

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Sebagai pusat informasi, perpustakaan berkembang dan memberikan manfaat sebagai sumber ilmu pengetahuan, penelitian, serta pemeliharaan warisan budaya nasional. Dalam Undang-Undang Nomor 43 tahun 2007 tentang Perpustakaan dijelaskan 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 [1]. Ilmu dan informasi disampaikan oleh pemustaka dapat digunakan untuk mendukung studi akademis, memperluas pengetahuan dalam berbagai bidang, dan menyediakan hiburan [2].

Perpustakaan UIN Sunan Gunung Djati Bandung merupakan salah satu fasilitas yang disediakan untuk menunjang kegiatan akademik. Buku yang tersedia di Perpustakaan UIN Sunan Gunung Djati cukup banyak mencakup berbagai cabang ilmu pengetahuan. Seiring berjalannya waktu, kebutuhan terhadap informasi dalam bentuk buku atau artikel ilmiah semakin meningkat, seperti tercermin dari jumlah buku yang melimpah di perpustakaan. Saat ini, pengelolaan buku di Perpustakaan UIN Sunan Gunung Djati Bandung sudah cukup optimal. Pada saat civitas akademik ingin mencari buku berdasarkan kategori, maka pengguna harus mencari satu persatu secara manual. Terhitung pada 28 Januari 2024, jumlah buku yang ada di Perpustakaan UIN Sunan Gunung Djati Bandung yaitu 79.430 eksemplar [3]. Semakin banyak buku atau dokumen yang tersedia, semakin banyak waktu dan tenaga yang dibutuhkan. Oleh karena itu, diperlukan suatu sistem yang mampu mengklasifikasikan dokumen secara otomatis. Untuk mencapai hasil optimal dalam mengklasifikasikan dokumen, digunakan dataset yang berasal dari data buku yang ada di perpustakaan.

Berdasarkan *website* lib.uinsgd.ac.id/research yang diakses pada 06 Juni 2024, pada kategori *keyword* pencarian terpopuler dalam periode satu tahun terakhir menunjukkan bahwa pada urutan pertama yaitu ‘Ekonomi Syariah’, kedua ‘Hukum’, ketiga ‘Novel’, keempat ‘Sejarah Peradaban Islam’, dan pada urutan kelima ‘Ilmu Tauhid’. *Keyword* tersebut menunjukkan kata untuk jurusan yang ada di lingkungan UIN Sunan Gunung Djati Bandung. Berdasarkan data tersebut, pengunjung perpustakaan mencari buku-buku untuk jurusan yang sesuai. Namun, Perpustakaan UIN Sunan Gunung Djati hanya menyediakan kategori buku berdasarkan DDC (*Dewey Decimal Classification*), berdasarkan bahasa dan berdasarkan jenis buku. Maka dari itu, diperlukan kategori buku berdasarkan jurusan di lingkungan UIN Sunan Gunung Djati Bandung agar pengunjung dengan mudah untuk menemukan buku yang sesuai dengan jurusan. Untuk melakukan kategorisasi buku tersebut, dapat dilakukan dengan metode klasifikasi.

Klasifikasi teks atau kategorisasi teks merupakan suatu pendekatan teknologi yang ditujukan untuk menentukan apakah suatu teks termasuk kedalam satu atau lebih kategori berdasarkan isi dari teks tersebut yang mengacu terhadap pemodelan kategori yang diberikan [4]. Proses dari klasifikasi teks itu sendiri bervariasi, hal tersebut dikarenakan beberapa penelitian memiliki kepentingan dan kebutuhan yang berbeda-beda. Namun secara umum proses klasifikasi teks terdiri dari enam tahapan, diantaranya yaitu: pengumpulan data, *pre-processing*, *feature selection*, *feature extraction*, pembangunan model dan evaluasi performansi [5].

Mengingat buku yang ada dapat dikategorikan lebih dari satu jurusan, maka dilakukan klasifikasi multi label. Klasifikasi multi label pada dasarnya merupakan suatu bagian permasalahan dari klasifikasi teks, yang mana masing-masing dokumen dapat tergolong kedalam beberapa kelas. Klasifikasi multi label ini berbeda dengan klasifikasi *single* label. Klasifikasi *single* label bertujuan untuk mengklasifikasikan suatu dokumen hanya kedalam satu kelas saja [6].

Telah dilakukan penelitian terdahulu menggunakan pendekatan serupa. Misalnya, penelitian oleh Rizkyani yang menerapkan metode *Multi-Label K-Nearest Neighbor* (ML-KNN) untuk penyakit kanker serviks dengan akurasi sebesar 93% [7].

