

## **BAB 1 PENDAHULUAN**

Pada bab ini dijelaskan latar belakang dari penelitian yang memberikan gambaran awal mengenai klasifikasi topik dan perannya dalam perkembangan teknologi informasi. Tujuan dan ruang lingkup pengerjaan tugas akhir memberikan penjelasan mengenai hasil yang ingin diketahui serta batasan dalam pengerjaan. Tahapan dalam metodologi penelitian dan sistematika penulisan laporan.

### **1.1 Latar Belakang**

Dengan berkembangnya teknologi, kebutuhan akan informasi semakin meningkat. Informasi yang dibutuhkan mengalami perkembangan mulai dari informasi yang bersifat umum hingga informasi yang bersifat khusus. Apabila jumlah data yang akan dicari sedikit, hal ini masih dapat dilakukan secara manual. Akan tetapi, dengan jumlah data yang banyak, proses pencarian secara manual akan menghabiskan waktu dan tenaga dalam jumlah yang banyak. Padahal waktu merupakan salah satu factor yang menentukan efektivitas dan tingkat bermanfaatnya suatu data. Hal ini dikarenakan terdapat data yang bila telah melewati suatu waktu, data tersebut sudah tidak berguna atau tidak valid. Oleh karena itulah muncul kebutuhan untuk memperoleh data secara cepat dan tepat.

Kebutuhan ini dapat terbantu dengan adanya klasifikasi topik. Klasifikasi topik akan mengelompokkan data yang dimilikinya sesuai dengan topik yang terkandung pada data tersebut. Setelah itu, apabila terdapat permintaan akan suatu dokumen dan diketahui bahwa dokumen tersebut termasuk dalam topik A, maka akan langsung dilakukan pencarian dokumen-dokumen pada topik A dan bukan topik yang lain. Dokumen dari topik lain tidak akan melalui pemrosesan karena data tersebut tidak relevan dengan yang dicari. Klasifikasi topik yang baik akan menghemat tenaga dan waktu dalam pencarian suatu dokumen.

Contoh, apabila terdapat suatu dokumen A yang ingin diterjemahkan, dokumen tersebut haruslah diketahui termasuk dalam bahasa mana. Setelah diketahui bahwa

dokumen tersebut ditulis dalam bahasa Inggris, maka cukup menggunakan kamus yang berhubungan dengan bahasa Inggris dan bukan kamus lain.

Kebutuhan informasi yang khusus mendorong penelitian mengenai klasifikasi topik dan juga memotivasi penelitian dengan domain bahasa Indonesia. Penelitian klasifikasi topik dengan bahasa Indonesia diharapkan dapat membantu menjawab kebutuhan informasi tersebut.

## **1.2 Permasalahan**

Permasalahan pada tugas akhir ini adalah belum adanya klasifikasi topik dalam bahasa Indonesia. Klasifikasi topik dalam bahasa Inggris telah dilakukan tetapi belum adanya penelitian dalam bahasa Indonesia. Selain itu, ingin dilihat juga apakah metode-metode yang telah digunakan pada (Nigam, Lavery, & McCallum, 1999) dan (Sebastiani, 2002) dapat diterapkan pada dokumen dengan bahasa Indonesia. Beberapa metode tersebut adalah Naïve Bayes dan Maximum Entropy. Permasalahan lain yang dihadapi adalah keterbatasan sumber daya berupa data dalam bahasa Indonesia. Pada tugas akhir ini, dokumen berbahasa Indonesia didapatkan dengan pencarian artikel dan pengambilan abstrak tulisan ilmiah seperti yang dijelaskan pada subbab 4.1.1.

## **1.3 Tujuan**

Tujuan utama dari tugas akhir ini adalah mengetahui kinerja dari penerapan metode *machine learning* Naïve Bayes dan Maximum Entropy dalam melakukan klasifikasi topik untuk dokumen berbahasa Indonesia. Selain itu, tugas akhir ini juga bertujuan untuk mengetahui kinerja klasifikasi topik tiap metode dengan melihat perbandingan nilai akurasi hasil klasifikasi topik dari aspek fitur, jumlah topik, dan jenis data yang digunakan.

## **1.4 Ruang Lingkup**

Ruang lingkup pengerjaan dari tugas akhir ini adalah sebagai berikut:

1. Klasifikasi topik menggunakan dua jenis data, yaitu artikel media massa dan abstrak tulisan ilmiah. Artikel media massa didapat dengan melakukan pengambilan data dari *website* kompas.com dan abstrak tulisan ilmiah didapat dari basis data sistem Lontar.

