

SISTEM PENGELOMPOKAN KATEGORI BUKU BERBASIS METODE BAYES PADA SISTEM INFORMASI PERPUSTAKAAN JURUSAN TEKNIK ELEKTRO UM

Annisa Putri Ayudhitama

Teknik Informatika, Universitas Negeri Malang
ichadhitama@gmail.com

Abstrak

Pengelompokan kategori buku dalam perpustakaan merupakan salah satu hal penting karena jumlah buku dalam perpustakaan sangat banyak dan hal itu dilakukan untuk memudahkan para pembaca dalam mencari jenis buku yang ingin dibaca ataupun dipinjam. Pengelompokan kategori buku oleh admin pastinya membutuhkan waktu lebih lama jika dilakukan secara manual. Pengelompokan kategori menggunakan metode bayes ditinjau dari prosesnya yang mengambil aksi berdasarkan data yang telah ada sebelumnya, dengan metode bayes proses pengelompokan dapat disesuaikan sesuai dengan sifat dan kebutuhan masing-masing orang. Untuk itu pada penelitian ini, diusulkan pengembangan sistem pengelompokan kategori buku digunakan untuk membuat pengelompokan kategori buku secara otomatis dari judul buku, dan pengelompokan ini menggunakan metode bayes dengan melihat nilai probabilitasnya. Tahapan metode bayes dalam pengelompokan kategori buku ini yaitu: (1) Memasukkan judul buku, (2) *preprocessing*, (3) *tokenizing*, (4) *stopwords*, (5) perhitungan probabilitas, (6) pengelompokan kategori buku. Pengujian metode ini menggunakan 50 data latih judul buku dengan 10 kategori yang ada dan 50 data uji, model pengujian menggunakan akurasi dengan hasil percobaan mencapai 90%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa metode bayes bisa memberikan keputusan yang tepat dalam memecahkan permasalahan dari banyaknya solusi yang ada

Kata kunci : *Bayes*, pengelompokan.

1. Pendahuluan

Sistem informasi digunakan dalam berbagai bidang baik perpustakaan, pendidikan, perekonomian dan bidang-bidang lainnya, untuk mempermudah dalam mencari atau membuat suatu informasi yang kita butuhkan. Salah satu penerapan sistem informasi pada paper ini yaitu dalam bidang perpustakaan. Sistem informasi perpustakaan merupakan salah satu jenis sistem informasi yang digunakan untuk pendataan buku, pencarian buku, proses peminjaman, dan semua transaksi yang dibutuhkan pada perpustakaan. Pada perpustakaan jurusan teknik elektro UM sistem informasi yang dibutuhkan bukan hanya sistem informasi seperti pada umumnya, diperlukan kecerdasan buatan agar sistem informasi tersebut bekerja lebih baik dan lebih menarik.

Sistem informasi perpustakaan jurusan teknik elektro UM memiliki banyak kategori buku, admin akan memasukkan setiap buku baru pada sistem informasi tersebut sebagai pendataan buku. Tetapi pada kenyataannya admin masih mengelompokkan kategori secara manual jika terdapat buku baru yang masuk pada perpustakaan. Hal itu cukup menyulitkan admin jika admin tidak begitu paham akan kategori yang cocok untuk buku baru tersebut. Pengelompokan kategori buku tersebut bisa

dilakukan dengan kecerdasan buatan, salah satunya menggunakan metode bayes.

Beberapa penelitian telah difokuskan dalam sistem informasi perpustakaan. Christiane (2015) dalam papernya yang berjudul "*Ranking Search Results in Library Information Systems- Considering Ranking Approaches Adapted From Web Search Engines*" mengembangkan sistem informasi perpustakaan dengan menentukan peringkat hasil pencarian, prosesnya dengan dikategorikan kedalam 6 grup yang berbeda yaitu statistik teks, popularitas, kesegaran, lokalitas dan ketersediaan, properti konten, dan latar belakang pengguna. Dengan adanya fasilitas peringkat hasil pencarian membuat sistem informasi perpustakaan menjadi lebih canggih.

