

BAB I PENDAHULUAN

1.1 Latar Belakang

Selain menggunakan internet untuk mendapatkan informasi, perkembangan pengguna media sosial di Indonesia meningkat sangat pesat berbarengan dengan data pertumbuhan pengguna Internet. Menurut Sekjen APJII, Henri Kasyfi dari total populasi penduduk Indonesia sebanyak 264 juta jiwa, sebanyak 171,17 juta jiwa atau sekitar 64,8 Persen terhubung ke Internet [1]. Data penetrasi Ponsel pintar terhadap Jumlah Penduduk Indonesia yang dipublikasikan oleh katadata.co.id yang didasarkan data Statista, pengguna telepon pintar diproyeksikan mencapai 28% dari total penduduk Indonesia pada 2019, naik 2% dari tahun 2018. Peningkatan ini diprediksi akan meningkat menjadi sekitar 33% dari total penduduk Indonesia [2]

Berdasarkan publikasi laporan yang dibuat Data Reportal , di Indonesia jumlah pengguna Twitter mencapai 10,65 juta, dengan persentase pengguna 5,1% dari total jumlah penduduk Indonesia yang berusia 13 tahun keatas [3]. Besarnya jumlah pengguna ini berbanding lurus dengan data yang dihasilkan. Berkat perkembangan teknologi, saat ini bermunculan pendekatan sistem pengolahan *big data*. Salah satu pemanfaatan *Big Data* adalah mengganti pendekatan proses koleksi dan pengolahan data melalui survei. Selain biayanya mahal, proses survei memakan waktu yang lama. Meskipun begitu, Sistem *big data* sendiri memiliki tantangan yaitu pengolahan data dinamis. Salah satu pemanfaatan *Big data* adalah *Opinion Mining*.

Pesatnya perumbuhan sosial media, *opinion mining* telah menjadi area penelitian yang paling aktif untuk *Natural Language Processing*. Secara garis besar, *opinion mining* digunakan untuk mengetahui apa yang masyarakat pikirkan dan rasakan akan suatu produk maupun jasa melalui media sosial tempat berkumpulnya banyak orang. Pada dunia virtual, orang mudah saja berkomunikasi tanpa mengenal batas-batas teritori, membicarakan topik yang sama Bersama-sama dengan mudah. Fenomena membagikan rasa kekecewaan yang dialami sangat sering dijumpai di Indonesia, begitu pula pengaduan terhadap layanan publik, kebijakan publik termasuk bidang perpajakan. [4]

Pada penelitian ini penulis membahas proses Analisa Sentimen masyarakat dengan alur kerja pada tahap pertama mendapatkan *data training* dan data uji menggunakan layanan Streaming data dari API Twitter, dilanjutkan dengan *Preprocessing* data, dan Proses Klasifikasi. Data latih yang dikumpulkan selama periode tertentu, dan melakukan implementasi model pada data yang dikumpulkan dalam satu hari untuk sebuah topik, dan Implementasi berupa Aplikasi Web yang dapat digunakan untuk melakukan *monitoring* dalam kurun waktu tertentu

Penelitian yang penulis lakukan mengacu pada analisis sentimen yang sudah pernah dilakukan terhadap penilaian pelayanan pajak secara umum dimana hasil percobaan menunjukkan bahwa klasifikasi otomatis keluhan pajak di twitter dapat dilakukan dengan dukungan penambahan teks. Metode klasifikasi SVM (Support Vector Machine) memiliki nilai akurasi 89,6%, presisi 89,3%, recall 89,6%, dan f-ukur 89,3%. Sedangkan *Naïve Bayes* memiliki nilai ROC (Receiver Operating Characteristic) terbaik yaitu sebesar 88,8%. [5]

Berdasarkan uraian diatas maka penulis tertarik untuk membuat Tugas Akhir dengan judul ***“Penerapan Algoritma Multinomial Naïve Bayes Classifier pada Analisis Sentimen Topik Perpajakan Pada Media Sosial Twitter”***.

1.2 Identifikasi Masalah

Menilai opini masyarakat terhadap suatu produk melalui metode survei konvensional memakan biaya yang tinggi, dan memerlukan waktu yang banyak bagi perusahaan sedangkan proses untuk memantau opini masyarakat akan suatu topik harus dilakukan terus menerus. Disisi lain, memberikan opini melalui survei, atau menyampaikan melalui layanan pelanggan akan memakan banyak tenaga dari sisi pengguna dalam hal ini diistilahkan sebagai masyarakat, salah satu contohnya adalah ketidakpuasan masyarakat atas pelayanan pemerintahan yang dibagikan ke media sosial, dibandingkan mengisi survei kepuasan pelanggan, atau membuat panggilan untuk menyampaikan keluhan.

