



Analisis Churn Nasabah Bank Dengan Pendekatan *Machine Learning* dan Pengelompokan Profil Nasabah dengan Pendekatan *Clustering*

Arief Sulistyو Wibowo

Program Studi Teknik Industri, Fakultas Teknik,
Universitas Pembangunan Nasional “Veteran” Jawa Timur

Email: 21032010074@student.upnjatim.ac.id

Rusindiyanto

Program Studi Teknik Industri, Fakultas Teknik,
Universitas Pembangunan Nasional “Veteran” Jawa Timur
Jl. Rungkut Madya Surabaya 60294

Email: rusindiyanto.ti@gmail.com

Korespondensi penulis: : 21032010074@student.upnjatim.ac.id

Abstract. Rapid technological developments encourage the banking sector to continue to innovate so as not to be left behind. Tight competition in this industry is caused by customers' freedom to choose products and services that are considered more profitable. This phenomenon is known as Customer Churn, which is a condition where customers choose not to continue subscribing to a particular company. The method applied uses a machine learning approach and customer segmentation approach. The churn analysis results show that the machine learning model, especially the random forest model, has the highest level of accuracy with an F1-Score of 91%. This model has the potential to reduce churn rates from 20.4% to 5.61%, illustrating its positive impact. Apart from that, for the clustering results, the K-Prototype model was obtained for the clustering model with the highest Silhouette Score number of 0.1557 and 4 clusters were obtained.

Keywords: Bank, Churn, Cluster, Machine Learning

Abstrak. Perkembangan teknologi yang pesat mendorong sektor perbankan untuk terus berinovasi agar tidak tertinggal. Persaingan yang ketat dalam industri ini disebabkan oleh kebebasan nasabah dalam memilih produk dan layanan yang dianggap lebih menguntungkan. Fenomena ini dikenal sebagai Customer Churn, yaitu kondisi di mana pelanggan atau nasabah memilih untuk tidak melanjutkan berlangganan pada perusahaan tertentu. Metode yang diterapkan menggunakan pendekatan machine learning dan segmentasi nasabah dengan pendekatan. Hasil analisis churn menunjukkan bahwa model machine learning, khususnya model random forest, memiliki tingkat akurasi tertinggi dengan nilai F1-Score sebesar 91%. Model ini berpotensi untuk menurunkan churn rate dari 20,4% menjadi 5,61%, menggambarkan dampak positifnya. Selain itu, untuk hasil clustering diperoleh model K-Prototype untuk model clustering dengan angka Silhouette Score tertinggi sebesar 0.1557 dan didapatkan 4 cluster.

Kata kunci: Bank, Churn, Cluster, Machine Learning

LATAR BELAKANG

Perkembangan teknologi yang pesat pada revolusi industri 4.0 yang kian mempengaruhi pola hidup masyarakat, turut menuntut bidang perbankan berinovasi agar tidak tertinggal. Terlebih lagi dunia perbankan dihadapkan dengan hadirnya pesaing dalam bidang finansial lain seperti industri keberhasilan perusahaan. Perusahaan menyadari bahwa strategi komunikasi pemasaran satu arah bagi semua orang sudah tidak mencapai target yang diinginkan

Received Desember 25, 2023; Accepted Januari 05, 2024; Published Januari 31, 2023

* Arief Sulistyو Wibowo, 21032010074@student.upnjatim.ac.id

perusahaan. Ekspektasi konsumen yang tinggi membuat perusahaan harus merencanakan strategi baru. *Financial dan Technology (fintech)* yang semakin. Persaingan dalam bidang ini semakin ketat karena. Dampak dari persaingan tersebut yaitu kemungkinan nasabah berpindah ke bank maupun fintech yang lain. Dalam dunia bisnis hal ini dikenal dengan *Customer Churn*. *Churn* merupakan kecenderungan pelanggan untuk berhenti menggunakan produk atau layanan dari suatu perusahaan kemudian berpindah ke perusahaan lain (Irmanda, Astriratma, and Afrizal 2019).

