

IMPLEMENTASI MACHINE LEARNING UNTUK MENINGKATKAN KUALITAS OPERASIONAL SERVICE KENDARAAN DENGAN METODE RANDOM FOREST DAN LOGISTIC REGRESSION

IMPLEMENTATION OF MACHINE LEARNING TO IMPROVE THE QUALITY OF VEHICLE SERVICE OPERATIONS WITH RANDOM FOREST AND LOGISTIC REGRESSION METHODS

Ni Made Ika Marini Mandenni¹, I Putu Bayu Adhya Wiratama, Ariyono Setiawan, Gusti Agung Ayu Putri, I Putu Ngurah Krisna Dana

¹Corresponding author, Email: made_ikamarini@unud.ac.id

Universitas Udayana, Jl. Kampus Unud Jimbaran Badung, Bali, 80361, Indonesia

Paper received: 30-07-2024; revised: 08-08-2024; accepted: 02-09-2024; published: 30-10-2024

How to cite (APA Style): Mandenni, N. M. I. M., Wiratama, I. P. B. A., Setiawan, A., Putri, G. A. A., & Dana, I. P. N. K. (2024). Implementasi machine learning untuk meningkatkan kualitas operasional service kendaraan dengan metode random forest dan logistic regression. *Jurnal Praksis dan Dedikasi Sosial (JPDS)*, 7(2), 206-213. DOI: 10.17977/um022v7i2p206-213

Abstract

Vehicle motor repairs (service) are an important aspect for motor vehicle owners to undertake. This activity is carried out by automotive workshops to ensure that the customer's vehicle is in prime condition. To boost sales, some automotive workshops offer various promotional packages to attract customer interest. However, in practice, this is done manually by workshop staff, resulting in suboptimal performance in offer presentations (customer calls). This research aims to build a recommendation system for package deals and offer dates to enhance the quality of customer calls in the operations of automotive workshops using Random Forest and Logistic Regression. The dataset used is operational data from customer calls at one automotive workshop in Bali. The Random Forest model achieves 91 percent accuracy, while Logistic Regression achieves 72 percent accuracy. The system developed can be used to recommend good package deals and offer dates to customers.

Keywords: classification; machine learning; random forest; logistic regression; customer call

Abstrak

Perbaikan kendaraan bermotor (*service*) merupakan hal penting untuk dilakukan bagi pemilik kendaraan bermotor. Kegiatan ini dilakukan oleh bengkel otomotif untuk memastikan kondisi kendaraan *customer* dalam kondisi prima. Untuk meningkatkan penjualan, beberapa bengkel otomotif menawarkan berbagai paket promo untuk menarik minat *customer*. Namun dalam pelaksanaannya, hal ini dilakukan secara manual oleh staff bengkel yang mengakibatkan performa penawaran (*customer call*) kurang optimal. Penelitian ini bertujuan untuk membangun sistem rekomendasi paket dan tanggal penawaran untuk meningkatkan kualitas *customer call* pada operasional bengkel otomotif menggunakan *Random Forest* dan *Logistic Regression*. Dataset yang digunakan adalah data operasional *customer call* salah satu bengkel otomotif di Bali. Model *Random Forest* mencapai akurasi 91 persen dan *Logistic Regression* mencapai akurasi 72 persen. Sistem yang dibangun dapat digunakan untuk merekomendasikan paket dan tanggal penawaran yang baik untuk ditawarkan kepada *customer*.

Kata kunci: klasifikasi; *machine learning*; *random forest*; *logistic regression*; *customer call*

PENDAHULUAN

Saat ini, kendaraan bermotor menjadi kebutuhan utama bagi manusia untuk menempuh jarak jauh dalam berbagai keperluan seperti bekerja dan menempuh pendidikan. Namun, agar tetap dalam kondisi optimal dan aman digunakan, kendaraan bermotor memerlukan perawatan berkala. Perawatan ini mencakup faktor-faktor seperti jarak waktu antara servis terakhir dan servis berikutnya, serta jumlah kilometer yang telah ditempuh kendaraan. Perbaikan dan perawatan rutin ini umumnya dilakukan di bengkel atau workshop yang menyediakan layanan khusus untuk menjaga performa kendaraan tetap prima.

