

PENANGANAN *MISSING VALUES* UNTUK MENINGKATKAN KINERJA MODEL *MACHINE LEARNING* PADA DATA TELEMARKETING

Anggit Ferdita Nugraha¹, Yoga Pristyanto², Irfan Pratama³

¹ Teknik Komputer, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta,

² Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta
Jl. Ring Road Utara, Condong Catur Sleman, Yogyakarta (telp: 0274-884201-207; 0274-884208)

³ Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta,
Jl. Jembatan Merah, No.84C. Gejayan, Yogyakarta (telp: 0274-584922)

¹anggitferdita@amikom.ac.id

²yoga.pristyanto@amikom.ac.id

³irfanp@mercubuana-yogya.ac.id

Abstrak: Telemarketing menjadi salah satu media pemasaran produk maupun jasa yang sampai saat ini masih banyak digunakan oleh berbagai perusahaan terutama di bidang perbankan, asuransi, dan telekomunikasi. Telemarketing dianggap lebih efektif dibandingkan dengan penggunaan baliho, flyer maupun brosur karena menghubungkan secara langsung antara perusahaan dengan calon pelanggan. Pada proses implementasinya, penggunaan telemarketing ini justru seringkali menemui kegagalan dan bahkan berujung pada sikap antipati calon pelanggan yang dihubungi bahkan sebelum proses komunikasi dilakukan. Salah satu penyebab terjadinya kegagalan dari proses telemarketing adalah tidak adanya proses seleksi untuk mengetahui potensi dari calon pelanggan. *Machine learning* menjadi salah satu alternatif yang dapat digunakan untuk membantu memilih dan menentukan calon pelanggan potensial dan sesuai dengan target perusahaan. Akan tetapi, *missing values* yang ada pada data telemarketing seringkali menjadi permasalahan baru yang harus ditangani agar penggunaan *machine learning* menjadi lebih efektif dan mendukung keberhasilan telemarketing. Pada penelitian ini, *K-means* digunakan sebagai salah satu cara untuk menangani adanya *missing values* pada data telemarketing. Hasilnya, terdapat peningkatan hingga sebesar 6% jika diukur berdasarkan nilai akurasi menggunakan model *Decision Tree*.

Kata Kunci: Telemarketing, *Machine Learning*, *Missing Values*, *K-Means*, Klasifikasi.

Abstract: Telemarketing is one of the marketing media for products and services, which is still widely used by various companies, especially in banking, insurance and telecommunications. Telemarketing is considered more effective than the use of billboards, flyers and brochures because it connects directly between companies and potential customers. In the implementation process, the use of telemarketing often encounters failure and even leads to antipathy attitudes to the prospect of being contacted even before the communication process is carried out. One of the causes of the failure of the telemarketing process is the absence of a selection process to determine the potential of prospective customers. Machine learning is an

alternative that can be used to help select and determine potential customers according to the company's target. However, the missing values that exist in telemarketing data are often a new problem that must be addressed so that the use of machine learning becomes more effective and supports the success of telemarketing. In this study, K-means is used as a way to deal with missing values in telemarketing data. As a result, there is an increase of up to 6% when measured based on the accuracy value using the Decision Tree model.

Keywords: Telemarketing, Machine Learning, Missing Values, K-Means, Classification

I. PENDAHULUAN

Era teknologi informasi dan komunikasi yang berkembang semakin pesat memberikan dampak yang sangat besar bagi kehidupan manusia. Perubahan dari pola dan cara hidup yang konvensional menjadi pola hidup yang serba *online* dan digital, tentu akan membawa budaya baru yang tidak hanya bagi pribadi manusia, melainkan juga bagi lingkungan personal, dan lingkungan organisasi [1], [2].

Salah satu lingkungan organisasi yang terkena dampak dari perubahan budaya tersebut diantaranya adalah lingkungan bisnis dan pemasaran [3]. Pola kerja yang sebelumnya memanfaatkan media fisik seperti *flyer*, brosur, maupun baliho untuk proses pemasaran produk maupun jasa kini berangsur ditinggalkan dan mulai berevolusi untuk memanfaatkan berbagai media digital dan media telekomunikasi. Salah satu contoh yang sampai saat ini masih digunakan adalah Telemarketing [4], [5].

