

Klasifikasi Judul Buku dengan Algoritma *Naive Bayes* dan Pencarian Buku pada Perpustakaan Jurusan Teknik Elektro

Vidya Rizqiyani¹, Anggraini Mulwinda², dan Riana Defi Mahadji Putri³

Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang

Kampus Sekaran, Gunungpati, Semarang, 50229, Indonesia

vrizqiyani@gmail.com¹, anggrainimulwinda@mail.unnes.ac.id², putri13.rdmp@gmail.com³

Abstrak— Berdasarkan Undang-undang Republik Indonesia Nomor 43 Tahun 2007 tentang Perpustakaan menyebutkan bahwa penyelenggaraan pendidikan nasional memerlukan pengembangan layanan perpustakaan berbasis teknologi informasi dan komunikasi sehingga Universitas Negeri Semarang khususnya Jurusan Teknik Elektro perlu adanya pengembangan layanan tersebut. Ketiadaan pengembangan layanan tersebut di perpustakaan Jurusan Teknik Elektro mengakibatkan pihak pengelola dan mahasiswa pengunjung perpustakaan mengalami kesulitan dalam mengelola dan menemukan referensi buku yang dicari. Pada penelitian ini, metode yang digunakan yaitu *waterfall* dengan pengujian *performance measure*, uji *black box*, dan uji sistem oleh ahli. Adapun pengujian *performance measure* memperoleh nilai *precision* 94,56%, *recall* 88,20%, *f-measure* 90,46%, dan akurasi 97,78%. Sedangkan uji sistem oleh ahli menghasilkan persentase rata-rata 87,3%. Dari hasil-hasil tersebut, disimpulkan bahwa sistem berfungsi dengan baik serta algoritma *Naive Bayes* dapat digunakan untuk pengklasifikasian buku. Untuk penelitian selanjutnya, diperlukan penataan buku koleksi perpustakaan sesuai kategori dan penelitian ini dapat dikembangkan dengan algoritma klasifikasi lainnya untuk melakukan perbandingan hasil *performance measure*.

Kata kunci— klasifikasi, pencarian, buku, *Naive Bayes*

I. PENDAHULUAN

Undang-undang Republik Indonesia Nomor 43 Tahun 2007 tentang Perpustakaan, menyebutkan bahwa perpustakaan adalah institusi pengelola koleksi karya tulis, karya cetak, dan/atau karya rekam secara profesional dengan sistem yang baku guna memenuhi kebutuhan pendidikan, penelitian, pelestarian, informasi, dan rekreasi para pemustaka. Guna mendukung penyelenggaraan pendidikan nasional yang mencukupi, yaitu dengan mendukung pelaksanaan pendidikan, penelitian, dan pengabdian kepada masyarakat, setiap universitas harus memiliki pengembangan layanan perpustakaan berbasis teknologi informasi dan komunikasi [1].

Berdasarkan wawancara dengan staf bidang Teknologi Informasi Unit Pelaksana Teknis Perpustakaan Universitas Negeri Semarang pada bulan April 2016, menyatakan bahwa dari 20 perpustakaan hanya 4 perpustakaan yang sudah menggunakan layanan perpustakaan berbasis teknologi informasi dan komunikasi. Hal ini dikarenakan masih terbatasnya sumber daya manusia dalam mengelola perpustakaan dengan sistem komputerisasi.

Perpustakaan Jurusan Teknik Elektro merupakan salah satu perpustakaan yang belum memiliki sistem klasifikasi dan pencarian buku. Menurut studi pendahuluan pada bulan Januari 2016 ditemukan permasalahan, yaitu pihak pengelola perpustakaan tidak mengetahui semua judul buku yang tersedia di perpustakaan. Hal ini menyebabkan pengelola perpustakaan mengalami kesulitan saat pengunjung

menanyakan keberadaan judul buku referensi yang sedang dicari.

