Jurnal Infortech
Volume 7 No. 2 Desember 2025
E-ISSN: 2715-8160

Algoritma Random Forest Sebagai Uplift Modeling pada Prediksi Retensi Pelanggan Layanan Telekomunikasi

Rizki Tri Prasetio[1], Yudi Ramdhani[2], Vito Hafizh Cahaya Putra[3], Pratiwi[4]

[1,2,3]Universitas Satu
e-mail: [1]rizki.prasetio@univ.satu.ac.id, [2]yudi.ramdhani@univ.satu.ac.id, [3]vito.putra@univ.satu.ac.id

[4]SMK Negeri 13 Bandung
e-mail: [4]pratiwi06@guru.smk.belajar.id

Diterima	Direvisi	Disetujui
07-07-2025	10-09-2025	22-12-2025

Abstrak - Persaingan penyedia layanan telekomunikasi mengakibatkan potensi pelanggan berpindah layanan ke penyedia lain atau disebut dengan *customer churn*. Hal ini dapat mengancam keberlangsungan hidup provider tersebut. Oleh karena itu perlu dilakukan upaya prediksi pelanggan yang mungkin akan melakukan *churn* agar perusahaan dapat mempersiapkan strategi untuk membuat pelanggan dapat kembali berlangganan pada provider tersebut atau disebut dengan *customer retention*. Oleh karena itu pada penelitian ini akan digunakan algoritma *random forest* sebagai *uplift modeling* untuk prediksi retensi pelanggan layanan telekomunikasi. *Uplift modeling* akan berfokus pada mereformulasi target variabel dari pelanggan yang akan melakukan *churn* menjadi pelanggan yang mungkin akan melakukan retensi. Sementara algoritma *random forest* dipilih karena mampu membagi kriteria pemecahan kedalam ruang/segmen yang sesuai dirancang untuk *uplift modeling*. Metode penelitian yang digunakan pada penelitian ini adalah CRISP-DM yang merupakan kerangka penelitian *data mining* untuk penelitian lintas industri. Hasil penelitian ini menunjukan bahwa algoritma yang diusulkan menghasilkan akurasi sebesar 91,87%.

Kata Kunci: *customer churn, customer retention*, *uplift modeling*, *random forest*

Abstract *- Competition in telecommunication service providers causes potential customers to switch services to other providers or known as customer churn. It is necessary to make an effort to predict customers who might churn so that the company can prepare a strategy to get customers to re-subscribe to the provider or called customer retention. Random forest algorithm will be used as an uplift model for predicting customer retention of telecommunications services. Uplift modeling will focus on the formulation of target variables from customers who will churn to customers who are likely to do retention. Meanwhile, the random forest algorithm was chosen because it was able to divide the solving criteria into appropriate spaces/segments designed for uplift modeling. The research method used in this research is CRISP-DM which is a data mining research framework for cross-industry research. The results of this study indicate that the proposed algorithm produces an accuracy of 91.87%.*

Keywords: *customer churn, customer retention, uplift modeling, random forest*

PENDAHULUAN

Bagian paling esensial dari bisnis telekomunikasi ialah pelanggan (Kaharudin, Pradana, & Kusrini, 2019). Pelanggan merupakan sumber pendapatan utama bagi perusahaan telekomunikasi. Kesetiaan pelanggan sangat bergantung dari kualitas layanan provider serta tawaran fitur maupun program yang menarik bagi pelanggan (Suryana, Pratiwi, & Prasetio, 2021). Apabila pelayanan dan penawaran dari provider dinilai tidak menarik bagi pelanggan, maka ada kemungkinan pelanggan akan beralih ke provider lain, tindakan pelanggan ini disebut dengan *customer churn*.

Customer churn dapat berpengaruh pada keberlangsungan hidup provider telekomunikasi (Jain, Khunteta, & Srivastava, 2020). Beberapa upaya dapat dilakukan untuk dapat mengatasi pelanggan yang akan melakukan *churn* yakni memprediksi pelanggan yang mungkin akan melakukan *churn* dan memprediksi pelanggan yang mungkin akan kembali menggunakan layanan provider atau disebut dengan retensi pelanggan (*customer retention*) (Lalwani, Mishra, Chadha, & Sethi, 2021).

