Jurnal Bisnis dan Manajemen
ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

PERBANDINGAN PREDIKSI KEBERLANJUTAN POLIS ASURANSI DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES, KNN, DAN SVM

Empu Galih Gondo Sukmo, Munawar, Gerry Firmansyah, Budi Tjahjono

Universitas Esa Unggul, Indonesia

*Corresponding Author:
Empu Galih Gondo Sukmo

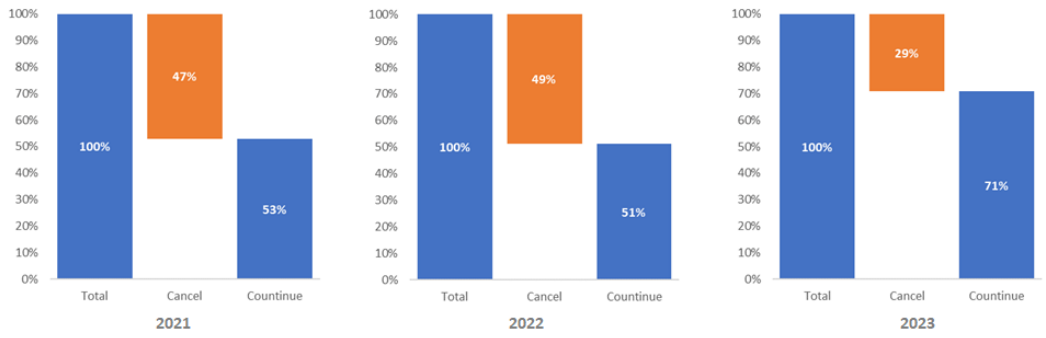
Abstract.

This research investigates insurance policy continuation using Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) algorithms. Data from PT Asuransi ABC, consisting of 35,420 data points, was analyzed to predict policy cancellations. The findings suggest that KNN outperforms other algorithms in handling non-linear data. KNN successfully identified reasons for policy cancellations with greater accuracy, while Gaussian Naive Bayes and SVM require further adjustments to enhance their accuracy. This research provides valuable insights for insurance companies in developing strategies to increase customer retention and reduce policy cancellation rates. This is achieved by understanding policy cancellation patterns, enabling companies to take proactive steps to provide better services and tailor their insurance products to meet customer needs.

Keywords: Insurance policy continuation, Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) algorithms

PENDAHULUAN

Industri asuransi saat ini dihadapkan pada berbagai tantangan, termasuk meningkatnya kompleksitas risiko, perubahan regulasi, dan kemajuan teknologi. Mitigasi risiko merupakan salah satu fokus utama dalam industri asuransi. Mitigasi Risiko adalah tindakan terencana dan berkelanjutan yang dilakukan oleh pemilik risiko agar bisa mengurangi dampak dari suatu kejadian yang berpotensi atau telah merugikan atau membahayakan pemilik risiko tersebut. Oleh sebab itu, perusahaan asuransi memiliki tanggung jawab untuk melakukan penilaian risiko yang akurat, memastikan keberlanjutan polis tanpa menimbulkan kerugian finansial yang berdampak besar (Ajeng Mega Pratiwi & Mutaqin, 2021). Hal ini tidak terkecuali juga berlaku pada PT Asuransi ABC.



Gambar 1. Persentase Nasabah yang Melakukan Pembatalan Polis di PT Asuransi ABC pada Tahun 2021-2023

ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

Data persentase pembatalan polis menunjukkan tren dari tahun 2021 hingga 2023 dengan total sebesar 41%. Pembatalan polis asuransi pada tahun 2021 sebesar 47%. Pada tahun 2022, terjadi peningkatan dengan persentase sebesar 49%. Sementara pada tahun 2023 jumlah pembatalan polis asuransi sebesar 29%.

Pembatalan polis asuransi memiliki beberapa dampak negatif bagi perusahaan asuransi. Salah satu dampak utama adalah kehilangan pendapatan premi yang diharapkan dari polis tersebut. Kehilangan nasabah juga berarti kehilangan potensi pendapatan jangka panjang. Selain itu, pembatalan polis bisa jadi mencerminkan ketidakpuasan nasabah akan pelayanan jasa asuransi. Kejadian pembatalan polis yang sering terjadi dapat merusak citra perusahaan dan menurunkan kepercayaan nasabah akan pelayanan jasa asuransi tersebut.

Perusahaan asuransi, seperti PT Asuransi ABC, perlu memprediksi tren dan fluktuasi masa depan. Prediksi ini menjadi krusial dalam menyusun strategi dan rencana mitigasi risiko. Dengan memahami pola pembatalan polis yang terjadi, perusahaan dapat meningkatkan respons dan membuat kebijakan yang lebih adaptif. Sebagai contoh, sinyal peningkatan persentase pembatalan polis dapat menjadi peringatan bagi perusahaan untuk meningkatkan komunikasi dengan nasabah, menawarkan insentif, atau menyusun program retensi nasabah yang lebih efektif.

Prediksi ini juga memungkinkan perusahaan untuk memiliki kesiapan yang lebih baik dalam menghadapi fluktuasi dan mengurangi dampak finansial. Dengan demikian, perusahaan dapat mengimplementasikan strategi yang lebih proaktif untuk mempertahankan nasabah dan memitigasi risiko pembatalan polis yang tinggi. Teknologi informasi dapat menjadi salah satu solusi dalam hal ini. Algoritma prediktif dapat digunakan oleh perusahaan asuransi untuk memprediksi keberlanjutan polis asuransi.

Dewasa ini ada beberapa algoritma yang dapat digunakan untuk memprediksi suatu kejadian, termasuk memprediksi keberlanjutan polis asuransi. Algoritma prediktif seperti *Naive Bayes*, KNN, dan SVM menjadi salah satu pilihan untuk menganalisis data historis dan tren, membantu perusahaan asuransi dalam mengambil keputusan yang lebih cerdas. Alasan di balik pemilihan algoritma *Naive Bayes*, *K-Nearest Neighbors* (KNN), dan *Support Vector Machine* (SVM) dalam penelitian ini melibatkan pertimbangan fitur-fitur spesifik dari masing-masing algoritma yang dapat memberikan performa yang baik dalam memprediksi keberlanjutan polis.

Naive Bayes dikenal efisien dan dapat mengatasi asumsi independensi fitur, KNN sederhana dan efektif untuk menemukan pola lokal, sementara SVM efektif untuk menangani data dengan dimensi tinggi dan pemisahan yang kompleks. Pemilihan ini didasarkan pada kebutuhan untuk menjawab tantangan spesifik dalam prediksi keberlanjutan polis yang melibatkan kompleksitas dan variasi yang signifikan dalam data nasabah.

Naive Bayes, sebagai algoritma klasifikasi yang berbasis pada teorema *Bayes*, sering digunakan dalam prediksi berbasis teks (Harianto, 2019). Di sisi lain, algoritma *K-Nearest Neighbors* atau bergantung pada pola kemiripan data untuk membuat prediksi (Cahyanti *et al*., 2020), sementara algoritma *Support Vector Machine* atau SVM berfokus pada pencarian garis pemisah optimal antara kelas data (Muhathir *et al*., 2021). Perbandingan ketiga algoritma ini tidak hanya akan memberikan pemahaman tentang kinerja mereka tetapi juga mengetahui kelebihan dan kekurangan ketiga algoritma tersebut dalam memprediksi keberlanjutan polis asuransi.

