

BAB I

PENDAHULUAN

1.1 Latar Belakang

Internet telah menjadi sarana berbagi informasi dan mengekspresikan opini dengan respon cepat oleh pengguna di media sosial (Gupta, 2012). Media sosial memungkinkan orang-orang dapat saling berinteraksi satu sama lain tanpa batasan ruang dan waktu. Pada saat ini media sosial digunakan sebagai sarana menyampaikan opini masyarakat tentang berbagai topik atau isu-isu yang sedang terjadi. Salah satu media sosial yang populer dan banyak digunakan untuk menyampaikan opini adalah Twitter. Berdasarkan data dari PT Bakrie Telcom, Indonesia memiliki 19,5 juta pengguna Twitter dari total 500 juta pengguna global. Hal ini membuat Indonesia menempati peringkat lima pengguna Twitter terbesar di dunia.

Twitter merupakan microblog yang memungkinkan penggunanya dapat mengirim dan membaca informasi ataupun opini terhadap isu-isu yang bermacam-macam dalam bentuk 280 karakter pesan twitter yang dinamakan *tweet*. Microblog adalah salah satu jenis alat komunikasi online dimana pengguna dapat memperbarui pesan. Twitter dikenal sebagai sarana penebaran informasi dan mengekspresikan opini. Contohnya yaitu terjadi peristiwa besar seperti bencana alam, pemilihan umum, pendapat tentang tokoh maupun aplikasi. Hal – hal tersebut selalu di tanggapinya oleh pengguna twitter (Gupta A. L., 2013).

Tweet yang disebar oleh pengguna twitter ini mengandung konten yang berbeda-beda sesuai dengan kebiasaan dan gaya menulis pengguna. Konten yang disebar oleh pengguna dapat dikategorikan secara umum yaitu informasi/berita,

opini, promosi, random *tweet*, obrolan, pernyataan dan *tweet* mengenai pengguna tersebut (Christopher, 2010). Pada penelitian ini opini yang disampaikan oleh pengguna twitter dapat digunakan sebagai sumber data yaitu mengenai pendapat tentang aplikasi E-money opini yang dianalisis dapat dijadikan masukan bagi perusahaan dalam menentukan pengembangan pada aplikasi tersebut.

Analisis sentiment dapat menggunakan teknik *data mining*. Yang dapat mengklasifikasikan konten opini dari sumber data yang sangat banyak. Metode text mining untuk melakukan analisis terhadap *tweet* yaitu menggunakan algoritma *naïve bayes* dan *Decision Tree* (C4.5) karena proses analisis bersifat klasifikasi. Akurasi dalam klasifikasi yang dilakukan oleh Abdan Syakuro pada tahun 2017 menggunakan algoritma naive bayes mencapai 88.8% dan pada penelitian yang dilakukan oleh Arif dan Rifqi pada tahun 2019 menggunakan algoritma c4.5 pada klasifikasi sentimen mencapai akurasi 80.00%, Kedua Algoritma tersebut memiliki tingkat akurasi yang cukup tinggi. Hal ini merupakan alasan penelitian ini dibuat, karena ingin menggabungkan output dari masing-masing algoritma untuk dianalisis tingkat akurasinya .

Algoritma C4.5 adalah salah satu sistem untuk membangun model klasifikasi yang umum digunakan (W.Xindong, 2007) . Algoritma ini sudah banyak digunakan dalam implementasi penentuan keputusan karena memiliki banyak kelebihan. Salah satu kelebihan ini misalnya dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain (J. Purnomo, 2014).

Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan (Pattekari, 2012). Selain itu mudah di pahami dan pengkodean yang sederhana.

Dalam penelitian ini pesan-pesan twitter akan dikumpulkan terlebih dahulu. Proses pengumpulan data tersebut dinamakan *crawling*. Tweet yang telah dikumpulkan akan memasuki tahap *preprocessing*. Selanjutnya tweet akan memasuki tahap pembobotan data menggunakan information gain , agar dapat dianalisis dengan menggunakan algoritma C.5 dan *Naive Bayes*. Dan hasil dari masing-masing algoritma disimpan di database, lalu dalam penelitian ini akan membandingkan kedua hasil seleksi menggunakan metode *Voting –by-majority* yaitu dengan memberikan kelas baru pada setiap kelas yang terdapat pada sebuah dataset. hasilnya menggambarkan bahwa metode *Voting –by-majority* lebih baik dari yang individu (Alaa Khudhair Abbas, 2020).

1.2 Rumusan Masalah

Adapun rumusan masalah dalam makalah ini dapat kita ambil dari latar belakang proposal ini, yaitu sebagai berikut:

- a. Bagaimana cara menyiapkan twitter sehingga diperoleh data yang siap untuk dianalisis?
- b. Bagaimana cara menggabungkan hasil klasifikasi algoritma naïve bayes dan Decision Tree (C4.5) menggunakan metode Voting –by-majority ?

- c. Bagaimana cara membandingkan akurasi algoritma dengan metode yang telah diusulkan (voting by majority)?

1.3 Batasan Masalah

Berdasarkan rumusan masalah di atas maka permasalahan yang akan di bahas akan dibatasi dengan batasan masalah berikut :

- a. Tweet yang akan dianalisis mengandung kata tentang pembayaran online atau E-money setelah itu dilakukan pengambilan data menggunakan Twitter Streaming API.
- b. Opini yang dianalisis adalah opini terhadap 3 aplikasi E-Money yaitu OVO, Dana dan Go-Pay.
- c. Metode untuk klasifikasi tweet menggunakan pohon keputusan yang dibentuk algoritma C4.5 dan Naïve Bayes lalu output masing-masing algoritma akan dibandingkan dan menghasilkan output akhir yang digunakan untuk menghitung akurasi penggabungan kedua algoritma tersebut

1.4 Tujuan

Tujuan dari tugas akhir adalah membangun suatu sistem pengolahan text mining dengan Analisis Sentimen Pengguna E-Money Pada Twitter Menggunakan Algoritma C4.5 Dan Naïve Bayes menggunakan text sebagai sumber data.

- a. Dapat mengetahui tingkat akurasi Algoritma C4.5 dan Naïve Bayes dalam mengklasifikasi tweet sentimen.

- b. Dapat menjadi referensi tambahan masyarakat terhadap e-money.

1.5 Manfaat

Manfaat dari penelitian ini adalah :

- a. Mengetahui tingkat akurasi penggabungan algoritma C4.5 dan Naïve Bayes.
- b. Membantu perusahaan agar mengetahui ulasan dari E-Money dan referensi untuk masyarakat terhadap E-Money.