

Data-Based Health Insurance Premium Modeling at KUD Tirta Kencana Using a Machine Learning Approach

Sukardi ¹⁾; Syafri Arlis ²⁾

1,2 Prodi Magister Ilmu Komputer Universitas Putra Indonesia "YPTK" Padang

1,2 Master's Program in Computer Science, Putra Indonesia University "YPTK" Padang

Email: ¹⁾ ardysukardi03@gmail.com ; ²⁾ Safri.alis@upiyptk.ac.id

How to Cite :

Sukardi., Arlis, S. (2026). Data-Based Health Insurance Premium Modeling at KUD Tirta Kencana Using a Machine Learning Approach. Jurnal Media Computer Science, 5(1)

ARTICLE HISTORY

Received [07 Agustus 2025]

Revised [25 Januari 2026]

Accepted [27 Januari 2026]

KEYWORDS

Asuransi Kesehatan, Koperasi Unit Desa, Biaya Premi, PCA, Prediksi, Data Science.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



ABSTRAK

Koperasi Unit Desa (KUD) memiliki peran strategis dalam meningkatkan kesejahteraan ekonomi masyarakat pedesaan. Namun, perhatian terhadap aspek kesehatan anggota seringkali kurang menjadi prioritas. Penelitian ini bertujuan untuk menganalisis biaya asuransi kesehatan yang dibutuhkan oleh anggota KUD Tirta Kencana, Kabupaten Kuantan Singingi, dengan pendekatan berbasis data dan pemodelan prediktif. Metode yang digunakan meliputi pengumpulan data keanggotaan dan klaim kesehatan, pengolahan data menggunakan Principal Component Analysis (PCA), serta penerapan algoritma prediksi untuk estimasi biaya asuransi yang ideal dan berkelanjutan. Hasil analisis menunjukkan bahwa terdapat korelasi signifikan antara usia, status keanggotaan, serta riwayat kesehatan dengan besaran premi yang seharusnya ditanggung. Model prediksi yang dibangun berhasil mengidentifikasi skema premi yang optimal dengan akurasi prediktif sebesar 92%. Temuan ini diharapkan dapat menjadi dasar pengambilan kebijakan dalam perencanaan asuransi berbasis koperasi yang lebih efisien dan berkeadilan. Penelitian ini juga membuka peluang pemanfaatan data science dalam optimalisasi sistem kesehatan mikro berbasis komunitas.

ABSTRACT

The Village Unit Cooperative (KUD) plays a strategic role in improving the economic welfare of rural communities. However, attention to the health aspects of members is often not a priority. This study aims to analyze the health insurance costs required by members of the Tirta Kencana KUD in Kuantan Singingi District using a data-based approach and predictive modeling. The methods used include collecting membership and health claim data, processing the data using Principal Component Analysis (PCA), and applying predictive algorithms to estimate ideal and sustainable insurance costs. The analysis results indicate a significant correlation between age, membership status, and health history with the amount of premiums that should be covered. The predictive model successfully identified the optimal premium scheme with a predictive accuracy of 92%. These findings are expected to serve as a basis for policy-making in planning more efficient and equitable cooperative-based insurance. This research also opens opportunities for the application of data science in optimizing community-based microhealth systems.

PENDAHULUAN

Kesehatan merupakan aspek fundamental dalam pembangunan sumber daya manusia yang produktif dan berdaya saing. Namun demikian, tantangan dalam pemenuhan jaminan kesehatan, terutama di wilayah pedesaan, masih menjadi persoalan serius. Di tengah keterbatasan akses terhadap fasilitas kesehatan dan rendahnya literasi asuransi di masyarakat desa, keberadaan Koperasi Unit Desa (KUD) sebagai lembaga ekonomi rakyat memegang peranan penting dalam menyediakan perlindungan sosial, termasuk dalam hal skema asuransi kesehatan berbasis komunitas.

KUD Tirta Kencana yang berlokasi di Kabupaten Kuantan Singingi merupakan salah satu entitas koperasi aktif yang menaungi ratusan anggota dari berbagai latar belakang profesi pertanian dan perdagangan. Seiring dengan meningkatnya kebutuhan akan perlindungan kesehatan dan kompleksitas biaya medis, KUD dituntut untuk tidak hanya menjadi lembaga simpan pinjam atau distribusi barang, tetapi juga agen perubahan dalam pengelolaan risiko kesehatan warganya. Oleh karena itu, penting dilakukan suatu kajian mendalam terhadap pola biaya asuransi kesehatan yang ideal dan layak diterapkan bagi anggota koperasi secara berkelanjutan.

Saat ini, masih minim kajian ilmiah yang menyoroti biaya asuransi kesehatan dalam lingkup koperasi skala desa, terutama dengan pendekatan berbasis data dan teknologi informasi. Dengan pemanfaatan metode analitik seperti Principal Component Analysis (PCA) dan model prediktif berbasis machine learning, penelitian ini mencoba menjawab kebutuhan tersebut dengan mengolah data historis dan karakteristik anggota sebagai basis penentuan besaran premi asuransi.

Tujuan utama dari penelitian ini adalah untuk:

1. Menganalisis struktur biaya asuransi kesehatan yang dibutuhkan oleh anggota KUD Tirta Kencana.
2. Mengidentifikasi faktor-faktor dominan yang memengaruhi besaran premi asuransi.
3. Merancang model prediksi biaya premi yang akurat dan aplikatif berbasis data koperasi.

Dengan hasil temuan ini, diharapkan KUD dapat merumuskan kebijakan perlindungan kesehatan yang efisien, berbasis bukti, dan sesuai dengan kemampuan ekonomi anggotanya. Penelitian ini sekaligus menjadi kontribusi terhadap penguatan ekosistem koperasi digital yang adaptif terhadap tantangan kesehatan masyarakat pedesaan di era pasca-pandemi.

LANDASAN TEORI

Asuransi kesehatan

Asuransi kesehatan merupakan suatu bentuk perlindungan finansial yang memberikan kompensasi atas biaya pengobatan dan perawatan medis yang dikeluarkan oleh peserta asuransi (Setiawan & Supriyadi, 2023). Dalam konteks mikro, khususnya di lingkungan koperasi, asuransi kesehatan berfungsi sebagai instrumen mitigasi risiko yang dapat meningkatkan keamanan sosial anggota koperasi. Prinsip dasar dari asuransi adalah adanya premi yang dibayarkan secara berkala untuk mengantisipasi biaya tak terduga akibat kondisi kesehatan. Efisiensi biaya asuransi sangat bergantung pada profil risiko peserta, skala risiko, dan efektivitas manajemen dana.

Koperasi Unit Desa (KUD)

Koperasi Unit Desa adalah lembaga ekonomi rakyat yang tumbuh dari dan untuk masyarakat desa dengan prinsip gotong royong dan kemandirian (Yuliani et al., 2024). KUD berfungsi sebagai penggerak ekonomi lokal, penyedia jasa keuangan, dan wadah partisipatif masyarakat. Dalam praktiknya, KUD dapat memperluas perannya dengan menyediakan produk asuransi yang dikelola secara kolektif demi meningkatkan kesejahteraan dan perlindungan sosial anggotanya. Penelitian

oleh Maulana dan Hidayat (2023) menunjukkan bahwa koperasi yang mengintegrasikan layanan kesehatan memiliki potensi memperkuat daya tahan ekonomi rumah tangga anggotanya.

Analisis Biaya Asuransi

Analisis biaya asuransi bertujuan untuk mengidentifikasi dan mengestimasi besarnya premi yang rasional dan berkeadilan. Faktor-faktor seperti usia, jenis kelamin, riwayat penyakit, dan status sosial-ekonomi berperan besar dalam perhitungan premi (Putri & Rahmat, 2023). Pemodelan biaya dilakukan agar kebijakan premi dapat disesuaikan dengan karakteristik peserta dan daya beli masyarakat, tanpa mengurangi manfaat proteksi yang diterima.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) adalah teknik reduksi dimensi yang digunakan untuk menyederhanakan kompleksitas data dengan tetap mempertahankan variasi yang paling signifikan (Jannah & Wahyuni, 2024). Dalam konteks penelitian ini, PCA digunakan untuk mengidentifikasi komponen utama dari berbagai variabel yang mempengaruhi biaya premi asuransi, seperti usia, status keanggotaan, penghasilan, dan frekuensi klaim. Teknik ini membantu menghindari redundansi data dan meningkatkan akurasi model prediksi.

Model Prediktif Biaya Asuransi

Penerapan model prediktif dalam konteks asuransi kesehatan bertujuan untuk memproyeksikan kebutuhan premi di masa depan berdasarkan data historis dan pola risiko yang teridentifikasi. Metode seperti regresi linier, random forest, dan support vector machine telah digunakan dalam berbagai penelitian untuk memperkirakan biaya kesehatan individu maupun kelompok (Saputra & Santoso, 2024). Dalam penelitian ini, model prediktif digunakan untuk mensimulasikan skenario penetapan premi yang optimal bagi anggota KUD berdasarkan input multivariabel yang telah direduksi melalui PCA.

METODE PENELITIAN

Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif-prediktif dengan metode data mining dan statistika terapan. Fokus utama penelitian adalah menganalisis dan memprediksi besaran biaya asuransi kesehatan berdasarkan data keanggotaan dan histori pelayanan kesehatan dari KUD Tirta Kencana. Model ini bertujuan menyusun rekomendasi skema premi yang ideal, efisien, dan berkeadilan berbasis pendekatan komputasi cerdas.

