

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Media sosial sudah sangat umum digunakan saat ini. Media sosial banyak dimanfaatkan untuk kepentingan promosi, jual-beli, media penyampaian informasi, juga sebagai media untuk para pengguna dapat mengekspresikan diri. Saat ini media sosial lebih mudah karena kemajuan teknologi memungkinkan media sosial dalam bentuk aplikasi *mobile* yang mudah untuk dapat digunakan atau diakses dimanapun dan kapanpun melalui sambungan internet [1]. Berdasarkan penelitian atau survey oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) yang dilakukan tahun 2017, sebanyak 143,26 juta jiwa dari seluruh penduduk yang ada di Indonesia yaitu sebanyak 262 juta jiwa yang menggunakan akses internet dan 87,13% diantaranya mengakses media sosial. Media sosial yang pertama muncul dan banyak diminati adalah *Facebook* di tahun 2004. Pada tahun 2006 muncul media sosial yang tidak kalah peminatnya bernama *Twitter* [2].

*Statista Research Department* di tahun 2021 menyajikan statistik pengguna *Twitter* di seluruh dunia dan negara dengan pengguna *Twitter* terbanyak pertama adalah Amerika Serikat. Amerika Serikat memiliki pengguna *Twitter* terbanyak dengan jumlah 77,75 juta jiwa dan Indonesia menduduki peringkat keenam pengguna *Twitter* paling banyak yang ada di dunia dengan jumlah total 17,5 juta jiwa [3]. *Twitter* merupakan media sosial yang dapat mengunggah kalimat atau paragraf dengan istilah *Tweet*. *Tweet* sering dimanfaatkan sebagai sarana menyampaikan opini atau informasi tentang berbagai topik yang sangat beragam [4].

Portal berita CNN Indonesia mengemukakan pada 14 Oktober 2021 bahwa tagar atau *hashtag* “Percuma Lapor Polisi” menjadi perbincangan hangat. Tagar tersebut ramai ketika ada seorang ibu di Sulawesi Selatan melaporkan pelecehan seksual suaminya terhadap tiga orang anaknya. Warga Indonesia yang aktif bersosial media turut bersimpati karena kasus tersebut dihentikan dengan alasan bukti yang tidak kuat. Kasus tersebut membuka banyak pengalaman mengecewakan lainnya yang dituangkan melalui *Tweet* pada sosial media *Twitter* dengan

menggunakan tagar “Percuma Lapor Polisi”. *Tweet* berisi kekecewaan masyarakat terhadap polisi tersebut terangkum dalam tagar “Percuma Lapor Polisi” [5].

*Hashtag* “Percuma Lapor Polisi” dapat dilakukan analisis sentimen untuk mengidentifikasi positif negatifnya suatu *Tweet*. Analisis Sentimen dilakukan untuk mendapatkan gambaran persepsi dari masyarakat atau pengguna *Twitter* dengan mengelompokkan persepsi tersebut kedalam kelas sentimen positif atau sentiment negatif [6]. Adapun metode dari analisis sentimen yang umum digunakan adalah *Naïve Bayes Classifier* (NBC), K-Nearest Neighbor (KNN), *Support Vector Machine* (SVM) [7]. *Naïve Bayes Classifier* banyak digunakan karena kemudahan penggunaannya [8], KNN merupakan metode dengan keefektifan tinggi untuk permasalahan klasifikasi serta waktu pemrosesan yang lebih cepat, SVM memiliki nilai akurasi yang lebih besar. Pada penelitian yang akan dilakukan, menggunakan metode *Support Vector Machine* (SVM), salah satu metode yang tepat untuk *text classification* dengan pencarian *hyperplane* terbaik [9]. *Hyperplane* memiliki fungsi untuk memisahkan dua kelompok data [7], dengan adanya *hyperplane* pada SVM membuat metode SVM memiliki generalitas tinggi dengan akurasi terbaik dibandingkan metode lainnya [9].

Penelitian Ghulam Asrofi Buntoro pada tahun 2016 membandingkan metode *Naïve Bayes* dengan SVM, hasil yang diperoleh terbukti nilai akurasi tertinggi adalah dengan metode SVM [8]. Adapun penelitian yang membandingkan metode KNN dengan metode SVM dalam penelitian analisis sentimen *Twitter*, yang diteliti oleh Muhammad Rangga dan Mardhiya Hayaty di tahun 2019. Kesimpulan dari penelitian Muhammad Rangga dan Mardhiya adalah SVM memiliki nilai akurasi yang lebih unggul sedangkan keunggulan KNN adalah waktu proses yang lebih cepat [9].

