

BAB II TINJAUAN PUSTAKA

2.1 Kajian Pustaka

Pada penyusunan penelitian ini tentunya diperlukan kajian atau teori-teori yang berkaitan dengan permasalahan dan ruang lingkup pembahasan penelitian yang akan digunakan sebagai suatu landasan. Dengan mengkaji terhadap penelitian-penelitian terdahulu merupakan salah satu upaya untuk mendapatkan wawasan sehingga dapat dijadikan sebagai referensi dalam menyusun penelitian.

Adapun penelitian-penelitian terdahulu tersebut membahas mengenai tema yang memiliki keterkaitan dalam sentimen analisa menggunakan metode *KNN*, berikut merupakan referensi jurnal terdahulu. Seperti penelitian yang dilakukan oleh Herda Andriana, Shofa Shofia Hilabi, dan Agustia Hananto dengan judul “Penerapan Metode K-Nearest Neighbor pada Sentimen Analisis Pengguna Twitter Terhadap KTT G20 di Indonesia” [10].

Penelitian tersebut bertujuan untuk mengetahui tingkat akurasi metode *KNN* pada model klasifikasi data *tweet* KTT G20 yang dibedakan menjadi data berkategori positif, netral atau negatif dengan jumlah data sebanyak 2500 data. Hasil dari penelitian bahwa penyusunan menggunakan algoritma *KNN* mencapai tingkat akurasi sebesar 99%.

Penelitian lainnya dilakukan oleh Dyah Apriliani, Ardi Susanto, Muhammad Fikri Hidayattullah, dan Ginanjar Wiro Sasmito dengan judul “Sentimen Analisis Pandangan Masyarakat Terhadap Vaksinasi Covid 19 Menggunakan K-Nearest Neighbors” [11]. Dalam penelitiannya mengusulkan proses klasifikasi pandangan masyarakat terhadap vaksinasi Covid 19. Data penelitian yang digunakan diambil dari twitter sebanyak 2241 data. Data akan diklasifikasikan menjadi 2 kelas yaitu positif dan negatif. Proses klasifikasi akan dilakukan dengan menggunakan metode *K-Nearest Neighbors (KNN)*. Hasil dari penelitian bahwa penyusunan menggunakan algoritma *KNN* mencapai tingkat akurasi sebesar 79,25%.

Penelitian yang dilakukan oleh Jepi Supriyanto, Debby Alita, dan Auliya Rahman Isnain dengan judul “Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring” [12]. Tujuan dari penelitian adalah untuk menerapkan algoritma *KNN (K-Nearest Neighbor)* dalam melakukan sentimen analisis pengguna Twitter tentang isu terkait kebijakan pemerintah tentang Pembelajaran Daring.

Hasil dari penelitian bahwa penyusunan menggunakan algoritma *KNN* mencapai tingkat akurasi sebesar 84.65%.

Penelitian yang dilakukan oleh Muhamad Trian Diwandanu dan Lulu Mawaddah Wisudawati yang berjudul “Analisis Sentimen Terhadap Twit Maxim Pada Twitter Menggunakan R Programming Dan K-Nearest Neighbors” [13]. Tujuan penelitian ini dilakukan adalah melakukan analisis sentimen terhadap twit maxim pada Twitter. Metode klasifikasi analisis sentimen yang digunakan menggunakan *K-Nearest Neighbors (KNN)* untuk mengklasifikasi data sentimen positif, negatif dan netral. Hasil dari penelitian bahwa penyusunan menggunakan algoritma *KNN* mencapai tingkat akurasi sebesar 95.43%.

Selanjutnya penelitian yang dilakukan oleh Dickna Niken Larasakti, Abdul Aziz dan Danang Aditya dengan judul “Analisis Sentimen Komentar Video Youtube Dengan Metode K-Nearest Neighbor” [14]. Tujuan penelitian ini dilakukan adalah analisa sentimen terhadap data komentar youtube dengan menggunakan *K-Nearest Neighbors (KNN)* untuk mengklasifikasi data sentimen positif, dan negatif. Hasil dari penelitian bahwa penyusunan menggunakan algoritma *KNN* mencapai tingkat akurasi sebesar 92.71%.

