

BAB I

PENDAHULUAN

1.1 Latar Belakang

Instagram merupakan aplikasi media sosial untuk berbagi foto dan video yang memungkinkan pengguna mengunggah foto dan video dan menggunakan beberapa filter untuk foto ataupun video yang akan diunggah ke Instagram. Instagram bisa diakses melalui web dan mobile. Aplikasi Instagram dikembangkan untuk smartphone, baik platform iOS, Android, ataupun Windows Phone. Instagram dapat di download melalui smartphone secara gratis. Instagram berbasis web bisa diakses di <http://www.instagram.com>. Akan tetapi melalui web hanya bisa melihat foto dan video dari akun Instagram lain yang mereka ikuti, melihat dan mengedit profil pengguna aktif, dan melihat aktivitas pengguna Instagram lainnya, akan tetapi pengguna tidak dapat mengunggah foto/gambar dari web, melainkan harus menggunakan aplikasi pada smartphone.

Instagram merupakan situs media sosial yang semakin banyak digunakan oleh para artis/aktor Indonesia. Para pengguna Instagram biasanya mengunggah foto ataupun video kegiatan mereka kemudian masing-masing foto atau video akan diberi *caption*, *tagging* ke akun Instagram lainnya, bahkan lokasi tempat kegiatan foto ataupun video diunggah, dan menerapkan *filter edit* yang sudah disediakan oleh Instagram. Dan tak lupa untuk menambahkan *hashtag* tertentu agar foto atau video tersebut semakin banyak dilihat oleh orang lain.

Salah satu hal yang menyebabkan Instagram banyak digunakan adalah kemudahannya untuk mengunggah foto ataupun video langsung dari *smartphone*.

Pengguna Instagram kebanyakan adalah orang muda yang menyukai selfie, yang memiliki hobby dengan photography dan videography. Namun disamping kelebihan tersebut terdapat beberapa kekurangan yang cukup mengganggu yaitu banyaknya komentar yang dapat dikategorikan sebagai komentar spam pada suatu postingan foto maupun video yang sudah diunggah pada Instagram. Komentar spam akan semakin banyak dijumpai pada beberapa akun Instagram artis/orang terkenal karena followers-nya yang banyak.

Spam merupakan informasi yang dikirimkan ke beberapa pengguna dimana informasi yang dikirim tidak berhubungan dengan penerima informasi tersebut [1]. Tujuan pembuatan spam sendiri adalah promosi dan iklan yang dapat berupa link, berita, email, tulisan dan bahkan komentar [1]. *Spam* pada media sosial merupakan komentar yang berisi konten yang dikirim oleh *spammer* menggunakan fitur comment, dimana konten yang dikirimkan berisi informasi yang tidak relevan dengan pengguna yang menerima komentar tersebut [2]. Munculnya spam membuat pengguna Instagram merasa terganggu, khususnya para artis yang memiliki jumlah follower yang banyak dan menjadikan sasaran *spammer*. Bahkan beberapa pengguna pernah menjadi korban penipuan yang mengakibatkan terjadinya kerugian [3].

Adapun solusi untuk menghadapi komentar spam sudah ada, namun semuanya dilakukan secara manual. Pengguna Instagram dapat menghapus secara manual komentar spam tersebut akan tetapi hal itu akan membutuhkan tenaga dan waktu yang tidak efisien dalam memeriksa satu persatu komentar. Cara lain selain dihapus secara manual Instagram menyediakan fitur untuk melaporkan semua

komentar sebagai spam tetapi dilakukan secara manual juga, tetap saja tidak efisien karena dilakukan satu persatu. Ada lagi cara lain untuk meminimalisir komentar spam yaitu dengan mengubah akun Instagram menjadi *private*. Tentu hal ini sangat sulit dilakukan bagi akun artis/aktor, karena jika akun Instagram dibuat menjadi *private* tidak bisa langsung di *follow* oleh akun lain. Cara terakhir yang bisa dilakukan adalah menggunakan pengaturan mengaktifkan fitur Instagram untuk menghapus komentar yang mengandung kata-kata tertentu yang dimasukkan sendiri oleh pengguna yang dianggap sebagai komentar *spam*.

Berdasarkan latar belakang tersebut pada penelitian ini akan dibangun suatu sistem yang dapat mengklasifikasikan komentar spam dengan mengambil data *training* komentar-komentar *spam* pada instagram beberapa artis dan orang nomor satu di Indonesia. Terdapat beberapa metode untuk klasifikasi seperti *Naive Bayes*, *K-Nearest Neighbour*, *Decision Tree* dan *Random Forest*. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine*. *Support Vector Machine (SVM)* adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Teknik SVM digunakan untuk menemukan fungsi pemisah (*hyperplane*) yang optimal yang bisa memisahkan dua set data dari dua kelas yang berbeda. Penggunaan teknik *machine learning* tersebut, karena performansinya yang meyakinkan dalam memprediksi kelas suatu data baru. Metode ini dipilih karena menghasilkan akurasi yang lebih baik.

Penelitian ini akan mengumpulkan data berupa status foto dan komentar dari 10 akun artis/aktor di Indonesia yang memiliki *follower* lebih dari 10 juta. Mengambil 10 akun, setiap akun akan diambil 100 komentar dari 5 *posting* terbaru.

