

BAB I

PENDAHULUAN

1.1. Latar Belakang

Berita palsu atau hoax merupakan berita yang berisi informasi tidak benar dan berbahaya. Dikatakan berbahaya karena berita palsu dapat mengubah persepsi pembacanya terhadap sebuah informasi palsu yang disampaikan seolah-olah sebagai kebenaran [1]. Alasan penyebaran berita palsu ini bisa berdasarkan berbagai macam motif seperti politik atau komersil. Sementara efek yang ditimbulkan antara lain ialah dapat menyebabkan rusaknya kredibilitas seseorang ataupun organisasi, kerugian secara materi, bahkan menebar ketakutan di masyarakat [2].

Pesatnya penyebaran berita palsu saat ini, didukung pula oleh keberadaan sosial media. Orang-orang dengan mudahnya melakukan *share*. Padahal yang mereka baca adalah berita dengan sumber yang tidak jelas. Selain untuk mempengaruhi pembaca, berita palsu juga dibuat untuk mendapatkan keuntungan dari banyaknya berita tersebut dibaca. Apalagi jika menyangkut isu-isu sensitif seperti agama, politik, perselisihan, kesehatan, dan lain sebagainya [2]. Oleh karena itu, penelitian klasifikasi berita palsu ini dibuat, untuk membantu masyarakat dalam mendeteksi berita palsu secara otomatis menggunakan pendekatan pembelajaran mesin.

Pembuatan sistem klasifikasi berita palsu menggunakan pendekatan pembelajaran mesin ini akan memiliki beberapa tahapan. Pertama, yaitu tahap pengumpulan data berupa berita online yang diambil dari internet dan dilabeli secara manual terlebih dahulu. Kedua, ialah memproses data berita tersebut menggunakan teknik text mining. Data berita dibersihkan terlebih dahulu untuk membuang data tidak perlu (*preprocessing*), kemudian berita diubah ke dalam potongan-potongan kata untuk dijadikan sebuah fitur. Ketiga, fitur-fitur yang telah ditentukan kemudian diseleksi kembali dan dijadikan sebagai model agar nantinya dapat dijadikan sebagai acuan sebuah berita itu palsu atau tidak.

Secara umum, sistem klasifikasi teks yang berbasis pembelajaran mesin terdiri dari pra-proses, ekstraksi fitur, seleksi fitur dan klasifikasi. Beberapa algoritma diantaranya, yang biasa digunakan untuk klasifikasi *Naive Bayes* (Callum & Nigam, 1998), Model *Multivariat Bernoulli* (Joachim, 1997), *Rocchio* (Joachim, 1997), *K-*



Nearest Neighbor (Kotsiantis, 2005), dan *Support Vector Machine* (Manning et al, 2008).

Pada penelitian ini, tahap *preprocessing* dilakukan dengan cara *tokenizing*, *case folding*, penghapusan *stopwords*, dan *stemming*. Tahap ekstraksi fitur dilakukan dengan teknik unigram. Sementara tahap penyeleksian fitur, dilakukan dengan menggunakan teknik TF-IDF untuk pemberian bobot pada fitur [3], yang kemudian akan digabungkan dengan metode *Mutual Information* agar dapat meningkatkan recall dan presisi. Setelah itu, pada tahap klasifikasi berita palsu, metode yang akan digunakan ialah *Naive Bayes*.

Naive Bayes dipilih karena pada penelitian klasifikasi berita palsu sebelumnya, mendapatkan nilai akurasi, *precision* dan *recall* paling tinggi diantara metode lain [4]. Perhitungan akurasi, *precision* dan *recall* didapatkan dari pengujian model menggunakan metode *k-fold crossvalidation* dan *confussion matrix*. Di mana data yang telah dikumpulkan sebelumnya dibagi menjadi data training dan data testing untuk dilakukan pengujian. Namun, hal itu hanya berlaku apabila klasifikasi dilakukan pada kategori berita secara umum. Bila berita diklasifikasi berdasarkan kategori atau topik tertentu, sistem klasifikasi cenderung salah dalam mendeteksi berita palsu sehingga akurasi, *precision* dan *recall* menurun drastis. Oleh karena itu, pada penelitian ini, data berita akan diklasifikasikan secara otomatis sesuai dengan topik tertentu terlebih dahulu sebelum dijadikan model. Metode yang digunakan pada tahap klasifikasi topik ini ialah *Support Vector Machine* (SVM).

SVM dipilih sebagai metode klasifikasi kategori berita karena pada penelitian sebelumnya metode ini berhasil memberikan akurasi, *precision* dan *recall* yang tinggi dalam mengklasifikasi berita online berbahasa Indonesia secara otomatis [5]. Selain penambahan tahap klasifikasi kategori, penelitian ini juga menggunakan judul berita sebagai bahan data training. Hal ini dikarenakan judul berita memiliki keterkaitan yang kuat dalam pendeteksian berita palsu [7].