Pengembangan teknologi informasi memberikan akses yang lebih besar terhadap data historis dan *real-time*. Dengan memanfaatkan data ini secara efektif, perusahaan asuransi dapat meningkatkan akurasi prediksi mereka dan mengurangi risiko kegagalan keberlanjutan polis. Penelitian ini berusaha untuk merinci potensi penerapan algoritma-algoritma ini dalam meningkatkan kualitas prediksi keberlanjutan polis asuransi. Hasil dari penelitian ini yaitu penentuan algoritma di antara *naive bayes*, KNN, dan SVM yang paling baik untuk memprediksi keberlanjutan polis asuransi. Dengan demikian jasa asuransi seperti PT Asuransi ABC dapat menerapkan algoritma tersebut untuk memprediksi keberlanjutan polis asuransi nasabah mereka. Dengan mengetahui prediksi tersebut, perusahaan dapat mengetahui nasabah mana yang mungkin akan membatalkan polis asuransi. Perusahaan kemudian dapat meningkatkan respons dan

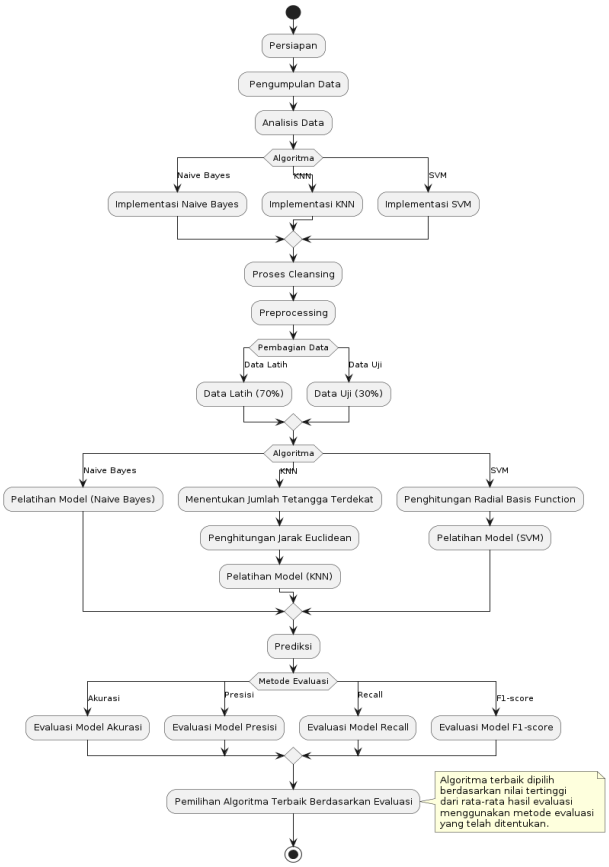
ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

membuat kebijakan yang lebih adaptif pada nasabah terkait agar risiko pembatalan polis asuransi oleh nasabah tersebut semakin berkurang.

Penelitian ini menggunakan tiga jenis produk asuransi untuk prediksi keberlanjutan polis: *Future Security*, *Capital Builder*, dan *Lifetime Assurance. Future Security* adalah produk asuransi yang fokus pada perlindungan jangka panjang terhadap risiko masa depan. *Capital Builder* adalah produk asuransi yang dirancang untuk membantu nasabah membangun modal atau tabungan jangka panjang. *Lifetime Assurance* adalah produk asuransi seumur hidup yang memberikan perlindungan dan manfaat selama masa hidup nasabah.

Alasan pembatalan polis yang dipertimbangkan dalam penelitian ini mencakup: *Family Reason* (alasan keluarga), yang meliputi perubahan situasi keluarga seperti pernikahan, kelahiran, atau kematian yang mempengaruhi kebutuhan asuransi; *Financial Instability* (ketidakstabilan keuangan), di mana nasabah menghadapi kesulitan keuangan yang membuat mereka tidak mampu membayar premi; *Found Better Alternative* (menemukan alternatif yang lebih baik), ketika nasabah menemukan produk asuransi lain yang menawarkan manfaat atau premi lebih baik; dan *Product Benefit* (manfaat produk), di mana nasabah memutuskan untuk tetap melanjutkan polis karena manfaat yang diberikan produk sesuai dengan kebutuhan mereka. Dengan memahami karakteristik produk dan alasan pembatalan, perusahaan asuransi dapat mengembangkan strategi yang lebih efektif untuk mempertahankan nasabah dan meningkatkan keberlanjutan polis.

METODE



Gambar 2. Kerangka Berpikir

Flowchart ini menggambarkan proses analisis data untuk prediksi keberlanjutan polis asuransi menggunakan tiga algoritma: Naive Bayes, K-Nearest Neighbors (KNN), dan Support Vector Machine (SVM). Proses dimulai dengan tahap persiapan, di mana data yang diperlukan

ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

dikumpulkan dari PT Asuransi ABC. Data tersebut diperoleh dari tahun 2021 hingga 2023 dengan total 35.420 rows.

Selanjutnya, data tersebut dianalisis untuk mempersiapkan implementasi algoritma yang sesuai. Berdasarkan pilihan algoritma yang dipilih, proses berlanjut dengan tahap implementasi, yang mencakup langkah-langkah khusus sesuai dengan algoritma yang dipilih.

Setelah itu, proses cleansing dan preprocessing dilakukan untuk mempersiapkan data sebelum pembagian menjadi dua bagian: data latih dan data uji. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk menguji performa model yang telah dilatih. Pemisahan data ini dilakukan dengan proporsi 80% data latih dan 20% data uji.

Selanjutnya, proses pelatihan model dilakukan sesuai dengan algoritma yang dipilih. Untuk Naive Bayes, proses pelatihan model dilakukan langsung setelah proses pembagian data. Sedangkan untuk KNN, terdapat langkah tambahan yaitu menentukan jumlah tetangga terdekat dan menghitung jarak Euclidean sebelum pelatihan model. Untuk SVM, langkah tambahan adalah penghitungan radial basis function sebelum pelatihan model.

Setelah model dilatih, proses prediksi dilakukan untuk memprediksi keberlanjutan polis asuransi. Selanjutnya, model dievaluasi menggunakan metode evaluasi yang telah ditentukan, seperti akurasi, presisi, recall, dan F1-score. Evaluasi ini dilakukan untuk mengukur sejauh mana model dapat memprediksi keberlanjutan polis dengan tepat.

Terakhir, algoritma terbaik dipilih berdasarkan nilai tertinggi dari rata-rata hasil evaluasi menggunakan metode evaluasi yang telah ditentukan sebelumnya. Hal ini membantu dalam menentukan algoritma yang paling efektif dalam memprediksi keberlanjutan polis asuransi.

HASIL DAN PEMBAHASAN

Deskripsi Data

Data yang digunakan dalam penelitian ini berasal dari PT Asuransi ABC, dengan jumlah total 35.420 rows. Data ini mencakup informasi mengenai tiga jenis produk asuransi yang berbeda, yaitu Future Security, Capital Builder, dan Lifetime Assurance. Masing-masing jenis produk memiliki karakteristik dan pola penggunaannya sendiri, yang akan dianalisis untuk memprediksi apakah polis asuransi akan terus berlanjut atau dibatalkan. Jika polis dibatalkan, analisis juga mencakup alasan pembatalannya. Deskripsi data yang digunakan dalam penelitian ini meliputi atribut-atribut penting yang mencakup berbagai aspek dari produk asuransi dan profil pemegang polis.

Produk asuransi Future Security dirancang untuk memberikan jaminan keamanan finansial di masa depan, sedangkan Capital Builder bertujuan untuk membantu pemegang polis dalam membangun modal melalui investasi yang terstruktur. Lifetime Assurance, di sisi lain, menawarkan jaminan seumur hidup dengan manfaat yang lebih komprehensif. Setiap jenis produk asuransi ini memiliki atribut yang serupa, tetapi dengan variasi dalam detail dan parameter spesifik. Atribut-atribut yang terdapat dalam data ini antara lain Nomor Polis, Payment Plan, Payment Method, Product, Sum Insured, Amount, Umur Polis, Province, Kategori Pekerjaan, Pendapatan, Periode Awal, Periode Akhir, Jenis Kelamin, Age, Age Generation, Cancel Reason, Insurance Status.

