

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Saat ini perkembangan *social* media semakin berkembang pesat dibuktikan dengan mudahnya akses internet menjadi salah satu alat komunikasi yang sangat populer dipengguna internet terutama generasi milenial. Dimana banyak muncul jutaan akses informasi dimana banyak orang yang bertukar pesan instan di berbagai situs sosial media terpopuler.

Dengan berkembang pesatnya internet banyak orang yang membuat postingan tentang opini dari berbagai topik dan membahas isu-isu yang terjadi saat ini ataupun keluh kesah tentang kehidupan. Kebebasan dalam membuat tautan dan kemudahan akses dari berbagai platform salah satunya *twitter* membuat pengguna internet semakin leluasa mengekspresikan sudut pandang mereka dalam tautan baik tentang politik, agama hingga trend terkini. Mengacu pada pesatnya penggunaan internet terutama pada situs media social *twitter* merupakan salah satu situs dengan pengguna aktif lebih dari 330 juta memungkinkan masyarakat menggunakan *social* media ini untuk berbagi tautan hingga lebih dari 400 juta *tweet* perhari. *Twitter* dapat menjadi sumber data opini dan sentiment masyarakat dan data tersebut dapat digunakan secara efisien untuk analisa strategi pemasaran ataupun studi sosial.

Kemudian timbul kebutuhan untuk mengatur dan mengelompokkan opini berdasarkan *sentiment* agar dapat dicari dan dinilai kualitas data dari opini publik, akan tetapi jika pengelompokkan berita dilakukan secara manual memakan waktu yang lama dan mahal. Oleh sebab itu diperlukan klasifikasi pada postingan, namun keberadaan kata berperan penting pada proses klasifikasi. Salah satu faktor yang mempengaruhi keberadaan kata yang tidak sesuai penulisannya atau tidak baku.

Analisis sentimen atau disebut juga *opinion mining* yaitu menganalisa pendapat orang, evaluasi, penilaian, sikap dan emosi terhadap suatu entitas seperti produk, jasa, pelayanan, organisasi, individu, isu-isu, peristiwa dan topik. Fokus utama dari analisis sentimen adalah menyatakan mana yang termasuk opini positif dan mana yang termasuk opini negatif. Sentimen analisa akan mengekstrak dan mengelompokkan teks yang tertaut dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat tersebut apakah bersifat positif, negative atau netral (Pang & Lee, 2008).

Salah satu contoh dari pengaplikasian analisis sentimen adalah ketika suatu perusahaan mengeluarkan suatu produk dan perusahaan tersebut menyediakan layanan untuk menerima opini dari konsumen mengenai produk tersebut. Analisis

sentimen digunakan untuk mengelompokkan opini-opini positif dan negatif dari konsumen yang menggunakan produk tersebut sehingga mempercepat dan mempermudah tugas perusahaan untuk meninjau kembali kekurangannya. Penelitian mengenai sentiment analisa sudah dilakukan oleh Davidson (2017) berupa *Automated Hate Speech Detection and Offensive Language*.

Penelitian tersebut menggunakan *Term Frequency - Index Document Frequency* (TF-IDF) untuk seleksi fitur dan menggunakan *logistic regression* sebagai pengklasifikasi dari *Tweet*. Hasil dari penelitian tersebut menghasilkan nilai *precision* 0.91 dan *recall* 0.90. Ada juga penelitian mengenai sentimen yang dilakukan oleh Waseem (2016) dengan judul *Predictive Feature for Hate Speech Detection on Twitter* dengan menggunakan n-gram untuk seleksi fitur dan *logistic regression* sebagai pengklasifikasi tweet. Badjatiya (2017) juga melakukan analisis sentimen dengan membandingkan beberapa metode klasifikasi, hasilnya metode klasifikasi yang memiliki nilai evaluasi paling tinggi dengan mengkombinasikan *Deep Neural Network* dengan *Gradient Boosted Decision Trees* (GBDT).

Metode yang digunakan dalam permasalahan ini yaitu metode *Naïve Bayes*. *Naïve Bayes* merupakan salah satu metode yang dapat membantu mengklasifikasikan untuk seleksi fitur dan klasifikasi terhadap *tweet*. Adapun manfaat dengan adanya pengaplikasian metode *Naïve Bayes* ini, yaitu mendapatkan tingkat persentase *sentiment* kalimat pada postingan twitter dan secara teknis mengetahui kalimat *sentiment* pada penyingkatan kata setiap postingan menggunakan metode klasifikasi *naïve bayes* kedalam aplikasi. Berdasarkan latar belakang yang telah didefinisikan, maka judul untuk penelitian ini yaitu **“PENERAPAN METODE NAÏVE BAYES UNTUK MENENTUKAN KLASIFIKASI SENTIMENT PADA POSTINGAN TWITTER”**.

