

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Beberapa jurnal dan penelitian berkaitan dengan analisis sentimen adalah sebagai berikut:

I Gede Suardika (2014) melakukan analisis sentimen dan analisis korelasi keramahan di Bali. Dataset yang digunakan adalah data hotel yang terpilih berdasarkan peringkat di Tripadvisor. Data diambil dengan mengambil penilaian untuk hotel dari area Jimbaran, Kuta, Nusa Dua, dan Seminyak. Sebagai representasi dari setiap area, dipilih satu hotel dengan penilaian tertinggi dan terendah. Sistem analisis yang dibangun menggunakan bahasa pemrograman PHP, database MySQL dan menggunakan metode klasifikasi Naive Bayes dengan akurasi 81%. Hasil dari analisis korelasi membuktikan bahwa hotel dengan penilaian lebih rendah di Tripadvisor mempunyai persentase komentar negatif lebih banyak.

Boy Utomo Manalu (2014) menggunakan 3300 data twitter tentang sentimen terhadap provider telekomunikasi. Data tersebut diklasifikasi secara manual dan dibagi ke dalam masing-masing 1000 data untuk sentimen positif, negatif dan netral. Kemudian 300 data digunakan untuk testing, dimana tiap jenis sentimen berjumlah 100 tweet. Hasil penelitian ini menghasilkan sebuah sistem yang dapat mengklasifikasi sentimen secara otomatis dengan hasil pengujian 100 tweet mencapai 93 % dengan 2700 data training.

Falahah dan Dyar Dwiki A N (2015) melakukan klasifikasi kata-kata dan difokuskan pada tweet dalam bahasa Indonesia. Aplikasi ini kemudian diterapkan untuk mengklasifikasikan opini publik pada Twitter terkait layanan pemerintah terhadap masyarakat, berdasarkan sentimen positif, negatif atau netral. Data latih diperoleh melalui aplikasi platform KNIME Analytic dan sumber text diperoleh dari akun Twitter Dinas pemerintah Kota Bandung. Proses klasifikasi dilakukan melalui serangkaian tahapan seperti praproses (case folding, parsing, dan transformasi) serta proses klasifikasi itu sendiri.

Rizal Setya Perdana dan Mochamad Ali Fauzi (2017) melakukan analisis sentimen terhadap tayangan berdasarkan opini masyarakat pada media sosial Twitter menggunakan metode K-Nearest Neighbor dan pembobotan jumlah retweet. Data yang digunakan berupa opini masyarakat terhadap tayangan televisi pada twitter sejumlah 400 data. Dari hasil pengujian akurasi dengan menggunakan pembobotan tekstual diperoleh 82, 50%, menggunakan pembobotan non-tekstual 60%, dan menggunakan penggabungan keduanya 83,33% dengan nilai $k=3$ dan konstanta perkalian yang tepat $\alpha=0,8$ dan $\beta=0,2$.

Nuke Y. A. Faradillah (2016) melakukan eksperimen sistem klasifikasi analisis sentimen Twitter pada akun resmi pemerintah Kota Surabaya berbasis pembelajaran mesin. Pengklasifikasian dilakukan dengan menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM). Data diambil dari twitter @e100ss dan @SapawargaSby selama periode 1 September 2015 sampai dengan 13 Oktober 2015, dimana akan dikelompokkan menjadi 3 kelas sentimen yaitu positif, negatif, dan netral. Model klasifikasi terbaik didapatkan dengan

menggunakan algoritma Support Vector Machine (SVM) dengan hasil akurasi sebesar 79,81%.

Prima Arfianda Putri dkk melakukan implementasi metode improved k-nearest neighbor pada analisis sentimen twitter berbahasa Indonesia. Hasil yang telah diperoleh melalui implementasi dan pengujian sistem adalah jumlah data latih, keseimbangan proporsi kategori data latih, dan nilai k berpengaruh terhadap ketepatan hasil analisis sentimen. Rata-rata precision yang diperoleh sistem sebesar 82%, rata-rata recall sebesar 87%, dan rata-rata F-measure sebesar 84%, sehingga dapat disimpulkan efektivitas sistem sudah berjalan dengan relatif baik.

Robert Habibi dkk (2016) melakukan analisis sentimen pada Twitter mahasiswa menggunakan metode backpropagation. Hasil klasifikasi di tes menggunakan WEKA dengan pengklasifikasi multilayer perceptron. Mengambil data kuesioner mahasiswa ATMAJAYA lalu memilih 30 responden yang memiliki akun Twitter yang tidak dalam mode protected dengan tweet lebih dari 200. Jumlah total data tweet dari 300 responden adalah 6.000 tweet. 25% digunakan sebagai data training. Hasil analisis sentimennya adalah adalah tendensi positif 33.33%, netral 53.33% dan negatif 13.33%.

