

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pemanfaatan dan kemajuan teknologi dinilai menguntungkan kehidupan manusia dengan perkembangan industri 4.0 yang begitu pesat. Kemudahan dan kecepatan pengiriman dan distribusi informasi, yang dikenal sebagai teknologi informasi merupakan salah satu keuntungan dari kemajuan teknologi yang dirasakan oleh manusia. Semua perangkat berbasis komputer yang digunakan oleh orang-orang untuk mengelola dan memenuhi kebutuhan informasi organisasi disebut teknologi informasi. Adapun kegiatan mengumpulkan, memproses, menyimpan, menganalisis, dan menyebarkan data adalah tugas sistem informasi [1].

Adapun dalam perkembangan teknologi di industri 4.0 tidak lepas dari hadirnya media sosial, sebagai contoh adalah media sosial X. X berganti nama dari Twitter semenjak 2022. X sebagai platform jejaring sosial *online* terkenal, menjadi tempat kolaborasi dan komunikasi bagi jutaan pengguna harian sejak dirilis pada tahun 2006. Di tahun 2018, dengan lebih dari 320 juta pengguna aktif bulanan, X digunakan untuk berbagai tujuan, termasuk berbagi pesan singkat yang disebut *tweet*, dengan batasan maksimal 140 karakter. Selain *tweet*, pengguna juga dapat berbagi berbagai konten multimedia, termasuk video, sebagai bentuk komunikasi di platform ini [2]. Peningkatan signifikan dalam penggunaan X, yang dipicu oleh pertumbuhan interaksi sosial sehari-hari antara pengguna di *platform* ini. Pertumbuhan ini juga menarik perhatian penjahat dunia maya yang memanfaatkan hubungan kepercayaan antar pengguna untuk menyebarkan konten berbahaya kepada sejumlah besar korban di jaringan. Penjahat dunia maya di X, yang sering disebut sebagai spammer sosial atau akun *suspicious* berkolaborasi dengan berbagai cara untuk meningkatkan visibilitas akun mereka [2]. *Spammer* memanfaatkan X untuk mengirimkan pesan kekerasan, tautan *phishing*, perundungan siber dan

kejahatan siber lainnya. Dalam perkembangannya, akun spam ini juga menyebarkan berita palsu (*fake news*) yang dipandang sebagai salah satu ancaman terbesar terhadap demokrasi, jurnalisme, dan perekonomian [3]. Kelompok pelaku spam ini mengandalkan layanan yang disediakan oleh X untuk menyebarkan konten berbahaya mereka. X memungkinkan penggunaan simbol "#" untuk mengelompokkan *tweet* berdasarkan topik. Perkembangan yang cepat dari berita palsu dan dampak negatifnya terhadap demokrasi, keadilan, serta tingkat kepercayaan masyarakat telah meningkatkan permintaan akan pendeteksian berita palsu. Penyebaran berita palsu saat ini sangat cepat, dan dampaknya dapat mengakibatkan pengetahuan yang salah. Dengan menggunakan akun spam, penyerang dapat mengubah lingkungan Jaringan Sosial *Online* (OSN) tepercaya melawan penggunaannya dengan menyebarkan pesan spam di jaringan tersebut. Meningkatnya serangan berbahaya di situs jejaring sosial secara eksponensial menghadirkan kebutuhan akan solusi untuk mencegah serangan ini dan melindungi pengguna dari penipuan [3]. Diperlukan suatu kerangka kerja yang komprehensif untuk secara konsisten memahami dan mengidentifikasi akun spam, dengan tujuan untuk mengumpulkan dan menyatukan peneliti dari berbagai latar belakang dalam upaya penelitian terkait penyebaran berita palsu [4].

Pelaksanaan Pemilihan Presiden (Pilpres 2024) tidak dapat dilepaskan dengan kehadiran media sosial seperti X dan *spammer*-nya. Dengan adanya pengalaman dan data historis dari Pemilihan Presiden dua periode terakhir di tahun 2014 dan 2019, fenomena antusiasme oleh masyarakat Indonesia tidak terlepas dari kontribusi kedua pendukung kandidat calon presiden, khususnya di media sosial Twitter [5]. Antusiasme dukungan ini tidak sepenuhnya merupakan fakta, melainkan juga penggiringan opini negatif yang mengarah juga kepada kampanye hitam atau *black campaign*. Media sosial berupa X menjadi ruang atau wadah bukan hanya bagi publik dalam menyampaikan fakta dan opini positif namun juga menjadi media penyebaran hoaks atau berita palsu [9].

