

# BAB I

## PENDAHULUAN

### A. Latar Belakang

Pemanfaatan AI (*Artificial Intelligence*) di berbagai bidang seperti pendidikan, kesehatan, keamanan, dan lain - lain, kini sangatlah marak ditemukan. Hal ini dikarenakan tingkah laku AI yang dapat mengurangi kesalahan dalam bekerja dengan mempelajari data-data yang telah ada sebelumnya, lalu akan mengimprovisasi dirinya sendiri seiring dengan banyaknya data yang dipelajari. Salah satu teknologi turunan dari AI adalah *Machine Learning*. *Machine Learning* adalah sebuah pendekatan yang fokus pengaplikasiannya adalah memprediksi [1]. *Machine Learning* juga banyak diterapkan di berbagai bidang, seperti di bidang kesehatan untuk mendeteksi pasien yang terkena diabetes, di bidang pendidikan untuk pengembangan *Automatic Assessment*, dan lain-lain.

Dalam penelitian ini, penulis akan berfokus pada penyelesaian masalah menggunakan *Machine Learning*. *Machine Learning* dipilih karena dapat menyelesaikan dengan baik suatu permasalahan yang memiliki banyak data sebagai informasinya. Kasus yang akan diteliti menggunakan *machine learning* oleh penulis adalah *SMS Spamming*. Singkatan SMS berasal dari kependekan dalam bahasa Inggris yaitu *Short Message Service*. SMS adalah salah satu metode komunikasi perangkat bergerak (*mobile*) yang paling umum [2]. SMS spam secara spesifik adalah jenis pesan teks yang tidak diinginkan atau tidak diminta yang dikirim ke ponsel pengguna, seringkali untuk tujuan komersial. Pesan tersebut dapat berupa pesan sederhana, tautan ke nomor untuk menelepon atau SMS, tautan ke situs web untuk informasi lebih lanjut, atau tautan ke situs web untuk mengunduh aplikasi [3].

Metode yang dapat digunakan untuk mengatasi *sms spamming* adalah *machine learning*. *Machine learning* dapat mengekstraksi informasi dari teks dengan menggunakan berbagai tipe algoritma statistik dengan proses yang dikelompokkan ke dalam *text mining*, dan *text analytics* [4]. Secara khusus,

algoritma yang akan digunakan untuk mengatasi masalah SMS *Spamming* adalah *Logistic Regression*. *Logistic Regression* dapat menyelesaikan masalah SMS Spam karena algoritma ini sangat baik dalam melakukan klasifikasi untuk *binary classification*, yaitu mengklasifikasi label untuk nilai 1 dan 0. Pada kasus ini, permasalahan yang akan diteliti adalah memprediksi sms mana yang spam (1) dan mana yang tidak spam (0).

Hal ini juga diperkuat dengan adanya penelitian sebelumnya mengenai prediksi pasien pengidap diabetes melitus yang ditulis oleh Muhammad Ichsan Gunawan, Dedy Sugiarto, dan Is Mardianto pada tahun 2020, dalam jurnal penelitiannya yang berjudul *Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Search Pada Algoritma Logistic Regression* [5]. Pada jurnal tersebut menghasilkan model *Logistic Regression* dengan *Grid Search* yang memiliki rata – rata akurasi model pada *Classification Report* sekitar 79% dan pada akurasi *data check* sebesar 83,33%. Penelitian lain yang diteliti oleh Lang Wu, dan Menggang Li yang berjudul *Mengaplikasikan CG-Logistic Regression untuk Memprediksi Masalah Customer Churn* [6]. Pada penelitian tersebut memberikan hasil dari 3 dataset *customer churn* yang dievaluasi menggunakan *Logistic Regression*, skor yang didapat pada ketiga dataset tersebut berada diatas kisaran diatas 89%.

Penulis nantinya akan mengimplementasikan model dari *machine learning* dengan algoritma *Logistic Regression* tersebut ke dalam bentuk sebuah aplikasi web sederhana. Web memiliki pengertian, yaitu fasilitas internet yang menghubungkan dokumen dalam lingkup lokal maupun jarak jauh yang menyimpan informasi tertentu seperti iklan, *link* bersponsor, *headers*, dan *footers* [7]. Pembuatan aplikasi web akan menggunakan bantuan Flask. Flask adalah sebuah kerangka kerja (*framework*) pada bahasa pemrograman Python [8]. Aplikasi web dipilih sebagai media pengaplikasian *machine learning* karena mempermudah dalam melakukan pengecekan hasil output dari prediksi *machine learning*, hanya dengan memasukan pesan teks, lalu hasil prediksi akan muncul. Aplikasi web *machine learning* ini nantinya akan menerima hasil input berupa pesan sms, lalu sistem akan memprediksi apakah input tersebut spam atau tidak spam. Untuk

menentukan sms mana yang spam dan tidak spam, hal tersebut dapat ditentukan dengan mempertimbangkan seberapa banyak kata / *term* yang sama muncul dalam satu dokumen (*text*). Jika dalam satu dokumen terdapat 1 kata yang sama muncul beberapa kali, maka itu adalah spam. Sebaliknya, SMS itu adalah bukan spam.

## **B. Rumusan Masalah**

Rumusan masalah yang peneliti akan coba teliti adalah mengenai masalah SMS spam yang didapat pengguna, terutama untuk pengguna ponsel. Pesan tersebut memuat jenis pesan yang tidak diinginkan pengguna dan seringkali untuk tujuan komersial. Solusi yang digunakan untuk menyelesaikan masalah ini adalah dengan menggunakan *machine learning*.

## **C. Batasan Masalah**

Batasan masalah diperlukan untuk membuat topik penelitian yang dibahas lebih terarah. Beberapa batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Kasus yang diteliti adalah SMS *Spamming*.
2. Algoritma *machine learning* yang digunakan adalah *Logistic Regression*.
3. Pengimplementasian dari model *machine learning* akan diterapkan pada aplikasi web.
4. Dataset untuk penelitian diperoleh dari penelitian skripsi peneliti lain [9].
5. Jumlah data pesan teks yang ada pada dataset, berjumlah 1140 pesan.
6. Data pesan teks berasal dari tahun 2016.

## **D. Pertanyaan Masalah**

Berdasarkan rumusan masalah dan batasan masalah yang ada, berikut adalah beberapa pertanyaan untuk menanggapi kasus ini:

1. Bagaimana *machine learning* dapat mendeteksi *sms spamming* ?
2. Seberapa akurat algoritma *Logistic Regression* menggunakan model evaluasi *Accuracy* pada *machine learning* dalam memprediksi pesan spam dan tidak spam?

3. Bagaimana pengimplementasian model *machine learning* ke dalam aplikasi web dilakukan ?

#### **E. Tujuan Penelitian**

Adapun tujuan penelitian ini adalah sebagai berikut :

1. Untuk mengetahui bagaimana *machine learning* mendeteksi SMS mana yang spam dan tidak spam.
2. Untuk mengetahui tingkat akurasi yang dihasilkan algoritma *Logistic Regression* pada klasifikasi sms spam.
3. Untuk mengetahui pengimplementasian model *machine learning* ke dalam bentuk aplikasi web.

#### **F. Manfaat Penelitian**

Berikut adalah manfaat – manfaat yang diharapkan oleh penulis dari penelitian ini :

1. Memberikan opsi penyelesaian masalah untuk sms spam, yaitu dengan menggunakan *machine learning*
2. Memberikan salah satu contoh untuk mengimplementasikan model *machine learning* ke dalam bentuk aplikasi yang bisa diakses banyak orang.