Menurut wawancara dengan 10 mahasiswa pengunjung perpustakaan Jurusan Teknik Elektro, mereka mengalami kesulitan dalam mencari buku referensi. Tata letak buku yang tidak beraturan dikarenakan koleksi buku belum mempunyai kategori serta belum tertata berdasarkan kategori yang sesuai sehingga menyulitkan mahasiswa pengunjung perpustakaan dalam mencari buku yang dibutuhkan dan membutuhkan waktu yang cukup lama untuk mencari buku tersebut. Ketiadaan sistem klasifikasi dan pencarian buku mengharuskan para mahasiswa pengunjung perpustakaan untuk melakukan pencarian buku secara manual. Pengelola perpustakaan juga belum dapat menata buku berdasarkan kategori buku yang sesuai karena ketidaktahuan kategori yang dimiliki pada tiap buku yang tersedia.

Penelitian mengenai perancangan maupun pengembangan sistem klasifikasi dan pencarian dokumen sudah banyak dilakukan sebelumnya. Salah satunya yaitu penelitian [2] yang menyimpulkan bahwa dalam perbandingan algoritma SVM, *Naive Bayes*, dan *Decision Tree* dalam mengklasifikasikan serangan (*Attacks*) pada sistem pendeteksi intrusi, secara keseluruhan algoritma *Decision Tree* merupakan teknik yang paling sederhana dalam mengelompokkan kasus IDS dan memiliki kecenderungan tingkat akurasi yang tinggi. Akan tetapi, algoritma *Naive Bayes* merupakan algoritma yang paling baik dalam hal waktu

komputasi (waktu yang dibutuhkan untuk membangun sebuah model). Waktu komputasi algoritma *Naive Bayes* tidak perlu diragukan karena dibandingkan dengan algoritma lain pun *Naive Bayes* merupakan yang tercepat. Kecepatan komputasi adalah salah satu kelebihan yang baik untuk diterapkan dalam sebuah sistem klasifikasi buku. Sedangkan pada penelitian ini akan digunakannya algoritma *Naive Bayes* pada pengkategorian/ pengklasifikasian buku pada perpustakaan Jurusan Teknik Elektro.

Penelitian selanjutnya [3] menyatakan bahwa dari perbandingan kinerja algoritma *Naive Bayes* terhadap *K-Nearest Neighbor* dan gabungan *K-Means* dan *LVQ* dalam pengkategorian buku komputer berbahasa Indonesia berdasarkan judul dan sinopsis mendapatkan hasil akurasi tertinggi sebesar 98% dan stabil. Perancangan sistem dibuat menggunakan bahasa pemrograman Java dan *database* yang digunakan adalah *JavaDB* (*Derby*). Masalah yang ditangani yaitu klasifikasi buku komputer berbahasa Indonesia. Sedangkan pada penelitian ini akan dirancang dengan menggunakan bahasa pemrograman PHP dan *database MySQL*. Adapun penelitian ini tidak hanya mencakup tentang buku komputer namun beberapa kategori lain berdasarkan koleksi buku perpustakaan Jurusan Teknik Elektro, baik yang berbahasa Indonesia maupun berbahasa Inggris. Penelitian ini akan memiliki 2 hak akses, yaitu admin dalam pengklasifikasian buku dan pengunjung dalam pencarian buku.

Penelitian lain yang serupa yaitu penelitian [4] yang menjelaskan bahwa dalam pengklasifikasian dokumen terdapat beberapa metode atau algoritma yang dapat digunakan. Metode atau algoritma tersebut antara lain adalah *Naive Bayes*, *Support Vector Machine*, *Neural Network*, dan *Decision Tree*. Dalam penelitiannya mendapatkan hasil akurasi tertinggi diperoleh dari algoritma *Naive Bayes* dengan nilai sebanyak 97% dan dalam waktu komputasi sebanyak 0,19 detik yang menjadikan algoritma *Naive Bayes* menjadi algoritma terbaik dalam pengklasifikasian dokumen. Sehingga pada penelitian ini akan menggunakan algoritma *Naive Bayes* untuk pengklasifikasian buku.