Atribut 'Product' mengidentifikasi jenis produk asuransi yang dimiliki oleh pemegang polis. Ini penting untuk analisis karena setiap produk memiliki karakteristik dan risiko yang berbeda yang mempengaruhi keberlanjutan polis. 'Payment Plan' mencakup rencana pembayaran yang dipilih oleh pemegang polis, yang dapat berupa pembayaran bulanan, triwulanan, atau tahunan. Rencana pembayaran ini dapat mempengaruhi tingkat keberlanjutan polis, karena frekuensi dan jumlah pembayaran dapat menjadi beban finansial bagi pemegang polis.

'Amount' merujuk pada jumlah uang yang diasuransikan atau nilai polis asuransi. Nilai ini penting dalam analisis karena polis dengan jumlah yang lebih tinggi mungkin memiliki risiko yang berbeda dibandingkan dengan polis dengan jumlah yang lebih rendah. 'Kategori Pekerjaan' mengindikasikan pekerjaan pemegang polis, yang dapat memberikan wawasan tentang stabilitas

ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

keuangan dan kemampuan untuk mempertahankan polis asuransi. Misalnya, pekerjaan dengan pendapatan yang stabil dan tinggi mungkin lebih mungkin untuk mempertahankan polis asuransi dibandingkan dengan pekerjaan yang memiliki pendapatan yang tidak stabil.

'Jenis Kelamin' dan 'Age' adalah atribut demografis yang penting dalam analisis asuransi. Jenis kelamin dapat mempengaruhi jenis produk asuransi yang dipilih dan risiko yang terkait dengannya. Usia pemegang polis juga merupakan faktor kunci, karena usia yang lebih tua mungkin memiliki risiko kesehatan yang lebih tinggi yang dapat mempengaruhi keputusan untuk melanjutkan atau membatalkan polis asuransi. 'Generation' mengklasifikasikan pemegang polis ke dalam kelompok generasi yang berbeda, seperti Baby Boomers, Generation X, Millennials, dan Generation Z. Generasi yang berbeda mungkin memiliki pandangan yang berbeda tentang asuransi dan risiko, yang dapat mempengaruhi keputusan mereka terkait polis asuransi.

'Insurance Status' adalah atribut yang menunjukkan status keberlanjutan polis asuransi. Status ini mencakup apakah polis tersebut aktif (continue) atau telah dibatalkan (cancel). Jika polis dibatalkan, atribut ini juga mencakup alasan pembatalannya, seperti ketidakmampuan membayar premi, perubahan situasi finansial, atau ketidakpuasan dengan layanan asuransi. Analisis status ini penting untuk memahami faktor-faktor yang mempengaruhi keberlanjutan polis dan untuk mengidentifikasi pola yang dapat membantu dalam merancang strategi untuk meningkatkan retensi pelanggan.

'Provinsi' adalah atribut yang menunjukan perbandingan geografis kita bisa memperkaya data dengan informasi lokasi mengenai provinsi di mana polis asuransi tersebut berlaku. Ini dapat memberikan wawasan tambahan tentang variasi geografis dalam keberlanjutan polis asuransi. Dengan demikian, menambahkan elemen provinsi ke analisis "Insurance Status" dapat memberikan pandangan yang lebih holistik dan membantu mengidentifikasi peluang untuk meningkatkan layanan dan retensi berdasarkan faktor geografis.

Data yang digunakan dalam penelitian ini mencakup berbagai kombinasi atribut-atribut di atas untuk setiap jenis produk asuransi. Distribusi data menunjukkan bahwa ada variasi yang signifikan dalam setiap atribut, yang memungkinkan analisis yang mendalam untuk memahami pola dan hubungan antara atribut-atribut tersebut. Sebagai contoh, analisis awal menunjukkan bahwa polis dengan jumlah yang lebih tinggi cenderung lebih sering dibatalkan, terutama jika pemegang polis memiliki pekerjaan dengan pendapatan yang tidak stabil. Demikian pula, pemegang polis yang lebih tua lebih mungkin untuk mempertahankan polis mereka dibandingkan dengan pemegang polis yang lebih muda.

Pra-pemrosesan data melibatkan pengkodean variabel kategorikal menggunakan LabelEncoder untuk mengubah nilai teks menjadi nilai numerik yang dapat digunakan dalam model prediksi. Pengkodean ini penting untuk memastikan bahwa model dapat mengolah data dengan benar dan menghasilkan prediksi yang akurat. Selain itu, data dibagi menjadi set pelatihan dan pengujian untuk memastikan bahwa model dapat dievaluasi dengan benar dan hasil prediksi dapat diandalkan.

Dengan menggunakan berbagai model klasifikasi seperti Gaussian Naive Bayes, K-Nearest Neighbors, dan Support Vector Classifier, analisis ini berfokus pada prediksi status keberlanjutan polis asuransi. Setiap model dilatih menggunakan set pelatihan dan kemudian dievaluasi menggunakan set pengujian untuk menentukan kinerja model dalam memprediksi apakah polis asuransi akan terus berlanjut atau dibatalkan. Hasil evaluasi menunjukkan bahwa ada perbedaan kinerja antara model-model tersebut, dengan beberapa model menunjukkan tingkat akurasi yang lebih tinggi untuk jenis produk tertentu.

Selain itu, analisis asosiasi digunakan untuk mengidentifikasi pola dan hubungan antara atribut-atribut dalam data. Analisis ini membantu dalam memahami faktor-faktor yang berkontribusi terhadap keputusan pemegang polis untuk mempertahankan atau membatalkan polis asuransi. Sebagai contoh, analisis asosiasi mungkin mengungkapkan bahwa pemegang polis dengan rencana pembayaran bulanan lebih cenderung membatalkan polis mereka dibandingkan

ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

dengan mereka yang memilih rencana pembayaran tahunan, karena beban finansial bulanan yang lebih tinggi.

Hasil analisis ini memberikan wawasan yang berharga bagi PT Asuransi ABC dalam merancang strategi untuk meningkatkan retensi pelanggan dan mengurangi tingkat pembatalan polis. Dengan memahami faktor-faktor yang mempengaruhi keputusan pemegang polis, perusahaan dapat mengambil langkah-langkah proaktif untuk memberikan layanan yang lebih baik dan menyesuaikan produk asuransi sesuai dengan kebutuhan pelanggan. Misalnya, perusahaan dapat menawarkan opsi pembayaran yang lebih fleksibel atau memberikan insentif bagi pemegang polis yang mempertahankan polis mereka dalam jangka waktu tertentu.

Secara keseluruhan, deskripsi data ini memberikan gambaran komprehensif tentang karakteristik dan pola yang ada dalam data asuransi dari PT Asuransi ABC. Dengan menggunakan pendekatan analisis yang sistematis dan berbagai metode prediksi, penelitian ini berusaha untuk memberikan pemahaman yang lebih baik tentang faktor-faktor yang mempengaruhi keberlanjutan polis asuransi dan untuk mengidentifikasi strategi yang dapat membantu perusahaan dalam meningkatkan retensi pelanggan dan mengoptimalkan kinerja bisnis.