Mario (2015) dalam papernya yang berjudul "*Visual articulation of navigation and search systems for digital libraries*" mengembangkan sistem informasi perpustakaan khusus jurnal dan lainnya dengan mengusulkan artikulasi navigasi dan sistem pencarian. Prosesnya pada navigasi disaring melalui kriterianya yaitu penulis, tahun publikasi, jurnal atau penerbit dan subjek. Pada pencarian disaring berdasarkan bidang yaitu judul, deksripsi, kata kunci. Dengan adanya artikulasi navigasi dan sistem pencarian pada sistem informasi perpustakaan akan memudahkan pengguna dan

membuat sistem informasi lebih diminati.

Irene (2013) dalam papernya yang berjudul "An Automatic System For Identifying Authorities In Digital Libraries" mengembangkan sistem informasi perpustakaan dengan membuat sistem untuk mengidentifikasi karya-karya penulis secara otomatis. Prosesnya kita hanya perlu mengetikkan nama penulis pada kolom pencarian yang tersedia, maka karya-karya penulis tersebut akan dimunculkan dan kita tinggal memilih karya mana yang akan kita gunakan.

Berdasarkan beberapa penelitian di atas, sistem informasi perpustakaan dikembangkan menggunakan berbagai metode kecerdasan buatan untuk memudahkan segala aktifitas pengguna pada sistem informasi perpustakaan tersebut. Dengan adanya pengembangan tersebut maka sistem informasi perpustakaan akan lebih bermanfaat dan lebih banyak diminati.

Menurut G.P. Amaya Cruz dan G.Beliakov, pendekatan probabilistik Bayesian untuk memodelkan ketidakpastian merupakan bidang penelitian dari banyak ilmuwan. Dengan metode bayes bisa menghubungkan perbedaan solusi dan menyelesaikan bagian tersulit dari berbagai solusi yang ada.

Berdasarkan penjelasan di atas dapat disimpulkan bahwa sistem informasi perpustakaan membutuhkan metode kecerdasan buatan untuk mengembangkan sistem informasi perpustakaan agar lebih baik. Sementara itu, *Bayes* merupakan salah satu metode yang dapat membantu mengatasi suatu permasalahan dengan menggunakan fungsi probabilitas yang sangat baik. Untuk itu, pada penelitian ini diusulkan Sistem pengelompokan kategori buku berbasis metode bayes pada sistem informasi perpustakaan jurusan teknik elektro UM. Pengelompokan kategori buku pada sistem informasi perpustakaan ini diharapkan bisa membantu admin untuk mengelompokkan kategori buku secara otomatis berdasarkan judul bukunya.

2. Tinjauan Pustaka

2.1 Bayes

Suatu sistem biasanya digunakan untuk memecahkan berbagai permasalahan yang kita temukan sehari-hari seperti penentuan keputusan, jika biasanya kita dalam menentukan keputusan pasti mendapatkan berbagai solusi. Kebanyakan dari kita akan kebingungan memutuskan satu solusi tepat dari banyaknya solusi yang diberikan. Salah satu metode yang dapat diterapkan sebagai penentuan solusi tepat adalah metode *bayes*.

Teorema Bayes adalah sebuah teorema dengan dua penafsiran berbeda. Dalam penafsiran Bayes, teorema ini menyatakan seberapa jauh derajat kepercayaan subjektif harus berubah secara rasional ketika ada petunjuk baru. Dalam penafsiran frekuentis teorema ini menjelaskan representasi

invers probabilitas dua kejadian. Teorema ini merupakan dasar dari statistika Bayes dan memiliki penerapan dalam sains, rekayasa, ilmu ekonomi (terutama ilmu ekonomi mikro), teori permainan, kedokteran dan hukum. Penerapan teorema Bayes untuk memperbarui kepercayaan dinamakan inferens Bayes. (M. Yani, 2014).

