

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada saat ini pemanfaatan internet di Indonesia berkembang dengan sangat baik. Berbagai pemanfaatan internet dalam pelayanan publik yang diselenggarakan oleh pihak swasta atau pemerintah. Salah satu pelayanan publik yang memanfaatkan internet adalah pengaduan masyarakat yang dapat menggunakan berbagai media sosial, SMS, website. Saat ini terdapat sistem laporan pengaduan masyarakat yang sudah terintegrasi ke berbagai pihak dan instansi. Sistem ini dibentuk oleh Pemerintah Republik Indonesia bernama Sistem Pengelolaan Pengaduan Pelayanan Publik Nasional (SP4N) – Layanan Aspirasi dan Pengaduan Online Rakyat (LAPOR!) [1]. Sistem tersebut dapat diakses melalui website www.lapor.go.id. Berdasarkan sistem SP4N-LAPOR! rata-rata jumlah laporan masuk harian adalah 570 laporan, total pengguna yang terdaftar per Januari 2019 adalah 801.257 pengguna, dan jumlah laporan masuk sejak 2012 berjumlah 1.389.891 laporan. Sistem SP4N-LAPOR! telah terhubung dengan 34 Kementerian, 96 Lembaga, dan 493 pemerintah daerah.

Dengan sistem SP4N-LAPOR! memudahkan dalam pengelolaan laporan pengaduan karena terpusat dan terintegrasi ke berbagai Kementerian, Lembaga, dan Pemerintah Daerah. Alur proses laporan di sistem tersebut dimulai dengan pengguna memasukkan laporan, laporan diverifikasi minimal 3 hari, laporan diteruskan ke pihak terkait dan akan diproses minimal 5 hari. Lamanya proses verifikasi dikarenakan perlu pengecekan laporan oleh admin untuk mengetahui isi dan pihak yang dituju dari laporan. Selain itu jumlah rata-rata harian 570

laporan yang masuk membuat selalu bertambah data yang harus diverifikasi sehingga membuat proses verifikasi bertambah lama. Berdasarkan permasalahan tersebut terdapat beberapa penelitian dengan memanfaatkan machine learning dalam melakukan klasifikasi.

Pada penelitian [2]. Menggunakan metode C4.5 berbasis forward selection. Dari penelitian diketahui bahwa algoritma C4.5 dengan menggunakan forward selection bisa meningkatkan akurasi klasifikasi yakni sebesar 85,27%. Penggunaan dataset juga tidak berpengaruh pada kenaikan akurasi pada klasifikasi dengan seleksi fitur, hal ini bisa dilihat pada pengujian 275 data, nilai akurasi tertinggi 81,09%. Sedangkan pada pengujian 550 data nilai akurasi tertinggi 85,27%. Tetapi pada pengujian tanpa seleksi fitur, jumlah tidak berpengaruh pada hasil akurasi, hal ini bisa dilihat dari hasil pengujian 1100 data, nilai tertinggi 78,81%, kemudian naik pada pengujian 550 data sebesar 81,09%, tetapi turun kembali pada pengujian 275 data menjadi 71,27%.

Pada penelitian [3]. Menggunakan algoritma Naive Bayes Classifier dalam klasifikasi laporan pengaduan. Dari penelitian tersebut diperoleh beberapa kesimpulan diantaranya jumlah kategori mempengaruhi kinerja klasifikasi menggunakan Naive bayes. Tingkat kemiripan diantara kategori satu dengan yang lain mempengaruhi tingkat klasifikasi jika tingkat kemiripan diantara dua kategori tinggi maka akan sulit membedakan kedua kategori tersebut. Penggunaan stopwords dan stemming dapat meningkatkan tingkat akurasi klasifikasi. Pada penelitian tersebut menggunakan 79 dokumen latih dan 29 data uji, diperoleh 27 dokumen relevan dan 2 dokumen tidak relevan dengan hasil

tingkat *accuracy* dengan *recall* sebesar 93%, *precision* sebesar 93%, *f-measure* sebesar 92% , dan rata-rata akurasi sebesar 92,67%.

Bedasarkan penelitian diatas maka *text mining* merupakan proses yang tepat karena data laporan yang masuk berupa teks. *Text mining* merupakan proses menggali informasi secara intensif dimana pengguna berhadapan dengan sekumpulan dokumen menggunakan tools analisis data mining [1]. Dengan *text mining* maka akan dilakukan *text classification* untuk mengklasifikasikan setiap laporan ke kategori yang sesuai. Dalam penelitian diatas belum adanya tahap perlakuan terhadap kata slang/gaul sedangkan laporan pengaduan dapat berasal darimana saja dan semua kalangan yang memungkinkan terdapat bahasa slang/gaul dalam isi pengaduan. Pada penelitian ini peneliti akan menggunakan *Artificial Neural Networks* dalam melakukan klasifikasi karena mampu mempelajari model non-linier dan data yang besar dan melakukan proses *replace slang words* pada tahap *preprocessing data*. Selain itu dalam penelitian ini menggunakan tiga kelas kategori pengaduan yaitu kategori pendidikan, kesehatan, dan infrastruktur.

