

BAB I PENDAHULUAN

1.1 Latar Belakang

Saat ini, segala macam komunikasi sangat sering dilakukan melalui media sosial di internet. Salah satu media sosial yang digunakan di Indonesia adalah *Twitter*. Untuk berbagi informasi seluas mungkin, *twitter* menyediakan akses programatik ke data *twitter* kepada perusahaan, pengembang, dan pengguna lewat API (*application programming interface*). *Twitter* mengizinkan kemudahan akses ke bagian dari layanannya melalui API untuk memungkinkan orang-orang membangun perangkat lunak yang terintegrasi dengan *twitter* seperti solusi yang membantu sebuah perusahaan menjawab umpan balik pelanggan di *twitter*. Platform API *Twitter* menyediakan akses luas ke data *twitter* publik yang telah dipilih pengguna untuk dibagikan secara umum. *Twitter* juga mendukung API yang memungkinkan pengguna mengelola informasi yang termasuk non-publik dan memberikan informasi ini ke pengembang yang telah diizinkan pengguna untuk melakukannya[1]. Platform media sosial *Twitter* menghubungkan pengguna supaya dapat berinteraksi secara pribadi maupun terbuka. Berbagai macam topik aktif dibicarakan seperti halnya dengan politik, sosial, ekonomi, hukum dan budaya. Para pengguna *Twitter* dapat mengetahui topik yang dibahas melalui fitur *hashtag* secara *real-time*. Pembahasan dari para pengguna perihal media *online* yang diperbincangkan di *Twitter* beberapa kali menjadi *headline* dari sebuah berita[2].

Pemindahan ibu kota negara adalah salah satu topik yang sempat *trending* diperbincangkan di *Twitter*. Wacana ini sudah pernah muncul di era presiden-presiden sebelumnya, kemudian kembali ramai diperbincangkan setelah pemerintahan presiden Joko Widodo membahas kembali hal tersebut. Isu pemindahan ibu kota tentu bukan hal yang mudah dan menjadi perhatian besar bagi negara kita terutama dalam menghadapi persaingan *global*. Namun, topik ini memang selalu menimbulkan kontroversi dari warganet yang *pro* maupun kontra. Topik ini tidak terlepas dari perhatian warganet *Twitter* dan menjadi sumber perdebatan dari sudut pandang yang berbeda[3]. Penyampaian opini warganet pada

media sosial *Twitter* tersebut kemudian dikumpulkan untuk melangsungkan proses analisis sentimen dalam menjalankan tahapan proses *text mining*[4].

Analisis sentimen adalah proses otomatis dalam memahami, mengekstraksi, dan memproses data tekstual untuk mendapatkan informasi sentimen yang terkandung dalam kalimat opini[5]. Analisis sentimen diperlukan untuk mengetahui pendapat atau kecenderungan warganet mengenai pemindahan ibukota negara Indonesia yang memuat sentimen positif dan negatif. Sentimen yang dimaksud biasanya merujuk pada perasaan, emosi, sikap maupun pendapat. Dengan menggunakan metode *Naïve Bayes Classifier* dan *K-nearest neighbor* yang dijadikan pilihan karena memiliki model yang sederhana namun dapat bersaing dengan model algoritma lainnya[6]. Penentuan klasifikasi menggunakan *K-Nearest Neighbor* pada implementasinya mengklasifikasikan objek menggunakan jarak terdekat atau kemiripan dari suatu objek pada kumpulan data. Sedangkan *Naïve Bayes* merupakan pembelajaran *machine learning* untuk pengelompokan (klasifikasi) statistik dalam perhitungan probabilitas. *Naïve* berdasar pada penerapan teorema atau aturan bayes dengan asumsi independensi yang kuat pada fitur, artinya bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. Dengan perbedaan tersebut akan dibandingkan performa yang dihasilkan pada penerapannya terhadap klasifikasi sentimen.

