



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 2, No. 1b, Januari 2025

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 2, No. 1b, Januari 2025

Pages : 1853–1862

Pengaruh *Dimensionality Reduction* Menggunakan *Self-Organizing Maps* terhadap Klasifikasi *Naive Bayes* pada Data Kampanye Pemasaran

Abi Candra Arta Meviangga, Hasbi Firmansyah, Satria Ajis Saputra,
Mukhamad Falah

Prodi Informatika Fakultas Teknik dan Ilmu Komputer Universitas
Pancasakti Tegal

Article in Journal of MISTER

Available at : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

DOI : <https://doi.org/10.32672/mister.v2i1b.2765>

How to Cite this Article

APA : Arta Meviangga, A. C., Firmansyah, H., Saputra, S. A. ., & Falah, M. (2024). Pengaruh Dimensionality Reduction Menggunakan Self-Organizing Maps terhadap Klasifikasi Naive Bayes pada Data Kampanye Pemasaran. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 2(1b), 1853 – 1862. <https://doi.org/10.32672/mister.v2i1b.2765>

Others Visit : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



9 773032 710001 9 773032 601002

Pengaruh Dimensionality Reduction Menggunakan Self-Organizing Maps terhadap Klasifikasi Naive Bayes pada Data Kampanye Pemasaran

Abi Candra Arta Meviangga¹, Hasbi Firmansyah², Satria Ajis Saputra³, Mukhamad Falah⁴

Prodi Informatika Fakultas Teknik dan Ilmu Komputer Universitas Pancasakti Tegal

*Email: 6623600019@upstegal.ac.id ; hasbifirmansyah@upstegal.ac.id ; 6623600001@upstegal.ac.id ; 6623600020@upstegal.ac.id

Diterima: 27-12-2024

| Disetujui: 28-12-2024

| Diterbitkan: 29-12-2024

ABSTRACT

This study aims to evaluate the performance of the Naive Bayes algorithm in predicting the success of a bank marketing campaign, both with and without dimensionality reduction using Self-Organizing Maps (SOM). The dataset used originates from the UCI Machine Learning Repository, comprising the target attribute "y" and predictor attributes such as age, occupation, education, and credit default status. The research process includes data cleaning, attribute transformation, and normalization, followed by two main approaches: (1) direct application of Naive Bayes without SOM and (2) using SOM for dimensionality reduction before applying Naive Bayes. The results indicate that the model without SOM achieved an accuracy of 80.95% and an AUC of 0.850. However, the precision for predicting the "yes" class was only 34.63%. In contrast, the approach with SOM yielded higher precision (63.24%) for the "yes" class, though the accuracy and recall decreased to 62.10% and 57.80%, respectively, with an AUC of 0.667. Further analysis revealed a trade-off between overall accuracy and the precision of positive predictions, depending on the campaign's objectives. Visualization using SOM scatter plots helped uncover complex data distribution patterns and potential overlaps between classes. This study provides insights into the benefits and limitations of dimensionality reduction using SOM in enhancing the efficiency of probabilistic classification models.

Keywords: Dimensionality Reduction; Self-Organizing Maps; Naive Bayes

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi performa algoritma Naive Bayes dalam memprediksi keberhasilan kampanye pemasaran bank, baik dengan maupun tanpa pengurangan dimensi menggunakan Self-Organizing Maps (SOM). Dataset yang digunakan berasal dari UCI Machine Learning Repository, mencakup atribut target "y" dan atribut prediktor seperti usia, pekerjaan, pendidikan, serta status kredit macet. Proses penelitian mencakup pembersihan data, transformasi atribut, dan normalisasi, diikuti dengan dua pendekatan utama: (1) penerapan langsung Naive Bayes tanpa SOM dan (2) penggunaan SOM untuk mereduksi dimensi sebelum penerapan Naive Bayes. Hasil menunjukkan bahwa model tanpa SOM mencapai akurasi 80,95%, dengan nilai AUC 0,850. Namun, precision untuk prediksi kelas "yes" hanya sebesar 34,63%. Sebaliknya, pendekatan dengan SOM menghasilkan precision yang lebih tinggi (63,24%) untuk kelas "yes", tetapi akurasi dan recall menurun menjadi masing-masing 62,10% dan 57,80%, dengan AUC 0,667. Analisis lebih lanjut mengidentifikasi adanya trade-off antara akurasi keseluruhan dan ketepatan prediksi positif, yang bergantung pada tujuan kampanye. Visualisasi menggunakan scatter plot SOM membantu mengungkap pola distribusi data yang kompleks dan potensi tumpang tindih antar kelas. Penelitian ini memberikan wawasan tentang manfaat dan keterbatasan reduksi dimensi menggunakan SOM dalam meningkatkan efisiensi model klasifikasi berbasis probabilistik.