B. Permasalahan

Proses penggunaan *text analytics* untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada *platform* tersebut. Setiap hari, internet dibanjiri oleh miliaran data dari berbagai sumber. *Sentiment analysis* berperan sebagai alat yang dapat menghubungkan seluruh data tersebut. Dengan begitu, perusahaan dapat memperoleh masukan inti dari pengguna atau konsumen secara efisien untuk mengenali dan mengekstraksi opini dalam bentuk teks. Informasi berbentuk teks saat di internet dalam format forum, blog, media sosial, serta situs berisi *review*. Dengan bantuan *sentiment analysis*, informasi yang tadinya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur. Data

tersebut dapat menjelaskan opini masyarakat mengenai produk, merk, layanan, politik, atau topik lainnya. Perusahaan, pemerintah, maupun bidang lainnya kemudian memanfaatkan data-data tersebut untuk membuat analisis *marketing*, *review* produk, umpan-balik produk, dan layanan masyarakat. Bahkan ketika membuka atau melihat suatu opini dari pengguna satu layanan pada postingan *twitter* seringkali mendapati kalimat sentimen suatu postingan tersebut. Opini tersebut berisikan kata yang tidak baku dan tidak beraturan sehingga mengakibatkan terjadi kelemahan pada proses *text mining*, dimana dapat menyulitkan fitur yang diambil serta mengurangi persentase sentiment kalimat pada postingan *twitter*, hal ini mempengaruhi hasil pada saat pengklasifikasian data.. Maka diidentifikasi masalah, yaitu:

1. Identifikasi Masalah

Berdasarkan permasalahan diatas maka dapat diidentifikasi masalah yaitu:

- a. Tingkat persentase sentiment kalimat pada postingan suatu layanan atau produk sebuah perusahaan mempengaruhi terhadap tingkat produktifitas layanan atau produk perusahaan.
- b. Sulit untuk mengetahui persentase pada kalimat *sentimen* berdasarkan *tweet* postingan.

2. Rumusan Masalah

Berdasarkan identifikasi masalah diatas maka dapat ditetapkan bahwa masalah dari penelitian ini adalah kurang akurat dan kurang efektif dalam menganalisa dan mengklasifikasikan sebuah postingan pada twitter baik postingan maupun tagar terhadap produk layanan. Pertanyaan penelitian yang dapat diajukan adalah

- a. Bagaimana penerapan metode naïve bayes dalam pengklasifikasian sentimen pada postingan twitter.
- b. Seberapa tingkat kelayakan penerapan metode naïve bayes dalam pengklasifikasian postingan pada twitter.

C. Maksud Dan Tujuan Penelitian

1. Menerapkan algoritma pada proses pengklasifikasian text minning untuk mempermudah klasifikasi dari postingan twitter.
2. Tujuan :
 - a. Mengembangkan aplikasi untuk mengetahui tingkat persentase sentiment pada postingan *twitter*.

- b. Menerapkan metode naïve bayes untuk mengklasifikasikan kalimat *sentimen* pada setiap postingan

D. Spesifikasi Hasil Yang Diharapkan

Adanya metode yang dapat diterapkan dalam sebuah proses pengklasifikasian data yang dapat memperoleh hasil klasifikasi tersebut secara cepat. Berupa proses klasifikasi dengan menggunakan algoritma pada proses data mining untuk memudahkan proses pengklasifikasian dan pelabelan kata.

E. Signifikansi Penelitian

Adapun manfaat yang akan dihasilkan dari penelitian ini adalah:

1. Manfaat teoritis; untuk memberikan sumbangan pengetahuan terhadap penelitian khususnya klasifikasi *data mining* menggunakan metode naïve bayes
2. Manfaat praktis; secara teknis mengetahui hasil klasifikasi menggunakan metode klasifikasi naïve bayes kedalam aplikasi.
3. Manfaat kebijakan; penerapan Metode *Naive Bayes* mampu menjadi alat untuk penklasifikasian *sentiment* terhadap postingan di *twitter*.

F. Asumsi Dan Keterbatasan

Kondisi awal yang dapat dijadikan asumsi dari penelitian ini yaitu:

1. Adanya postingan yang selalu mengarah pada kalimat sentiment negatif pada postingan di *twitter*.
2. Adanya kebebasan internet dan mengekspresikan opini pada media social *twitter*.

Penelitian ini memiliki keterbatasan pengembangan yaitu:

1. Penelitian ini hanya digunakan untuk mengklasifikasikan sentiment pada postingan berbahasa indonesia.
2. Dalam perancangan dan pengembangan pada penelitian ini hanya menggunakan bahasa pemrograman php untuk proses *text mining* dan pengolahan pengklasifikasian data.

G. Definisi Istilah Dan Definisi Operasional

Adapun istilah-istilah yang digunakan pada penelitian ini adalah sebagai berikut:

1. Media social:
Media *online* (daring) yang dimanfaatkan sebagai sarana pergaulan sosial secara online di internet. Di media sosial, para penggunanya dapat saling berkomunikasi, berinteraksi, berbagi, *networking*, dan berbagai kegiatan lainnya.
2. Data minning
Sebuah aktifitas dan bukanlah sebuah algoritma atau program. Dalam pelaksanaan aktifitas Data Mining maka seringkali digunakan berbagai teknik ataupun algoritma yang berasal dari berbagai disiplin ilmu misalnya statistik, artificial intelligence ataupun machine learning.
3. Text minning
Proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar sumber data tak terstruktur.
4. Twitter
Sebuah layanan jejaring sosial (media sosial) dan juga mikroblog yang memungkinkan penggunanya berkirim dan membaca pesan yang tidak lebih dari 280 karakter yang disebut sebagai tweet. Sebelumnya, pesan di Twitter hanya sampai 140 karakter tetapi pada tanggal 7 November 2017 ditambah menjadi 280 karakter.
5. Tweet
Adalah apa yang kita tulis atau posting di dalam twitter (layaknya status dalam facebook).
6. Sentimen
Pendapat atau pandangan yang didasarkan pada perasaan yang terhadap sesuatu produk jasa atau layanan tertentu.

[Halaman sengaja dikosongkan]