Churn nasabah terjadi ketika nasabah memutuskan untuk tidak menggunakan sebagai pengguna Bank tersebut lagi. Fenomena ini dikenal dengan istilah *churn* karena nasabah memutuskan untuk mengakhiri hubungan dengan Bank tertentu dan mencari pelayanan di Bank lain. Hal ini menunjukkan bahwa nasabah merasa bahwa lebih mendapatkan keuntungan menggunakan Bank yang lain. Masalah *customer churn* ini menjadi krusial, karena biaya yang dikeluarkan untuk mendapatkan pelanggan baru, untuk iklan, marketing, komisi, dan lain-lain akan jauh lebih besar dibandingkan biaya yang harus dikeluarkan untuk menjaga pelanggan yang sudah ada. Ditambah lagi belum kebanyakan pelanggan baru cenderung tidak lebih menghasilkan keuntungan dibandingkan pelanggan yang sudah lama dan bertahan. Sehingga mempertahankan pelanggan yang sudah ada merupakan prioritas utama (Mulia et al. 2023).

Dengan demikian, penelitian ini dilakukan untuk mengurangi tingkat kehilangan nasabah di institusi perbankan dengan menerapkan model *machine learning*. Selanjutnya, penelitian ini juga melibatkan segmentasi nasabah melalui analisis *clustering*, yang nantinya akan digunakan sebagai dasar untuk memberikan saran dan rekomendasi bisnis berdasarkan temuan dari data. Hasil penelitian ini diharapkan dapat menjadi pedoman bagi perusahaan perbankan dalam merancang kebijakan terkait keberlanjutan dan loyalitas nasabah, serta meningkatkan kualitas layanan yang diberikan kepada mereka

KAJIAN TEORITIS

A. *Churn* Nasabah

Customer churn merupakan kondisi dimana pelanggan tidak melanjutkan berlangganan pada perusahaan yang dipilih dan berpindah pada perusahaan pesaing. *Customer churn* merupakan istilah yang digunakan untuk mendefinisikan bahwa seorang *customer* (pelanggan) menghentikan hubungan bisnis dengan perusahaan (Arifin, 2018). *Customer churn* atau kehilangan pelanggan merupakan istilah yang digunakan dalam bisnis untuk menyebut

kehilangan pelanggan atau terputusnya hubungan antara customer dengan pemilik bisnis. Dalam industri teknologi informasi, *customer churn* mengacu kepada customer yang meninggalkan bisnis untuk berpindah kepada pesaing bisnis (Yulianto, 2021).

B. *Machine Learning*

Machine learning adalah metode analisis statistik yang efisien untuk menangkap hubungan internal linear atau nonlinier dengan belajar dari data empiris. Kompleksitas masalah teknik telah mendorong peningkatan penerapan *machine learning*. Algoritma matematika mampu belajar cepat dari pola yang diperkenalkan sebelumnya, dan teknik ini dapat berhasil memperoleh hubungan timbal balik yang kompleks antara beberapa parameter dan dengan cepat memprediksi output yang diinginkan (Leni et al. 2023). *Machine Learning* atau Mesin Pembelajaran adalah cabang dari AI yang berfokus pada pembelajaran dari data. Ini menekankan pengembangan sistem yang mampu belajar secara "mandiri" tanpa harus diulang-ulang diprogram oleh manusia. Jika *Machine Learning* diibaratkan sebagai kendaraan bermotor, maka data menjadi bahan bakar utamanya. Hal ini karena *Machine Learning* membutuhkan data untuk menciptakan metode penyelesaian masalah (Alfarizi et al. 2023).

C. *Clustering*

Clustering mengacu pada pengelompokan seperti *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan dari record yang memiliki kemiripan satu sama lain, dan berbeda dengan *record* di klaster lain. *Clustering* mencoba untuk membagi seluruh kumpulan data menjadi kelompok-kelompok yang relatif memiliki kemiripan, di mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal (Nabila, Rahman Isnain, and Abidin 2021). *Clustering* merupakan teknik dalam data mining yang bertujuan untuk mengelompokkan data-data (objek) ke dalam beberapa *cluster* atau kelompok sehingga objek dengan pola serupa disatukan kedalam *cluster* yang sama, sedangkan objek dengan pola berbeda harus menjadi bagian dari cluster yang berbeda. Ada dua jenis data *clustering* yang sering dipergunakan dalam proses pengelompokan data, yaitu hierarchial data *clustering* dan *non-hierarchical data clustering* (Susanto 2022).