2. Fitur yang digunakan adalah fitur unigram dengan melakukan variasi pada pemilihan informasi fitur dan jumlah *token* yang digunakan.
3. Klasifikasi topik dilakukan dengan metode *machine learning* Naïve Bayes dan Maximum Entropy dengan *tools* yang sudah tersedia. Pekerjaan yang dilakukan mencakup pengambilan data, pemilihan fitur, persiapan data masukan untuk masing-masing tools, dan pemanfaatan *tools* atau *library* yang tersedia.

### 1.5 Metodologi Penelitian

Metodologi yang digunakan dalam pengerjaan tugas akhir ini adalah metode eksperimental dan dilaksanakan dalam tahapan-tahapan sebagai berikut:

1. Studi literatur – Pencarian informasi mengenai klasifikasi topik, metode-metode yang dapat digunakan, dan pembelajaran mengenai metode-metode tersebut
2. Perancangan – Melakukan perancangan percobaan dengan mempersiapkan data, penentuan variabel percobaan, dan perancangan klasifikasi topik menggunakan *machine learning*
3. Implementasi – Pada bagian implementasi dilakukan penerapan dari perancangan yang dilakukan dengan pengolahan data dan penggunaan metode *machine learning* dalam klasifikasi topik.
4. Analisis hasil – Melakukan perbandingan nilai akurasi yang didapat dengan menggunakan metode *machine learning* dilihat dari aspek metode, jenis fitur, jumlah *token*, jumlah topik, dan jenis data yang digunakan.

### 1.6 Sistematika Penulisan

Sistematika penulisan laporan mengikuti tahapan-tahapan yang dilakukan untuk menyelesaikan tugas akhir, sebagai berikut:

- BAB 1 PENDAHULUAN - Pada bab ini dijelaskan latar belakang dari penelitian yang memberikan gambaran awal mengenai klasifikasi topik dan perannya dalam perkembangan teknologi informasi. Tujuan dan ruang lingkup pengerjaan tugas akhir memberikan penjelasan mengenai hasil

yang ingin diketahui serta batasan dalam pengerjaan. Tahapan dalam metodologi penelitian dan sistematika penulisan laporan.

- **BAB 2 LANDASAN TEORI** - Pada bab ini dijelaskan landasan teori dari pekerjaan dan metode yang digunakan dalam tugas akhir untuk melakukan klasifikasi topik. Pembahasan ini dimulai dengan penjelasan klasifikasi topik, lalu dilanjutkan mengenai metode-metode pada *machine learning* yang akan digunakan pada tugas akhir ini.
- **BAB 3 PERANCANGAN** - Pada bab ini dijelaskan perancangan untuk melakukan klasifikasi topik pada artikel media massa dan abstrak tulisan ilmiah. Klasifikasi topik dilakukan dengan mengelompokkan dokumen ke dalam salah satu topik yang ada. Perancangan klasifikasi topik ini meliputi persiapan data, penentuan variabel percobaan, dan perancangan klasifikasi topik menggunakan *machine learning*.
- **BAB 4 IMPLEMENTASI** - Pada bab ini dijelaskan secara rinci penerapan dari perancangan yang telah dilakukan untuk klasifikasi topik. Implementasi yang dijelaskan berupa persiapan data, proses pemilihan fitur, dan implementasi klasifikasi topik menggunakan metode *machine learning*. Implementasi persiapan data dan pemilihan fitur dilakukan dengan menggunakan PERL dan Java. Sementara, klasifikasi topik dengan *machine learning* dilakukan dengan menggunakan *tools* dan *library* yang sudah tersedia, yang dibuat dalam bahasa Java. Hasil yang didapat setelah melakukan implementasi persiapan data dan pemilihan fitur adalah *input* data berupa data *training* dan data *testing* yang sesuai dengan masing-masing *tools*.
- **BAB 5 HASIL DAN PEMBAHASAN** - Pada bab ini diberikan hasil dari percobaan yang dilakukan dalam melakukan klasifikasi topik dengan menggunakan Naïve Bayes dan Maximum Entropy. Pembahasan dari hasil

mencakup perbandingan antara metode *machine learning* yang digunakan serta pembahasan klasifikasi topik dengan variasi penggunaan fitur, nilai fitur, dan jenis data yang digunakan

- **BAB 6 PENUTUP** - Bab ini merupakan penutup dari laporan tugas akhir yang berisi kesimpulan dan kendala dari percobaan yang dilakukan dengan metode Naïve Bayes, Naïve Bayes Multinomial, dan Maximum Entropy. Selain itu, pada subbab terakhir juga diberikan saran-saran untuk pengembangan lebih lanjut dalam penelitian klasifikasi topik.

