

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Internet pada masa sekarang ini berkembang dengan pesat. Bisa dikatakan bahwa hampir sebagian besar masyarakat telah mampu mengoperasikan internet baik melalui ponsel pintar maupun komputer pribadi. Kemajuan internet ini merupakan salah satu kontribusi dari globalisasi yang menghasilkan perubahan pola hidup masyarakat. Pada zaman sebelum penggunaan internet sudah umum, Masyarakat hanya dapat memperoleh informasi melalui media cetak, radio serta televisi. Sekarang masyarakat lebih cenderung menggunakan gawai untuk mengakses media sosial dalam rangka untuk berkomunikasi antar sesama maupun memperoleh informasi secara instan.

Penggunaan media sosial di Indonesia merupakan salah satu yang terbesar di dunia. Menurut data hasil riset Wearesosial Hootsuite yang, pengguna internet di Indonesia pada awal tahun 2019 mencapai 150 juta atau sekitar 56% dari total populasi. 130 juta diantaranya merupakan pengguna media sosial [1].

Pengguna situs media sosial Youtube di Indonesia mencakup 88% dari total pengguna. Oleh karena itu, Youtube dapat disebut sebagai media sosial dengan jumlah pengguna terbanyak [2]. Berkembangnya media social di Indonesia ini menuai pro dan kontra di kalangan masyarakat. Salah satu efek buruk dari perkembangan tersebut adalah *cyberbullying*. Hasil survey APJII (Asosiasi Penyelenggaraan Jasa Internet Indonesia) memperlihatkan bahwa setidaknya 49% dari *sample* mengaku pernah dirundung . Hal ini membuat *cyberbullying* menjadi

suatu fenomena yang cukup mengkhawatirkan bagi perkembangan manusia di Indonesia dan juga dunia. *Cyberbullying* dapat menyebabkan efek buruk bagi korban mulai dari perasaan malu hingga berujung dengan bunuh diri [3]. Untuk menghindari hal – hal tersebut, dibutuhkan langkah preventif seperti penghapusan komentar perundungan terhadap orang lain pada media social. Sebelum dilakukan penghapusan, komentar perundungan dan bukan perundungan harus diklasifikasi terlebih dahulu. Metode yang dapat digunakan untuk mengenali hal tersebut adalah *Machine Learning* seperti yang telah dilakukan pada sebuah penelitian perbandingan beberapa algoritma[4].

Dalam penelitian *Perbandingan Kinerja Algoritma K-Nearest Neighbor, Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Tingkah Laku Bully pada Aplikasi Whatsapp*, KNN, *Naïve Bayes* dan *Support Vector Machine* dibandingkan. Hasil penelitian tersebut menunjukkan bahwa *Support Vector Machine* merupakan algoritma paling optimal untuk klasifikasi teks [4]. Selain itu, penelitian *Advantage and drawback of support vector machine functionality* menyimpulkan bahwa SVM memiliki performa terbaik dalam penerapan pada dataset besar [5]. Dari pernyataan dari penelitian diatas dapat ditarik kesimpulan bahwa SVM merupakan opsi terbaik dalam klasifikasi komentar youtube. Kolom komentar di youtube pada video dengan *view* lebih dari satu juta cenderung memiliki komentar hingga ribuan.

Berdasarkan uraian yang telah diungkapkan oleh penulis, penulis ingin melakukan penelitian terhadap filtrasi komentar social media menggunakan

*Support Vector Machine* dengan beberapa fungsi kernel. Penelitian ini diharapkan menghasilkan sistem filtrasi berbasis SVM yang memiliki akurasi yang baik.

## 1.2 Rumusan Masalah

1. Bagaimana melakukan klasifikasi serta filtrasi komentar *cyberbullying* pada kolom komentar video youtube dengan model *Support Vector Machine*?
2. Bagaimana kinerja dari SVM beserta fungsi *kernel* dalam klasifikasi komentar *cyberbullying* komentar youtube?
3. Bagaimanakah hasil filtrasi dari implementasi *Support Vector machine* terhadap comment youtube ?

## 1.3 Batasan Masalah

1. Algoritma yang diterapkan pada data yaitu *Support Vector Machine* dengan metode kernel *Polynomial*, *RBF*, serta *Sigmoid*.
2. Parameter yang digunakan dalam mengukur performa dari algoritma adalah *Confusion Matrix*( *precision*, *recall*, *accuracy* dan *f-measure*)
3. Bahasa pemrograman yang digunakan yaitu *Python* dengan library *sci-kit learn* (*sklearn*) , framework *flask-bootstrap*, *Youtube API*, dan *python sastrawi*.
4. Proses *preprocessing* terdiri atas *Data Cleaning*, *Case Folding*, *Stopword Removal*, *Stemming*, *Normalization*, dan *Tokenization*.
5. Metode *stemming* yang digunakan yaitu *stemming Nazief* dan *Adriani*
6. Nilai *C* yang digunakan dalam fungsi *Polynomial* dan *RBF* adalah 0.1, 0.6, 1, dan 2

7. Nilai  $\gamma$  untuk RBF dan *Sigmoid* yaitu 0.1, 0.3, 1, 2, 2.5, dan 3
8. Nilai  $r$  untuk *Sigmoid* yaitu 0.125, 0.5, 1, 2, dan 10

#### 1.4 Maksud dan tujuan penelitian

1. Mengimplementasikan Support Vector Machine dengan fungsi *Kernel* untuk Klasifikasi komentar *Cyberbullying* pada kolom komentar youtube.
2. Mengukur kinerja dari ketiga kernel tersebut untuk mengetahui fungsi kernel SVM yang memiliki performa yang paling baik dalam kasus Klasifikasi komentar dan juga klasifikasi teks pada umumnya.

