

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan platform media sosial yang pesat dalam satu dekade terakhir telah mengubah secara fundamental cara manusia berinteraksi, mengonsumsi informasi, dan mengalokasikan waktu hariannya. Pada tahun 2024, jumlah pengguna media sosial aktif secara global melampaui 5,1 miliar orang, setara dengan sekitar 63% total populasi dunia. Di Indonesia sendiri, data We Are Social (2025) melaporkan bahwa rata-rata pengguna internet menghabiskan 7 jam 38 menit per hari di depan layar digital, dengan lebih dari 3 jam secara khusus untuk media sosial [1].

Salah satu dampak kognitif yang semakin menarik perhatian para peneliti adalah perubahan pada kemampuan mempertahankan perhatian atau *attention span*. Attention span didefinisikan sebagai durasi seseorang mampu memfokuskan perhatian secara aktif pada suatu stimulus atau aktivitas tanpa mengalami pengalihan yang tidak disengaja [2]. Penelitian-penelitian terkini mengindikasikan bahwa paparan berulang terhadap konten digital berformat singkat dan pergantian stimulus yang cepat di media sosial dapat melatih otak untuk mengharapkan perubahan konstan, yang pada jangka panjang berpotensi menurunkan kapasitas *sustained attention* [3][4].

Dataset Social Media & Mental Health (SMMH) yang tersedia secara publik di Kaggle menyediakan basis data survei komprehensif dari 481 responden yang mencakup pola penggunaan media sosial, kondisi demografis, dan 12 indikator kesehatan mental yang dinilai menggunakan skala Likert 1–5. Dataset ini menjadi fondasi empiris yang relevan karena mengandung variabel-variabel yang secara teoretis berkorelasi langsung dengan *attention span*, khususnya skor distraksi, kecenderungan mudah teralihkan, dan kesulitan berkonsentrasi.

Dalam konteks machine learning, Support Vector Machine (SVM) adalah algoritma supervised learning yang diperkenalkan oleh Cortes dan Vapnik pada tahun 1995 dan telah terbukti sangat efektif untuk masalah klasifikasi data berdimensi tinggi dengan jumlah sampel yang relatif terbatas [5]. Keunggulan SVM terletak pada prinsip *structural risk minimization* yang mencari hyperplane dengan margin terlebar, memberikan jaminan teori generalisasi yang lebih kuat dibandingkan algoritma lain [6]. Dalam beberapa tahun terakhir, SVM telah berhasil

diaplikasikan pada berbagai domain kesehatan digital dan analisis perilaku, menunjukkan akurasi yang kompetitif bahkan pada data survei [7][8][9].

Penelitian sebelumnya oleh Hawi dan Samaha (2017) menemukan korelasi signifikan antara intensitas penggunaan media sosial dan gangguan perhatian, namun tidak menggunakan pendekatan machine learning [10]. Studi lain oleh Twenge et al. (2018) secara longitudinal mendokumentasikan penurunan attention span, namun masih terbatas pada analisis statistik deskriptif [11][12]. Kesenjangan ini menunjukkan kebutuhan akan pendekatan komputasional yang lebih sistematis untuk mengklasifikasikan tingkat attention span berbasis data perilaku media sosial.

Penelitian ini mengisi kesenjangan tersebut dengan membangun model klasifikasi SVM berbasis data SMMH untuk mengklasifikasikan attention span pengguna ke dalam tiga kelas (Tinggi, Sedang, Rendah) menggunakan 19 fitur perilaku digital dan demografis. Penelitian ini juga membandingkan tiga varian kernel SVM untuk mengidentifikasi konfigurasi terbaik, serta menganalisis feature importance untuk mengidentifikasi variabel perilaku media sosial yang paling determinatif terhadap attention span.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, penelitian ini merumuskan tiga permasalahan utama:

1. Bagaimana kinerja model SVM dengan kernel Linear, RBF, dan Polynomial dalam mengklasifikasikan attention span pengguna media sosial ke dalam kelas Tinggi, Sedang, dan Rendah berdasarkan dataset SMMH (478 responden, 19 fitur)?
2. Kernel SVM manakah yang menghasilkan performa terbaik diukur menggunakan akurasi, F1-score, dan cross-validation 5-fold pada data Social Media & Mental Health?
3. Variabel perilaku media sosial manakah yang memiliki kontribusi paling signifikan terhadap prediksi attention span berdasarkan analisis permutation importance?

1.3 Tujuan Penelitian

Penelitian ini memiliki tiga tujuan utama:

1. Membangun dan mengevaluasi model klasifikasi attention span menggunakan SVM dengan tiga kernel (Linear, RBF, Polynomial) pada dataset SMMH serta mengukur kinerjanya melalui accuracy, F1-score, confusion matrix, dan cross-validation 5-fold.

2. Membandingkan performa ketiga kernel SVM dan melakukan optimasi hyperparameter melalui GridSearchCV untuk mengidentifikasi konfigurasi model terbaik.
3. Mengidentifikasi fitur-fitur perilaku media sosial yang paling berpengaruh terhadap attention span melalui analisis permutation importance.

1.4 Manfaat Penelitian

1.4.1 Manfaat Teoritis

Penelitian ini berkontribusi pada pengembangan literatur di bidang *computational mental health informatics* dengan mendemonstrasikan aplikasi SVM untuk klasifikasi konstruk psikologis berbasis data survei media sosial.

1.4.2 Manfaat Praktis

Bagi pengembang platform dan pembuat kebijakan digital wellbeing, model SVM yang dibangun dapat menjadi prototipe sistem skrining awal yang mengidentifikasi pengguna berisiko mengalami penurunan attention span.

1.4.3 Manfaat Akademis

Bagi civitas akademika program studi Sistem Informasi, penelitian ini mendemonstrasikan implementasi lengkap pipeline machine learning dari pengumpulan dan pembersihan data, rekayasa fitur, pelatihan model dengan multiple kernel, evaluasi komprehensif, hingga interpretasi hasil menggunakan bahasa Python dan library scikit-learn yang menjadi standar industri [13].

1.5 Batasan Penelitian

1. Dataset yang digunakan adalah SMMH dari Kaggle dengan 481 responden dan 21 variabel. Penelitian tidak melakukan pengumpulan data primer baru.
2. Label Attention Span dibentuk secara komputasional dari skor komposit tiga pertanyaan kognitif menggunakan pendekatan berbasis persentil, bukan dari pengukuran psikometri terstandar.
3. Algoritma yang digunakan terbatas pada SVM dengan tiga kernel: Linear, RBF, dan Polynomial.
4. Evaluasi mencakup accuracy, F1-score, confusion matrix, dan cross-validation 5-fold.