

BAB I PENDAHULUAN

1.1. Latar Belakang

Pengkategorian buku merupakan kegiatan penting yang harus dilakukan di dalam perpustakaan. Pengkategorian buku adalah kegiatan mengelompokkan buku ke dalam golongan sesuai dengan kesamaan ciri dengan buku yang lain. Pengkategorian buku dilakukan karena memberikan banyak manfaat bagi pustakawan maupun bagi pengguna, seperti menunjukkan isi pokok yang dibahas dalam buku tersebut, memudahkan peletakan dan pencarian suatu buku, memandu pengguna menemukan buku dengan subjek yang relevan, dan lain sebagainya [HAM-95].

Dalam proses mengkategorikan buku, tak jarang pustakawan menemui beberapa kendala. Pustakawan akan menemui kendala waktu dan tenaga jika pustakawan tersebut harus mengkategorikan buku dalam jumlah besar, karena selama ini pustakawan melakukan cara manual dalam mengkategorikan buku. Selain itu, tak jarang pula pustakawan menemui kendala dalam menentukan golongan / kelas dari suatu buku, karena buku tersebut mencakup pokok bahasan dari beberapa golongan buku.

Seiring dengan berkembangnya teknologi informasi, komputer seharusnya dapat menjadi alat bantu pustakawan dalam menentukan kategori buku. Pustakawan tentunya mengharapkan sistem pengkategorian buku otomatis memiliki tingkat keakuratan yang tinggi yang dapat meminimalisir kerancuan dalam pengkategorian buku.

Untuk mengoptimalkan kinerja sistem pengkategorian buku secara otomatis, harus digunakan suatu metode pengkategorian yang baik. Metode pengkategorian buku otomatis dapat menggunakan metode yang dipakai dalam pengkategorian dokumen teks, karena buku juga merupakan dokumen teks. Namun metode yang dapat dipilih sangat banyak dan beragam. Untuk memilih metode pengkategorian yang terbaik, dapat dilakukan perbandingan antara beberapa metode pengkategorian dokumen teks. Dalam penelitian Cahyo Darujati pada tahun 2010 [DAR-10]. dan J. Sreemathy dan P. S. Balamurugan [SRE-12]

pada tahun 2012, dibandingkan metode *Naïve Bayes* dan *k-Nearest Neighbor* (KNN) dalam pengkategorian dokumen teks berbahasa Indonesia, dan disimpulkan bahwa metode *Naïve Bayes* adalah metode yang memiliki tingkat akurasi terbaik. Namun, dalam penelitian Vaibhav C. Gandhi dan Jignesh A.Prajapati pada tahun 2012, dibandingkan metode *Support Vector Machine* (SVM), *Naïve Bayes*, dan KNN dalam pengkategorian dokumen teks berbahasa Inggris, didapatkan hasil yang berbeda, yaitu KNN dan SVM lebih unggul dibandingkan metode *Naïve Bayes* [GAN-12].

Karena perbedaan hasil dari beberapa penelitian, belum dapat dihasilkan kesimpulan metode mana yang paling cocok untuk pengkategorian buku komputer berbahasa Indonesia. Untuk itu, penelitian ini melakukan perbandingan terhadap beberapa metode pengkategorian dokumen teks untuk mendapatkan metode yang terbaik dalam menyelesaikan permasalahan pengkategorian buku komputer. Metode yang dipilih dalam penelitian ini adalah KNN dan *Naïve Bayes*, karena berdasarkan beberapa penelitian, *Naïve Bayes* dan KNN merupakan metode yang cukup efektif dalam menyelesaikan masalah pengkategorian dokumen teks. Selain itu, metode baru hasil penelitian Dian Eka Ratnawati, dkk, pada tahun 2012, yaitu metode gabungan *K-Means* dan LVQ juga dipilih karena metode baru ini sebelumnya belum pernah dibandingkan dengan metode KNN dan *Naïve Bayes*, sehingga dapat diketahui keefektifan metode ini terhadap pengkategorian buku komputer berbahasa Indonesia [RAT-12].

Dalam penelitian ini, pengkategorian buku dibatasi pada pengkategorian buku komputer berbahasa Indonesia, karena jenis buku yang terlampau banyak, sehingga tidak memungkinkan jika seluruh buku diambil sebagai bahan uji. Selain itu, sumber data buku komputer yang relatif mudah didapat karena banyak toko buku digital yang mempublikasikan data buku komputernya. Buku komputer juga cukup representatif dalam menggambarkan masalah pengkategorian buku, karena cukup banyak buku komputer yang mencakup pokok bahasan dari beberapa golongan buku komputer. Selain itu, variabel yang diambil sebagai bahan acuan pengkategorian buku dibatasi pada judul dan sinopsis, karena menurut Towa P. Hamakonda, variabel yang dipakai dalam pengkategorian buku mulai dari yang terpenting adalah judul, sinopsis, daftar isi / kata pengantar / pendahuluan, dan

sebagian isi buku [HAM-95]. Judul dan sinopsis diambil karena kedua variabel tersebut adalah variabel terpenting dalam pengkategorian buku. Selain itu, data judul dan sinopsis lebih mudah didapat, karena biasanya toko buku online mempublikasikan data judul dan sinopsis.