Berdasarkan uraian latar belakang tersebut, maka dirancanglah sebuah sistem analisis sentimen dengan *python* berbasis *text mining*. Penelitian ini berjudul “Analisis Sentimen Pindah Ibu Kota Negara Berbasis *Text Mining* Menggunakan Algoritma *K-nearest neighbors* Dan *Naïve Bayes*” sebagai objek dalam menyelesaikan Tugas Akhir.

1.2 State of The Art

State of the art adalah bentuk keaslian karya ilmiah yang dibuat sehingga tidak ada Tindakan plagiat sebagai bentuk pembajakan terhadap karya orang lain. Dalam hal ini, *state of the art* menjelaskan perbandingan terhadap riset yang telah dilakukan sebelumnya dan menjadi acuan pembuatan tugas akhir. Perbandingan tersebut yaitu:

Tabel 1.1 Referensi.

Judul	Peneliti	Tahun	Deskripsi
Sentimen Analisis <i>Stay Home</i> menggunakan metode klasifikasi <i>Naïve Bayes</i> , <i>Support Vektor Machine</i> , dan <i>k-nearest neighbor</i>	Ikhwanul Hakim, Arifin Nugroho, Sulaeman Hadi Sukmana, Windu Gata	2020	Penelitian ini fokus pada sentimen analisis stay home pada pengguna <i>twitter</i> , untuk dapat melihat efek dari kebijakan tersebut terhadap kehidupan mereka. Karena hampir diseluruh negara yang terkena pandemi ini mengeluarkan kebijakan seperti itu. Data yang diperoleh akan diolah menggunakan tiga metode klasifikasi yaitu <i>Naive bayes</i> (NB), <i>Support Vector Machine</i> (SVM), dan <i>k-nearest neighbor</i> (k-NN). Dengan ketiga metode klasifikasi tersebut, akan dicari metode mana yang akan menghasilkan akurasi yang paling baik terkait dengan stay home dari tweets para penggunanya. Setelah dilakukan percobaan, algoritma <i>Support Vector Machine + Smote</i> mendapatkan hasil akurasi yang paling baik jika dibandingkan dengan dua algoritma lainnya. Hasil akurasi yang didapat sebesar 80,05%.
Analisa Sentimen Terhadap Review Fintech Dengan Metode <i>Naive bayes classifier</i> Dan <i>K- Nearest Neighbor</i>	Surohman, Sopian Aji, Rousyati, Fanny Fatma Wati	2020	Penelitian yang dipakai adalah metode eksperimen dengan menggunakan Algoritma <i>Naive bayes classifier</i> dan <i>K-nearest neighbor</i> . Hasilnya pengujian pada Aplikasi Dana memiliki nilai akurasi terbaik sebesar 84,76%. Maka dapat menyimpulkan review tersebut dapat meyakinkan dari review positif oleh pengguna lain dan jika terdapat review negatif maka akan menjadi masukan kepadapihak perusahaan untuk lebih mengembangkan maupun memperbaiki produk tersebut.
		2020	Sentimen yang dianalisis yakni berupa sentimen positif, negatif, dan netral. Untuk mendapatkan data tweet pada <i>Twitter</i> , perlu dilakukan pengumpulan data terlebih dahulu. Kemudian untuk melakukan analisis sentimen, data perlu diolah dengan text mining. Pada studi kasus ini,

