

1 PENDAHULUAN

1.1 Latar Belakang

Pada era globalisasi ini, pertukaran informasi terjadi dengan begitu mudah melalui media sosial, salah satunya adalah twitter. Menurut eBizMBA, twitter berada di urutan kedua sebagai media sosial terpopuler setelah Facebook di dalam Top 15 Most Popular Social Networking Sites dengan perkiraan pengunjung bulanan yang unik berjumlah 310.000.000 [18]. Twitter menjadi tempat sebagian besar masyarakat untuk mengemukakan opini mereka terkait isu yang sedang hangat dibicarakan pada saat tertentu dengan bebas. Opini-opini yang ada di twitter dapat digunakan untuk menilai sentimen atas suatu topik tertentu, seperti produk, film, jasa, tokoh publik, kebijakan pemerintah dan sebagainya. Salah satu kebijakan pemerintah yang banyak menarik perhatian masyarakat adalah amnesti pajak. Amnesti pajak merupakan kebijakan pemerintah di bidang perpajakan yang berlaku hingga 31 Maret 2017 [2].

Pada Tugas Akhir ini, *tweet* yang mengandung opini masyarakat terhadap amnesti pajak diklasifikasikan menggunakan metode klasifikasi *naïve Bayesian classification*. *Tweet* yang berisi amnesti pajak diambil dengan menggunakan NodeXL. Namun, data yang telah dikumpulkan masih belum dapat digunakan untuk klasifikasi karena terdapat kata-kata yang tidak baku dan terdapat banyak *noise* di dalam data tersebut sehingga perlu dilakukan *preprocessing*. Tahapan *preprocessing* dilakukan untuk menghilangkan hal yang tidak dibutuhkan (*url*, *mention*), *tokenization*, *stopword removal*, dan *stemming*. Berdasarkan penelitian yang dilakukan Ledy Agusta, *stemming* dengan algoritma nazief dan adriani memiliki hasil akurasi yang lebih baik dibandingkan dengan *stemming* menggunakan algoritma porter [5]. Setelah itu, data tersebut diklasifikasikan menggunakan *naïve Bayesian classification*. Pendekatan *naïve Bayesian classification* merupakan pendekatan yang mengacu pada teorema Bayes yang menggunakan prinsip peluang statistika untuk mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru untuk menyelesaikan masalah klasifikasi [6]. Pada penelitian ini, data twitter yang digunakan adalah data yang berisi opini sehingga data yang telah dikumpulkan terlebih dahulu melalui *filtering* manual dimana *tweet* dengan *username* kemenkeuri, dirjenpajakri, akun kantor pajak daerah, dan media-media berita dihapuskan. Kata-kata yang tidak baku pada data yang telah di-*filter* diubah menjadi kata baku secara manual. Pada penelitian tugas akhir ini, sistem mampu mengklasifikasikan data twitter ke dalam tiga kelas yaitu positif, negatif, dan netral dengan ekstraksi fitur menggunakan unigram dan frekuensi kata (*term frequency*). Tetapi akurasi yang dihasilkan hanya mencapai 53,45% dengan data pembelajaran sebesar 80% dari 578 data twitter.

1.2 Perumusan Masalah

Permasalahan yang dapat dirumuskan berdasarkan latar belakang yang telah diuraikan di atas adalah sebagai berikut :

- a. Bagaimana cara menerapkan metode *naïve Bayesian classification* dalam melakukan pengklasifikasian *tweet* yang berisi informasi tanggapan masyarakat mengenai kebijakan pemerintah?

- b. Bagaimana tingkat akurasi dari penggunaan metode *naive Bayesian classification* dalam mengklasifikasikan tweet yang berisi informasi tanggapan masyarakat mengenai kebijakan pemerintah?