Beberapa bengkel yang beroperasi memiliki strategi pemasaran (*marketing*) untuk menarik pelanggan agar bersedia untuk melakukan perbaikan secara berkala. Salah satu strategi yang dilakukan adalah dengan melakukan panggilan untuk mengingatkan pelanggan akan jadwal perbaikan berikutnya (*Customer Call*). Selain itu, bengkel juga menawarkan berbagai promosi dan paket yang memungkinkan pelanggan untuk membayar biaya yang lebih murah daripada harga normal sehingga menarik minat pelanggan untuk melakukan perbaikan di bengkel tersebut. Strategi ini dilakukan oleh bagian MRA dengan berbagai pertimbangan seperti jenis kendaraan, kilometer yang telah ditempuh, dan lain-lain.

Terdapat berbagai kekurangan dan kendala yang selama ini dialami oleh bagian MRA dalam melakukan *Customer Call*. Tidak semua panggilan yang dilakukan menghasilkan pelanggan atau *customer* yang melakukan *booking* atau memesan jadwal perbaikan berkala. Proses ini juga dilakukan secara manual yang memakan cukup waktu dalam menentukan paket yang akan ditawarkan. Berdasarkan kendala ini, perlu adanya sebuah solusi untuk meningkatkan kualitas *Customer Call* agar tingkat keberhasilan dari strategi ini dapat menjadi lebih baik.

Perkembangan teknologi semakin pesat mengikuti perkembangan zaman modern. Salah satu penerapan teknologi informasi yang banyak dipakai saat ini adalah *Machine Learning*. *Machine Learning* merupakan metode pembelajaran yang luar biasa di mana sistem secara otomatis memperoleh pengetahuan baru (Bagheri & Fazlikhani, 2023). Metode pembelajaran tersebut bergantung pada data yang besar untuk menghasilkan pengetahuan (Capelli et al., 2023). Kemampuan *machine learning* untuk memecahkan masalah bergantung pada jenis algoritma yang digunakan (Mahesh, 2020). Salah satu kemampuan *machine learning* yang banyak dimanfaatkan oleh perusahaan adalah prediksi. Kemampuan ini memungkinkan *machine learning* untuk memprediksi berbagai hal seperti memprediksi inflasi yang berpengaruh pada ekonomi (Elshewey, 2023), memprediksi harga rumah berdasarkan berbagai faktor, memprediksi apakah kreditur layak untuk diberikan pinjaman, dan lain-lain. Hasil dari prediksi ini mampu digunakan untuk membantu manusia dalam mengambil keputusan.

Penelitian sebelumnya yang menggunakan metode *Random Forest* dan *Logistic Regression* untuk klasifikasi dan prediksi telah banyak dilakukan. Beberapa di antaranya mencoba mengembangkan metode *Random Forest* dalam berbagai bidang. Sreejith et al., misalnya, dalam penelitiannya menggunakan *Random Forest Classifier* dan *Red Deer Algorithm* untuk merancang *clinical decision support system* yang membantu tenaga kesehatan dalam memonitor *Sindrom Polycystic Ovarian* (PCOS). *Algoritma Red Deer* (RDA) digunakan untuk menemukan fitur optimal, sementara *Random Forest Classifier* digunakan dalam proses evaluasi. Penelitian ini mencapai akurasi 89,81% dengan dataset PCOS dari Kaggle (Sreejith et al., 2022).

Penelitian lain menggunakan *Random Forest Classifier* untuk memprediksi faktor yang mempengaruhi tingkat penerimaan pembukaan kembali Pembangkit Listrik Tenaga Nuklir Bataan (Ong et al., 2022). Selain *Random Forest Classifier*, *Algoritma Decision Tree*, dan *Artificial Neural Network* (ANN) juga digunakan (Fu et al., 2023). Dataset penelitian ini berasal dari kuesioner yang diisi oleh 1.252 warga Filipina, menghasilkan 46.324 data dari 37 pertanyaan. Algoritma *Random Forest Classifier* dan ANN menghasilkan akurasi masing-masing sebesar 97% dan 93,44%.