Beberapa penelitian telah dilakukan untuk memprediksi *customer churn* dengan berbagai pendekatan dan varian algoritma diantaranya *naïve bayes*, *k-nearest neighbor* (Kaharudin, Pradana, & Kusrini, 2019) (Prasetio R. T., 2020), *extreme learning* (Lalwani, Mishra, Chadha, & Sethi, 2021), *artificial neural network* (Jain, Khunteta, & Srivastava, 2020), *decision tree* (Geiler, Affeldt, &

Nadif, 2022) (Caigny, Coussement, & Bock, 2018) dan *random forest* (Geetha, Punitha, Nandhini, Shakila, & Sushmitha, 2020).

Namun prediksi retensi pelanggan dinilai akan lebih efektif dalam mengembalikan keuntungan perusahaan (Yudiana, Agustina, & Khofifah, 2023) melalui kembalinya pelanggan menggunakan layanan perusahaan provider. Oleh karena itu, perlu modifikasi pada target variabel yang sebelumnya memprediksi pelanggan yang akan melakukan *churn* menjadi memprediksi pelanggan yang mungkin dapat dibujuk untuk kembali menggunakan layanan provider, proses ini dapat dilakukan dengan metode *uplift modeling* (Nyberg & Klami, 2023).

Uplift modeling dinilai lebih baik jika dibandingkan dengan model prediksi *customer churn* serta meningkatkan keuntungan perusahaan melalui program retensi pelanggan melalui penawaran dan kampanye (Yudiana, Agustina, & Khofifah, 2023).

Beberapa penelitian menggunakan pendekatan *uplift modeling* telah dilakukan menggunakan beberapa algoritma yakni *decision tree* (Rafla, Voisine, & Cremilleux, 2023), *ensemble methods* (Vairetti, Marfan, & Maldonado, 2023) dan *contextual treatment selection algorithm* (Saito, Sakata, & Nakata, 2020).

Diantara beberapa algoritma yang telah diimplementasikan untuk *uplift modeling*, algoritma dengan *tree-based algorithm* seperti *decision tree* dinilai lebih baik digunakan pada *uplift modeling* karena merupakan pendekatan natural dengan pembagian segmen berdasarkan kriteria yang dinilai cocok dalam implementasi *uplift modeling* (Rafla, Voisine, & Cremilleux, 2023). Hasil penelitian menggunakan *decision tree* menunjukan hasil yang menarik dengan kelebihan dapat menghindari *overfitting* pada dataset agar diperoleh hasil yang lebih maksimum (Prasetio & Riana, 2015), namun diperlukan analisis yang lebih mendalam agar memperoleh kemungkinan prediksi yang lebih tepat (Riana, Ramdhani, Prasetio, & Hidayanto, 2018).

Oleh karena itu pada penelitian ini akan digunakan algoritma *random forest*, *random forest* menjanjikan analisis dan pilih prediksi yang lebih dalam karena merupakan kumpulan dari beberapa *decision tree* yang nantinya pilihan prediksi ditentukan melalui *majority voting* (Nyberg & Klami, 2023).

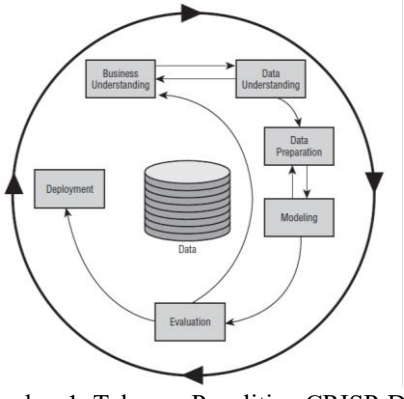
Sumber data yang digunakan pada penelitian ini bersumber dari dataset publik yang didapatkan dari situs BigML dan Kaggle dengan fokus pada dataset *customer churn*. Dengan modifikasi menyesuaikan kebutuhan *uplift modeling* yakni perubahan target variabel dari *customer churn* menjadi *customer retention*. Dataset ini memiliki 3.333 sampel data pelanggan yang merepresentasikan pelanggan yang melakukan *churn* maupun tidak. Berdasarkan uraian tersebut maka manfaat umum dari penelitian ini adalah membantu provider telekomunikasi untuk dapat memprediksi pelanggan yang mungkin dapat diajak kembali menggunakan layanan provider melalui kampanye *customer retention* yang tepat. Selain itu, manfaat khusus penelitian ini guna mendapatkan pendekatan yang paling optimal dalam mengatasi permasalahan *customer churn* melalui strategi kampanye retensi pelanggan.