Menurut Nyoman (2008) bayesian theory mempunyai berbagai keuntungan jika dibandingkan dengan beberapa teori lainnya, yaitu:

1. **Interpolation.** Bayesian method menghubungkan segala hal dengan teori-teori engineering. Bayesian method menghubungkan perbedaan yang besar karena bayesian prior dapat menjadi sebuah delta function dari suatu model yang luas.
2. **Language.** Bayesian method mempunyai bahasa tersendiri untuk menetapkan hal-hal yang prior dan posterior. Hal ini secara signifikan membantu pada saat menyelesaikan bagian yang sulit dari sebuah solusi.
3. **Intuitions.** Bayesian method melibatkan prior dan integration, dua aktivitas yang berguna secara luas

Bayesian *probability* adalah teori terbaik dalam menghadapi masalah estimasi dan penarikan kesimpulan. Bayesian *method* dapat digunakan untuk penarikan kesimpulan pada kasus-kasus dengan *multiple source of measurement* yang tidak dapat ditangani oleh metode lain seperti model hierarki yang kompleks. Kekurangan bayesian *theory*, yaitu:

1. **Information Theoretically Infeasible.** Pada kenyataannya menentukan prior pada Bayesian method merupakan hal yang cukup sulit. Kita harus menentukan angka yang riil untuk semua parameter pada model keseluruhan.
2. **Computationally Infeasible.** Walaupun dapat ditentukan prior secara akurat, namun proses perhitungan posterior kemungkinan sangatlah sulit. Kesulitan ini membutuhkan perkiraan komputasional.
3. **Unautomatic.** Selama terdapat masalah-masalah baru, selalu terdapat kebutuhan akan adanya ahli-ahli Bayesian untuk menyelesaikannya.
4. Dibutuhkan banyak hitungan komputasional yang sulit untuk menjalankan metode ini.

3. Metodologi

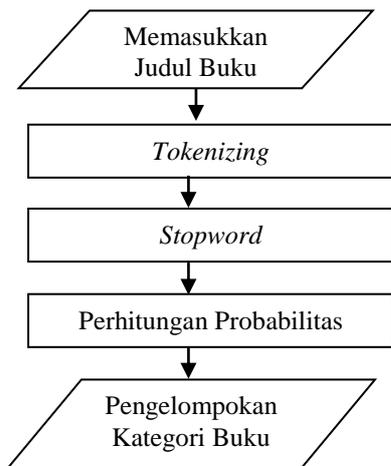
3.1 Data Penelitian

Data uji coba yang digunakan adalah 100 data buku pada perpustakaan jurusan teknik elektro UM dengan 10 kategori yang berbeda. Data buku tersebut menggunakan variabel judul dan dari judul tersebut masing-masing diekstrak menjadi keyword.

Pada penelitian ini, diusulkan sistem pengelompokan kategori buku berbasis metode bayes

pada sistem informasi perpustakaan jurusan teknik elektro UM. Pengelompokan kategori buku dalam penelitian ini adalah mengelompokkan kategori setiap buku berdasarkan judul bukunya. Dari judul setiap kategori tersebut ditentukan keywordnya atau kata kunci. Secara umum, tahapan algoritma bayes yang diterapkan pada penelitian ini, ditunjukkan pada gambar 1.

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



Gambar 1. Algoritma Bayes Pengelompokan Kategori Buku

3.2 Tokenizing

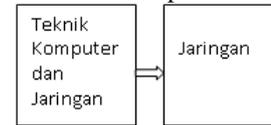
Tokenizing adalah proses memotong sebuah urutan karakter menjadi sebuah potongan-potongan, yang disebut *token*. Proses ini juga melakukan pembersihan karakter-karakter tertentu tanda baca dengan melakukan fungsi penghilangan tanda baca serta merubah huruf menjadi huruf kecil.