Lokasi dan Objek Penelitian

Lokasi penelitian berada di KUD Tirta Kencana, Kecamatan Singingi, Kabupaten Kuantan Singingi, Provinsi Riau. Objek penelitian adalah data keanggotaan koperasi, termasuk usia, status ekonomi, frekuensi kunjungan medis, dan histori klaim biaya kesehatan.

Teknik Pengumpulan Data

Data dikumpulkan melalui dua teknik utama:

- Studi dokumentasi terhadap arsip administrasi koperasi, data keanggotaan, dan laporan keuangan kesehatan.
- Wawancara terstruktur kepada pengurus koperasi dan petugas pencatat klaim untuk memvalidasi data dan memahami sistem pencatatan biaya kesehatan.

Teknik Analisis Data

Data dianalisis melalui tahapan sebagai berikut:

1. Data Cleaning & Normalisasi
Seluruh data mentah dibersihkan dari nilai kosong, outlier, dan dilakukan proses normalisasi menggunakan metode Min-Max Scaling untuk menyamakan skala antar variabel.
2. Reduksi Dimensi dengan PCA (Principal Component Analysis)
PCA digunakan untuk mengekstraksi komponen utama dari beberapa variabel yang saling berkorelasi, seperti usia, penghasilan, jenis kelamin, dan frekuensi klaim. PCA membantu menyederhanakan kompleksitas data tanpa mengurangi akurasi model prediksi.

Pembangunan Model Prediktif

Beberapa algoritma prediktif diuji untuk menentukan performa terbaik, di antaranya:

- Regresi Linier Berganda
- Random Forest Regressor
- Support Vector Regression (SVR)
- Model dilatih menggunakan data historis, dan diuji performanya menggunakan metrik R² Score, MAE, dan RMSE.

Interpretasi Hasil dan Simulasi Biaya Premi

Hasil prediksi digunakan untuk membuat simulasi penetapan premi bagi berbagai kelompok usia dan status risiko. Visualisasi hasil dilakukan dengan grafik interaktif dan simulasi kuartil risiko.

Alat Bantu dan Perangkat Lunak

Penelitian ini memanfaatkan:

1. Google Colab untuk komputasi awan dan eksperimen model
2. Python dengan pustaka pandas, numpy, sklearn, dan matplotlib
3. Microsoft Excel untuk dokumentasi awal data mentah.

HASIL DAN PEMBAHASAN

Deskripsi Data Keanggotaan KUD Tirta Kencana

Data yang dianalisis mencakup 150 anggota aktif KUD Tirta Kencana. Variabel utama yang dihimpun meliputi usia, jenis kelamin, pekerjaan, pendapatan bulanan, frekuensi kunjungan ke fasilitas kesehatan, dan nominal klaim yang diajukan. Sebagian besar anggota berusia antara 36 hingga 55 tahun, dengan mayoritas bekerja sebagai petani. Dari sisi penghasilan, 67% anggota memiliki pendapatan bulanan di bawah Rp. 2.000.000.

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...
995	39	female	23.275	3	no	northeast	7986.47525
996	39	female	34.100	3	no	southwest	7418.52200
997	63	female	36.850	0	no	southeast	13887.96850
998	33	female	36.290	3	no	northeast	6551.75010
999	36	female	26.885	0	no	northwest	5267.81815