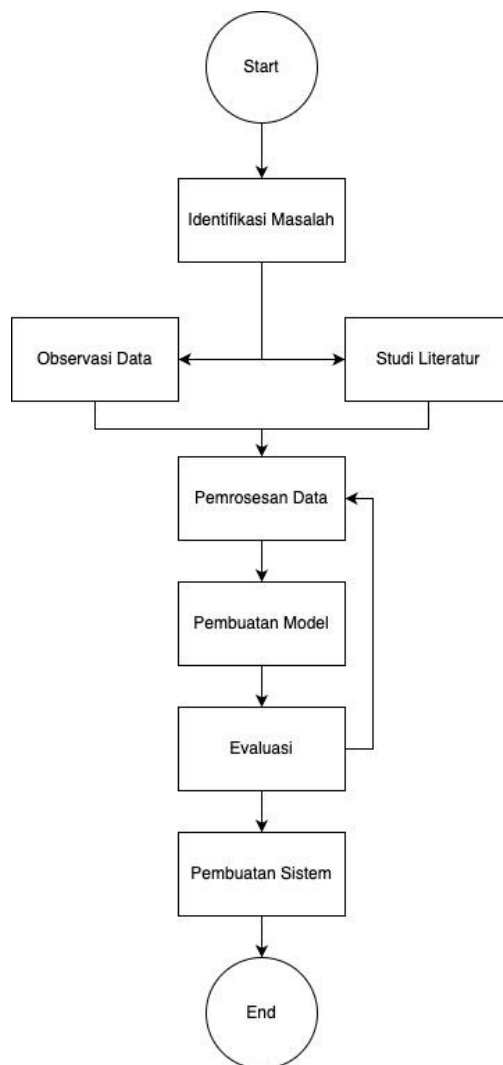
Penelitian lain memanfaatkan *Random Forest Classifier* untuk klasifikasi kanker payudara. Metode ini melibatkan kombinasi bobot fitur (*feature weight*) (Lin et al., 2024) dan *Hyper Parameter Tuned Random Forest Classifier* (Kumar & Nair, 2021). *C-Means Clustering* digunakan untuk mengalokasikan bobot besar pada fitur yang relevan dan bobot kecil pada fitur yang tidak relevan (Hussain et al., 2023). Kemudian, *Tuning Hyperparameter* dilakukan pada model *Random Forest* dengan bantuan algoritma optimasi Bayesian (Wu et al., 2024). Model ini digunakan untuk klasifikasi dataset Kanker Payudara Prognostik Wisconsin (WPBC). Hasil akurasi dari kombinasi metode tersebut (FW + BOA-RDF) menunjukkan peningkatan sebesar 6,66%, 12,659%, dan 37,618% dibandingkan metode lain seperti FW + ALO-BPNN, FW + SSA-SVM, dan FW + GA-SVM.

Berdasarkan latar belakang tersebut, tujuan dari penelitian ini adalah untuk mengembangkan sebuah sistem berbasis *machine learning* yang dapat meningkatkan efektivitas strategi *Customer Call*, dengan fokus pada peningkatan akurasi prediksi paket penawaran yang paling menarik bagi pelanggan, serta mengurangi waktu dan upaya manual yang diperlukan dalam proses tersebut.

METODE

Penelitian dimulai dengan melakukan identifikasi masalah yang akan diselesaikan. Hal ini dicapai dengan melakukan diskusi dengan pembimbing lapangan dan pihak-pihak terkait. Setelah masalah teridentifikasi, selanjutnya dilakukan observasi data dan studi literatur untuk menentukan pendekatan apa yang baik untuk dilakukan. Setelah pendekatan dan metode ditetapkan, dilakukan pengumpulan data lebih lanjut untuk melengkapi kebutuhan data. Hal tersebut dilakukan agar memperoleh data yang reliabel dan obyektif (Sugiyono, 2014). Setelah data dikumpulkan, dilanjutkan dengan pembuatan gambaran umum kerangka kerja yang akan dikerjakan agar proses pengerjaan terstruktur, terukur, dan terarah.