Berdasarkan uraian di atas, maka didapatkan jawaban atas permasalahan yang ada yaitu merancang dan membuat suatu sistem komputerisasi yang dapat mengklasifikasikan buku sesuai kategorinya serta membuat suatu sistem pencarian buku dari hasil klasifikasi buku tersebut dengan menggunakan algoritma *Naive Bayes*. Algoritma *Naive Bayes* merupakan algoritma yang mempunyai banyak kelebihan, dilihat dari tingkat akurasi yang tinggi dan tingkat waktu komputasi yang cepat sehingga algoritma *Naive Bayes* adalah algoritma yang tepat untuk digunakan dalam sistem klasifikasi buku. Kemudian hasil dari pengklasifikasian tersebut dapat digunakan pada pencarian buku bahwa buku yang tersedia sudah berdasarkan kategori yang tepat. Sistem ini merupakan salah satu bentuk layanan perpustakaan berbasis teknologi informasi dan komunikasi yang dibutuhkan di lingkungan

Pendidikan. Sistem klasifikasi dan pencarian buku ini bersifat *online*.

II. METODE PENELITIAN

A. Metode Algoritma *Naive Bayes*

Dalam *information retrieval* ada beberapa tahap yang dilakukan untuk menyelesaikan permasalahan pada sistem klasifikasi dan pencarian dokumen yaitu pengumpulan dokumen (*documents collection*), *preprocessing*, *indexing*, perhitungan algoritma *Naive Bayes*, dan *performance measure* (*recall*, *precision*, *f-measure*, dan akurasi).

1) *Documents Collection*

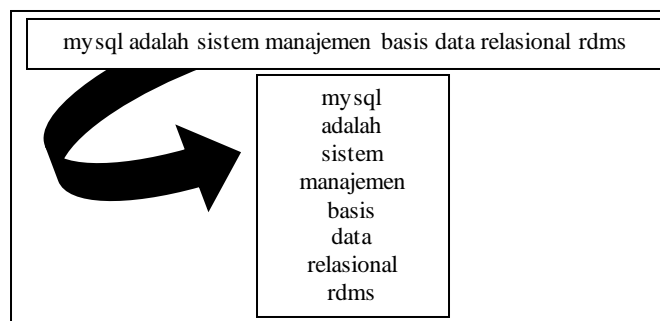
Pada tahap ini digunakan dalam pengumpulan dokumen dalam berbagai format yang dapat digunakan seperti *.html*, *.pdf*, *.doc*, *web content*, dan lain-lain [5].

2) *Preprocessing*

Preprocessing adalah tahapan pertama dalam tahapan *text mining* untuk mempersiapkan teks menjadi data yang akan diolah pada tahapan selanjutnya [6]. Atribut buku yang digunakan pada *preprocessing* adalah judul buku. Tahapan dalam *preprocessing* adalah *tokenizing*, *stopword*, dan *stemming*.

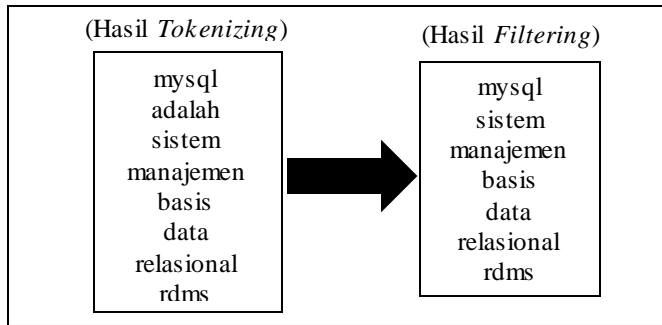
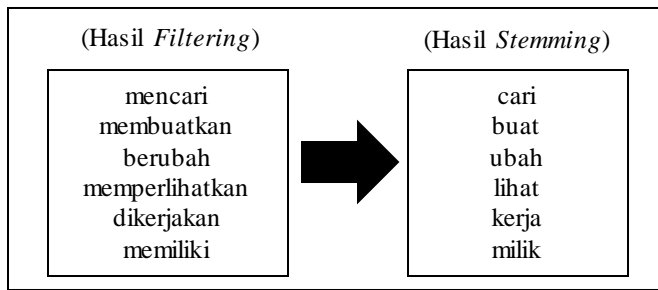
Tokenizing adalah proses memotong sebuah urutan karakter menjadi sebuah potongan-potongan, yang disebut *token* [7]. Proses ini juga melakukan pembersihan karakter-karakter tertentu seperti tanda baca dengan melakukan fungsi penghilangan tanda baca serta merubah huruf menjadi huruf kecil. Contoh dari tahapan ini dapat dilihat pada Gambar 1

