

BAB I

PENDAHULUAN

1.1. Latar Belakang

Sentimen analisis adalah proses untuk mendapatkan berbagai sumber data dari internet melalui berbagai platform media sosial menggunakan analisis teks. Tujuannya adalah untuk mengumpulkan pendapat dari pengguna di platform sosial media. Sentimen analisis merupakan bagian dari *text mining*, data kumpulan sentimen yang akan dianalisis adalah data berupa teks yang dapat diambil dari kolom komentar [1]. Setiap hari internet dibanjiri miliaran data dari berbagai sumber. Analisis sentimen bertindak sebagai alat yang dapat menghubungkan data apa pun. Hal ini memungkinkan bisnis untuk secara efisien mendapatkan masukan inti dari pengguna atau konsumen.

Twitter merupakan salah satu media sosial yang sangat populer dikalangan pengguna internet, hal ini dikarenakan kesederhanaan dan kemudahan dalam penggunaannya, serta pengguna dapat dengan bebas mengeluarkan pendapat atau opini mereka [2]. *Twitter* di luncurkan pertama kali pada tahun 2006, awal kemunculan *Twitter* masyarakat merespon cukup baik karena fasilitas yang ada pada *Twitter* terbilang cukup segar pada waktu itu yakni fitur *following* yang memungkinkan pengguna dapat menentukan pertemanan di jejaring tersebut.

Naive Bayes adalah klasifikasi dengan metode *probabilistic* dan statistik diusulkan oleh ilmuwan Inggris Thomas Bayes, *Naive Bayes* menghitung untuk setiap kelas keputusan probabilitas yang memiliki syarat bahwa kelas keputusan benar, sesuai vektor dari informasi objek. Algoritma *Naive Bayes* memperkirakan

peluang masa depan sesuai *experience* seseorang pada masa lalu hal tersebut dikenal dengan *Teorema Bayes*. Ciri-ciri *Naïve Bayes Classifier* ini adalah praduga yang amat kuat (naïf) akan inden pada tiap-tiap situasi kondisi [1].

Investasi adalah informasi tentang bagaimana menggunakan sebagian dana atau sumber daya yang Anda miliki untuk mendapatkan keuntungan di masa depan [3]. Umumnya masyarakat melakukan investasi untuk mendapatkan keuntungan yang cukup dan sesuai dengan yang diharapkan, di era yang serba digital ini investasi dapat dilakukan dimana saja dan kapan saja melalui beberapa aplikasi investasi yang sudah resmi atau terdaftar Otoritas Jasa Keuangan (OJK). Bentuk dari investasi pun beragam mulai dari saham, obligasi, pasar uang dan yang sedang berkembang pesat saat ini adalah *Cryptocurrency* seperti *bitcoin*, *Ethereum*, *binance* dan masih banyak lagi.

Media sosial merupakan salah satu pangsa pasar yang memungkinkan masyarakat mengkonsumsi berita atau promosi dari suatu layanan investasi, karena data sumber tersebut sangat melimpah dan *update* tiap detiknya di internet dan hal tersebut menjadikan media sosial rentan berisi berita atau informasi suatu platform investasi dan bentuk investasi yang ilegal atau disebut investasi bodong, sehingga untuk mengatasi hal tersebut diperlukan “Sentimen Analisis Thread Investasi Pada *Twitter* Menggunakan *Naive Bayes*” hasil yang diperoleh berupa komentar, kritik dan saran baik positif ataupun negative pada suatu bentuk, model produk atau aplikasi investasi yang beredar dan di bahas pada *Twitter*. Diharapkan mampu menjadi gambaran pada masyarakat mengenai bentuk, model produk, atau aplikasi investasi yang aman dan terpercaya.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, maka perlu dirumuskan suatu masalah yang akan diselesaikan pada penelitian ini:

1. Berapa akurasi metode *Naïve Bayes* dalam analisis sentimen pada *Twitter* mengenai investasi.
2. Bagaimana klasifikasi opini di media sosial (*Twitter*) sehingga dapat dilakukan evaluasi terhadap *thread* investasi untuk memperoleh *feedback* positif dan negatif.