Katakunci: Dimensionality Reduction; Self-Organizing Maps; Naive Bayes

PENDAHULUAN

Kemajuan teknologi informasi dan ketersediaan data yang semakin besar telah memberikan peluang bagi organisasi untuk menggunakan pendekatan berbasis data dalam pengambilan keputusan. Salah satu tantangan utama adalah bagaimana memanfaatkan data tersebut untuk mendukung analisis prediktif yang akurat, terutama dalam konteks yang kompleks seperti kampanye pemasaran bank. Analisis data mining telah menjadi solusi yang efektif untuk menemukan pola tersembunyi dalam data, yang dapat digunakan untuk memahami faktor-faktor yang memengaruhi keberhasilan kampanye pemasaran.

Dalam konteks ini, teknik klasifikasi seperti Naive Bayes sering digunakan karena efisiensinya dalam menangani data yang kompleks. Algoritma ini menawarkan pendekatan probabilistik yang dapat menghubungkan atribut prediktor dengan atribut target untuk memprediksi keberhasilan kampanye. Namun, keakuratan model klasifikasi dapat dipengaruhi oleh dimensi data yang tinggi dan kehadiran atribut yang kurang relevan, sehingga diperlukan teknik tambahan untuk meningkatkan efisiensi model.

Self-Organizing Maps (SOM) adalah salah satu metode klastering berbasis jaringan saraf tiruan yang mampu mereduksi dimensi data dengan memetakan pola kompleks ke dalam ruang dua dimensi. Teknik ini tidak hanya menyederhanakan analisis, tetapi juga membantu dalam visualisasi distribusi data yang lebih intuitif. Kombinasi SOM dengan Naive Bayes menawarkan pendekatan yang menjanjikan untuk meningkatkan akurasi prediksi, terutama dalam konteks kampanye pemasaran yang memiliki distribusi kelas tidak seimbang.

Penelitian ini bertujuan untuk mengevaluasi performa algoritma Naive Bayes dalam memprediksi keberhasilan kampanye pemasaran bank, baik dengan maupun tanpa reduksi dimensi menggunakan SOM. Dengan menggunakan dataset dari UCI Machine Learning Repository, penelitian ini berfokus pada dua pendekatan utama:

- (1) Penerapan langsung Naive Bayes tanpa reduksi dimensi(2)
- (2) Kombinasi SOM untuk mereduksi dimensi sebelum menerapkan Naive Bayes.

Studi ini tidak hanya mengukur keakuratan dan ketepatan model, tetapi juga menganalisis trade-off antara berbagai metrik performa, seperti precision, recall, dan Area Under Curve (AUC), untuk memberikan wawasan tentang manfaat dan keterbatasan reduksi dimensi dalam analisis prediktif.

METODOLOGI PENELITIAN

Data Mining

Data mining adalah proses penggalian informasi atau pola berharga dari kumpulan data besar melalui penggunaan teknik matematika, statistik, dan pembelajaran mesin. Proses ini bertujuan untuk menemukan pola tersembunyi yang dapat membantu pengambilan keputusan dan memberikan wawasan baru dari data yang kompleks (Simanjuntak et al., 2022; Wati et al., 2024).

Data mining adalah inti dari penelitian ini, yang bertujuan untuk menggali pola tersembunyi dalam dataset kampanye pemasaran bank. Teknik data mining memungkinkan eksplorasi data secara sistematis untuk memahami hubungan antar atribut yang memengaruhi keberhasilan kampanye. Dalam penelitian ini, data mining diterapkan untuk membangun model prediksi menggunakan algoritma Naive Bayes, dengan dan tanpa pengurangan dimensi menggunakan Self-Organizing Maps (SOM).

