

BAB 2

LANDASAN TEORI

2.1 Sentiment Analysis

Sentimen analisis adalah suatu metode analisis yang digunakan untuk mengevaluasi sentimen di balik teks yang ditulis oleh pengguna. Sentimen yang dianalisis dapat berupa sentimen positif, negatif, atau netral. [4]. Sentiment analysis dapat dilakukan pada berbagai jenis teks, seperti ulasan produk, tweet, artikel berita, atau posting media sosial. Tujuan dari sentiment analysis adalah untuk memahami dan menganalisis perasaan dan pandangan orang terhadap suatu topik, merek, atau produk. Menurut Clayton dalam penelitiannya Sentimental Analysis di bagi dapat dibedakan menjadi dua [5]. Yakni Sentimental analysis dokumen dan data dimana Berdasarkan Level sumber data, Sentimen analisis dapat di bagi menjadi dua yaitu *Coarse-grained Sentiment Analysis* dan *Fined-grained Sentiment Analysis* [6]:

1. Coarse-grained Sentiment Analysis

Coarse-grained Sentiment Analysis adalah teknik analisis sentimen yang membagi sentimen ke dalam beberapa kategori atau level yang lebih umum. Biasanya, ada tiga kategori umum yang digunakan dalam *coarse-grained sentiment analysis*: positif, negatif, dan netral. Dalam melakukan analisis terkait opini publik terhadap suatu product biasanya *Coarse grained* biasa di gunakan suatu perusahaan untuk mengetahui positif atau negatifnya respon publik terhadap product atau layanan perusahaan.

2. Fined-grained Sentiment Analysis

berbeda dengan *coarse-grained*, *Fined-grained Sentiment Analysis* adalah teknik analisis sentimen yang lebih detail seperti marah, senang, sedih, kecewa dan lainnya. Analisis menggunakan *Fined-grained Sentiment* relatif digunakan saat melakukan analisis di sosial media survei atau data teks yang mampu direpresentasikan sebagai sebuah respon konsumen terhadap suatu produk yang dikeluarkan oleh sebuah perusahaan

2.2 Intent Sentiment Analysis

Intent Sentiment Analysis adalah suatu metode sentimen analisis yang digunakan untuk mengidentifikasi sentimen di balik intent atau tujuan di balik

teks yang ditulis oleh pengguna[7]. Dalam pengolahan bahasa alami, intent merujuk pada tujuan atau niat dari teks yang ditulis. Selain itu, pendekatan ini juga mempertimbangkan tujuan atau niat penulis dalam menulis teks, seperti apakah mereka bermaksud memberikan ulasan atau komentar positif atau negatif. Terdapat beberapa langkah yang umumnya dilakukan dalam melakukan intent sentiment analysis, yakni :

1. *Preprocessing*

tahapan ini adalah tahap awal dimana data yang sudah ada di bersihkan seperti menghapuskan karakter khusus huruf besar kecil yang tidak rapih dan lain-lain. Dalam Tahapan ini akan dilakukan Data Cleaning, Data Integration Transformasi Data, Mengurangi Data.[8]

2. *Feature Selection*

dalam proses ini mulai melakukan ekstrak fitur atau ciri-ciri yang dimaksud atau ingin di identifikasi. Proses ini berfungsi menerapkan efisiensi dalam proses klasifikasi dengan mengatur jumlah fitur yang sudah diberikan bobot[9].

3. *Model Training*

setelah fitur ditentukan, selanjutnya model *mechine Learning* atau *Deep Learning* dilatih untuk mengenali pola yang sudah di tentukan.

4. *Testing and Evaluation*

Setelah proses *Model Training* selesai di lakukan, maka data pengujian dapat digunakan untuk mengevaluasi hasil model yang sudah di buat.

Untuk mendukung penggunaan intent sentiment analisis algoritma Naive Bayes Classification dapat digunakan untuk membantu dalam mengklasifikasikan data yang ingin digunakan.

2.3 Naive Bayes

Naive Bayes merupakan algoritma yang digunakan dalam machine learning untuk mengklasifikasikan suatu data probabilitas. dalam penerapannya algoritma ini digunakan untuk memprediksi kemunculan kelas target dari suatu data. Cara kerja dari *Naive Bayes* sendiri dengan menghitung probabilitas kelas target pada suatu

data. Salah satu pendekatan dari *Naive Bayes* adalah *Multinomial Naive Bayes* dimana Algoritma dari *Multinomial Naive Bayes (MNB)* digunakan untuk data yang berbentuk data teks dengan menghitung probabilitas untuk memprediksi suatu kelas[10]. Adapun Perhitungan yang digunakan adalah

$$P(C) = \frac{\text{Count}(c) + K}{N + K|\text{Classes}|} \quad (2.1)$$

Dimana :

1. P = merupakan Probabilitas dari Variable C
2. Count (c) = Jumlah kemunculan dari nilai sample c
3. K = Nilai Parameter
4. N = jumlah total dari nilai sample c
5. Classes = Jumlah kelas pada Sample

Setelah melakukan perhitungan multinomial dilanjutkan dengan menghitung nilai atribut. Jumlah atribut yang sudah diberi kelas ditambah satu dan dibagi dengan jumlah kelas tertentu ditambah dengan hasil kali antara jumlah atribut dikali satu[11].

$$P(W|C) = \frac{\text{Count}(c) + K}{\text{Count}(C)|V|.1} \quad (2.2)$$

Dimana :

1. Count = Jumlah munculnya atribut pada suatu kelas
2. Count (c) = Jumlah kemunculan dari nilai sample c
3. K = Nilai Parameter
4. N = jumlah total dari nilai sample c
5. V = Jumlah atribut pada Sample