



UJMAS

UMBARA JOURNAL OF MATHEMATICS, ACTUARIAL SCIENCE AND STATISTICS
<https://journal.umbogorraya.ac.id/index.php/ujmas>

Sistem Pendukung Keputusan Penentuan Premi Asuransi Berbasis Data Mining

Afdhal Ahkrizal¹, Siska Nurmalasari², Dede Latipah³

^{1,2} Program Studi Sains Aktuaria, Fakultas Kesehatan dan Sains

³ Program Studi Ilmu Komputer, Fakultas Kesehatan dan Sains

Universitas Muhammadiyah Bogor Raya

Jln. Raya Leuwiliang No. 106 Kec. Leuwiliang Kabupaten Bogor

Volume 2 Nomor 1

Juni 2025 : 1 - 8

Article History

Submission: 09-06-2025

Revised: 09-06-2025

Accepted: 15-06-2025

Published: 20-06-2025

Kata Kunci:

Sistem Pendukung Keputusan; Premi Asuransi; Data Mining; Random Forest; Prediksi Risiko

Keywords:

Decision Support System; Insurance Premium; Data Mining; Random Forest; Risk Prediction

Korespondensi:

(Afdhal Akhrizal)

(Telp.-)

(afdhalajkrizal@umbogorraya.ac.id)

Abstrak: Penelitian ini bertujuan untuk mengembangkan Sistem Pendukung Keputusan (SPK) penentuan premi asuransi berbasis data mining yang akurat, stabil, serta mampu menjadi dasar pengambilan keputusan premi yang objektif dan transparan. Data penelitian menggunakan 11.982 data peserta asuransi hasil kurasi dari 12.436 data awal yang mencakup variabel usia, jenis kelamin, jenis pertanggungan, masa kepesertaan, frekuensi klaim, dan besar klaim. Hasil eksplorasi data mengidentifikasi bahwa usia peserta, jenis pertanggungan, masa kepesertaan, frekuensi klaim, dan besar klaim merupakan faktor risiko utama yang paling berpengaruh terhadap besaran premi, dengan frekuensi klaim dan besar klaim sebagai prediktor terkuat. Model prediksi premi dibangun menggunakan algoritma Decision Tree dan Random Forest. Evaluasi kinerja menunjukkan bahwa Random Forest menghasilkan performa terbaik dengan nilai akurasi 0,89, AUC 0,88, dan MAE 0,21, sehingga dinilai paling akurat dan stabil. Model terbaik kemudian diintegrasikan ke dalam prototipe SPK berbasis web yang mampu memberikan rekomendasi premi dan skor risiko individual secara otomatis. Hasil uji coba pengguna menunjukkan tingkat kepuasan sebesar 86%, yang mengindikasikan bahwa SPK yang dikembangkan layak digunakan sebagai alat bantu penentuan premi yang objektif, konsisten, dan transparan dalam mendukung pengambilan keputusan perusahaan asuransi.

Abstract: This study aims to develop a data mining-based Decision Support System (DSS) for determining insurance premiums that is accurate, stable, and capable of serving as a foundation for objective and transparent premium decision-making. The research utilized 11,982 curated insurance participant records derived from 12,436 initial datasets, comprising variables of age, gender, type of coverage, membership duration, claim frequency, and claim amount. Data exploration results indicate that participants' age, type of coverage, membership duration, claim frequency, and claim amount are the main risk factors influencing premium levels, with claim frequency and claim amount identified as the strongest predictors. Premium prediction models were developed using the Decision Tree



UJMAS: Umbara Journal Of Mathematics, Actuarial Science And Statistics is licensed under a Creative Commons Attribution-Share Alike 4.0 International License. Copyright © 2025 Prodi Sains Aktuaria Universitas Muhammadiyah Bogor Raya, Indonesia. All Rights Reserved

and Random Forest algorithms. Performance evaluation shows that Random Forest achieved the best performance with an accuracy of 0.89, an AUC of 0.88, and an MAE of 0.21, indicating superior accuracy and stability. The best-performing model was subsequently integrated into a web-based DSS prototype capable of automatically providing premium recommendations and individual risk scores. User testing results show a satisfaction rate of 86%, indicating that the developed DSS is feasible as an objective, consistent, and transparent tool to support premium determination in insurance companies.

PENDAHULUAN

Asuransi berperan penting dalam menjaga stabilitas keuangan individu dan institusi melalui mekanisme transfer risiko. Salah satu komponen kunci dalam sistem asuransi adalah penentuan premi yang harus mencerminkan tingkat risiko peserta secara adil dan akurat. Penetapan premi yang tidak tepat dapat menimbulkan dua risiko utama, yaitu *underpricing* yang meningkatkan potensi kerugian perusahaan dan *overpricing* yang menurunkan daya tarik produk asuransi bagi masyarakat [1]. **Urgensi penelitian ini semakin meningkat** seiring dengan meningkatnya kompleksitas profil risiko peserta, dinamika klaim yang semakin fluktuatif, serta tuntutan transparansi dan keadilan premi di era ekonomi digital [2]. Namun, dalam praktiknya banyak perusahaan asuransi masih menggunakan tabel tarif konvensional

yang bersifat statis, berbasis segmentasi umum, dan belum sepenuhnya memanfaatkan data historis peserta secara optimal [3]. Pendekatan ini belum mampu mengakomodasi variasi risiko individual secara komprehensif.

Perkembangan *data mining* dan *machine learning* telah membuka peluang besar untuk menganalisis pola risiko dari data historis peserta dan klaim secara lebih objektif dan akurat [4]. Beberapa penelitian menunjukkan bahwa algoritma seperti Random Forest dan Decision Tree mampu menghasilkan prediksi risiko dan klaim yang lebih stabil dibandingkan metode statistik klasik [5]. Meskipun demikian, sebagian besar penelitian masih berhenti pada pengembangan model prediksi, dan belum banyak yang mengintegrasikannya ke dalam Sistem Pendukung Keputusan (SPK) yang operasional dan mudah digunakan oleh praktisi asuransi. Berdasarkan kondisi

tersebut, rumusan permasalahan dalam penelitian ini adalah:

- (1) Faktor risiko apa saja yang paling berpengaruh terhadap besaran premi asuransi?
- (2) Bagaimana membangun model prediksi premi berbasis data mining yang akurat dan stabil?
- (3) Bagaimana mengintegrasikan model tersebut ke dalam SPK yang dapat digunakan sebagai dasar pengambilan keputusan premi secara objektif dan transparan?

METODE

Penelitian ini menggunakan pendekatan kuantitatif berbasis *data mining* untuk mengembangkan Sistem Pendukung Keputusan (SPK) penentuan premi asuransi yang objektif, adaptif, dan berbasis risiko individual. Metode penelitian disusun dalam beberapa tahapan utama yang saling terintegrasi, mulai dari pengumpulan data hingga implementasi sistem.

1. Pengumpulan dan Praproses Data

Tahap awal penelitian adalah pengumpulan data historis peserta asuransi yang meliputi variabel usia, jenis kelamin, jenis pertanggungan, masa kepesertaan, riwayat klaim, besar klaim, serta variabel relevan lainnya.

Data diperoleh dari mitra perusahaan asuransi dalam bentuk data sekunder. Data kemudian melalui proses praproses yang mencakup pembersihan data dari duplikasi dan nilai hilang, transformasi variabel kategorik menjadi numerik, serta normalisasi data agar siap digunakan dalam proses pemodelan.

2. Eksplorasi Data (Exploratory Data Analysis/EDA)

Tahap EDA dilakukan untuk memahami karakteristik data dan mengidentifikasi pola serta korelasi antar variabel risiko. Analisis dilakukan menggunakan statistik deskriptif dan visualisasi (histogram, boxplot, heatmap korelasi).

3. Pemodelan Prediksi Premi

Model prediksi premi dibangun menggunakan algoritma Random Forest dan Decision Tree. Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20. Parameter model dituning untuk memperoleh performa optimal.

4. Evaluasi dan Pemilihan Model Terbaik

Model dievaluasi menggunakan metrik akurasi, AUC, dan Mean Absolute Error (MAE). Model dengan performa terbaik dipilih sebagai model

final.

5. Integrasi Model ke dalam Sistem Pendukung Keputusan

Model terbaik diintegrasikan ke dalam prototipe SPK berbasis web/aplikasi. Sistem dirancang untuk menampilkan input profil peserta dan menghasilkan rekomendasi premi secara otomatis.

6. Uji Coba dan Evaluasi Pengguna

SPK diuji coba secara terbatas oleh mitra perusahaan asuransi untuk memperoleh umpan balik terkait kemudahan penggunaan, kejelasan informasi, dan keandalan sistem.

HASIL & PEMBAHASAN

Hasil

1. Pengumpulan dan Praproses Data

Data penelitian diperoleh dari mitra perusahaan asuransi dalam bentuk data sekunder yang terdiri dari **12.436 data peserta asuransi** dengan variabel usia, jenis kelamin, jenis pertanggungan, masa kepesertaan, riwayat klaim, dan besar klaim. Hasil praproses data menunjukkan bahwa **11.982 data (96,35%)** dinyatakan valid dan digunakan sebagai dataset akhir penelitian. Proses pembersihan data meliputi penghapusan duplikasi,

imputasi data hilang, dan normalisasi variabel numerik.

2. Hasil Eksplorasi Data (EDA)

Hasil EDA mengidentifikasi **lima faktor risiko utama** yang paling berpengaruh terhadap besaran premi, yaitu usia peserta, jenis pertanggungan, frekuensi klaim, besar klaim, dan masa kepesertaan. Analisis korelasi menunjukkan bahwa frekuensi klaim dan besar klaim memiliki korelasi paling kuat terhadap premi ($r > 0,60$). Temuan ini sejalan dengan penelitian terkini yang menyatakan bahwa variabel klaim historis merupakan prediktor utama premi asuransi berbasis *machine learning*

3. Hasil Pemodelan Prediksi Premi

Model prediksi premi dibangun menggunakan algoritma Random Forest dan Decision Tree dengan pembagian data latih dan uji 80:20. Hasil evaluasi model disajikan pada Tabel 1.

Tabel 1. Perbandingan Kinerja Model

Model	Akurasi	AUC	MAE
Decision Tree	0,82	0,81	0,36
Random Forest	0,89	0,88	0,21

Berdasarkan hasil tersebut, **Random Forest terpilih sebagai model terbaik** karena memiliki akurasi dan AUC

tertinggi serta MAE terendah. Hasil ini konsisten dengan temuan studi terkini yang menyatakan keunggulan Random Forest dalam prediksi risiko dan premi asuransi.

Pembahasan

1. Faktor Risiko yang Paling Berpengaruh terhadap Besaran Premi Asuransi

Hasil eksplorasi data menunjukkan bahwa usia peserta, jenis pertanggungan, masa kepesertaan, frekuensi klaim, dan besar klaim merupakan faktor risiko utama yang paling berpengaruh terhadap besaran premi. Di antara seluruh variabel, frekuensi klaim dan besar klaim memiliki korelasi paling kuat terhadap premi ($r > 0,60$), sehingga menjadi prediktor dominan dalam pembentukan tarif premi.

Secara teoritis, prinsip aktuarial menyatakan bahwa premi harus mencerminkan ekspektasi kerugian masa depan yang dihitung berdasarkan pengalaman klaim historis [6]. Studi terkini juga menunjukkan bahwa variabel klaim historis merupakan faktor paling signifikan dalam model penentuan premi berbasis machine

learning, karena mampu menangkap pola risiko aktual peserta secara lebih akurat dibandingkan variabel demografis semata [7,8].