Proses berikutnya yaitu tahap pembuatan model *machine learning*. Tahap ini melibatkan beberapa langkah diantaranya persiapan dan pembersihan data, *preprocessing* data, pembuatan model, dan evaluasi. Metode yang digunakan pada penelitian ini adalah *Random Forest* (Tiwari, 2022) dan *Logistic Regression* (Peng et al., 2002). Setelah model berhasil dibuat dan mencapai performa yang diinginkan, dilanjutkan pada tahap *deployment* atau peluncuran model *machine learning* agar dapat digunakan oleh *user/stakeholder*. Tahap ini melibatkan beberapa langkah diantaranya pembuatan *backend* dan *frontend* dari *platform* yang digunakan (*website*). Tahap analisis dan pembuatan sistem memberikan kontribusi paling besar terhadap nilai dari produk akhir (Casteren, 2017). Setelah *website* dibuat, dilanjutkan pada tahap pengujian untuk melihat apakah fitur-fitur yang ada dapat digunakan dengan lancar dan tanpa hambatan. Pengujian juga memastikan bahwa sesuai dengan standar yang ditetapkan pada tahap sebelumnya (Mudassar & Khan, 2023). Setelah memastikan seluruh fitur dapat digunakan dengan lancar, *website* pun siap untuk digunakan oleh *user/stakeholder*. Gambar 1 merupakan diagram alur yang digunakan sebagai landasan proses penelitian.



Gambar 1. Diagram Alur Penelitian

Data primer yang digunakan pada penelitian ini bersumber dari *Big Data* salah satu perusahaan otomotif di Bali. Data yang digunakan terdiri dari 767 *record*, 8 variabel independen (*fitur*) dan 1 variabel dependen (*target* atau *label*).

HASIL DAN PEMBAHASAN

Exploratory Data Analysis

Tahap *exploratory data analysis* (EDA) (Landge et al., 2023) bertujuan untuk mendapatkan pemahaman terhadap data yang digunakan. Pada penelitian ini, proses EDA yang dilakukan adalah *data summarization* dan *univariate analysis* (Guembe et al., 2022). *Data summarization* dilakukan dengan melihat gambaran besar dari data. Gambar 2 menunjukkan 5 *record* teratas dari data yang digunakan. Terdapat 9 variabel yang terdiri dari 8 variabel independen (MODEL, ASPEK_CALL, BOOKING, BOOKING_REASON, REPAIR_TYPE, TYPE_JOB, LAST_KM, dan COUNT_SBE) dan 1 variabel dependen (PENAWARAN).

Tabel 1 menjelaskan *summary* dari dataset yang digunakan diantaranya penjelasan serta tipe data dari masing-masing variabel. Sebagian besar tipe data dari variabel yang terdapat dalam *dataset* bertipe *object* yang mana tidak dapat langsung digunakan oleh model *machine*

learning. Dari pengamatan ini, proses pengolahan data yang dapat dilanjutkan adalah melakukan *encoding* terhadap variabel bertipe *object* agar bernilai *integer* atau angka.

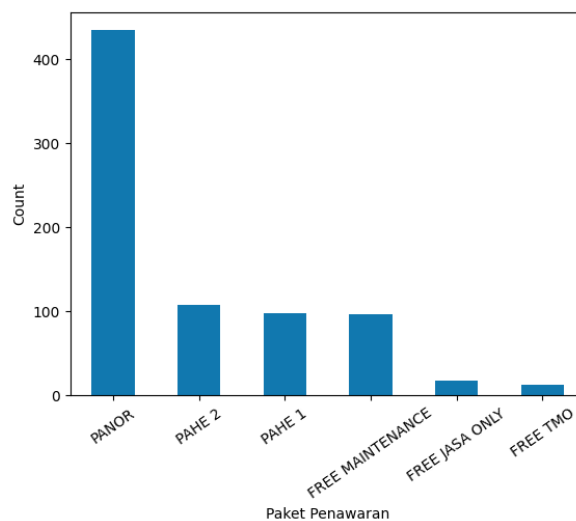
	MODEL	ASPEK_CALL	PENAWARAN	BOOKING	BOOKING_REASON	REPAIR_TYPE	TYPE_JOB	LAST_KM	COUNT_SBE
0	ALPHARD	Call Out	PANOR	N	Pemilik sibuk	SBE	90000	84011.0	1
1	ALPHARD	Call Out	PANOR	N	Konfirmasi pemilik	SBE	90000	84011.0	1
2	VELLFIRE	Call Out	PANOR	N	Pemilik sibuk	SBE	60000	33441.0	1
3	ALPHARD	Call Out	PAHE 1	N	Pemilik sibuk	SBE	80000	36971.0	2
4	ALPHARD	Call Out	FREE TMO	N	Pemilik sibuk	SBE	80000	36971.0	2