Filtering dilakukan untuk mengeliminasi kata yang dianggap tidak memiliki makna, sebagai contoh kata atau, di, kemudian, dan lain sebagainya dengan melakukan fungsi *stopword*. Artinya jika kata yang ada pada tabel kata ditemukan satu atau lebih pada tabel *stopword* maka kata tersebut dieliminasi atau dihilangkan dengan melakukan fungsi *stopword*. *Stopword* dikeluarkan dari dokumen karena kata-kata itu tidak diukur sebagai kata kunci dalam aplikasi *text mining* [8]. Contoh dari tahapan ini dapat dilihat pada Gambar 2.



Gambar 1. Proses *tokenizing* [9]

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 dapat dilihat pada Gambar 3.

Gambar 2. Proses *filtering* [9]Gambar 3. Proses *stemming* [9]

3) Indexing

Teks dokumen yang telah melalui proses *tokenizing*, *filtering*, dan *stemming* kemudian diindeks ke dalam *database*. *Indexing* disini berupa pembobotan pada *term* (kata) hasil akhir *preprocessing*.

4) Naive Bayes

Bayesian Classification (NBC) adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class* [10]. Metode NBC menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap pengujian/ klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan *vocabulary*, yaitu kata yang mungkin muncul dalam koleksi dokumen sampel, sedapat mungkin dapat menjadi representasi dokumen. Selanjutnya adalah penentuan probabilitas *prior* bagi tiap kategori berdasarkan sampel dokumen. Pada tahap pengujian/ klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan *term* yang muncul dalam dokumen yang diklasifikasi.

Lebih konkritnya jika diasumsikan dimiliki koleksi dokumen $D = \{d_i \mid i = 1, 2, \dots, |D|\} = \{d_1, d_2, \dots, d_{|D|}\}$ dan koleksi kategori $V = \{v_j \mid j=1, 2, \dots, |V|\} = \{v_1, v_2, \dots, v_{|V|}\}$. Klasifikasi NBC dilakukan dengan cara mencari probabilitas $P(V = v_j \mid D = d_i)$, yaitu probabilitas kategori v_j jika diketahui dokumen d_i . Dokumen d_i dipandang sebagai *tuple* dari kata-kata dalam dokumen, yaitu $\langle a_1, a_2, \dots, a_n \rangle$, yang frekuensi

kemunculannya diasumsikan sebagai variabel random dengan distribusi probabilitas Bernoulli. Selanjutnya klasifikasi dokumen adalah mencari nilai maksimum dari :

$$V_{\text{MAP}} = \underset{v_j \in V}{\text{argmax}} P(v_j \mid a_1, a_2, \dots, a_n) \quad (1)$$

Teorema Bayes tentang probabilitas bersyarat menyatakan :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (2)$$

Keterangan :

A = Data dengan *class* yang belum diketahui

B = Hipotesis data A merupakan suatu *class* spesifik

$P(B|A)$ = Probabilitas hipotesis B berdasar kondisi A (*conditional/posterior probability*)

$P(B)$ = Probabilitas hipotesis B (*prior probability*)

$P(A|B)$ = Probabilitas A berdasar kondisi pada hipotesis B

$P(A)$ = Probabilitas dari A

argmax = fungsi yang mengembalikan indeks dari nilai maksimum dari sekumpulan himpunan data.