1.2 Rumusan Masalah

Bedasarkan latar belakang yang telah disampaikan, maka diperoleh rumusan masalah antara lain sebagai berikut:

1. Bagaimana mengklasifikasikan laporan dengan menggunakan *neural network* ?
2. Perfoma algoritma *neural network* dalam klasifikasi laporan pengaduan ?
3. Pengaruh penambahan *replace slang words* pada *preprocessing data* terhadap perfoma *neural network* dalam klasifikasi?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini sebagai berikut:

1. Data pengaduan yang digunakan berasal dari LAPOR! yang diakses melalui situs laporan www.lapor.go.id.
2. Kategori laporan yang digunakan adalah Pendidikan, Kesehatan, dan Infrastruktur.
3. Jumlah dataset sebanyak 750 data aduan yang dibagi sama banyak di setiap kategori yaitu 250 data.
4. Dataset yang digunakan dengan format CSV.
5. Data aduan yang diambil mempunyai rentang waktu tahun 2018 sampai oktober 2021.
6. Menggunakan dua kamus bahasa slang / gaul dari <https://raw.githubusercontent.com/fendiirfan/Kamus-Alay/main/Kamu-Alay.csv> dan <https://raw.githubusercontent.com/nasalsabila/kamus-alay/master/colloquial-indonesian-lexicon.csv>.
7. Evaluasi model menggunakan Confusion Matrix.

1.4 Maksud dan Tujuan Penelitian

Tujuan Penelitian ini diantara lain sebagai berikut:

1. Melakukan pengujian algoritma *neural network* sebagai klasifikasi pengaduan laporan.
2. Mengetahui performa algoritma *neural network* dalam melakukan klasifikasi laporan pengaduan.
3. Mengetahui pengaruh *replace slang words* terhadap performa model klasifikasi.

1.5 Manfaat Penelitian

Manfaat yang diperoleh setelah melakukan penelitian ini sebagai berikut:

- a. Dapat mengetahui hasil algoritma *neural network* dalam mengklasifikasikan laporan pengaduan.
- b. Dapat mengetahui pengaruh *replace slang word* pada preprocessing terhadap klasifikasi pengaduan dengan *neural network*.

1.6 Metodologi Penelitian

Untuk memperoleh data yang benar serta relevan dalam pemecahan masalah sesuai dengan kebutuhan untuk penelitian ini, maka diperlukannya suatu metode yang tepat, agar dapat tercapai sesuai dengan tujuan penelitian.

1.6.1 Metode Pengumpulan data

Metode pengumpulan data dengan memanfaatkan fasilitas internet untuk mengunjungi website atau literature berupa jurnal ilmiah, e-book dan berbagai referensi yang berhubungan dengan penelitian.

1.6.2 Metode Analisis

Metode analisis yang digunakan dalam penelitian ini dimulai tahap 1 dengan mengidentifikasi masalah, tujuan, batasan, dan metodologi dalam penelitian. Tahap 2 melakukan kajian pustaka dan studi literature mengenai penelitian yang dilakukan. Tahap 3 mengumpulkan dataset dan memproses dataset. Tahap 4 melakukan perancangan model klasifikasi. Tahap 5 melakukan pengujian dan evaluasi model.

1.6.3 Metode Perancangan

Metode perancangan yang digunakan dalam penelitian mempunyai beberapa tahapan. Tahap pertama pengumpulan data pengaduan sebagai dataset. Tahap kedua *preprocessing data* meliputi *data cleaning*, *case folding*, *replace slang words*, *stopword removal*, *stemming*, dan *tokenizing*. Tahap ketiga pembobotan kata menggunakan TF-IDF. Tahap keempat membagi *data training* dan *data test* dengan

menggunakan *split validation*. Tahap kelima pembuatan model *neural network* dan *training* model dengan *data training*. Tahap keenam melakukan pengujian model dengan *data test* dan melakukan evaluasi dengan *confusion matrix*.

1.6.4 Metode Pengujian

Metode pengujian dengan cara menggunakan skenario uji model yang menggunakan *preprocessing replace slang word* dan tidak menggunakan *preprocessing replace slang word*. Skenario pengujian model dilakukan sebanyak 3 kali dengan epoch yang berbeda. Hasil pengujian model akan di evaluasi dengan *confusion matrix*.

1.7 Sistematika Penulisan

Penyusunan dalam skripsi ini dengan menggunakan sistematika pembahasan yang tersusun sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan dalam skripsi.

BAB II LANDASAN TEORI

Bab ini menjelaskan tentang dasar teori serta penelitian sebelumnya yang berkaitan dengan penelitian ini dan literature yang digunakan seperti *Text Mining*, *Klasifikasi Teks*, *Preprocessing Text*, *Machine Learning*, *Neural Network*, *Confusion matrix*, *Keras*.

BAB III ANALISIS dan PERANCANGAN

Bab ini menjelaskan metode dan langkah-langkah yang digunakan dalam penelitian. Dalam bab ini terdapat alur penelitian dan alur

klasifikasi, dengan ini maka diharapkan penelitian dapat berjalan dengan baik.

BAB IV HASIL dan PEMBAHASAN

Bab ini menjelaskan mengenai hasil dari analisis yang dilakukan dan pembahasan dari hasil yang didapatkan yaitu berupa hasil model jaringan *neural network* dari dataset yang sudah dilatih. Hasil tersebut berupa performa yang didapat serta evaluasi model.

BAB V PENUTUP

Bab ini berisikan mengenai kesimpulan dari hasil penelitian yang telah didapatkan dari hasil keseluruhan yang telah dilakukan, serta saran untuk peneliti selanjutnya yang akan mengembangkan penelitian ini.