Yane Marita Febrianti dkk melakukan analisis sentimen pada ulasan "Lazada" berbahasa Indonesia menggunakan K-Nearest Neighbor (K-NN) dengan perbaikan kata menggunakan Jaro Winkler Distance. Pengujian ini menggunakan data latih secara bertahap dari 250 data, 300 data, 350 data, 400 data, hingga 450 data dengan setiap data latih yang diberikan diujikan dengan 50 data uji. Hasil

pengujian dengan nilai accuracy yang baik terdapat pada analisis sentimen dengan perbaikan kata yakni 76 %, dengan nilai precision 0.76, dengan nilai recall 1.

Dalam penelitian ini, peneliti akan mengkasifikasi tweet dengan hashtag #Malioboro yang diambil dengan API dari media sosial Twitter ke dalam positif, netral, atau negatif. Metode yang digunakan dalam pengklasifikasian adalah k-Nearest Neighbour. Data twitter yang akan diambil adalah tweet yang mengandung kata Malioboro dengan rentang satu tahun yaitu Juli 2017 sampai dengan Juli 2018.

Secara detail, perbedaan penelitian ini dengan penelitian yang sudah ada dapat dilihat pada tabel berikut:

Tabel 2. 1 Tinjauan Pustaka

Peneliti	Tahun	Data	Sumber	Metode	Hasil
Boy Utomo Manalu	2017	Provider telekomunikasi	Twitter	<i>Naive Bayes</i>	sentimen positif, negatif dan netral
I Gede Suardika	2014	Ulasan mengenai hotel	Trip Advisor	<i>Naive Bayes</i>	sentimen positif, negatif dan netral dengan korelasi nilai peringkat hotel
Falahah dan Dyar Dwiki A N	2015	Opini publik terhadap layanan pemerintah	Twitter	<i>KNIME</i>	sentimen positif, negatif atau netral
Nuke Y. A. F.	2016	akun resmi pemerintah Kota Surabaya	twitter @e100ss dan @Sapawar gaSby	<i>SVM</i>	sentimen positif, negatif atau netral

Prima Arfianda Putri dkk		bertema penyedia layanan GSM	Twitter	<i>Improved KNN</i>	Sentimen positif atau negatif
Rizal S. P. dan Mochamad Ali F.	2017	Opini masyarakat terhadap tayangan televisi	twitter	K-NN dan pembo botan jumlah retweet	sentimen positif, negatif atau netral
Robet Habibi	2016	Tweet mahasiswa Atmajaya Yogyakarta	Twitter	<i>Multilayer Perceptro n</i>	Emosi positif, netral dan negatif
Yane Marita Febrianti dkk	2018	Ulasan “Lazada” berbahasa Indonesia	Lazada	<i>kNN</i>	nilai <i>precission</i> 0,76, dengan nilai <i>recall</i> 1
Richa Chlarramita	2018	Malioboro	Twitter	<i>KNN</i>	sentimen positif, negatif dan netral

Penelitian yang akan diajukan adalah analisis sentimen yang berkaitan Malioboro dengan metode klasifikasi kNN dengan hasil sentimen positif, negatif dan netral.

2.2 Dasar Teori

2.1 Python

Python adalah sebuah bahasa pemrograman interpretatif tingkat tinggi yang multiguna. Python dapat berjalan pada banyak varian sistem operasi, seperti Unix, Linux, dan Windows. Python adalah bahasa berorientasi objek yang ditafsirkan yang mendukung operator overloading dan menawarkan antarmuka lintas-platform untuk fungsionalitas sistem operator (Xing Cai dkk, 2005).

2.2.1 MySQL

MySQL adalah sebuah relational database. Database relasional adalah struktur data yang memungkinkan pengguna untuk menghubungkan informasi dari berbagai 'tabel', atau berbagai jenis data yang berbeda.

MySQL adalah sistem manajemen basis data relasional open source yang populer (RDBMS) yang didistribusikan, dikembangkan, dan didukung oleh Oracle Corporation. Sistem relasional seperti, MySQL menyimpan data dalam bentuk tabel dan menggunakan bahasa query terstruktur (SQL) untuk mengakses data. Di MySQL, kita harus menentukan sebelumnya skema berdasarkan persyaratan dan mengatur aturan untuk mengontrol hubungan antar bidang dalam catatan. Di MySQL, informasi terkait dapat disimpan dalam tabel yang berbeda, tetapi mereka terkait dengan penggunaan gabungan. Dengan demikian, duplikasi data dapat diminimalkan (Dipina Damodaran, 2016).

2.2.2 Analisis Sentimen

Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau fitur/tingkat aspek dan menentukan apakah pendapat yang dikemukakan dalam dokumen, kalimat atau fitur entitas/aspek bersifat positif, negatif atau netral. Lebih lanjut analisis sentimen dapat menyatakan emosional sedih, gembira, atau marah (Liu B, 2012).

Ekspresi atau sentimen mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subject yang berbeda. Sebagai contoh, adalah hal yang baik untuk mengatakan alur film tidak terprediksi, tapi adalah hal yang tidak baik jika 'tidak terprediksi'

dinyatakan pada kemudi dari kendaraan. Bahkan pada produk tertentu, kata-kata yang sama dapat menggambarkan makna kebalikan, contoh adalah hal yang buruk untuk waktu start-up pada kamera digital jika dinyatakan “lama”, namun jika “lama” dinyatakan pada usia baterai maka akan menjadi hal positif. Oleh karena itu pada beberapa penelitian, didahului dengan menentukan elemen dari sebuah produk yang akan ditinjau sebelum memulai proses analisis sentimen.

2.2.3 K-Nearest Neighbor Classifier

K-Nearest Neighbor (k-NN) merupakan metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode Klasifikasi yang dilakukan berdasarkan pembelajaran data yang sudah terklarifikasikan sebelumnya. Termasuk dalam supervised learning, dimana hasil query instance yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam k-NN (Oliver Sutton, 2012)

Dalam penelitian ini yang menjadi data uji adalah tweet pada Twitter. Ada dua tahap pada klasifikasi tweet. Tahap pertama adalah pelatihan terhadap tweet yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi tweet yang belum diketahui kategorinya.

Dalam metode kNN setiap tweet direpresentasikan dengan pasangan atribut “ $x_1, x_2, x_3, \dots, x_n$ ” dimana x_1 adalah kata pertama, x_2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori tweet. Pada saat klasifikasi, akan dicari probabilitas tertinggi dari semua kategori dokumen yang diujikan (V_{map}).

Algoritma KNN adalah sebagai berikut:

1. Tentukan parameter K
2. Hitung jarak antara data yang akan dievaluasi dengan semua pelatihan
3. Urutkan jarak yang terbentuk (urut naik)
4. Tentukan jarak terdekat sampai urutan K
5. Pasangkan kelas yang bersesuaian
6. Cari jumlah dari kelas tetangga yang terdekat dan tetapkan kelas tersebut

sebagai kelas data yang akan dievaluasi

Sebuah fungsi jarak diperlukan untuk membandingkan kesamaan data, menggunakan rumus Euclidean

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \dots\dots\dots(2.1)$$

Keterangan:

X1 : Sampel data

X2 : Data Uji/Testing

i : variabel Data

d : Jarak

p : dimensi data.

2.2.4 API

API (Application Programming Interface) merupakan sekumpulan perintah, fungsi, dan protokol yang dapat digunakan dalam membangun perangkat lunak untuk sistem operasi tertentu, juga merupakan suatu dokumentasi yang

terdiri dari antar muka, fungsi, kelas, struktur untuk membangun software (Ade Suryansyah , 2014).

API memungkinkan aplikasi atau layanan untuk berkomunikasi dengan aplikasi atau layanan lain berdasarkan protokol untuk bertukar data baik secara internal maupun eksternal ke bisnis dan pengembang lain.

API diimplementasikan dengan menuliskan function calls di dalam program, yang memberikan koneksi kepada subroutine yang dibutuhkan untuk eksekusi. Sehingga, sebuah API menunjukkan bahwa beberapa modul program telah tersedia di dalam komputer untuk menjalankan operasi atau bahwa modul program tersebut dapat dihubungkan ke sebuah program yang telah ada untuk melakukan tugas yang telah ditentukan.

2.2.5 Twitter

Twitter adalah situs jejaring sosial yang menawarkan layanan blogging mikro untuk berinteraksi melalui posting Twitter, juga disebut tweet, pada Smartphone, laptop, iPod, dan perangkat apa pun dengan akses Internet. Twitter adalah sebuah media sosial mikroblog setiap pengguna dapat mengunggah pesan berbasis 140 karakter.

Dalam beberapa tahun terakhir, para peneliti telah secara signifikan mempelajari situs jejaring sosial online (misalnya, Facebook, Twitter, LinkedIn, dan YouTube) untuk menguji hubungan antara hasil pendidikan dan jejaring sosial (Bista, 2014; Gouseti, 2010; Kirschner & Karpinski, 2010).

Twitter dapat diakses melalui website dan aplikasi yang terinstall di perangkat mobile. Twitter dapat digunakan dengan gratis. Melalui Twitter

pengguna dapat mengetahui apa yang sedang terjadi di dunia dan apa yang sedang orang bicarakan saat ini.