

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

*Knowledge Discovery in Database* (KDD) atau yang dikenal dengan *Data Mining* adalah proses pengumpulan data yang bertujuan untuk menemukan pola, pengetahuan, dan juga informasi. Namun, pasti ada algoritma atau teknik yang digunakan untuk menemukan pola tersebut. *Output* yang dapat dihasilkan dari proses data mining digunakan sebagai pilihan untuk pengambilan keputusan (A'yuniyah & Reza, 2023). Tujuan dari *Knowledge Discovery in Database* (KDD) dan *data mining* adalah untuk menerapkan metode saintifik pada data mining. *Data Mining* (DM) adalah inti dari proses KDD, melibatkan kesimpulan dari algoritma yang mengeksplorasi data, mengembangkan model dan menemukan pola yang sebelumnya tidak diketahui. Tujuan KDD dan data mining adalah untuk menggali informasi tersembunyi dari sebuah basis data yang sangat besar (Oktavia et al., 2023).

Salah satu bidang khusus dalam *data mining* adalah *teks mining*, yang berarti menambang data berupa teks. Sumber data biasanya berasal dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dokumen sehingga dapat dilakukan analisis hubungan antar dokumen. Di sisi lain, teks mining dapat dilakukan dalam komputer dengan tujuan mengolah informasi yang sangat lama secara efektif, yang menghasilkan temuan informasi baru (Djamaludin et al., 2022).

Klasifikasi adalah teknik untuk mendapatkan fungsi untuk membedakan jenis kategori atau kelas data. Banyak teknik yang tersedia dalam bidang data *mining* yang dapat digunakan untuk mengolah sejumlah besar data menjadi informasi bermanfaat.

Tujuan klasifikasi adalah untuk memprediksi atau memperkirakan kelas dari data baru yang belum memiliki merek. Untuk mencapai tujuan ini, perlu dibuat cara untuk membedakan kelas data dengan metode tertentu (Paul et al., 2023). Salah satunya yaitu menggunakan *Support Vector Machine* (SVM) yang merupakan salah satu algoritma pengajaran mesin yang paling banyak digunakan untuk klasifikasi. Dalam sepuluh tahun terakhir, SVM telah berkembang menjadi alat yang kuat untuk pola klasifikasi dengan tingkat keberhasilan yang tinggi saat digunakan di berbagai industri. Karena SVM dapat menangani berbagai masalah pembelajaran, banyak komunitas pembelajaran mesin tertarik untuk mempelajari dan mengembangkannya. Metode pembelajaran mesin yang dikenal sebagai SVM bertujuan untuk menemukan *hyperplane* terbaik yang dapat memisahkan dua kelas pada ruang input. Dengan menggunakan data training, algoritma klasifikasi SVM membuat model klasifikasi. Model ini digunakan untuk memprediksi kelas data baru yang belum pernah ada sebelumnya, yang disebut data testing (Suryati et al., 2023).

Sedangkan pada penggunaan klasifikasi *Naive Bayes* adalah metode klasifikasi statistik yang digunakan untuk memprediksi kemungkinan keanggotaan kelas tertentu, menghitung kemungkinan untuk suatu hipotesis, dan menghitung kemungkinan kelas dari setiap kelompok atribut yang ada, serta menentukan kelas mana yang paling optimal (Amaliah & Dwi Nuryana, 2022).

Terdapat beberapa penelitian yang merujuk pada penelitian ini. Hasil penelitian analisis sentimen pemindahan ibu kota negara Indonesia menggunakan tiga algoritma: *naïve bayes*, *knn*, dan *random forest* didapatkan nilai akurasi metode Algoritma *Naïve Bayes Classifier* yang didapati nilai keakuratan sebesar 65.26%, Algoritma *K-Nearest Neighbor* sebesar 58.25%, serta Algoritma Random Forest sebesar 45.05% (Muliawan & Dazki, 2023).

Sedangkan penelitian dengan judul Analisis Sentimen Masyarakat Indonesia Terhadap *Metaverse* menggunakan Algoritma *Support Vector Machine* dengan jumlah data sebanyak 2504 data mendapatkan hasil akurasi tertinggi sebesar 81% (Sumayah et al., 2023).

