

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Perkembangan industri 4.0 yang sangat berkaitan dengan inovasi memiliki salah satu tren teknologi yaitu *Artificial Intelligence* (AI) atau yang disebut kecerdasan buatan. AI kemudian menjadi penggerak revolusi industri 4.0 yang menjanjikan banyak kemudahan dalam berbagai sektor [1]. Dengan bantuan *Big Data*, otomatisasi dapat terealisasi dengan meningkatkan aspek mobilitas, fleksibilitas, dan efisiensi energi. Sehingga, pengaplikasian *data mining* dapat digunakan untuk analisis, *modelling*, simulasi, komputasi, dan pengambilan keputusan [2].

Dewasa ini, banyak peneliti yang menggunakan AI dalam meningkatkan kemampuan prediksi, diagnosis, serta pengobatan penyakit. Hasil penerapan AI yaitu *machine learning* merupakan sistem yang mampu belajar secara otomatis dari data yang telah ada. Salah satu aplikasi dari *machine learning* yang paling menarik yaitu di bidang ilmu psikologi. Dalam bidang psikologi, *machine learning* dapat dikembangkan menjadi alat penelitian yang dapat memprediksi perilaku manusia dan ciri-ciri kepribadian [3].

Kepribadian merupakan identitas terkompleks yang dimiliki oleh manusia. Kepribadian manusia juga dapat menandakan keunikan dari setiap orang [4]. Kepribadian seseorang diartikan pula sebagai pola dari pengaruh, perilaku, kesadaran, dan keinginan yang kemudian menandakan keunikan setiap individu. Adanya keterkaitan antara kepribadian seseorang dengan pemilihan tipe kata yang digunakan memungkinkan perbedaan ciri kepribadian seseorang dapat diprediksi dengan menganalisis bahasa mereka. Contohnya seseorang dengan nilai kepribadian cenderung *extrovert* lebih sering menggunakan kata-kata yang menunjukkan proses sosial. Sedangkan, penggunaan kata ganti orang pertama tunggal seperti "saya" dan "-ku" menunjukkan refleksi diri yang mengarah ke *neuroticism* dan depresi [5].

Ada beberapa teori yang dapat digunakan mengenai ciri kepribadian seseorang, salah satunya yaitu Myers-Briggs Type Indicator (MBTI). MBTI merupakan adaptasi dari teori tipe psikologis oleh C. G. Jung yang bertujuan untuk dipelajari dan dapat berguna bagi kehidupan manusia. Teori ini menjelaskan bahwa perilaku manusia yang terjadi merupakan perilaku yang teratur dan konsisten. Hal ini

dikarenakan perbedaan dari setiap individu dalam menggunakan persepsi dan cara menilai mereka. Ada empat pembagian dasar yang dapat menentukan tipe kepribadian berdasarkan teori MBTI yaitu *favorite world* (menentukan dimensi kepribadian *extraversion* atau *introversion*), *information* (menentukan dimensi kepribadian *sensing* atau *intuition*), *decisions* (menentukan dimensi kepribadian *thinking* atau *feeling*), *structure* (menentukan dimensi kepribadian *judging* atau *perceiving*) [6].

Support Vector Machines (SVM) adalah *supervised learning model* yang dapat bekerja dengan baik dan memiliki kemampuan generalisasi dalam menganalisis data serta mengenali pola yang digunakan dalam analisis klasifikasi dan regresi [7]. SVM merupakan metode *machine learning* yang bekerja dengan prinsip Structural Risk Minimization (SRM) dengan tujuan untuk menemukan *hyperplane* terbaik yang memisahkan data menjadi dua kelas [8]. Penelitian [9], membuktikan bahwa SVM menghasilkan performa yang baik dalam klasifikasi teks dibandingkan algoritma lain seperti Naïve Bayes dan KNN. Hal ini dikarenakan pada klasifikasi teks memiliki banyak sekali fitur dan SVM memiliki perlindungan terhadap *overfitting* sehingga SVM memiliki kemampuan untuk menghadapi fitur yang sangat banyak. Selain itu, kebanyakan kasus klasifikasi teks merupakan kasus yang dapat dipisah secara linear sehingga cocok menggunakan algoritma SVM. Pada penelitian yang berjudul Sentiment Analysis of Twitter Data: A Survey of Techniques [10], hasil klasifikasi dengan tingkat akurasi tertinggi menggunakan SVM yaitu dengan akurasi 76,68%, sedangkan pada Maximum Entropy yaitu sebesar 74,93% dan Naïve Bayes Classifier (NBC) sebesar 74,65%.

