

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Sepakbola merupakan salah satu olahraga dengan jumlah penggemar terbanyak di Indonesia. Berdasarkan *survey* yang dilakukan oleh Skala Survei Indonesia (SSI) pada tahun 2015 membuktikan, bahwa 90.8% publik Indonesia tahu olahraga sepakbola. Dari yang tahu ini, 47.6% menyukainya [1]. PSSI merupakan induk organisasi yang bertanggungjawab akan segala hal yang berkaitan dengan sepakbola Indonesia. Sebagai induk organisasi kinerja PSSI dalam mengelola sepakbola Indonesia menjadi sorotan publik. Setiap kinerja PSSI tidak lepas dari pro dan kontra dari masyarakat. Penyampaian pendapat masyarakat kerap diunggah di media sosial salah satunya Twitter.

Untuk mengetahui respon masyarakat terhadap PSSI perlu dilakukan sentimen analisis. Sentimen analisis merupakan ilmu yang berguna untuk menganalisis pendapat seseorang, sentimen seseorang, evaluasi seseorang, sikap seseorang dan emosi seseorang ke dalam bahasa tertulis [2]. Analisis sentimen pada Twitter cukup populer dilakukan oleh banyak peneliti di dunia [3], karena Twitter merupakan salah satu media sosial yang populer dan tempat yang bebas untuk berekspresi seperti memberikan opini terhadap suatu hal. Sentimen analisis dilakukan dengan cara mengambil data tweet. Kemudian, dilakukan analisis sentimen pada data untuk memperoleh informasi apakah tweet tersebut mengandung unsur positif atau negatif.

Dalam melakukan analisis sentimen dilakukan pendekatan menggunakan teknik klasifikasi. Teknik klasifikasi yang umum digunakan untuk analisis sentimen, seperti Naive Bayes, Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN) [4]. Diantara teknik teknik klasifikasi tersebut, SVM merupakan teknik yang paling sering untuk melakukan analisis sentimen. SVM adalah sebuah klasifikasi algoritma linear yang memiliki peranan utama menentukan pemisah linear yang paling utama di ruang pencarian terbaik yang dapat memisahkan dua kelas berbeda [5]. Kelebihan SVM adalah mampu mengidentifikasi hyperplane yang terpisah yang memaksimalkan margin antara dua kelas [5][6].

Akan tetapi, Masalah utama saat melakukan analisis sentimen adalah peningkatan dimensi data yang tidak diinginkan. Untuk mengurangi dimensi data, perlu dilakukan preprocessing data [7]. *Preprocessing* dilakukan sebelum melakukan pembobotan data dan klasifikasi. Preprocessing data bertujuan untuk menghilangkan data yang tidak lengkap, data yang mengandung noise, dan data yang tidak konsisten[8]. *Preprocessing* data yang sering kali digunakan adalah *Case folding*, *stopwords removal*, *Stemming*, *Tokenizing*, dan *Tokenizing*. Seringkali ekspresi diungkapkan dengan sebuah simbol emoticon dan kata slang dalam twitter menyebabkan perlu adanya pengkonversian ke dalam bentuk string yang dapat diartikan maknanya[9].

Dari uraian tersebut maka diperlukan adanya penanganan yang ekstra pada saat tahap preprocessing. Pada penelitian ini akan menggunakan algoritma Support Vector Machine pada tahap klasifikasi dengan menambah *convert emoticon* dan

*slang word* pada tahap *preprocessing* untuk melakukan analisis sentimen terhadap kepengurusan PSSI menggunakan data yang diperoleh dari twitter.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan di atas, maka rumusan masalah penelitian ini adalah

1. Bagaimana gambaran umum tweet masyarakat terhadap kinerja PSSI dalam mengelola sepakbola Indonesia ?
2. Bagaimana kinerja metode Support Vector Machine setelah dan sebelum penambahan *convert emoticon* dan *slang word* pada tahap *preprocessing* dalam mengklasifikasikan tweet masyarakat terhadap kinerja PSSI dalam mengelola sepakbola Indonesia?

## 1.3 Batasan Masalah

Batasan masalah pada penelitian ini antara lain:

1. Penelitian ini hanya mengukur tingkat akurasi Support Vector Machine setelah dan sebelum diterapkan *convert emoticon* dan *slang word* pada tahap *preprocessing*.
2. Data sentimen mengenai sepakbola Indonesia diperoleh dari media sosial twitter dengan kata kunci PSSI.
3. Jumlah data yang dipakai pada penelitian ini berjumlah 1000 data tweets, dimana 80% digunakan untuk data latih dan 20% digunakan untuk data uji, yang diambil dari tanggal 1 Maret 2019 – 30 Juni 2019.