Telemarketing menjadi salah satu metode untuk memperkenalkan produk atau jasa dari suatu perusahaan memanfaatkan media telekomunikasi seperti telepon. Prosesnya, bagian pemasaran dari perusahaan akan menghubungi calon pelanggan untuk kemudian berkomunikasi dan menjalin kerjasama terkait produk maupun jasa yang ditawarkan. Telemarketing sampai saat ini masih dianggap sebagai metode yang efektif karena prosesnya yang secara langsung berkenaan dengan atensi calon pelanggan [4]. Harapannya, muncul timbal-balik antara perusahaan dengan calon pelanggan dalam menemukan kebutuhan yang cocok dan sesuai dengan apa yang diinginkan oleh calon pelanggan. Disisi lain, penggunaan telemarketing sebagai media pemasaran diharapkan mampu menjadi sarana evaluasi bagi

perusahaan terutama yang berkenaan dengan pengembangan produk maupun jasa guna meningkatkan efektifitas dan keuntungan perusahaan.

Meskipun dianggap sebagai metode pemasaran yang efektif, namun, tidak sedikit dari perusahaan yang menggunakan metode telemarketing ini justru mengalami kegagalan dan penolakan dari calon pelanggan, Umumnya penolakan tersebut justru terjadi sebelum adanya proses komunikasi antara pihak perusahaan untuk memperkenalkan barang atau jasa kepada calon pelanggan.

Salah satu faktor yang menjadi penyebab kegagalan dari proses Telemarketing adalah pemilihan calon pelanggan yang akan dihubungi oleh perusahaan dalam rangka untuk menawarkan produk maupun jasa. Proses komunikasi yang dilakukan secara acak, dan banyaknya data calon pelanggan yang tidak tersortir dengan baik oleh perusahaan menjadi penyebab utama proses telemarketing ini menemui hasil yang kurang maksimal. Oleh karena itu, pemanfaatan *machine learning* menjadi satu hal perlu dilakukan guna mengetahui dan menentukan calon pelanggan yang sesuai dengan kriteria perusahaan.

Sampai saat ini, penggunaan model *machine learning* yang populer dan banyak digunakan pada penelitian yang berkaitan dengan data telemarketing diantaranya adalah *Decision Tree* [5], [6], *Naïve Bayes* [7]–[9], *Logistic Regression* [6], [10], dan *Support Vector Machine* [6], [11], [12], dimana secara umum penggunaan model tersebut menunjukkan kinerja yang cukup baik berdasarkan nilai akurasi yang dihasilkan dalam menentukan calon pelanggan yang potensial sekaligus meningkatkan keberhasilan dalam proses Telemarketing.

Pada prosesnya, penggunaan model *machine learning* untuk data Telemarketing masih menerapkan model klasifikasi tunggal dan seringkali mengabaikan adanya *missing values* pada data yang akan diproses [13].

Disisi lain adanya penanganan *missing values* sebelum diproses menggunakan berbagai model *machine learning* diketahui dapat memaksimalkan kinerja klasifikasi [14], [15]. Oleh karena itu, pada penelitian ini dilakukan penanganan *missing value* memanfaatkan proses klasterisasi menggunakan K-Means model pada data Telemarketing yang bersumber dari Bank Marketing Dataset UCI Repository [16], untuk selanjutnya diklasifikasikan menggunakan metode *Decision Tree* sebagai model *machine learning*. Tujuannya adalah untuk mengetahui seberapa besar peningkatan kinerja klasifikasi yang dihasilkan dari penanganan *missing values* yang dilakukan sebelum diproses menggunakan model *machine learning*. dengan adanya peningkatan kinerja klasifikasi diharapkan dapat digunakan sebagai referensi, terutama dalam rangka pengembangan berbagai sistem yang dapat meningkatkan keberhasilan proses Telemarketing sehingga meningkatkan efektifitas penjualan dan keuntungan perusahaan.

II. TINJAUAN PUSTAKA

A. Bank Marketing Dataset

Bank Marketing dataset merupakan dataset utama yang digunakan dalam penelitian ini. Bank Marketing Dataset termasuk dalam kategori *benchmark* dataset untuk Telemarketing yang dapat diperoleh secara mudah melalui halaman *UCI Machine Learning Repository* [16]. Bank Marketing dataset ini memuat informasi Telemarketing yang diakusisi berdasarkan data nasabah *Portuguese Retail Bank* pada tahun 2008

hingga tahun 2013. Dataset ini memuat respon dari calon pelanggan *Potuguese Retail Bank* terhadap proses langganan tabungan deposito jangka panjang dengan instance yang ada pada dataset ini berjumlah 45211, memiliki 16 feature dan 2 kelas yaitu “ya”, dan “tidak”.