Terdapat pula penelitian mengenai analisis sentimen terhadap pemindahan ibu kota negara Indonesia dengan membandingkan dua algoritma, yaitu *Naïve Bayes*, dan *K-Nearest Neighbor*. Dari pengujian ini, algoritma *Naïve Bayes* memiliki skor akurasi

63.09% dan algoritma *K-Nearest Neighbor* mendapatkan skor akurasi sebesar 69.23% (Pramana et al., 2023).

Penelitian selanjutnya dengan judul *Sentiment Analysis On Reviews Of Women's Tops On Shopee Marketplace Using Naive Bayes Algorithm* dapat disimpulkan hasil pengujian menggunakan *naive bayes* menunjukkan nilai akurasi sebesar 89% (Subarkah et al., 2023).

Penelitian selanjutnya dengan judul *Comparison Of Random Forest, Support Vector Machine And Naive Bayes Algorithms To Analyze Sentiment Towards Mental* dapat disimpulkan hasil pengujian model SVM menunjukkan akurasi sebesar 86.11%, model Random Forest menunjukkan akurasi sebesar 82.55%, sedangkan model Naive Bayes menunjukkan akurasi sebesar 78.19%. Oleh karena itu, dapat disimpulkan bahwa SVM memiliki performa yang paling baik dalam mengklasifikasikan tweet yang mengandung stigma kesehatan mental (Elisa & Rahman Isnain, 2024).

Penelitian lainnya juga meneliti dengan judul *Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine*. Didapati hasil pengujian model SVM menunjukkan akurasi sebesar 91.61%, sedangkan model *Naive Bayes* menunjukkan akurasi sebesar 98.80%, yang berarti model *Naive Bayes* lebih akurat dibanding model *Support Vector Machine* (Sarimole & Septian, 2024).

Melalui pelaksanaan penelitian ini, diharapkan dapat mengungkap pemahaman tentang respons masyarakat terhadap pemindahan Ibu Kota Negara Indonesia, apakah cenderung bersifat positif, negatif, atau netral. Selain itu, tujuan penelitian ini adalah untuk menemukan kata-kata yang sering muncul dalam setiap sentimen, baik itu positif, negatif, atau netral. Dengan perbandingan ini, diharapkan dapat digunakan untuk membedakan kedua metode algoritma *Support Vector Machine* dan *Naive Bayes Classifier*, yang lebih akurat dalam menilai sentimen topik pembicaraan, terutama yang berkaitan dengan pemindahan atau relokasi ibu kota negara ke IKN.

Berdasarkan uraian latar belakang diatas, maka penelitian ini akan diangkat kedalam bentuk tesis yang berjudul: **“ANALISIS SENTIMEN PUBLIK TERHADAP IBU KOTA NUSANTARA (IKN) MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES”**

## 1.2 Rumusan Masalah

Dari latar belakang yang sudah dijabarkan, maka dapat diuraikan dalam rumusan masalah sebagai berikut:

1. Bagaimana penerapan metode *Support Vector Machine* dan *Naïve Bayes* dalam menganalisis sentimen masyarakat *Tiktok* terhadap Ibu Kota Nusantara (IKN) menggunakan bahasa pemrograman *python*?
2. Bagaimana efektivitas metode *Support Vector Machine* dan *Naïve Bayes* dalam menganalisis sentimen masyarakat *Tiktok* terhadap Ibu Kota Nusantara (IKN) ?
3. Apa saja kata – kata populer yang muncul dalam perbincangan di forum komentar *Tiktok* mengenai Ibu Kota Nusantara (IKN)?