Gambar 2. 5 Baris Teratas Dataset

Tabel 1. Summary Dataset

Variabel	Penjelasan	Tipe Data
MODEL	Jenis model kendaraan yang digunakan oleh customer.	Object
ASPEK_CALL	Tipe <i>call</i> atau panggilan yang terjadi, apakah dilakukan oleh staff MRA atau oleh <i>customer</i> itu sendiri.	Object
BOOKING	Kondisi apakah <i>customer</i> melakukan <i>book</i> ketika dihubungi oleh MRA atau tidak.	Object
BOOKING_REASON	Alasan yang diberikan oleh <i>customer</i> ketika menolak melakukan <i>book</i> .	Object
REPAIR_TYPE	Tipe perbaikan yang dilakukan apakah berkala atau perbaikan biasa.	Object
TYPE_JOB	Tipe <i>service</i> berdasarkan berdasarkan kilometer yang telah ditempuh.	Int64
LAST_KM	Kilometer yang telah ditempuh oleh kendaraan <i>customer</i> .	Float64
COUNT_SBE	Kali ke berapa kendaraan <i>diservice</i> secara berkala.	Int64
PENAWARAN (Target)	Jenis penawaran yang ditawarkan oleh staff MRA.	Object

Selanjutnya Gambar 3 menunjukkan hasil *univariate analysis* yang dilakukan pada variabel dependen dengan menggunakan plotting distribusi (*distribution plot*). Plotting Distribusi berguna untuk melihat persebaran data, apakah seimbang atau condong ke satu arah. Berdasarkan hasil plotting di atas, terlihat bahwa mayoritas data condong pada satu nilai yaitu PANOR. Hal ini berarti paket penawaran PANOR merupakan paket yang sering ditawarkan oleh staff MRA kepada customer.



Gambar 3. Univariate Analysis Variabel Dependen

Kemudian disusul oleh paket PAHE 2, PAHE 1, dan FREE MAINTENANCE dan FREE JASA ONLY serta FREE TMO sebagai paket yang paling sedikit ditawarkan oleh staff MRA. Hal ini menunjukkan bahwa dataset yang digunakan merupakan *imbalance* dataset, yaitu dataset yang memiliki variabel dependen yang tidak seimbang. Hal ini memiliki kemungkinan besar untuk menggecohkan hasil evaluasi dari model *Machine Learning* yang dibangun nanti. Tabel 2 menunjukkan hasil evaluasi metode random forest dan logistic regression.

Tabel 2. Hasil Evaluasi Metode Random Forest dan Logistic Regression

Model	Kelas	Precision	Recall	Accuracy
Random Forest	FREE JASA ONLY	0.98	0.98	0.98
	FREE MAINTENANCE	0.92	0.93	0.93
	FREE TMO	0.99	0.99	0.99
	PAHE 1	0.91	0.90	0.90
	PAHE 2	0.88	0.86	0.87
	PANOR	0.81	0.82	0.82
	FREE JASA ONLY	0.50	0.50	0.50
Logistic Regression	FREE MAINTENANCE	0.68	0.60	0.64
	FREE TMO	0.33	0.33	0.33
	PAHE 1	0.81	0.59	0.68
	PAHE 2	0.47	0.56	0.51
	PANOR	0.79	0.85	0.82

Untuk mengatasi hal tersebut, maka dapat dilakukan proses *sampling* untuk menyamaratakan jumlah sampel dari masing-masing kelas.

Model Evaluation

Evaluasi dilakukan dalam dua (2) tahap, yaitu evaluasi menggunakan data validasi dan menggunakan data latih. Data validasi didapatkan dari melakukan *cross validation* (Nakatsu, 2023) dengan jumlah *fold* sebanyak 10. Tabel 2 menunjukkan hasil evaluasi model pada tiga (3) metrik, diantaranya *precision*, *recall*, dan *accuracy*. Berdasarkan hasil tersebut, didapatkan akurasi rata-rata model *Random Forest* sebesar 91% dan *Logistic Regression* sebesar 72% dalam melakukan klasifikasi terhadap 6 buah kelas. Hasil ini menunjukkan bahwa model *Random Forest* menghasilkan performa yang lebih baik dibandingkan *Logistic Regression*, sehingga model *Random Forest* akan digunakan pada tahap *deployment* model.