1.3. Batasan Masalah

Untuk menghindari pembahasan yang diluar penelitian utama maka dibuatlah batasan masalah sebagai berikut:

1. Algoritma yang digunakan adalah *naïve bayes* dan tidak membandingkan atau menggunakan algoritma lain.
2. Data penelitian adalah *tweets* dari berbagai sumber yang membahas mengenai investasi *cryptocurrency* dan berbahasa Indonesia dan bahasa Inggris.

1.4. Maksud dan Tujuan Penelitian

Maksud dan tujuan dari penelitian yang dilakukan adalah:

1. Mengetahui akurasi metode *naïve bayes* dalam analisis sentimen pada *Twitter* mengenai investasi.
2. Mengklasifikasi opini di media sosial (*Twitter*) sehingga dapat dilakukan evaluasi terhadap *thread* investasi untuk memberikan gambaran masyarakat mengenai investasi yang aman dan terpercaya.

1.5. Manfaat Penelitian

Berdasarkan perumusan masalah yang sudah dijabarkan maka manfaat yang diharapkan dari penelitian ini adalah:

1. Manfaat Teoritis
 - a. Menambah ilmu pengetahuan dalam bidang analisis sentimen dan pengelolaan data.
 - b. Menambah ilmu pengetahuan tentang *thread* investasi yang beredar di media sosial *Twitter*.
 - c. Menambah wawasan tentang system analisis data terutama sentimen analisis menggunakan *naïve bayes*.
2. Manfaat Aplikatif
 - a. Sebagai salah satu syarat untuk menyelesaikan studi pada program studi SI Informatika, UNIVERSITAS AMIKOM YOGYAKARTA.
 - b. Memberikan aplikasi tentang *thread* investasi dengan analisis sentiment.
 - c. Memperoleh hasil analisis sentiment terhadap *thread* investasi menggunakan metode *naïve bayes*.

1.6. Metode Penelitian

Untuk melakukan penelitian dibutuhkan beberapa cara untuk mendapatkan hasil penelitian yang sesuai harapan, berikut metode penelitian yang digunakan:

1.6.1 Metode Pengumpulan Data

1.6.1.1 Metode Studi Pustaka

Studi Pustaka merupakan metode pengumpulan data menggunakan bantuan jurnal-jurnal atau sumber literatur lain seperti sumber dari internet dan buku yang berkaitan dengan penelitian. Metode ini dapat dijadikan langkah awal untuk menentukan tema penelitian atau objek yang akan diteliti. Peneliti mencari sumber-sumber baik berupa jurnal dan informasi dari internet yang membahas mengenai sentiment analisis serta metode *naïve bayes*.

1.6.1.2 Metode Observasi

Observasi merupakan metode pengumpulan data dengan cara mengumpulkan data secara langsung pada objek yang akan diteliti. Peneliti melakukan observasi pada *thread* investasi yang ada di platform sosial media *Twitter*, mencari dan melihat pada kolom komentar berupa kritik saran baik positif maupun negatif, dan melihat *reply* dari *tweet* tersebut sehingga mendapatkan data komentar yang lebih banyak. Data yang sudah diperoleh di tabulasi dan diberikan label positif atau negatif pada masing-masing komentar.

1.6.2 Metode Analisis

Metode analisis menggunakan metode *naïve bayes* yang dibantu dengan bahasa pemrograman *Python* untuk memperoleh data sentiment analisis yang diinginkan. Berikut alur dari metode *naïve bayes*:

1. *Read Training Data*

Dimulai dari membaca data *training* yang diperoleh dan diolah, data *training* didapatkan pada pengumpulan data komentar dari *Twitter* mengenai *thread* investasi.

2. Hitung Jumlah dan Probabilitas

Jumlah dan probabilitas akan terdata pada tabel probabilitas, apabila data numerik maka urutannya adalah mencari mean tiap parameter, kemudian standar deviasi tiap parameter, dan didata pada tabel mean dan standar deviasi

3. Mendapatkan nilai mean

Nilai mean didapat apabila data numerik, apabila bukan numerik maka akan melewati tahap ini dan langsung mendapat hasil solusi.

4. Perolehan Data

Data yang diperoleh berupa kritik dan saran secara positif dan negatif.

1.6.3 Metode Perancangan

Metode perancangan menggunakan pemrograman *Python* dengan IDE *Jupyter Notebook*. Berikut *classification* yang akan digunakan pada sentiment analisis penelitian ini.



```

import Perancangan
...
# Download Twitter dataset
...
# Classify the tweets using Naive Bayes
...

```

Gambar 1. 1. Twitter Classification

1. Langkah Pertama Persiapan

Persiapan dengan *import package* yang diperlukan yakni NLTK dan *NumPy* serta mengambil *datasheet Twitter* yang telah dikumpulkan dengan cara *crawling tweeter data* secara manual.

2. *Pre-process Tweets*

Preprocess Tweets diperlukan untuk mempersiapkan *tweet* yang telah di data, dengan mengimpor kan *package re* (*regular expression operations*), *string* (*string operation*).

3. *Tokenize The String*

Pada tahap ini *string* atau tulisan dari data yang telah di *import* dan di persiapan maka dilakukan *tokenizes string* yang bertujuan untuk memisahkan kata pada kalimat, dan kata tersebut sebagai hasil ekstraksi.

4. *Remove stop words and punctuations*

Tahap ini untuk menghapus kata-kata imbuhan yang tidak penting atau tidak digunakan seperti 'saya', 'kamu', atau 'hahaha' dan lain sebagainya.

5. *Stemming*

Tahap ini mengubah kata ke bentuk yang paling umum missal 'belajar menjadi belajar', 'pelajaran menjadi belajar', 'dipelajari-saja menjadi belajar' dan sebagainya.

6. Membuat Kamus Frekuensi

Kamus frekuensi dibuat untuk menyimpan kata atau *string* yang digunakan pada saat pemrosesan dan pengujian metode.

7. Pembagian data (*split data*)

Pembagian data menjadi dua bagian *training* dan *testing*. Tahap ini termasuk pengujian data yang telah diproses.

8. Penerapan Metode *Naïve Bayes*

Tahap ini adalah tahap akhir dimana setelah kamus frekuensi kata terkumpul maka dilakukan penerapan metode *naïve bayes* dengan menghitung jumlah dan probabilitas kemudian mendapatkan nilai akurasi serta solusi dari evaluasi yang dibutuhkan.

1.7. Sistematika Penulisan

Pada bagian ini dituliskan urutan dan sistematika penulisan yang dilakukan.

BAB I

PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan, manfaat penelitian, metode penelitian dan sistematika penulisan

BAB II

LANDASAN TEORI

Bab ini terdiri dari kajian pustaka yang berisi penelitian sebelumnya serta sumber-sumber yang berisi teori dasar dan teori kutipan yang berkaitan dengan penelitian.

BAB III

METODE PENELITIAN

Bab ini terdiri dari alat dan bahan yang digunakan pada penelitian berkaitan dengan kebutuhan *hardware* dan *software* serta alur penelitian yang berisi alur perancangan pada penelitian.

BAB IV

HASIL DAN PEMBAHASAN

Bab ini terdiri dari hasil penelitian dan pembahasan bab ini merupakan inti dari penelitian dari penyajian data, analisis data, penggunaan *tools*, dan penerapan metode yang digunakan.

BAB V

PENUTUP

Bab ini terdiri dari kesimpulan dan saran dari penelitian yang dilakukan, kesimpulan penelitian merupakan penyajian hasil penelitian sedangkan saran merupakan pemberian solusi yang diperlukan guna menyempurnakan penelitian tersebut.