METODE PENELITIAN

Bahan yang digunakan pada analisis *churn* yaitu berupa kumpulan data *churn* banking yang berasal dari situs Kaggle terdiri dari 10.000 baris dan 14 kolom meliputi *RowNumber*, *CustomerId*, *Surname*, *CreditScore*, *Geography*, *Gender*, *Age*, *Tenure*, *Balance*, *NumOfProducts*, *HasCrCard*, *IsActiveMember*, *EstimatedSalary* dan *Exited*.

RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
1	15634602	Hargrave	619	France	Female	42	2	0	1	1	1	101348.88	1
2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
3	15619304	Onio	502	France	Female	42	8	159660.8	3	1	0	113931.57	1
4	15701354	Boni	699	France	Female	39	1	0	2	0	0	93826.63	0
5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.1	0
6	15574012	Chu	645	Spain	Male	44	8	113755.78	2	1	0	149756.71	1
7	15592531	Bartlett	822	France	Male	50	7	0	2	1	1	10062.8	0
8	15656148	Obinna	376	Germany	Female	29	4	115046.74	4	1	0	119346.88	1
9	15792365	He	501	France	Male	44	4	142051.07	2	0	1	74940.5	0
10	15592389	Ilip	684	France	Male	27	2	134603.88	1	1	1	71725.73	0
11	15767821	Beauce	528	France	Male	31	6	102016.72	2	0	0	80181.12	0
12	15737173	Andrews	497	Spain	Male	24	3	0	2	1	0	76390.01	0
13	15632264	Kay	476	France	Female	34	10	0	2	1	0	26260.98	0
14	15691483	Chiu	549	France	Female	25	5	0	2	0	0	190857.79	0
15	15600882	Scott	635	Spain	Female	35	7	0	2	1	1	65951.65	0
16	15643966	Goforth	616	Germany	Male	45	3	143129.41	2	0	1	64327.26	0
17	15737452	Romes	653	Germany	Male	58	1	132602.88	1	1	0	5097.67	1
18	15788218	Henderson	549	Spain	Female	24	9	0	2	1	1	14406.41	0
19	15661507	Muldrov	587	Spain	Male	45	6	0	1	0	0	158684.81	0
20	15568982	Ilao	726	France	Female	24	6	0	2	1	1	54724.03	0

Gambar 1. Dataset churn modelling

Metode yang digunakan dalam menemukan dan memprediksi *churn* nasabah pada dataset terbagi menjadi 5 tahapan



Gambar 2. Tahapan pengerjaan

Background digunakan untuk memahami data dan dapat menentukan *problem statement*, *goal*, *objective* dan *bussines metric*. *Data Pre-processing* berguna untuk membersihkan data dengan tujuan untuk membuat data semakin sederhana dan memastikan *dataset* tidak ada yang duplikat, data yang hilang, *outlier* dan menentukan fitur pada data yang memiliki penting yang berperan untuk keperluan analisis dan prediksi. *Data Insight* bertujuan untuk memahami interpretasi atau informasi penting yang dapat diambil dari *dataset*. Selanjutnya *Modelling dan Experiments*, pada tahap ini akan dibuat algoritma klasifikasi dan juga klustering, dengan

melakukan *trial error* pada beberapa algoritma yang dirasa sesuai dengan *dataset*. Lalu *Bussines Recommendation* Pada tahap ini berdasarkan *modelling*, diberikan saran atau rekomendasi bisnis yang bertujuan menangani indikasi nasabah yang *churn* berdasarkan data. Pada analisis *churn* ini untuk proses pengolahan data menggunakan bantuan *software* anaconda dengan menggunakan Bahasa pemrograman *python*