Berdasarkan beberapa hal yang dipaparkan di atas, diadakanlah penelitian Prediksi Tipe Kepribadian Berdasarkan Konten Menggunakan Algoritma Support Vector Machines. Pada penelitian yang berjudul Peningkatan Akurasi pada Prediksi Kepribadian MBTI Pengguna Twitter Menggunakan Augmentasi Data [11], akurasi yang dicapai dengan algoritma Random Forest dengan *preprocessing*, *feature extraction* TF-IDF, serta tambahan fitur *following*, *followers*, dan jumlah *tweet* yaitu sebesar 40%. Lalu akurasi ditingkatkan menjadi 70% dengan menerapkan teknik augmentasi data dengan Random Deletion. *Dataset* yang sedikit kemudian diperkaya dengan menghapus beberapa kata dan dijadikan data yang baru dengan teknik ini. Pada penelitian ini, algoritma yang akan digunakan yaitu Support Vector Machine dengan *preprocessing*, *feature extraction* TF-IDF, serta teknik memperkaya *dataset* secara *oversampling* dengan SMOTE (Synthetic Minority Over-sampling Technique). SMOTE merupakan teknik *oversampling* yang menambah data kelas minoritas dengan menciptakan data sintesis berdasarkan tetangga

terdekat [12]. Dengan menggunakan teknik *oversampling* ini, diharapkan dapat meningkatkan hasil prediksi model dengan akurasi yang lebih tinggi. Demikian, dengan adanya sistem ini, diharapkan dapat membantu manusia memahami individu lain berdasarkan konten yang ditulisnya.

## **1.2 Rumusan Masalah**

Pada subbab ini, beberapa rumusan masalah yang akan dibahas yaitu sebagai berikut.

1. Bagaimana cara memprediksi tipe kepribadian MBTI seseorang berdasarkan konten media sosial dengan algoritma Support Vector Machines?
2. Bagaimana akurasi pada hasil prediksi tipe kepribadian MBTI seseorang berdasarkan konten media sosial dengan algoritma Support Vector Machines?

## **1.3 Batasan Permasalahan**

Batasan-batasan masalah yang ada dalam penelitian ini yaitu sebagai berikut.

1. Dataset yang digunakan diperoleh dari Kaggle (<https://www.kaggle.com/datasets/datasnaek/mbti-type>).
2. Jumlah data yang digunakan sebanyak 8675 baris data.

## **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah di atas, tujuan dari penelitian adalah sebagai berikut.

1. Mengetahui cara memprediksi tipe kepribadian MBTI seseorang berdasarkan konten media sosial dengan algoritma Support Vector Machines.
2. Mengukur akurasi dari hasil prediksi tipe kepribadian MBTI seseorang berdasarkan konten media sosial dengan algoritma Support Vector Machines.

## 1.5 Manfaat Penelitian

Manfaat yang didapatkan dari penelitian ini adalah sebagai berikut.

1. Membantu pengguna untuk mengetahui tipe kepribadian MBTI seseorang berdasarkan konten media sosial.
2. Mengetahui akurasi dari sistem prediksi tipe kepribadian MBTI seseorang berdasarkan konten media sosial dengan algoritma Support Vector Machines.
3. Menambah referensi dalam pengimplementasian algoritma Support Vector Machines untuk prediksi tipe kepribadian MBTI bagi pembaca yang ingin meneliti topik yang berhubungan.

## 1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN

Bab ini berisi latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, serta manfaat penelitian.

- Bab 2 LANDASAN TEORI

Bab 2 menjelaskan mengenai landasan teori yang digunakan selama penelitian. Landasan teori yang digunakan yaitu Myers-Briggs Type Indicator (MBTI) dan Support Vector Machines (SVM).

- Bab 3 METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai metodologi yang digunakan selama penelitian.

- Bab 4 HASIL DAN DISKUSI

Bab 4 membabarkan mengenai tahap penelitian, hasil penelitian, serta diskusi mengenai kemungkinan yang terjadi selama penelitian.

- Bab 5 SIMPULAN DAN SARAN

Bab 5 merupakan kesimpulan dari penelitian yang telah dilakukan dan saran yang dapat diterapkan untuk penelitian selanjutnya.