

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Penelitian ini menggunakan beberapa sumber pustaka yang berhubungan dengan kasus atau metode yang akan diteliti. Diantaranya yaitu:

Ulfa Nur Fajri Maharani (2022) melakukan penelitian dengan judul Analisis sentimen terhadap kasus vaksin covid-19 menggunakan metode naïve bayes berdasarkan twitter. Dengan memanfaatkan data dari twitter, penelitian ini memberikan informasi hasil dari analisis sentimen dari data dengan kata kunci “Vaksin COVID-19”. Penelitian ini mengambil 2005 data dan menghasilkan klasifikasi sentimen analisis terhadap data uji didapatkan 100% positif dan tidak ada negatif dengan akurasi sebesar 91.272%, nilai presisi sebesar 100% dan nilai recall sebesar 91.271%.

Awaluddin (2021) melakukan penelitian dengan judul Analisis sentimen pada akun Twitter BNI Customer care menggunakan metode naïve bayes classifier (NBC). Sistem yang dibuat tidak hanya menentukan sentimen dari setiap keluhan nasabah, tetapi juga memberikan info grafik WordCloud atau kat-kata yang sering dikeluhkan oleh nasabah pada akun Twitter.

Artanti Inez Tanggraeni dan Melklor N. N. Sitokdama (2022) melakukan penelitian dengan judul Analisis Aplikasi E-Government pada google play menggunakan algoritma naïve bayes. Penelitian ini memberikan informasi bahwa dari proses pelabelan menunjukkan bahwa aplikasi Sentuk Tanahku cenderung mendapatkan respon positif dari pengguna dengan

perbandingan 407 ulasan positif dan 235 ulasan negatif. dan dari hasil pengujian analisis sentimen menggunakan algoritma naïve bayes dengan pembuatan TF-IDF menghasilkan akurasi sebesar 89%, precision sebesar 83% dan recall sebesar 87%.

Akhmad Deviyanto dan M. Didik R. Wahyudi (2018) melakukan penelitian dengan judul Penerapan Analisis Sentimen pada pengguna Twitter menggunakan metode K-Nearest Neighbor. Penelitian ini menggunakan 200 data dan menghasilkan nilai akurasi terbesar adalah 67,2% ketika k=5, presisi tertinggi 56,94% ketika k=5 dan recall 76,24% dengan k=5.

Tiara Rahmadhani (2021) melakukan penelitian dengan judul Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat Pada Media Sosial Twitter Menggunakan Metode K-Nearest Neighbor. Penelitian ini menggunakan 2047 data yang menghasilkan akurasi terbesar yaitu 72.56% dengan nilai k=3. Dari hasil analisis sentimen dari setiap stasiun televisi didapat respon positif sebesar 69.47 % untuk stasiun televisi Net Tv, 72.63% untuk RCTI, 60.53% untuk SCTV, dan 74.32% untuk Trans Tv.

Tabel 2.1 Rangkuman tinjauan pustaka

No	Peneliti	Objek	Metode	Perbedaan
1	Ulfa Nur Fajri Maharani (2022)	kasus vaksin covid-19	Naïve bayes	Perbedaan dengan penelitian saat ini adalah dari penelitian Ulfa Nur Fajri Maharani tentang kasus vaksin covid-19 dengan metode naïve bayes, sedangkan untuk penelitian saat ini menggunakan objek

				QRIS dengan metode K-Nearest Neighbor.
2	Awaluddin (2021)	BNI Customer care	naïve bayes classifier (NBC)	Perbedaan dengan penelitian saat ini adalah dari penelitian Awaluddin tentang BNI Customer care dengan metode naïve bayes, sedangkan untuk penelitian saat ini menggunakan objek QRIS dengan metode K-Nearest Neighbor.
3	Artanti Inez Tanggraeni, Melkior N. N. Sitokdana (2022)	Aplikasi E-Government	Naïve bayes	Perbedaan dengan penelitian saat ini adalah dari penelitian Artanti Inez Tanggraeni dan Melkior N. N. Sitokdana tentang aplikasi Government dengan metode naïve bayes, sedangkan untuk penelitian saat ini menggunakan objek QRIS dengan metode K-Nearest Neighbor.
4	Akhmad Deviyanto, M. Didik R. Wahyudi (2018)	Pilkada DKI 2017	K-Nearest Neighbor	Perbedaan dengan penelitian saat ini adalah dari penelitian Akhmad Deviyanto dan M. Didik R. Wahyudi tentang Pilkada DKI 2017, sedangkan untuk penelitian saat ini menggunakan objek QRIS.
5	Tiara Rahmadhani (2021)	Tayangan Televisi	K-Nearest Neighbor	Perbedaan dengan penelitian saat ini adalah dari penelitian Tiara Rahmadhani tentang

				Tayangan Televisi, sedangkan untuk penelitian saat ini menggunakan objek QRIS.
--	--	--	--	--

2.2 Landasan Teori

2.2.1 QR Code

QR *code* adalah singkatan dari *quick response code*. Kode ini berisi dua dimensi yang dapat memberikan berbagai jenis informasi secara langsung. Untuk membuka informasi dalam QR *code*, dibutuhkan *scan* atau pemindaian. Umumnya kode QR dapat menyimpan 2089 digit atau sekitar 4289 karakter. Jenis QR *code* ada 2 yaitu kode QR statistik dan kode QR dinamis. Kode QR statistik adalah kode QR yang tidak dapat diedit dan diubah lagi setelah selesai dibuat. Informasi yang ada didalamnya tidak bisa diganti. Contoh penggunaan QR code statis pada akses wifi, transaksi bitcoin, teks biasa, email. Kode QR dinamis adalah kode QR yang dapat diedit, diubah, dan diperbarui sesuai dengan kebutuhan. Contoh penggunaan QR code dinamis pada pembayaran, kupon, berbagai media sosial, informai web bisnis (UMA, 2022).

2.2.2 QRIS

QRIS adalah standar kode QR Nasional untuk memfasilitasi pembayaran kode QR di Indonesia yang dikeluarkan oleh Bank Indonesia dan Asosiasi Sistem Pembayaran Indoensia (ASPI) pada tanggal 19 Agustus 2019 (Telkom Indonesia, 2023). Diketahui dari website Bank Indonesia,

jumlah pengguna uang elektronik meningkat pada setiap tahunnya. Data Bank Indonesia menunjukkan nilai transaksi uang digital dapat dilihat pada Tabel 2.2

Tabel 2.2 Tabel transaksi uang elektronik

Tahun	Volume (Dalam satuan transaksi)	Nominal (Dalam Juta Rupiah)
2017	943,319,933	12,375,468.72
2018	2,922,698,905	47,198,616.11
2019	5,226,699,919	145,165,467.60
2020	4,625,703,561	204,909,170
2021	602,293,039	35,100,099.84

2.2.3 Twitter

Twitter merupakan salah satu media sosial yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks yang dikenal dengan sebutan kicauan (tweet). Menurut laporan *We Are Social*, jumlah pengguna Twitter di Indonesia mencapai 18,45 juta pada 2022. Angka tersebut menepatkan Indonesia di peringkat kelima negara pengguna Twitter terbesar di dunia. (Rizaty, 2022).

Twint merupakan *tools scrapping* untuk twitter yang dibuat secara khusus dalam bahasa python, *tools* ini dapat digunakan tanpa API dari Twitter dengan keterbatasan hanya dapat melakukan *scrapping* data 1500 tweet dalam setiap bulan. *Tools* ini juga dapat melakukan *scrapping user, followers, retweet* dsb.

2.2.4 Analisis Sentimen

Analisis sentimen adalah proses menganalisis teks digital untuk menentukan apakah nada emosional pesan tersebut positif, negatif, atau netral. Analisis sentimen dikenal sebagai alat kecerdasan bisnis penting yang membantu perusahaan meningkatkan produk dan layanan mereka (Amazon, 2023).

2.2.5 TF-IDF

Pembobotan *Term Frequency Inversed Document Frequency* (TF-IDF) adalah sebuah metode yang digunakan untuk mencari memberikan bobot hubungan suatu kata (term) terhadap dokumen. Penghitungan bobot sebuah dokumen dengan menggunakan perkalian nilai *tf* dan *idf* menunjukkan bahwa deskripsi terbaik dari dokumen adalah term yang banyak muncul dalam dokumen tersebut dan sangat sedikit muncul pada dokumen yang lain. *Term frequency* menyatakan frekuensi (tingkat keseringan) munculnya suatu *term* dalam suatu dokumen. Perhitungan bobot *term* dijelaskan pada persamaan 2.1.

$$W_{tf_{t,d}} = \begin{cases} 1 + \log_{10} tf_{t,d} , & \text{jika } tf_{t,d} > 0 \\ 0 & , \text{jika } tf_{t,d} = 0 \end{cases} \quad (2.1)$$

Keterangan:

$W_{tf_{t,d}}$ = hasil pembobotan *term frequency*

$tf_{t,d}$ = jumlah kemunculan kata *t* dalam dokumen *d*

Inversed Document Frequency digunakan untuk menghitung banyaknya jumlah dokumen dimana sebuah term itu muncul. Rumus perhitungan IDF pada persamaan 2.2.

$$idf_t = \log\left(\frac{N}{df_t + 1}\right) \quad (2.2)$$

Keterangan:

idf_t = hasil inverse dari df_t

N = jumlah data

df_t = jumlah dokumen yang mengandung *term t*

Penambahan 1 digunakan untuk menghindari pembagian terhadap 0 jika $df(t)$ tidak ditemukan pada corpus. Nilai TF-IDF dari sebuah kata merupakan kombinasi dari nilai TF dan nilai IDF dalam perhitungan bobot. Perhitungan TF-IDF dijelaskan dalam pada persamaan 2.3.

$$W_{tf} = W_{tf_{t,d}} \times \log_{10} \frac{N}{Idf_t} \quad (2.3)$$

Keterangan:

W_{tf} = pembobotan TF-IDF

$W_{tf_{td}}$ = hasil pembobotan tf_{td}

Idf_t = *Invers document Frequency*

N = jumlah data

2.2.6 K-Nearest Neighbor

K-Nearest Neighbor atau biasa disebut dengan KNN adalah salah satu metode paling sederhana untuk memecahkan masalah klasifikasi.

Alogaritma sering digunakan untuk klasifikasi teks dan data. Pada metode ini dilakukan klasifikasi terhadap obyek berdasarkan data yang jaraknya paling dekat dengan obyek tersebut (Nurjanah, 2017). Tujuan dari penggunaan metode ini adalah untuk mengklasifikasikan obyek berdasarkan atribut dan *training sample*. Metode ini menggunakan klasifikasi ketetanggan sebagai nilai prediksi dari *query intance* yang baru.

Klasifikasi teks menggunakan metode KNN akan menghasilkan nilai yang lebih optimal jika menggunakan rumus *cosine similarity* untuk pembobotan tiap-tiap kata pada dokumen teks yang akan diproses. Metode *Cosine Similarity* adalah metode untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah vektor dengan menggunakan kata kunci dari sebuah dokumen sebagai ukuran. Semakin banyak jumlah kata istilah yang muncul pada suatu dokumen, maka semakin tinggi nilai *Cosine Similarity* yang diperoleh. Persamaan dari *Cosine Similarity* ditunjukkan pada persamaan 2.4.

$$\text{CosSim}(q, d_j) = \frac{d_j \cdot q}{|d_j| \cdot |q|} = \frac{\sum_{i=1}^t (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^t w_{ij}^2} \cdot \sqrt{\sum_{i=1}^t w_{iq}^2}} \quad (2.4)$$

Keterangan:

$\text{CosSim}(q, d_j)$ = Nilai kemiripan antara dokumen uji (q) dengan dokumen latih ke j (d)

t = jumlah term (kata)

d = dokumen

q = kata kunci (query)

w_{ij} = bobot term (kata) ke I pada dokumen latih j

w_{iq} = bobot term (kata) ke I pada dokumen uji q

Setelah dilakukan perhitungan *cosine similarity*, maka dilanjutkan dengan perhitungan nilai k-values baru atau biasa disebut dengan nilai n. Persamaan 2.5 merupakan perhitungan dalam menentukan k-values baru.

$$n = \left\lceil \frac{k * N(c_m)}{\max\{n(c_m) | j = 1 \dots N_c\}} \right\rceil \quad (2.5)$$

Keterangan:

n = nilai k-values baru

k = nilai k-values yang ditetapkan diawal

$N(c_m)$ = jumlah dokumen/data latih pada kategori m

$\max\{n(c_m) | j = 1 \dots N_c\}$ = jumlah dokumen/ data latih terbanyak pada semua kategori yang ada.

Tahap akhir dari klasifikasi menggunakan persamaan 2.6 yaitu menghitung nilai probabilitas dan diambil nilai maksimum dari hasil perhitungan probabilitas tersebut pada masing-masing kelas. Hasil akhir perhitungan probabilitas akan menjadi penentuan terhadap dokumen uji.

$$p(x, c_m) = \operatorname{argmax}_m \frac{\sum_{d_j \in \text{top } n \text{ kNN}(c_m)} \text{sim}(x, d_j) y(d_j, c_m)}{\sum_{d_j \in \text{top } n \text{ kNN}(c_m)} \text{sim}(x, d_j)} \quad (2.6)$$

Keterangan:

$p(x, c_m)$ = probabilitas dokumen X menjadi anggota kategori c_m

$\text{sim}(x, d_j)$ = kemiripan antara dokumen X dengan dokumen latih

d_j

top n kNN = top n tetangga

$y(d_j, c_m)$ = fungsi atribut dari kategori tertentu yang memenuhi persamaan 2.7.

$$y(d_j, c_m) = \begin{cases} 1, & d_j \in c_m \\ 0, & d_j \notin c_m \end{cases} \quad (2.7)$$