<p>Analisis Sentimen : Pemindahan Ibu Kota Indonesia Pada <i>Twitter</i></p>	<p>Erica Mas'udah, Eka Dyar W., Amalia Anjani</p>		<p>text mining yang dilakukan menggunakan metode Naïve Bayes. Metode Naïve Bayes yang digunakan ada 3 macam, dengan tujuan untuk mendapatkan nilai akurasi yang terbaik. Ditambah lagi dengan melakukan percobaan beberapa skenario. Dari penelitian yang dilakukan, hasil akurasi terbaik didapatkan oleh Multinomial Naïve Bayes.</p>
<p>Analisis Sentimen Topik Pindah Ibu Kota Pada <i>Twitter</i> Menggunakan Metode <i>Modified K-nearest neighbor</i></p>	<p>A.D.Dian Kurniawan</p>	<p>2021</p>	<p>Analisis sentimen dilakukan untuk menentukan apakah opini atau komentar terhadap suatu permasalahan, memiliki kecenderungan bernilai positif atau negatif dan dapat dijadikan sebagai acuan dalam meningkatkan suatu pelayanan. Analisis sentimen menggunakan algoritma M-KNN (<i>Modified K-nearest neighbor</i>) dengan data <i>twitter</i> sebanyak 4000 data tweet dan menghasilkan akurasi terbaik sebesar 56.16%, <i>precision</i> sebesar 100% dan <i>recall</i> sebesar 56,16% dengan nilai k=13 dengan perbandingan data latih dan data uji 70:30.</p>
<p>Analisis Pindah Ibu Kota Berbasis <i>Naïve Bayes Classifier</i></p>	<p>P. Arsi, B.A Kusuma, A, Nurhakim</p>	<p>2021</p>	<p>Pada penelitian ini diusulkan metode <i>Naive bayes classifier</i> (NBC) untuk menganalisa sentimen terhadap wacana pemerintah di media massa online <i>Twitter</i> pada topik pemindahan ibukota Indonesia dengan cara mengklasifikasikan menjadi positif, dan negatif. Hasil penelitian ini menunjukkan bahwa nilai akurasi yang diperoleh sebesar 94,33%. Dengan dilakukannya analisa sentimen ini diharapkan dapat diketahui permasalahan yang terdapat pada kontroversi topik pemindahan ibukota, sehingga dapat dijadikan sebagai bahan evaluasi untuk kepentingan lebih lanjut.</p>

Penelitian mengenai analisis sentimen menggunakan *machine learning* telah banyak dilakukan dan banyak publikasi riset tentang topik tersebut. Sebagaimana yang tertera dalam Tabel 1.1 telah diusulkan beberapa macam metode mengenai analisis sentimen menggunakan *machine learning*. Penelitian [7] menggunakan tiga metode klasifikasi yakni *Naïve Bayes (NB)*, *Support Vector Machine (SVM)*, dan *k-*

nearest neighbor (k-NN) untuk mengolah data yang diperoleh dari social media *Twitter*. Untuk mendapatkan hasil akurasi yang paling baik diantara ketiga metode klasifikasi tersebut maka dilakukan sebuah percobaan. Sesudahnya didapatkan algoritma *Support Vector Machine + Smote* sebagai hasil akurasi yang paling baik dengan angka 80,05%. Artikel [6] menggunakan metode *k-NN* dan *Naïve Bayes Classifier* pada Aplikasi Dana dengan hasil uji sebesar 84,7% sebagai akurasi terbaik terhadap review positif dari pengguna Aplikasi Dana. Dapat disimpulkan bahwa review tersebut dapat meyakinkan pengguna lain dan review negatif dapat menjadi masukan terhadap pihak perusahaan untuk melakukan pengembangan dan perbaikan produk mereka.

Analisis sentimen terhadap pemindahan Ibu Kota Indonesia telah dibahas oleh beberapa artikel. Diantaranya [8] menggunakan tiga macam metode *Naïve Bayes* dengan menambahkan percobaan beberapa skenario. Kemudian [9] memakai 4000 data *tweet* dan mengolah opini tersebut menggunakan metode M-KNN (*Modified K-nearest neighbor*) dengan hasil akurasi terbaik 56,1%. Penelitian [3] melakukan penelitian di tahun yang sama menggunakan algoritma *Naïve Bayes Clasifier* pada kontroversi topik pemindahan Ibu Kota Negara Indonesia. Dari penelitian tersebut didapat kesimpulan bahwa nilai *recall* yang tinggi sangat baik diaplikasikan pada sistem ini karena politik adalah sebuah topik yang sensitif.

