

BAB I

PENDAHULUAN

1.1 Latar Belakang

Industri asuransi memiliki peran strategis dalam memberikan perlindungan finansial bagi individu maupun perusahaan melalui mekanisme pengalihan risiko. Dalam konsep asuransi, pihak tertanggung membayarkan sejumlah premi untuk memperoleh perlindungan terhadap risiko tertentu, sehingga ketika risiko yang diasuransikan terjadi, pihak tertanggung dapat mengajukan klaim berdasarkan ketentuan polis yang telah disepakati[1]. Oleh karena itu, pengelolaan klaim menjadi salah satu aspek penting dalam operasional perusahaan asuransi karena berkaitan dengan pengendalian risiko, stabilitas keuangan, serta pelayanan kepada nasabah.

Seiring dengan meningkatnya perkembangan industri asuransi, kompleksitas dalam pengelolaan data polis dan klaim juga semakin meningkat. Perusahaan asuransi menghadapi berbagai tantangan seperti penilaian risiko, klasifikasi pemegang polis, prediksi klaim, serta potensi fraud yang sulit dianalisis hanya menggunakan pendekatan konvensional[2]. Selain itu, karakteristik klaim yang memiliki variasi nilai berbeda menyebabkan perusahaan membutuhkan metode analisis yang mampu memahami pola klaim secara lebih akurat, terutama karena klaim dengan nilai besar dapat berpengaruh terhadap pengelolaan cadangan perusahaan[3].

Selain proses klaim, polis asuransi juga memiliki peran penting sebagai sumber informasi dalam melakukan analisis risiko. Data polis yang berisi informasi mengenai karakteristik tertanggung, jenis perlindungan, premi, serta histori klaim dapat digunakan untuk membantu perusahaan memahami hubungan antara profil risiko dengan potensi nilai klaim. Kemampuan perusahaan dalam mengelola risiko berdasarkan informasi tersebut menjadi faktor penting untuk mengurangi potensi kerugian dan meningkatkan efektivitas proses bisnis asuransi[4].

Namun, meningkatnya jumlah dan kompleksitas data menyebabkan metode analisis tradisional memiliki keterbatasan dalam mengidentifikasi hubungan antarvariabel secara mendalam. Kesalahan dalam memperkirakan nilai klaim dapat berdampak terhadap perencanaan dana klaim, penentuan premi, serta proses *underwriting*. Oleh karena itu, pendekatan berbasis data seperti metode statistik dan *Machine Learning* diperlukan untuk membantu menghasilkan prediksi yang lebih objektif dan akurat[5].

Perkembangan *Artificial Intelligence (AI)* dan *Machine Learning* memberikan peluang besar dalam meningkatkan kemampuan analisis data pada

industri asuransi. Metode berbasis *AI* dapat membantu proses prediksi dan pengolahan klaim dengan mempelajari pola dari data historis. Namun, penerapan model prediktif tetap perlu mempertimbangkan aspek transparansi dan interpretasi agar hasil yang diperoleh dapat digunakan secara bertanggung jawab dalam proses pengambilan keputusan[6].

Penerapan *Machine Learning* dalam bidang asuransi telah digunakan untuk berbagai kebutuhan seperti prediksi nilai klaim, deteksi fraud, dan peningkatan efisiensi proses operasional. Pendekatan ini memungkinkan perusahaan untuk menemukan pola tersembunyi dari data historis sehingga dapat mendukung proses pengambilan keputusan yang lebih akurat dibandingkan pendekatan manual atau analisis sederhana[7].

Dalam prediksi nilai klaim asuransi, pemilihan metode menjadi aspek penting karena data klaim umumnya memiliki karakteristik kompleks, seperti hubungan antarvariabel yang tidak selalu linear dan distribusi nilai klaim yang bervariasi. Metode *Machine Learning* seperti *Random Forest*, *Gradient Boosting*, dan *XGBoost* mulai banyak digunakan karena mampu menangkap pola data yang kompleks dan menghasilkan performa prediksi yang baik[8].

Meskipun berbagai metode *Machine Learning* berkembang, metode statistik seperti *Linear Regression* masih banyak digunakan sebagai pendekatan dasar dalam permasalahan prediksi. *Linear Regression* memiliki keunggulan dari sisi interpretabilitas karena mampu menjelaskan hubungan antara variabel independen dan variabel dependen, sehingga dapat digunakan untuk mengetahui pengaruh karakteristik polis terhadap nilai klaim asuransi[9].

Selain interpretabilitas, evaluasi performa model prediksi juga menjadi aspek penting dalam menentukan metode terbaik. Beberapa penelitian prediksi klaim asuransi menggunakan metrik evaluasi seperti *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan *R-Squared (R^2)* untuk mengukur tingkat kesalahan dan kemampuan model dalam menjelaskan variasi data. Model berbasis *ensemble* juga terbukti mampu meningkatkan akurasi prediksi pada beberapa kasus asuransi[10].

