

# BAB I PENDAHULUAN

## A. Latar Belakang

Saat ini penggunaan media sosial merupakan kegiatan yang tidak dapat dipisahkan dari kehidupan kita sehari-hari. Media sosial dapat digunakan sebagai sarana komunikasi, berbagi informasi dan menyampaikan pendapat. Efek penggunaan media sosial dapat memberikan efek positif atau negatif yang dapat mempengaruhi kehidupan seseorang. Contoh media sosial yang umum digunakan adalah Twitter, Instagram, Tiktok, Facebook dll, untuk kali ini yang akan diteliti adalah Twitter. Menurut laporan *We Are Social*, jumlah pengguna Twitter di Indonesia akan mencapai 18,45 juta pada tahun 2022. Jumlah ini setara dengan 4,3 % pengguna Twitter yang mencapai 436 juta di seluruh dunia[1].

Ujaran kebencian adalah ujaran langsung atau tidak langsung terhadap seseorang atau kelompok yang mengandung kebencian berdasarkan sesuatu yang melekat pada orang atau kelompok tersebut. Faktor-faktor yang sering digunakan sebagai dasar kebencian meliputi etnis, agama, disabilitas, jenis kelamin, dan orientasi seksual. Penyebaran ujaran kebencian adalah tindakan yang sangat berbahaya yang dapat memiliki efek negatif seperti diskriminasi, konflik sosial, dan bahkan genosida manusia.[2]

Seiring bertambah banyaknya pengguna Internet, semakin bertambah juga jumlah ujaran kebencian di media sosial, hal ini menimbulkan korban dan pengguna media sosial menjadi tidak nyaman.

Apa yang membuat ujaran kebencian begitu sulit untuk dikategorikan adalah bahwa hal itu tidak memiliki standar yang benar benar baku. beberapa pihak mengatakan bahwa sebuah komentar dari seseorang di media sosial sangat menyakitkan, tapi mungkin saja itu tidak masalah bagi orang lain. Tidak banyak juga korban ujaran kebencian yang melaporkan hal tersebut ke pihak yang berwajib

karena ketakutan atau ketidakpedulian yang membuat kebencian sering sekali ada di Indonesia.

Beberapa cara untuk menghentikan ujaran kebencian adalah dengan memfilter komentar dan cuitan. Indonesia sudah memiliki undang-undang ITE (UU Informasi Transaksi Elektronik) yang memungkinkan pelaku ujaran kebencian diadili. Cara termudah untuk membatasi ujaran kebencian adalah dengan memblokir cuitan yang mengandung kata atau frasa yang dianggap sebagai bagian dari ujaran kebencian. Namun cara tersebut kurang efektif karena tidak semua komentar dan cuitan menggunakan kata/frasa tersebut tidak sesuai konteks. Misalnya, penggunaan kata babi dalam konteks hewan peliharaan sangat berbeda dengan ketika kata babi digunakan untuk menyatakan kutukan atau kebencian.

*Machine Learning* diperlukan media sosial untuk melacak ujaran kebencian dan mengambil tindakan pencegahan untuk mencegah ujaran kebencian muncul di media sosial. Alasan digunakannya *Machine Learning* adalah karena memungkinkan sistem melakukan analisis untuk melihat apakah komentar atau cuitan seseorang berisi ujaran kebencian. Dataset tentang ujaran kebencian juga diperlukan sehingga mesin dapat mempelajari apa itu ujaran kebencian. Untuk meningkatkan akurasi, maka akan dilakukan pengujian dengan berbagai algoritma *Machine Learning* untuk menemukan algoritma pendeteksian ujaran kebencian yang mana yang paling akurat.

## **B. Rumusan Masalah**

1. Bagaimana membuat model untuk mendeteksi ujaran kebencian.
2. Bagaimana mengevaluasi model dan menemukan algoritma yang terbaik.

### **C. Batasan Penelitian**

1. Dataset ujaran kebencian yang dibuat berasal dari media sosial Twitter.
2. Dataset terkait dengan tweet berbahasa Indonesia.
3. Parameter dalam dataset yang akan digunakan untuk mendeteksi Ujaran kebencian adalah cuitan
4. Algoritma *Machine Learning* yang digunakan adalah Multinomial Naive Bayes, Bernoulli Naive Bayes, Logistic Regression dan SVM
5. Dataset menggunakan *feature extraction* TF-IDF, Word2vec dan N-gram word & char.