Teknik Klastering

Teknik klastering adalah metode pengelompokan data berdasarkan kemiripan karakteristik, sehingga data yang mirip berada dalam satu kelompok (*cluster*), sementara data yang berbeda dikelompokkan secara terpisah. Algoritma seperti K-Means memanfaatkan nilai *centroid* untuk menentukan pengelompokan, sedangkan Self-Organizing Maps (SOM) menggunakan jaringan saraf tiruan untuk memetakan data melalui mekanisme pengaturan diri (Anjelita et al., 2020; Wati et al., 2024).

Teknik klastering digunakan untuk mengelompokkan data berdasarkan kesamaan pola. Dalam penelitian ini, SOM digunakan untuk mereduksi dimensi dataset yang semula memiliki banyak atribut menjadi hanya dua atribut utama. Reduksi ini bertujuan untuk menyederhanakan analisis data dan mengurangi noise, sehingga model dapat fokus pada pola penting yang relevan dengan prediksi.

Teknik Klasifikasi

Teknik klasifikasi adalah metode dalam *data mining* yang bertujuan untuk mengelompokkan data ke dalam kelas tertentu berdasarkan label yang telah ada sebelumnya. Salah satu metode yang populer adalah Naive Bayes, yang memanfaatkan teorema Bayes dan pendekatan probabilitas untuk memprediksi kelas data dengan asumsi bahwa setiap fitur bersifat independen (Simanjuntak et al., 2022; Anjelita et al., 2020).

Teknik klasifikasi digunakan untuk memetakan hubungan antara atribut prediktor dan atribut target. Naive Bayes, sebagai algoritma berbasis probabilistik, dipilih karena kemampuannya yang efisien dalam menangani data besar dengan struktur yang kompleks. Klasifikasi ini membantu dalam memahami bagaimana setiap atribut memengaruhi keberhasilan kampanye.

Self-Organizing Maps (SOM)

Self-Organizing Maps (SOM) adalah teknik berbasis jaringan saraf tiruan yang digunakan untuk klastering data. SOM bekerja dengan memetakan data ke ruang dua dimensi menggunakan mekanisme pengaturan diri, di mana neuron dengan jarak terdekat terhadap data akan diperbarui untuk mencerminkan karakteristik data tersebut. Keunggulan SOM terletak pada stabilitas klastering dan kemampuannya untuk menghasilkan nilai *centroid* yang konsisten (Wati et al., 2024).

Dalam penelitian ini, SOM digunakan untuk memproyeksikan data ke ruang dua dimensi, yang mempermudah visualisasi dan analisis pola. Keunggulan metode ini terlihat dari kemampuannya mengidentifikasi pola distribusi yang kompleks, seperti yang ditunjukkan oleh Annas et al. (2022) dalam studi mereka mengenai distribusi polusi udara di Kota Makassar. Penelitian tersebut menyoroti bagaimana SOM dapat mendukung interpretasi hasil melalui visualisasi intuitif yang terintegrasi dengan sistem informasi geografis (GIS), menjadikannya alat yang efektif dalam analisis data skala besar.

Naive Bayes

Naive Bayes adalah metode klasifikasi yang menggunakan teorema Bayes untuk memprediksi probabilitas suatu data masuk ke dalam kelas tertentu. Metode ini mengasumsikan bahwa fitur-fitur dalam data bersifat independen satu sama lain. Naive Bayes sering digunakan untuk tugas-tugas klasifikasi karena sifatnya yang sederhana dan efisien dalam menangani data besar (Simanjuntak et al., 2022).

Naive Bayes dipilih dalam penelitian ini karena kesederhanaan dan efisiensinya dalam membangun model prediksi, terutama untuk dataset berskala besar. Algoritma ini mengasumsikan independensi antar atribut secara kondisional terhadap kelas target, yang membuatnya sangat cepat dan ringan. Dalam

penelitian ini, Naive Bayes diterapkan dalam dua skenario: langsung pada dataset asli tanpa reduksi dimensi, dan setelah reduksi dimensi menggunakan SOM.

Sebagai algoritma berbasis probabilistik, Naive Bayes memiliki berbagai aplikasi lintas domain. Sebagai contoh, Watratan et al. (2020) menunjukkan efektivitasnya dalam memprediksi tingkat penyebaran COVID-19. Studi tersebut memanfaatkan Naive Bayes untuk menganalisis data kuantitatif dan menghasilkan prediksi yang efisien, meskipun dengan asumsi independensi yang kuat antar variabel. Hal ini menunjukkan potensi algoritma dalam memberikan hasil yang andal di berbagai konteks aplikasi.