Prediksi nilai klaim asuransi, khususnya pada asuransi kendaraan, memiliki peran yang sangat krusial dalam industri asuransi karena berkaitan langsung dengan kemampuan perusahaan dalam mengelola risiko, menentukan premi, serta menjaga kecukupan cadangan dana. Ketepatan dalam memprediksi nilai klaim memungkinkan perusahaan untuk mengalokasikan sumber daya secara lebih efektif[9].

Selain itu, prediksi nilai klaim juga berkaitan erat dengan pengelolaan sumber daya perusahaan serta tantangan data klaim yang bersifat tidak seimbang (*low frequency dan high severity*), sehingga diperlukan pendekatan yang lebih *advanced* untuk meningkatkan performa prediksi[9].

Penggunaan *Machine Learning* dalam prediksi biaya dan nilai klaim juga semakin berkembang dengan memanfaatkan berbagai variabel seperti usia, jenis kelamin, kebiasaan merokok, wilayah, serta faktor lainnya yang memengaruhi biaya asuransi. Sebelum proses pemodelan, *exploratory data analysis (EDA)* dilakukan untuk memahami karakteristik data, menemukan pola, serta mengidentifikasi hubungan antar variabel. Selain itu, metode statistik seperti *Linear Regression* tetap digunakan karena interpretabilitasnya yang tinggi, termasuk melalui pengujian *ANOVA* untuk mengidentifikasi variabel yang signifikan. Di sisi lain, model *Machine Learning* berbasis *ensemble* menunjukkan keunggulan dalam meningkatkan akurasi prediksi dengan evaluasi menggunakan metrik seperti *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *R-Squared (R^2)*[10].

Penelitian lainnya menunjukkan bahwa algoritma *Machine Learning* seperti *Gradient Boosting*, *Artificial Neural Network (ANN)*, *XGBoost*, serta *Random Forest* memiliki performa yang berbeda tergantung karakteristik dataset yang digunakan. Model berbasis *ensemble* cenderung unggul dalam menangani hubungan data yang kompleks karena mampu mempelajari pola *non-linear* yang sulit ditangkap oleh pendekatan statistik tradisional[11].

Berdasarkan penelitian sebelumnya, masih terdapat *research gap* yaitu sebagian besar penelitian lebih berfokus pada perbandingan performa algoritma menggunakan dataset umum, sedangkan implementasi prediksi nilai klaim berdasarkan karakteristik polis menggunakan data historis aktual perusahaan asuransi nasional masih terbatas. Selain itu, penelitian sebelumnya belum banyak menggabungkan evaluasi performa model dengan analisis faktor yang memengaruhi nilai klaim serta implementasi model prediksi.

Berdasarkan permasalahan tersebut, penelitian ini melakukan implementasi metode *Linear Regression* dan *Random Forest* untuk memprediksi nilai klaim asuransi berdasarkan karakteristik polis pada PT Jasaraharja Putera. *Linear Regression* digunakan sebagai model dasar karena memiliki interpretabilitas yang baik dalam menjelaskan hubungan antarvariabel, sedangkan *Random Forest* digunakan karena mampu menangani hubungan data yang kompleks dan *non-linear* serta menyediakan analisis *feature importance*.

Kontribusi dari penelitian ini adalah menghasilkan model prediksi nilai klaim berbasis *Machine Learning* menggunakan data historis PT Jasaraharja Putera, mengetahui faktor karakteristik polis yang paling berpengaruh terhadap nilai klaim, serta melakukan *deployment* model menggunakan *Streamlit* sehingga hasil prediksi dapat digunakan dalam bentuk aplikasi berbasis *web* sederhana sebagai pendukung pengambilan keputusan.

Pada penelitian ini, permasalahan yang menjadi fokus adalah bagaimana data historis polis dan klaim pada PT Jasaraharja Putera dapat dimanfaatkan untuk menghasilkan prediksi nilai klaim yang lebih informatif. Data yang

sebelumnya digunakan sebagai dokumentasi historis memiliki potensi untuk dianalisis lebih lanjut agar perusahaan dapat memahami faktor-faktor karakteristik polis yang memiliki pengaruh terhadap nilai klaim. Dengan adanya pendekatan prediktif, hasil analisis dapat digunakan sebagai pendukung dalam proses evaluasi risiko dan pengambilan keputusan berbasis data.

