

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pantai Talang Siring merupakan salah satu pantai yang berada di pulau Madura di kabupaten Pamekasan. Pada kabupaten Pamekasan, kawasan pariwisata pantai Talang Siring menjadi tempat favorit untuk didatangi oleh wisatawan, dengan data pengunjung tiap tahun mengalami ketidakstabilan pada pantai Talang Siring [1]. Twitter merupakan sosial media yang cukup sering digunakan oleh kaum milenial untuk mengungkapkan sesuatu atau informasi pada objek tertentu, media twitter berfokus pada tampilan halaman website yaitu teks, dengan begitu dapat memudahkan untuk pengambilan sebuah informasi.

Analisis Sentimen merupakan teknik untuk mengekstrak data dan mengolah secara otomatis dengan tujuan mendapatkan berbagai informasi dari suatu kalimat sentimen atau pendapat seseorang[2].Proses pengumpulan data tersebut menggunakan metode yang dinamakan *crawling*. *Crawling* adalah proses mengumpulkan informasi pada halaman website secara otomatis.

Metode yang digunakan untuk penelitian yaitu metode *Decision Tree*, metode *Decision Tree* merupakan proses klasifikasi menggunakan struktur pohon sebagai faktor membuat keputusan untuk memecahkan suatu masalah, salah satu algoritma *decision tree* adalah *C4.5*. Pada penelitian sebelumnya oleh Yunita, dkk[3]. Penelitian tersebut menggunakan perbandingan metode *C4.5 Decision Tree* dan *Naïve Bayes* untuk mendapatkan nilai tingkat akurasi yang lebih baik. Data yang digunakan sebanyak 1000 responden dengan dikurangi sebanyak 120 data, untuk data latih dengan sedikit manipulasi pada data tersebut. Hasil dari penelitian tersebut untuk metode *C4.5 Decision Tree* mendapatkan nilai *accuracy* sebesar 94,17%, lebih besar dibanding *Naïve Bayes* dengan nilai sebesar 85,83%.

Penelitian yang dilakukan oleh Fajriansyah sebelumnya [4]. Pada penelitian tersebut menggunakan metode klasifikasi menggunakan algoritma *C4.5 Decision Tree*. Penelitian tersebut menggunakan data berasal dari *twitter* dengan jumlah data berjumlah 150, 100 data latih dan 50 data uji, penelitian dilakukan dengan berbagai pengujian sebanyak tiga kali iterasi dan didapatkan hasil rata rata *error* sebesar 36%.

Penelitian sebelumnya oleh Syarifuddin [5]. Penelitian tersebut dilakukan menggunakan metode *Decision Tree*, *KNN* dan *Naïve Bayes*. Peneliti tersebut menggunakan data berasal dari twitter dengan jumlah data berjumlah 170 *tweet*, dengan menggunakan *cross validation* pengujian dilakukan sebanyak 10 kali guna evaluasi hasil analisis dan klasifikasi yang lebih baik. Hasil dari penelitian tersebut didapatkan hasil nilai *Decision* mendapatkan nilai tertinggi dengan *accuracy* sebesar 83,3%, *precision* 79%, *recall* 87,17% dan *K-NN* mendapatkan *accuracy* 80,80%, *precision* 80%, *recall* 74,41% sedangkan *Naïve Bayes* mendapatkan *accuracy* 80,03%, *precision* 86%, *recall* 62,71%.

Dengan hasil beberapa penelitian sebelumnya didapatkan bahwa *Decision Tree* metode *C4.5* memiliki hasil yang lebih baik dan akurat, oleh karena itu *algoritma C4.5 Decision Tree* dipilih sebagai metode penelitian dengan target Pantai Talang Siring, dengan judul Penerapan Metode Algoritma *C4.5 Decision Tree* untuk Analisis Sentimen pengguna Twitter terhadap Pantai Talang Siring di Pulau Madura.

