

BAB I

PENDAHULUAN

1.1 Latar Belakang

Asosiasi Penyelenggaraan Internet Indonesia (APJII) pada tahun 2018 telah melakukan survei, Ada sekitar 171,17 Juta pengguna internet di Indonesia atau sekitar 64,8 persen yang sudah terhubung ke internet dari total 264 juta jiwa penduduk Indonesia [1]. Angka ini mengalami kenaikan dari tahun 2017 sekitar 144,83 juta pengguna internet di Indonesia. Mengacu pada survei, sekitar 85,7% dari pengguna internet menggunakannya untuk mengakses konten media sosial. Media sosial itu diantaranya Facebook, Instagram, Youtube dan Twitter

Sosial media Twitter memiliki pengguna 7.05% atau sekitar 12,06 juta pengguna di Indonesia. Twitter menjadi platform media sosial paling aktif ke 4 di Indonesia dengan angka keaktifan mencapai 52% [2]. Dengan keaktifan tersebut terdapat banyak informasi dan *tweet* dari para penggunanya tentang berbagai hal : pendidikan, pekerjaan, hiburan, politik termasuk juga kebijakan pemerintah. Salah satu kebijakan pemerintah yang *trending topic* di Twitter pada eptember 2019 adalah tentang rencana pengesahan RKUHP.

Rencana pengesahan RKUHP (Rancangan Kitab Undang Undang Hukum Pidana) sempat masuk *trending topic* pada pertengahan bulan September 2019. Para pengguna Twitter memberikan opininya baik pro maupun kontra terhadap poin-poin yang ada pada RKUHP. Opini-opini tersebut dapat diolah, dicari dan dianalisis orientasi sentimennya. Analisis Sentimen adalah proses untuk menentukan apakah suatu tulisan seseorang termasuk dalam katageori positif, negatif. Dikenal juga

sebagai opinion mining, bertujuan untuk mendapatkan pendapat atau sikap masyarakat [3].

Penulis pada penelitian ini akan melakukan analisis sentimen terhadap pengguna Twitter mengenai rencana pemerintah dalam pengesahan RKUHP 2019. Dengan masukan berupa data *tweet* dalam Bahasa Indonesia, akan dilakukan klasifikasi dengan algoritma KNN (K-Nearest Neighbors). Algoritma ini dapat dijelaskan secara sederhana sebagai metode yang digunakan untuk mengelompokkan objek berdasarkan contoh pelatihan terdekat di ruang fitur [4]. KNN digunakan untuk menentukan apakah *tweet* tersebut berorientasi sentimen positif, netral atau negatif.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, dapat dirumuskan permasalahan yang diselesaikan dalam penelitian ini adalah bagaimana mengamati dan menganalisis opini pengguna twitter mengenai rencana pengesahan Rancangan Kitab Undang Undang Hukum Pidana (RKUHP) 2019 dan kemudian mengklasifikasikan data *tweets* tersebut, bagaimana tingkat akurasi K-NN dalam sentiment analisis Bahasa Indonesia.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah :Data yang digunakan pada penelitian ini adalah data *tweets* dalam Bahasa Indonesia

1. Data tweets yang digunakan sebanyak 3000 yang telah di filter dan diambil pada tanggal 20 september sampai 30 September 2019.

2. Data *tweets* yang akan dinilai adalah *tweets* yang mengandung *keyword* berikut: rkuhp, ruukuhp, demo mahasiswa, pasal ngawur, tolakrkuhp, tundarkuhp, revisirkuhp.
3. Metode pengklasifikasian yang digunakan dalam penelitian ini adalah metode K-Nearest Neighbor (KNN).
4. Metode stemming menggunakan algoritma Nazief & Adriani
5. Hasil keluaran penelitian ini dari pengklasifikasian sentimen menggunakan K-Nearest Neighbor (KNN) berupa sentimen positif dan negatif serta akurasi dari metode ini.

1.4 Tujuan Penelitian

1. Menerapkan Algoritma K-Nearest Neighbor untuk klasifikasi sentimen masyarakat dari data twitter dengan topik rencana pengesahan RKUHP 2019.
2. Mengetahui sentimen pengguna twitter terhadap rencana pengesahan RKUHP 2019.
3. Mengetahui seberapa cocok Algoritma K-Nearest Neighbor untuk analisis sentimen Bahasa Indonesia.