1.2 Rumusan Masalah

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

1. Apakah karakteristik polis berpengaruh signifikan terhadap nilai klaim asuransi pada PT Jasaraharja Putera dan variabel mana yang paling dominan?
2. Bagaimana kinerja model *Linear Regression* dalam memprediksi nilai klaim asuransi?
3. Bagaimana kinerja model *Random Forest* dalam memprediksi nilai klaim asuransi?
4. Model manakah yang memberikan performa terbaik berdasarkan hasil evaluasi dan optimisasi

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan tidak menyimpang dari tujuan utama, maka ditetapkan beberapa batasan masalah sebagai berikut:

1. Data yang digunakan merupakan data historis polis dan klaim pada PT Jasaraharja Putera dalam periode Januari 2022–Desember 2023 sesuai dengan ketersediaan dataset.
2. Variabel yang digunakan terbatas pada karakteristik polis yang tersedia dalam dataset penelitian.
3. Metode yang digunakan dalam penelitian ini adalah *Linear Regression* dan *Random Forest*.
4. Evaluasi model dilakukan menggunakan metrik regresi seperti *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *Coefficient of Determination (R^2)*.
5. Penelitian ini tidak membahas perhitungan aktuarial lanjutan seperti cadangan klaim atau perhitungan premi murni.
6. Implementasi model terbatas pada simulasi prediksi nilai klaim dan tidak langsung digunakan dalam sistem produksi perusahaan.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menganalisis pengaruh karakteristik polis terhadap nilai klaim asuransi pada PT Jasaraharja Putera serta mengidentifikasi variabel yang paling dominan memengaruhi nilai klaim.
2. Menganalisis kinerja model *Linear Regression* dalam memprediksi nilai klaim asuransi berdasarkan karakteristik polis.
3. Menganalisis kinerja model *Random Forest* dalam memprediksi nilai klaim asuransi berdasarkan karakteristik polis.
4. Membandingkan performa model *Linear Regression* dan *Random Forest* untuk menentukan model prediksi yang paling optimal berdasarkan hasil evaluasi model.

1.4.2 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Manfaat Teoritis

Memberikan kontribusi dalam pengembangan penerapan metode regresi statistik dan *Machine Learning* pada industri asuransi, khususnya dalam prediksi nilai klaim asuransi.

2. Manfaat Praktis

Memberikan insight bagi PT Jasaraharja Putera dalam:

- Mengidentifikasi faktor risiko yang memengaruhi nilai klaim
- Meningkatkan akurasi prediksi klaim
- Mendukung pengambilan keputusan berbasis data dalam manajemen risiko

3. Manfaat Akademis

Menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan implementasi klaim asuransi menggunakan pendekatan komparasi model statistik dan *Machine Learning*.

1.5 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun untuk memberikan gambaran umum mengenai isi setiap bab yang terdapat dalam skripsi. Adapun sistematika penulisan adalah sebagai berikut:

- Bab I: Pendahuluan

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan

- Bab II: Landasan Teori

Berisi teori-teori yang mendukung penelitian, meliputi konsep asuransi dan klaim, karakteristik polis, regresi linier, *Random Forest*, serta penelitian terdahulu yang relevan.

- Bab III: Metodologi Penelitian

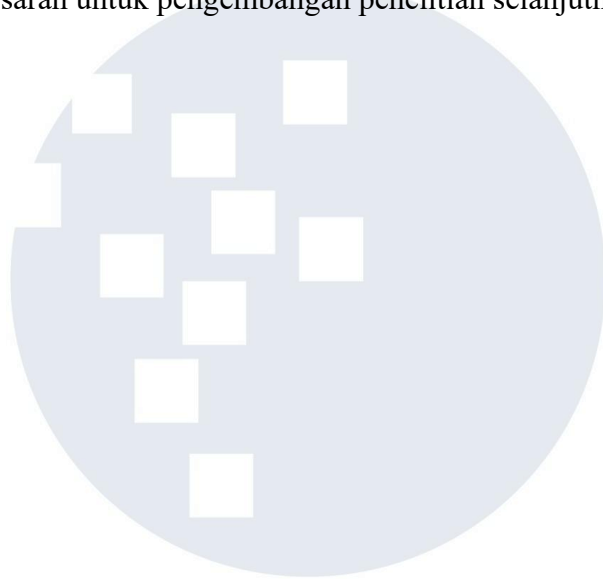
Berisi gambaran umum objek penelitian, deskripsi dataset, tahapan pengolahan data, metode implementasi, serta proses pembangunan dan optimisasi model.

- **Bab IV: Implementasi dan Hasil Penelitian**

Berisi hasil implementasi data, evaluasi model, komparasi performa *Linear Regression* dan *Random Forest*, serta interpretasi hasil penelitian.

- **Bab V: Simpulan dan Saran**

Berisi kesimpulan dari penelitian yang telah dilakukan serta saran untuk pengembangan penelitian selanjutnya.



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA