograman	3
berbasis	basis	jaringan	4
membangun	bangun	jaringan	16
aplikasi	aplikasi	database	10
mysql	mysql	database	18
basis	basis	database	6
php	php	database	3
membangun	bangun	database	2
aplikasi	aplikasi	multimedia	3
aplikasi	aplikasi	sistem operasi	1
berbasis	basis	sistem operasi	2
membangun	bangun	sistem operasi	2

Gambar 22. Hasil Pengecekan Kata pada Data Latih

Proses selanjutnya yaitu menentukan probabilitas bobot kata sesuai kategori buku dengan menghitung nilai  $P(w_k|v_j)$  yang ditentukan dengan rumus:  $(w_k|v_j) = \frac{r_k+1}{n+|kata|}$ . Nilai  $P(w_k|v_j)$  pada masing-masing kata dapat dilihat pada tabel 2.

Tabel 2. Nilai  $P(w_k|v_j)$  Masing-Masing Kata

Kata	Nilai $P(w_k v_j)$ setiap kategori				
	Pemrograman	Jaringan	Database	Multimedia	Sistem Operasi
mysql	7/1639	1/1667	19/1657	1/1706	1/1605
php	13/1639	1/1667	4/1657	1/1706	1/1605
basis	4/1639	5/1667	7/1657	1/1706	3/1605
aplikasi	11/1639	1/1667	11/1657	1/1706	2/1605
bangun	7/1639	16/1667	3/1657	1/1706	3/1605

Selanjutnya menentukan nilai  $V_{MAP}$  pada setiap kategori.  $V_{MAP}$  merupakan nilai probabilitas hasil perhitungan *Naive Bayes Classifier*. Nilai  $V_{MAP}$  didapatkan dari hasil perkalian nilai  $P(v_j)$  pada masing-masing kategori dan nilai  $P(w_k|v_j)$  pada masing-masing kata disetiap kategori. Nilai  $V_{MAP}$  pada setiap kategori ini dapat dilihat pada Tabel 3.

Tabel 3. Penentuan Nilai  $V_{MAP}$

Kategori	Nilai $P(v_j)$	Nilai $P(w_k v_j)$					Nilai $V_{MAP}$
		mysql	php	basis	aplikasi	bangun	
Pemrograman	0.2	7/1639	13/1639	4/1639	11/1639	7/1639	$4.7 \times 10^{-13}$
Jaringan	0.2	1/1667	1/1667	5/1667	1/1667	16/1667	$1.3 \times 10^{-13}$
Database	0.2	19/1657	4/1657	7/1657	11/1657	3/1657	$2.8 \times 10^{-13}$
Multimedia	0.2	1/1706	1/1706	1/1706	1/1706	1/1706	$5.5 \times 10^{-17}$
Sistem Operasi	0.2	1/1605	1/1605	3/1605	2/1605	3/1605	$3.3 \times 10^{-16}$

Dari nilai  $V_{MAP}$  tersebut kemudian diambil nilai terbesar yang menjadi acuan kategori dari

judul tersebut. Pada judul tersebut, didapat bahwa nilai  $V_{MAP}$  untuk kategori Pemrograman, memiliki nilai tertinggi dibandingkan dengan kategori lainnya. Sehingga, judul tersebut diklasifikasikan sebagai kategori **Pemrograman**.

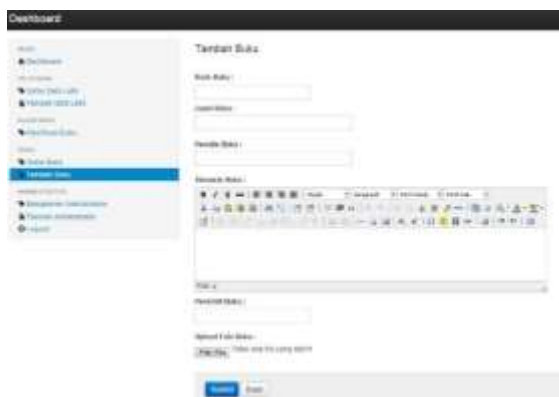
Pada penelitian ini data uji untuk klasifikasi berasal dari koleksi buku yang terdapat pada Perpustakaan Daerah Provinsi Kalimantan Timur. Data Uji yang digunakan berjumlah 150 buku. Berikut ini adalah tabel 4 hasil klasifikasi buku-buku tersebut.