HASIL DAN PEMBAHASAN

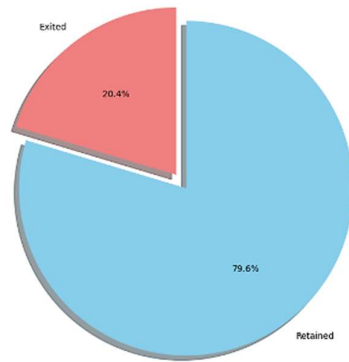
Pada dataset diatas masih belum jelas arti dari tiap kolom, sehingga dilakukan Langkah pendefinisian atau representasi dari tiap kolom agar memudahkan untuk memahami *dataset*.

Customer ID	Merepresentasikan identifikasi unik nasabah
Surname	Merepresentasikan last name nasabah
CreditScore	Merepresentasikan angka yang mengestimasi kemampuan nasabah dalam melunasi utangnya tepat waktu
Geography	Merepresentasikan wilayah nasabah
Gender	Merepresentasikan jenis kelamin nasabah (male or female)
Age	Merepresentasikan usia nasabah
Tenure	Merepresentasikan jangka waktu pinjaman nasabah
Balance	Merepresentasikan saldo pada akun nasabah
NumOfProducts	Merepresentasikan jumlah produk yang digunakan atau dimiliki oleh nasabah
HasCrCard	Merepresentasikan kepemilikan nasabah akan kartu kredit
IsActiveMember	Merepresentasikan apakah member nasabah aktif atau tidak
EstimatedSalary	Merepresentasikan estimasi gaji nasabah
Exited	Merepresentasikan nasabah churn atau tidak oleh nasabah

Gambar 3. Representasi Dari Tiap Kolom *Dataset*

Setelah mengerti representasi tiap kolom, bisa dilihat juga dataset berjenis *supervised learning*. Hal ini karena Pada dataset diatas terdapat kolom *exited (churn)* yang merupakan label sehingga dataset berjenis *supervised learning*. Algoritma *supervised learning* adalah algoritma yang bergantung pada data *input* berlabel untuk mempelajari fungsi yang menghasilkan *output* yang sesuai ketika diberi data baru tanpa label. Selanjutnya menentukan proporsi nasabah yang churn atau tidak.

Proporsi Customer Churn dan Tidak Churn



Gambar 4. Proporsi nasabah

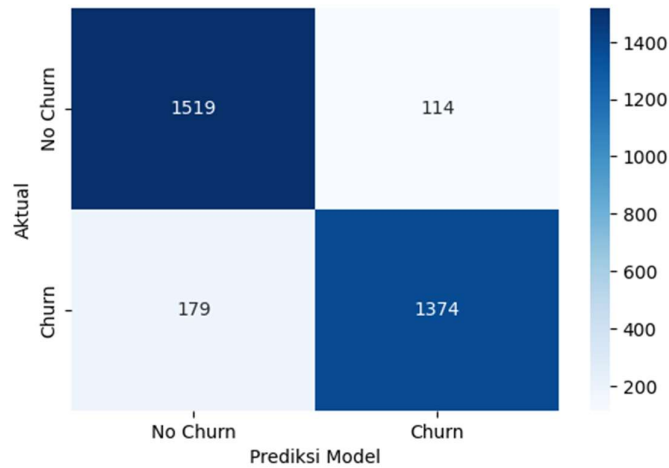
Pada gambar terlihat persentase nasabah churn menunjukkan angka 20,4% sehingga data perlu dilakukan imbalanced supaya dalam proses modelling model yang di buat lebih akurat. Setelah melakukan pemahaman dan penentuan pada *background* data serta pembersihan serta memastikan fitur penting pada data. Selanjutnya pada hasil korelasi pearson menunjukkan bahwa fitur *credit score*, *balance*, *estimated salary*, *age* dan *tenure*, *num of product*, *has credit card*, *is active member*, *exited*, *geography* dan *gender*. Dengan demikian fitur fitur tersebut bida dilanjutkan ke proses penentuan model dengan menguji pada 4 model machine learning yang berbeda yaitu Random Forest, Decision Tree, SVM dan Naïve Bayes.