Pra-Pemrosesan Data

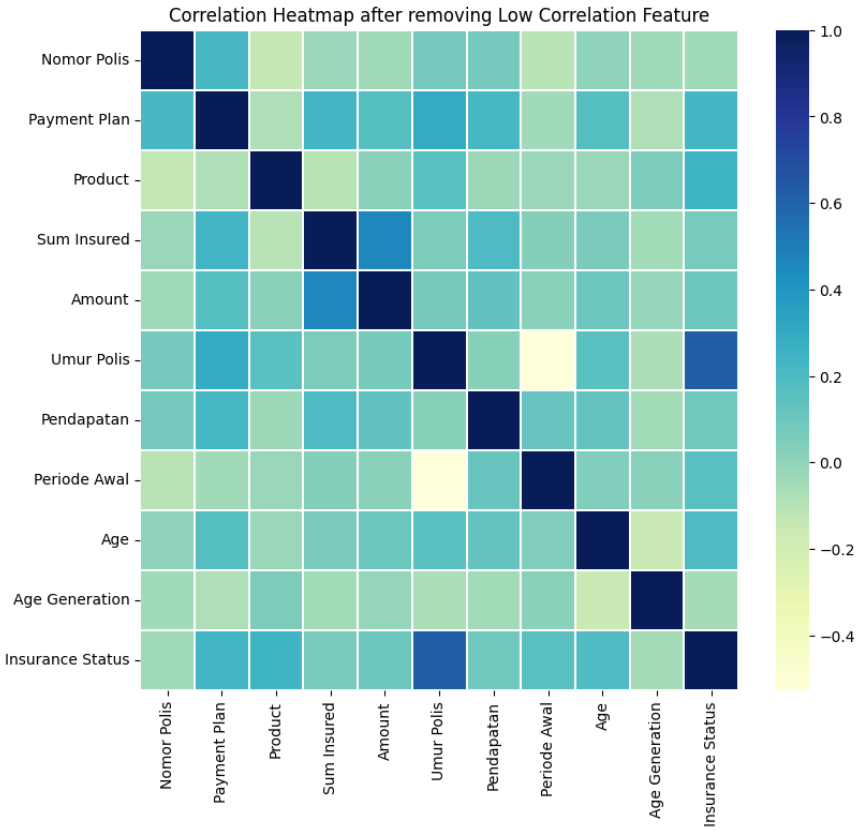
Pra-pemrosesan data merupakan langkah krusial dalam analisis data dan pengembangan model prediktif. Langkah ini memastikan bahwa data dalam kondisi yang optimal untuk diproses lebih lanjut oleh algoritma pembelajaran mesin. Pada penelitian ini, data asuransi dari PT Asuransi ABC dengan jumlah 35.420 catatan diproses untuk memprediksi keberlanjutan polis asuransi berdasarkan beberapa atribut seperti jenis produk, rencana pembayaran, jumlah asuransi, pekerjaan, jenis kelamin, usia, generasi, dan status asuransi. Langkah-langkah pra-pemrosesan data yang dilakukan meliputi pembersihan data, pengkodean variabel kategorikal, dan pembagian data menjadi set pelatihan dan pengujian.

Langkah pertama dalam pra-pemrosesan data adalah pembersihan data. Data mentah seringkali mengandung nilai-nilai yang hilang, duplikat, atau tidak konsisten yang dapat mempengaruhi kinerja model. Oleh karena itu, data dibersihkan untuk memastikan bahwa setiap catatan memiliki informasi yang lengkap dan akurat. Pada tahap ini, catatan dengan nilai yang hilang atau tidak valid dihapus atau diimputasi berdasarkan strategi yang sesuai. Misalnya, jika ada catatan yang tidak memiliki informasi tentang usia pemegang polis, catatan tersebut dapat dihapus atau usia dapat diimputasi berdasarkan rata-rata usia dari dataset.

Setelah data dibersihkan, langkah selanjutnya adalah pengkodean variabel kategorikal. Variabel kategorikal adalah variabel yang mengandung nilai-nilai non-numerik, seperti jenis produk asuransi, pekerjaan, jenis kelamin, dan generasi. Karena algoritma pembelajaran mesin memerlukan input numerik, variabel-variabel ini perlu dikonversi menjadi format numerik yang dapat digunakan oleh model pembelajaran mesin.

Langkah selanjutnya dalam pra-pmprosesan data adalah melihat hubungan antara atribut (*Features Correlation)* Hubungan antara atribut ini berguna untuk melihat hubungan yang rendah dan hubungan yang tinggi. Jika suatu atribut memiliki hubungan yang rendah dengan atribut *Insurance Status* maka bisa dihilangkan atau dibersihkan.

Berikut ini hasil dari heatmap atribut yang sudah dihapus atau dibersihkan karena memiliki hubungan yang rendah dengan *Insurance Status.*



Gambar 3. Heatmap Setelah Atribut Low Correlation Dihapus

Langkah selanjutnya adalah data Balancing (Penyeimbang data). Dengan balancing data untuk meningkatkan akurasi dan memperbaiki ketidakseimbangan. Ketidakseimbangan data terjadi ketika jumlah objek dalam satu kelas data lebih besar dari kelas data lainnya. Ini terjadi ketika objek dalam kelas data utama banyak dan objek dalam kelas data minor sedikit. Penggunaan data yang tidak seimbang sangat memengaruhi hasil model. Sebagian besar algoritma pemrosesan mengabaikan kelas minor dan kewalahan jika mengabaikan ketidakseimbangan data (V. N. Chawla, K. W. Bowyer, 2002). Solusi untuk menangani data yang tidak seimbang adalah pendekatan SMOTE (V. N. Chawla, K. W. Bowyer, 2002). Meskipun prinsipnya berbeda dari pendekatan oversampling yang disarankan di atas, prinsip SMOTE meningkatkan pengamatan acak sehingga setara.

Data balancing dimaksudkan untuk mengatasi data tidak seimbang pada setiap kelas klasifikasi, sehingga hasil klasifikasi tersebut tidak akan menampilkan hasil kelas paling dominan (Putra, N. S., Hutabarat, B. F., & Khaira, 2023).

Selain itu, ketidakseimbangan data antara kelas mayoritas dan minoritas dapat menyebabkan pengklasifikasi bergantung pada kelas mayoritas (Mutmainah, 2021). SMOTE telah terbukti berhasil dalam berbagai aplikasi dalam berbagai bidang (Fernández, A., García, S., Herrera, F., & Chawla, 2018).

Langkah selanjutnya dalam pra-pemrosesan data adalah pembagian data menjadi set pelatihan dan pengujian. Pembagian ini penting untuk mengevaluasi kinerja model dengan benar. Data pelatihan digunakan untuk melatih model, sementara data pengujian digunakan untuk menguji model dan mengukur akurasinya. Pembagian data dilakukan sedemikian rupa sehingga set pelatihan dan pengujian memiliki distribusi yang serupa untuk memastikan bahwa model dapat generalisasi dengan baik ke data yang belum pernah dilihat sebelumnya. Dalam penelitian ini, data yang dibagi menjadi set pelatihan dan pengujian telah disiapkan sebelumnya dan dimuat dari file CSV.

ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

Setelah data dibagi menjadi set pelatihan dan pengujian, model klasifikasi dapat dilatih menggunakan data pelatihan. Tiga model klasifikasi yang digunakan dalam penelitian ini adalah Gaussian Naive Bayes, K-Nearest Neighbors, dan Support Vector Classifier. Setiap model dilatih untuk memprediksi status keberlanjutan polis asuransi berdasarkan fitur yang ada. Model-model ini dipilih karena mereka mewakili berbagai pendekatan dalam pembelajaran mesin dan memiliki kinerja yang baik dalam berbagai situasi.