Penelitian terdahulu telah membuktikan bahwa *accuracy* dari SVM lebih tinggi dibandingkan dengan NBC dan KNN, untuk itu peneliti memutuskan untuk menggunakan SVM. Penggunaan metode dengan *accuracy* tinggi untuk mendapatkan persentase sentimen dengan prediksi yang tepat, sehingga mendapatkan analisis terbaik. Tahapan dari penelitian ini secara garis besar adalah pengumpulan data, *pre-processing*, pembobotan kata, pembagian data, klasifikasi

menggunakan SVM, dan evaluasi hasil sentimen [10].

*Pre-processing* adalah proses untuk menyiapkan data tidak terstruktur menjadi terstruktur. Tahapan *pre-processing* salah satunya ialah *normalization* yaitu mengubah kata tidak baku menjadi kata baku dengan konteks yang sesuai. Terdapat 85% *Tweet* pada *Twitter* menggunakan bahasa yang tidak baku atau *slang*, yang kemudian dalam Bahasa Indonesia bahasa tidak baku tersebut dikenal dengan bahasa “alay”. Bahasa atau kata tidak baku umumnya kurang sesuai untuk dapat diproses menggunakan *Natural Language Processing* (NLP) seperti *Information Retrieval*, *Machine Translation* dan *Opinion Mining* karena ketidakteraturan dalam corak bahasanya. Penelitian ini terbagi menjadi dua skenario untuk menunjukkan pengaruh penggunaan proses *normalization* terhadap *accuracy* metode SVM [11].

## 1.2 Rumusan Masalah

Rumusan yang diambil pada latar belakang penelitian, dirangkum menjadi rumusan masalah seperti di bawah ini:

1. Belum terdapat penelitian mengenai analisis sentimen terhadap *hashtag* “Percuma Lapor Polisi” di *Twitter* menggunakan metode SVM.
2. Belum diketahui nilai dari *accuracy* dari penggunaan metode SVM terhadap analisis sentimen *hashtag* “Percuma Lapor Polisi” di *Twitter*.

## 1.3 Batasan Masalah

Batasan masalah pada penelitian ini dibuat agar fokus penelitian jelas dan terarah. Adapun batasan masalah dalam penelitian ini ditentukan:

1. Dataset berjumlah 3.533 *Tweet* yang terkumpul dalam *hashtag* “Percuma Lapor Polisi” di *Twitter* semenjak tanggal 11 November hingga 17 November 2021. Dataset diperoleh menggunakan *Twitter API*.
2. Proses analisis sentimen menggunakan *Support Vector Machine* (SVM).
3. Langkah tahapan *pre-processing* berupa *case folding*, *cleaning*, *normalization*, *tokenizing*, *stopword removal*, *stemming* [12].

4. Proses *normalization* menggunakan kamus dari hasil penelitian [13].
5. Proses pelabelan pada penelitian ini menggunakan *lexicon based-features*.
6. Penelitian dilakukan dengan *Jupyter Notebook* sebagai *tools* untuk menulis serta menyimpan kode dan alur program, *Python* sebagai bahasa pemrograman yang digunakan.
7. Output penelitian berupa analisis sentimen, persentase sentimen positif dan negatif, nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

#### **1.4 Tujuan Penelitian**

Penelitian ini dilakukan memiliki tujuan seperti di bawah ini:

1. Menentukan hasil persentase dari analisis sentimen dari *hashtag* “Percuma Lapor Polisi” di *Twitter* dengan metode SVM.
2. Mengetahui nilai *accuracy*, *precision*, *recall*, dan *f1-score* SVM metode dalam penelitian analisis sentimen *hashtag* “Percuma Lapor Polisi” di *Twitter*.

#### **1.5 Manfaat Penelitian**

Manfaat penelitian yang dapat diambil dari penelitian tugas akhir ini disimpulkan sebagai berikut:

1. Mengetahui perbandingan persentase sentimen positif dan negatif.
2. Penelitian dapat menentukan nilai *accuracy*, *precision*, *recall*, dan *f1-score* metode SVM untuk analisis sentimen pada *hashtag* “Percuma Lapor Polisi” di *Twitter*.