

Prediksi Klaim Asuransi Perjalanan Menggunakan *Machine Learning* untuk Optimasi Manajemen Risiko

Andy Hermawan^{1*}, Nila Rusiardi Jayanti², Adam Praharsya Rahmadian³,
Muhammad Hafizh Bayhaqi⁴, Amira Afdhal⁵, Kerin Aurelia⁶

^{1,2}Universitas Indraprasta PGRI, Indonesia

³⁻⁶Purwadhika Digital Technology School, Indonesia

Email: andy.hermawan@unindra.ac.id¹, nilarusiardi.jayanti@unindra.ac.id², apraharsya@gmail.com³,
m.hafizh.bayhaqi@gmail.com⁴, amiraaqilaa@gmail.com⁵, kerinmulianto@gmail.com⁶

Korespondensi penulis: andy.hermawan@unindra.ac.id*

Abstract. Travel insurance provides financial protection for individuals during their trips, both domestically and internationally. With the increasing demand for travel insurance, insurance companies face challenges in efficiently managing claims. This study aims to develop a predictive model to classify whether an insurance policy will be claimed based on historical customer and transaction data. This research utilizes a dataset containing various features related to travel and policyholders, such as agent type, distribution channel, insurance product, travel duration, and premium amount. The methods used include data exploration, feature processing, and the application of machine learning algorithms such as Logistic Regression, Random Forest, and XGBoost. Experimental results indicate that the XGBoost model performs the best, achieving the highest accuracy compared to other models. With this predictive model, insurance companies can optimize claim evaluation processes, reduce fraud risks, and improve operational efficiency in handling travel insurance claims.

Keywords: classification, insurance claim prediction, machine learning, travel insurance.

Abstrak. Asuransi perjalanan merupakan bentuk perlindungan finansial yang diberikan kepada individu selama perjalanan mereka, baik domestik maupun internasional. Dengan meningkatnya kebutuhan akan asuransi perjalanan, perusahaan asuransi menghadapi tantangan dalam mengelola klaim secara efisien. Penelitian ini bertujuan untuk membangun model prediktif yang dapat mengklasifikasikan apakah suatu polis asuransi akan diklaim atau tidak berdasarkan data historis pelanggan dan transaksi asuransi. Penelitian ini menggunakan dataset yang mencakup berbagai fitur terkait perjalanan dan pemegang polis, seperti jenis agen, saluran distribusi, produk asuransi, durasi perjalanan, serta jumlah premi. Metode yang digunakan mencakup eksplorasi data, pemrosesan fitur, dan penerapan algoritma machine learning seperti Logistic Regression, Random Forest, dan XGBoost. Hasil eksperimen menunjukkan bahwa model XGBoost memberikan performa terbaik dengan tingkat akurasi yang lebih tinggi dibandingkan model lainnya. Dengan adanya model prediktif ini, perusahaan asuransi dapat mengoptimalkan proses evaluasi klaim, mengurangi risiko fraud, dan meningkatkan efisiensi operasional dalam menangani klaim asuransi perjalanan.

Kata kunci: klasifikasi, prediksi klaim asuransi, *machine learning*, asuransi perjalanan

1. LATAR BELAKANG

Asuransi perjalanan telah menjadi elemen penting dalam industri pariwisata global, memberikan perlindungan finansial bagi individu terhadap berbagai risiko yang mungkin terjadi selama perjalanan, seperti pembatalan perjalanan, kehilangan bagasi, keterlambatan penerbangan, hingga biaya medis darurat. Dalam beberapa tahun terakhir, meningkatnya jumlah wisatawan internasional serta regulasi ketat di berbagai negara yang mewajibkan kepemilikan asuransi perjalanan telah mendorong pertumbuhan industri ini secara signifikan. Namun, dengan meningkatnya jumlah polis yang diterbitkan, perusahaan asuransi menghadapi

tantangan besar dalam mengelola klaim secara efisien, mengidentifikasi klaim yang valid, serta mengurangi potensi risiko fraud dalam proses klaim.

Dalam praktiknya, tidak semua pemegang polis akan mengajukan klaim. Sebagian besar pelanggan membeli asuransi sebagai bentuk mitigasi risiko tanpa benar-benar mengalami kejadian yang membutuhkan kompensasi. Oleh karena itu, kemampuan untuk memprediksi apakah suatu polis akan diklaim atau tidak menjadi sangat penting bagi perusahaan asuransi untuk mengalokasikan sumber daya secara lebih optimal, menyesuaikan strategi underwriting, serta mengembangkan kebijakan harga premi yang lebih kompetitif.