Tabel 1. *Evaluation Metric*

Model	Accuracy	Precision	Recall	F1	AUC
Random Forest	0,91	0,93	0,89	0,91	0,97
Decision Tree	0,83	0,83	0,82	0,83	0,91
SVM	0,87	0,88	0,84	0,86	0,94
Naïve Bayes	0,81	0,80	0,80	0,80	0,88

Berdasarkan uji coba model, bisa dilihat model dengan performa terbaik adalah Random Forest. Random Forest memiliki nilai *F1-Score* tertinggi di antara model lainnya. *F1-Score* adalah rata-rata harmonis antara presisi (*precision*) dan *recall* (sensitivitas). Pemilihan *F1-Score* sebagai *evaluation metric* karena dataset mengalami ketidakseimbangan kelas sehingga perlu mengidentifikasi nasabah yang akan *churn* dengan akurasi yang tinggi dan meminimalkan jumlah nasabah yang salah diklasifikasikan sebagai *churn*. Selanjutnya dari model *random forest* juga bisa di hasilkan *confusion matrix feature importance*. *Confusion*

matrix merupakan suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. *Confusion matrix* digambarkan dengan tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan



Gambar 5. *Confusion matrix*

Seperti tertera pada gambar bisa dilihat hasil dari *modelling random forest* didapatkan hasil dari 1.519 data nasabah di prediksi tetap berlangganan dan ternyata memang tetap berlangganan (*True Positif*), lalu sebanyak 114 data nasabah diprediksi model berhenti berlangganan tetapi kenyataan tetap berlangganan (*False Negatif*). Sedangkan sebanyak 179 data nasabah diprediksi tetap berlangganan tetapi kenyataannya berhenti berlangganan (*False Positif*) dan sebanyak 1374 data baik prediksi maupun kenyataan berhenti berlangganan (*True Negatif*). Kemudian didapat *churn rate* dengan persamaan nilai *false negative* dibagi dengan total nasabah sehingga nasabah sehingga menghasilkan angka pada penggunaan model sebesar 5,61%. Dengan demikian apabila model digunakan akan menurunkan angka *churn rate* dari 20,4% turun sebesar 14,79% menjadi 5,61%.

Tabel 2. *Feature importance*

Feature	Feature Importance
Age	0,2252352
Num of Products	0,18396855
Is Active Member	0,09250019
Balance	0,09296625
Estimated Salary	0,084213
Credit Score	0,07821962
Tenure	0,0627265
France	0,05348106
Gender	0,04822749
Spain	0,04100564
Germany	0,02391431
Has Cr Card	0,01354218

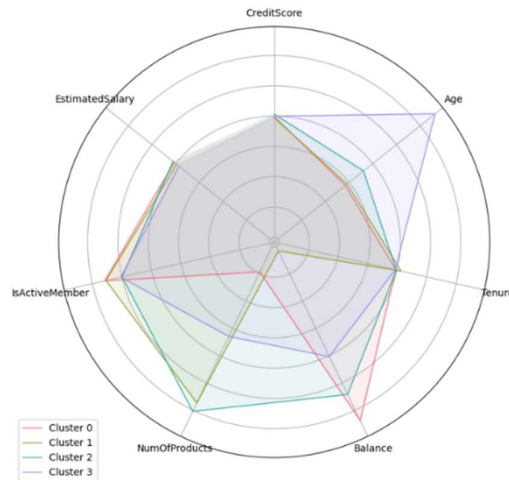
Berdasarkan proses *feature importance* secara berurutan feature importance dari yang tertinggi ke terendah beserta coefisien-nya terlihat pada tabel di atas. Ke tujuh *feature* yang memiliki *importance* tertinggi adalah *Age*, *Num of products*, *Is Active Member*, *Balance*, *Estimated Salary*, *Credit Score* dan *Tenure*. Feature tersebut akan dilakukan segmentasi (*clustering*) untuk melihat segmentasi profil nasabah bank.

