

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan era digital telah meningkatkan ketergantungan masyarakat terhadap teknologi informasi, terutama dalam mengakses berita secara daring yang jumlah dan kecepatannya terus meningkat. Salah satu aktivitas utama masyarakat Indonesia dalam berinternet adalah membaca berita online, yang mencapai 15,5% dari total aktivitas internet menurut data Kominfo [1]. Berita pariwisata merupakan salah satu jenis konten yang cukup banyak diminati, mencakup berbagai topik seperti wisata alam, wisata buatan, wisata budaya hingga wisata religi [2]. Penelitian ini berfokus pada klasifikasi otomatis berita pariwisata berdasarkan isi kontennya, dengan menggunakan data yang dikumpulkan dari situs *Kompas.com* melalui teknik web scraping. Tujuannya adalah untuk mengelompokkan berita – berita tersebut ke dalam kategori tertentu sehingga lebih mudah ditemukan dan dimanfaatkan oleh pembaca.

Namun, meskipun akses informasi digital semakin luas, tingkat literasi digital di Indonesia masih tergolong rendah [1]. Banyak pengguna membaca berita dari berbagai sumber tanpa memperhatikan kredibilitas dan latar belakang media tersebut, sehingga berpotensi menyebabkan penyebaran misinformasi [1]. Di sisi lain, tingginya volume berita pariwisata yang beredar menimbulkan tantangan tersendiri dalam hal pengelolaan dan pencarian informasi. Pengguna seringkali kesulitan menemukan berita yang sesuai dengan kebutuhan karena tidak adanya sistem klasifikasi yang efektif berdasarkan kategori pariwisata, seperti wisata alam, wisata buatan, wisata budaya dan wisata religi.

Masalah utama yang diangkat dalam penelitian ini adalah belum adanya sistem klasifikasi otomatis untuk berita pariwisata khususnya pada website *Kompas.com* yang mampu mengelompokkan berita sesuai isi kontennya secara akurat dan efisien. Misalnya, berita berjudul “Eksotisme Pantai Ora di Maluku Tengah” seharusnya termasuk kategori wisata alam, tetapi tanpa sistem klasifikasi, berita tersebut tercampur dengan artikel lain, menyulitkan pencarian spesifik. Dengan meningkatnya jumlah konten digital setiap harinya, permasalahan ini menjadi semakin relevan untuk ditangani.

Untuk mengatasi masalah tersebut, telah tersedia berbagai pendekatan klasifikasi teks berbasis algoritma pembelajaran mesin. Beberapa algoritma yang umum digunakan meliputi K-Nearest Neighbors (KNN), Gaussian Naïve Bayes, Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (MLP), Ridge Regression, serta Stochastic Gradient Descent (SGD). Setiap algoritma tersebut memiliki keunggulan tersendiri dalam proses klasifikasi, tergantung pada jenis data dan kompleksitas relasi antar fitur.

Kendati demikian, masing-masing metode juga memiliki keterbatasan. KNN cenderung lambat saat digunakan pada dataset besar [3]. Gaussian Naïve Bayes memiliki asumsi independensi antar fitur yang sering kali tidak sesuai dengan realitas data teks [4]. Random Forest cukup stabil, namun proses pelatihannya memerlukan waktu lebih lama [5]. SVM bekerja optimal pada data berdimensi tinggi tetapi sangat sensitif terhadap parameter [6]. Ridge dan SGD memiliki efisiensi tinggi namun tidak cukup kuat dalam memodelkan relasi non-linier [7]. Di sisi lain, MLP mampu memodelkan hubungan non-linier secara efektif dan melakukan ekstraksi fitur secara otomatis, menjadikannya pilihan yang menjanjikan dalam klasifikasi teks [8].

Penelitian ini memilih algoritma MLP sebagai solusi karena keunggulannya dalam menangani data kompleks dan non-linier. MLP merupakan jenis jaringan saraf tiruan berlapis yang terdiri dari lapisan input, minimal satu lapisan tersembunyi, dan lapisan output. Berdasarkan penelitian Sudianto dkk, MLP menunjukkan hasil akurasi 78% pada klasifikasi topik berita, lebih tinggi dibandingkan SVM yang hanya mencapai 74%. Penelitian lainnya oleh Rindri dan Fitriyani (2023) menunjukkan bahwa MLP menghasilkan akurasi 91% dalam klasifikasi jenis migrain, mengungguli KNN yang hanya memperoleh 72% [1]. Hal ini memperkuat keyakinan bahwa MLP merupakan algoritma yang cocok untuk digunakan dalam penelitian ini.

Penelitian ini diharapkan dapat menghasilkan sistem klasifikasi otomatis berita pariwisata yang mampu mempermudah pembaca untuk mencari berita pariwisata sesuai dengan keinginannya. Selain itu, sistem ini ditargetkan mampu menghasilkan klasifikasi dengan akurasi tinggi dan proses yang efisien dalam membantu pengguna mengakses informasi yang relevan secara cepat dan tepat,

sehingga memudahkan pengguna dalam mengakses informasi yang relevan secara cepat dan tepat. Dengan klasifikasi yang akurat, pengguna akan lebih mudah menemukan berita sesuai minat, seperti berita seputar wisata alam, buatan, budaya, atau religi sekaligus mengurangi potensi misinformasi akibat pencarian berita yang tidak terarah. Selain itu penelitian ini juga diharapkan untuk memberikan kontribusi terhadap literasi digital dan pengembangan teknologi klasifikasi berita berbasis machine learning khususnya di Indonesia.

