

BAB I

PENDAHULUAN

1.1 Latar Belakang

Asuransi adalah sebuah mekanisme perlindungan terhadap pihak tertanggung apabila mengalami resiko di masa yang akan datang dimana pihak tertanggung akan membayar premi guna mendapatkan ganti rugi dari pihak penanggung. Sejatinya, asuransi berfungsi sebagai perlindungan jika hal-hal yang tidak diinginkan terjadi. Namun beberapa pihak yang tidak bertanggung jawab melihat asuransi adalah sebagai sebuah kesempatan untuk mendapatkan keuntungan dengan memanipulasi data atau keterangan untuk mendapatkan keuntungan atau manfaat dari klaim asuransi. Inilah yang biasa disebut sebagai *fraud* dalam asuransi.

Banyaknya nasabah atau dalam kasus kali ini koperasi yang tidak jujur dalam memberikan informasi membuat asuransi melakukan berbagai cara dalam memvalidasi data atau penyebab meninggal atau klaim terjadi. Salah satu variabel data yang dilabeli atau dicek oleh asuransi adalah kronologi kematian. Saat ini, perusahaan asuransi melakukan pengecekan kronologi secara manual. Dicocokkan data ketika nasabah mendaftar dan penyebab meninggal nya. Hal utama yang dicek adalah apakah nasabah memiliki penyakit kronis atau tidak ? Karena aturan yang ada dalam proyek asuransi ini adalah, jika nasabah memiliki atau meninggal karena penyakit kronis, maka klaim asuransi tidak dapat dicairkan.

Pada saat mendaftar, nasabah diberikan pertanyaan apakah anda memiliki riwayat penyakit kronis ? Kebanyakan kasus yang terjadi adalah nasabah menjawab tidak saat mendaftar, namun penyebab meninggal nya adalah penyakit yang digolongkan kronis. Perbedaan data ini biasanya dilabeli oleh karyawan asuransi sebagai potensi penyakit kronis dan nanti akan diberikan treatment tambahan terhadap polis asuransi nya. Pengecekan kronologi klaim akan terlihat mudah jika polis yang mengajukan klaim tidak dalam jumlah banyak. Namun, pada saat ini perusahaan asuransi menghadapi banyak nya laporan klaim yang jumlah nya mencapai ribuan dan harus dicek satu per satu secara detail. Pada saat ini saja, ada total 10.000 laporan klaim dengan akumulasi nilai klaim mencapai Rp 43.380.813.206 dengan yang sedang di proses. Bukti bahwa fraud itu ada, bisa dilihat dari *loss ratio* asuransi yang mencapai 82,54% yang menunjukkan bahwa klaim nya ini tidak normal. Batas normal *loss ratio* klaim asuransi adalah 40% menurut perusahaan asuransi ini.

Untuk mengatasi permasalahan diatas, saya ingin memberikan solusi untuk perusahaan asuransi dalam hal membantu untuk mengidentifikasi atau melabeli serta mendeteksi pola pengisian data riwayat penyakit kronis nasabah berdasarkan kronologi yang dicantumkan. Dalam penelitian ini, akan dibandingkan dua metode text mining yaitu k-nearest neighbor dan naïve bayes classifier untuk menentukan metode mana yang lebih efektif dan akurat untuk diterapkan di kasus kali ini. Perbandingan ini didasarkan pada penelitian yang dilakukan oleh Asilah Salma dan Widya Silfianti pada jurnal ilmiah yang berjudul Sentiment Analysis of User Reviews on Covid-

19 Information Applications Using Naïve Bayes Classifier, and K-Nearest Neighbor menyimpulkan naïve bayes classifier mendapatkan hasil yang lebih akurat sebesar 72,3% sementara k-nearest neighbor sebesar 59,1%. Sementara di penelitian lain, yang dilakukan oleh Asdar Mustofa dan Rice Novita dalam jurnal nya yang berjudul Klasifikasi Sentimen Masyarakat Terhadap Pemberlakuan Pembatasan Kegiatan Masyarakat Menggunakan Text Mining Pada Twitter menyimpulkan bahwa k-nearest neighbor memberikan hasil akurasi yang lebih baik dengan nilai 79,67% sementara naïve bayes classifier sebesar 78,86%.