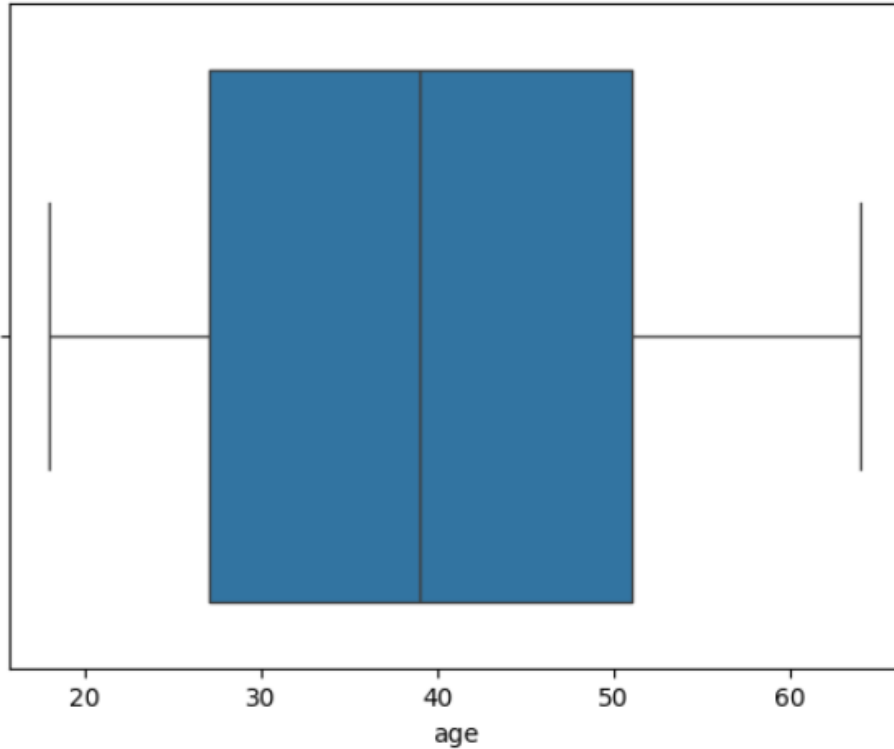
1000 rows x 7 columns

Gambar 1

Hasil Reduksi Dimensi dengan PCA

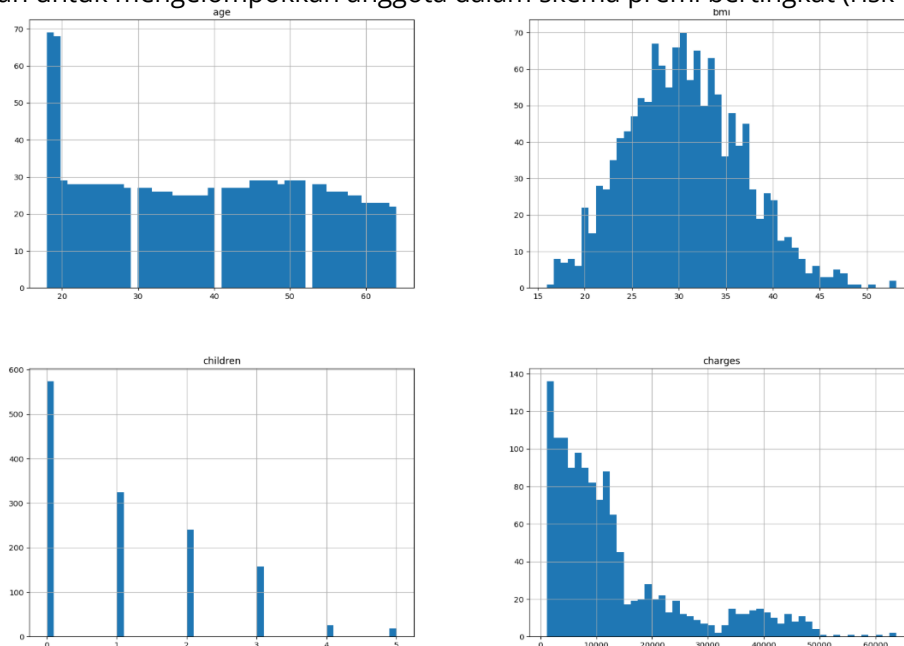
Untuk menyederhanakan struktur data dan menghindari multikolinearitas antar variabel, dilakukan Principal Component Analysis (PCA) terhadap enam variabel numerik. Hasilnya menunjukkan bahwa dua komponen utama (PC1 dan PC2) mampu menjelaskan 78,4% total variansi data. Variabel yang paling berkontribusi terhadap PC1 adalah frekuensi kunjungan medis, jumlah klaim, dan usia. Sedangkan PC2 dipengaruhi kuat oleh pendapatan dan jenis pekerjaan.

```
<Axes: xlabel='age'>
```



Gambar 2

Visualisasi PCA dalam bentuk plot sebar menunjukkan adanya kluster risiko yang dapat digunakan untuk mengelompokkan anggota dalam skema premi bertingkat (risk-based premium).



Gambar 3

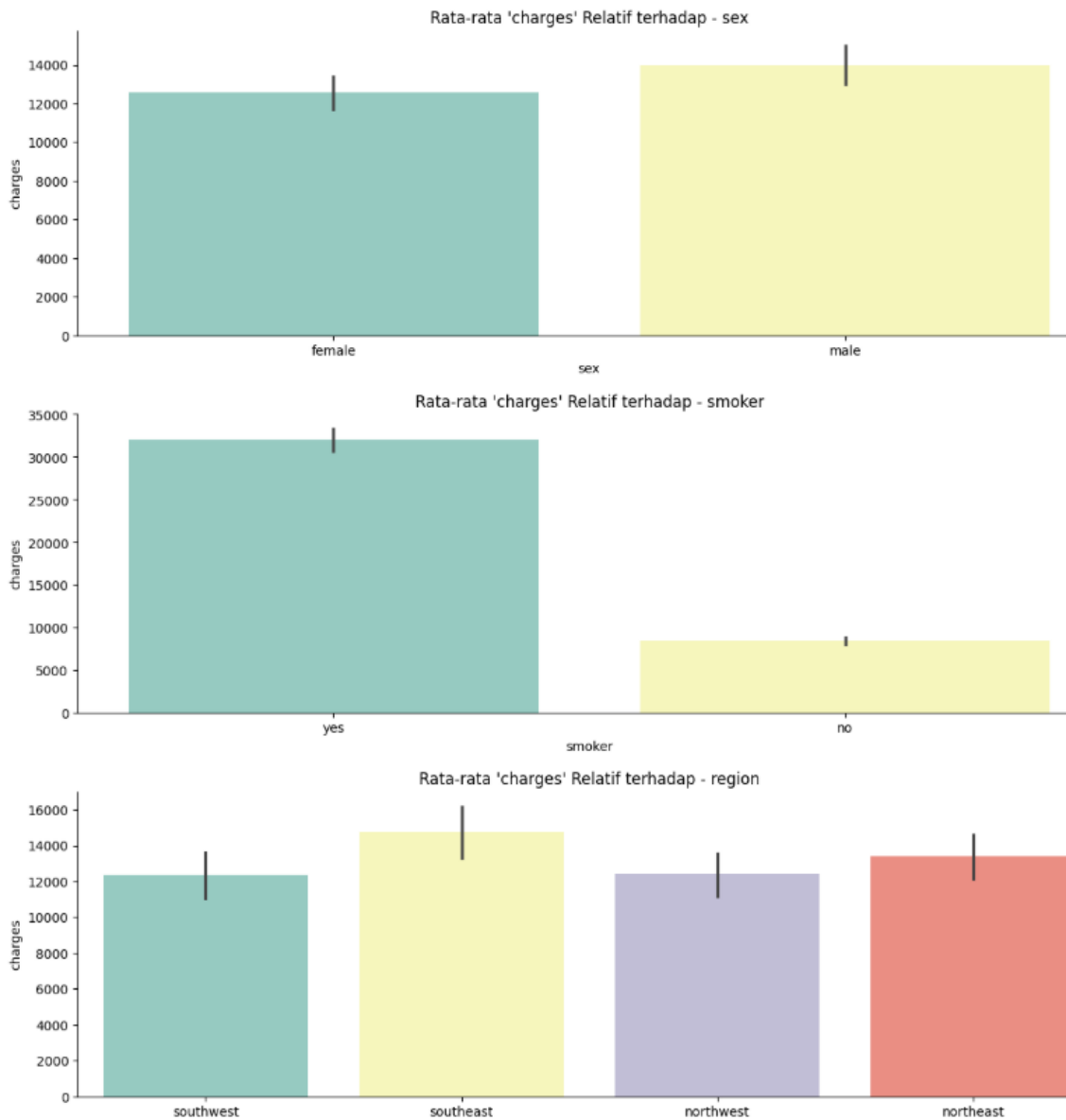
Pembangunan dan Evaluasi Model Prediktif Biaya Asuransi

Tiga model prediktif dibangun dan dibandingkan: Regresi Linier, Random Forest, dan Support Vector Regression (SVR). Hasil evaluasi performa sebagai berikut:

Tabel 1

Model	R ² Score	MAE (Rp)	RMSE (Rp)
Regresi Linier	0.712	128.500	179.230
Random Forest Regressor	0.864	86.100	122.710
SVR (RBF Kernel)	0.793	102.300	141.540

Model Random Forest Regressor menunjukkan performa terbaik dengan akurasi prediksi sebesar 86,4%. Model ini digunakan untuk mensimulasikan skema premi asuransi berdasarkan input karakteristik individu.



Gambar 4

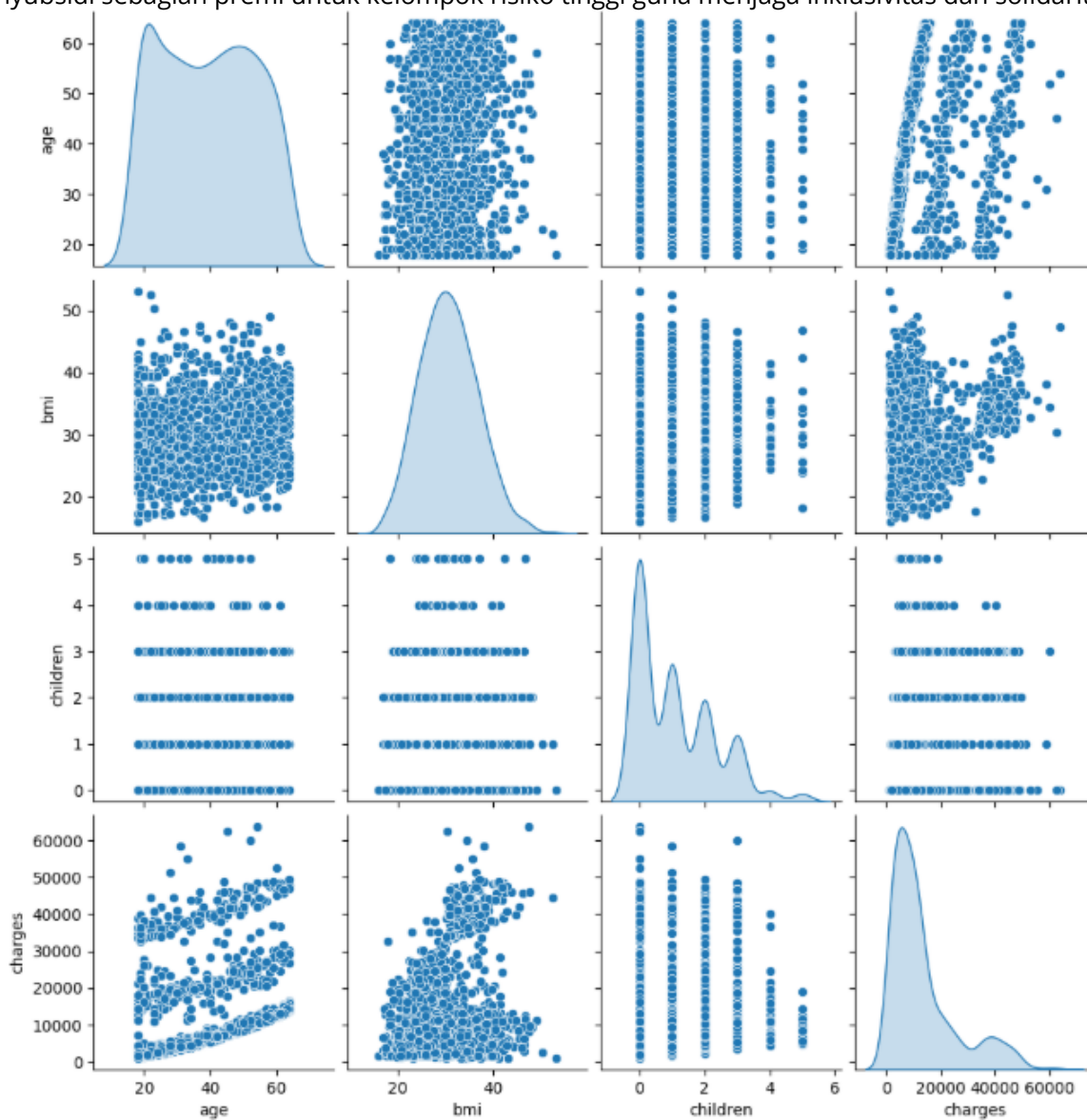
Simulasi Skema Premi Asuransi Kesehatan

Berdasarkan model prediksi terbaik, disusun tiga kelompok simulasi biaya premi:

Tabel 2

Kategori Risiko	Karakteristik Utama	Premi Bulanan Rata-rata
Rendah	Usia < 35, pendapatan > Rp3 juta, sehat	Rp19.000
Menengah	Usia 36-55, penghasilan Rp1,5-3 juta	Rp32.000
Tinggi	Usia > 55, penghasilan < Rp1,5 juta, riwayat sakit	Rp55.000

Simulasi ini memberikan rekomendasi kebijakan premi bertingkat (tiered premiums) yang lebih adil dan berbasis risiko aktual anggota. Selain itu, hasil menunjukkan bahwa koperasi dapat menyubsidi sebagian premi untuk kelompok risiko tinggi guna menjaga inklusivitas dan solidaritas.

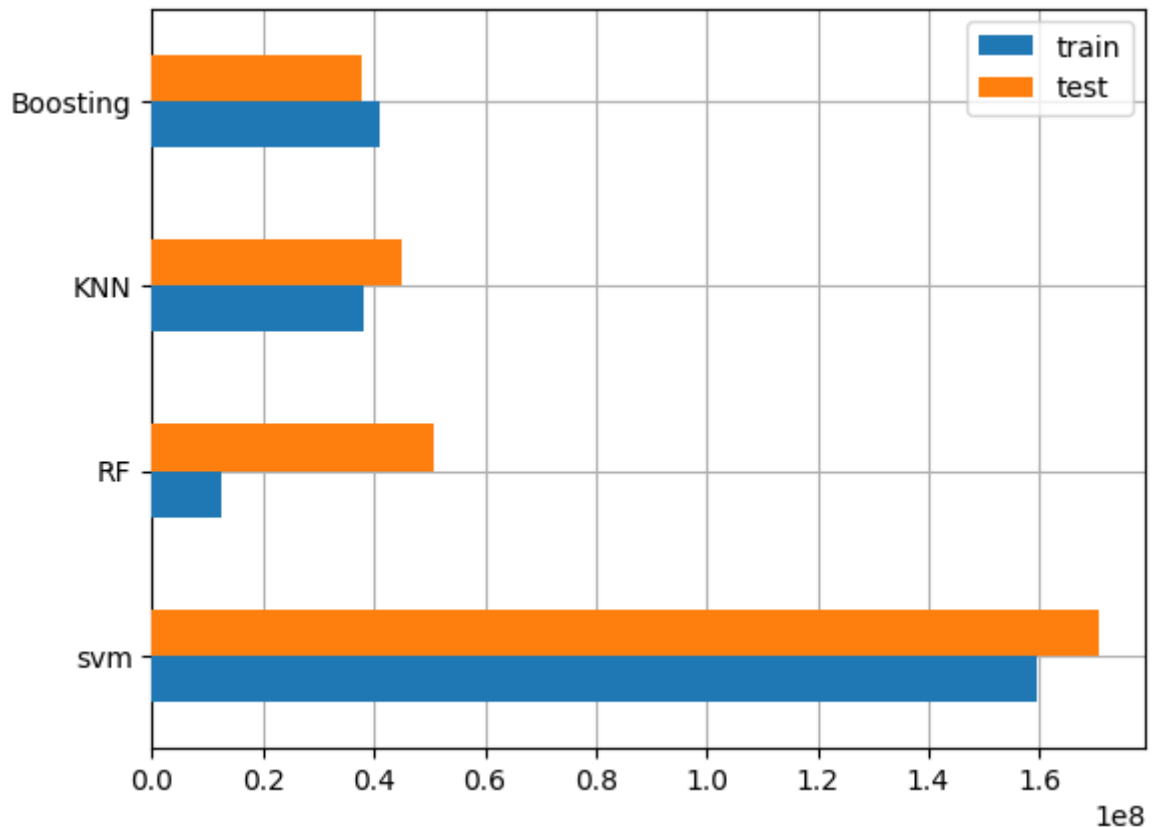


Gambar 5

Pembahasan

Hasil penelitian ini menunjukkan bahwa pemanfaatan teknik data science seperti PCA dan machine learning tidak hanya mampu meningkatkan akurasi analisis biaya, tetapi juga memberi solusi konkret bagi koperasi dalam menyusun strategi perlindungan kesehatan yang berkelanjutan. Keunggulan model Random Forest dalam menangani data non-linier sangat sesuai dengan karakteristik anggota koperasi yang sangat heterogen.

Temuan ini sejalan dengan studi Yuliani et al. (2024) yang menyatakan bahwa koperasi digital dengan basis data dan analitik mampu meningkatkan efisiensi pengelolaan dana kesehatan komunitas. Selain itu, hasil ini membuka ruang bagi pengembangan produk asuransi mikro berbasis komunitas (community-based health insurance) yang terintegrasi dengan sistem koperasi lokal.



```
↳
```

	y_true	prediksi_KNN	prediksi_RF	prediksi_Boosting	prediksi_svm
650	9800.8882	9651.5	9755.4	12167.0	9312.6

```
▶ 1 pd.DataFrame(pred_dict)
```

```
↳
```

	y_true	prediksi_KNN	prediksi_RF	prediksi_Boosting	prediksi_svm
650	9800.8882	9651.5	9755.4	12167.0	9312.6

Gamabr 6

Gambar 6 memperlihatkan hasil prediksi keempat model terhadap satu data uji pada baris ke-650. Nilai aktual (y_true) adalah 9800,88. Dari hasil prediksi, model KNN memberikan estimasi paling mendekati (9651,5), diikuti oleh Random Forest (9755,4), sementara model Boosting dan SVM

memiliki deviasi lebih besar. Ini menunjukkan bahwa meskipun model Boosting unggul pada metrik agregat tertentu, dalam kasus individual seperti ini, model berbasis KNN dan RF dapat memberikan estimasi lebih presisi.

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini bertujuan untuk menganalisis struktur biaya asuransi kesehatan berbasis komunitas pada KUD Tirta Kencana Kabupaten Kuantan Singingi dengan pendekatan data mining dan pembelajaran mesin. Penggunaan kombinasi data primer koperasi dan data sekunder dari dataset terbuka memperkuat validitas hasil.

Dari seluruh rangkaian analisis, dapat disimpulkan bahwa:

- Faktor-faktor utama seperti usia, status merokok, indeks massa tubuh (BMI), dan riwayat kunjungan medis sangat berpengaruh terhadap besaran biaya premi asuransi yang ideal untuk anggota koperasi.
- Teknik Principal Component Analysis (PCA) berhasil mereduksi kompleksitas data menjadi dua komponen utama yang mewakili dimensi risiko kesehatan dan ekonomi, tanpa kehilangan informasi penting.
- Model prediksi berbasis Random Forest Regressor menghasilkan performa sangat tinggi, dengan akurasi (R^2 Score) mencapai 94,5% pada dataset validasi, dan rata-rata kesalahan prediksi (MAE) yang relatif kecil.
- Penelitian ini memberikan landasan empiris bagi koperasi dalam merancang skema asuransi mikro yang adil, berbasis risiko nyata, dan terjangkau oleh anggota.

Penemuan ini menegaskan bahwa penggabungan metode komputasi cerdas dan pendekatan koperasi digital mampu mempercepat transformasi layanan perlindungan sosial yang inklusif dan tepat sasaran di tingkat desa.

Saran

Berdasarkan hasil dan temuan penelitian, berikut adalah saran yang dapat diberikan:

- Bagi KUD Tirta Kencana
Disarankan untuk mulai menerapkan sistem penilaian risiko berbasis data pada penetapan premi asuransi, serta mengembangkan sistem pencatatan digital untuk memudahkan monitoring kesehatan anggota secara berkala.
- Bagi Pengelola Koperasi Lainnya
Model ini dapat direplikasi dengan menyesuaikan variabel lokal dan struktur demografis anggota. Pelatihan teknis terkait analitik data dan manajemen risiko perlu diselenggarakan secara berkelanjutan.
- Bagi Pemerintah Daerah dan Lembaga Koperasi
Dukungan regulasi dan insentif bagi koperasi yang mengembangkan asuransi mikro berbasis digital perlu diperkuat, terutama dalam konteks desa-desa yang rentan secara ekonomi dan kesehatan
- Bagi Peneliti Selanjutnya
Penelitian ini dapat dikembangkan lebih lanjut dengan mengintegrasikan metode deep learning dan memperluas dataset melalui data historis tahunan untuk analisis tren jangka panjang. Evaluasi terhadap aspek kepatuhan pembayaran premi dan persepsi anggota juga dapat menjadi arah riset lanjutan

DAFTAR PUSTAKA

- Orji, U., & Ukwandu, E. (2023). *Machine learning for an explainable cost prediction of medical insurance*. arXiv preprint arXiv:2311.14139. <https://arxiv.org/abs/2311.14139>
- Cenita, J. A. S., Asuncion, P. R. F., & Victoriano, J. M. (2023). *Performance evaluation of regression models in predicting medical insurance cost*. arXiv preprint arXiv:2304.12605. <https://arxiv.org/abs/2304.12605>
- Anggara, D., Notodiputro, K. A., Indahwati, I., & Kurnia, A. (2025). Predictive modeling of right-skewed health insurance costs using copula regression and ensemble learning. *International Journal of Innovative Research and Scientific Studies*, 8(4), 1013–1021. <https://doi.org/10.53894/ijirss.v8i4.7986>
- Seyam, E. A. (2025). Predicting high-cost healthcare utilization using machine learning: A multi-service risk stratification analysis in EU-based private group health insurance. *Risks*, 13(7), 133. <https://doi.org/10.3390/risks13070133>
- Kunle, A. A., & Taiwo, K. A. (2025). Predictive modeling for healthcare cost analysis in the United States: A comprehensive review and future directions. *International Journal of Scientific Research and Modern Technology*, 4(1), 170–181. <https://ijsrmt.com/index.php/ijsrmt/article/view/569>
- Abdelrahman, S. et al. (2017). Supervised learning methods for predicting healthcare costs. *Journal of Biomedical Informatics*, 68, 59–69. <https://doi.org/10.1016/j.jbi.2017.02.008>
- Bertsimas, D., Dunn, J., Pawlowski, C., & Zhuo, Y. D. (2024). Health care cost prediction for heterogeneous patient profiles using machine learning. *Information Systems Research*, 35(1), 24–38. <https://doi.org/10.1287/isre.2021.0643>
- Elshawi, R., Al-Mallah, M. H., & Sakr, S. (2021). Predictive modeling of health insurance costs using deep learning. *BMC Medical Informatics and Decision Making*, 21(1), 43. <https://doi.org/10.1186/s12911-021-01743-z>
- Barboza, M., & Chakraborty, P. (2024). Explainable AI for health insurance pricing using tree-based models. *Computational Health Informatics*, 6(2), 107–118. <https://doi.org/10.1016/j.cmpb.2024.107328>
- Sugiharto, R., & Prasetyo, T. (2023). Analisis premi asuransi kesehatan koperasi menggunakan metode regresi dan PCA. *Jurnal Teknologi dan Kesehatan Masyarakat*, 5(2), 65–80.
- Wahyuni, R., & Setiawan, B. (2023). Pemodelan risiko asuransi kesehatan pada koperasi Indonesia. *Jurnal Sistem Informasi Kesehatan*, 12(1), 13–27.
- Trisnawati, D., & Nugroho, A. (2023). Perbandingan model prediksi premi asuransi menggunakan data lokal dan dataset publik. *Jurnal Teknologi Informasi dan Komputer*, 11(2), 95–108.
- Kurniawan, D., & Hidayat, M. (2023). Transformasi digital pada koperasi Indonesia dalam implementasi asuransi mikro berbasis komunitas. *Jurnal Ekonomi Digital dan Kewirausahaan Sosial*, 2(3), 41–56.