Pra-pemrosesan data juga melibatkan penanganan data yang tidak seimbang. Dalam beberapa kasus, distribusi kelas dalam target mungkin tidak seimbang, yang berarti bahwa beberapa kelas memiliki lebih banyak contoh daripada yang lain. Hal ini dapat mempengaruhi kinerja model karena model cenderung lebih baik dalam memprediksi kelas mayoritas. Untuk mengatasi masalah ini, teknik seperti oversampling, undersampling, atau penggunaan metrik evaluasi yang disesuaikan dapat digunakan. Dalam penelitian ini, distribusi kelas dalam status asuransi dianalisis untuk memastikan bahwa model tidak bias terhadap kelas mayoritas.

Pra-pemrosesan data adalah langkah penting yang memastikan bahwa data dalam kondisi yang optimal untuk diproses lebih lanjut oleh algoritma pembelajaran mesin. Langkah-langkah ini mencakup pembersihan data, pengkodean variabel kategorikal, pembagian data menjadi set pelatihan dan pengujian, dan penanganan data yang tidak seimbang. Dengan melakukan pra-pemrosesan yang menyeluruh, penelitian ini dapat memastikan bahwa model yang dikembangkan memiliki kinerja yang baik dan dapat menghasilkan prediksi yang akurat tentang keberlanjutan polis asuransi. Analisis ini memberikan wawasan yang berharga bagi PT Asuransi ABC dalam merancang strategi untuk meningkatkan retensi pelanggan dan mengurangi tingkat pembatalan polis.

Dalam penelitian ini, berbagai teknik pra-pemrosesan data digunakan untuk memastikan bahwa data dalam kondisi yang optimal untuk analisis lebih lanjut. Teknik-teknik ini mencakup penghapusan nilai yang hilang, pengkodean variabel kategorikal, pembagian data, dan penanganan data yang tidak seimbang. Setiap langkah dalam pra-pemrosesan data dirancang untuk memaksimalkan kinerja model prediktif dan menghasilkan wawasan yang akurat dan dapat diandalkan. Dengan melakukan pra-pemrosesan data yang menyeluruh, penelitian ini dapat memberikan hasil yang berarti dan bermanfaat bagi PT Asuransi ABC dalam upaya mereka untuk meningkatkan keberlanjutan polis asuransi dan meminimalkan pembatalan polis.

Training dan Testing Data

Pelatihan dan pengujian data menggunakan algoritma pembelajaran mesin adalah langkah dalam analisis data untuk membangun model prediktif yang akurat. Dalam penelitian ini, data asuransi dari PT Asuransi ABC dengan jumlah 35.420 catatan digunakan untuk melatih model yang memprediksi keberlanjutan polis asuransi. Tiga algoritma pembelajaran mesin yang digunakan adalah Gaussian Naive Bayes (NB), K-Nearest Neighbors (KNN), dan Support Vector Machine (SVM). Setiap algoritma dipilih berdasarkan karakteristiknya yang unik dan kemampuannya dalam menangani berbagai jenis data.

Gaussian Naive Bayes (NB) adalah algoritma pembelajaran mesin yang didasarkan pada Teorema Bayes dengan asumsi bahwa fitur-fitur yang digunakan untuk prediksi bersifat independen. Meskipun asumsi ini jarang terpenuhi sepenuhnya dalam praktik, NB seringkali memberikan kinerja yang baik dalam banyak kasus. Algoritma ini sederhana namun efektif dan dapat menangani data dengan baik bahkan dalam jumlah data yang besar seperti dalam penelitian ini. NB cocok digunakan untuk masalah klasifikasi yang melibatkan variabel kategorikal dan numerik. Dalam pelatihan menggunakan NB, data yang telah diproses diinput ke dalam model, dan model tersebut dilatih untuk mengenali pola-pola dalam data yang mengarah pada keberlanjutan atau pembatalan polis asuransi.

K-Nearest Neighbors (KNN) adalah algoritma pembelajaran mesin berbasis instance yang mengklasifikasikan data baru berdasarkan kedekatan dengan data pelatihan yang sudah ada. Algoritma ini tidak membuat asumsi kuat tentang distribusi data dan bekerja dengan baik pada

ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

data yang tidak linier. KNN menghitung jarak antara data baru dan data pelatihan menggunakan metrik seperti jarak Euclidean, dan menentukan kelas data baru berdasarkan mayoritas kelas dari K tetangga terdekatnya. Algoritma ini intuitif dan mudah dipahami, namun memerlukan banyak memori dan komputasi saat melakukan prediksi pada data yang besar. Dalam konteks penelitian ini, KNN digunakan untuk memahami bagaimana kedekatan antara catatan polis asuransi yang berbeda dapat membantu dalam memprediksi keberlanjutan polis.

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. SVM bekerja dengan mencari hyperplane terbaik yang memisahkan kelas-kelas data dalam ruang fitur. Algoritma ini efektif dalam ruang fitur yang tinggi dan menggunakan kernel trick untuk menangani data yang tidak linier. SVM memiliki kemampuan untuk menemukan margin pemisah yang maksimal antara kelas-kelas, yang membuatnya sangat kuat dalam menangani data yang kompleks dan memiliki outlier. Dalam penelitian ini, SVM digunakan untuk mempelajari batasan optimal antara polis asuransi yang berlanjut dan yang dibatalkan, serta alasan pembatalan jika ada.

Proses ini dimulai dengan memfilter data pelatihan dan pengujian berdasarkan jenis produk. Data yang telah difilter kemudian digunakan untuk melatih model, dan model tersebut digunakan untuk membuat prediksi pada data pengujian. Hasil prediksi kemudian dikonversi kembali ke format teks asli untuk memudahkan interpretasi hasil.

Evaluasi Model

Evaluasi algoritma adalah bagian penting dalam proses pengembangan model pembelajaran mesin karena memberikan gambaran tentang seberapa baik model tersebut bekerja dalam memprediksi hasil yang diinginkan. Dalam penelitian ini, tiga algoritma pembelajaran mesin—Gaussian Naive Bayes (NB), K-Nearest Neighbors (KNN), dan Support Vector Machine (SVM)—dievaluasi untuk memprediksi keberlanjutan polis asuransi dari PT Asuransi ABC. Evaluasi ini bertujuan untuk menentukan seberapa efektif masing-masing algoritma dalam memprediksi apakah polis asuransi akan berlanjut atau dibatalkan, dan jika dibatalkan, alasan di balik pembatalan tersebut.

Gaussian Naive Bayes (NB) adalah algoritma klasifikasi yang mengasumsikan bahwa fitur-fitur dalam data saling independen. Meskipun asumsi ini jarang sepenuhnya terpenuhi dalam praktik, NB seringkali memberikan hasil yang baik karena kesederhanaannya dan kemampuannya untuk menangani data dengan baik bahkan ketika fitur tidak benar-benar independen. Evaluasi model NB dilakukan dengan menghitung berbagai metrik evaluasi seperti confusion matrix, classification report, dan accuracy score.

Confusion matrix memberikan informasi mengenai jumlah prediksi yang benar dan salah untuk setiap kelas. Dalam konteks model NB, confusion matrix digunakan untuk mengevaluasi seberapa baik model dalam mengklasifikasikan polis sebagai "continue" atau "cancel" serta alasan pembatalan jika ada. Classification report, di sisi lain, memberikan metrik tambahan seperti precision, recall, dan F1-score yang membantu dalam memahami kinerja model pada setiap kelas secara lebih mendetail. Precision menunjukkan proporsi prediksi yang benar dari semua prediksi yang dilakukan untuk kelas tertentu, recall menunjukkan proporsi data yang benar-benar termasuk dalam kelas tersebut yang berhasil diprediksi, dan F1-score adalah rata-rata harmonis dari precision dan recall. Accuracy score mengukur persentase prediksi yang benar dari total prediksi yang dilakukan oleh model.