Berdasarkan Tabel 1.1 sudah pernah dilakukan penelitian mengenai topik Analisis Sentimen terhadap Pemindahan Ibu Kota Negara Indonesia. Oleh karena itu, pada penelitian tugas akhir ini akan dilakukan penelitian menggunakan data yang baru pada rentang bulan Juni hingga Juli 2022. Penelitian ini menggunakan algoritma *K-Nearest Neighbor* dan *Naïve Bayes* untuk melakukan analisis sentimen. Pembagian data pada data latih akan dilakukan dengan perbandingan 7:3, 8:2 dan 9:1 serta penambahan metode *SMOTE* untuk memperbaiki perbandingan kelas data yang tidak seimbang. Penelitian ini diberi judul “Analisis Sentimen Pindah Ibu Kota Negara Berbasis *Text Mining* Menggunakan Algoritma *K-Nearest Neighbors* Dan *Naïve Bayes*” dengan memuat teori yang sudah ada sebelumnya.

1.3 Rumusan Masalah

Berdasarkan pada latar belakang diatas, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana penerapan metode *K-nearest neighbors* dan *Naïve Bayes* dalam mengklasifikasi sentimen terhadap topik pemindahan Ibu Kota Negara pada *Twitter*.
2. Bagaimana tingkat perbandingan nilai akurasi metode *K-nearest neighbors* dan *Naïve Bayes*.

1.4 Tujuan

Tujuan yang ingin dicapai dari penelitian ini yaitu:

1. Penerapan metode *K-nearest neighbors* dan *Naïve Bayes* dalam mengklasifikasikan sentimen terhadap topik pemindahan Ibu Kota pada *Twitter*.
2. Menguji tingkat akurasi terhadap metode *K-nearest neighbors* dan *Naïve Bayes* dalam mengklasifikasikan sentimen terhadap topik pemindahan Ibu Kota pada *Twitter*.

1.5 Manfaat

Hasil penelitian ini diharapkan memiliki manfaat penelitian sebagai berikut:

1. Manfaat bagi bidang Akademis
Memperkaya khazanah saah satu bidang ilmu pengetahuan yaitu *Machine Learning*, khususnya penerapan *text mining* menggunakan metode *K-nearest neighbors* dan *Naïve Bayes* untuk mengklasifikasikan sentimen pada topik pindah Ibu Kota dengan *platform Social media Twitter*.
2. Manfaat Praktis
Dengan penerapan analisis sentimen pindah Ibu Kota ini dapat digunakan sebagai alat bantu untuk melihat respon masyarakat terhadap kebijakan pemerintah tersebut, sehingga dapat segera diambil langkah-langkah strategis berikutnya.