## 1.3 Batasan Masalah

Batasan masalah bertujuan untuk memberikan ruang lingkup penelitian yang akan dilakukan. Batasan masalah akan membantu dalam mengidentifikasi masalah, membatasi ruang lingkup dan menjadi gambaran terkait hal yang akan diteliti sehingga penelitian menjadi lebih terarah dan lebih fokus. Untuk membatasi ruang lingkup penelitian, maka batasan masalah dibatasi sebagai sebagai berikut:

1. Penelitian Data yang digunakan dalam penelitian ini adalah data komentar dari video yang di posting pada media sosial *tiktok* dengan dengan ketentuan 3 video teratas yang muncul dengan penonton terbanyak yang membahas tentang topik pembangunan Ibu Kota Nusantara.
2. Pengumpulan data dilakukan hanya pada media sosial *tiktok*.
3. Data yang digunakan sebanyak 1500 komentar pada 3 video tersebut.

## 1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mencapai tiga sasaran utama yang telah dirumuskan, dengan perincian sebagai berikut:

1. Menerapkan metode *Support Vector Machine* dan *Naïve Bayes* dalam menganalisis sentimen masyarakat *Tiktok* terhadap Ibu Kota Nusantara (IKN) menggunakan bahasa pemrograman *python*.

2. Mengevaluasi efektivitas metode *Support Vector Machine* dan *Naïve Bayes* dalam menganalisis sentimen masyarakat Tiktok terhadap Ibu Kota Nusantara (IKN).
3. Mengidentifikasi kata – kata populer yang muncul dalam perbincangan di Tiktok mengenai Ibu Kota Nusantara (IKN).

### **1.5 Manfaat Penelitian**

Hasil penelitian yang dilakukan diharapkan dapat dimanfaatkan baik bagi kepentingan peneliti, organisasi, maupun ilmu pengetahuan bagi peneliti selanjutnya. Manfaat penelitian ini juga dapat dikatakan jawaban atas tujuan penelitian yang dilakukan. Adapun manfaat yang akan didapatkan dalam penelitian ini adalah sebagai berikut :

1. Manfaat bagi peneliti adalah dapat mengimplementasikan ilmu yang di dapat saat belajar di Universitas Putra Indonesia YPTK Padang, melatih dalam berfikir secara sistematis dan ilmiah.
2. Manfaat bagi pembaca diharapkan dapat mengetahui fenomena apa yang sedang terjadi dan apa saja yang sedang dibahas oleh warganet mengenai ibu kota nusantara (IKN).
3. Diharapkan penulisan ini dapat menjadi referensi bagi pembaca dalam penggunaan metode *Support Vector Machine* dan *Naïve Bayes*.

### **1.6 Sistematika Penulisan**

Sistematika Penulisan merupakan gambaran secara umum tentang apa yang akan dibahas setiap bab pada tesis ini yang terdiri dari 6 (enam) bab, antara lain:

#### **BAB I : PENDAHULUAN**

Dalam bab ini, penulis menguraikan tentang latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian dan sistematika penulisan.

#### **BAB II : LANDASAN TEORI**

Bab ini menjelaskan tentang beberapa hal yang mendasar tentang masalah yang akan dibahas, yang mana nantinya akan digunakan sebagai landasan dasar dalam melakukan pemecahan masalah pada penelitian yang dilakukan.

### **BAB III : METODOLOGI PENELITIAN**

Bab ini berisikan tentang langkah atau kerangka kerja dalam melakukan analisis sentimen dengan menggunakan metode *Support Vector Machine* dan *Naïve Bayes*.

### **BAB IV : ANALISA DAN PERANCANGAN**

Bab ini memuat tentang analisa data, pengolahan data komentar tiktok mengenai opini terhadap Ibu kota nusantara. Selain itu bab ini juga memuat hasil dari penggunaan metode *Support Vector Machine* dan *Naïve Bayes*.

### **BAB V : IMPLEMENTASI DAN HASIL**

Pada bab ini berisi penerapan metode *Support Vector Machine* dan *Naïve Bayes* menggunakan Bahasa pemrograman *python* sebagai aplikasi pengklasifikasian data yang diambil dari tiktok.

### **BAB VI : PENUTUP**

Bab ini menjelaskan tentang kesimpulan dari hasil hasil penelitian dan saran-saran untuk pihak lain dalam melakukan penelitian dan pengembangan untuk selanjutnya.