Tabel 2.1 Penelitian Terdahulu

No	Judul	Tahun	Peneliti	Sumber Data	Hasil
1.	Penerapan Metode K-Nearest Neighbor pada Sentimen Analisis Pengguna Twitter Terhadap KTT G20 di Indonesia	2023	Herda Andriana, Shofa Shofia Hilabi, dan Agustia Hananto	Twitter	Hasil dari penelitian bahwa penyusunan menggunakan algoritma <i>KNN</i> mencapai tingkat akurasi sebesar 99%
2.	Sentimen Analisis Pandangan Masyarakat Terhadap Vaksinasi Covid 19 Menggunakan K-Nearest Neighbors	2023	Dyah Apriliani, Ardi Susanto, Muhammad Fikri Hidayattullah, dan Ginanjar Wiro Sasmito	Twitter	Hasil dari penelitian bahwa penyusunan menggunakan algoritma <i>KNN</i> mencapai tingkat akurasi sebesar 79,25%.
3.	Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring	2023	Jepi Supriyanto, Debby Alita, dan Auliya Rahman	Twitter	Hasil dari penelitian bahwa penyusunan menggunakan algoritma <i>KNN</i> mencapai tingkat akurasi sebesar 84.65%.
4.	Analisis Sentimen Terhadap Twit Maxim Pada Twitter Menggunakan R Programming Dan K-Nearest Neighbors	2023	Muhamad Trian Diwandanu dan Lulu Mawaddah Wisudawati	Twitter	Hasil dari penelitian bahwa penyusunan menggunakan algoritma <i>KNN</i> mencapai tingkat akurasi sebesar 95.43%.
5.	Analisis Sentimen Komentar Video Youtube Dengan Metode K-Nearest Neighbor	2023	Dickna Niken Larasakti, Abdul Aziz dan Danang Aditya	Twitter	Hasil dari penelitian bahwa penyusunan menggunakan algoritma <i>KNN</i> mencapai tingkat akurasi sebesar 92.71%.

Berdasarkan Tabel 2.1 merupakan penelitian terdahulu yang dijadikan sebagai referensi dalam mengembangkan skema penulisan penelitian ini karena berkaitan dengan penelitian.

2.2 Dasar Teori

2.2.1 Pemindahan Ibu Kota

Menurut Kepala Badan Perencanaan Pembangunan Nasional (Bapennas), upaya pemindahan ibu kota Indonesia dimulai pada tahun 2019, pada masa kepresidenan Jokowi. Pada tanggal 29 April 2019, melalui rapat terbatas pemerintah, Presiden Jokowi memutuskan untuk memindahkan ibu kota negara ke luar pulau Jawa. Rencana ini kemudian diwujudkan dalam rencana pembangunan jangka menengah nasional 2020-2024. Pada tanggal 26 Agustus 2019, Presiden Jokowi mengumumkan bahwa ibu kota baru akan dibangun di wilayah administratif Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara, Kalimantan Timur [15].

2.2.2 Twitter

Twitter adalah sebuah situs web yang dioperasikan oleh Twitter Inc. Platform ini menawarkan jejaring sosial yang memungkinkan pengguna untuk mengirim pesan dalam bentuk cuitan atau yang biasa disebut sebagai "tweets" [16]. pengguna dapat berbagi informasi tentang kehidupan sehari-hari mereka. Twit-twit yang ditulis oleh pengguna Twitter merupakan sumber data yang valid dan relevan untuk melakukan analisis sentimen. Karena platform ini memungkinkan pengguna untuk secara spontan mengekspresikan pemikiran, perasaan, dan opini mereka dalam waktu nyata, komentar-komentar di Twitter dapat memberikan wawasan yang berharga tentang sentimen publik terhadap berbagai topik atau peristiwa. Oleh karena itu, analisis sentimen terhadap *twit-twit* ini dapat memberikan pemahaman yang lebih mendalam tentang pandangan dan persepsi pengguna terhadap suatu masalah atau peristiwa tertentu [13].

2.2.3 Machine Learning

Machine learning adalah suatu pendekatan dalam bidang komputer yang sangat membantu dalam menyelesaikan berbagai masalah dan membuat proses pengerjaan menjadi lebih mudah. Machine learning memungkinkan mesin untuk

belajar dan meningkatkan kinerjanya sendiri tanpa harus dikontrol secara langsung oleh pengguna [17].