1.2 Rumusan Masalah

Berdasarkan pada uraian latar belakang maka dapat dipaparkan rumusan masalah sebagai berikut:

1.2.1 Permasalahan

Menganalisis *tweet* pada *twitter* untuk mengetahui *tweet* masyarakat terhadap Pantai Talang Siring di sosial media *twitter* berupa sentimen *positif negatif*, dan *netral*.

1.2.2 Solusi Permasalahan

Berdasarkan masalah tersebut, peneliti menggunakan metode klasifikasi *algoritma C4.5 Decision Tree* pada klasifikasi komentar dengan harapan mendapatkan nilai akurasi yang optimal terhadap klasifikasi analisis sentimen sosial media *twitter* terhadap Pariwisata di Pantai Talang Siring Pulau Madura.

1.2.3 Pertanyaan Penelitian

1. Berapa nilai akurasi yang diperoleh dengan melakukan percobaan parameter *max_depth* yang berbeda pada *Decision Tree C4.5*?
2. Bagaimana pengaruh jumlah pembagian data dengan *ratio* yang berbeda terhadap nilai akurasi yang didapat ?

1.3 Batasan Masalah

Dari pembahasan sebelumnya, penulis mendapatkan hasil untuk di gunakan ruang lingkup penelitian sehingga menghasilkan output yang diinginkan dan optimal dengan Batasan yang diterapkan sebagai berikut :

1. Data yang digunakan penelitian hanya berasal dari *twitter*.
2. Data yang diperoleh berasal dari *tweet* pada Pariwisata di Pantai Talang Siring Pulau Madura.
3. Algoritma yang digunakan untuk pengklasifikasian adalah *C4.5 Decision Tree*.
4. Menghilangkan semua emoticon pada setiap kalimat *tweet*.
5. *Tweet* yang diambil hanya menggunakan bahasa Indonesia.
6. Sistem dirancang menggunakan bahasa pemrograman *python* dengan bantuan *library nltk* dan *sastrawi*.
7. Data *crawling twitter* menggunakan *library snsrape*,
8. Kamus *spelling corection* berasal dari *github*.
9. Klasifikasi pohon keputusan dibangun menggunakan *library* dari *skitch learn decisiontreeClassifier*.
10. Langkah *stopwords* menggunakan *library* dari *nltk* berbahasa Indonesia, pada penelitian ini memodifikasi list *stopwords* pada *library nltk* sesuai dengan kasus yang dilakukan.

1.4 Manfaat dan Tujuan

1.4.1 Tujuan

Berdasarkan rumusan masalah tersebut, tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui nilai akurasi yang diperoleh ketika melakukan skenario yang berbeda pada *max_depth C4.5 Decision Tree*.
2. Untuk mengetahui nilai akurasi yang didapat dari pembagian data dengan ratio yang berbeda.

1.4.2 Manfaat

Diharapkan memberikan manfaat yang berguna bagi masyarakat tentang gambaran sentimen terhadap Pariwisata di pantai Talang Siring pulau Madura.

1.5 Sistematika Proposal

Sistematika penulisan proposal skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas latar belakang mengenai ruang lingkup masalah yang dibahas, solusi yang diusulkan untuk menyelesaikan permasalahan, tujuan serta manfaat dengan diterapkannya solusi, dan sistematika penulisan proposal.

BAB II KAJIAN PUSTAKA

Bab ini akan menjelaskan mengenai materi dan metode yang dijadikan landasan teori dalam penelitian ini. Diantaranya materi yang dibahas adalah algoritma Decision Tree C4.5, dan penelitian sebelumnya terkait analisis sentimen.

BAB III USULAN METODE

Pada bab ini dijelaskan tentang dataset yang akan digunakan, alur sistem, tahapan pengujian, skenario uji coba dan jadwal pengerjaan.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan tentang pembahasan penelitian terkait dengan menjelaskan penelitian berlangsung dan mendapatkan hasil pada penelitian,

BAB V PENUTUP

Pada bab ini dijelaskan tentang kesimpulan pada penelitian yang menghasilkan beberapa kesimpulan pada penelitian ini dan saran yang dapat dilakukan dan dikembangkan untuk penelitian kali ini.