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini adalah :

1. Mendapatkan hasil akurasi dari Algoritma K-Nearest Neighbor pada klasifikasi sentimen dengan topik rencana pengesahan RKUHP 2019.
2. Memberikan informasi tentang sentiment twitter terhadap rencana pengesahan RKUHP 2019.

3. Menjadi referensi untuk penelitian selanjutnya yang berkaitan dengan Analisis Sentimen Bahasa Indonesia.

1.6 Metode Penelitian

Agar menghasilkan data dan laporan yang akurat dan relevan, metode penelitian yang digunakan dalam penelitian ini meliputi beberapa tahap yaitu sebagai berikut :

1.6.1 Metode Pengumpulan Data

1. Metode Deskriptis

Penulis melakukan crawling data *tweets* pada trending twitter tentang RKUHP menggunakan tools TAGS dengan mengarsipkan data langsung ke google sheet. Kemudian file hasil crawling tersebut disimpan dalam bentuk file csv. Data yang telah diambil selanjutnya difilter untuk mendapatkan data yang siap untuk dianalisis.

2. Studi Pustaka

Penulis menghimpun informasi dengan mengumpulkan referensi diantaranya: buku, jurnal-jurnal, dokumen skripsi maupun website mengenai metode K-Nearest Neighbor (KNN) guna menunjang penelitian.

1.6.2 Metode Analisis

Tahap analisis digunakan untuk menganalisis dan mempelajari data maupun informasi yang sudah didapatkan pada proses pengumpulan data. Data maupun informasi yang telah di analisis digunakan untuk mengembangkan metode yang dapat mengklasifikasi data twitter. Data yang telah didapatkan akan melewati serangkaian preprocessing text (Cleaning, Case Folding, Tokenization, Filtering,

Stemming, Normalisazion). Data yang telah melalui tahap preprocessing akan dibagi menjadi 2, yaitu data latih dan data uji. Dalam penelitian ini, proses K Nearest Neighbor meliputi 3 proses, yaitu: Menghitung bobot kata (TF-IDF), Menghitung tingkat kemiripan (Cosine Similarity) dan Menentukan nilai sentiment. Selanjutnya implementasi metode K-Nearest Neighbors dalam bentuk program computer untuk mengetahui kemampuan dari metode K-NN tersebut.

1.6.3 Metode Perancangan

Tahapan dalam merancang aplikasi ini terdiri dari beberapa tahap yaitu:

1. Perancangan Sistem berisikan tahapan yang dilewati (input data, preprocessing, klasifikasi K-NN) sehingga menghasilkan akurasi.
2. Perancangan Antar Muka (Interface) Agar terjalin hubungan baik antara pengguna (user) dan aplikasi, maka perlu dirancang tampilan antarmuka.

1.6.4 Metode Pengujian

Metode pengujian yang dilakukan untuk menguji sistem dengan dilakukan perhitungan akurasi dari model pengklasifikasian yang dihasilkan. Maka tahap ini menggunakan metode confusion matrix. Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi dalam perhitungan akurasi pada konsep data mining.

1.7 Sistematika Penelitian

Sistematika penulisan laporan ini terdiri dari pokok-pokok permasalahan yang dibahas pada masing-masing yang diuraikan menjadi beberapa bagian:

BAB I PENDAHULUAN

Bab ini membahas tentang gambaran umum isi tugas akhir yang meliputi latar belakang yang menguraikan konsep-konsep dasar tentang penelitian ini. Pada pendahuluan ini juga diuraikan rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penelitian.

BAB II LANDASAN TEORI

Pada bab ini akan dibahas mengenai segala sesuatu hal yang berkaitan dengan proses penulisan tugas akhir ini dan dijadikan sebagai sebuah landasan dalam penulisan dan penelitian.

BAB III METODE PENELITIAN

Bab ini berisi tentang penjelasan metode dalam menganalisa perhitungan yang digunakan dalam penelitian, permodelan data yang digunakan untuk penelitian, dan tahapan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi implementasi algoritma yang diusulkan terhadap dataset, penguraian tahapan peneliti dalam melakukan penelitian, dan hasil dari penelitian.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian yang sudah dilakukan, penarikan hasil dari penelitian, dan berisi saran yang dapat digunakan peneliti lain untuk mengembangkan hasil dari penelitian yang sudah dilakukan.

DAFTAR PUSTAKA

Berisikan tentang sumber-sumber, jurnal ataupun buku-buku yang menjadi referensi sebagai acuan dalam penyusunan skripsi ini.