K-Nearest Neighbors (KNN) adalah algoritma yang mengklasifikasikan data berdasarkan kedekatannya dengan data pelatihan. Algoritma ini bekerja dengan cara menghitung jarak antara data baru dan data pelatihan, kemudian menentukan kelas data baru berdasarkan mayoritas kelas dari K tetangga terdekatnya. Evaluasi model KNN melibatkan pengukuran seberapa baik model ini dalam memprediksi kelas yang benar dengan membandingkan hasil prediksi terhadap data aktual.

ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

Confusion matrix untuk model KNN menunjukkan distribusi hasil prediksi dibandingkan dengan label sebenarnya. Ini memungkinkan analisis lebih dalam tentang bagaimana model KNN mengklasifikasikan data pada setiap kelas. Classification report untuk KNN memberikan metrik yang serupa dengan NB, namun karena KNN tidak membuat asumsi tentang distribusi data, hasilnya mungkin berbeda terutama dalam data yang tidak linier. Accuracy score untuk KNN memberikan gambaran tentang seberapa sering model memprediksi kelas yang benar dari total prediksi yang dilakukan.

Support Vector Machine (SVM) adalah algoritma yang mencari hyperplane terbaik yang memisahkan kelas dalam ruang fitur. SVM menggunakan kernel trick untuk menangani data yang tidak linier dan dapat menemukan margin pemisah optimal antara kelas-kelas. Evaluasi model SVM dilakukan dengan menghitung confusion matrix, classification report, dan accuracy score seperti pada algoritma lainnya.

Confusion matrix untuk model SVM membantu dalam mengidentifikasi bagaimana model membagi data antara kelas-kelas yang ada. Classification report memberikan metrik yang menunjukkan performa model pada setiap kelas, sementara accuracy score mengukur proporsi prediksi yang benar. SVM sering kali menunjukkan kinerja yang baik terutama dalam data yang kompleks dan memiliki outlier, sehingga hasil evaluasi model ini dapat memberikan wawasan yang mendalam tentang bagaimana model tersebut menangani pola yang rumit dalam data.

Model kemudian digunakan untuk membuat prediksi pada data pengujian. Hasil prediksi dikonversi kembali ke format teks asli untuk kemudahan interpretasi. Confusion matrix, classification report, dan accuracy score dihitung dan ditampilkan untuk memberikan gambaran tentang kinerja model pada setiap jenis produk. Hasilnya adalah sebagai berikut:

Evaluasi Metode Naïve Bayes

Tabel 1. Gaussian Naive Bayes (NB) Tanpa Optimasi

Insurance Status	***Precision***	***Recall***	***F1-Score***
Countinue	0.52	0.98	0.68
Cancel	0.85	0.12	0.21
Akurasi			0.541054

Pada tabel di atas dapat dilihat bahwa evaluasi model dengan metode Naïve Bayes tanpa optimasi menghasilkan akurasi sebesar 0.541054 dengan nilai *precision* pada status *countinue* bernilai 0.52. Sedangkan untuk status *cancel* bernilai 0.85. *Recall* pada status *countinue* bernilai 0.98, untuk status *cancel* bernilai 0.12. *F1-Score* pada status *countinue* bernilai 0.68, untuk status *cancel* bernilai 0.21.

Tabel 2. Gaussian Naïve Bayes (NB) Dengan Optimasi

Insurance Status	***Precision***	***Recall***	***F1-Score***
Countinue	0.64	0.96	0.77
Cancel	0.92	0.48	0.63

ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

Akurasi		0.714904

Pada tabel di atas dapat dilihat bahwa evaluasi model dengan metode Naïve Bayes dengan optimasi menghasilkan akurasi sebesar 0.714904 dengan nilai *precision* pada status *countinue* bernilai 0.64. Sedangkan untuk status *cancel* bernilai 0.92. *Recall* pada status *countinue* bernilai 0.96, untuk status *cancel* bernilai 0.48. *F1-Score* pada status *countinue* bernilai 0.77, untuk status *cancel* bernilai 0.63.

Evaluasi Metode K-Nearest Neighbors (KNN)

Tabel 3. K-Nearest Neighbors (KNN) Tanpa Optimasi

Insurance Status	***Precision***	***Recall***	***F1-Score***
Countinue	0.68	0.76	0.72
Cancel	0.74	0.66	0.70
Akurasi			0.709125

Pada tabel di atas dapat dilihat bahwa evaluasi model dengan metode KNN tanpa optimasi menghasilkan akurasi sebesar 0.709125 dengan nilai *precision* pada status *countinue* bernilai 0.68. Sedangkan untuk status *cancel* bernilai 0.74. *Recall* pada status *countinue* bernilai 0.76, untuk status *cancel* bernilai 0.66. *F1-Score* pada status *countinue* bernilai 0.72, untuk status *cancel* bernilai 0.70.

Tabel 4. K-Nearest Neighbors (KNN) Dengan Optimasi

Insurance Status	***Precision***	***Recall***	***F1-Score***
Countinue	0.69	0.76	0.72
Cancel	0.74	0.68	0.71
Akurasi			0.715988

Pada tabel di atas dapat dilihat bahwa evaluasi model dengan metode KNN dengan optimasi menghasilkan akurasi sebesar 0.715988 dengan nilai *precision* pada status *countinue* bernilai 0.69. Sedangkan untuk status *cancel* bernilai 0.74. *Recall* pada status *countinue* bernilai 0.76, untuk status *cancel* bernilai 0.68. *F1-Score* pada status *countinue* bernilai 0.72, untuk status *cancel* bernilai 0.71.

ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

Evaluasi Metode Support Vector Machine (SVM)

Tabel 5. Support Vector Machine (SVM) Tanpa Optimasi

Insurance Status	***Precision***	***Recall***	***F1-Score***
Countinue	0.47	0.48	0.47
Cancel	0.48	0.47	0.48
Akurasi			0.474837

Pada tabel di atas dapat dilihat bahwa evaluasi model dengan metode SVM tanpa optimasi menghasilkan akurasi sebesar 0.474840 dengan nilai *precision* pada status *countinue* bernilai 0.47. Sedangkan untuk status *cancel* bernilai 0.48. *Recall* pada status *countinue* bernilai 0.48, untuk status *cancel* bernilai 0.47. *F1-Score* pada status *countinue* bernilai 0.47, untuk status *cancel* bernilai 0.48.

Tabel 6. Support Vector Machine (SVM) Dengan Optimasi

Insurance Status	***Precision***	***Recall***	***F1-Score***
Countinue	0.47	0.49	0.48
Cancel	0.48	0.47	0.48
Akurasi			0.477365

Pada tabel di atas dapat dilihat bahwa evaluasi model dengan metode SVM dengan optimasi menghasilkan akurasi sebesar 0.477365 dengan nilai *precision* pada status *countinue* bernilai 0.47. Sedangkan untuk status *cancel* bernilai 0.48. *Recall* pada status *countinue* bernilai 0.49, untuk status *cancel* bernilai 0.47. *F1-Score* pada status *countinue* bernilai 0.48, untuk status *cancel* bernilai 0.48.