Tabel 3. Model Clustering

Clustering (Data Without Handle Outlier)				
Evaluation Metrics	K-Medoids	DBSCAN	Complete Linkage	GMM
Silhouette Score	0.0725	-0.1761	0.1786	0.0993
Davies-Bouldin	3.5634	1.7588	1.940	2.7758

Calinski-Harabasz	738.5761	7.7491	913.0237	894.9819
Evaluation Metrics	Clustering (Handle Outlier With IQR)		Clustering (Drop Outlier)	
	K-Means	K-Prototype	K-Means	K-Prototype
Silhouette Score	0.1556	0.1557	0.1378	0.1366
Davies-Bouldin	1.9757	1.9740	2.2109	2.2190
Calinski-Harabasz	1458.3229	1454.9214	1513.0909	1503.2883

Berdasarkan hasil uji coba, model dengan performa terbaik adalah K-Prototype dengan data *outlier* di *handle* dengan IQR. K-Prototype dipilih berdasarkan nilai *Silhouette Score* tertinggi di antara model lainnya. *Silhouette Score* bernilai positif menunjukkan bahwa objek tersebut cocok dengan kelompoknya sendiri lebih baik daripada kelompok lain, sedangkan nilai negatif menunjukkan sebaliknya.



Gambar 6. Clustering

Berdasarkan gambar diatas terdapat 4 *cluster* dari hasil K-Prototype. *Cluster 0* profil nasabah bank adalah nasabah yang aktif, dengan *balance* paling tinggi, *tenure* lebih lama, dan memiliki *credit score* serta *estimated salary* yang tinggi tinggi. *Cluster* ini memiliki rentang usia yang cenderung beragam. Sementara, untuk kepemilikan produk lebih rendah. Pada *Cluster 1* profil nasabah bank adalah nasabah aktif dengan kepemilikan jumlah produk yang

tinggi, *tenure* sedang, *estimated salary credit score* sedang. Selain itu, nasabah pada *cluster* ini memiliki *balance* rendah dan usia nasabah cenderung beragam. Pada *Cluster 2* profil nasabah adalah nasabah pasif yang kepemilikan produknya tinggi dibandingkan dengan *cluster* lain. Nasabah pada *cluster* ini juga memiliki *balance* lebih tinggi dibanding *cluster 1* dan *3*, dengan *tenure* sedang, *estimated salary* dan *credit score* sedang. Untuk usia, nasabah pada *cluster* ini cenderung beragam. Pada *Cluster 3* nasabah cenderung pasif. Nasabah pada *cluster* ini memiliki usia lebih tua. Kepemilikan produk dan *balance* sedang. Sementara itu, lama *tenure*, *estimated salary*, dan *credit score* juga sedang.

Berdasarkan modelling *classification* dan juga *clustering*, diberikan saran atau rekomendasi bisnis yang bertujuan menangani indikasi nasabah yang *churn* berdasarkan data rekomendasi bisnis tersebut antara lain :

- a. Meningkatkan komunikasi antara bank dengan nasabah sehingga dapat meningkatkan keaktifan dan loyalitas nasabah
- b. Meningkatkan retensi pelanggan dengan menawarkan insentif atau promosi pada nasabah yang memiliki banyak produk
- c. Melakukan survey kepuasan atas produk yang dimiliki nasabah yang mencakup evaluasi produk, sehingga bank dapat mengembangkan produk sesuai dengan preferensi nasabah.
- d. Memberikan limit kredit yang panjang bagi nasabah yang membayar tepat waktu.
- e. Memberi reward kepada nasabah yang setia agar memperkuat loyalitas nasabah.
- f. Menghadirkan produk dan layanan sesuai umur sehingga meningkatkan efektivitas dan efisiensi pemasaran
- g. Memberikan layanan yang lebih mudah dan cepat, seperti layanan perbankan digital yang lebih canggih dan efisien

KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian pada 4 model *machine learning* didapatkan evaluasi metric F1-score yang paling unggul dengan angka 91% selanjutnya dengan *confusion metric* menghasilkan *churn rate* dengan persamaan nilai *false negative* dibagi dengan total nasabah sehingga menghasilkan angka pada penggunaan model sebesar 5,61%. Dengan demikian

apabila model digunakan akan menurunkan angka *churn rate* dari 20,4% turun sebesar 14,79% menjadi 5,61%. Kemudian untuk segmentasi profil nasabah dengan *clustering* didapatkan dari beberapa metode dengan 3 kondisi terhadap *outlier* didapatkan performa terbaik yaitu K-Prototype untuk model clustering dengan angka *Silhouette Score* tertinggi sebesar 0.1557 dan didapatkan 4 *cluster*. Lalu diberikan rekomendasi bisnis yang bertujuan menangani indikasi nasabah yang *churn* berdasarkan

Saran untuk penelitian berikutnya adalah bisa melakukan tahap deployment. Deployment sendiri adalah tahap evaluasi implementasi seluruh model dengan detail, menyesuaikan tahapan model hingga menghasilkan suatu capaian yang sesuai dengan target sehingga bisa terlihat jelas hasil prediksi dari model yang telah dibuat.

UCAPAN TERIMA KASIH

Dengan tulus dan rendah hati, saya ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada orang tua, teman-teman terutama kepada dosen pembimbing atas dukungan, bimbingan, dan kerja keras yang diberikan selama pengerjaan jurnal ini. Tanpa bantuan dan inspirasi dari berbagai pihak, pencapaian ini tidak akan menjadi mungkin. Setiap kontribusi dan dorongan telah menjadi pendorong yang luar biasa. Semoga karya ini dapat memberikan manfaat dan memberikan sumbangsih positif. Terima kasih sekali lagi atas semua bantuan dan dukungan yang luar biasa.

DAFTAR REFERENSI

- Alfarizi, M. Riziq Sirfatullah, Muhamad Zidan Al-farish, Muhamad Taufiqurrahman, Ginan Ardiansah, and Muhamad Elgar. 2023. "Penggunaan Python Sebagai Bahasa Pemrograman Untuk Machine Learning Dan Deep Learning." *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)* 2(1):1-6.
- Arifin, S. (2018). Analisis Faktor-Faktor Yang Mempengaruhi Churn Rate Pada Perusahaan Telekomunikasi Menggunakan Metode Support Vector Machines (Studi Kasus: PT Telekomunikasi XYZ)
- Irmanda, Helena Nurramdhani, Ria Astriratma, and Sarika Afrizal. 2019. "Perbandingan Metode Jaringan Syaraf Tiruan Dan Pohon Keputusan Untuk Prediksi Churn." *JSI: Jurnal Sistem Informasi (E-Journal)* 11(2):1817-25. doi: 10.36706/jsi.v11i2.9286.
- Leni, Desmarita, Helga Yermadona, Ade Usra Berli, Ruzita Sumiati, and Haris Haris. 2023. "Pemodelan Machine Learning Untuk Memprediksi Tensile Strength Aluminium Menggunakan Algoritma Artificial Neural Network (ANN)." *Jurnal Surya Teknika*

10(1):625–32. doi: 10.37859/jst.v10i1.4843.

- Mulia, Chiekal, Aliyah Kurniasih, Program Studi, Ilmu Komputer, and Cilandak Timur. 2023. “Teknik SMOTE Untuk Mengatasi Imbalance Class Dalam Klasifikasi Bank Customer Churn Menggunakan Algoritma Naïve Bayes Dan Logistic Regression.” 0:552–59.
- Nabila, Zulfa, Auliya Rahman Isnain, and Zaenal Abidin. 2021. “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means.” *Jurnal Teknologi Dan Sistem Informasi (JTISI)* 2(2):100.
- Susanto, Edy. 2022. “Analisis Cluster Pasien Covid-19 Berdasarkan Jumlah.” 9(2):817–26.
- Yulianto, A. (2021). Prediksi Customer Churn Pada Bisnis Retail Menggunakan Algoritma Naïve Bayes. *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, 6(1), 41-47.