

BAB I

PENDAHULUAN

1.1 Latar Belakang

Era digital seperti sekarang ini, perkembangannya sedemikian pesat menyentuh hampir di seluruh bidang dan tatanan kehidupan yang sangat terbantu sekali oleh teknologi digital. Internet telah memberikan begitu banyak manfaat, dapat mengetahui informasi tertentu dengan cepat, mudah, bahkan real time. Berkaitan dengan perkembangan teknologi saat ini, pendidikan merupakan salah satu bidang yang paling berpengaruh dalam proses perkembangan teknologi. Di antara lembaga pendidikan, universitas atau perguruan tinggi adalah pusat pengembangan ilmu pengetahuan yang paling kompleks dan besar, salah satu bentuk pemanfaatan teknologi di dunia pendidikan yaitu sistem klasifikasi.

Sistem klasifikasi merupakan proses penentuan sebuah model atau fungsi yang mendeskripsikan perbedaan konsep atau kelas data untuk mengevaluasi kelas dari suatu objek yang labelnya tidak diketahui. Untuk mencapai tujuan ini, proses klasifikasi membuat model yang mampu membedakan data ke dalam kelas yang berbeda berdasarkan aturan atau fungsi tertentu (Faid et al., 2019). Pemanfaatan sistem klasifikasi memiliki bermacam-macam jenis berdasarkan tipe datanya seperti klasifikasi berdasarkan suara menggunakan pendekatan *support vector machine* (Handoko & Suyanto, 2019), klasifikasi berdasarkan gambar menggunakan pendekatan *regionprops* dan *decision tree* (Pamungkas et al., 2020), klasifikasi berdasarkan video menggunakan pendekatan convolutional neural network (CNN) (Yohannes et al., 2022), klasifikasi berdasarkan citra medis menggunakan pendekatan convolutional neural network (CNN) (Gunawan & Setiawan, 2022), klasifikasi SMS berbasis Web menggunakan algoritma Logistic Regression (Pramakrisna et al., 2022) dan klasifikasi text menggunakan pendekatan *Naïve Bayes* (Darujati & Gumelar, 2012). Salah satu contoh lain dari sistem klasifikasi adalah sistem klasifikasi Dewey Decimal

Classification (DDC) dengan pendekatan naïve bayes (Mulyani et al., 2021).

Naïve Bayes adalah salah satu metode machine learning yang digunakan untuk klasifikasi teks dengan kecepatan pemrosesan yang cukup tinggi dalam data besar. Pendekatan ini memiliki probabilistik sederhana yang diterapkan pada teorema bayes dengan ketergantungan yang kuat (Puspita & Widodo, 2021).

Logistic Regression merupakan pendekatan machine learning yang cukup terkenal digunakan untuk tugas klasifikasi (Gunawan, Muhammad Ichan Sugiarto & Mardianto, 2020). Logistic Regression ialah klasifikasi lini yang sudah teruji menciptakan klasifikasi yang powerful dengan statistik probabilitas serta bisa menanggulangi permasalahan klasifikasi multi kelas. Kelemahan dari Logistic Regression merupakan rentan terhadap underfitting pada dataset yang kelasnya tidak balance, sehingga mempunyai akurasi yang rendah (Alfarobi et al., 2023).

Proses klasifikasi DDC yang dilakukan seringkali menerapkan model-model tertentu didalamnya. Model-model yang diterapkan memberikan hasil yang berbeda-beda untuk kegiatan klasifikasi DDC yang dilakukan. Dalam penelitian ini, kegiatan klasifikasi DDC yang dilakukan menggunakan model *Naïve Bayes* dan *Logistic Regression*. Penelitian ini dilakukan untuk menguji seberapa baik model tersebut dalam melakukan kegiatan klasifikasi DDC. Dikarenakan hasil akurasi dari model naïve bayes dan logistic regression belum diketahui seberapa baik dalam melakukan klasifikasi DDC, untuk itulah dilakukan pengujian model agar dapat mengetahui hasil akurasi pengujian dari model tersebut. Melalui permasalahan tersebut, peneliti tertarik untuk melakukan penelitian yang berjudul **“PENERAPAN ALGORITMA NAÏVE BAYES DAN LOGISTIC REGRESSION DALAM KLASIFIKASI DATA DEWEY DECIMAL CLASIFICATION (DDC) INDONESIA”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah pada penelitian ini yaitu “Bagaimana cara melakukan klasifikasi data Dewey Decimal Clasification (DDC) dengan menerapkan Algoritma *Naïve Bayes* dan *Logistic Regression*?”.

1.3 Tujuan Penelitian

Tujuan dari penelitian ini ialah untuk mengetahui sejauh mana kemampuan model Naive Bayes dan Logistic Regression dalam melakukan pengklasifikasian data Dewey Decimal Classification (DDC).

1.4 Batasan Masalah

Dalam penelitian ini diperlukan adanya batasan-batasan masalah sehingga permasalahan pada penelitian ini dapat terfokus. Adapun batasan masalah yang akan diterapkan adalah klasifikasi berdasarkan informasi judul dan kata kunci dengan menerapkan Algoritma Naïve Bayes sehingga menemukan nilai akurasi tidak sampai pada pembuatan sistemnya.

1.5 Manfaat Penelitian

Dengan dilakukannya penelitian ini diharapkan bisa memberikan manfaat bagi pengguna sebagai berikut :

- a. Mengetahui kemampuan model naïve bayes dan logistic regression dalam melakukan klasifikasi data *Dewey Decimal Classification* (DDC)
- b. Model yang dilatih dapat menjadi referensi bagi peneliti selanjutnya dalam mengembangkan sistem klasifikasi *Dewey Decimal Classification* (DDC).