Dataset yang terbentuk akan dilakukan pelabelan secara manual untuk dapat digunakan sebagai data latih sistem machine learning deteksi komentar spam menggunakan *Support Vector Machine*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, maka perlu dirumuskan suatu masalah yang akan dipecahkan/diselesaikan pada penelitian/perancangan ini adalah bagaimana klasifikasi komentar *spam* yang ada di Instagram dengan menggunakan metode *Support Vector Machine*.

1. Bagaimana cara mengklasifikasi komentar *spam* dan *ham* (not spam) dengan menggunakan metode *Support Vector Machine*.
2. Mencari nilai akurasi dari algoritma *Support Vector Machine* dalam klasifikasi komentar *spam* pada Instagram.

1.3 Batasan Masalah

Berdasarkan rumusan masalah yang telah disampaikan diatas, terdapat beberapa pembatasan masalah pada beberapa pokok bahasan, yaitu:

1. Penelitian ini hanya mengambil 5000 komentar yang berasal dari 10 akun artis.
2. Bahasa yang di ambil hanyalah Bahasa Indonesia.
3. Klasifikasi menggunakan *Support Vector Machine kernel RBF*.
4. Pengklasifikasian menggunakan bahasa Python.
5. Untuk proses text mining menggunakan bahasa PHP.

1.4 Maksud dan Tujuan Penelitian

Tujuan utama dari penelitian ini adalah untuk mengetahui kinerja dari penerapan metode *Support Vector Machine kernel RBF* dalam klasifikasi komentar *spam* pada Instagram.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah :

1. Untuk mengetahui hasil akurasi dari klasifikasi dokumen pada komentar *spam* Instagram menggunakan metode *Support Vector Machine*.
2. Membangun sistem pengklasifikasian menggunakan *Support Vector Machine* yang nantinya diharapkan dapat memilah komentar *spam* dalam komentar agar informasi yang diperoleh lebih akurat dan terpercaya.

1.6 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah metode eksperimental dan dilaksanakan dalam tahapan-tahapan berikut:

1.6.1 Metode Pengumpulan Data

Pada tahapan pengumpulan data ini dilakukan pengambilan data pelatihan dari Instagram menggunakan Instagram API (*Grabber*) berupa data dari 10 akun 5 postingan terbaru dan 100 komentar setiap postingan. Setelah data diperoleh maka akan dilakukan pelabelan data secara manual dan kemudian dilakukan implementasi deteksi *spam* menggunakan algoritma yang bisa digunakan untuk klasifikasi teks yaitu *Support Vector Machine*.

1.6.2 Metode Analisis

Pada tahap ini dilakukan beberapa bagian yaitu pengumpulan *dataset* yang dilakukan dengan cara *crawling* pada penyedia layanan viewer media social Instagram, analisis karakteristik spam yang ada, perancangan metode klasifikasi, serta merancang cara mengimplementasikan komponen-komponen sistem seperti antarmuka untuk pembuatan data latih pada algoritma *Support Vector Machine*.

1.6.3 Metode Perancangan

Pada tahap ini dilakukan perancangan cara mengimplementasikan komponen-komponen sistem seperti antar muka untuk pembuatan data latih pada algoritma *Support Vector Machine*.

1.6.4 Implementasi

Pada tahap ini akan dilakukan pembuatan kode program sampai proses klasifikasi sentimen dokumen.

1.6.5 Pengujian

Pada tahap ini dilakukan perhitungan akurasi dari implementasi yang telah dilakukan. Metode yang digunakan yaitu *K-Fold Cross Validation*. Dalam penelitian ini, tahap pengujian menggunakan *10-Fold* sebagai parameter untuk melakukan *Cross Validation*. Pengujian yang dilakukan adalah perbandingan akurasi klasifikasi yang dihasilkan SVM. Tahap yang dibandingkan adalah dengan

membandingkan hasil klasifikasi sebelum menggunakan algoritma SVM dan sesudah menggunakan algoritma SVM.

1.7 Sistematika Penulisan

Agar dokumentasi penelitian ini sistematis, berikut adalah sistematika penulisan skripsi ini:

BAB I. PENDAHULUAN

Pada bab ini akan membahas mengenai Latar Belakang Masalah, Rumusan Masalah, Batasan Masalah, Maksud dan Tujuan Penelitian, Manfaat Penelitian, Metode Penelitian serta Sistematika Penulisan.

BAB II. LANDASAN TEORI

Pada bab ini akan dijelaskan dan diuraikan tentang uraian teori-teori yang mendasari pembahasan beserta tinjauan pustaka.

BAB III. ANALISIS DAN PERANCANGAN

Pada bab ini akan dijelaskan dan diuraikan mengenai pengenalan sistem secara umum dan perangkat lunak yang akan digunakan dan alur dari penelitian dalam penyusunan skripsi.

BAB IV. IMPLEMENTASI DAN PEMBAHASAN

Memuat tentang implementasi sistem dengan mengacu pada analisis dan perancangan yang telah dibuat serta pembahasan masing-masing komponen yang membentuk sistem.

BAB V. PENUTUP

Memuat tentang kesimpulan dari proses dan hasil penelitian, serta saran dari pihak kedua terhadap sistem yang telah dibuat.