1. Bagaimana mengimplementasikan Algoritma Multinomial *Naïve Bayes Classifier* pada analisis sentimen?
2. Bagaimana performansi model analisis sentimen dengan Algoritma Multinomial *Naïve Bayes*?

3. Bagaimana hasil penerapan Algoritma Multinomial *Naïve Bayes Classifier* pada topik perpajakan ?

1.3 Tujuan Tugas Akhir

Tujuan dari penelitian ini adalah

1. Mengimplementasikan Algoritma Multinomial *Naïve Bayes* pada analisa sentiment terkait topik perpajakan
2. Mengklasifikasikan komentar dari sosial media Twitter dengan topik perpajakan menggunakan Algoritma Multinomial *Naïve Bayes*.

1.4 Manfaat Tugas Akhir

Adapun manfaat dari tugas akhir ini adalah sebagai berikut :

1. Memberikan gambaran penilaian publik terhadap suatu produk, atau layanan, sehingga dapat menjadi landasan untuk perbaikan produk dan pengembangan produk lainnya.
2. Membantu proses Analisis Sentimen Publik berupa data visualisasi.
3. Memberikan gambaran pemanfaatan media sosial sebagai sumber analisis sentimen.

1.5 Ruang Lingkup Tugas Akhir

Mengingat cukup luasnya cakupan yang dikemukakan serta waktu pengerjaan yang relatif singkat, maka penulis memfokuskan masalah tanpa mengurangi nilai pokok permasalahan yang telah dirumuskan. Pada penititan analisis sentimen ini dibatasi dengan hal-hal sebagai berikut :

1. Sumber data yang digunakan sebagai dataset penelitian adalah Dataset Twitter, dengan jumlah dataset 3000 tweet dengan topik Perpajakan Berbahasa Indonesia.
2. Algoritma klasifikasi yang digunakan adalah Multinomial *Naïve Bayes*.
3. Mengklasifikasikan sentimen kedalam kategori negatif, netral, dan positif, tanpa mempertimbangkan kalimat merupakan kalimat ironi, sarkasme, satir, dan paradox.
4. Penelitian analisa sentimen dibatasi dengan studi kasus topik Insentif pajak dengan objek penelitian P2 Humas DJP.

5. Preprocessing pada penelitian ini terdiri dari normalisasi *Slangword*, *Stemming*, Menghapus *Stopword*, dan *Part of Speech Tagging*.
6. Pembuatan Aplikasi *monitoring* sentimen bertujuan menunjukkan model yang dibuat dapat diimplementasikan klasifikasi sentimen dan mendukung pengembangan aplikasi lebih lanjut.

1.6 Kerangka Berpikir

Pada Penelitian yang digunakan adalah *experimental research* dan *case study research*. *Case study* atau studi kasus yang diambil adalah mengenai topik Perpajakan dengan *experimental reaserach* menggunakan probabilitas statistik algoritma multinomial naïve bayes.

terdapat 3 tahapan utama dalam penelitian ini, yaitu pengumpulan data yang penulis dapat selama Melakukan wawancara, dan Studi pustaka. Dari observasi tersebut penulis mengidentifikasikan masalah dan masuk ke tahap perancangan. Dari identifikasi masalah tersebut penulis mengajukan pembuatan aplikasi untuk memonitoring sentimen publik sebagai aplikasi usulan. Tahap selanjutnya dibagi menjadi dua, yang dilakukan secara paralel yaitu pembuatan model yang mengacu pada *Machine learning Life Cycle*, dan aplikasi berbasis website

1.7 Sistematika Penulisan

Sistematika penulisan dapat dibagi menjadi tiga bagian, yaitu awal, isi, dan akhir. Berikut adalah sistematika penulisannya :

BAB I PENDAHULUAN

Pada bab ini dijelaskan mengenai hal yang terdiri dari latar belakang, perumusan masalah, batasan masalah, maksud dan tujuan penelitian, sistematika penulisan penelitian.

BAB II TINJAUAN PUSTAKA

Pada bab ini berisi Landasan teori yang digunakan dalam tugas akhir, dan konsep dalam menyelesaikan masalah pada penelitian ini.

BAB III METODE PENELITIAN

Pada bab ini dijelaskan mengenai alur rencana penelitian, objek penelitian, teknik dan waktu pengumpulan data, dan perancangan implementasi yang diusulkan dalam menyelesaikan masalah.

BAB IV PEMBAHASAN

Bab ini memaparkan hasil penerapan Algoritma Multinomial Naïve Bayes Classifier berupa hasil klasifikasi, dan hasil ukuran performansi sistem.

BAB V KESIMPULAN

Bab ini berisikan kesimpulan mengenai hasil analisis pada sistem, memberikan saran untuk pengembangan lebih lanjut.