Dengan dasar acuan diatas, penulis ingin mengambil judul **“PERBANDINGAN METODE K-NEAREST NEIGHBOR DAN NAÏVE BAYES UNTUK TEXT MINING PADA KLAIM ASURANSI UNTUK MENDETEKSI POLA RIWAYAT PENYAKIT KRONIS”** yang diharapkan dapat membantu perusahaan asuransi dalam mendeteksi Riwayat penyakit kronis dan mempercepat pengecekan data. Serta perbandingan metode yang dilakukan akan menentukan metode mana yang akan diterapkan pada permasalahan kali ini.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas, maka rumusan masalah dalam penelitian dan pembuatan aplikasi ini adalah sebagai berikut :

1. Bagaimana perbandingan kinerja K-Nearest Neighbors (KNN) dan Naive Bayes Classifier (NBC) dalam melakukan klasifikasi pada dataset yang diberikan?

2. Metode mana yang lebih sesuai untuk diterapkan dalam pembuatan aplikasi text mining pada kasus pendeteksian pola riwayat penyakit kronis?

1.3 Batasan Masalah

Untuk mengkhususkan ruang lingkup penelitian, maka akan diberikan suatu batasan-batasan masalah sebagai berikut :

1. Aplikasi membandingkan beberapa variable perbandingan diantaranya dari sisi performa waktu, akurasi hasil, precision, recall, specificity, dan f1 score.
2. Aplikasi ini berbasis web dengan PHP, HTML, CSS, Javascript, framework Laravel, React, library PHP-ML, dan Sastrawi.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Tujuan pembuatan program aplikasi pada penelitian ini adalah sebagai berikut.

1. Untuk mengetahui kelebihan dan kekurangan dari masing-masing metode klasifikasi.
2. Untuk menentukan metode yang lebih sesuai untuk digunakan dalam pembuatan aplikasi text mining.
3. Membantu perusahaan asuransi dalam melakukan pengecekan potensi riwayat penyakit kronis secara otomatis.

1.4.2 Manfaat

1. Hasil penulisan laporan tugas akhir ini dapat menjadi tambahan referensi untuk penulisan dan penelitian selanjutnya.

2. Hasil penelitian ini dapat membantu perusahaan asuransi dalam melakukan pengecekan data klaim secara lebih cepat dan akurat.

1.5 Metodologi Penelitian

Untuk menyusun laporan tugas akhir yang baik, maka diperlukan data-data yang akurat agar dapat menghasilkan suatu laporan yang baik dan benar. Adapun metode pengumpulan data yang digunakan oleh peneliti adalah:

1. Metode Observasi

Metode observasi dilakukan dengan pengamatan langsung ke perusahaan asuransi terkhusus ke tim atau karyawan yang melakukan pengecekan data klaim

2. Metode Wawancara

Metode wawancara dilakukan dengan melakukan dialog dengan karyawan yang melakukan pengecekan data klaim termasuk kronologi.

3. Metode Pengambilan Data

Data diambil langsung melalui sistem yang digunakan oleh perusahaan asuransi tentu saja dengan seizin pemangku jabatan perusahaan.

4. Metode Studi Pustaka

Tahapan ini adalah kebutuhan penulis dalam melengkapi data-data yang sudah didapat dan dipelajari yaitu dengan mempelajari catatan jurnal serta buku-buku referensi agar dapat menunjang hasil laporan.

1.6 Sistematika Penulisan

Pada penulisan skripsi ini, akan dipergunakan sistematika penulisan sebagai berikut :

BAB I : PENDAHULUAN

Bagian ini berisikan informasi mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat, metode yang digunakan dan sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini menjelaskan tentang teori – teori yang berhubungan dengan penulisan laporan tugas akhir, yaitu definisi UML (Unified Modeling Language), Algoritma Collaborative Filtering dan lain sebagainya

BAB III : DESAIN DAN PERANCANGAN SISTEM

Pada bab ini berisi tentang implementasi program yang telah dihasilkan, gambaran umum system dan evaluasi mengenai system yang telah dirancang dan dibuat.

BAB IV : PENUTUP

Bagian ini berisi mengenai kesimpulan yang dapat diambil dari penyusunan tugas akhir, serta saran-saran penulis yang diharapkan dapat bermanfaat bagi pihak-pihak lain yang berkepentingan.