Saat ini, penelitian tentang deteksi akun spam, termasuk akun yang dicurigai sebagai akun spam (*suspicious account*) menggunakan model pembelajaran mesin (*machine learning*) dan kecerdasan buatan (*artificial intelligence*) telah dilakukan. Penelitian tersebut membantu pembaca berita menghindari berita palsu [4]. Beberapa penelitian terdahulu yang dilakukan yaitu penelitian yang dilakukan oleh Kayode Sakariyah Adewole di tahun 2018 yang menggunakan kombinasi algoritma klasifikasi dan *Random Forest* [6], penelitian oleh Supanya membahas deteksi berita palsu dengan menggunakan model *machine learning* dengan menggunakan metode klasifikasi melalui algoritma Naïve Bayes, SVM, dan *Neural Network* [7]. Penelitian lainnya yang dilakukan pada tahun 2022 juga menggunakan pendekatan K-Means *clustering* dan SVM untuk mendeteksi berita hoaks melalui judul berita [8]. Penelitian yang dimaksud menguji dua algoritma pengelompokan utama, di antaranya adalah K-Medoids dan K-Means, pada satu set data. Hasil uji menunjukkan bahwa K-Means unggul dibandingkan K-Medoids. Selanjutnya, penelitian ini juga menerapkan metode *supervised learning* seperti Regresi Logistik, *K-Nearest Neighbours*, dan *Support Vector Classifier*, untuk melatih judul berita yang sama yang digunakan untuk pengelompokan. Hasilnya dicatat dalam kolom 'Prediksi', dan kemudian teknik dengan akurasi tertinggi dipilih. Berdasarkan pengujian, diketahui bahwa SVM mencapai akurasi maksimum sebesar 94,93 persen. Kesimpulannya, penelitian ini merancang sebuah model hibrid yang terdiri dari algoritma pengelompokan K-Means *unsupervised learning* dan algoritma klasifikasi *supervised learning Support Vector*. Dalam model ini, algoritma K-Means mengelompokkan judul berita berdasarkan penggunaan kata-kata tertentu, sementara algoritma vektor dukungan belajar dari kelompok-kelompok tersebut untuk memprediksi kategori judul berita yang tidak terlihat. Hasil kinerja menunjukkan bahwa SVM mencapai presisi yang lebih tinggi yaitu sebesar 94.93%, dibandingkan dengan pengklasifikasi DT dan NB. Akan tetapi, pada penelitian ini mendeteksi dari dari segi judul berita yang dicurigai berita palsu dan tidak melakukan optimasi pada algoritma SVM. Maka dalam penelitian ini, dikembangkan algoritma yang sama dengan optimasi

*hyperparameter* pada SVM untuk mendeteksi *suspicious account* dan *non-suspicious account* pada media sosial X dengan kata kunci “pilpres”.

Dari adanya penelitian yang telah dilakukan terdahulu pada pemaparan di atas, penelitian ini juga mengadopsi metode pengimplementasian kombinasi algoritma unsupervised learning dan supervised learning. Pada penelitian ini, dilakukan pengimplementasian algoritma SVM dari data yang sudah dilalui algoritma PCA dan K-Means menjadi dua kluster, yaitu *suspicious* dan *non-suspicious*. Setelah itu, dilakukan optimasi kombinasi algoritma optimasi melalui *setting hyperparameter* dengan harapan dapat terciptanya kerangka kerja yang komprehensif untuk mengidentifikasi dan mencegah berbagai bentuk serangan berbahaya di lingkungan media sosial.