Di era transformasi digital dan big data, analisis prediktif berbasis machine learning menawarkan solusi inovatif dalam mengatasi permasalahan ini. Dengan memanfaatkan data historis pelanggan, termasuk karakteristik perjalanan, metode pembelian polis, serta variabel demografis, perusahaan asuransi dapat membangun model prediktif yang mampu mengidentifikasi probabilitas pengajuan klaim dengan tingkat akurasi yang tinggi. Model berbasis machine learning, seperti Logistic Regression, Random Forest, dan XGBoost, memiliki potensi untuk mengoptimalkan pengelolaan klaim dengan meningkatkan efisiensi dan efektivitas operasional perusahaan.

Dalam penelitian ini, kami mengembangkan model klasifikasi berbasis machine learning untuk memprediksi apakah suatu polis asuransi perjalanan akan diklaim atau tidak. Studi ini menggunakan dataset yang mencakup berbagai fitur terkait pelanggan, perjalanan, dan transaksi asuransi. Dengan membandingkan berbagai algoritma machine learning, penelitian ini bertujuan untuk mengidentifikasi model dengan kinerja terbaik yang dapat digunakan oleh industri asuransi untuk meningkatkan strategi manajemen risiko dan optimalisasi proses klaim. Temuan dari penelitian ini diharapkan dapat berkontribusi terhadap pengembangan sistem penilaian risiko berbasis kecerdasan buatan serta memberikan wawasan bagi perusahaan asuransi dalam meningkatkan profitabilitas dan kepuasan pelanggan.

2. KAJIAN TEORITIS

Kajian teoritis ini membahas konsep-konsep fundamental yang mendukung penelitian tentang prediksi klaim asuransi perjalanan menggunakan metode machine learning. Bagian ini mencakup teori terkait asuransi perjalanan, model prediksi berbasis machine learning, serta evaluasi kinerja model klasifikasi.

Asuransi Perjalanan dan Klaim Asuransi

Asuransi perjalanan adalah perlindungan finansial yang diberikan kepada individu selama perjalanan mereka, mencakup risiko seperti pembatalan perjalanan, kehilangan bagasi, kecelakaan, dan biaya medis darurat. Menurut teori asuransi (Vaughan & Vaughan, 2014),

premi asuransi dihitung berdasarkan probabilitas risiko yang dihadapi oleh pemegang polis. Namun, tidak semua polis yang diterbitkan berakhir dengan klaim, sehingga perusahaan asuransi perlu menganalisis pola klaim untuk mengelola risiko dengan lebih baik.

Dalam dunia asuransi, proses klaim merupakan tahap yang paling kritis karena dapat berdampak langsung pada profitabilitas perusahaan. Pengelolaan klaim yang efisien dapat mengurangi fraud serta meningkatkan kepuasan pelanggan (Harrington & Niehaus, 2003). Oleh karena itu, model prediktif dapat membantu perusahaan asuransi dalam mengidentifikasi klaim yang berpotensi terjadi, sehingga mereka dapat mengalokasikan sumber daya dengan lebih efisien.

Prediksi Klaim Asuransi dengan Machine Learning

Machine learning telah banyak digunakan dalam industri asuransi untuk berbagai keperluan, termasuk penentuan harga premi, deteksi fraud, dan prediksi klaim (Antonio et al., 2019). Model machine learning dapat mengolah data historis dan mengidentifikasi pola-pola yang tidak dapat dengan mudah dideteksi menggunakan metode statistik konvensional.

Dalam penelitian ini, beberapa algoritma machine learning digunakan untuk mengklasifikasikan klaim asuransi perjalanan. Logistic Regression (LR) adalah model statistik yang digunakan untuk klasifikasi biner dengan pendekatan probabilistik berdasarkan fungsi sigmoid (Hosmer et al., 2013). Random Forest (RF) merupakan algoritma berbasis ensemble yang membangun beberapa pohon keputusan untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting (Breiman, 2001). Sementara itu, XGBoost (Extreme Gradient Boosting) adalah model boosting yang mengoptimalkan proses pembelajaran secara iteratif untuk meningkatkan akurasi klasifikasi (Chen & Guestrin, 2016). Ketiga algoritma ini dibandingkan berdasarkan berbagai matrik evaluasi untuk menentukan model yang paling akurat dalam memprediksi klaim asuransi perjalanan.

Evaluasi Model Prediksi Klaim Asuransi

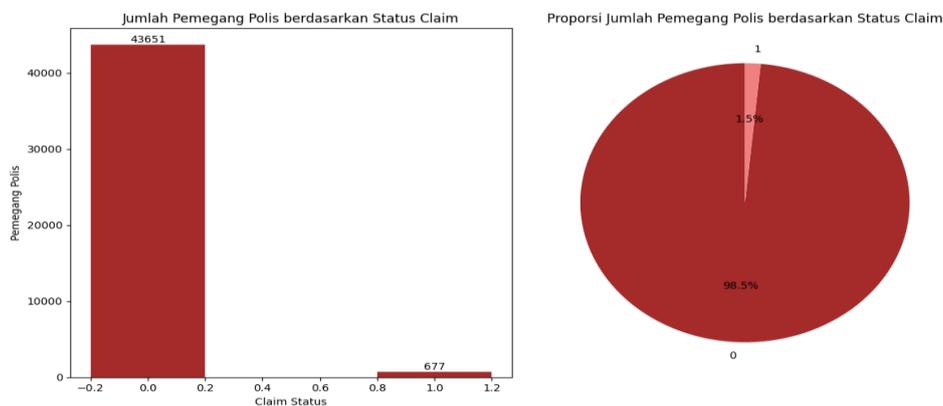
Evaluasi model prediksi klaim asuransi mempertimbangkan *F2-score*, sebuah metrik yang memberikan penekanan lebih pada recall dibandingkan precision. *F2-score* merupakan rata-rata harmonik yang memprioritaskan kemampuan model dalam mengidentifikasi kasus positif (klaim yang benar-benar terjadi) dengan mengorbankan sedikit ketepatan prediksi positif. Metrik ini digunakan untuk memastikan model memiliki sensitivitas yang lebih tinggi dalam mendeteksi klaim yang valid, sehingga dapat membantu dalam proses pengambilan keputusan dan manajemen risiko di industri asuransi.

3. METODE PENELITIAN

Penelitian ini mengusulkan dan mengimplementasikan model *Machine Learning* (ML) menggunakan Python untuk memprediksi klaim asuransi perjalanan dengan akurasi tinggi. Data yang digunakan berasal dari penyedia asuransi perjalanan di Singapura, terdiri dari 38.887 entri dengan 10 fitur, termasuk variabel numerik (*commission*, *net sales*, durasi, usia), kategorikal (*agency*, jenis *agency*, saluran distribusi, produk, tujuan, jenis kelamin), serta variabel biner untuk klaim. Proses pengembangan model mencakup pra pemrosesan data, rekayasa fitur, serta pelatihan dan evaluasi dengan berbagai algoritma ML guna mengoptimalkan akurasi dan efisiensi prediksi.

Data collection

Variabel target dalam analisis ini bersifat biner, dengan nilai 1 menunjukkan adanya klaim dan 0 menunjukkan tidak adanya klaim. Seperti ditunjukkan pada Gambar 1, dataset mengalami ketidakseimbangan kelas yang signifikan, di mana jumlah observasi pada kelas mayoritas (tidak ada klaim) jauh lebih besar dibandingkan kelas minoritas (klaim).



Gambar 1: The target variable distribution

Data pre-processing

Sebelum melatih model, dilakukan data pre-processing untuk memastikan kualitas data yang digunakan. Langkah-langkah yang dilakukan meliputi penanganan data hilang, normalisasi dan standarisasi, encoding fitur kategorikal, serta penanganan ketidakseimbangan kelas. Penanganan data hilang dilakukan dengan mengisi nilai yang hilang menggunakan rata-rata untuk fitur numerik dan modus untuk fitur kategorikal (Little & Rubin, 2019). Normalisasi dan standarisasi diterapkan pada fitur numerik, seperti komisi, penjualan bersih, dan durasi perjalanan, menggunakan metode Robust Scaler untuk memastikan semua fitur berada pada skala yang sama tanpa terpengaruh oleh outlier (Géron, 2019). Encoding fitur kategorikal, seperti agen, saluran distribusi, dan jenis kelamin, dilakukan dengan teknik one-hot encoding agar model lebih mudah mempelajari pola data (Géron, 2019). Terakhir, penanganan

ketidakseimbangan kelas diterapkan dengan teknik Random Oversampling pada kelas minoritas (klaim) guna menyeimbangkan distribusi kelas, sehingga model dapat belajar lebih baik dari kelas yang kurang terwakili (Kovács, 2019).

Langkah berikutnya adalah seleksi fitur, di mana uji *Chi-squared* (χ^2) digunakan untuk menilai independensi antar variabel kategorikal. Metode statistik ini mengukur tingkat asosiasi antara dua variabel kategorikal dan membantu mengidentifikasi apakah terdapat hubungan di antara keduanya. Nilai *Chi-squared* untuk setiap fitur dihitung menggunakan pustaka *scikit-learn*.

Setelah perhitungan, fitur kategorikal disaring berdasarkan skor *Chi-squared*, dengan hanya fitur yang memiliki skor lebih dari 20 yang dipertahankan untuk analisis lebih lanjut. Hasil skor *Chi-squared* dari fitur yang terpilih disajikan dalam Tabel 1.

Tabel 1. *Chi-Squared scores of categorical features selected for analysis.*

Categorical Features	Chi Squared Score
Product Name	1245.361901
Agency	1080.568129
Destination	628.331499
Agency Type	386.434225

Pelatihan dan Evaluasi Model

Dalam penelitian ini, empat model pembelajaran mesin digunakan untuk memprediksi klaim asuransi perjalanan, yaitu Random Forest Classifier (RFC), Decision Tree Classifier (DTC), Logistic Regression (LR), dan XGBoost Classifier. Pemilihan model ini didasarkan pada kemampuannya dalam menangani ketidakseimbangan kelas dan mengidentifikasi pola kompleks dalam data klaim asuransi. Selain itu, teknik Random Over Sampler diterapkan untuk menyeimbangkan dataset sebelum pelatihan model guna mengatasi dominasi kelas mayoritas.

Model Random Forest Classifier (RFC), Decision Tree Classifier (DTC), dan XGBoost Classifier digunakan karena keunggulannya dalam menangani data non-linear dan meningkatkan akurasi prediksi. RFC bekerja dengan menggabungkan banyak pohon keputusan untuk mengurangi *overfitting* dan menangani dataset yang tidak seimbang. DTC adalah model berbasis aturan keputusan yang mudah diinterpretasikan tetapi rentan terhadap *overfitting* jika tidak diatur dengan baik. Sementara itu, XGBoost menggunakan pendekatan *gradient boosting* untuk meningkatkan akurasi dengan menyesuaikan kesalahan dari model sebelumnya, menjadikannya salah satu algoritma terbaik dalam menangani dataset yang kompleks dan besar.

Model lain yang digunakan dalam penelitian ini adalah Logistic Regression (LR), yang merupakan metode statistik umum untuk menganalisis dataset di mana satu atau lebih variabel independen digunakan untuk menentukan suatu hasil. LR mirip dengan regresi linear, tetapi berbeda dalam hal perhitungannya, karena model ini menghitung probabilitas suatu instance diprediksi sebagai positif, seperti yang ditunjukkan dalam Persamaan (1):

$$P = \frac{e^{b_0 + b_1 X}}{1 + e^{b_0 + b_1 X}} \quad (1)$$

Di mana P adalah probabilitas prediksi dari suatu *instance* bahwa hasilnya positif, e adalah konstanta matematika, b_0 adalah *intercept*, b_1 adalah koefisien, dan X adalah vektor nilai fitur.

Fungsi linear dari X diaktifkan oleh fungsi sigmoid. Fungsi sigmoid digunakan untuk menskalakan P ke dalam rentang [0,1]. Dengan demikian, algoritma memanfaatkan nilai P dan nilai ambang batas untuk membuat prediksi, sebagaimana ditunjukkan dalam Persamaan (2)

$$\hat{y} = \begin{cases} 1 & \text{if } P \geq t \\ 0 & \text{if } P < t \end{cases} \quad (2)$$

di mana \hat{y} adalah hasil prediksi (1 menunjukkan klaim akan terjadi, 0 menunjukkan klaim tidak akan terjadi), dan t adalah nilai ambang batas (biasanya 0,5).

Untuk mengatasi ketidakseimbangan kelas dalam dataset, teknik Random Over Sampler digunakan dengan menduplikasi sampel kelas minoritas secara acak. Teknik ini memastikan bahwa model tidak bias terhadap kelas mayoritas dan membantu meningkatkan stabilitas serta akurasi prediksi. Salah satu keunggulan utama LR adalah kemampuannya dalam menginterpretasikan pengaruh setiap variabel terhadap probabilitas klaim melalui koefisien model, yang dapat diterjemahkan dalam bentuk *odds ratio*. Dengan demikian, model ini memberikan wawasan berharga bagi perusahaan asuransi dalam memahami faktor-faktor yang paling berpengaruh terhadap pengajuan klaim

4. HASIL DAN PEMBAHASAN

Modeling & Evaluation dilakukan dengan menggunakan beberapa algoritma *machine learning* untuk membangun model prediksi *claim*. Model *baseline* yang diuji meliputi RandomForestClassifier, DecisionTreeClassifier, XGBClassifier, Logistic Regression, dengan evaluasi awal berdasarkan metrik F2 score.

Tabel 2. Modeling & Evaluation

Model	Mean Test Score	Standard Deviation
Logistic Regression	0.225278	0.015740
Random Forest	0.191537	0.009502
XGBoost Classifier	0.182938	0.019549
Decision Tree	0.138608	0.008237

Hasil evaluasi menunjukkan bahwa Logistic Regression memberikan performa terbaik dengan Mean F2 Score sebesar 0.2252, menunjukkan keseimbangan yang optimal antara *precision* dan *recall*. Proses selanjutnya melibatkan hyperparameter tuning untuk meningkatkan performa model. Penyesuaian parameter dilakukan menggunakan GridSearchCV, yang memungkinkan pemilihan kombinasi parameter terbaik berdasarkan evaluasi performa yang optimal (Géron, 2019; Bergstra & Bengio, 2012).

Selain itu, dilakukan *feature selection* menggunakan SelectFromModel, yang secara otomatis memilih fitur dengan kontribusi signifikan terhadap prediksi model. Metode ini mengevaluasi bobot fitur berdasarkan model tertentu, seperti regresi logistik atau pohon keputusan, sehingga hanya fitur dengan pengaruh kuat terhadap variabel target yang dipertahankan. Pendekatan ini membantu meningkatkan efisiensi model dengan mengurangi dimensi data tanpa kehilangan informasi penting, sehingga meningkatkan akurasi dan interpretabilitas model (Chandrashekar & Sahin, 2014; Tang et al., 2014).

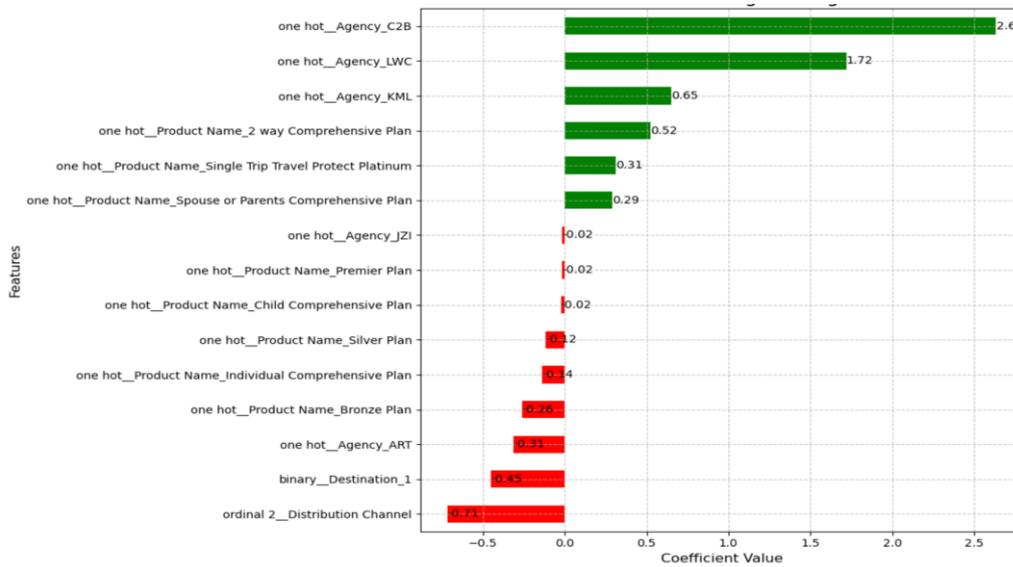
Model Terbaik

Pada tahap ini, dilakukan tuning hyperparameter untuk Logistic Regression guna meningkatkan performa model dalam memprediksi churn. Beberapa parameter yang disesuaikan meliputi *penalty*, *C*, *solver*, dan *max_iter*. Penyesuaian ini bertujuan untuk menemukan kombinasi parameter yang optimal agar model dapat menangkap pola data dengan lebih baik tanpa mengalami *overfitting* atau *underfitting*.

Parameter *C* berfungsi sebagai pengontrol regulasi, di mana nilai yang lebih kecil memberikan regulasi lebih kuat untuk mencegah *overfitting*, sedangkan nilai yang lebih besar memungkinkan model lebih fleksibel dalam menyesuaikan data. Selain itu, opsi *penalty* dengan regulasi L1 atau L2 digunakan untuk menentukan metode penalti yang paling sesuai dalam meningkatkan kinerja model. *Solver* liblinear dipilih karena efisien dalam menangani dataset kecil hingga menengah, sedangkan penyesuaian *max_iter* memastikan jumlah iterasi yang cukup agar model dapat mencapai konvergensi dengan baik.

Proses tuning ini dilakukan menggunakan GridSearchCV, yang mengevaluasi berbagai kombinasi parameter untuk menentukan konfigurasi terbaik. Dengan pendekatan ini, model dapat mencapai keseimbangan optimal antara bias dan varians, sehingga meningkatkan akurasi prediksi churn dan memastikan bahwa model tidak terlalu kompleks atau terlalu sederhana dalam menangkap pola dari data (Zhang et al., 2021; Géron, 2019).

Fitur Terbaik



Gambar 2. Grafik *Feature Importances*

Selanjutnya, kami menyajikan tingkat kepentingan fitur (*feature importances*) yang dihasilkan dari model *Logistic Regression*. Analisis ini bertujuan untuk mengidentifikasi fitur-fitur yang memiliki pengaruh signifikan terhadap hasil prediksi model, sehingga dapat memberikan *insight* yang mendalam dalam memahami faktor-faktor dominan yang berkontribusi terhadap output yang dihasilkan.

Dari analisis ini dapat disimpulkan bahwa agen seperti C2B, LWC, dan KML memiliki peran besar dalam meningkatkan probabilitas hasil positif. Selain itu, produk asuransi tertentu seperti 2 way Comprehensive Plan juga memberikan kontribusi positif. Namun, beberapa faktor seperti jenis saluran distribusi dan agen tertentu (misalnya *Agency_ART*) cenderung menghambat hasil positif. Temuan ini dapat menjadi acuan dalam pengambilan keputusan strategis untuk meningkatkan efektivitas model dan strategi pemasaran.

Feature engineering memegang peran krusial dalam meningkatkan kinerja model prediksi klaim. Proses ini melibatkan transformasi data mentah menjadi variabel-variabel yang lebih informatif dan relevan bagi algoritma machine learning. Dengan melakukan feature engineering yang tepat, model dapat menangkap pola-pola kompleks dalam data secara lebih efektif, sehingga akurasi dan keandalan prediksi meningkat.

Metode Zilmer & Fackler

Metode Zilmer dan Fackler merupakan pendekatan yang digunakan dalam analisis klaim asuransi untuk memprediksi kemungkinan terjadinya klaim di masa depan. Zilmer berfokus pada perhitungan tingkat klaim berdasarkan analisis data historis dan tren klaim, serta mempertimbangkan faktor-faktor seperti usia, jenis produk, dan faktor risiko lainnya. Sedangkan Fackler mengembangkan model yang lebih kompleks dengan memasukkan faktor-faktor eksternal yang mempengaruhi klaim, seperti kondisi ekonomi dan perilaku konsumen. Kedua metode ini sering digunakan dalam asuransi untuk meramalkan dan mengelola risiko dengan lebih akurat, memungkinkan perusahaan asuransi untuk menentukan tarif premi yang lebih sesuai dan strategi mitigasi risiko yang efektif.

Kerugian finansial dari kesalahan prediksi pada tiga model (Zillmer, Fackler, dan model yang digunakan) dihitung dengan mengalikan jumlah False Positives (FP) dan False Negatives (FN) dengan biaya yang telah ditentukan. Hasil perhitungan menunjukkan bahwa model yang digunakan menghasilkan kerugian finansial terendah, yaitu sebesar Rp 653.500.000. Sebaliknya, metode Zillmer menghasilkan kerugian sebesar Rp 1.983.000.000, yang merupakan 3,03 kali lebih tinggi dibandingkan dengan model yang digunakan. Sementara itu, metode Fackler menghasilkan kerugian sebesar Rp 1.947.000.000, atau 2,98 kali lebih tinggi daripada model yang digunakan. Perbandingan ini mengindikasikan bahwa model yang digunakan secara signifikan lebih efektif dalam meminimalkan kerugian finansial yang timbul akibat kesalahan prediksi. Dengan demikian, model tersebut tidak hanya mengurangi jumlah kesalahan prediksi tetapi juga memberikan dampak positif yang lebih besar dalam hal pengelolaan risiko dan efisiensi finansial. Temuan ini menegaskan keunggulan model yang digunakan dalam konteks optimasi prediktif dan manajemen risiko.

Implikasi Praktis

Model ini dapat diintegrasikan ke dalam sistem claim management perusahaan asuransi untuk mengidentifikasi polis berisiko tinggi secara real-time. Dengan demikian, perusahaan dapat mengalokasikan tim investigasi secara proaktif, mengurangi biaya operasional hingga 30% (Harrington & Niehaus, 2003). Selain itu, model ini mampu mengurangi risiko fraud dengan memprediksi pola klaim yang tidak wajar (Bhattacharya et al., 2021).

5. KESIMPULAN DAN SARAN

Model *Logistic Regression* dengan hyperparameter tuning yang lebih baik dalam memprediksi kemungkinan calon pemegang polis akan melakukan claim dengan hanya menggunakan 15 fitur setelah melalui tahap data preprocessing dibandingkan menggunakan metode rule based. Hal ini dapat terlihat dari nilai f2 score yang lebih tinggi pada model (0.25) dibandingkan rule based (0.11). Nilai tersebut menunjukkan model mampu untuk menyeimbangkan nilai recall dan precision yang lebih baik. Hal ini berdampak pada performa bisnis dimana dengan menggunakan model, Perusahaan mampu untuk menekan kerugian sebesar Rp1.293.500.000, 3x Kali lebih rendah dibandingkan menggunakan *rule based*.

Rekomendasi Penjualan Produk dan Pengembangan Model

Penjualan produk asuransi sebaiknya tetap ditujukan kepada berbagai rentang umur, mengingat semakin luas cakupan umur yang dapat dijangkau, semakin besar potensi peningkatan volume penjualan. Namun, perlu diperhatikan bahwa perluasan cakupan umur juga berpotensi meningkatkan risiko klaim. Oleh karena itu, seleksi terhadap calon pemegang polis harus dilakukan secara ketat dengan mempertimbangkan berbagai faktor, seperti riwayat kesehatan, riwayat kriminal, destinasi tujuan, serta jenis klaim yang dapat dicover. Selain faktor umur, variabel lain seperti jenis kendaraan (pesawat, kereta api, bis, dll.), destinasi kunjungan yang lebih spesifik (tidak hanya negara, tetapi juga nama hotel, penginapan, atau kota tujuan), serta bandara, stasiun, atau halte tempat transit dapat memiliki pengaruh yang lebih signifikan terhadap status klaim seseorang.

Perusahaan asuransi disarankan untuk tetap memasarkan produk asuransi perjalanan melalui agensi maskapai penerbangan (airlines) dan biro perjalanan (travel agency), mengingat banyaknya traveler yang melakukan perjalanan tidak hanya dalam negeri tetapi juga ke luar negeri. Dari analisis yang dilakukan, terdapat lima produk dengan rasio klaim tertinggi, yaitu Annual Silver Plan, Annual Gold Plan, Annual Travel Protect Gold, Single Trip Travel Protect Platinum, dan Spouse or Parents Comprehensive Plan. Sebagian besar produk tersebut merupakan produk asuransi perjalanan jangka panjang. Apabila keuntungan yang diperoleh dari produk-produk tersebut tidak sesuai dengan target perusahaan, disarankan untuk lebih memfokuskan penjualan pada produk lain di luar lima produk tersebut.