3.3 Stopword

Filtering dilakukan untuk mengeliminasi kata yang dianggap bukan kata kunci dari kategori tersebut dengan melakukan fungsi *stopword*. Artinya jika mengandung kata yang ada pada *stopword* maka kata tersebut dihilangkan dengan melakukan fungsi *stopword*. Jika kata yang tidak terdapat pada *stopword* maka dianggap kata kunci pada kategori tersebut. Contohnya :



Gambar 2. Tahap *Tokenizing*



Gambar 3. Tahap *Filtering*

3.4 Perhitungan Probabilitas

Teorema *Bayes* tentang probabilitas bersyarat menyatakan :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (1)$$

Dimana :

$P(B|A)$: probabilitas hipotesis B terjadi jika evidence A terjadi, pada penelitian ini hipotesis B merupakan keyword seluruh kategori. Evidence A pada penelitian ini merupakan keyword dari judul yang dicari.

$P(A|B)$: probabilitas munculnya evidence A, jika hipotesis B terjadi

$P(B)$: probabilitas hipotesis B tanpa memandang evidence apapun

$P(A)$: probabilitas evidence E tanpa memandang apapun

Persamaannya dalam mencari nilai maksimum adalah:

$$VMAP = \arg \max P(V_j|a_1, a_2, \dots, a_n) \quad (2)$$

Menurut persamaan (2), maka persamaan (1) dapat ditulis:

$$VMAP = \underset{v_j \in V}{\operatorname{argmax}} \frac{P(V_j|a_1, a_2, \dots, a_n|v_j)P(v_j)}{P(a_1, a_2, \dots, a_n)} \quad (3)$$

$P(a_1, a_2, \dots, a_n)$ konstan, sehingga dapat dihilangkan menjadi

$$VMAP = \underset{v_j \in V}{\operatorname{argmax}} P(V_j|a_1, a_2, \dots, a_n|v_j)P(v_j) \quad (4)$$

Karena $P(a_1, a_2, \dots, a_n | v_j)$ sulit untuk dihitung, maka akan diasumsikan bahwa setiap kata pada dokumen tidak mempunyai keterkaitan.

$$VMAP = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \prod P(a_j|v_j) \quad (5)$$

Keterangan :

$$P(v_j) = \frac{|doc_j|}{|contoh|} \quad (6)$$

Dimana $|doc_j|$ adalah banyaknya dokumen yang memiliki kategori j dalam pelatihan, sedangkan $|Contoh|$ banyaknya dokumen dalam contoh yang digunakan untuk pelatihan. Untuk nilai $P(w_k | v_j)$, yaitu probabilitas kata w_k dalam kategori j ditentukan dengan:

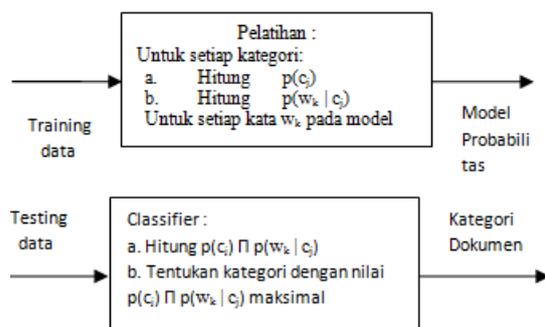
$$P(w_k|v_j) = \frac{n_{k+1}}{n+|kata|} \tag{7}$$

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 $|kata|$ adalah banyaknya kata dalam contoh pelatihan.

Tahap terakhir adalah perhitungan probabilitas setiap kategori. Langkah-langkah klasifikasi adalah :

Step 1 : Hitung $P(v_j) \prod P(w_k | v_j)$ untuk setiap kategori

Step 2 : Tentukan kategori dengan nilai $P(v_j) \prod P(w_k | v_j)$ maksimal.



3.5 Uji Coba Perhitungan Metode Bayes

Setelah mengembangkan sistem pengelompokan kategori buku pada sistem perpustakaan jurusan teknik elektro UM, dilakukan pengujian akurasi perhitungan metode bayes untuk menentukan kategori buku dari judul buku.