Detail fitur dari dataset ini akan dijelaskan menggunakan tabel.1.

Tabel 1. Detail Fitur Bank Marketing Dataset.

Variabel	Tipe	Deskripsi
<i>Age</i>	<i>Numeric</i>	Fitur ini berisi usia dari target pelanggan, dimana data yang tercatat pada variabel / fitur ini memiliki rentang antara 18 – 95 tahun
<i>Job</i>	<i>Categorical</i>	Fitur ini berisi data pekerjaan dari target pelanggan, dimana pada data ini tercatat ada 11 pekerjaan, yakni: <i>admin, blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, students, technicians, unemployed</i> dan <i>unknown</i> . Untuk data <i>unknown</i> , akan dianggap sebagai <i>missing values</i> yang akan ditangani.
<i>Marital</i>	<i>Categorical</i>	Fitur yang berisi status perkawinan dari target pelanggan. Dimana pada fitur ini berisi status seperti: <i>single, married, dan divorced</i>
<i>Education</i>	<i>Categorical</i>	Merupakan fitur yang mengelompokkan jenjang Pendidikan target pelanggan kedalam tiga kelompok, yakni <i>primary, secondary, and tertiary</i> . Pada fitur ini terdapat nilai keempat yakni <i>unknown</i> dimana nantinya akan dianggap sebagai <i>missing values</i> yang perlu ditangani.
<i>Default</i>	<i>Categorical</i>	Fitur yang mencatat data pelanggan berdasarkan kepemilikan kartu kredit. Dimana fitur ini berisi <i>yes</i> , atau <i>no</i>
<i>Balance</i>	<i>Numeric</i>	Fitur yang mencatat besaran saldo pada rekening yang dimiliki target pelanggan
<i>Housing</i>	<i>Categorical</i>	Fitur yang mencatat apakah pelanggan memiliki pinjaman dana perumahan. Pada fitur ini berisi <i>yes</i> atau <i>no</i>
<i>Loan</i>	<i>Categorical</i>	Fitur yang berisi <i>yes</i> atau <i>no</i> , dimana menggambarkan apakah pelanggan memiliki pinjaman dana perseorangan
<i>Contact</i>	<i>Categorical</i>	Fitur yang mencatat bagaimana bentuk komunikasi dengan target pelanggan, fitur ini berisi: <i>cellular, telephone or unknown</i> .

		Fitur <i>unknown</i> akan dianggap sebagai <i>missing values</i> dan akan ditangani
Day	Numeric	Fitur yang mencatat kapan tanggal terakhir berkomunikasi dengan pelanggan
Month	Categorical	Berisi bulan (januari – desember) dimana komunikasi terakhir dengan pelanggan dilakukan
Duration	Numeric	Berisi lama durasi panggilan yang dilakukan sebagai bentuk promosi kepada pelanggan
Campaign	Numeric	Fitur yang mencatat berapa kali panggilan dilakukan selama proses penawaran kepada pelanggan dilakukan melalui telemarketing
Pdays	Numeric	Fitur yang mencatat jeda waktu untuk proses promosi berdasarkan kapan waktu terakhir dihubungi
Previous	Numeric	Fitur yang berisi informasi tentang seberapa sering melakukan kontak dengan pelanggan sebelum penawaran dilakukan
Poutcome	Categorical	Fitur yang berisi respon pelanggan terhadap penawaran produk sebelumnya. Dimana pencatatan fitur ini dilakukan berdasarkan nilai: <i>failure</i> , <i>success</i> , <i>other</i> , dan <i>unknown</i> Dimana <i>unknown</i> merupakan missing value yang perlu untuk ditangani.

B. Klasterisasi

Klasterisasi merupakan istilah yang merujuk pada pengelompokan data berdasarkan kesamaan dan kedekatan karakteristik [17]. Klasterisasi atau *clustering* termasuk dalam kategori *unsupervised learning* dimana prosesnya tidak memerlukan *training set* dan *target output*. Proses pengelompokan yang dilakukan pada model klasterisasi ini dilakukan berdasarkan keanggotaan data, besaran jumlah kelompok yang ditentukan, dan struktur dari kelompok yang terbentuk.