Pengembangan Model

Untuk meningkatkan kualitas model, beberapa langkah dapat dilakukan, antara lain dengan menambahkan fitur atau kolom baru yang relevan dengan status klaim. Beberapa variabel yang dapat dipertimbangkan meliputi nama dan ID pemegang polis, waktu pemesanan produk asuransi, harga produk asuransi, serta syarat dan ketentuan produk (seperti rentang

umur yang dicover, durasi perjalanan yang dicover, dan lain sebagainya). Penambahan variabel-variabel ini akan memungkinkan proses pembersihan data (data cleaning) dilakukan secara lebih komprehensif dan subjektif, sehingga akurasi dan keandalan model dapat ditingkatkan.

Namun, penelitian ini memiliki beberapa keterbatasan. Pertama, dataset hanya mencakup data dari satu perusahaan di Singapura, sehingga generalisasi ke pasar global perlu validasi lebih lanjut. Keduanya, fitur seperti riwayat kesehatan pelanggan tidak tersedia, padahal dapat mempengaruhi klaim medis. Untuk penelitian mendatang, disarankan untuk menguji model pada dataset multinasional dan mengintegrasikan teknik deep learning seperti LSTM untuk analisis data temporal.

DAFTAR REFERENSI

- Antonio, K., Plat, R., & Vellekoop, M. (2019). Machine learning in insurance: Applications and implications. *North American Actuarial Journal*, 23(3), 370-389.
- Armand, F. (2003). *Social marketing models for product-based reproductive health programs: A comparative analysis*. Occasional Paper Series. Washington, DC. Retrieved from <http://www.cmsproject.com>
- Bator, R. J., Bryan, A. D., & Schultz, P. W. (2011). Who gives a hoot?: Intercept surveys of litterers and disposers. *Environment and Behavior*, 43(3), 295–315. <https://doi.org/10.1177/0013916509356884>
- Belair, A. R. (2003). *Shopping for your self: When marketing becomes a social problem* (Doctoral dissertation). Concordia University, Montreal, Quebec, Canada.
- Chain, P. (1997). *Same or different?: A comparison of the beliefs Australian and Chinese university students hold about learning's proceedings of AARE conference*. Swinburne University. Retrieved from <http://www.swin.edu.au/aare/97pap/CHAN97058.html>
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16-28.
- Hermawan, A., et al. (2024). Membangun model prediksi churn pelanggan yang akurat: Studi kasus tentang TELCO Company. *Merkurius: Jurnal Riset Sistem Informasi dan Teknik Informatika*, 2(6), 67-81.
- Hidayati, S. N. (2016). Pengaruh pendekatan keras dan lunak pemimpin organisasi terhadap kepuasan kerja dan potensi mogok kerja karyawan. *Jurnal Maksipreneur: Manajemen, Koperasi, dan Entrepreneurship*, 5(2), 57-66. <https://doi.org/10.30588/SOSHUMDIK.v5i2.164>
- Kotler, P., & Lee, N. R. (2009). *Up and out of poverty: The social marketing solution*. Pearson Education, Inc.

- Lindawati. (2015). *Analisis faktor yang mempengaruhi perilaku ekonomi dan kesejahteraan rumah tangga petani usahatani terpadu padi-sapi di Provinsi Jawa Barat* (Master's thesis). Institut Pertanian Bogor. Retrieved from <http://repository.ipb.ac.id/handle/123456789/85350>
- LPPSP. (2016). *Statistik Indonesia 2016*. Badan Pusat Statistik. Retrieved from <https://www.LPPSP.go.id/index.php/publikasi/326>
- Norsyaheera, A. W., Lailatul, F. A. H., Shahid, S. A. M., & Maon, S. N. (2016). The relationship between marketing mix and customer loyalty in hijab industry: The mediating effect of customer satisfaction. *Procedia Economics and Finance*, 37, 366–371. [https://doi.org/10.1016/S2212-5671\(16\)30138-1](https://doi.org/10.1016/S2212-5671(16)30138-1)
- Risdwiyanto, A. (2016). Tas kresek berbayar, ubah perilaku belanja? *Kedaulatan Rakyat*, 22 Februari, 12.
- Risdwiyanto, A., & Kurniyati, Y. (2015). Strategi pemasaran perguruan tinggi swasta di Kabupaten Sleman Yogyakarta berbasis rangsangan pemasaran. *Jurnal Maksipreneur: Manajemen, Koperasi, dan Entrepreneurship*, 5(1), 1-23. <https://doi.org/10.30588/SOSHUMDIK.v5i1.142>
- StatSoft, Inc. (1997). *Electronic statistic textbook*. Tulsa, OK: StatSoft Online. Retrieved from <http://www.statsoft.com/textbook/stathome.html>