Hasil Analisis Rekomendasi Tanggal Penawaran

Rekomendasi tanggal penawaran dilakukan berdasarkan perhitungan rata-rata selisih jarak hari antara *customer call* yang tidak menghasilkan *book* dengan *customer call* yang menghasilkan *book* sesuai kategori alasan *customer*. Perhitungan rata-rata tersebut menggunakan rumus rata-rata (*mean*) (Sreejith et al., 2022). Tabel 3 menunjukkan analisis rekomendasi tanggal penawaran per reason.

$$y = \frac{\sum x_1 - x_0}{n} \tag{1}$$

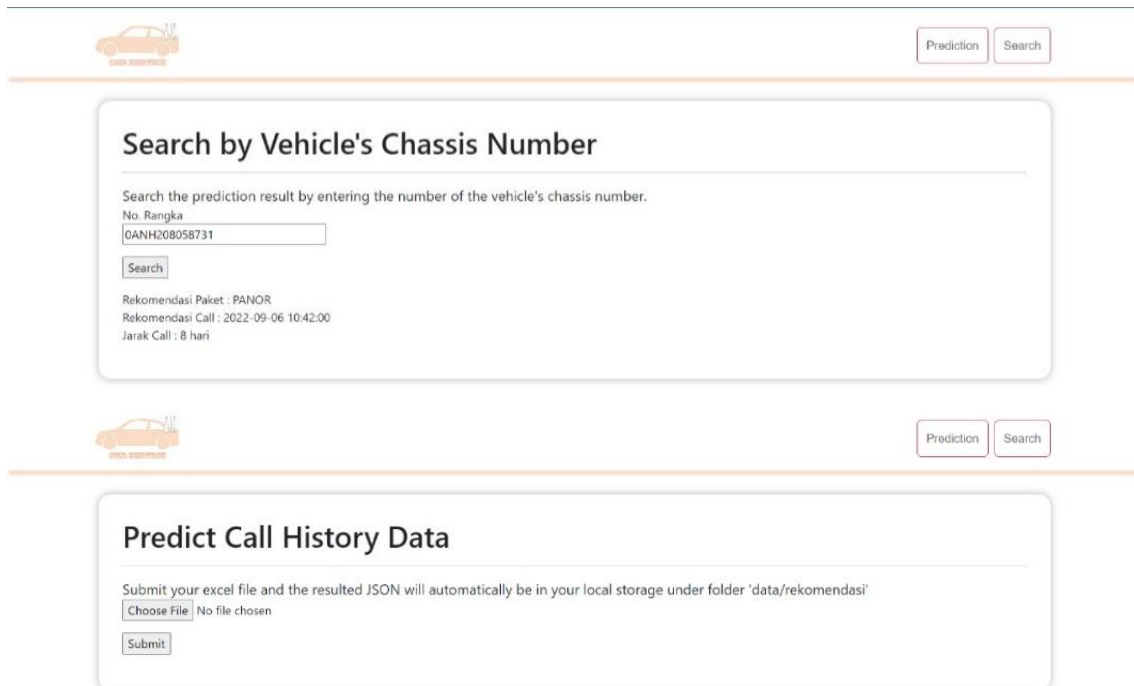
Dimana y adalah rata-rata, x_1 adalah tanggal customer melakukan book, x_0 adalah tanggal customer menolak melakukan book, dan n adalah jumlah penawaran yang berhasil melakukan book.

Tabel 3. Analisis Rekomendasi Tanggal Penawaran per Reason

Alasan Menolak	Rata-Rata Selisih Jarak (hari)
Pemilik Sibuk	9
Konfirmasi Pemilik	8
KM masih rendah	9
Kendaraan operasional luar kota	7
Sudah melakukan service	10
Kendaraan sibuk	11
Customer masih diluar kota	8
Kendaraan sudah dijual	16
Belum ada dana	4
Customer Fleet	15

Website Testing

Model Machine Learning yang telah dibuat kemudian dideploy pada platform website agar dapat digunakan oleh user. Website yang dibangun menggunakan bahasa pemrograman Python dengan framework Flask. Gambar 3 menunjukkan hasil pengujian *website* berdasarkan beberapa skenario.