Dengan menerapkan teorema Bayes persamaan (1) dapat ditulis :

$$V_{\text{MAP}} = \underset{v_j \in V}{\text{argmax}} \frac{P(a_1, a_2, \dots, a_n \mid v_j)P(v_j)}{P(a_1, a_2, \dots, a_n)} \quad (3)$$

Karena nilai $P(a_1, a_2, \dots, a_n)$ untuk semua v_j besarnya sama maka nilainya dapat diabaikan, sehingga persamaan (3) menjadi:

$$V_{\text{MAP}} = \underset{v_j \in V}{\text{argmax}} P(a_1, a_2, \dots, a_n \mid v_j)P(v_j) \quad (4)$$

Dengan mengasumsikan bahwa setiap kata dalam $\langle a_1, a_2, \dots, a_n \rangle$ adalah *independent*, maka $P(a_1, a_2, \dots, a_n \mid v_j)$ dalam persamaan (4) dapat ditulis sebagai:

$$P(a_1, a_2, \dots, a_n \mid v_j) = \prod_i P(a_i \mid v_j) \quad (5)$$

Sehingga persamaan (4) dapat ditulis:

$$V_{\text{MAP}} = \underset{v_j \in V}{\text{argmax}} P(v_j) \prod_i P(a_i \mid v_j) \quad (6)$$

Nilai $P(v_j)$ ditentukan pada saat pelatihan, yang nilainya didekati dengan:

$$P(v_j) = \frac{|doc_j|}{|Contoh|} \quad (7)$$

Dimana $|doc_j|$ adalah banyaknya dokumen yang memiliki kategori j dalam pelatihan, sedangkan $|Contoh|$ banyaknya dokumen dalam contoh yang digunakan untuk pelatihan.

Untuk nilai $P(w_k \mid v_j)$, yaitu probabilitas kata w_k dalam kategori j ditentukan dengan:

$$P(w_k \mid v_j) = \frac{n_k + 1}{n + |vocabulary|} \quad (8)$$

Dimana n_k adalah frekuensi munculnya kata w_k dalam dokumen yang berkategori v_j , sedangkan nilai n adalah banyaknya seluruh kata dalam dokumen berkategori v_j , dan $|vocabulary|$ adalah banyaknya kata dalam contoh pelatihan [11].

B. Analisis Kebutuhan *Software*

Dalam pembuatan sistem klasifikasi dan pencarian buku ini, *software* yang dibutuhkan antara lain :

1) XAMPP

XAMPP adalah sebuah *software* yang berfungsi untuk menjalankan *website* berbasis PHP dan menggunakan pengolah data MySQL di komputer lokal. XAMPP berperan sebagai *server web* pada komputer. XAMPP juga dapat disebut sebuah *CPanel server virtual*, yang dapat membantu melakukan *preview* sehingga dapat memodifikasi *website* tanpa harus *online* atau terakses dengan internet. XAMPP merupakan paket *software* yang didalamnya sudah terkandung *Web Server Apache*, *database MySQL* dan *PHP Interpreter* [12].

2) MySQL

MySQL adalah salah satu jenis *database server* yang sangat terkenal. Kepopuleran disebabkan karena MySQL menggunakan SQL sebagai bahasa dasar untuk mengakses basis datanya. MySQL merupakan *server basis data* yang menggunakan teknik relasional untuk menghubungkan antara tabel-tabel dalam basis datanya atau mendukung RDBMS (*Relational Database Management System*).

3) PHP (*Hypertext Preprocessor*)

PHP adalah akronim dari *Hypertext Preprocessor*, yaitu suatu bahasa pemrograman berbasiskan kode-kode (*script*) yang digunakan untuk mengolah suatu data dan mengirimkannya kembali ke *web browser* menjadi kode HTML [13].

4) *Sublime Text*

Sublime text adalah teks editor berbasis *Python*, sebuah teks editor yang elegan, kaya fitur, *cross platform*, mudah, dan simpel yang cukup terkenal dikalangan *developer* (pengembang), penulis, dan desainer. Para *programmer* biasanya menggunakan *sublime text* untuk menyunting *source code* yang sedang dikerjakan. Sampai saat ini *sublime text* sudah mencapai versi 3. *Sublime text* mempunyai beberapa keunggulan-keunggulan yang dapat membantu dalam membuat sebuah *web development* antara lain *multiple selection*, *command pallete*, *distraction free mode*, *find in project*, *plugin API switch*, *drag and drop*, *split editing*, dan *multi platform* [14].