K-Means menjadi salah satu metode klasterisasi yang populer digunakan. K-Means bekerja dengan membagi data kedalam dua atau lebih sesuai dengan kesamaan karakteristik dengan tujuan agar meminimalkan fungsi objektif yang ada pada suatu kelompok serta memaksimalkan nilai variansi antar kelompok yang terbentuk [18].

Dasar dari penggunaan K-Means digambarkan menggunakan algoritme sebagai berikut [19]:

1. Tentukan nilai K sebagai konstanta yang menggambarkan jumlah *cluster* yang akan terbentuk.
2. Inisialisasi nilai K secara acak sebagai *centroid* atau titik pusat cluster.
3. Hitung jarak untuk setiap data dengan titik *centroid* menggunakan persamaan *Euclidian distance* (d)

$$d(x, y) = \sum_{i=1}^n (x_i - y_i)^2 \quad (1)$$

Keterangan;

x: data yang akan di cluster

y: data *centroid*

4. Kelompokkan setiap data berdasarkan nilai yang terdekat dengan centroidnya
5. Perbarui posisi *centroid* dan lakukan langkah ke-3 sampai ke-5, hingga titik pusat *cluster* tidak lagi berubah (konvergen).

C. Decision Tree

Decision Tree merupakan salah satu algoritma populer dalam data mining dan banyak diimplementasikan untuk menyelesaikan berbagai permasalahan terutama yang berkaitan dengan proses klasifikasi dan prediksi [12].

Decision Tree memiliki struktur kerja yang bersifat rekursif, dan sederhana untuk mengekspresikan sebuah proses klasifikasi data. *Decision Tree* menggunakan representasi bentuk struktur pohon dimana setiap *node* dari pohon tersebut akan berperan sebagai *attribute*, cabang menggambarkan nilai dari *attribute* dan daun berperan sebagai representasi dari kelas data yang akan diproses [20].

Keunggulan yang diperoleh dari penggunaan *Decision Tree* ini terletak pada pembentukan modelnya yang cukup mudah, dan sangat cocok untuk digunakan dalam memodelkan data baik numerik maupun kategorikal tentunya dengan

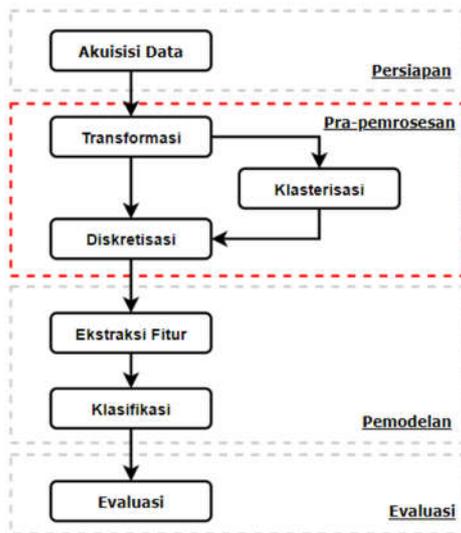
kinerja yang sebanding dengan berbagai macam model *machine learning*, meskipun digunakan untuk pemrosesan data dengan jumlah yang besar [18].

III. METODE PENELITIAN

Secara umum, proses penelitian akan terbagi dalam 4 fase utama yakni:

1. Fase Persiapan
2. Fase Pra-Pemrosesan
3. Fase Pemodelan
4. Fase Evaluasi

Proses yang terdapat dalam masing-masing fase penelitian perlu dilakukan secara berurutan sebagaimana digambarkan menggunakan Gambar 1.



Gambar 1. Tahapan Proses Penelitian.

Tahapan yang dilakukan pertama kali adalah proses akuisisi data yang termasuk dalam fase persiapan. Pada tahapan ini, segala kebutuhan penelitian seperti alat dan bahan yang berupa dataset penelitian akan dipersiapkan dan dipastikan tersedia untuk proses penelitian selanjutnya.

Bank Marketing Dataset yang menjadi bahan utama penelitian ini terlebih dahulu diakuisisi

dengan cara diunduh secara langsung melalui halaman *UCI Machine Learning repository* [16].

Tahapan kedua merupakan tahapan yang berada pada fase pra-pemrosesan. Dimana pada fase ini, dataset yang telah dipersiapkan akan dipelajari terlebih dahulu karakteristiknya untuk menentukan kebutuhan dari data yang nantinya perlu untuk digunakan dalam pemrosesan di tahapan berikutnya, atau perlu dihilangkan karena informasinya bersifat mengganggu atau *noise*.

