

BAB I

PENDAHULUAN

1.1 Latar Belakang

Surat kabar atau koran merupakan salah satu bentuk media massa yang memuat berita-berita terkini dengan harga yang terjangkau oleh masyarakat. Seiring perkembangan zaman, trend dari media ini juga berkembang. Perusahaan-perusahaan dari media massa surat kabar sudah memanfaatkan teknologi internet sehingga dapat menyebarkan berita-berita terkini secara *online* melalui media massa *online*. Media massa *online* ini dapat diakses oleh masyarakat dengan lebih cepat dan lebih murah. Akan tetapi kita harus memilih media massa mana yang akan kita gunakan yang dapat dipercaya. Sebagaimana firman Allah Subhanahu Wa Ta'ala dalam Qur'an Surat Al Hujurat ayat 6.

فَعَلَّمْ مَا عَلَىٰ فَتَصْبِحُوا بِجَهْلَةٍ قَوْمًا تُصِيبُوا أَنْ فَتَنِينَا بِنَبَأٍ فَاسِقٌ جَاءَكُمْ إِنْ ءَامَنُوا الَّذِينَ يَأْتِيهَا
تَدْمِينٌ

Artinya: “Hai orang-orang yang beriman, jika datang kepadamu orang fasik membawa suatu berita, maka periksalah dengan teliti agar kamu tidak menimpakan suatu musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu.” (Q.S. Al-Hujurat : 6)

Dalam surat lain Allah Subhanahu Wa Ta'ala berfirman.

مِنْهُمْ أُولَىٰ وَإِلَىٰ الرَّسُولِ إِلَىٰ رُدُّوهُ وَلَوْ ۖ بِإِذَاعُوا الْخَوْفِ أَوْ الْأَمْنِ مِنْ أَمْرٍ جَاءَهُمْ وَإِذَا
قَلِيلٌ إِلَّا الشَّيْطَانُ لَا تَبِعْتُمْ وَرَحْمَتُهُ عَلَيْكُمْ اللَّهُ فَضْلٌ وَلَوْلَا ۖ مِنْهُمْ يَسْتَنْبِطُونَهُ الَّذِينَ لَعَلِمَهُ

Artinya: “Dan apabila datang kepada mereka suatu berita tentang keamanan ataupun ketakutan, mereka lalu menyiarkannya. Dan kalau mereka menyerahkannya kepada Rasul dan Ulil Amri di antara mereka, tentulah orang-orang yang ingin mengetahui kebenarannya (akan dapat) mengetahuinya dari mereka (Rasul dan Ulil Amri). Kalau tidaklah karena karunia dan rahmat Allah kepada kamu, tentulah kamu mengikut setan, kecuali sebahagian kecil saja (di antaramu).” (Q.S An-Nisa : 83)

Berbagai fasilitas disediakan pada media massa *online* tersebut sehingga lebih mempermudah masyarakat dalam penggunaannya. Salah satu fasilitas yang biasa diberikan adalah pencarian berita berdasarkan topik atau kategorinya.

Pada zaman ini dunia teknologi informasi dan komunikasi berkembang sangat pesat, ini menyebabkan bertambahnya data atau *file* elektronik yang merupakan hasil pertukaran informasi user di seluruh dunia. Pertambahan data yang sangat cepat menyebabkan penumpukan data yang belum dimanfaatkan yang dianggap sebagai *spam*, padahal dari sekumpulan data tersebut dapat diperoleh informasi yang bermanfaat untuk digali. Salah satu ilmu yang menangani hal ini yakni *data mining* [7].

Data mining adalah suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar. Proses *data mining* seringkali menggunakan metode statistika, matematika, hingga memanfaatkan teknologi *artificial intelligence*. Nama alternatifnya yaitu *Knowledge discovery (mining) in databases (KDD)*, *knowledge extraction*, *data/pattern analysis*, *data archeology*, *data dredging*, *information harvesting*, *business intelligence*, dan lain-lain.