Tabel 1. Data Kategori

Kategori	Data Latih	Data Uji
Jaringan	5	5
Multimedia	5	5
Statistika	5	5
Pendidikan	5	5
Elektronika Dasar	5	5
Power	5	5
Mekatronika	5	5
Instrumentasi	5	5
Mikrokontroler	5	5
Sistem Informasi	5	5

Tabel 2. Data Confusion Matrix

Data Prediksi	Data Sebenarnya		
		T	F
	T	TP	FN
F	FP	TN	

Keterangan :

TP = Data uji bayes yang benar dan terindikasi benar oleh sistem

FP = Data uji bayes yang benar namun terindikasi salah oleh sistem

TN = Data uji salah yang terindikasi benar oleh

sistem

FN = Data uji salah yang terindikasi salah oleh sistem

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

Pada penelitian ini pengujian yang diambil adalah pengujian akurasi. Kesimpulan yang diperoleh yaitu dari 50 data uji diperoleh akurasi kecocokan data sebenarnya terhadap data prediksi dengan teorema bayes sebesar 90%.

4. Uji Coba Perhitungan Metode Bayes

4.1 Deskripsi Sistem Pengelompokan Kategori Buku

Pada sistem pengelompokan kategori buku yang dikembangkan, sistem tersebut hanya terdapat 1 halaman karena nantinya akan digabungkan kedalam sistem informasi perpustakaan jurusan teknik elektro UM. Pada halaman pertama terdapat menu untuk memasukkan judul buku yang akan dicari kategorinya. Pada menu ini judul diketikan pada text area yang telah disediakan, lalu untuk mencari kategorinya tinggal klik button cek kategori yang terletak di sebelah text area. Menu ditunjukkan pada Gambar 4.

Pada halaman pertama sistem pengelompokan kategori buku selain terdapat menu pengelompokan kategori judul juga terdapat menu daftar buku dari 10 kategori, masing-masing kategori terdapat 5 judul buku.

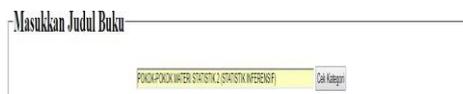
Pengujian untuk mengelompokkan kategori buku ditunjukkan pada gambar 6. Pada gambar tersebut dimasukkan salah satu judul untuk mengetahui kategorinya yaitu POKOK-POKOK MATERI STATISTIK 2 (STATISTIK INFERENSIF).



Gambar 4. Menu Pengelompokan Kategori Judul Daftar Buku

NO	document	category
1	AC THEORY	power
2	Analisis & Desain Sistem Informasi	sistem_informasi
3	Analisis & Perancangan Sistem Informasi	sistem_informasi
4	ANALISIS DATA PENELITIAN DENGAN STATISTIKA	statistika
5	Analisa Sistem Informasi	sistem_informasi
6	ANALYSIS OF FAULTED POWER SYSTEMS	power
7	Belajar dan Pembelajaran berbasis Komputer	pendidikan
8	Belajar Mikrokontroler AT89C51 52 55 Teori dan Aplikasi Edisi 2	mikrokontroler
9	BELAJAR MIKROKONTROLER AT89S51 DENGAN BAHASA C	mikrokontroler
10	Belajar Sendiri Microcontroller AVR Seri ATMega8/16/32 Simulasi, Hardware dan Aplikasi	mikrokontroler
11	BUKU PINTAR ROBOTIKA - BAGAIMANA MERANCANG & MEMBUAT ROBOT SENDIRI	mekatronika
12	Cisco Router Konfigurasi Voice, Video, dan Fax	jaringan
13	dari Begguner jadi Master Corel Draw	multimedia
14	Dasar Elektronika	elektronika_dasar
15	Dasar Teknik Instalasi Jaringan Komputer	jaringan
16	Dasar-Dasar Evaluasi Pendidikan	pendidikan
17	Dasar-Dasar Transistor	elektronika_dasar
18	DC THEORY	power
19	E-Learning Konsep dan Aplikasi	pendidikan
20	ELECTRICAL POWER AND CONTROLS	power
21	ELECTRICAL POWER AND CONTROLS	power
22	Elektromagnetika	elektronika_dasar

Gambar 5. Menu Daftar Kategori dan Judul Buku



Gambar 6. Memasukkan Salah Satu Judul Buku



Gambar 7. Muncul Kategori dari Judul Buku

Kemudian untuk menentukan kategori judul tersebut, maka ditekan button cek kategori, dan muncul kategori statistika. Maka judul tersebut menurut klasifikasi metode bayes termasuk kategori statistika seperti pada gambar 7. Hasil pengujian untuk masing-masing kategori ditunjukkan pada tabel 3.