METODE PENELITIAN

Metode penelitian yang digunakan adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM digunakan karena penelitian ini merupakan penelitian lintas industri antara manajemen pelayanan dan pemasaran dengan pembelajaran mesin. CRISP-DM merupakan standar yang bertujuan untuk melakukan analisa dari industri sebagai strategi pemecahan masalah dari bisnis atau suatu penelitian (Wu, Kumar, Qunlan, Gosh, & Yang, 2008).

Tidak ada karakteristik tertentu untuk data yang dapat diproses karena data tersebut diproses kembali pada setiap fase penelitian. Terdapat enam tahapan atau fase dalam CRISP-DM ini yakni dijelaskan pada Gambar 1.

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)
   Tahapan ini menentukan masalah penelitian secara detil dalam unit penelitian. Masalah dalam penelitian ini adalah memprediksi kemungkinan pelanggan yang masih dapat dibujuk untuk kembali berlangganan pada layanan telekomunikasi menggunakan pembelajaran mesin yang bertujuan untuk menjadi dasar strategi perusahaan dalam kampanye produk layanan telekomunikasi.
2. Fase Pemahaman Data (*Data Understanding Phase*)
   Tahapan ini mengumpulan serta melakukan evaluasi kualitas data. Data yang digunakan dalam penelitian ini merupakan data publik yang didapatkan dari BigML dan Kaggle. Dataset berisi data kondisi pelanggan yang masih menggunakan layanan telekomunikasi dan pelanggan yang mengundurkan diri.



Gambar 1. Tahapan Penelitian CRISP-DM
Sumber: (Wu, Kumar, Qunlan, Gosh, & Yang, 2008)

3. Fase Pengolahan Data (*Data Preparation Phase*)
   Tahapan ini dilakukan persiapan data seperti *remove duplicate* dan *remove missing values* guna meningkatkan kualitas data. Kondisi dataset yang mengalami ketidakseimbangan kelas tidak ditangani di level dataset, namun ditangani di level algoritma, sehingga pada penelitian ini tidak dilakukan teknik resampling. Selain itu, normalisasi dilakukan dengan menggunakan teknik *z-transformation* sehingga masing-masing data dikonversi pada skala tertentu untuk mempermudah analisa data. *Z-transformation* dipilih karena *Z-score* bekerja berdasarkan distribusi data bukan nilai min/max, sehingga lebih efektif jika rentang data di masa depan tidak diketahui atau berubah-ubah (Siregar, et al., 2024). Kemudian membagi dataset menjadi data latih (*data training*) dan data uji (*data testing*) menggunakan *cross validation*.
4. Fase Permodelan (*Modelling Phase*)
   Tahapan ini melakukan pemilihan serta penerapan teknik pemodelan algoritma yang sesuai. Algoritma yang digunakan pada penelitian ini adalah *uplift modeling* menggunakan *random forest*. *Uplift modeling* digunakan untuk mereformulasi target pelanggan yang akan melakukan retensi. Sementara *random forest* diimplementasikan untuk memprediksi *customer retention* pada data latih sehingga menghasilkan model algoritma. Model algoritma yang dihasilkan kemudian akan diterapkan pada data uji sehingga menghasilkan prediksi *customer retention*. Hasil prediksi ini kemudian dievaluasi menghasilkan model yang sesuai dengan tujuan pada tahap awal.
5. Fase Evaluasi (*Evaluation Phase*)
   Tahap ini melakukan evaluasi terhadap model algoritma dengan cara melakukan beberapa eksperimen. Hasil eksperimen kemudian akan dibandingkan dan dievaluasi menggunakan *confussion matrix* untuk mengetahui apakah metode yang diusulkan menghasilkan prediksi yang lebih baik jika dibandingkan dengan penelitian lain.
6. Fase Penyebaran (*Deployment Phase*)
   Fase ini mengimplementasikan model yang dihasilkan pada fase sebelumnya.

HASIL DAN PEMBAHASAN

Penelitian dilakukan dengan cara menguji beberapa skenario algoritma pada dataset retensi pelanggan. Hasil penelitian didapatkan dari dua eksperimen yaitu, eksperimen terhadap algoritma random forest sederhana tanpa optimasi apapun dan menggunakan algoritma random forest sebagai model uplift.