Dataset

Tabel 1 ExampleSet dari Dataset

No	Y	Age	Job	Marital	Education	Default	...	pOutcome
1	no	58	management	married	tertiary	no	...	unknown
2	no	44	technician	single	secondary	no	...	unknown
3	no	33	entrepreneur	married	secondary	no	...	unknown
4	no	47	blue-collar	married	unknown	no	...	unknown
5	no	33	unknown	single	unknown	no	...	unknown
6	no	35	management	married	tertiary	no	...	unknown
7	no	28	management	single	tertiary	no	...	unknown
8	no	42	entrepreneur	divorced	tertiary	yes	...	unknown
9	no	58	retired	married	primary	no	...	unknown
10	no	43	technician	single	secondary	no	...	unknown
...	...	...	...	...	...	...	...	...
45211	no	37	entrepreneur	married	secondary	no	...	other

Dataset yang digunakan berasal dari UCI Machine Learning Repository dan berisi informasi tentang kampanye pemasaran bank di Portugal. Dataset ini memiliki atribut target "y" yang menyatakan keberhasilan kampanye pemasaran dengan nilai "yes" atau "no". Atribut prediktor lainnya meliputi:

- **Age (umur):** Usia responden.
- **Job (pekerjaan):** Profesi responden.
- **Marital (status perkawinan):** Status pernikahan responden.
- **Education (pendidikan):** Tingkat pendidikan terakhir responden.
- **Default (status kredit macet):** Indikator apakah responden memiliki kredit macet.
- **Balance (saldo):** Jumlah saldo rekening nasabah.
- **Housing (pinjaman perumahan):** Kepemilikan pinjaman rumah ("yes" atau "no").
- **Loan (pinjaman pribadi):** Kepemilikan pinjaman pribadi ("yes" atau "no").
- **Contact (jenis kontak):** Jenis kontak terakhir dengan nasabah ("cellular" atau "telephone").
- **Day (hari kontak):** Hari dalam sebulan ketika nasabah dihubungi.
- **Month (bulan kontak):** Bulan ketika nasabah dihubungi.
- **Duration (durasi):** Lama waktu kontak berlangsung (dalam detik).
- **Campaign (jumlah kontak):** Jumlah total kontak selama kampanye saat ini.

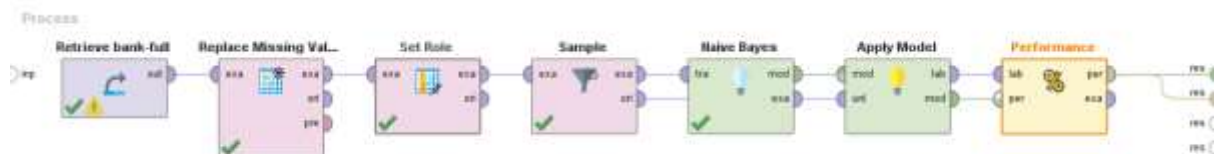
- **Pdays (hari sejak kontak terakhir):** Jumlah hari sejak terakhir kali nasabah dihubungi pada kampanye sebelumnya (999 jika nasabah belum pernah dihubungi sebelumnya).
- **Previous (jumlah kontak sebelumnya):** Jumlah total kontak pada kampanye sebelumnya.
- **Poutcome (hasil kampanye sebelumnya):** Hasil dari kampanye pemasaran sebelumnya ("success", "failure", "other", atau "unknown").

Sebelum analisis, dataset melalui proses persiapan, termasuk pembersihan data, transformasi atribut nominal ke numerik, dan normalisasi.

Tahapan Penelitian

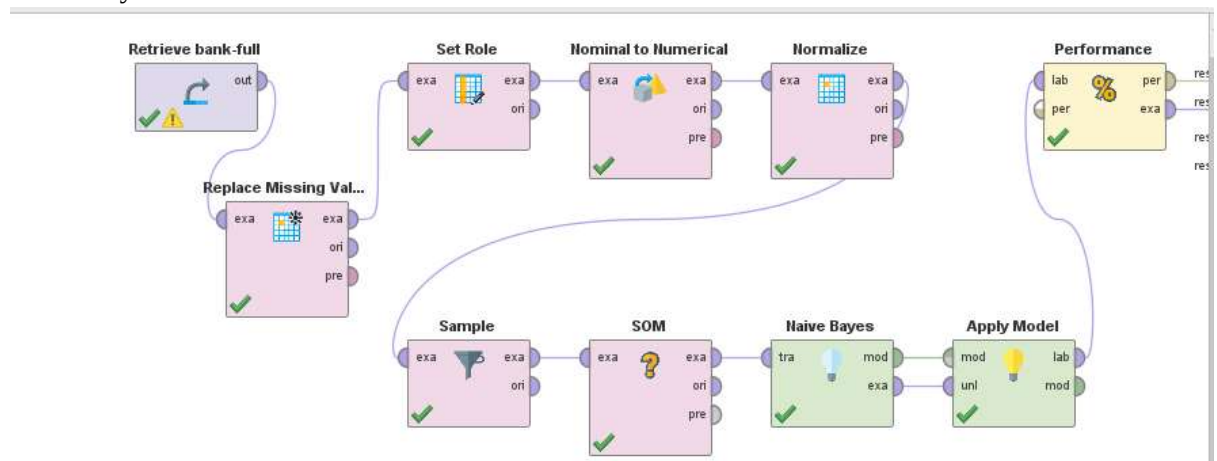
Penelitian ini terdiri dari beberapa tahapan:

1. **Persiapan Data:** Meliputi pembersihan data, transformasi atribut, dan normalisasi dalam rentang 0 dan 1.



Gambar 1. Jalur tanpa SOM

2. **Jalur Tanpa SOM:** Dataset langsung digunakan untuk membangun model prediksi menggunakan Naive Bayes.



Gambar 2. Jalur dengan SOM

3. **Jalur Dengan SOM:** Dataset direduksi dimensinya menggunakan SOM sebelum diaplikasikan pada Naive Bayes. SOM diimplementasikan dengan:
 - Net size: default ($5 * \sqrt{\text{training sample number}}$)
 - Training round: 8000
 - Initial learning rate: 0.3
 - Learning rate function: linear

- Menggunakan toroid network
4. **Evaluasi Kinerja:** Hasil dari kedua jalur dievaluasi berdasarkan metrik akurasi, precision, recall, dan AUC.

HASIL DAN PEMBAHASAN

Hasil Tanpa SOM

Pada pengujian pertama, Naive Bayes diaplikasikan langsung pada dataset yang telah melalui tahap persiapan data tanpa reduksi dimensi menggunakan SOM. Hasil menunjukkan bahwa model memiliki tingkat akurasi sebesar 80.95%. Berikut rincian hasil yang diperoleh:

1. Tabel Performa

Tabel 2 Performa Tanpa SOM

	true no	true yes	class precision
Prediksi "no"	32,854	1,544	95.51%
Prediksi "yes"	7,068	3,745	34.63%
class recall	82.30%	70.81%	

2. Confusion Matrix

- **True Negative (no/no):** 32,854 data berhasil diklasifikasikan dengan benar sebagai "no".
- **False Positive (no/yes):** 1,544 data salah diklasifikasikan sebagai "yes".
- **False Negative (yes/no):** 7,068 data salah diklasifikasikan sebagai "no".
- **True Positive (yes/yes):** 3,745 data berhasil diklasifikasikan dengan benar sebagai "yes".

3. Precision, Recall, dan AUC

- Precision untuk kelas "yes" adalah 34.63%, menunjukkan bahwa dari semua prediksi "yes", hanya 34.63% yang benar-benar positif.
- Recall untuk kelas "yes" mencapai 70.81%, menunjukkan sensitivitas model terhadap data "yes".
- Nilai AUC adalah 0.850, menunjukkan kemampuan model untuk membedakan antara kelas "yes" dan "no".

4. **Analisis** Meskipun recall dan AUC cukup tinggi, precision rendah menunjukkan banyak prediksi positif palsu. Hal ini kemungkinan disebabkan oleh ketidakseimbangan distribusi data antara kelas "yes" dan "no".

Hasil Dengan SOM

Setelah reduksi dimensi menggunakan SOM, dataset yang semula memiliki banyak atribut menjadi dua atribut utama. Hasil pengujian menunjukkan bahwa:

1. Tabel Peforma

Tabel 3 Peforma dengan SOM

	true no	true yes	class precision
prediksi no	664	422	61.14%
prediksi yes	336	578	63.24%
class recall	66.40%	57.80%	

Confusion Matrix

- **True Negative (no/no):** 664 data berhasil diklasifikasikan dengan benar sebagai "no".
- **False Positive (no/yes):** 422 data salah diklasifikasikan sebagai "yes".
- **False Negative (yes/no):** 336 data salah diklasifikasikan sebagai "no".
- **True Positive (yes/yes):** 578 data berhasil diklasifikasikan dengan benar sebagai "yes".

