

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan media sosial telah mengubah secara signifikan cara masyarakat memperoleh dan mengonsumsi berita. Media sosial tidak lagi hanya berfungsi sebagai sarana komunikasi dan hiburan, tetapi juga menjadi kanal utama penyebaran informasi, khususnya berita lokal yang bersifat cepat, ringkas, dan kontekstual. Instagram merupakan salah satu platform yang banyak dimanfaatkan sebagai media distribusi informasi karena kemampuannya menyajikan konten visual yang disertai teks singkat (*caption*) yang mudah diakses oleh pengguna [1].

Seiring meningkatnya peran Instagram sebagai media informasi, muncul berbagai akun berbasis komunitas yang berfungsi sebagai penyedia berita lokal. Salah satu akun yang memiliki peran tersebut adalah Merapi Uncover, yang secara aktif membagikan informasi seputar wilayah Yogyakarta dan sekitarnya. Konten yang disajikan mencakup berbagai topik, antara lain aktivitas Gunung Merapi, kondisi lalu lintas, peristiwa kriminal, kegiatan sosial dan pemerintahan, serta hiburan, budaya, dan wisata. Keberagaman topik ini menjadikan Merapi Uncover sebagai sumber informasi lokal yang relevan dan banyak digunakan oleh masyarakat.

Namun, tingginya volume unggahan dan variasi topik berita dalam satu akun menyebabkan informasi sulit dikelola dan disaring secara manual. Unggahan disajikan secara kronologis tanpa pengelompokan topik, sehingga pengguna kerap mengalami kesulitan dalam menemukan informasi yang sesuai dengan kebutuhan tertentu, seperti berita lalu lintas atau bencana. Selain itu, dari sisi analisis data, banyaknya kategori awal dengan jumlah data yang tidak seimbang menyulitkan proses pengelompokan dan evaluasi konten secara sistematis. Kondisi ini menunjukkan perlunya sistem klasifikasi topik berita secara otomatis untuk meningkatkan efisiensi penyajian dan pengelolaan informasi [2].

Klasifikasi berita berbasis teks pada media sosial memiliki tantangan yang berbeda dibandingkan teks berita konvensional. Caption Instagram umumnya bersifat pendek, tidak terstruktur, menggunakan bahasa informal, singkatan, serta variasi gaya penulisan yang tinggi. Karakteristik tersebut menyebabkan proses klasifikasi menjadi lebih kompleks dan menuntut metode pembelajaran mesin yang mampu menangani data berdimensi tinggi dengan tingkat generalisasi yang baik [3].

Salah satu algoritma machine learning yang banyak digunakan dalam klasifikasi teks adalah *Support Vector Machine* (SVM). Algoritma ini dikenal memiliki performa yang stabil pada data berdimensi tinggi dan mampu menghasilkan akurasi yang baik dalam klasifikasi berita berbahasa Indonesia. Beberapa penelitian menunjukkan bahwa SVM mampu mengungguli metode lain seperti *Naïve Bayes* dan *K-Nearest Neighbor* dalam tugas klasifikasi berita digital [4], [5]. Selain itu, SVM memiliki keunggulan dari sisi efisiensi komputasi dibandingkan model *deep learning*, sehingga sesuai untuk diterapkan pada dataset berukuran menengah seperti *caption* media sosial.

Performa SVM sangat dipengaruhi oleh tahapan pra-proses teks dan teknik ekstraksi fitur yang digunakan. *Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan metode representasi teks yang umum digunakan karena mampu menonjolkan kata-kata yang relevan sekaligus menekan pengaruh kata umum yang tidak informatif. Penelitian sebelumnya menunjukkan bahwa kombinasi preprocessing teks yang tepat dengan TF-IDF dapat meningkatkan performa klasifikasi berita secara signifikan [6].

Meskipun penelitian mengenai klasifikasi berita telah banyak dilakukan, sebagian besar studi masih berfokus pada teks berita formal dari portal daring atau judul berita yang relatif terstruktur. Penelitian yang secara khusus mengkaji klasifikasi topik berita pada *caption* Instagram akun komunitas lokal seperti Merapi Uncover masih sangat terbatas. Selain itu, belum banyak penelitian yang membahas dampak pemetaan kategori awal yang sangat beragam menjadi

sejumlah kelas besar, serta pengaruh ketidakseimbangan kelas terhadap performa SVM pada konteks berita lokal media sosial.

Novelty penelitian ini terletak pada penerapan dan analisis performa algoritma SVM dengan representasi fitur TF-IDF pada caption Instagram akun Merapi Uncover yang memiliki karakteristik unik, yaitu teks pendek, bahasa informal, dengan kategori lima kelas utama, serta distribusi data yang tidak seimbang. Penelitian ini tidak hanya mengukur akurasi keseluruhan, tetapi juga menganalisis performa model pada setiap kategori topik menggunakan metrik *precision*, *recall*, dan *F1-score*.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menganalisis performa algoritma *Support Vector Machine* dalam melakukan klasifikasi topik berita pada akun Instagram Merapi Uncover. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem klasifikasi berita otomatis berbasis media sosial berbahasa Indonesia, serta menjadi referensi bagi penelitian selanjutnya dalam bidang pengolahan teks dan *machine learning*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini adalah bagaimana performa algoritma *Support Vector Machine* dalam melakukan klasifikasi topik berita pada Merapi Uncover?

1.3 Batasan Masalah

Adapun masalah dalam penelitian ini meliputi :

1. *Dataset* yang digunakan berasal dari postingan Instagram dengan nama akun *merapi_uncover*.
2. Algoritma yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM).

3. Topik berita yang digunakan dalam penelitian ini dibatasi dalam 5 kategori utama yaitu:

1. Hiburan / Budaya / Wisata
2. LaluLintas & Transportasi
3. Bencana & Kebencanaan
4. Sosial & Pemerintahan
5. Kriminal & Keamanan Publik

kernel yang di gunakan adalah *kernel linear* (LinearSVC). *Kernel non-linear* seperti *Radial Basis Function* (RBF), *polynomial*, dan *sigmoid* tidak digunakan dalam penelitian ini.

1.4 Tujuan Penelitian

Tujuan utama pada penelitian ini adalah untuk menganalisis performa algoritma *Support Vector Machine* dalam melakukan klasifikasi topik berita pada akun instagram bernama *merapi_uncover*.

1.5 Manfaat Penelitian

Menambah pengetahuan tentang *Performa Support Vector Machine* pada klasifikasi berita untuk mempermudah mencari berita berdasarkan kategori.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini terdiri dari 5 bab, dengan susunan sebagai berikut :

BAB I PENDAHULUAN, berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, sistematika penulisan.

BAB II TINJAUAN PUSTAKA, berisi pembahasan dasar teori yang berkaitan dengan penelitian serta penelitian-penelitian sebelumnya yang relevan.

BAB III METODELOGI PENELITIAN, bab ini menguraikan tahapan yang akan dilaksanakan dalam penelitian.

BAB VI HASIL DAN PEMBAHASAN, bab ini berisi hasil dan pembahasan dari penerapan metode yang di jelaskan pada bab sebelumnya.

BAB V PENUTUP, berisi kesimpulan dan saran yang diperoleh dari penelitian dan saran untuk penelitian selanjutnya.