Tabel 1. Hasil Ekeperimen dengan Algoritma *Random Forest* Sederhana

Algoritma	Akurasi
Random Forest	**79,62%**
Naïve Bayes	72,67%
Decision Tree	76,42%
k-Nearest Neighbor	76,27%
Deep Learning	77,25%

A. Hasil Pengujian dengan Algoritma Random Forest Sederhana

Hasil pengujian ini didapat dari algoritma *random forest* tanpa dilakukan optimasi menggunakan *model uplift* untuk klasifikasi retensi pelanggan. Setelah itu hasil pengujian dari algoritma *random forest* dibandingkan dengan algoritma lain. Algoritma tersebut diantaranya *naïve bayes, decision tree, k-nearest neighbor* dan *deep learning*. Validasi yang digunakan pada pengujian ini menggunakan *cross validation* dengan jumlah fold sebanyak 10. Untuk mengetahui hasil terbaik ditunjukan oleh besarnya nilai akurasi yang dihitung menggunakan confussion matrix untuk masing-masing algoritma klasifikasi. Hasil pengujian pertama ini dapat dilihat pada tabel 1.

Hasil pengujian pada tabel 1 menunjukan bahwa akurasi yang dihasilkan dari algoritma random forest merupakan akurasi terbaik yang didapatkan pada eksperimen sederhana tanpa optimasi dengan akurasi 79.62%. Meskipun secara umum hasil pengujian ini belum memuaskan, akan tetapi jika dibandingkan dengan algoritma klasifikasi lain, algoritma random forest merupakan algoritma yang dapat menghasilkan akurasi yang paling baik. Nilai akurasi algoritma random forest pun diatas nilai rata-rata akurasi yang dapat dihasilkan oleh algoritma klasifikasi sederhana yakni sebesar 76,45%.

B. Hasil Pengujian dengan Algoritma *Random Forest* sebagai *Model Uplift*

Hasil pengujian ketiga dilakukan dengan menerapkan random forest sebagai model uplift untuk klasifikasi retensi pelanggan. Validasi yang digunakan pada pengujian ini menggunakan cross validation. Untuk dapat melihat peningkatan akurasi pada klasifikasi retensi pelanggan menggunakan model uplift ini, pengujian dilakukan juga kepada algoritma klasifikasi lain sebagai model uplift-nya, diantaranya naïve bayes, decision tree, k-nearest neighbor dan deep learning.

Tabel 2. Hasil Eksperimen dengan Algoritma *Random Forest* dengan *Model Uplift*

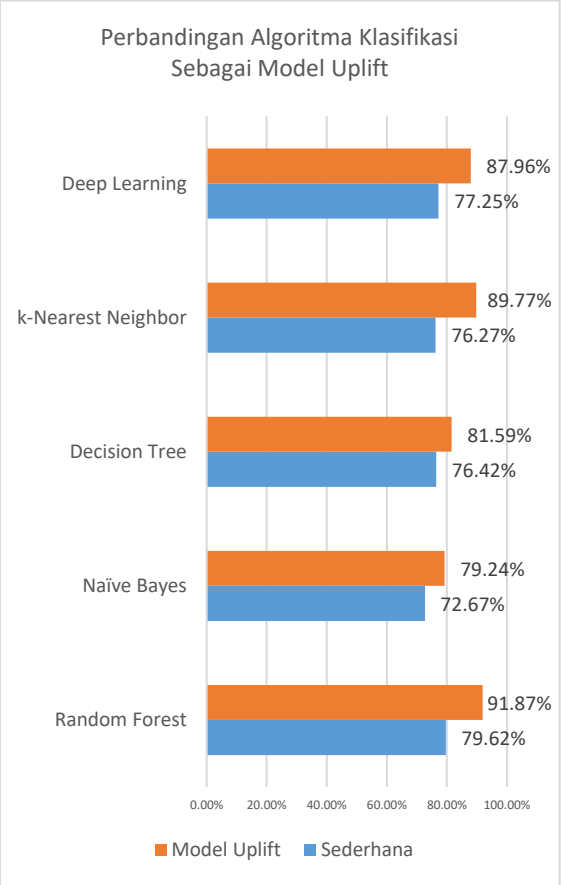
Algoritma	Standar	Uplift
Random Forest	**79,62%**	**91,87%**
Naïve Bayes	72,67%	79,24%
Decision Tree	76,42%	81,59%
k-Nearest Neighbor	76,27%	89,77%
Deep Learning	77,25%	87,96%

Untuk mengetahui hasil terbaik ditunjukan oleh besarnya nilai akurasi yang dihitung menggunakan confussion matrix untuk masing-masing algoritma klasifikasi sebagai model uplift. Hasil pengujian ketiga ini dapat dilihat pada tabel 2.