2. Precision, Recall, dan AUC

- Precision untuk kelas "yes" meningkat menjadi 63.24%.
- Recall untuk kelas "yes" menurun menjadi 57.80%.
- Nilai AUC rata-rata adalah 0.667, lebih rendah dibandingkan tanpa SOM.

3. **Analisis** Meskipun precision meningkat, akurasi dan recall menurun. Hal ini menunjukkan bahwa reduksi dimensi membantu meningkatkan ketepatan prediksi positif tetapi mengurangi sensitivitas model terhadap data "yes".

Analisis Perbandingan

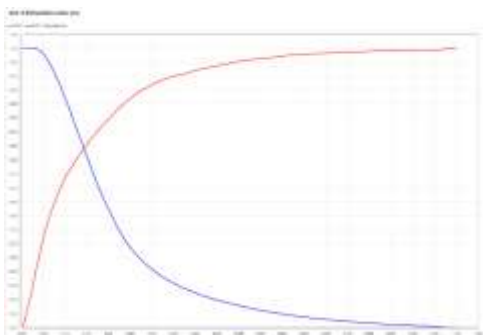
Tabel 4 Analisis Perbandingan

Peforma	Akurasi	Precision	Recall	AUC
Dengan SOM	62.10%	63.24%	57.80%	0.667
Tanpa SOM	80.95%	34.63%	70.81%	0.85

Trade-off antara kedua pendekatan terlihat jelas. Pendekatan tanpa SOM lebih sesuai untuk menemukan target sebanyak mungkin (prioritas recall), sementara pendekatan dengan SOM lebih baik untuk mengurangi prediksi positif palsu (prioritas precision).

Visualisasi Hasil

Kurva ROC



Gambar 3 Kurva ROC tanpa SOM

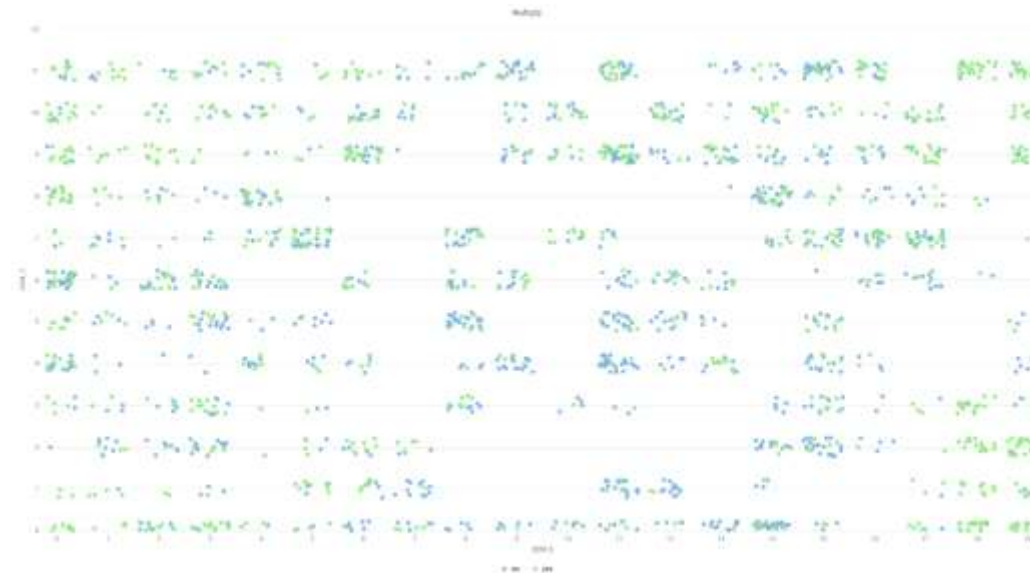


Gambar 4 Kurva ROC dengan SOM

Kurva Receiver Operating Characteristic (ROC) digunakan untuk mengevaluasi performa model klasifikasi. Area Under Curve (AUC) tanpa SOM memiliki nilai yang lebih tinggi (0.850) dibandingkan dengan setelah menggunakan SOM (0.667). Hal ini menunjukkan bahwa model tanpa SOM memiliki kemampuan yang lebih baik dalam membedakan antara kelas "yes" dan "no".