4. Data yang diambil dari twitter merupakan data tweet berbahasa Indonesia.
5. Data sentimen diklasifikasikan menjadi 2 kelas yaitu positif dan negatif.

#### 1.4 Maksud dan Tujuan Penelitian

1. Penelitian ini bertujuan untuk membandingkan antara akurasi metode support vector machine sebelum dan setelah ditambahkan *convert emoticon* dan *slangword* pada tahap preprocessing.
2. Mengetahui gambaran umum sentimen masyarakat terhadap PSSI dalam mengelola sepakbola Indonesia di media sosial Twitter.

#### 1.5 Manfaat Penelitian

Manfaat Penelitian ini adalah sebagai berikut:

1. Hasil penelitian dapat dijadikan referensi bagi para peneliti berikutnya yang akan membahas mengenai analisis sentimen (*opinion mining*) terutama pada jejaring sosial Twitter.
2. Memperoleh informasi dan menambah pengetahuan mengenai mengelola data teks untuk melakukan sentimen analisis.
3. Memperoleh informasi tentang sentimen sepakbola indonesia.
4. Meningkatkan kemampuan dalam text mining dan analisis.

## 1.6 Metode Penelitian

### 1. Pengumpulan Data

Peneliti menjabarkan cara-cara memperoleh data-data yang digunakan untuk kebutuhan penelitian.

#### a. Studi Literatur

Studi literatur dilakukan untuk mempelajari penelitian-penelitian dengan tema sejenis yang pernah dilakukan sebelumnya. Penelitian-penelitian ini dijadikan referensi pada penelitian ini. Sumber yang dipakai dapat berupa buku, jurnal, *paper*, *website*, dan lain-lain. Pada penelitian ini, referensi yang dicari adalah mengenai analisis sentimen dan metode klasifikasi SVM, *Preprocessing*, *Convert Emoticon*, *Slangword*.

#### b. Pengumpulan Data Tweet

Data *input* yang digunakan pada penelitian ini berupa *tweet* dengan kata kunci "PSSF". Data yang sudah dikumpulkan disimpan dalam format \*.csv.

### 2. Eksperimen

#### a. Alat dan Bahan

Penelitian ini menggunakan bahasa Python 3.7 dengan IDE JetBrains Pycharm Community Edition 2018 ver. 2.3, dengan *library* Scikit-learn untuk *machine learning* dan *library* NLTK untuk *preprocessing data*. Sedangkan bahan untuk penelitian ini

adalah data *tweet* yang diambil dari Twitter menggunakan *GetOldTweets3* untuk pengumpulan data tweet.

b. Tahap penelitian

Pada tahap ini dilakukan perancangan dengan menerapkan *preprocessing data*, seperti melakukan *cleaning*, *tokenization*, *penghapusan stopwords*, *stemming*, *convert emoticon*, *slangword* serta memberi bobot pada kata menggunakan *TF-IDF*, dan juga klasifikasi sentimen menggunakan *SVM*. Serta melakukan pengujian model menggunakan *confusion matrix* dan *k-fold cross validation*.

3. Hasil dan Pembahasan

Pada tahap ini memaparkan dan membahas hasil dari eksperimen yang telah dilakukan. Pembahasan yang dilakukan mengenai akurasi model serta pengaruh penerapan teknik *convert emoticon* dan *slangword*.

4. Penulisan Laporan

Hasil penelitian kemudian dituliskan ke dalam sebuah laporan. Dalam penulisan laporan penelitian akan ditarik pula suatu kesimpulan berdasarkan hasil pembahasan dan pengujian, serta saran untuk penelitian berikutnya.

## 1.7 Sistematika Penulisan

Rincian sistematika penulisan yang digunakan dalam penelitian ini adalah sebagai berikut:

## BAB I PENDAHULUAN

Memuat latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan.

## BAB II LANDASAN TEORI

Menguraikan kajian pustaka dan dasar teori yang berkaitan dengan penelitian, seperti referensi-referensi dengan topik yang sama serta definisi yang berhubungan dengan ilmu dan permasalahan yang sedang diteliti.

## BAB III METODOLOGI PENELITIAN

Menguraikan alat dan bahan yang dipakai dalam penelitian, alur penelitian, dan strategi pemecahan masalah.

## BAB IV HASIL DAN PEMBAHASAN

Membahas tentang implementasi dari cara kerja analisis sentimen mulai dari *pre-processing* data, TF-IDF, SVM dan hasil yang didapat dari penelitian serta pembahasannya.

## BAB V PENUTUP

Bab ini menyajikan kesimpulan dari hasil penelitian yang didapat dan juga saran untuk penelitian selanjutnya.

## DAFTAR PUSTAKA