

BAB I

PENDAHULUAN

1.1. Latar Belakang

Short Message Service (SMS), yang juga dikenal sebagai layanan pesan singkat, tetap menjadi salah satu alat komunikasi jarak jauh yang banyak digunakan di era saat ini untuk mengirim pesan singkat. Seiring dengan kemajuan layanan pesan singkat ini, timbul dampak negatif berupa serangan pada perangkat seluler dalam bentuk SMS spam. SMS spam adalah pesan singkat yang tidak diinginkan oleh penerima, termasuk pesan berisi iklan dan upaya penipuan (*fraud*) (DwiYansaputra, 2021).

Menurut laporan yang dipublikasikan dalam *Truecaller Insights Report* (2020), Indonesia mencatatkan jumlah pesan spam tertinggi di Asia pada tahun 2020. Angka pesan spam di Indonesia menurun sebanyak 34% dari 27,9% pada tahun 2019 menjadi 18,3%, dan mengalami penurunan peringkat ke urutan keenam di antara 20 negara dengan tingkat pesan spam tertinggi. Jenis pesan spam yang paling umum di Indonesia adalah layanan keuangan, mencapai 52% dari total pesan spam, diikuti oleh asuransi sebesar 25%, operator seluler 11%, penipuan (*scam*) 9%, dan tagihan hutang 3% (Herwanto, 2021). Selain itu, berdasarkan laporan terbaru yang disajikan dalam *Truecaller Insights Report* (2021), posisi Indonesia tetap tak berubah dari tahun sebelumnya, yaitu masih berada di urutan keenam dari 20 negara di Asia. Jenis pesan spam yang dominan di Indonesia, terutama dalam layanan keuangan, mengalami peningkatan signifikan, mencapai 80% dari sebelumnya 52%.

Dalam konteks ini tindakan proaktif dan langkah-langkah yang efektif dapat membantu meminimalkan dampak dari pesan-pesan spam yang mengganggu. Fitur

linguistik dan statistik memiliki peran yang signifikan dalam mengklasifikasi SMS spam. Fitur linguistik melibatkan analisis struktur, pola, dan penggunaan bahasa dalam teks, sementara fitur statistik merepresentasikan teks menjadi angka agar mengetahui bobot yang ada pada teks (AL-Jumaili, 2020).

Dalam konteks pengklasifikasian SMS spam. KNN merupakan teknik yang sering digunakan untuk klasifikasi data. Cara kerja dari metode ini adalah dengan mengklasifikasikan data berdasarkan seberapa dekat jarak antara satu data dengan data lain (Laksono, 2020). Dalam kasus SMS spam, KNN akan mengidentifikasi dan mengklasifikasi pesan baru berdasarkan fitur linguistik dan statistik.

Di sisi lain, *Neural Network* adalah model yang terinspirasi dari jaringan saraf manusia. Model ini terdiri dari *input layer*, satu atau lebih *hidden layer*, dan *output layer*. Setiap neuron dalam jaringan ini mengolah informasi menggunakan fungsi aktivasi dan menyesuaikan bobotnya berdasarkan data yang diberikan (Yuliana, 2019). Dalam konteks SMS spam, NN dapat mempelajari pola kompleks dalam data, memungkinkan identifikasi pesan spam dengan lebih akurat berdasarkan fitur linguistik dan statistik

Berdasarkan latar belakang tersebut kontribusi penelitian ini akan berfokus pada pemodelan untuk membandingkan fitur statistik dan linguistik pada algoritma KNN dan *Neural Network* dalam mengklasifikasi SMS spam dan evaluasi perbandingan algoritma menggunakan *classification report* guna mengetahui performa dari masing masing model.

1.2. Rumusan Masalah

Berdasarkan latar belakang diatas, berikut rumusan masalah:

1. Bagaimana tahapan perbandingan fitur statistik dan linguistik pada algoritma KNN dan NN terhadap SMS spam?

2. Bagaimana hasil evaluasi perbandingan fitur statistik dan linguistik pada algoritma KNN dan NN terhadap SMS spam?

1.3. Batasan Masalah

Berdasarkan masalah yang dibahas, penelitian ini memiliki batasan-batasan tertentu sebagai berikut:

1. Terdapat jenis bahasa pada SMS yang digunakan yaitu Bahasa Indonesia maupun *slank*.
2. Fitur statistik dan fitur linguistik diterapkan pada masing masing algoritma klasifikasi.
3. Hasil penelitian dari perbandingan fitur statistik dan linguistik pada algoritma *K-Nearest Neighbors* dan *Neural Network* dalam pengklasifikasian SMS spam yaitu berupa *classification report*, dan mengetahui SMS berupa spam (penipuan dan promo) atau ham.
4. *Output* dari keluaran yaitu spam penipuan, spam promo dan ham (normal).

1.4. Tujuan Penelitian

Adapun tujuan penelitian yang diangkat. Berikut adalah tujuannya:

1. Mengetahui tahapan perbandingan fitur statistik dan linguistik pada algoritma KNN dan NN terhadap SMS spam.
2. Mengetahui hasil evaluasi perbandingan fitur statistik dan linguistik pada algoritma KNN dan NN terhadap SMS spam.

1.5. Manfaat Penelitian

Berdasarkan latar belakang yang dijelaskan, Adapun manfaat penelitian yaitu:

1. Menambah wawasan terhadap algoritma yang digunakan.
2. Dengan adanya penelitian ini, kita dapat meminimalisir gangguan dan kebingungan yang disebabkan oleh pesan-pesan tidak diinginkan.

1.6. Sistematika Penulisan

Untuk memudahkan pembahasan dalam skripsi ini, sistematika penulisan dibagi menjadi 5 (lima) bab yang terdiri dari:

BAB I PENDAHULUAN

Bab ini mencakup konteks dan alasan di balik penelitian (latar belakang), termasuk perumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bagian ini menguraikan berbagai teori dan konsep dari sumber-sumber terkait yang menjadi dasar referensi utama dalam melakukan penelitian ini.

BAB III METODOLOGI PENELITIAN

Dalam bab ini, dijelaskan tentang pendekatan dan teknik yang digunakan oleh penulis dalam menangani penelitian sesuai dengan masalah yang diangkat.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil yang diperoleh berupa hasil evaluasi dari model yang dibuat serta pengujian dalam mengklasifikasi sebuah pesan baru.

BAB V PENUTUP

Bab ini berisi kesimpulan yang didapatkan dari analisis model yang didapatkan serta saran untuk peneliti yang ingin melanjutkan penelitian ini.