Dalam penelitian ini, penulis menganalisis keberlanjutan polis asuransi dengan menggunakan metode klasifikasi dan asosiasi. Fokus utama dari penelitian ini adalah untuk mengevaluasi seberapa baik berbagai model klasifikasi dapat memprediksi apakah polis asuransi akan berlanjut atau dibatalkan, serta alasan di balik pembatalan tersebut. Dengan menggunakan dataset dari PT Asuransi ABC yang mencakup tiga jenis produk asuransi—Future Security, Capital Builder, dan Lifetime Assurance. Penulis menerapkan model Gaussian Naive Bayes (NB), K-Nearest Neighbors (KNN), dan Support Vector Machine (SVM) untuk mengevaluasi performa mereka dalam hal akurasi, precision, recall, dan F1-score. Pembahasan ini bertujuan untuk memberikan wawasan mendalam mengenai hasil evaluasi dan implikasinya terhadap strategi manajemen polis asuransi.

Evaluasi Model Naïve Bayes

Untuk model Gaussian Naive Bayes tanpa optimasi menghasilkan akurasi sebesar 54.10% yang menunjukkan performa kurang optimal. Sedangkan model Naïve Bayes dengan optimasi

ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

menghasilkan akurasi 71.49% yang menunjukkan performa yang cukup optimal. Model yang menggunakan optimasi mengalami kenaikan akurasi yang relatif tinggi sehingga sudah lebih baik untuk melakukan prediksi dibandingkan model Naïve Bayes tanpa optimasi.

Di sisi lain, model K-Nearest Neighbors (KNN) tanpa menunjukkan performa yang lebih baik dari Naïve Bayes dengan akurasi 70.91%. Sedangkan untuk model KNN dengan optimasi menghasilkan akurasi 71.59%. Model yang menggunakan optimasi mengalami kenaikan akurasi sebesar 0.68%. KNN memiliki kemampuan untuk menangani data dengan lebih baik, seperti yang terlihat dari precision dan recall yang lebih tinggi. Model ini tampaknya lebih efektif dalam mengidentifikasi *insurance status.* Akurasi KNN dengan optimasi lebih tinggi dari model Naïve Bayes dengan optimasi.

Support Vector Machine (SVM) tanpa optimasi memiliki akurasi 47.48%. Sedangkan model SVM dengan optimasi menghasilkan akurasi sebesar 47.73%. Ada sedikit kenaikan akurasi pada SVM dengan optimasi. Meskipun SVM hanya dapat mengenali kategori "Continue" dengan baik, precision dan recall untuk kategori lain sangat rendah. Hal ini menunjukkan bahwa meskipun SVM dapat bekerja dengan baik dalam klasifikasi biner, performanya mungkin kurang optimal dalam kasus di mana data memiliki banyak kelas atau di mana data tidak dapat dipisahkan secara linier.

Perbandingan dan Implikasi

Tabel 7. Perbandingan Model

No	Model	Data Training	Data Testing
1	Gaussian Naïve Bayes	54.75%	54.10%
2	K-Nearest Neighbors (KNN)	79.71%	70.91%
3	Support Vector Machine (SVM)	48.25%	47.48%
4	Gaussian Naïve Bayes Optimasi	72.49%	71.49%
5	K-Nearest Neighbors (KNN) Dengan Optimasi	84.70%	71.59%
6	Support Vector Machine (SVM)	48.52%	47.73%

Dari tabel di atas maka dapat dilihat secara keseluruhan, K-Nearest Neighbors (KNN) tampak sebagai model yang paling konsisten dalam hal akurasi dan kemampuan identifikasi alasan pembatalan pada berbagai produk asuransi. KNN memiliki keunggulan dalam menangani data yang kompleks dan tidak linier. Akurasi yang dimiliki oleh KNN dengan optimasi menjadi akurasi tertinggi dibandingkan dengan model lain yaitu 84.70% untuk data training dan 71.59% untuk data testing.

Gaussian Naive Bayes, mudah diimplementasikan dan cepat dalam pelatihan, menunjukkan keterbatasan dalam menangani data dengan banyak fitur dan kategori ketika belum dioptimasi. Namun setelah optimasi, Naïve bayes menunjukan hasil akurasi yang mendekati akurasi dengan KNN. Akurasi yang didapat dengan optimasi adalah 72.49% untuk data training dan 71.49% untuk data testing.

Support Vector Machine (SVM) menunjukkan potensi dalam klasifikasi, terutama dalam kasus di mana data dapat dipisahkan dengan jelas. Namun, kinerja SVM yang beragam menunjukkan bahwa model ini tidak efektif untuk melakukan prediksi. Akurasi dengan optimasi adalah 48.52% untuk data training dan 47.73% untuk data testing.

ISSN: 2477-178x (Print)
ISSN: 2477-17xx (Online)

Hasil dari penelitian ini menunjukkan pentingnya pemilihan model yang tepat berdasarkan karakteristik data dan tujuan analisis. KNN dengan optimasi mungkin lebih cocok untuk data asuransi yang kompleks dengan interaksi fitur yang tidak linier, sementara Gaussian Naive Bayes dan SVM memerlukan penyesuaian tambahan untuk meningkatkan performa mereka.

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, berikut adalah kesimpulan yang dapat ditarik dari penelitian ini mengenai implementasi dan perbandingan algoritma dalam memprediksi keberlanjutan polis asuransi:

1. Penulis berhasil melakukan implementasi algoritma *Gaussian Naive Bayes*, *K-Nearest Neighbors* (KNN), dan *Support Vector Machine* (SVM) dilakukan untuk memprediksi keberlanjutan polis asuransi. Langkah pertama dalam implementasi melibatkan *preprocessing* data, termasuk pembersihan data, pengisian nilai yang hilang, dan *encoding* fitur kategorikal menjadi format numerik yang dapat diproses oleh algoritma. Data kemudian dibagi menjadi set pelatihan dan set pengujian untuk melatih dan mengevaluasi model. Selanjutnya, setiap algoritma diterapkan pada set pelatihan untuk membangun model klasifikasi yang dapat memprediksi apakah polis asuransi akan berlanjut atau dibatalkan. Model yang telah dilatih diuji menggunakan set pengujian untuk memprediksi status polis asuransi dan hasilnya digunakan untuk menilai akurasi dan efektivitas model dalam konteks prediksi pembatalan polis. Dengan informasi ini, perusahaan asuransi dapat mengidentifikasi nasabah yang berisiko membatalkan polis dan merancang strategi mitigasi yang lebih efektif, seperti intervensi awal atau penyesuaian kebijakan untuk mempertahankan nasabah.
2. Berdasarkan hasil evaluasi metrik performa dari ketiga algoritma *Gaussian Naive Bayes*, *K-Nearest Neighbors* (KNN), dan *Support Vector Machine* (SVM) dapat disimpulkan bahwa performa masing-masing algoritma bervariasi tergantung pada jenis produk asuransi. *K-Nearest Neighbors* (KNN) dengan optimasi menunjukkan hasil terbaik dengan akurasi 71.59% dan performa F1-*score* yang lebih baik dibandingkan *Gaussian Naive Bayes* (NB) dan *Support Vector Machine* (SVM). Secara keseluruhan, KNN terbukti sebagai algoritma yang paling efektif dalam memprediksi keberlanjutan polis asuransi di ketiga jenis produk, dengan konsistensi performa yang lebih tinggi dalam hal akurasi dan F1-*score.* Hal ini menunjukkan bahwa KNN lebih mampu menangani data dengan kompleksitas dan variabilitas tinggi dibandingkan *Gaussian Naive Bayes* dan SVM dalam konteks studi ini.