Disisi lain, penelitian oleh Moh. Mahdi Sya'bani juga menggunakan metode *Cosine Similarity* untuk klasifikasi pada synopsis buku menghasilkan nilai *precision* sebesar 90,91% [8].

Hasil-hasil penelitian tersebut menunjukkan bahwa algoritma *K-Nearest Neighbor* dapat efektif digunakan dalam konteks klasifikasi multi-label. Oleh karena itu, penelitian ini melibatkan penggunaan algoritma *K-Nearest Neighbor* dan *Cosine Similarity* dalam klasifikasi buku. Pada penelitian ini juga menggunakan dataset yang lebih besar agar kinerja model lebih optimal. Dalam konteks ini, penelitian diharapkan dapat memberikan wawasan tentang bagaimana *K-Nearest Neighbor* dapat digunakan dalam klasifikasi multi label pada buku secara efisien dan efektif.

1.2 Perumusan Masalah

Berdasarkan latar belakang tersebut, dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana menerapkan algoritma *K-Nearest Neighbor* dalam klasifikasi multi label untuk menentukan kategori buku?
2. Bagaimana menganalisis kinerja algoritma *K-Nearest Neighbor* dalam klasifikasi multi label untuk menentukan kategori buku?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan pendekatan algoritma *K-Nearest Neighbor* dalam klasifikasi multi label untuk menentukan kategori buku.
2. Menganalisis kinerja pendekatan algoritma *K-Nearest Neighbor* dalam klasifikasi multi label untuk menentukan kategori buku.

1.4 Manfaat Penelitian

Penulisan penelitian ini diharapkan mampu memberikan manfaat sebagai berikut:

1. Bagi Pengguna

Penelitian ini akan memberikan manfaat langsung bagi pengguna. Dengan adanya sistem klasifikasi judul buku untuk menentukan kategori buku akan

memudahkan pengguna untuk mengetahui buku sesuai preferensi jurusan. Selain itu, model klasifikasi judul buku juga membuat kumpulan buku lebih terorganisir.

2. Bagi Penulis

Penelitian ini akan memberikan kontribusi dalam pengembangan teknologi klasifikasi teks dengan menggunakan algoritma *K-Nearest Neighbor* (KNN), Penulis akan memiliki kesempatan untuk berkontribusi pada penelitian lanjutan di bidang ini dan mengembangkan pemahaman yang lebih mendalam tentang *data mining*

3. Bagi Akademik

Penelitian ini akan menjadi sumber referensi dan dasar untuk penelitian lebih lanjut dalam bidang *data mining*, khususnya dalam konteks klasifikasi yang terdapat dalam kumpulan data besar. Hal ini akan membantu memperluas pengetahuan akademik tentang data mining dan algoritma *K-Nearest Neighbor*.