III. HASIL DAN PEMBAHASAN

A. Sistem Klasifikasi dan Pencarian Buku

Sistem ini terdiri atas dua *level* pengguna, yaitu *level user* dan *level administrator*. Halaman *user* terdiri dari beberapa menu diantaranya menu *Home*, *Cari*, dan *Kategori*. Hak akses *user* meliputi mencari buku, melihat koleksi buku per kategori, dan melihat detail buku. Tampilan halaman pencarian buku oleh *user* dapat dilihat pada Gambar 4. Halaman *administrator* terdiri dari beberapa menu diantaranya adalah menu *Admin*, *Klasifikasi Buku*, *Klasifikasi*, *Buku Data Latih*, *Stopword*, *Kata Dasar*, *Ganti Password*, dan *Logout*. Hak

akses *administrator* adalah meliputi segala aspek yang ada dalam sistem, seperti mengelola data admin, dan mengelola data buku. Tampilan halaman utama *administrator* dapat dilihat pada Gambar 5.

Pengklasifikasian buku dengan algoritma *Naive Bayes* diproses pada menu *Klasifikasi Buku* tepatnya pada tombol *Save* yang digunakan pada saat *peng-input-an* buku seperti pada Gambar 6.



Gambar 4. Tampilan Halaman Pencarian Buku oleh *user*



Gambar 5. Tampilan Halaman Utama Admin

Gambar 6. Tampilan *Input* Pengklasifikasian Buku

B. Hasil Pengujian *Performance Measure*

Klasifikasi buku pada perpustakaan Jurusan Teknik Elektro ini melalui 2 tahapan utama, yaitu tahap pelatihan (*training*) dan tahap pengujian (*testing*) yang masing-masing data dibagi dengan perbandingan 70% data pelatihan dan 30% data pengujian seperti pada Tabel I. Kemudian setelah proses klasifikasi selesai, hasil dari klasifikasi oleh sistem dibandingkan dengan hasil klasifikasi manual untuk mengetahui data yang relevan sesuai kategori atau tidak. Langkah selanjutnya yaitu perhitungan nilai *performance measure* dari hasil klasifikasi tersebut.

Tabel II merupakan hasil dari perbandingan klasifikasi manual dengan klasifikasi sistem untuk mengetahui banyaknya data relevan dan tidak relevan yang akan digunakan dalam perhitungan *performance measure*. Tabel III menunjukkan hasil dari perhitungan *performance measure* dengan masing-masing nilai *precision*, *recall*, *f-measure*, dan akurasi.

TABEL I. JUMLAH DATA PELATIHAN DAN DATA PENGUJIAN

No.	Kategori	Data Pelatihan	Data Pengujian
1	Jaringan Komputer	18	8
2	Pemrograman Komputer	13	6
3	Pendidikan	40	16
4	Matematika	27	12
5	Fisika	27	12
6	Teknik Elektro	49	21
Jumlah		174	75
Total		249	

TABEL II. HASIL PENGUJIAN DOKUMEN RELEVAN KLASIFIKASI BUKU

Hasil Uji	Relevan	Tidak Relevan
Manual	75	0
Sistem	70	5

TABEL III. HASIL PERFORMANCE MEASURE KLASIFIKASI BUKU (%)

Evaluasi	JK	PK	PEND	MTK	FIS	TE	Rata-rata
<i>Precision</i>	83,3	100	100	100	100	84	94,56
<i>Recall</i>	62,5	66,7	100	100	100	100	88,20
<i>F-measure</i>	71,4	80	100	100	100	91,3	90,46
Akurasi	94,7	97,3	100	100	100	94,7	97,78

C. Pengujian *Black Box*

Pengujian *black box* berupa uji fungsi. Uji fungsi digunakan untuk mengecek berjalan atau tidaknya suatu fungsi pada sistem. Pengujian dilakukan dengan membuat skenario yang telah disesuaikan dengan komponen sistem yang telah dibuat, kemudian sistem diuji berdasarkan skenario yang telah dibuat. Pada pengujian *blackbox*, semua fungsi dapat menunjukkan hasil yang diterima/ sukses, sehingga dapat dikatakan bahwa semua fungsi baik tombol maupun menu dalam sistem klasifikasi dan pencarian buku perpustakaan Jurusan Teknik Elektro berfungsi dengan baik.