Gambar 3. Hasil Pengujian Website

KESIMPULAN

Sistem rekomendasi berbasis *machine learning* telah dikembangkan untuk membantu bagian MRA meningkatkan efisiensi dalam menentukan jadwal dan paket penawaran *servis* kepada *customer* saat melakukan *reminder servis*. Menggunakan metode *Random Forest* yang dideploy dalam bentuk website, sistem ini bertujuan meningkatkan efektivitas operasional *customer call* dengan menghemat waktu, tenaga, dan sumber daya. Model ini menunjukkan performa yang baik dengan akurasi 91% pada tahap training dan 72% pada tahap testing. Analisis data juga memberikan rekomendasi jarak call yang disesuaikan berdasarkan alasan penolakan *servis* oleh *customer*, yang berguna dalam menentukan paket dan tanggal penawaran. Setelah performa model memenuhi standar, sistem ini siap digunakan oleh user atau stakeholder melalui platform website.

DAFTAR PUSTAKA

- Bagheri, I., & Fazlikhani, M. (2023). A deeper review on applications of machine learning in data mining. *Conference: 5th International Conference on Science, Engineering, and Role of Technology in New Businesses*.
- Capelli, G., Verdi, D., Frigerio, I., Rashidian, N., Ficorilli, A., Grasso, V., Majidi, D., Gumbs, A. A., & Spolverato, G. (2023). White paper: Ethics and trustworthiness of artificial intelligence in clinical surgery. *Artificial Intelligence Surgery*, 3(2), 111–122.
- Elshewey, A. M. (2023). An optimized ensemble model for inflation prediction in Egypt. *American Journal of Business and Operations Research (AJBOR)*, 10(2), 61–73.
- Fu, M., Zhang, C., Hu, C., Wu, T., Dong, J., & Zhu, L. (2023). Achieving verifiable decision tree prediction on hybrid blockchains. *Entropy*, 25(7), 1058.
- Guembe, B., Azeta, A., Misra, S., & Garg, L. (2022). Multivariate and univariate anomaly detection in machine learning: A bibliometric analysis. *International Conference on Information Systems and Management Science*, 341–363.
- Hussain, I., Sinaga, K. P., & Yang, M.-S. (2023). Unsupervised multiview fuzzy c-means clustering algorithm. *Electronics*, 12(21), 4467.
- Kumar, P., & Nair, G. G. (2021). An efficient classification framework for breast cancer using hyper parameter tuned Random Decision Forest Classifier and Bayesian Optimization. *Biomedical Signal Processing and Control*, 68, 102682.
- Landge, H. P. J., Patel, K., Bhanushali, A., Rajaram, S., & Mittal, S. (2023). A review on data visualization for exploratory data analysis. *Journal of Data Acquisition and Processing*, 38(4), 478–494.
- Lin, S., Wang, C., Mao, Y., & Lin, Y. (2024). Feature selection for label distribution learning under feature weight view. *International Journal of Machine Learning and Cybernetics*, 15(5), 1827–1840.
- Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*, 9(1), 381–386.
- Nakatsu, R. T. (2023). Validation of machine learning ridge regression models using Monte Carlo, bootstrap, and variations in cross-validation. *Journal of Intelligent Systems*, 32(1), 20220224.
- Ong, A. K. S., Prasetyo, Y. T., Velasco, K. E. C., Abad, E. D. R., Buencille, A. L. B., Estorninos, E. M., Cahigas, M. M. L., Chuenyindee, T., Persada, S. F., & Nadlifatin, R. (2022). Utilization of random forest classifier and artificial neural network for predicting the acceptance of reopening decommissioned nuclear power plant. *Annals of Nuclear Energy*, 175, 109188.
- Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1), 3–14.
- Sreejith, S., Nehemiah, H. K., & Kannan, A. (2022). A clinical decision support system for polycystic ovarian syndrome using red deer algorithm and random forest classifier. *Healthcare Analytics*, 2, 100102.
- Wu, K., Kim, K., Garnett, R., & Gardner, J. (2024). The behavior and convergence of local bayesian optimization. *Advances in Neural Information Processing Systems*, 36.