Berdasarkan latar belakang studi kasus dan metode yang telah dipaparkan, maka judul skripsi yang diambil adalah **"Perbandingan Kinerja Metode *Naïve Bayes*, *K-Nearest Neighbor*, dan Metode Gabungan *K-Means* dan LVQ dalam Pengkategorian Buku Komputer Berbahasa Indonesia Berdasarkan Judul dan Sinopsis"**.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka dapat dirumuskan beberapa permasalahan, yaitu :

1. Bagaimana cara mengkategorikan buku komputer berbahasa Indonesia dengan menggunakan metode KNN, *Naïve Bayes*, dan gabungan *K-Means* dan LVQ.
2. Bagaimana tingkat akurasi metode KNN, *Naïve Bayes*, dan gabungan *K-Means* dan LVQ dalam pengkategorian buku komputer berbahasa Indonesia.

1.3. Batasan Masalah

Agar permasalahan yang dirumuskan lebih terfokus, maka penelitian ini dibatasi dalam hal :

1. Data yang digunakan dalam penelitian ini adalah data buku komputer berbahasa Indonesia yang diambil dari beberapa sumber, yaitu grahailmu.co.id, biobses.com, dan andipublisher.com.
2. Variabel buku yang dipakai untuk pengkategorian golongan buku adalah judul dan sinopsis.
3. Pengujian dilakukan dengan mengukur akurasi sistem.

1.4. Tujuan Penelitian

Penelitian ini dibuat dengan tujuan :

1. Menerapkan metode KNN, *Naïve Bayes*, dan gabungan *K-Means* dan LVQ dalam pengkategorian buku komputer berbahasa Indonesia.
2. Mengetahui tingkat akurasi metode KNN, *Naïve Bayes*, dan gabungan *K-Means* dan LVQ dalam pengkategorian buku komputer berbahasa Indonesia.

1.5. Manfaat Penelitian

Hasil penelitian ini diharapkan :

1. Mendapatkan metode yang paling efektif dalam pengkategorian buku komputer berbahasa Indonesia.
2. Menyediakan sistem pengkategorian buku komputer berbahasa Indonesia otomatis yang dapat membantu meringankan pekerjaan pustakawan dalam pengkategorian buku komputer berbahasa Indonesia.
3. Menjadi referensi bagi pembaca dalam merancang sistem pengkategorian buku komputer berbahasa Indonesia otomatis dengan menggunakan metode KNN, *Naïve Bayes*, dan metode gabungan *K-Means* dan LVQ.
4. Menjadi referensi bagi pembaca dalam mengetahui perbandingan kinerja metode KNN, *Naïve Bayes*, dan gabungan *K-Means* dan LVQ dalam kasus pengkategorian buku komputer berbahasa Indonesia.

1.6. Sistematika Penulisan

Sistematika penulisan penelitian ini adalah sebagai berikut :

1. BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan untuk memberikan gambaran umum mengenai penelitian.

2. BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan kajian pustaka dan dasar teori yang meliputi buku komputer; konsep umum klasifikasi buku yang terdiri dari klasifikasi buku, tujuan klasifikasi buku, manfaat klasifikasi buku, dan cara

penentuan klasifikasi Buku; *teks mining*; *text preprocessing* yang terdiri dari *case folding*, *tokenizing*, *filtering*, dan *stemming*; pembobotan Tf-Idf; *Naive Bayes*, *k-Nearest Neighbor*, *Cosine Similarity*, *Learning Vector Quantization (LVQ)*, *k-Means*, *euclidean distance*, metode gabungan *k-Means* dan LVQ; dan akurasi sistem.

3. BAB III METODOLOGI PENELITIAN DAN PERANCANGAN

Bab ini menjelaskan langkah-langkah yang dilakukan dalam pengerjaan penelitian yang terdiri dari studi pustaka, pengumpulan data, perancangan, implementasi, pengujian, dan evaluasi sistem. Selain itu, Bab ini juga membahas tentang perancangan sistem, perancangan *user interface* (tatap muka pengguna), perancangan *database*, dan contoh perhitungan manual.

4. BAB IV IMPLEMENTASI

Bab ini menjelaskan lingkungan implementasi, pembahasan dari implementasi program dan implementasi *user interface* yang telah dirancang sebelumnya.

5. BAB V PENGUJIAN DAN ANALISA

Bab ini menjelaskan skenario pengujian terhadap metode yang digunakan, hasil pengujian, dan analisa hasil uji coba yang dilakukan dalam penelitian.

6. BAB VI PENUTUP

Bab ini menyimpulkan hal-hal yang diperoleh dari keseluruhan penelitian dan saran yang dapat dikembangkan untuk kemajuan penelitian.