Apabila dilihat dari detail informasi fitur yang ada pada dataset, terdapat nilai “*unknown*” di berbagai *instance*. Padahal nilai tersebut sejatinya merupakan “*null*” data yang apabila tidak ditangani justru akan mengganggu pemrosesan pada tahap berikutnya. Oleh karena itu, pada pemrosesan awal ini perlu dilakukan adanya proses transformasi untuk mengubah nilai “*unknown*” menjadi “*null*” value.

Setelah proses transformasi dilakukan, berikutnya akan diberikan 2 buah skenario yang akan menjadi fokus pada penelitian ini, dimana skenario pertama adalah dengan membiarkan *null value* pada dataset untuk langsung diproses pada tahapan dan fase pemrosesan selanjutnya, sedangkan skenario kedua adalah dengan melakukan proses klasterisasi untuk penanganan *missing value* menggunakan metode K-Means sebelum diproses lebih lanjut pada tahapan dan fase pemrosesan selanjutnya.

Hal lain yang juga dilakukan untuk melengkapi tahapan proses pada fase pra-pemrosesan ini adalah dengan melakukan proses diskretisasi pada berbagai fitur bertipe numerik dan memiliki sifat kontinu seperti *Age*, *Balance* dan *Duration*. Diskretisasi adalah sebuah proses untuk mengelompokkan nilai numerik kontinu kedalam beberapa kelompok nilai. Proses diskretisasi ini

dilakukan dengan tujuan agar memudahkan dalam pemrosesan data pada fase pemodelan dan analisis.

Fitur yang telah terbentuk dari berbagai tahapan sebelumnya kemudian akan diproses pada fase pemodelan dan analisis menggunakan *Decision tree* sebagai *learning model*nya. Hasil dari penggunaan *learning model* tersebut kemudian akan dilakukan evaluasi berdasarkan pembagian data menggunakan *10-fold cross validation* untuk mengetahui kinerja terbaik dari skenario penelitian yang telah ditentukan berdasarkan nilai akurasi yang dihasilkan.

IV. HASIL DAN PEMBAHASAN

Pada bagian ini, akan dijelaskan dan ditunjukkan berbagai hasil dari penelitian yang dilakukan berdasarkan skenario penelitian yang telah dirancang.

Pada skenario pertama, proses klasifikasi data dari *Bank Marketing dataset* dilakukan tanpa adanya penanganan *missing value* yang ada pada dataset, sedangkan pada skenario kedua, proses klasifikasi dilakukan dengan memberikan penanganan *missing values* menggunakan *K-Means clustering*.

Nilai akurasi yang menjadi acuan utama dalam menentukan kinerja klasifikasi terbaik antara kedua skenario dilakukan berdasarkan pengujian data yang mengimplementasikan *10-fold cross validation* untuk membagi data kedalam bentuk *training* dan *testing* data.

Dengan bantuan dari perangkat lunak *Weka* dan *Orange Machine Learning*, nilai akurasi yang dihasilkan untuk kedua skenario penelitian ditunjukkan menggunakan Gambar 2.



Gambar 2. Perbandingan Kinerja Klasifikasi

Dari Gambar 2. terlihat bahwa pada skenario pertama dimana *missing values* pada dataset tidak diberikan penanganan, menghasilkan kinerja klasifikasi yang baik dengan nilai akurasi sebesar 90.3%. Sedangkan pada skenario kedua, atau skenario yang menangani *missing values* berdasarkan proses klusterisasi menggunakan *K-Means model* diketahui dapat meningkatkan dan memaksimalkan kinerja klasifikasi hingga 6% dengan nilai akurasi yang mencapai 96.5%.

V. KESIMPULAN

Dari penelitian yang telah dilakukan pada kedua skenario penelitian, diketahui bahwa penanganan *missing values* pada suatu dataset perlu untuk dilakukan guna meningkatkan dan memaksimalkan kinerja klasifikasi menggunakan model *machine learning*. Dengan adanya penanganan *missing values* yang memanfaatkan proses klusterisasi menggunakan *K-Means model* terutama untuk data *Telemarketing* dapat meningkatkan nilai akurasi hingga sebesar 6% jika dibandingkan dengan hasil pemrosesan tunggal yang hanya menghasilkan nilai akurasi sebesar 90.3%. dengan adanya peningkatan nilai tersebut diharapkan penelitian ini mampu menjadi referensi untuk pengembangan berbagai sistem yang mendukung dan memaksimalkan tingkat keberhasilan proses *telemarketing*.