1.3 Tujuan

Tujuan dari penyusunan Tugas Akhir ini adalah :

- a. Melakukan implementasi algoritma *naive Bayesian classification* dalam melakukan pengklasifikasian tweet yang berisi informasi mengenai kebijakan pemerintah.
- b. Menentukan tingkat akurasi dari penggunaan metode *naive Bayesian classification* dalam mengklasifikasikan tweet yang berisi informasi mengenai kebijakan pemerintah.

1.4 Batasan Masalah

Batasan masalah pada pengerjaan Tugas Akhir ini yaitu :

1. *Tweet* yang digunakan adalah *tweet* berbahasa Indonesia.
2. Hasil pengklasifikasian adalah kelas positif, negatif dan netral.
3. Data dikumpulkan dari tanggal 27 Juli 2016 sampai 17 Oktober 2016.
4. Data yang digunakan berjumlah 578 data twitter berisi opini.
5. Kata kunci yang digunakan dalam mengumpulkan data amnesti pajak adalah #amnestipajak, #pengampunanpajak, #taxamnesty, amnesti pajak, pengampunan pajak, dan *tax amnesty*.
6. Data *tweet* dengan *username* kemenkeuri, ditjenpajakri, klinik_pajak, dan media berita dihapuskan.
7. Kata-kata tidak baku pada data *tweet* yang digunakan diubah menjadi baku secara manual.

1.5 Metodologi Penyelesaian Masalah

Pembuatan Tugas Akhir ini dilakukan dengan metodologi penyelesaian masalah sebagai berikut :

1. Studi Literatur

Studi literatur adalah suatu proses yang dimaksudkan untuk melakukan pencarian, pengumpulan, penyaringan, pembelajaran, pemahaman dan pendalaman literatur dengan mempelajari buku, jurnal, paper, dan mengunjungi *website*. Studi literatur ini bertujuan untuk mencari informasi yang berhubungan dengan klasifikasi, algoritma *naive Bayesian classification*, dan data mining.

2. Pengumpulan Data

Mengumpulkan *dataset* mengenai tanggapan masyarakat mengenai kebijakan pemerintah dengan cara mengambil *tweet* yang mengandung hashtag #amnestipajak, #pengampunanpajak, #taxamnesty. Pengambilan *tweet* dilakukan dengan menggunakan aplikasi NodeXL. Data yang telah didapat disimpan ke dalam Microsoft Excel.

3. Analisis Awal dan Perancangan Sistem

Pada tahap ini dilakukan analisis untuk mengetahui kebutuhan sistem. Setelah itu, dilakukan perancangan sistem yang sesuai dengan kebutuhan dan masalah yang ada.

4. Implementasi

Pada tahap ini dilakukan implementasi dari rancangan yang telah dibuat. Dalam implementasi dilakukan *preprocessing* dan klasifikasi.

5. Pengujian

Pengujian dilakukan untuk menguji sistem dengan menggunakan data yang telah didapatkan. Dari pengujian yang telah dilakukan, maka akan dianalisis hasil dari pengujian tersebut.

6. Pembuatan Laporan

Laporan dibuat secara menyeluruh untuk melakukan dokumentasi dan pelaporan dari semua kegiatan yang telah dilakukan dalam pelaksanaan Tugas Akhir ini.

1.6 Sistematika Penulisan

Sistematika penulisan pada buku ini adalah sebagai berikut.

1. **Pendahuluan**

Berisi penjelasan mengenai latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi, dan sistematika penulisan.

2. **Tinjauan Pustaka**

Berisi landasan teori yang berkaitan dengan Tugas Akhir ini seperti data mining, twitter, dan naive bayesian classification.

3. **Perancangan Sistem**

Berisi tahap-tahap perancangan sistem yang akan digunakan dalam Tugas Akhir ini.

4. **Pengujian dan Analisis Sistem**

Berisi hasil pengujian yang dilakukan dan analisis yang didapatkan dari hasil pengujian yang telah dilakukan.

5. **Kesimpulan dan Saran**

Berisi kesimpulan yang diambil dari hasil analisis yang telah dilakukan dan saran untuk pengembangan selanjutnya.