Jika dilihat dilihat pada gambar dalam proses KDD tersebut, Banyak konsep dan teknik yang digunakan dalam proses *data mining*. Proses tersebut membutuhkan beberapa langkah untuk mendapatkan sebuah data yang diinginkan [2].

Informasi merupakan sekumpulan data yang sudah diproses dan dikelola sehingga dapat dipahami dan bermanfaat bagi penerimanya. Informasi juga merupakan sarana komunikasi antara sesama manusia. Informasi adalah hasil pengolahan data yang memberikan arti dan manfaat tertentu bagi orang yang menerimanya (Susanto, 2017). Pada zaman dahulu penyebaran informasi sangat dibatasi dengan jarak dan waktu karena teknologi pada saat itu masih terbatas. Di era teknologi modern saat ini kendala jarak dan waktu dalam penyebaran informasi sudah teratasi. Apalagi pada era teknologi modern saat ini informasi dapat disebar dengan mudah melalui berbagai macam media. Salah satunya yaitu berita [4].

Berita adalah laporan mengenai suatu peristiwa atau kejadian yang terbaru; laporan mengenai fakta-fakta yang aktual, menarik perhatian, dinilai penting atau

luar biasa (Budiman, 2005). Sedangkan menurut Dyoti (2003), berita adalah semua peristiwa yang sudah atau akan terjadi yang perlu diketahui oleh manusia pembacanya/pendengarnya/pemirsanya. Secara umum, penulisan berita menggunakan model deduktif, artinya pembeberan fakta dimulai dari hal-hal yang bersifat umum ke hal yang khusus. Dari yang paling penting ke hal yang kurang penting. Struktur tulisannya dikenal dengan bentuk piramida terbalik. Karena struktur demikian ini cukup menarik untuk diteliti pengaruh pembobotan berdasarkan posisi kata terhadap keakuratan hasil klasifikasi [5].

Text Mining adalah proses mengekstraksi informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi pola yang menarik dari kumpulan dokumen yang berupa data tekstual yang tidak terstruktur [8].

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 *training set*[2].

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 [2].

Oleh karena itu diperlukan metode klasifikasi untuk mengkategorikan berita-berita yang sudah ada. Ada beragam teknik klasifikasi dokumen, di antaranya adalah *Naive Bayes classifier*, *Decision Trees*, dan *Support Vector Machines*. Pada skripsi ini penulis membatasi pengkajian pada metode *Naive*

Bayes classifier. Metode *Naive Bayes classifier* ini merupakan metode yang paling populer digunakan dalam pengklasifikasian dokumen.

Lalu pengklasifikasi *bayes* merupakan salah satu pengklasifikasi statistik, di mana pengklasifikasi ini dapat memprediksi probabilitas keanggotaan kelas suatu data *tuple* yang akan masuk ke dalam kelas tertentu, sesuai dengan perhitungan probabilitas. Pengklasifikasi *Bayes* didasari oleh teorema *bayes* yang ditemukan oleh Thomas Bayes pada abad ke-18. Dalam studi perbandingan algoritma klasifikasi telah ditemukan *simple bayesian* atau yang biasa dikenal dengan *Naive Bayes classifier*. *Naive Bayes classifier* menunjukkan akurasi dan kecepatan yang tinggi bila diterapkan pada *database* yang besar. Metode ini sering digunakan dalam menyelesaikan masalah dalam bidang mesin pembelajaran karena metode ini dikenal memiliki tingkat akurasi yang tinggi dengan perhitungan sederhana [2].

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah penulis sampaikan sebelumnya, penulis merumuskan masalah dalam penulisan skripsi ini, yaitu:

1. Mengkategorikan berita islam menggunakan metode *Naive Bayes Classifier* (NBC).
2. Membandingkan kinerja *Naive Bayes Classifier* (NBC) antara model kejadian *Multinomial*, *Bernoulli* dan *Gaussian* untuk mengkategorikan berita islam.