Tabel 3. Data Uji Kategori

Uji	Data Sebenarnya	Data Prediksi
D1	Jaringan	Jaringan
D2	Multimedia	Multimedia
D3	Statistika	Statistika
D4	Pendidikan	Pendidikan
D5	Elektronika Dasar	Elektronika Dasar & Multimedia
D6	Power	Power
D7	Mekatronika	Mekatronika & Instrumentasi
D8	Instrumentasi	Instrumentasi, Power, Jaringan
D9	Mikrokontroler	Mikrokontroler
D10	Sistem Informasi	Sistem Informasi

Tabel 4. Data Confusion Matrix

Data Prediksi	Data Sebenarnya	
	T	F
T	45	5
F	5	45

5. Pembahasan

Pengujian sistem ini dilakukan pada laptop toshiba dengan sistem operasi windows 7 home premium 32 bit, processor dual core. Pengujian metode tetap berjalan lancar, berdasarkan hasil pengujian akurasi yang didapat yaitu 90%. Dari 50 data latih dan 50 data yang diuji terdapat 45 judul buku yang kategorinya benar duji dengan sistem dan cocok dengan data aslinya, namun terdapat 5 judul buku yang kategorinya salah saat diuji dengan sistem atau tidak cocok dengan data aslinya.

6. Kesimpulan dan Saran

Berdasarkan hasil tersebut, dapat disimpulkan bahwa metode bayes bisa digunakan sebagai sistem pengelompokan kategori buku, jika terdapat banyak kemungkinan kategori dari satu judul maka dengan metode bayes ini bisa membantu menentukan kategori yang tepat. Kecepatan akses sistem tergantung dari jumlah data yang kita sediakan, semakin banyak data maka proses akses akan bisa jadi akan melambat karena metode ini perlu untuk menghitung setiap datanya untuk dibandingkan.

Daftar Pustaka :

Achmad Solichin. 2017. “Mengukur Kinerja Algoritma Klasifikasi Dengan Confusion Matrix”.<http://achmatim.net/2017/03/19/mengukur-kinerja-algoritma-klasifikasi-dengan-confusion-matrix/>. Diakses pada tanggal 1 April 2018.

Aida Indriani. 2014. “Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier”. STMIK PPKIA. Indonesia.

Christiane Behnert, Dirk Lewandowski. 2015. *Rangking Search Results in Library Information Systems-Considering Ranking Approaches Adapted From Web Search Engines*. Hamburg University. Germany.

G.P. Amaya Cruz, G.Beliakow. 1996.*On the interpretation of certainty factors in expert systems*. Los Andes University. Colombia.

Irene Diaz Valenzuela, Maria J.Martin Bautista, Maria Ampalo Vila, Jesus R.Campana. 2013. *An Automatic System For Identifying Authorities In Digital Libraries*. University if Granada. Spain.

M Yani. 2014. “Makalah terorema bayes”.<http://muhammadyaniishak.blogspot.co.id/2014/08/makalah-teorema-bayes.html>. Diakses pada tanggal 31 Maret 2018.

Mario Perez Montoro, Jaume Nualart. 2015. *Visual artivulation of navigation and search systems for digital libraries*. University of Canberra. Australia.

Nyoman. 2008. “Kelebihan & Kekurangan dari metode Bayesian Theory”.<https://nyo2nyo.wordpress.com/2008/04/05/kelebihan-kekurangan-dari-metode-bayesian-theory/>. Diakses pada tanggal 31 Maret 2018.