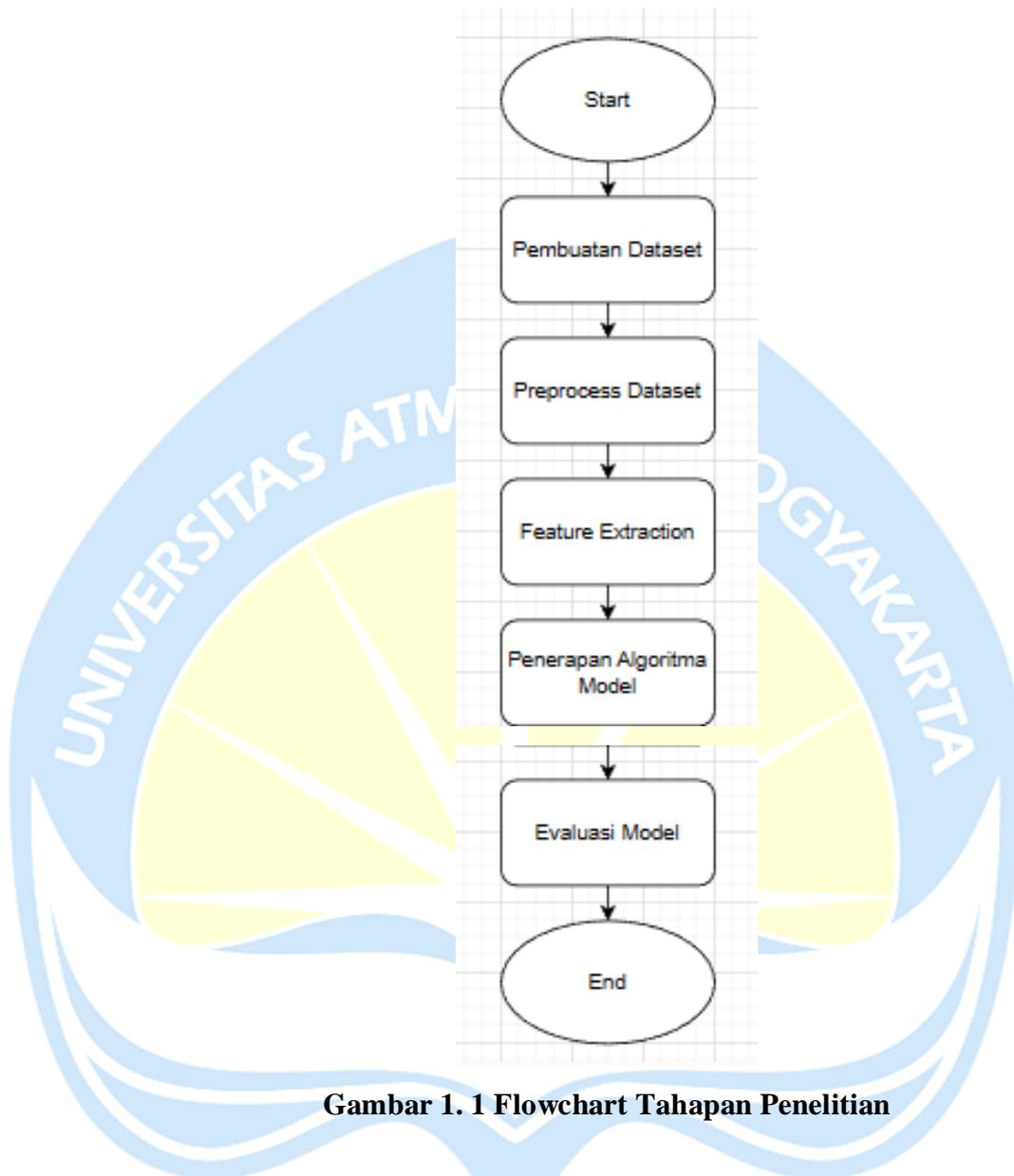
### **D. Tujuan Penelitian**

1. Membuat model untuk mendeteksi ujaran kebencian
2. Mengevaluasi model dan menemukan algoritma yang terbaik.

### **E. Metode Penelitian**

#### **1. Tahapan Penelitian**

Penelitian ini merupakan model pembelajaran mesin untuk mendeteksi ujaran kebencian dalam bahasa Indonesia. Saat membuat model, beberapa pengujian algoritma dilakukan untuk memastikan bahwa model yang dibuat dapat mendeteksi ujaran kebencian dengan tingkat ketepatan yang tinggi. Hal yang akan dilakukan adalah sebagai berikut:



**Gambar 1. 1 Flowchart Tahapan Penelitian**

### **1.1 Pembuatan Dataset**

Pembuatan dataset dilakukan dengan mengambil dataset yang telah ada dari Kaggle. Dataset tersebut akan digunakan sebagai bahan utama untuk melatih dan menguji model klasifikasi. Proses pengumpulan dataset ini merupakan langkah awal untuk memastikan ketersediaan data yang memadai dan bervariasi untuk membangun model klasifikasi

## 1.2 Pembersihan Dataset

Pembersihan dataset adalah langkah krusial dalam mempersiapkan data untuk analisis lebih lanjut. Proses pembersihan meliputi penghapusan data yang tidak relevan, penanganan nilai-nilai yang hilang atau tidak valid, normalisasi teks, dan eliminasi duplikat. Tujuannya adalah untuk memastikan integritas data dan menghasilkan dataset yang bersih, terstruktur, dan siap digunakan.