2.2.4 Text Preprocessing

Text preprocessing merupakan langkah awal dalam persiapan dokumen teks sebelum dapat digunakan dalam proses pemodelan. Pada tahap *text preprocessing*, data teks mentah diolah dengan membersihkan karakter khusus, mengubah teks menjadi huruf kecil, menghapus kata penghubung, dan mengubah kata-kata menjadi bentuk dasarnya. Tujuannya adalah untuk menyederhanakan teks dan mempersiapkannya agar siap untuk analisis lebih lanjut. Dengan melakukan *text preprocessing* yang baik, dokumen teks siap digunakan dalam analisis sentimen, klasifikasi, atau pencarian informasi, dan dapat menghasilkan hasil yang akurat [18].

2.2.5 Term - Frequency Inverse Document Frequency (TF – IDF)

TF-IDF merupakan metode pembobotan kata yang melibatkan proses *text preprocessing*. Dalam metode ini, pentingnya suatu kata dalam sebuah dokumen ditentukan berdasarkan frekuensi kemunculannya. Tahap awal melibatkan pemisahan kata-kata dalam dokumen menggunakan *tokenisasi*, diikuti dengan penghapusan kata-kata umum yang tidak memberikan informasi penting (*stopwords*), dan pengubahan kata-kata ke bentuk dasarnya melalui *stemming*. Selanjutnya, dilakukan perhitungan frekuensi kemunculan kata-kata dalam dokumen. Kata-kata yang sering muncul akan diberi bobot yang lebih tinggi, namun bobot juga dipengaruhi oleh seberapa sering kata tersebut muncul dalam seluruh koleksi dokumen. Dengan menggunakan metode *TF-IDF*, kita dapat mengidentifikasi kata-kata penting yang memberikan kontribusi besar dalam sebuah dokumen. Metode ini berguna dalam berbagai aplikasi seperti analisis teks, rekomendasi, dan klasifikasi dokumen [19].

2.2.6 K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor (KNN)* merupakan metode yang termasuk dalam kategori algoritma *supervised*. Dalam *supervised learning*, perbedaannya dengan *unsupervised learning* terletak pada tujuannya. Dalam *supervised learning*, tujuannya adalah untuk menemukan pola baru dalam data dengan menghubungkan pola

data yang sudah ada dengan data baru yang akan diprediksi. Sedangkan dalam *unsupervised learning*, data belum memiliki pola yang jelas, dan tujuannya adalah untuk menemukan pola dalam data tersebut. Tujuan dari algoritma *KNN* adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan sampel pelatihan (*training samples*). Prosesnya melibatkan perbandingan atribut dari objek baru dengan atribut dari sampel pelatihan yang sudah ada. Hasil klasifikasi dari objek baru didasarkan pada mayoritas kategori dari sampel pelatihan yang memiliki atribut yang mirip dengan objek baru tersebut. Dengan menggunakan algoritma *KNN*, kita dapat melakukan pengklasifikasian objek baru berdasarkan kesamaan atribut dengan sampel pelatihan yang sudah ada. Algoritma ini banyak digunakan dalam berbagai aplikasi seperti pengenalan pola, analisis teks, dan sistem rekomendasi [20].

2.2.7 *Confusion Matrix*

Confusion matrix adalah metode yang umum digunakan untuk menghitung akurasi. Ia memberikan representasi hasil klasifikasi dengan empat istilah yang berbeda. Istilah-istilah tersebut meliputi *True Positive (TP)*, yang merupakan jumlah data positif yang terdeteksi dengan benar. Selanjutnya, *True Negative (TN)* adalah jumlah data negatif yang terdeteksi dengan benar. Kemudian, *False Positive (FP)* merujuk pada data yang sebenarnya negatif, namun terdeteksi sebagai positif. Sedangkan *False Negative (FN)* menggambarkan data yang sebenarnya negatif, tetapi terdeteksi sebagai negatif [21]. *Confusion Matrix* memiliki rumus accuracy sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Gambar 2.1 Rumus *Accuracy Confusion Matrix*