

BAB I

PENDAHULUAN

1.1. Latar Belakang

Mayoritas penduduk Indonesia beragama Islam. Untuk itu diperlukannya memilihkan model yang baik dalam memproses klasifikasi dokumen hadis. Teknologi membuat berkembangnya tulisan-tulisan dalam bentuk dokumen elektronik. Dokumen elektronik yang berkembang adalah dokumen dengan bahasa arab, yang umumnya dibuat oleh umat Islam salah satunya Indonesia yang mayoritas muslim. Karena perkembangan dokumen bahasa arab yang begitu besar membuat proses pengelompokan data tidak bisa dilakukan secara manual sehingga peneliti membuat proses untuk pengelompokan data lebih mudah *data mining* adalah teknologi yang memadukan metode analisis data tradisional dengan algoritma canggih untuk memproses data dalam volume besar. Ini juga membuka peluang menarik untuk mengeksplorasi dan menganalisis tipe data baru dan menganalisis tipe data lama dengan cara baru. Dalam bab pendahuluan ini, penulis menyajikan tinjauan umum tentang *data mining* dan menguraikan topik utama yang akan dibahas dalam buku ini. Kami mulai dengan deskripsi beberapa aplikasi terkenal yang membutuhkan teknik baru untuk analisis data.

Proses pengelompokan dokumen atau teks bisa dilakukan dengan dua cara yaitu dengan klasifikasi dan *clustering*. Klasifikasi dan *clustering* telah dipelajari selama bertahun-tahun oleh para peneliti pencarian informasi, dengan tujuan meningkatkan efektivitas, atau dalam beberapa kasus efisiensi, aplikasi pencarian. Dari perspektif lain, dua tugas ini adalah masalah pembelajaran mesin klasik. Dalam pembelajaran mesin, algoritma pembelajaran biasanya dicirikan sebagai diawasi atau tanpa pengawasan. Dalam pembelajaran terbimbing (*Supervised learning*), model dipelajari menggunakan satu set item sepenuhnya berlabel, yang sering disebut pelatihan data.

Klasifikasi teks merupakan suatu kegiatan penggolongan teks-teks bahasa alami dengan seperangkat kategori tematik yang ditentukan sebelumnya. Oleh karena itu, klasifikasi dokumen elektronik terutama mengacu pada pengkategorian dokumen berdasarkan rekan-rekan mereka ke dalam kelompok yang relevan. Pertumbuhan informasi teks yang tinggi di internet dan lingkungan data yang besar menghasilkan miliaran dokumen elektronik untuk dibuat, diedit dan disimpan secara digital.

Para peneliti pencarian informasi melakukan penelitian teks dengan tujuan meningkatkan efektivitas, atau dalam beberapa kasus efisiensi, aplikasi pencarian. Dari perspektif lain, dua tugas ini *clustering* dan klasifikasi adalah masalah pembelajaran mesin klasik. Dalam pembelajaran mesin, algoritma pembelajaran biasanya dicirikan sebagai diawasi atau tanpa pengawasan. Dalam pembelajaran diawasi, model dipelajari menggunakan satu set item sepenuhnya berlabel, yang sering disebut training set. Setelah model dipelajari, dapat diterapkan untuk satu set item berlabel, disebut tes set, untuk secara otomatis menerapkan label. Klasifikasi sering dilemparkan sebagai masalah belajar diawasi. Misalnya, diberi satu set email yang telah diberi label sebagai “spam” atau “bukan spam” (pelatihan set), model klasifikasi dapat dipelajari.

Oleh karena itu skripsi ini akan membahas tentang proses klasifikasi dimana klasifikasi atau pengkategorian data dapat membantu dalam mengelompokan suatu teks bahasa Arab.

1.2.Rumusan Masalah

Berdasarkan dari uraian latar belakang yang telah dipaparkan sebelumnya, maka rumusan masalah dalam penulisan skripsi ini, yaitu:

1. Apa pengaruh fungsi kernel terhadap hasil klasifikasi kernel naïve bayes?
2. Bagaimana hasil klasifikasi yang diperoleh berdasarkan nilai terbaik pada *f-measure* dan *purity*?

1.3. Batasan Masalah

Adapun batasan masalah pada penelitian tugas akhir ini adalah sebagai berikut:

1. Data yang digunakan berupa dokumen hadis Bukhari dan hadis Muslim pada bab shalat, iman, haji, zakat dan puasa saja.

1.4. Tujuan

Berdasarkan latar belakang dan rumusan masalah yang telah dipaparkan sebelumnya, maka tujuan dalam penulisan skripsi adalah sebagai berikut:

1. Untuk mengetahui pengaruh dan proses kernel terhadap hasil klasifikasi kernel naïve bayes.
2. Untuk memperoleh dokumen yang berkorelasi berdasarkan nilai terbaik pada *f-measure* dan *purity*.

1.5. Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah dengan mengkaji sumber pustaka berupa buku, jurnal, dan skripsi yang berkaitan dengan *Kernel Naive Bayes*, serta pembahasan lain yang mendukung dalam penelitian ini. Data yang digunakan dalam penelitian ini berupa teks bahasa arab berupa hadis dimana hadis yang diambil berupa hadis bukhari dan muslim.

Pada analisis data kontinu maka metode yang digunakan adalah metode *kernel naive bayes* dimana metode tersebut merupakan metode *naive bayes* dengan masuknya parameter penghalus berupa kernel kemudian akan dianalisis fungsi kernel yang berpengaruh terhadap klasifikasi *kernel naïve bayes*.

1.6 Sistematika Penelitian

Berdasarkan sistematika penulisannya, skripsi ini terdiri atas empat bab serta daftar pustaka, dimana dalam setiap bab terdapat beberapa subbab yang dirincikan sebagai berikut.

BAB I : PENDAHULUAN

Bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penelitian.

BAB II : LANDASAN TEORI

Bab ini berisi tentang teori-teori yang mendasari pembahasan dalam studi literatur ini. Secara garis besar, mencakup klasifikasi teks arab, metode kernel *naive bayes*.

BAB III : ANALISIS FUNGSI KERNEL PADA METODE KERNEL NAÏVE BAYES CLASSIFIER DENGAN DATA HADIS

Bab ini berisi pembahasan utama dari studi yang meliputi pembahasan mengenai klasifikasi Teks Arab berdasarkan pada penggolongan kernel *naive bayes*.

BAB IV : STUDI KASUS DAN SIMULASI PERHITUNGAN

Bab ini merupakan intisari dari bab-bab sebelumnya yang terdiri kesimpulan dan saran untuk pengembangan penelitian yang lebih baik.

BAB V : PENUTUP