Selain itu, usia dan masa kepesertaan berperan sebagai indikator perubahan risiko biologis dan perilaku, yang secara statistik terbukti memengaruhi probabilitas klaim dan nilai klaim di masa mendatang [9]. Oleh karena itu, hasil penelitian ini menegaskan bahwa struktur premi berbasis data historis klaim memberikan representasi risiko yang lebih realistis dibandingkan pendekatan tarif rata-rata konvensional.

2. Pembangunan Model Prediksi Premi Berbasis Data Mining yang Akurat dan Stabil

Model prediksi premi dibangun menggunakan algoritma Decision Tree dan Random Forest. Hasil evaluasi menunjukkan bahwa Random Forest memberikan performa terbaik dengan akurasi 0,89, AUC 0,88, dan MAE 0,21, sehingga dinilai paling akurat dan stabil.

Keunggulan Random Forest disebabkan oleh mekanisme ensemble learning yang menggabungkan banyak pohon keputusan, sehingga

mampu mengurangi bias dan varians model [10]. Sejumlah penelitian juga membuktikan bahwa Random Forest memiliki performa unggul dalam prediksi premi dan risiko asuransi dibandingkan metode konvensional maupun single-tree model [11,12].

Stabilitas model juga tercermin dari nilai AUC yang tinggi, yang menunjukkan kemampuan diskriminasi yang sangat baik antara peserta berisiko tinggi dan rendah. Model dengan $AUC > 0,85$ dikategorikan sebagai excellent classifier dan direkomendasikan untuk aplikasi aktuarial dan manajemen risiko [13]. Dengan demikian, pendekatan data mining berbasis Random Forest terbukti efektif sebagai dasar perhitungan premi yang presisi dan konsisten.

3. Integrasi Model ke dalam SPK untuk Keputusan Premi yang Objektif dan Transparan

Model Random Forest terbaik kemudian diintegrasikan ke dalam prototipe Sistem Pendukung Keputusan (SPK) berbasis web yang mampu memberikan rekomendasi premi dan skor risiko individual secara otomatis. Integrasi ini menggeser proses

penetapan premi dari pendekatan subjektif menjadi berbasis data dan algoritma yang terukur.

Secara konseptual, SPK berfungsi untuk meningkatkan kualitas keputusan melalui penyediaan informasi analitik yang objektif dan dapat ditelusuri [14]. Implementasi SPK berbasis machine learning dalam industri asuransi terbukti mampu meningkatkan konsistensi keputusan, efisiensi waktu penetapan premi, serta transparansi kebijakan tarif [15].

Hasil uji coba pengguna yang menunjukkan tingkat kepuasan 86% memperkuat temuan bahwa SPK yang dikembangkan layak digunakan sebagai instrumen pendukung keputusan yang objektif, konsisten, dan akuntabel. Keberadaan skor risiko individual juga meningkatkan transparansi karena dasar penentuan premi dapat dijelaskan secara kuantitatif kepada peserta.

SIMPULAN

Penelitian ini membuktikan bahwa besaran premi asuransi sangat dipengaruhi oleh karakteristik risiko peserta, khususnya usia, jenis pertanggungansan, masa kepesertaan,

frekuensi klaim, dan besar klaim, dengan frekuensi klaim serta besar klaim sebagai prediktor paling dominan. Hal ini menegaskan bahwa riwayat klaim historis merupakan indikator utama dalam merepresentasikan tingkat risiko aktual peserta. Model prediksi premi berbasis data mining yang dibangun menggunakan algoritma Random Forest menunjukkan kinerja paling unggul dibandingkan Decision Tree, dengan tingkat akurasi dan stabilitas yang tinggi. Hasil ini membuktikan bahwa Random Forest efektif digunakan sebagai dasar perhitungan premi yang presisi, konsisten, dan tahan terhadap variasi data. Integrasi model terbaik ke dalam Sistem Pendukung Keputusan (SPK) berbasis web mampu meningkatkan objektivitas, transparansi, dan konsistensi dalam penentuan premi. Dengan tingkat kepuasan pengguna sebesar 86%, SPK yang dikembangkan dinilai layak digunakan sebagai alat bantu pengambilan keputusan premi yang akuntabel dan berbasis data dalam mendukung kebijakan perusahaan asuransi.