## 1.3 Pemberian Label

Dataset yang sudah dibersihkan akan diberikan label yaitu 0 untuk melambangkan cuitan tersebut tidak mengandung ujaran kebencian, sedangkan 1 melambangkan cuitan tersebut mengandung ujaran kebencian. Cuitan yang mengandung Ujaran kebencian harus mengandung unsur provokasi, diskriminasi, penghinaan, olok-an kepada suatu suku, agama, dan ras tertentu.

## 1.4 Feature extraction

Sebuah model dibuat dengan memasukkan dataset dan pembelajaran mesinnya dilatih menggunakan *Feature extraction* dan Algoritma. Model akan diuji dengan mengimplementasikan Preprocessing, *Feature extraction* dan penerapan Algoritma. *Feature extraction* yang digunakan adalah TF-IDF, n-gram dan word2vec yang akan diimplementasikan pada masing-masing algoritma.

TF-IDF bekerja dengan memberikan bobot pada setiap kata atau kalimat pada cuitan, berdasarkan frekuensi kemunculan kata atau kalimat tersebut. Semakin sering sebuah kata muncul, semakin besar bobotnya

Word2Vec mengubah kata-kata menjadi vektor yang merepresentasikan maknanya, memungkinkan kata-kata dengan makna mirip memiliki vektor yang dekat. Word2Vec bekerja dengan memberikan vektor pada setiap kata atau kalimat, menggabungkan seluruh cuitan

menjadi sebuah corpus, lalu memberikan bobot dengan rentang nilai rata-rata diantara nol sampai satu.

N-gram bekerja dengan cara memotong-motong cuitan menjadi beberapa suku kata maupun kalimat dari n kata dalam teks, di mana n mewakili jumlah kata atau item yang diambil untuk membentuk pola. N-gram berguna untuk menganalisis hubungan antar kata dalam konteks.

### **1.5 Penerapan Algoritma**

Algoritma akan diterapkan dengan cara memilih library yang ingin digunakan, lalu mengimport dataset yang telah dibuat, kemudian memisahkannya untuk membuat data training dan data testing. Pemisahan tersebut akan dilakukan dengan K-fold. Tiga algoritma digunakan: Support Vector Machines, Logistic Regression, dan Naive Bayes

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. SVM memiliki tujuan untuk membuat garis atau batas keputusan terbaik yang dapat membagi ruang n-dimensi ke dalam kelas-kelas

Naive Bayes adalah algoritma klasifikasi berdasarkan teorema Bayes. Algoritma ini mengasumsikan bahwa setiap fitur adalah independen terhadap yang lain. Algoritma ini memiliki efisiensi komputasi yang tinggi terlepas dari kesederhanaannya

Logistic Regression adalah model regresi yang digunakan untuk memprediksi probabilitas bahwa sebuah instafence akan termasuk dalam kelas tertentu. Algoritma ini menggunakan fungsi logistik untuk menghasilkan nilai probabilitas antara 0 dan 1, dan kemudian mengambil keputusan berdasarkan probabilitas tersebut.

### **1.6 Evaluasi Model**

Melakukan evaluasi terhadap model yang sudah dibuat, membandingkan suatu hasil algoritma dengan algoritma yang lain yang

bertujuan untuk menguji seberapa akuratnya algoritma dan preprocessing yang digunakan. *Feature extraction* juga akan dievaluasi dalam tahap ini dan dapat dilakukan penyesuaian ulang jika kurang akurat.

## **F. Sistematika Penelitian**

Secara sistematis isi dari laporan ini dibuat sebagai berikut:

### **BAB 1 : Pendahuluan**

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan, metodologi penelitian dan sistematika penulisan laporan.

### **BAB II : Tinjauan Pustaka**

Bab ini berisi penjelasan penelitian yang pernah dilakukan sebelumnya yang berkaitan dengan topik yang akan dibahas dan perbandingan dengan topik yang penulis bahas.

### **BAB III : Landasan Teori**

Bab ini berisi penjelasan mengenai dasar teori yang berhubungan dengan masalah yang akan dibahas.

### **BAB IV : Dataset dan Perancangan Model**

Bab ini berisi tentang dataset dari penelitian yang akan diteliti serta perancangan model yang akan digunakan dalam penelitian ini.

### **BAB V : Implementasi dan Pengujian Sistem**

Bab ini berisi penjelasan mengenai implementasi perangkat lunak yang dibuat dan pengujian perangkat lunak.

### **BAB VI : Kesimpulan dan Saran**

Bab ini berisi kesimpulan dari pembahasan.