#### 1.5 Manfaat Penelitian

1. Memberikan wawasan dalam pengimplementasian SVM dengan Metode Kernel Polynomial, RBF, dan Sigmoid untuk klasifikasi komentar *cyberbullying* pada kolom komentar youtube.
2. Dapat mengetahui fungsi kernel terbaik dalam klasifikasi komentar.
3. Sebagai rujukan untuk penelitian selanjutnya mengenai klasifikasi komentar youtube dan klasifikasi teks pada umumnya.
4. Pengetahuan yang ada pada penelitian ini dapat digunakan untuk membangun sebuah sistem menghapus spam dan juga *cyberbullying*.
5. Aplikasi yang dihasilkan dapat digunakan untuk sekedar penyaring komentar *cyberbullying* guna menghindari efek buruk dari komentar – komentar yang ada

## 1.6 Metodologi Penelitian

Metode penelitian pada dasarnya merupakan suatu cara ilmiah yang bertujuan untuk mendapatkan data sesuai dengan tujuan dari penelitian yang dilakukan. Dalam rangka untuk mendapatkan data yang diinginkan, dibutuhkan metode-metode yang relevan dengan tujuan penelitian. Pada penelitian yang penulis lakukan menggunakan pendekatan kuantitatif. Pendekatan kualitatif melibatkan generasi data dalam bentuk kuantitatif (atau angka). Menurut (Kothari 1990:5), pendekatan kuantitatif dapat diklasifikasi menjadi 3 subbagian yaitu *inferential*, *experimental*, serta pendekatan *simulation*. Dari ketiga subbagian tersebut, penulis memanfaatkan pendekatan *experimental* atau eksperimentasi dalam bahasa Indonesia.

### 1.6.1 Sumber Data

Untuk menunjang penelitian ini digunakan dua jenis data yaitu data primer dan data sekunder.

#### 1. Data Primer

Data ini diperoleh dari hasil pencarian kolom komentar youtube beberapa video berbahasa indonesia secara manual.

#### 2. Data Sekunder

Data ini diperoleh dari studi literature seperti jurnal, buku, dataset, serta *thesis*.



## 1.6.2 Tahapan Penelitian

### 1. Studi Literatur

Pada tahapan ini, penulis melakukan pencarian, penggalian serta pendalaman terhadap informasi – informasi yang berkaitan dengan penelitian. Informasi tersebut didapatkan dari berbagai sumber diantaranya e-jurnal, buku pembelajaran, skripsi dan *website*.

### 2. Analisis dan Perancangan Sistem

Pada tahapan ini dilakukan berbagai analisa terhadap metode – metode serta *framework* aplikasi yang telah diperoleh dari tahapan studi literatur. Dari hasil analisa tersebut, sistem kemudian dirancang. Adapun bagian yang dirancang meliputi interface, data dan sistem filtrasi secara keseluruhan

### 3. Implementasi dan Pengujian

Pada tahapan ini dilakukan implementasi rancangan sistem menjadi sebuah aplikasi berbasis web dengan memanfaatkan bahasa pemrograman python.

Kemudian pada tahapan ini juga dilakukan pengujian. Pengujian dilakukan dengan memanipulasi parameter – parameter pada algoritma *Support Vector Machine*. Parameter – parameter tersebut adalah  $C$ ,  $\gamma$  dan  $r$ .

## 1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan pada penelitian ini mengikuti sistematika yang ditetapkan oleh Universitas Amikom Yogyakarta

## BAB I PENDAHULUAN

Bagian ini menguraikan latar belakang permasalahan, rumusan masalah, batasan masalah, tujuan dari penelitian, manfaat penelitian, metodologi penelitian serta sistematika penulisan

## BAB II LANDASAN TEORI

Bab landasan teori memuat tentang teori – teori pendukung untuk penelitian ini. Bagian ini terdiri dari tinjauan pustaka, penjelasan lebih lanjut mengenai *cyberbullying*, penjelasan mengenai bahasa pemrograman python, penjelasan singkat terhadap framework flask serta librari sci-kit learn, Flowchart, penjelasan terhadap klasifikasi, uraian algoritma *Support Vector Machine* beserta jenis-jenis metode *kernel*, *Confusion Matrix* (*precision*, *accuracy*, *recall* dan *f-measure*), Flowchart dan DFL.

## BAB III ANALISIS DAN PERANCANGAN

Bab ini menguraikan tentang analisis data, analisis *pre-processing*, analisis *kernel linear*, *polynomial*, *rbf* dan *sigmoid*, analisis rancangan perangkat lunak yang menjadi alat untuk eksperimentasi dan analisis proses kerja.

## BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini menguraikan hasil implementasi dari rancangan perangkat lunak dan hasil pengujian *support vector machine* dengan metode kernel pada data penyakit jantung, pengukuran performa dari algoritma terhadap data yang disajikan dengan *Covariance Matrix*

## BAB V PENUTUP

Bab ini menguraikan poin-poin kesimpulan yang dihasilkan dari penelitian ini berdasarkan bab-bab sebelumnya