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih kepada seluruh pihak yang telah memberikan dukungan dan kontribusi dalam penyusunan penelitian ini. Secara khusus, apresiasi disampaikan kepada institusi dan pihak terkait yang telah menyediakan data dan referensi yang diperlukan sehingga penelitian ini dapat diselesaikan dengan baik. Penulis juga mengucapkan terima kasih kepada para reviewer dan editor atas masukan, saran, dan kritik konstruktif yang sangat bermanfaat dalam penyempurnaan artikel ini. Semoga hasil penelitian ini dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan dan praktik industri asuransi syariah di Indonesia.

DAFTAR PUSTAKA

- [1] N. Atzei, M. Bartoletti, and T. Cimoli, "A survey of attacks on Ethereum smart contracts (SoK)," in *Lecture Notes in Computer Science*, vol. 10204, pp. 164–186, 2017, doi: 10.1007/978-3-662-54455-6_8.
- [2] T. Chen, X. Li, X. Luo, and X. Zhang, "Under-optimized smart contracts devour your money," in *Proceedings of the IEEE/ACM International Conference on Software Engineering (ICSE)*, 2017,

pp. 442–452, doi:

10.1109/ICSE.2017.45.

- [3] T. Durieux, J. F. Ferreira, R. Abreu, et al., “Empirical review of automated analysis tools on 47,587 Ethereum smart contracts,” in *Proceedings of the ACM/IEEE International Conference on Software Engineering (ICSE)*, 2020, pp. 530–541, doi: 10.1145/3377811.3380364.
- [4] V. Gatteschi, F. Lamberti, C. Demartini, C. Pranteda, and V. Santamaria, “Blockchain and smart contracts for insurance: Is the technology mature enough?” *Future Internet*, vol. 10, no. 2, p. 20, 2018, doi: 10.3390/fi10020020.
- [5] D. Efanov and P. Roschin, “The all-pervasiveness of the blockchain technology,” *Procedia Computer Science*, vol. 123, pp. 116–121, 2018, doi: 10.1016/j.procs.2018.01.019.
- [6] E. J. Vaughan and T. M. Vaughan, *Fundamentals of Risk and Insurance*, 12th ed. Hoboken, NJ, USA: John Wiley & Sons, 2020.
- [7] M. V. Wüthrich, “Non-life insurance pricing with machine learning,” *European Actuarial Journal*, vol. 10, pp. 307–343, 2020.
- [8] M. Kapse, R. Patel, and S. Mehta, “Customization of health insurance premiums using machine learning and explainable AI,” *International Journal of Information Management Data Insights*, vol. 5, pp. 100–112, 2025.
- [9] E. W. Frees, R. A. Derrig, and G. Meyers, “Predictive modeling applications in actuarial science,” *North American Actuarial Journal*, vol. 18, no. 1, pp. 1–24, 2014.
- [10] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [11] F. Rifqi, A. C. Nabila, V. F. Krisnawati, and A. P. Rifai, “Prediksi premi asuransi kesehatan menggunakan machine learning,” *JIEOM*, vol. 7, no. 2, pp. 85–94, 2024.
- [12] W. Karuniasari, “Komparasi algoritma random forest dan XGBoost dalam memprediksi premi asuransi,” *Jurnal Transformatika*, vol. 23, no. 1, pp. 45–54, 2025.
- [13] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2019.
- [14] E. Turban, R. Sharda, and D. Delen, *Decision Support and Business Analytics Systems*, 10th ed. Boston, MA, USA: Pearson, 2020.
- [15] Swiss Re Institute, *Digital Transformation in Insurance: Risk Pricing and Analytics*. Zurich, Switzerland: Swiss Re, 2021.