Hasil pengujian pada tabel 2 menunjukan bahwa akurasi algoritma random forest sebagai model uplift sangat memuaskan yakni 91,87%. Meningkat sekitar 15% dari algoritma random forest sederhana pada pengujian sebelumnya sebesar 79,62%.

Sama seperti pengujian sebelumnya, algoritma random forest sebagai model uplift masih tetap unggul jika dibandingkan algoritma klasifikasi lain sebagai model uplift. Bahkan lebih tinggi dari rata-rata hasil eksperimen yakni sebesar 86,90%. Perbandingan hasil akurasi yang dihasilkan oleh beberapa algoritma klasifikasi sebagai model uplift dapat dilihat pada Gambar 2.

Untuk mengetahui apakah model uplift dapat secara signifikan meningkatkan hasil akurasi dari berbagai algoritma klasifikasi yang telah dilakukan pada eksperimen sebelumnya, dilakukan pengujian t-Test. Hasil pengujian t-Test dapat dilihat pada tabel 3. Dari hasil pengujian t-Test tersebut, menghasilkan hasil P two-tail dibawah 0,05 yakni 0,00397 yang menunjukan bahwa model uplift terbukti dapat meningkatkan akurasi algoritma klasifikasi secara signifikan.



Gambar 2. Perbandingan Algoritma Klasifikasi Sebagai Model Uplift

Tabel 3. Hasil Pengujian t-Test

	Variable 1	*Variable 2*
Mean	0,76446	0,86086
Variance	0,000625	0,002941
Observations	5	5
Pearson Correlation	0,833074	
Hypothesized Mean	0	
df	4	
T Stat	-5,96345	
P(T<=t) one-tail	0,001985	
T Critical one-tail	2,131847	
P(T<=t) two-tail	**0,00397**	
T Critical two-tail	2,776445	

C. Pembahasan Hasil Penelitian

Hasil penelitian ini memiliki implikasi praktis yang signifikan terhadap strategi retensi pelanggan pada perusahaan layanan telekomunikasi. Berbeda dengan pendekatan klasifikasi konvensional yang hanya memprediksi apakah pelanggan akan churn atau tidak, penggunaan random forest sebagai model uplift memungkinkan perusahaan mengidentifikasi pelanggan yang benar-benar responsif terhadap intervensi retensi. Dengan demikian, perusahaan dapat memfokuskan sumber daya promosi, insentif, atau program loyalitas hanya pada segmen pelanggan yang memiliki probabilitas peningkatan retensi akibat perlakuan tertentu, bukan sekadar pelanggan dengan risiko churn tinggi. Pendekatan ini berpotensi mengurangi biaya retensi yang tidak efektif, seperti pemberian insentif kepada pelanggan yang akan tetap bertahan tanpa perlakuan atau pelanggan yang cenderung churn meskipun telah diberikan intervensi.

Secara operasional, hasil uplift modeling dapat digunakan sebagai dasar pengambilan keputusan dalam perancangan kampanye retensi yang lebih presisi dan berbasis data. Perusahaan dapat mengelompokkan pelanggan ke dalam kategori seperti *persuadable*, *sure things*, *lost causes*, dan *do not disturb*, sehingga strategi yang diterapkan menjadi lebih terarah dan adaptif. Temuan bahwa model uplift berbasis random forest menunjukkan peningkatan akurasi yang signifikan juga mengindikasikan bahwa pendekatan ini lebih andal dalam mendukung keputusan bisnis strategis dibandingkan model klasifikasi tradisional.

Dengan mengintegrasikan hasil uplift modeling ke dalam sistem manajemen pelanggan, perusahaan telekomunikasi dapat meningkatkan efektivitas retensi, memaksimalkan return on investment (ROI) dari program pemasaran, serta memperkuat hubungan jangka panjang dengan pelanggan.

