

## **BAB I PENDAHULUAN**

### **1.1 Latar Belakang**

Pada era digital saat ini, ketersediaan informasi yang berlimpah, khususnya berita online. Berita adalah laporan tercepat tentang fakta atau ide terbaru yang penting, menarik, atau relevan bagi banyak orang. Laporan ini disampaikan melalui berbagai media berkala, termasuk surat kabar, radio, televisi, dan internet. *BBC News* adalah divisi berita dari *British Broadcasting Corporation* (BBC), lembaga penyiaran publik utama di Inggris. Sebagai salah satu divisi terbesar di BBC, *BBC News* mengkhususkan diri pada liputan berita internasional, nasional, dan lokal. Melalui berbagai platform seperti televisi, radio, dan internet, BBC News menawarkan cakupan berita yang luas dan menyeluruh ke seluruh dunia. Berita yang berasal dari BBC yang bahasannya meliputi business, teknologi, politik, olahraga dan hiburan. *Update* berita atau informasi menjadi penting karena informasi dapat menjadi dasar bagi pembuatan keputusan yang dibutuhkan untuk dalam beberapa bidang seperti bisnis, ekonomi, dan politik. Sangat penting untuk memiliki metode efektif untuk mengakses dan memanfaatkan informasi tersebut secara efisien.

Klasifikasi berita atau informasi yang populer dilakukan dengan tujuan untuk memudahkan pengaksesan dan mengorganisasi informasi yang banyak dan beragam. Klasifikasi ini berguna untuk mempermudah pengaksesan informasi, melancarkan efisiensi dalam pencarian, dan mengurangi kesulitan dalam mengelompokkan berita yang banyak dan beragam.

Dataset yang digunakan adalah data berita BBC News yang memiliki dokumentasi resmi dan penelitian terkait yang dapat membantu penelitian ini, dataset ini juga dalam bentuk format yang tidak terlalu rumit yang memberikan kemudahan dalam proses preprocessing tanpa memerlukan langkah-langkah ekstraksi data yang rumit. Selain itu, data berita *BBC News* telah terkategori ke dalam lima kategori utama yaitu bisnis, teknologi, politik, olahraga dan hiburan.

Dengan menggunakan dataset ini, penelitian bertujuan untuk mengevaluasi kinerja kedua metode klasifikasi *Naïve Bayes* dan *Convolutional Neural Network (CNN)* berdasarkan berbagai matrik evaluasi seperti akurasi, presisi, recall, dan F1-score.

Tujuan utama dari penelitian ini adalah untuk menentukan pendekatan mana yang lebih efisien dan akurat dalam mengklasifikasikan berita ke dalam kategori yang sesuai dengan menggunakan metode *Naïve Bayes* yang merupakan metode probabilistik sederhana, atau *Convolutional Neural Network (CNN)* yang merupakan model deep learning yang lebih kompleks. Melalui analisis yang mendalam terhadap hasil evaluasi, penelitian ini juga bertujuan untuk mengidentifikasi kelebihan dan keterbatasan dari masing-masing metode klasifikasi dalam konteks data teks berita.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang berharga tentang pilihan metode klasifikasi untuk berita online. Meskipun CNN mungkin menawarkan kinerja yang superior, penelitian ini mungkin menunjukkan bahwa *Naïve Bayes* masih relevan untuk beberapa kasus penggunaan tertentu. Sekian itu, penelitian ini juga dapat membuka peluang untuk eksplorasi lebih lanjut tentang bagaimana kombinasi atau modifikasi dari kedua metode tersebut dapat lebih meningkatkan efektivitas klasifikasi berita di masa depan.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang sudah dijelaskan di atas, maka rumusan masalah yang didapat pada penelitian ini sebagai berikut:

1. Bagaimana perbandingan kinerja antara metode klasifikasi teks *Naïve Bayes* dan *Convolutional Neural Network (CNN)* dalam mengklasifikasikan berita dari *BBC News*?
2. Apakah penggunaan metode *Naïve Bayes* dan *Convolutional Neural Network (CNN)* dapat mengklasifikasi berita lebih akurat?

## **1.3 Batasan Masalah**

Dalam penelitian ini memiliki Batasan masalah sebagai berikut:

1. Pengambilan data berita *BBC News* melalui *Kaggle* berupa *category*, *title*, dan *content* yang bersumber dari: <https://www.kaggle.com/dheemantbbhat/bbc-full-text-preprocessed/data>.
2. Mengukur nilai matrik evaluasi akurasi, presisi, recall, dan *F1-score* pada data berita *BBC News* melalui *Kaggle* sebagai media *Data Science* dan *Machine Learning*.

#### 1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah:

1. Membandingkan efisiensi dan keakuratan pada metode klasifikasi *Naive Bayes* dan *Convolutional Neural Network (CNN)* dalam mengklasifikasikan berita dari *BBC News*.
2. Menjelaskan kelebihan dan kekurangan masing-masing metode klasifikasi *Naive Bayes* dan *Convolutional Neural Network (CNN)*.

#### 1.5 Manfaat Penelitian

Berdasarkan dari tujuan penelitian ini, adapun manfaat dari penelitian yaitu:

1. Hasil dari penelitian ini akan memberikan wawasan tentang pilihan metode klasifikasi untuk berita, membantu organisasi media dan pembuat keputusan untuk memilih pendekatan yang sesuai.
2. Penelitian ini akan membantu meningkatkan efisiensi pengaksesan informasi dengan memberikan metode yang lebih efektif dalam mengklasifikasikan berita.

#### 1.6 Sistematika Penulisan

Susunan yang telah ditulis secara sistematis dalam penelitian ini memuat masalah-masalah pada setiap bab sebagai berikut:

BAB I PENDAHULUAN, berisi Latar belakang masalah, rumusan masalah, Batasan masalah, manfaat penelitian, tujuan penelitian dan sistematika penelitian.

BAB II TINJAUAN PUSTAKA, berisi tinjauan pustaka, dasar-dasar teori yang digunakan berupa jurnal, buku, dan situs web resmi.

BAB III METODE PENELITIAN, didalamnya terdapat tinjauan umum tentang objek penelitian, dan alur penelitian yang meliputi pengumpulan data, Exploratory Data Analysis (EDA), pembuatan model, hasil model, evaluasi model.

BAB IV HASIL DAN PEMBAHASAN, bab ini merupakan tahapan yang penulis lakukan dalam riset klasifikasi data berita.

BAB V PENUTUP, berisi kesimpulan dan saran yang dapat peneliti rangkum selama proses penelitian.

