

BAB III TINJAUAN PUSTAKA

3.1 Subjek dan Objek Penelitian

Subjek penelitian merupakan orang, tempat, atau benda yang diamati. Subjek pada penelitian ini ialah para pengguna atau berbagai orang yang memiliki media social Twitter dan melakukan *tweet* mengenai IKN. Objek penelitian ialah suatu atribut yang berasal dari orang atau kegiatan yang ditetapkan oleh peneliti. Objek penelitian ini ialah data *tweet* yang berkaitan dengan pemindahan IKN.

3.2 Alat dan Bahan

Alat dan Bahan yang digunakan dalam penelitian ini antara lain:

3.2.1 Perangkat keras

Perangkat keras yang dibutuhkan dan digunakan dalam pengembangan penelitian ini dengan sebuah laptop yang memiliki spesifikasi:

1. Prosesor Intel® Core™ i3-5005U
2. RAM 4 GB
3. Storage 500 GB HDD
4. Display 14” Full HD
5. Graphics Intel HD

3.2.2 Perangkat lunak

Perangkat lunak yang dibutuhkan dan digunakan dalam pengembangan penelitian ini, sebagai berikut:

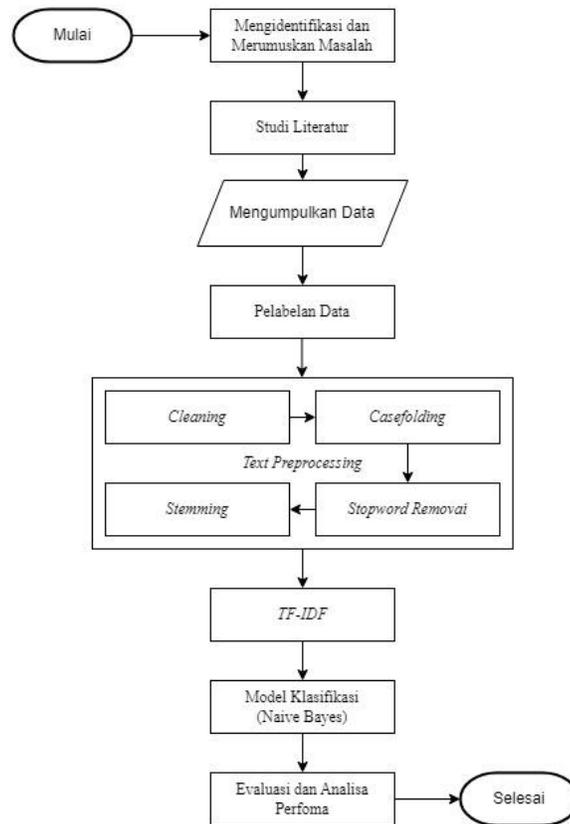
Tabel 3.1 Perangkat Lunak dan Kegunaan

No	Nama	Kegunaan
1	Sistem operasi Windows 10	Akses ke berbagai perangkat lunak
2	Jupyter Notebook (Anaconda)	Pemrograman, eksplorasi data, analisis data dan visualisasi
3	Python 3.8	Automasi tugas dan skrip
4	Browser	Penelusuran data, akses ke media sosial dan pengumpulan data teks

3.2.3 Bahan

Bahan penelitian yang digunakan untuk penelitian ini ialah kumpulan data *tweet* pada media sosial Twitter terkait IKN sebanyak 1000 data.

3.3 Diagram Alir Penelitian



Gambar 3.1 Diagram Alir Penelitian

Pada Gambar 3.1 merupakan diagram alir penelitian secara umum. Berikut uraian dari gambar 3.1:

3.3.1 Mengidentifikasi dan merumuskan masalah

Dalam langkah ini, peneliti melakukan penentuan bidang, topik, dan masalah penelitian yang akan diteliti. Selain itu, peneliti juga mengusulkan metode yang akan digunakan dalam penelitian. Pada tahap ini, peneliti mempelajari masalah-masalah yang masih terjadi dalam kehidupan sehari-hari. Tujuan dari penelitian ditetapkan dengan jelas, termasuk ruang lingkup yang akan dibahas.

3.3.2 Studi Literatur

Langkah selanjutnya adalah melaksanakan studi literatur yang melibatkan pengumpulan data terkait dengan topik permasalahan yang sedang diteliti. Data yang dikumpulkan meliputi data *tweet* klasifikasi pemindahan IKN di Twitter, dan algoritma yang akan diusulkan, yaitu *Naive Bayes*. Untuk mendapatkan data tersebut, peneliti merujuk kepada jurnal ilmiah, buku elektronik, situs internet, serta sumber media elektronik lainnya. Melalui studi literatur, tujuan yang ingin dicapai adalah untuk memperkuat pemahaman tentang permasalahan yang sedang dibahas dalam penelitian ini, serta menyediakan dasar yang solid untuk pengembangan penelitian selanjutnya.