KESIMPULAN

Berdasarkan hasil eksperimen yang dilakukan sebanyak tiga kali, dapat disimpulkan bahwa algoritma random forest sebagai model uplift mampu memberikan kinerja yang memuaskan dengan tingkat akurasi sebesar 91,87%, meningkat sebesar 12,25%

dibandingkan penggunaan algoritma random forest tanpa pendekatan uplift. Kinerja ini juga melampaui rata-rata akurasi beberapa algoritma klasifikasi lain yang hanya mencapai 86,90%. Hasil pengujian signifikansi menggunakan uji t menunjukkan bahwa peningkatan kinerja yang dihasilkan oleh model yang diusulkan bersifat signifikan secara statistik, dengan nilai signifikansi two-tailed < 0,05, yaitu sebesar 0,00397.

Implikasi dari hasil penelitian ini menunjukkan bahwa penerapan uplift modeling berbasis random forest berpotensi mendukung perusahaan layanan telekomunikasi dalam merancang strategi retensi pelanggan yang lebih efektif dan efisien. Dengan kemampuan model dalam mengidentifikasi pelanggan yang responsif terhadap intervensi retensi, perusahaan dapat mengalokasikan sumber daya secara lebih tepat sasaran, sehingga tidak hanya meningkatkan tingkat retensi pelanggan, tetapi juga mengoptimalkan efektivitas program pemasaran dan biaya operasional.

REFERENSI

Caigny, A. D., Coussement, K., & Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 760-772.

Geetha, V., Punitha, A., Nandhini, A., Shakila, S., & Sushmitha, R. (2020). Customer Churn Prediction In Telecommunication Industry Using Random Forest Classifier. *IEEE International Conference on System, Computation, Automation and Networking (ICSCAN).* Pondicherry: IEEE.

Geiler, L., Affeldt, S., & Nadif, M. (2022). A survey on machine learning methods for churn prediction. *International Journal of Data Science and Analytics*, 217-242.

Jain, H., Khunteta, A., & Srivastava, S. (2020). Telecom churn prediction and used techniques, datasets and performance measures: a review. *Telecommunication Systems*, 613-630.

Kaharudin, Pradana, M. G., & Kusrini. (2019). PREDIKSI CUSTOMER CHURN PERUSAHAAN TELEKOMUNIKASI MENGGUNAKAN NAİVE BAYES DAN K-NEAREST NEIGHBOR. *Informasi Interaktif*.

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2021). Customer churn prediction system: a machine learning approach. *Computing*, 271-294.

Nyberg, O., & Klami, A. (2023). Exploring uplift modeling with high class imbalance. *Data Mining and Knowledge Discovery*, 736-766.

Prasetio, R. T. (2020). Genetic Algorithm to Optimize k-Nearest Neighbor Parameter for Benchmarked Medical Datasets Classification. *Jurnal Online Informatika, 5*(2), 153-160.

Prasetio, R. T., & Riana, D. (2015). A comparison of classification methods in vertebral column disorder with the application of genetic algorithm and bagging. *2015 4th international conference on instrumentation, communications, information technology, and biomedical engineering (ICICI-BME)* (pp. 163-168). Bandung: IEEE.

Rafla, M., Voisine, N., & Cremilleux, B. (2023). Parameter-Free Bayesian Decision Trees for Uplift Modeling. *Advances in Knowledge Discovery and Data Mining*, 309-321.

Riana, D., Ramdhani, Y., Prasetio, R. T., & Hidayanto, A. N. (2018). Improving Hierarchical Decision Approach for Single Image Classification of Pap Smear. *International Journal of Electrical and Computer Engineering*.

Saito, Y., Sakata, H., & Nakata, K. (2020). Cost-Effective and Stable Policy Optimization Algorithm for Uplift Modeling with Multiple Treatments. *Proceedings of the 2020 SIAM International Conference on Data Mining.* Ohio: Society for Industrial and Applied Mathematics.

Suryana, N., Pratiwi, & Prasetio, R. T. (2021). Penanganan Ketidakseimbangan Data pada Prediksi Customer Churn Menggunakan Kombinasi SMOTE dan Boosting. *IJCIT (Indonesian Journal on Computer and Information Technology), 6*(1).

Vairetti, C., Marfan, M. J., & Maldonado, S. (2023). Dealing With Class Imbalance in Uplift Modeling-Efficient Data Preprocessing via Oversampling and Matching. *IEEE Access*.

Wu, X., Kumar, V., Qunlan, J. R., Gosh, J., & Yang, Q. (2008). *Top 10 Algorithms in Data Mining.* London: Springer-Verlag.

Yudiana, Y., Agustina, A. Y., & Khofifah, N. (2023). Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan. *Indonesian Journal of Islamic Economics and Business*.