D. Pembahasan

Sistem klasifikasi dan pencarian buku merupakan sebuah sistem yang dirancang untuk penentuan kategori dan pencarian buku pada perpustakaan Jurusan Teknik Elektro yang dikelola oleh Jurusan Teknik Elektro untuk meringankan beban kerja pengelola perpustakaan dalam mengelola buku yang mana di dalam sistem tersebut berfungsi untuk mengklasifikasikan buku berdasarkan pengkategorian buku yang sesuai. Bagi pengunjung, sistem ini dapat memudahkan mereka dalam mencari buku referensi yang dibutuhkan.

Berdasarkan informasi yang disajikan pada Tabel I dapat dilihat bahwa hasil klasifikasi buku secara manual yaitu berupa judul buku Perpustakaan Jurusan Teknik Elektro dibagi menjadi 6 kategori yakni kategori Jaringan Komputer (JK), Pemrograman Komputer (PK), Pendidikan (PEND), Matematika (MTK), Fisika (FIS), dan Teknik Elektro (TE). Berdasarkan hasil klasifikasi manual tersebut terdiri dari 75 dokumen uji dengan pembagian masing-masing dokumen uji per kategori yaitu 8 dokumen uji Jaringan Komputer, 6 dokumen uji Pemrograman Komputer, 16 dokumen uji Pendidikan, 12 dokumen uji Matematika, 12 dokumen uji Fisika, dan 21 dokumen uji Teknik Elektro.

Tabel II merupakan hasil klasifikasi sistem menggunakan algoritma *Naive Bayes* dimana pada tahap pengujian melakukan kesalahan sebanyak 5 dokumen uji dari total keseluruhan sebanyak 75 dokumen uji. Kesalahan yang diperoleh yaitu 3 kesalahan pada dokumen uji kategori Jaringan Komputer, dan 2 kesalahan pada dokumen uji kategori Pemrograman Komputer. Sedangkan untuk hasil *performance measure* menunjukkan bahwa pengujian evaluasi sistem dalam pengklasifikasian buku Perpustakaan Jurusan Teknik Elektro dinyatakan sangat baik, valid, dan akurat ditunjukkan dengan hasil pengukuran mendapatkan nilai rata-rata *recall* 88,20%, *precision* 94,56%, *f-measure* 90,46%, dan akurasi 97,78% seperti data yang ditampilkan pada Tabel III. Berdasarkan standar tingkat akurasi, maka hasil akurasi tersebut menunjukkan nilai *excellent classification*. Hasil pengujian *black-box* menunjukkan bahwa keseluruhan fungsi yang ada dalam Sistem Klasifikasi dan Pencarian Buku telah berjalan dengan baik sehingga tidak lagi diperlukan perbaikan fungsi.

IV. PENUTUP

A. Simpulan

1. Sistem klasifikasi dan pencarian buku ini mempunyai 2 hak akses yaitu admin dan *user*. Sistem ini menggunakan algoritma *Naive Bayes* dalam pengklasifikasian judul buku.
2. Klasifikasi dan pencarian buku perpustakaan Jurusan Teknik Elektro dilakukan dengan menggunakan 249 dokumen. Hasil pengujian algoritma *Naive Bayes* diperoleh 70 dokumen relevan dan 5 dokumen tidak relevan dengan nilai masing-masing *recall* sebesar 88,20%, *precision* sebesar 94,56%, *f-measure* sebesar 90,46%, dan akurasi sebesar 97,78% sehingga dapat disimpulkan bahwa algoritma *Naive Bayes* dapat digunakan dalam otomatisasi klasifikasi judul buku Perpustakaan Jurusan Teknik Elektro.

3. Pengujian *black box* pada sistem menghasilkan uji fungsi menu maupun tombol dapat digunakan dengan baik dan tidak didapatkan *error* pada saat pengujian.