REFERENSI

- [1] N. Dewi, R. Purnama, and R. Andari, "Pengaruh Direct Mail dan Telemarketing Terhadap Keputusan Pembelian Paket Umrah Raka Tours and Travel," *Tour. Hosp. Essentials J. (THE Journal)*, vol. III, no. 2, pp. 633–648, 2013.
- [2] Teemu V, "Telemarketing Data Analysis and Predictive Modelling," no. December, 2014.
- [3] R. Vaidehi, "Predictive Modeling to Improve Success Rate of Bank Direct Marketing Campaign," *Int. J. Manag. Bus. Stud.*, vol. 6, no. 1, pp. 22–24, 2016.
- [4] K. C. Schneider, "Telemarketing as a promotional tool-its effects and side effects," *J. Consum. Mark.*, vol. 2, no. 1, p. 29, 1985.
- [5] J. Asare-Frempong and M. Jayabalan, "Predicting customer response to bank direct telemarketing campaign," *2017 Int. Conf. Eng. Technol. Technopreneurship, ICE2T 2017*, vol. 2017-Janua, pp. 1–4, 2017.
- [6] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decis. Support Syst.*, vol. 62, pp. 22–31, 2014.
- [7] Q. R. Zhuang, Y. W. Yao, and O. Liu, "Application of data mining in term deposit marketing," *Lect. Notes Eng. Comput. Sci.*, vol. 2, pp. 14–17, 2018.
- [8] S. Palaniappan, A. Mustapha, C. F. Mohd Foozy, and R. Atan, "Customer Profiling using Classification Approach for Bank Telemarketing," *JOIV Int. J. Informatics Vis.*, vol. 1, no. 4–2, p. 214, 2017.
- [9] M. Purnachary, B. S. S. P. Kumar, and H. Shaziya, "Performance Analysis of Bayes Classification Algorithms in WEKA Tool using Bank Marketing Dataset," *India*, vol. 5, no. 2, pp. 128–133, 2018.
- [10] L. Sing'oei and J. Wang, "Data Mining Framework for Direct Marketing: A Case Study of Bank Marketing," *IJCSI Int. J. Comput. Sci. Issues*, vol. 10, no. 2, pp. 198–203, 2013.
- [11] Y. Pristyanto, I. Pratama, and A. F. Nugraha, "Data level approach for imbalanced class handling on educational data mining multiclass classification," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, 2018, pp. 310–314.
- [12] L. Rahman, N. A. Setiawan, and A. E. Permasari, "Feature Selection Methods in Improving Accuracy of Classifying Students' Academic Performance," *2017 2nd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. (ICITISEE)*, no. 1, pp. 267–271, 2017.
- [13] M. Albayrak, K. Turhan, and B. Kurt, "A missing data imputation approach using clustering and maximum likelihood estimation," in *2017 Medical Technologies National Congress (TIPTEKNO)*, 2017, pp. 1–4.
- [14] Y. Pristyanto and I. Pratama, "Missing Values Estimation on Multivariate Dataset: Comparison of Three Type Methods Approach," in *2019 International Conference on Information and Communications Technology (ICOIACT)*, 2019, pp. 342–347.
- [15] S. Hua-Yan, L. Ye-Li, Z. Yun-Fei, and H. Xu, "Accelerating EM Missing Data Filling Algorithm Based on the K-Means," in *2018 4th Annual International Conference on Network and Information Systems for Computers (ICNISC)*, 2018, pp. 401–406.
- [16] U. Machine Learning, "UCI Machine Learning Repository: Bank Marketing Dataset," 2019. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/bank+marketing>.
- [17] J. O. Ong, "Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. vol.12, no. no. juni, pp. 10–20, 2013.
- [18] Kusriani and E. T. Luthfi, *Algoritma Data Mining*. Penerbit Andi, 2009.
- [19] S. Riama and Chafid, "Algoritma K-Means Clustering Strategi Pemasaran Penerimaan Mahasiswa Baru Universitas Satya Negara Indonesia [Algoritma K-Means Clustering Strategy Marketing Admission Universitas Satya Negara Indonesia]," *Semin. Nas. Cendekiawan ke 4*, no. 2, pp. 685–690, 2018.
- [20] A. F. Nugraha and L. Rahman, "Meta-Algorithms for Improving Classification Performance in the Web-phishing Detection Process," in *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2019, pp. 271–275.