1.2. Perumusan Masalah

Permasalahan yang diangkat dalam penelitian ini berkaitan dengan belum tersedianya sistem klasifikasi otomatis yang mampu mengelompokkan berita pariwisata berdasarkan isi kontennya pada situs web Kompas.com secara akurat dan efisien. Dalam konteks situs berita seperti Kompas.com, berita-berita pariwisata yang membahas berbagai topik seperti wisata alam, wisata buatan, wisata budaya, dan wisata religi masih tercampur tanpa pengelompokan sistematis, sehingga menyulitkan pengguna dalam menemukan informasi yang relevan dan spesifik.

Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan penggunaan algoritma Multilayer Perceptron (MLP), yaitu salah satu jenis jaringan saraf tiruan yang memiliki kemampuan dalam memodelkan relasi non-linier serta melakukan ekstraksi fitur secara otomatis dari data teks. MLP diharapkan dapat memberikan akurasi klasifikasi yang lebih baik dibandingkan beberapa metode klasifikasi lain yang telah digunakan sebelumnya.

Dengan latar belakang dan solusi yang diusulkan tersebut, maka perumusan masalah dalam penelitian ini dijabarkan ke dalam tiga subbagian berikut:

1.2.1. Masalah (Problem Statement)

Belum adanya sistem klasifikasi otomatis yang mampu mengelompokkan berita pariwisata dari *Kompas.com* secara akurat berdasarkan isi kontennya.

1.2.2. Metode Usulan (*Propose Method*)

Metode yang diusulkan untuk menyelesaikan permasalahan tersebut adalah penggunaan algoritma *Multilayer Perceptron* (MLP), yang merupakan jenis *Artificial Neural Network* (ANN) berlapis, dan dikenal efektif dalam menangani data teks yang kompleks dan bersifat non-linier.

1.2.3. Pertanyaan Penelitian (*Research Questions*)

Berapa nilai akurasi yang diperoleh algoritma Multilayer Perceptron (MLP) dalam mengklasifikasikan berita pariwisata di *kompas.com* berdasarkan kategori konten seperti wisata alam, wisata buatan, dan wisata religi?

1.3. Tujuan dan Manfaat Penelitian

1.3.1. Tujuan Penelitian

Penelitian ini bertujuan untuk menghasilkan sistem klasifikasi otomatis berita pariwisata dengan menggunakan algoritma Multilayer Perceptron (MLP) untuk mengelompokkan berita ke dalam kategori seperti wisata alam, wisata buatan, dan wisata religi.

1.3.2. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini antara lain:

1. Memberikan kemudahan bagi pengguna dalam mengakses informasi berita pariwisata yang relevan melalui sistem klasifikasi otomatis.
2. Memberikan kontribusi terhadap peningkatan literasi digital masyarakat melalui penyajian informasi berita yang lebih terstruktur dan terorganisir berdasarkan kategori.
3. Menjadi referensi bagi pengembang sistem informasi atau media daring dalam penerapan algoritma Multilayer Perceptron (MLP) untuk tugas klasifikasi teks.

1.4. Batasan Penelitian

Agar penelitian lebih terfokus dan terarah, beberapa batasan yang ditetapkan dalam penelitian ini adalah sebagai berikut:

1. Kategori klasifikasi berita dibatasi pada empat jenis utama, yaitu: wisata alam, wisata buatan, wisata budaya dan wisata religi.
2. Proses klasifikasi dilakukan menggunakan algoritma *Multilayer Perceptron* (MLP) tanpa melakukan perbandingan dengan algoritma lain.
3. Dataset yang digunakan merupakan kumpulan berita pariwisata dari *Kompas.com* yang telah diberi label secara manual, dan diproses dalam

preprocessing, termasuk *Case Folding*, *Cleaning*, *Stopword Removal*, *Stemming*, dan *Tokenization*.

4. Dataset berjumlah 1025 data dengan periode berita berada di rentang 2024 hingga 2025, dengan 370 data berlabel alam, 301 data berlabel buatan, 195 data berlabel budaya, dan 159 data berlabel religi.
5. Evaluasi kinerja model hanya difokuskan pada akurasi klasifikasi sebagai metrik utama.

1.5. Sistematika Penulisan

Sistematika penulisan laporan skripsi ini disusun sebagai berikut:

BAB I PENDAHULUAN

Berisi latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan penelitian, serta sistematika penulisan.

BAB II KAJIAN PUSTAKA

Membahas teori-teori yang mendasari penelitian, termasuk konsep dasar klasifikasi teks, algoritma *Multilayer Perceptron*, dan penelitian-penelitian terdahulu yang relevan.

BAB III METODE USULAN

Menjelaskan secara rinci metode yang digunakan dalam penelitian, mulai dari arsitektur sistem, dataset, tahapan penelitian, hingga skenario pengujian.

BAB IV HASIL DAN PEMBAHASAN

Menyajikan hasil implementasi dan pengujian sistem klasifikasi, analisis kinerja model, serta pembahasan terhadap hasil yang diperoleh.

BAB V PENUTUP

Berisi kesimpulan dari hasil penelitian dan saran untuk pengembangan atau penelitian selanjutnya.