Kurva ROC memberikan gambaran tentang trade-off antara tingkat positif benar (true positive rate) dan tingkat positif palsu (false positive rate), yang menunjukkan efisiensi model untuk berbagai threshold prediksi.

Scatter Plot Hasil SOM



Scatter plot yang dihasilkan dari proses Self-Organizing Maps (SOM) memberikan representasi dua dimensi dari dataset yang direduksi. Pada scatter plot ini:

- Titik hijau mewakili data dengan target "yes".
- Titik biru mewakili data dengan target "no".

Visualisasi scatter plot menunjukkan pola distribusi data berdasarkan hasil klustering SOM. Berdasarkan scatter plot, terlihat bahwa beberapa kluster cenderung mendominasi salah satu kelas tertentu (baik "yes" atau "no"). Namun, terdapat pula tumpang tindih antar kelas di beberapa kluster, yang dapat menjelaskan penurunan akurasi dan recall setelah penggunaan SOM.

Scatter plot ini mengilustrasikan bahwa meskipun reduksi dimensi membantu dalam visualisasi dan pengelompokan data, trade-off antara kesederhanaan dan kehilangan informasi dapat memengaruhi performa model.

KESIMPULAN

Penelitian ini mengeksplorasi prediksi keberhasilan kampanye pemasaran menggunakan dua pendekatan: tanpa reduksi dimensi dengan langsung menerapkan algoritma Naive Bayes dan dengan reduksi dimensi menggunakan Self-Organizing Maps (SOM). Pada pendekatan tanpa SOM, model mencapai akurasi 80.95%, precision 34.63% untuk kelas "yes", recall 70.81%, dan AUC 0.850, menunjukkan sensitivitas tinggi terhadap keberhasilan kampanye meskipun precision rendah akibat distribusi data yang tidak seimbang. Sementara itu, pendekatan dengan SOM menghasilkan akurasi 62.10%, precision 63.24%, recall 57.80%, dan AUC 0.667, yang menunjukkan peningkatan precision tetapi dengan penurunan sensitivitas dan akurasi secara keseluruhan. Visualisasi menggunakan scatter plot

menunjukkan bahwa SOM berhasil mengelompokkan data ke dalam kluster yang merepresentasikan pola tertentu, meskipun terdapat tumpang tindih antar kelas yang memengaruhi performa model. Berdasarkan hasil ini, pendekatan tanpa SOM lebih efektif untuk mendeteksi sebanyak mungkin keberhasilan kampanye (prioritas pada recall), sementara pendekatan dengan SOM lebih cocok untuk skenario yang memprioritaskan ketepatan prediksi (precision). Untuk peningkatan lebih lanjut, disarankan untuk mengatasi ketidakseimbangan data dengan teknik seperti oversampling atau algoritma yang lebih adaptif terhadap distribusi data yang tidak seimbang.

DAFTAR REFERENSI

- Anjelita, M., Windarto, A. P., Wanto, A., & Sudahri, I. "Pengembangan Datamining Klastering Pada Kasus Pencemaran Lingkungan Hidup." *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2020.
- Annas, S., Uca, I., Safei, R. H., & Rais, Z. (2022). Using k-Means and Self-Organizing Maps in Clustering Air Pollution Distribution in Makassar City, Indonesia. *Jambura Journal of Mathematics*, 4(1), 167–17
- Simanjuntak, A. Y., Simatupang, I. S., & Anita. "Implementasi Data Mining Menggunakan Metode Naive Bayes Classifier untuk Data Kenaikan Pangkat Dinas Ketenagakerjaan Kota Medan." *Journal of Science and Social Research*, 2022.
- Wati, R. K., Pratiwi, H., & Sulandari, W. "Perbandingan Algoritma Density-Based Spatial Clustering Algorithm with Noise (DBSCAN) dan Self-Organizing Map (SOM) untuk Clustering Data Gempa Bumi." *Statistika*, Vol. 24, No. 2, 2024.
- Watratana, A. F., Puspita, A. B., & Moeis, D. (2020). Implementasi Algoritma Naive Bayes untuk Memprediksi Tingkat Penyebaran COVID-19 di Indonesia. *Journal of Applied Computer Science and Technology (JACOST)*, 1(1), 7–14.