Tabel 4. Hasil Klasifikasi Buku Perpustakaan Kaltim

No.	Judul	Kategori	Hasil Klasifikasi
1	PHP dan MySQL untuk Pemula	Pemrograman	Pemrograman
2	Multimedia alat untuk meningkatkan keunggulan bersaing	Multimedia	Multimedia
3	Pemrograman Java menggunakan IDE Eclipse Callisto	Pemrograman	Pemrograman
4	Protokol-Protokol Esensial Internet	Jaringan	Jaringan
5	Instalasi dan konfigurasi Jaringan Komputer	Jaringan	Jaringan
6	Dasar Teknik Pengolahan Obyek di CorelDraw	Multimedia	Multimedia
7	Dasar Pemrograman web dinamis dengan JSP (Java Server Pages)	Pemrograman	Pemrograman
8	Pengenalan Unix dan Linux	Sistem Operasi	Sistem Operasi
9	Panduan Lengkap Menggunakan Mac OS X Leopard	Sistem Operasi	Sistem Operasi
10	Pembuatan Desain Grafis dengan Corel Draw 12	Multimedia	Multimedia
11	Panduan Praktis Borland Delphi 6.0	Pemrograman	Database
12	Komunikasi & Jaringan Nirkabel	Jaringan	Jaringan
13	Microsoft Access	Database	Database
14	Oracle 8i/9i Backup & Recovery	Database	Database
15	Data Recovery Teknik Praktis Menyimpan & Menyelamatkan Data Komputer	Sistem Operasi	Pemrograman
16	Menguasai Java 2 & Object Oriented Programming	Pemrograman	Pemrograman
17	PHP dan PostgreSQL	Pemrograman	Pemrograman
18	Macromedia Flash Professional 8	Multimedia	Multimedia
19	Just XML	Pemrograman	Tidak masuk kategori
20	Mudah Beralih ke Windows 7	Sistem Operasi	Sistem Operasi

Hasil klasifikasi yang menunjukkan “Tidak masuk kategori”, menjelaskan bahwa kata hasil Stemming yang terdapat pada judul tersebut tidak satupun terdapat dalam database pelatihan.

Fitur yang diperuntukkan bagi admin selanjutnya adalah Halaman Tambah Buku. Pada halaman ini, admin dapat mengunggah buku yang nantinya akan ditampilkan pada halaman pengunjung. Rincian buku yang diinputkan oleh admin antara lain: kode buku, judul buku, penulis buku, synopsis buku, penerbit buku, dan photo buku. Pada fitur Tambah Buku ini penerapan metode *Naive Bayes Classifier* (NBC) juga diimplementasikan, ketika buku tersebut berhasil diunggah, sistem secara otomatis akan melakukan klasifikasi untuk menentukan kategori buku. Sehingga admin tidak perlu lagi untuk menginputkan kategori buku tersebut. Jika judul yang diinputkan sudah terdapat dalam *database* maka akan muncul notifikasi bahwa judul tersebut sudah ada sehingga tidak dapat ditambahkan lagi. Halaman Tambah Buku dapat dilihat pada gambar 23.



Gambar 23. Halaman Tambah Buku

Admin juga dapat menyunting atau mengubah rincian buku dari daftar buku tersebut. Rincian buku yang dapat diubah adalah kode buku, penulis, dan penerbit. Sedangkan judul dan kategori buku tidak dapat diubah karena kedua hal tersebut didapat dari hasil proses klasifikasi.

Halaman admin berikutnya adalah Manajemen Admin. Pada halaman ini admin dapat melihat daftar admin yang ada beserta rinciannya berupa Nama dan *Username* admin. Halaman ini juga berfungsi untuk memajemen data admin, meliputi penghapusan dan mengubah data admin. Selain itu pada halaman ini admin dapat mengubah data admin serta menambah jumlah admin seperti tampak pada gambar 24.



Gambar 24. Halaman Tambah Admin

## KESIMPULAN

Kesimpulan yang dapat diambil berdasarkan penelitian mengenai Klasifikasi dan Pencarian Buku Referensi Akademik Menggunakan Metode *Naïve Bayes Classifier* antara lain:

1. Metode *Naïve Bayes Classifier* dapat digunakan untuk melakukan klasifikasi buku berbahasa Indonesia yang berada di Perpustakaan Daerah Provinsi Kalimantan Timur.
2. Jumlah data latih sangat berpengaruh dalam proses klasifikasi, karena semakin banyak data latih maka tingkat keakuratan dalam penentuan kategori akan semakin tinggi begitu juga sebaliknya.

## SARAN

Berdasarkan hasil penelitian dari Implementasi sistem Klasifikasi dan Pencarian Buku Referensi Akademik Menggunakan Metode *Naïve Bayes*

*Classifier*, penulis memiliki beberapa saran untuk pengembangan selanjutnya, antara lain:

1. Perlu adanya penelitian yang lebih mendalam untuk mengetahui bagaimana pengaruh kinerja metode *Naïve Bayes Classifier* ini terhadap kategori lain.
2. Memperbanyak jumlah data latih sehingga tingkat akurasi pada proses klasifikasi semakin tinggi serta menghindari kemungkinan hasil klasifikasi yang tidak masuk dalam kategori manapun.

## DAFTAR PUSTAKA

- [1] Even, Y and Zohar. 2002. *Introduction to Text Mining*. National Center for Supercomputing Applications University of Illinois. Chicago.
- [2] McCallum, A and Nigam, K. 1998. *A Comparison of Event Models for Naïve Bayes Text Classification*. Journal of Carnegie Mellon University. Pittsburgh.
- [3] Samodra. 2009. *Klasifikasi Dokumen Teks Berbahasa Indonesia dengan Menggunakan Naïve Bayes*. Prosiding Seminar Nasional Electrical, Informatics, and It's Educations. Institut Teknologi Sepuluh Nopember. Surabaya.
- [4] Suhartono, E. 2013. *Sistem Klasifikasi Dan Pencarian Dokumen Teks dengan Algoritma Naïve Bayes Classifier Dan Vector Space Model*. Thesis Program Studi Megister Sistem Informasi Universitas Diponegoro. Semarang.