3.3.3 Mengumpulkan Data

Langkah berikutnya adalah melakukan pengumpulan data yang terdiri dari kumpulan *tweet* yang berkaitan dengan pemindahan IKN di platform Twitter. Tujuan dari pengumpulan data ini adalah untuk memperoleh informasi yang relevan mengenai pemindahan IKN yang akan dianalisis melalui proses klasifikasi menggunakan metode *Naive Bayes*. Dalam proses pengumpulan data, dilakukan penggunaan teknik *crawling* dengan memanfaatkan *library tweepy*, yang memudahkan akses ke Twitter API untuk mengambil data yang diperlukan.

3.3.4 Pelabelan Data

Setelah mengumpulkan data, data tersebut masih berbentuk mentah, sehingga memerlukan proses pelabelan agar dapat dilakukan klasifikasi. Tujuan dari pelabelan ini adalah untuk menandai apakah setiap teks mengandung unsur pro terhadap pemindahan IKN (label "1"), netral terhadap pemindahan IKN (label "2"), atau menolak pemindahan IKN (label "3"). Proses pelabelan data dilakukan secara manual, di mana setiap teks akan ditinjau oleh peneliti untuk menentukan label yang sesuai.

3.3.5 Text Preprocessing

Setelah melabeli data, langkah selanjutnya adalah melakukan *preprocessing teks*. *Preprocessing* ini melibatkan serangkaian langkah yang meliputi *cleaning*, *case folding*, *tokenisasi*, penghapusan kata penghubung, dan *stemming*.

- *Cleaning* dilakukan untuk membersihkan teks dari elemen yang tidak relevan atau dapat mengganggu analisis, seperti *retweet*, *URL*, dan emotikon. Langkah ini bertujuan untuk mempertahankan hanya informasi teks yang penting.
- *Case Folding*, yaitu mengubah huruf kapital menjadi huruf kecil (*lowercase*). Hal ini dilakukan agar tidak ada perbedaan dalam pemrosesan teks berdasarkan kapitalisasi huruf.
- *Tokenisasi* adalah proses mengurai kalimat menjadi kata-kata yang terpisah. Ini memungkinkan pemrosesan lebih lanjut pada tingkat kata dan membantu dalam analisis teks.
- Penghapusan kata penghubung (*stopword removal*) dilakukan untuk menghilangkan kata-kata yang umum tetapi tidak memberikan kontribusi signifikan pada pemahaman konten teks. Contohnya adalah kata penghubung seperti "dan", "atau", "di", dan sebagainya.
- *Stemming* yang bertujuan untuk merubah kata-kata menjadi bentuk dasarnya atau kata dasar. Misalnya, kata "berlari" menjadi "lari", "bermain" menjadi "main", dan seterusnya. Hal ini dilakukan untuk mengurangi variasi kata yang serupa agar dapat menggambarkan makna yang sama. Dengan melakukan *preprocessing teks* ini, data menjadi lebih terstruktur dan siap untuk diolah lebih lanjut dalam tahap analisis.

3.3.6 Proses *TF-IDF*

Selanjutnya, data akan mengalami proses seleksi fitur menggunakan metode pembobotan *TF-IDF*. Pembobotan fitur ini merupakan langkah krusial sebelum melakukan proses klasifikasi. Dalam metode *TF-IDF*, setiap fitur atau kata dalam *dataset* akan diberikan bobot berdasarkan dua faktor: frekuensi kemunculan kata tersebut dalam dokumen tertentu (*TF*) dan inversi frekuensi kemunculan kata tersebut dalam seluruh *dataset* (*IDF*). Tujuan dari pembobotan fitur ini adalah untuk meningkatkan akurasi klasifikasi dengan memberikan fokus pada kata-kata yang memiliki kontribusi penting dalam mengklasifikasikan data. Dengan memberikan bobot yang lebih tinggi pada kata-kata yang jarang muncul secara umum tetapi sering muncul dalam dokumen tertentu, metode *TF-IDF* membantu

dalam mengidentifikasi kata-kata kunci yang dapat membedakan kelas atau topik yang berbeda. Dengan menggunakan pembobotan fitur *TF-IDF*, sistem klasifikasi dapat memperoleh representasi yang lebih informatif dari teks, yang kemudian digunakan dalam proses klasifikasi untuk memprediksi kelas atau label yang tepat untuk data yang belum dilihat sebelumnya.