B. Saran

1. Diperlukan penataan buku sesuai kategori yang dapat menunjang pengaplikasian sistem pada perpustakaan Jurusan Teknik Elektro.
2. Penelitian ini dapat dikembangkan dengan metode/ algoritma klasifikasi lainnya untuk melakukan perbandingan hasil *performance measure*.
3. Pengklasifikasian yang digunakan pada data latih dapat ditambah dengan menggunakan identitas buku yang lain, seperti penulis, penerbit, maupun daftar isi.
4. Penelitian ini belum terlalu detail dalam penggunaan kode DDC, sehingga dapat dikembangkan pada penelitian selanjutnya yaitu dengan penggunaan kode DDC yang lebih detail dalam proses klasifikasi buku.

REFERENSI

- [1] Undang-undang Republik Indonesia Nomor 43 Tahun 2007. Perpustakaan. Lembaran Negara Republik Indonesia Tahun 2007 Nomor 129. Jakarta.
- [2] Widiastuti, D. 2007. Analisa Perbandingan Algoritma SVM, *Naive Bayes*, dan *Decision Tree* dalam Mengklasifikasikan Serangan (*Attack*) pada Sistem Pendeteksi Intrusi. Jurusan Sistem Informasi Universitas Gunadarma, pp.1-8.
- [3] Santoso, D., D. E. Ratnawati, dan Indriati. 2014. Perbandingan Kinerja Metode *Naive Bayes*, *K-Nearest Neighbor*, dan Metode Gabungan *K-Means* dan LVQ dalam Pengkategorian Buku Komputer Berbahasa Indonesia berdasarkan Judul dan Sinopsis. Repositori Jurnal Mahasiswa PT IIK UB 4(9).
- [4] Ting, S. L., W. H. Ip, dan A. H. C. Tsang. 2011. *Is Naive Bayes a Good Classifier for Document Classification?*. International Journal of Software Engineering and Its Applications 5(3): 37-46.
- [5] Bhumika, S. S. Sehra, dan A. Nayyar. 2013. *A Review Paper on Algorithms Used for Text Classification*. International Journal of Application or Innovation in Engineering and Management (IJAIEM) 2(3): 90-99.
- [6] Krishnamoorthy, M. dan M. Mani. 2014. *A Brief Survey on Text Mining and its Applications*. Int.J. Computer Technology & Applications 5(5): 1637-1640.
- [7] Manning, C.D., P. Raghavan, dan H. Schütze. 2008. *Introduction to Information Retrieval*. New York: Cambridge University Press.
- [8] Vijayarani, S., J. Ilamathi, dan Nithya. 2011. *Preprocessing Techniques for Text Mining - An Overview*. International Journal of Computer Science & Communication Networks 5(1): 7-16.
- [9] Setiawan, A., I. F. Astuti, dan A. H. Kridalaksana. 2015. Klasifikasi dan Pencarian Buku Referensi Akademik menggunakan Metode *Naive Bayes Classifier* (NBC) (Studi Kasus: Perpustakaan Daerah Provinsi Kalimantan Timur). Jurnal Informatika Mulawarman 10(1): 1-10.
- [10] Kusriani dan E. T. Luthfi. 2009. *Algoritma Data Mining*. Edisi 1. Yogyakarta: C.V Andi Offset.
- [11] Hamzah, A. 2012. Klasifikasi Teks dengan *Naive Bayes Classifier* (NBC) untuk Pengelompokan Teks Berita dan Abstrak Akademis. Prosiding Seminar Nasional Aplikasi Sains dan Teknologi (SNAST) Periode III. Institut Sains dan Teknologi AKPRIND. Yogyakarta. 269-277.
- [12] Wardana. 2010. Menjadi Master PHP dengan *Framework Codeigniter*. Jakarta: PT Elex Media Komputindo.
- [13] Anhar. 2010. Panduan Menguasai PHP dan MySQL secara Otodidak. Jakarta Selatan: Mediakita.
- [14] Faridl, Miftah. 2015. *Fitur Dahsyat Sublime Text 3*. Edisi Pertama. Surabaya: LUG.