1.3 Batasan Masalah

Agar tidak menyimpang dari pembahasan, maka ada beberapa batasan masalah dalam skripsi ini, sebagai berikut:

1. Data yang digunakan merupakan data yang dikumpulkan dari *web* *republika.co.id* berupa berita islam. Pertanyaan yang digunakan dibangun berdasarkan isi konteks hadis.
2. Pada *Naive Bayes Classifier* (NBC) menggunakan model kejadian *Multinomial*, *Bernoulli* dan *Gaussian*.

1.4 Tujuan dan Manfaat Penelitian

Adapun tujuan dan manfaat dari penelitian skripsi ini diantaranya:

1. Untuk Menerapkan metode *Naive Bayes Classifier* (NBC) dalam mengkategorikan berita islam
2. Untuk mengetahui model kejadian yang lebih baik dan juga lebih buruk diantara model kejadian *Multinomial*, *Bernoulli* dan *Gaussian* dalam mengkategorikan berita islam.

Adapun manfaat yang diperoleh dari hasil penelitian Skripsi ini antara lain adalah sebagai berikut:

1. Memberikan pemahaman tentang cara *Naive Bayes Classifier* (NBC) mengkategorikan berita islam dengan dengan berbagai model kejadian.
2. Memberikan pengetahuan mengenai kinerja *Naive Bayes Classifier* (NBC) untuk kategorisasi berita islam.

1.5 Metode Penelitian

Metodologi dalam penelitian skripsi ini diantaranya:

1. Studi Literatur

Pada tahap ini dilakukan pengumpulan fakta-fakta dan teori penunjang yang berhubungan dengan metode *Naive Bayes Classifier* (NBC) dari berbagai sumber, diantaranya buku, jurnal, artikel dan lain sebagainya.

2. Analisis

Pada tahap ini penulis menganalisis dan melakukan penelitian terhadap metode KLASIFIKASI TEKS BERITA ISLAM DENGAN METODE *NAIVE BAYES MULTINOMIAL*, *BERNOULLI* DAN *GAUSSIAN*.

3. Simulasi

Pada tahap ini dilakukan simulasi *Naive Bayes Classifier* (NBC) menggunakan Google Colab agar didapatkan data hasil *output* untuk dilakukan perbandingan.

1.6 Sistematika Penulisan

Sistematika penulisan pada skripsi ini terdiri dari lima bab dan di setiap bab terdiri dari beberapa subbab.

BAB 1 : PENDAHULUAN

Pada bab ini memuat latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan metode penelitian, serta sistematika penulisan

BAB 2 : LANDASAN TEORI

Pada bab ini berisi uraian materi tentang hal-hal yang melandasi pembahasan masalah dan teori-teori yang digunakan sebagai pedoman untuk menyelesaikan permasalahan.

BAB 3 : KLASIFIKASI TEKS BERITA ISLAM DENGAN METODE *NAIVE BAYES MULTINOMIAL, BERNOULLI* DAN *GAUSSIAN*

Bab ini berisi tentang penerapan *Naive Bayes Classifier* (NBC) untuk klasifikasi teks. Mulai dari pengambilan data, prapemrosesan, pembobotan kata, klasifikasi hingga hasil akurasi.

BAB 4 : ANALISIS HASIL KLASIFIKASI TEKS BERITA ISLAM DENGAN METODE *NAIVE BAYES MULTINOMIAL, BERNOULLI* DAN *GAUSSIAN*

Pada bab ini akan dijelaskan tentang analisis hasil implementasi dan percobaan *Naive Bayes Classifier* (NBC) menggunakan model kejadian *Multinomial*, *Bernoulli* dan *Gaussian* pada teks berita islam.

BAB 5 : PENUTUP

Pada bab ini dipaparkan kesimpulan sebagai jawaban dari rumusan masalah yang diajukan serta saran yang berisi rekomendasi untuk penelitian selanjutnya pada topik yang serupa.