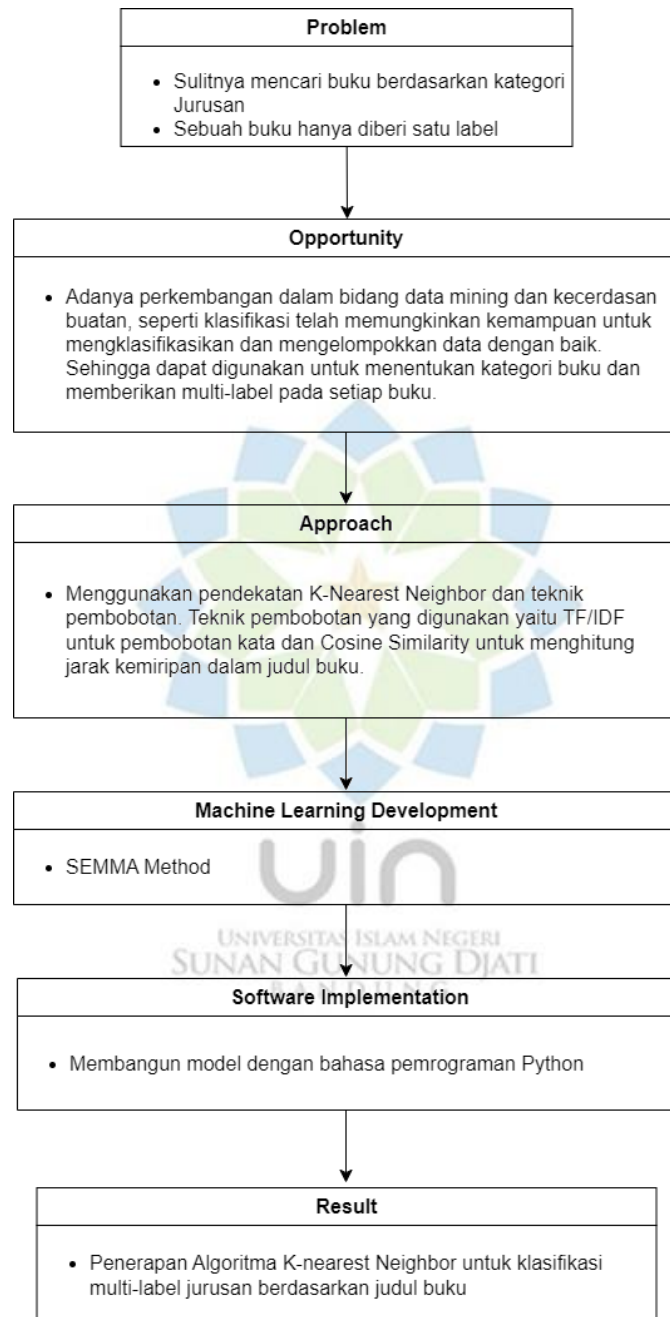
1.5 Batasan Masalah

Agar penelitian ini lebih terarah dan sesuai dengan tujuan yang diinginkan maka berikut adalah beberapa batasan masalah dari penelitian ini:

1. Algoritma *K-Nearest Neighbor* dengan menggunakan *Cosine Similarity*.
2. Data buku yang digunakan merupakan buku yang ada di Perpustakaan UIN Sunan Gunung Djati Bandung.
3. Dataset buku berjumlah 1.223 buku.
4. Labeling pada data dilakukan secara manual dengan mempertimbangkan label DDC.
5. Klasifikasi buku yang digunakan yaitu berjumlah 7 kelas. Dimana jumlah tersebut sama dengan jurusan yang ada di Fakultas Sains dan Teknologi UIN Sunan Gunung Djati Bandung.
6. Topik buku digunakan sebagai data pendukung opsional sehingga tidak dilakukan *preprocessing*.

1.6 Kerangka Pemikiran

Kerangka pemikiran dari tugas akhir ini dideskripsikan pada Gambar 1.1



Gambar 1.1 Kerangka Pemikiran

Pada Gambar 1.1 menjelaskan penelitian ini dilakukan untuk mengklasifikasikan buku dengan kategori jurusan yang ada di UIN Sunan Gunung Djati Bandung. Algoritma yang digunakan yaitu *K-Nearest Neighbor*. SEMMA sebagai metode

pengembangan yang digunakan dan *Python* sebagai bahasa pemrograman yang digunakan. Hasil akhir yaitu evaluasi model untuk klasifikasi multi label berdasarkan judul buku.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini disusun untuk memberikan gambaran umum mengenai model yang akan dibuat. Sisematika penulisan disajikan dalam beberapa bab yang masing-masing bab menguraikan beberapa pokok pembahasan. Adapun sistematika penulisannya adalah sebagai berikut.

BAB I: Pendahuluan

Bab ini berisi mengenai latar belakang masalah, perumusan masalah, tujuan penelitian, batasan masalah, kerangka pemikiran, metodologi penelitian dan sistematika penulisan.

BAB II: Kajian Literatur

Bab ini menjelaskan mengenai kajian literatur yang terdiri dari pembahasan penelitian sebelumnya dan penjelasan beberapa teori yang relevan untuk menyelesaikan masalah penelitian.

BAB III: Metodologi Penelitian

Bab ini membahas mengenai metode yang diterapkan pada penelitian ini, beserta penjelasan detail mengenai setiap langkah-langkah dan teknik yang digunakan dalam penelitian.

BAB V: Hasil dan Pembahasan

Pada bab ini menjelaskan tentang implementasi dari model dan pengujian model. Pengujian model meliputi beberapa skenario pengujian, langkah-langkah dalam uji coba model yang bertujuan untuk mengetahui apakah model tersebut telah menyelesaikan permasalahan sesuai dengan yang diharapkan atau belum.

BAB V: Kesimpulan dan Saran

Pada bab terakhir ini berisi pernyataan singkat berupa kesimpulan secara keseluruhan yang dapat menjawab rumusan masalah disertai dengan saran untuk penelitian selanjutnya yang bertujuan untuk mengembangkan penelitian lebih lanjut.