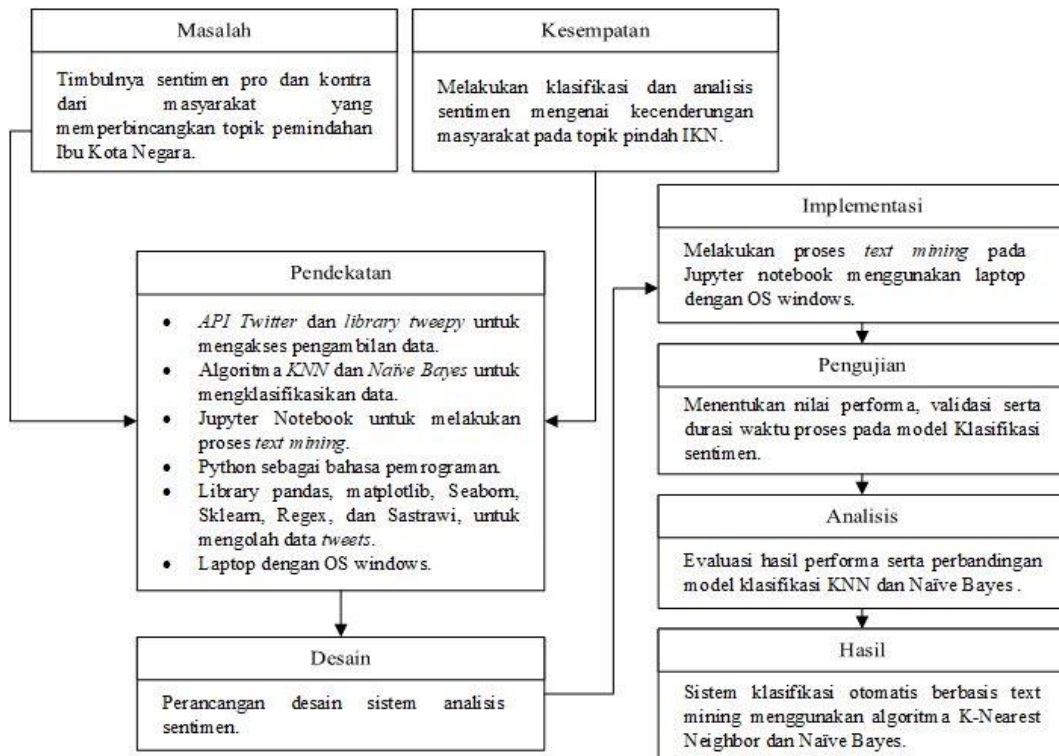
1.6 Batasan Masalah

Diperlukan Batasan masalah dalam pembuatan sistem analisis sentimen sehingga dapat diperoleh hasil yang sesuai dengan tujuan pembuatan/ Adapun pembatasan masalah pada penelitian ini adalah sebagai berikut:

1. Objek penelitian dilakukan terhadap topik pemindahan Ibu Kota pada *Twitter*.
2. Klasifikasi sentimen dibagi menjadi 2 kelas yaitu Positif dan Negatif.
3. Pelabelan terhadap kelas pada data latih dilakukan secara pribadi dan manual.
4. Data yang diambil adalah dari tanggal 20 Juni hingga 28 Juli 2022.
5. Data *Tweet* yang diambil yaitu *tweet* yang mengandung topik pindah Ibu Kota.
6. Jumlah data yang dilatih sebanyak 3.966 data *tweets*.
7. Pembagian data latih dilakukan dengan perbandingan 7:3, 8:2, dan 9:1.
8. Data *tweet* yang diambil berbahasa Indonesia.
9. Kata kunci yang digunakan Ketika pengambilan data adalah “IKN”, “Ibu kota Nusantara”.

1.7 Kerangka Berfikir

Kerangka berfikir berisi alur pemikiran yang memuat uraian sistematis tentang informasi hasil penelusuran atau perumusan masalah penelitian yang diduga dapat diselesaikan melalui pendekatan yang dilakukan dengan penelitian, membantu mempercepat pemahaman tentang alur logis penelitian, dan menjadi bentuk kasar dari struktur penelitian yang dilakukan. Kerangka berfikir penelitian ini dapat dijelaskan pada Gambar 1.1:



Gambar 1.1 Kerangka berfikir.

1.8 Sistematika Penulisan

Penulisan penelitian tugas akhir ini memiliki sistematika penulisan, berikut penjabarannya:

BAB I PENDAHULUAN

Pada bab ini meliputi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, *state of the art*, kerangka berfikir, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini membahas mengenai dasar teori dan pandangan umum tentang analisis sentimen, *twitter*, *text mining*, persiapan data, penilaian sentimen, *text transformation*, *feature selection*, *python*, *confusion matrix*, *K-nearest neighbor* dan *Naïve bayes*.

BAB III METODOLOGI RENCANA PENELITIAN

Pada bab ini menjelaskan metode dan tahapan – tahapan yang dilakukan ketika melakukan penelitian dan rencana dilakukannya penelitian.

BAB IV PERANCANGAN DAN IMPLEMENTASI

Pada bab ini menjelaskan mengenai tahapan perancangan dan hasil rancangannya kemudian melakukan implementasi dari rancangan tersebut baik *hardware* maupun *software*.

BAB V PENGUJIAN DAN ANALISIS

Pada bab ini membahas pengujian terhadap metode klasifikasi yang digunakan sehingga memperoleh hasil yang selanjutnya dianalisis berdasarkan pada teori dan standar penelitian sebelumnya.

BAB VI PENUTUP

Pada bab ini memuat elaborasi dan rincian kesimpulan dari hasil yang tercapai dalam tujuan. Serta memberikan saran atau rekomendasi terhadap penelitian lanjutan yang berkaitan dengan kajian.