3.3.7 Model Klasifikasi

Selanjutnya, langkah berikutnya adalah membangun model klasifikasi menggunakan metode *Naive Bayes*. Proses ini merupakan inti dari ekstraksi pengetahuan dari data yang telah dikumpulkan, dan akan diimplementasikan menggunakan bahasa pemrograman Python. Pada model klasifikasi, terdapat dua tahap utama yang harus dilakukan, yaitu tahap training dan tahap testing. Dataset yang telah terkumpul akan dibagi menjadi dua bagian, di mana 80% akan digunakan untuk tahap training, dan 20% sisanya akan digunakan untuk tahap testing. Pembagian dataset menjadi dua bagian, dengan menggunakan 80% untuk tahap training dan 20% sisanya untuk tahap testing, merupakan pendekatan yang umum digunakan dalam pengembangan model klasifikasi. Metode ini memastikan bahwa model memiliki sejumlah besar data untuk dipelajari selama tahap training sehingga dapat mengenali pola dan karakteristik yang mendasari pemindahan IKN dengan lebih baik. Selain itu, penggunaan dataset yang cukup besar untuk tahap training juga membantu mengurangi risiko overfitting. Overfitting terjadi ketika model terlalu spesifik dalam mempelajari data training dan mengabaikan pola umum, sehingga tidak dapat melakukan generalisasi dengan baik pada data yang belum pernah dilihat sebelumnya, seperti pada tahap testing. Dengan pembagian 80% untuk training dan 20% untuk testing, model klasifikasi akan memiliki lebih banyak peluang untuk belajar dari data latih yang mencakup berbagai variasi dan situasi yang mungkin terjadi dalam pemindahan IKN. Hal ini akan meningkatkan kemampuan model untuk mengklasifikasikan data baru secara akurat dan relevan, serta memberikan hasil klasifikasi yang lebih dapat diandalkan untuk mendukung proses pemindahan IKN yang efisien dan efektif. Selain itu pembagian ini telah banyak dilakukan dalam penelitian sentimen analisa. Proses training melibatkan penggunaan dataset yang telah dilabeli sebelumnya untuk melatih mesin agar dapat

memahami pola dan karakteristik data yang terkait dengan pemindahan IKN. Dalam tahap ini, model klasifikasi akan belajar mengenali pola yang ada dalam data latih. Setelah tahap training selesai, dilakukan proses testing menggunakan 20% dari total dataset yang tersedia. Tujuan dari tahap testing ini adalah untuk menguji tingkat akurasi dan kemampuan model klasifikasi yang telah dibangun dalam mengklasifikasikan data baru yang belum pernah dilihat sebelumnya. Melalui tahap ini, dapat dievaluasi sejauh mana model klasifikasi mampu mengklasifikasikan data dengan tingkat akurasi yang tinggi. Dengan melalui tahap training dan testing yang cermat, akan dapat dinilai sejauh mana model klasifikasi yang telah dibangun mampu memberikan hasil klasifikasi yang akurat dan relevan terhadap pemindahan IKN.

3.3.8 Evaluasi dan Analisa Performa Klasifikasi

Pada langkah evaluasi, metode yang digunakan adalah *confusion matrix*. Evaluasi ini bertujuan untuk mengukur nilai akurasi (*accuracy*), presisi (*precision*), *recall*, dan skor F1 (*F1 score*) dalam mengevaluasi performa model klasifikasi *Naive Bayes* dalam mengklasifikasikan teks yang berkaitan dengan pro dan kontra pemindahan IKN di Twitter. Dengan menganalisis hasil evaluasi ini, kita dapat membuat kesimpulan yang lebih kuat dari penelitian ini. Kesimpulan tersebut akan memberikan pemahaman yang mendalam tentang efektivitas dan keandalan model klasifikasi yang telah dibangun. Selain itu, kesimpulan ini akan memberikan wawasan tentang penggunaan algoritma klasifikasi yang tepat untuk kasus yang serupa di masa depan. Dengan demikian, evaluasi ini akan memberikan pemahaman yang lebih baik tentang performa model klasifikasi *Naive Bayes* dalam konteks pro dan kontra pemindahan IKN di Twitter, dan dapat menjadi dasar untuk rekomendasi atau pengembangan lebih lanjut dalam bidang ini.