Berdasarkan temuan dan analisis yang dilakukan dalam penelitian ini, beberapa langkah tambahan dapat diambil untuk lebih meningkatkan efektivitas prediksi keberlanjutan polis asuransi:

1. Untuk penelitian selanjutnya, disarankan untuk memperluas dataset dengan menambahkan lebih banyak variabel yang dapat mempengaruhi keberlanjutan polis asuransi. Variabel tambahan ini bisa mencakup data perilaku nasabah, faktor-faktor eksternal seperti kondisi ekonomi atau demografi regional, serta informasi historis yang lebih rinci tentang interaksi nasabah dengan perusahaan asuransi. Dengan memperluas dataset, model dapat menangkap lebih banyak aspek dari perilaku nasabah dan potensi penyebab pembatalan polis, yang dapat menghasilkan prediksi yang lebih akurat dan informatif.
2. Selain itu, eksperimen dengan teknik *preprocessing* data yang lebih canggih dan penalaan parameter model yang lebih mendalam dapat dilakukan untuk meningkatkan performa algoritma dalam memprediksi keberlanjutan polis. Ini termasuk penerapan teknik normalisasi dan transformasi fitur yang lebih kompleks, serta eksplorasi berbagai metode untuk menangani ketidakseimbangan kelas. Penalaan parameter model, seperti pengaturan hyperparameter untuk algoritma *Naive Bayes*, KNN, dan SVM, juga penting untuk mengoptimalkan kinerja model dan memperoleh hasil yang lebih robust. Evaluasi yang lebih

menyeluruh dengan metrik tambahan dan metode validasi silang dapat membantu dalam menentukan konfigurasi model yang paling efektif.

DAFTAR PUSTAKA

giyani, G., & Sukma Wati, A. (2022). Perbandingan Menggunakan Metode Exponential Smoothing Untuk Prediksi Jumlah Polis Asuransi Kendaraan Pada PT X Kota Palembang. *Journal of Information Technology Ampera*, *3*(3), 2774–2121. https://journal-computing.org/index.php/journal-ita/index

Ajeng Mega Pratiwi, & Mutaqin, A. K. (2021). Penerapan Algoritma Naïve Bayes Classifier dalam Memprediksi Status Keberlanjutan Polis Nasabah Asuransi PT.X. *Jurnal Riset Statistika*, *1*(2), 117–126. https://doi.org/10.29313/jrs.v1i2.435

Alvina Felicia Watratan, Arwini Puspita, D. M. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran. *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST)*, *Vol. 1 No.*(1), 7–14. https://doi.org/10.55606/jurritek.v1i1.127

Breiman, L. (2001). *Random Forests. Machine Learning*.

Cahyanti, D., Rahmayani, A., & Husniar, S. A. (2020). Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. *Indonesian Journal of Data and Science, 1*(2), 39–43. https://doi.org/10.33096/ijodas.v1i2.13

Daft, R. L. (2003). *Manajement* (Buku 1 Edi). Salemba Empat.

Educba, "Seaborn heatmap." (2022). https://www.educba.com/seaborn-heatmap

Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, *Vol. 61*.

Goodfellow, I., B. (2016). *Deep Learning*. New York: Mit Press.

Harahap, S. Z., & Nastuti, A. (2019). Teknik Data Mining Untuk Penentuan Paket Hemat Sembako. *Jurnal Ilmiah Fakultas Sains Dan Teknologi*, *7*(3), 111–119.

Harianto, B. (2019). Rancang Bangun Aplikasi Klasifikasi Prediksi Underwriting Data Calon Pemegang Polis Pada Asuransi Jiwa Syariah Menggunakan Algoritma Naive Bayes (Study Kasus: (Ajb Bumiputera 1912). *UG Jurnal*, *13*(4), 56–73.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements Of Statistica. Learning: Data Mining, Inference, dan Prediction*. Springer Science & Business Media.

Hudori, H. (2020). Resampling Neural Network Untuk Penanganan Class Imbalance Pada Prediksi Klaim Asuransi. *Teknois : Jurnal Ilmiah Teknologi Informasi Dan Sains*, *10*(1), 57–64. https://doi.org/10.36350/jbs.v10i1.78

Hyndman, R. J., &Amp; Athanasopoulos, G. (2018). *Forecasting: Principles dan Practice*. New York: Otexts.

I Komang Setia Buana. (2020). Implementasi Aplikasi Speech to Text untuk Memudahkan Wartawan Mencatat Wawancara dengan Python. *Jurnal Sistem Dan Informatika (JSI)*, *14*(2), 135–142. https://doi.org/10.30864/jsi.v14i2.293

Johnson, R. A. (2019). *Applied Multivariate Statistical Analysis*. Cambridge: Pearson.

Ma'arif, A. (2020). Buku Ajar Pemrograman Lanjut Bahasa Pemrograman Python Oleh : Alfian Ma ' Arif. *Universitas Ahmad Dahlan*, 62. http://eprints.uad.ac.id/32743/1/buku python.pdf

Meiriyama, M., Devella, S., & Adelfi, S. M. (2022). Klasifikasi Daun Herbal Berdasarkan Fitur Bentuk dan Tekstur Menggunakan KNN. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(3), 2573–2584. https://doi.org/10.35957/jatisi.v9i3.2974

Mondal, Z. (2022). *Application of Corporate Governance indices to predict Bankruptcy of Companies using Machine Learning*.

Muhammad Romzi, & Kurniawan, B. (2020). Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma. *JTIM: Jurnal Teknik Informatika Mahakarya*, *03*(2), 37–44.

Muhathir, M., Santoso, M. H., & Larasati, D. A. (2021). Wayang Image Classification Using

SVM Method and GLCM Feature Extraction. *Journal of Informatics and Telecommunication Engineering*, *4*(2), 373–382. https://doi.org/10.31289/jite.v4i2.4524

Mutmainah, S. (2021). PENANGANAN IMBALANCE DATA PADA KLASIFIKASI KEMUNGKINAN PENYAKIT STROKE. *Jurnal SNATi*, *1(1)*, 10–17.

Nugroho, A., & Religia, Y. (2021). Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(3), 504–510. https://doi.org/10.29207/resti.v5i3.3067

Ordila, R., Wahyuni, R., Irawan, Y., & Yulia Sari, M. (2020). PENERAPAN DATA MINING UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT DENGAN ALGORITMA CLUSTERING (Studi Kasus : Poli Klinik PT.Inecda). *Jurnal Ilmu Komputer*, *9*(2), 148–153. https://doi.org/10.33060/jik/2020/vol9.iss2.181

Pratama, R., Herdiansyah, M. I., Syamsuar, D., & Syazili, A. (2023). Prediksi Customer Retention Perusahaan Asuransi Menggunakan Machine Learning. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *12*(1), 96–104. https://doi.org/10.32736/sisfokom.v12i1.1507

Putra, N. S., Hutabarat, B. F., & Khaira, U. (2023). Implementasi Algoritma Convolutional Neural Network Untuk Identifikasi Jenis Kelamin Dan Ras. *Jurnal Pendidikan Teknologi Informasi*, *3(1)*, 82–93.

Putry, N. M. (2022). Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus. *EVOLUSI : Jurnal Sains Dan Manajemen*, *10*(1). https://doi.org/10.31294/evolusi.v10i1.12514

R. Shaikh. (2018). *"Feature Selection Techniques in Machine Learning with Python,."* Https://Towardsdatascience.Com/Feature-Selection-.

Rahayu, A. S., Fauzi, A., & Rahmat, R. (2022). Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *4*(2), 349. https://doi.org/10.30865/json.v4i2.5398

Retno Utari, D., & Wibowo, A. (2020). Pemodelan Prediksi Status Keberlanjutan Polis Asuransi Kendaraan dengan Teknik Pemilihan Mayoritas Menggunakan Algoritma-Algoritma Klasifikasi Data Mining. *Prosiding Seminar Nasional Teknoka*, *5*(2502), 19–24. https://doi.org/10.22236/teknoka.v5i.391

V. N. Chawla, K. W. Bowyer, L. O. H. dan W. P. K. (2002). SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*