1.2. Rumusan Masalah

Berdasarkan latar belakang diatas maka didapatkan perumusan masalah yaitu seberapa tinggi akurasi, *precision* dan *recall* dari klasifikasi menggunakan metode SVM dan *Naive Bayes* terhadap berita palsu bahasa Indonesia?



1.3. Tujuan

Mengacu pada latar belakang serta perumusan masalah diatas, maka tujuan dari penelitian ini adalah mengetahui seberapa tinggi akurasi, *precision* dan *recall* penggunaan metode SVM dan *Naive Bayes* untuk melakukan klasifikasi terhadap berita online asli atau palsu.

1.4. Manfaat

Manfaat dari hasil penelitian ini ialah adanya media berbasis pembelajaran mesin yang dapat memberikan pertimbangan lebih terhadap benar atau tidaknya sebuah berita. Sehingga dapat mengurangi resiko terjadinya kesalah pahaman atau beredarnya informasi yang tidak benar.

1.5. Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Data berita online yang digunakan merupakan berita yang dilabeli secara manual terlebih dahulu berdasarkan Forum Anti Fitnah dan Hoax (<http://turnbackhoax.id/>).
2. Bahasa pemrograman yang digunakan ialah Python, Flask dan database MySQL.
3. Topik berita yang diambil hanya 3 topik saja (politik, tokoh, serta kesehatan dan makanan).
4. Data berita diambil dari rentan bulan Mei 2014 hingga bulan Agustus 2018.
5. Jumlah total data ialah 150 berita, terdiri dari 75 berita palsu dan 75 berita benar.
6. Sistem tidak menggunakan acuan kumpulan kata yang memiliki potensi sebagai penciri berita palsu.
7. Data berita palsu yang digunakan ialah data berita yang banyak mengandung kata-kata ujaran kebencian, fitnah, dan hasutan yang kuat.

1.6. Metode Penelitian

Secara umum, penelitian ini dilakukan dengan beberapa tahapan, yaitu:

a. Studi Pustaka

Pada tahap ini, semua informasi meliputi penelitian seperti literatur tentang berita palsu, seleksi fitur, serta algoritma SVM



dan *Naive Bayes* sehingga dapat digunakan untuk membantu menyelesaikan proses pengerjaan riset tugas akhir.

b. Pengumpulan Data

Data yang dikumpulkan berupa kumpulan berita palsu dan asli dari Forum Anti Fitnah dan Hoax. Forum ini dikelola oleh MAFINDO yang sudah didirikan sejak tanggal 19 November 2016, berdasarkan akta Notaris Nomor 1 Tanggal 19 November 2016 oleh ISMA JANUARTI, SH., M.KN. SK (Surat Keputusan) Pendirian Perkumpulan NOMOR AHU-0078919.AH.01.07.TAHUN 2016 Menteri Hukum dan Hak Asasi Manusia.

c. Perancangan dan Implementasi Sistem

Data yang sudah terkumpul mulai diolah sesuai dengan tahapan serta metode yang direncanakan menggunakan *Python* dengan dukungan library *skicit.learn*, serta *framework Flask*.

d. Uji Coba Sistem

Proses uji coba sistem dilakukan menggunakan data *testing* dan data training yang sudah ditentukan menggunakan metode *k-fold crossvalidation*, untuk mengetahui apakah metode hasil penelitian sudah bekerja dengan baik dan benar.

e. Analisa dan Evaluasi

Menganalisa dan mengevaluasi hasil uji coba sistem untuk melihat tingkat akurasi dan performa sistem.

f. Dokumentasi

Dokumentasi dilakukan sebagai penyusunan laporan tugas akhir yang berlangsung mulai dari tahap studi pustaka hingga tahap evaluasi.

1.7. Sistematika Penulisan

Adapun sistematika pembahasan tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisikan latar belakang dari penelitian ini, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II STUDI PUSTAKA

Pada bab ini dijelaskan mengenai materi-materi yang menjadi landasan dalam penyusunan penelitian ini. Materi-materi tersebut



mengenai *preprocessing*, ekstraksi fitur, seleksi fitur, klasifikasi berikut metodenya, serta evaluasi klasifikasi menggunakan tabel confusion matrix dan *cross validation*.

BAB III PERANCANGAN SISTEM

Bab ini menjelaskan rancangan dari seluruh sistem yang akan dibuat dan alur dari sistem yang akan dikembangkan.

BAB IV IMPLEMENTASI SISTEM

Pada bab ini dijelaskan implementasi sistem dari rancangan yang telah dibuat pada bab sebelumnya, dimana sistem akan diuji coba berdasarkan skenario yang telah ditentukan.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dari hasil uji coba serta saran yang dapat digunakan untuk pengembangan selanjutnya.