## 1.2 Rumusan Masalah

Berdasarkan pemaparan latar belakang di atas, di bawah ini merupakan rumusan masalah dari penelitian yang dilakukan:

1. Bagaimana hasil pengembangan model klasifikasi *suspicious account* dengan menggunakan SVM berdasarkan label yang didapatkan dari proses *clustering* dengan PCA dan K-Means?
2. Bagaimana hasil optimasi *hyperparameter* model SVM untuk mendapatkan evaluasi yang optimal?

## 1.3 Batasan Masalah

Beberapa batasan masalah difiksasi agar pembahasan penelitian ini tetap pada arah yang sama dan tidak melenceng keluar dari batasan-batasan, di antaranya sebagai berikut:

1. *Tweet* yang dipakai dalam rangka melakukan analisis merupakan *tweet* dengan kata kunci “pilpres” yang berbahasa Indonesia dan diunggah oleh WNI.
2. Rentang waktu *Tweet* yang digunakan tanggal 1 Januari – 21 Mei 2023.

3. Analisis data tweet yang dilakukan mencakup interaksi percakapan melalui *tweet*, *retweet*, dan *mention* dari akun yang menyebarkan informasi terkait dengan proses pemilihan presiden.
4. Pengidentifikasian berita palsu dilakukan dengan memanfaatkan algoritma SVM dan optimasinya dengan *setting hyperparameter*.
5. Proses penelitian hanya mencakup tahapan evaluasi, tanpa melibatkan langkah-langkah berikutnya dalam penelitian.

#### **1.4 Tujuan dan Manfaat Penelitian**

##### **1.4.1 Tujuan Penelitian**

Di bawah ini merupakan tujuan dari dilakukannya penelitian ini:

1. Mendapatkan hasil dari pengembangan model klasifikasi *suspicious account* dengan menggunakan SVM berdasarkan label yang didapatkan dari proses *clustering* dengan PCA dan K-means.
2. Mendapatkan hasil dari melakukan optimasi *hyperparameter* model SVM untuk mendapatkan hasil evaluasi yang optimal.

##### **1.4.2 Manfaat Penelitian**

Di bawah ini merupakan manfaat dari penelitian yang dilakukan, yaitu sebagai berikut:

1. Memberikan wawasan terhadap performa algoritma SVM pada klasifikasi pendeteksian *suspicious account*.
2. Menjadi inspirasi dalam melakukan penelitian di masa yang akan datang untuk membantu meningkatkan kesadaran akan adanya akun yang dicurigai sebagai penyebar berita palsu dan bahayanya.

#### **1.5 Sistematika Penulisan**

1. BAGIAN 1: PENDAHULUAN

Bagian ini menguraikan fenomena permasalahan yang menjadi fokus penelitian, mencantumkan kajian literatur terkait, menetapkan tujuan penelitian, merinci manfaat yang diharapkan, dan merumuskan masalah beserta batasan-batasannya.

## 2. BAGIAN 2: LANDASAN TEORI

Bagian ini menyajikan secara komprehensif dasar-dasar teori yang menjadi landasan penelitian, mencakup teori umum dan khusus sesuai dengan konteks penelitian, serta mengeksplorasi penelitian terdahulu sebagai referensi utama.

## 3. BAGIAN 3: METODOLOGI

Bagian ini memberikan penjelasan mendalam mengenai metodologi penelitian dan gambaran umum proses penelitian yang diimplementasikan. Tujuan dari bagian ini adalah memastikan kesesuaian antara metode yang digunakan dengan tujuan penelitian yang telah ditetapkan.

## 4. BAGIAN 4: ANALISIS DAN HASIL

Bagian ini menyajikan penjelasan rinci mengenai hasil analisis dari penelitian yang telah selesai, mengikuti proses yang telah diuraikan sebelumnya. Selain itu, diuraikan pula secara komprehensif alur penelitian mulai dari pengumpulan data hingga prediksi dengan menggunakan algoritma yang diterapkan.

## 5. BAGIAN 5: SIMPULAN DAN SARAN

Bagian ini merangkum kesimpulan dari hasil penelitian dan memberikan saran serta rekomendasi untuk penelitian selanjutnya. Keseluruhan bagian ini disusun dengan kecermatan untuk memberikan kontribusi yang berarti dalam pengembangan pengetahuan dan praktik di bidang terkait.

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A