

Perbandingan Model *Logistic Regression* dan *Artificial Neural Network* pada Prediksi Pembatalan Hotel

Moch Shandy Tsalasa Putra ^{(1)*}, Yufis Azhar ⁽²⁾

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Malang, Malang
e-mail : shandytsalasa@gmail.com, yufis@umm.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 5 Juni 2020, direvisi 6 Juli 2020, diterima 6 Juli 2020, dan dipublikasikan 20 Januari 2021.

Abstract

Prediction for canceled booking hotels is an important part of hotel revenue management systems in the modern era. Because the predicted result can be used for the optimization of hotel performance. The application of machine learning will be very helpful for predicting canceled booking hotels because machine learning can process complex data. In this research, the proposed methods are Artificial Neural Network (ANN) and Logistic Regression. Later it will be done five times experiments with hyperparameter tuning to see which method is the most optimal to do prediction canceled booking hotel. From five times experiments, experiments number five (logistic regression with GridSearchCV) is the most optimal for predicting canceled booking hotels, with 79.77% accuracy, 85.86% precision, and 55.07% recall.

Keywords: *Artificial Neural Network, Logistic Regression, Prediction, Data Mining, Machine Learning*

Abstrak

Prediksi pembatalan pemesanan hotel adalah sebuah bagian yang penting dari *hotel revenue management systems* pada era modern ini. Karena dari hasil prediksi yang didapat, nantinya bisa dijadikan acuan untuk optimalisasi kinerja hotel. Penggunaan model *machine learning* akan sangat membantu dalam melakukan prediksi pembatalan pemesanan hotel, karena model *machine learning* bisa memproses data yang kompleks. Pada penelitian ini metode yang akan diusulkan adalah Jaringan Syaraf Tiruan (JST) dan Regresi Logistik. Nantinya akan dilakukan lima kali percobaan dan juga akan dilakukan *hyperparameter tuning* untuk melihat metode mana yang paling optimal untuk melakukan prediksi pemesanan hotel. Dari lima kali percobaan, percobaan yang kelima (*logistic regression with GridSearchCV*) adalah metode yang paling optimal dalam melakukan prediksi pembatalan pemesanan hotel, dengan nilai *accuracy* sebesar 79.77%, nilai *precision* 85.86% dan nilai *recall* 55.07%.

Kata Kunci: Jaringan Syaraf Tiruan, Regresi Logistik, Prediksi, Penggalian Data, Pembelajaran Mesin

1. PENDAHULUAN

Prediksi pembatalan pemesanan hotel adalah sebuah bagian yang penting dari *hotel revenue management systems* pada era modern ini. Karena dari hasil prediksi yang didapat, nantinya bisa dijadikan acuan untuk optimalisasi kinerja hotel (Lee, 2018; Rajopadhye et al., 2001). Penggunaan model *machine learning* sangat membantu dalam melakukan prediksi pembatalan pemesanan hotel, karena model *machine learning* bisa memproses data yang kompleks, seperti riwayat pemesanan hotel dan *hotel demand* (Zhang, 2019).

Penelitian tentang *hotel demand*, merupakan hal yang tidak baru. Pertama kali diteliti pada tahun 1966 di industri penerbangan, lalu diadopsi oleh industri lainnya, seperti hotel, penyewaan hotel, lapangan golf, dan kasino (Chiang et al., 2007). Penelitian yang dilakukan oleh (Weatherford et al., 2001) adalah untuk *forecast hotel revenue management*, metode yang digunakan untuk melakukan *forecast* adalah *aggregated forecast* dan *disaggregated forecast*. Tetapi pada penelitian yang dilakukan Weatherland et al. (2001) memiliki kelemahan, yaitu metode yang digunakan termasuk metode yang *outdated* pada tahun ini (2020).

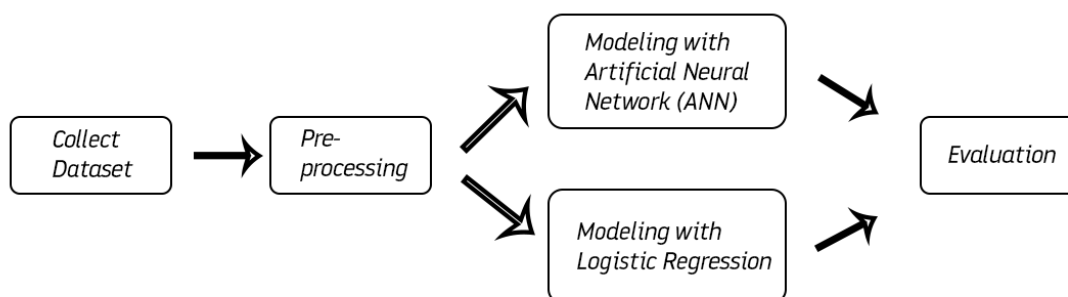


Pada penelitian yang dilakukan oleh (Antonio et al., 2017) adalah melakukan prediksi pembatalan pemesanan hotel. Metode pendekatan yang dipilih adalah *Boosted Decision Tree*, *Decision Forest*, *Decision Jungle*, *Locally Deep Support Vector Machine*, dan *Neural Network*. Hasil dari penelitian Antonio et al. (2017) metode yang terbaik untuk dipakai dalam prediksi pembatalan pemesanan hotel adalah menggunakan *Decision Forest* dan *Boosted Decision Tree*.

Artificial Neural Network atau yang lebih dikenal dengan ANN, merupakan bagian dari *artificial intelligence*. ANN adalah sebuah metode yang biasa dipakai untuk melakukan klasifikasi atau prediksi (El_Jerjawi & Abu-Naser, 2018). *Logistic Regression* atau yang lebih dikenal dengan *logistic regression analysis*, juga sering dipakai di dunia *data mining* dan digunakan untuk melakukan klasifikasi *independent variables* (Wang et al., 2019). Oleh karena itu, pada penelitian ini akan dilakukan perbandingan antara metode klasifikasi ANN dan *logistic regression* untuk melihat mana di antara kedua metode tersebut yang lebih baik dalam melakukan prediksi pada kasus pembatalan hotel.

2. METODE PENELITIAN

Alur proses yang akan dilakukan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Alur Proses Penelitian.

Gambar di atas merupakan alur proses dari penelitian yang akan dilakukan. Mulai dari pengumpulan *dataset*, melakukan *pre-processing*, lalu menggunakan dua metode untuk perbandingan performa prediksi yaitu *modeling* menggunakan metode *Artificial Neural Network* (ANN) dan *modeling* menggunakan metode *Logistic Regression*, dan *evaluation* dari model yang dibuat.

2.1. Pengumpulan *Dataset*

Dataset yang akan digunakan pada penelitian ini adalah menggunakan *dataset* dari penelitian Antonio et al. (2019). *Dataset* ini memiliki 31 *independent variables*, 1 *dependent variable* dan 119390 baris.

Untuk nama kolom dan tipe data dari *independent variables* adalah seperti pada Tabel 1 dan 2.

Tabel 1. Nama Kolom dan Tipe Data dari *Independent Variables*.

Nama Kolom	Tipe Data
<i>ADR</i>	<i>Numeric</i>
<i>Adults</i>	<i>Integer</i>
<i>Agent</i>	<i>Categorical</i>
<i>ArrivalDateDayOfMonth</i>	<i>Integer</i>
<i>ArrivalDateMonth</i>	<i>Categorical</i>
<i>ArrivalDateWeekNumber</i>	<i>Integer</i>
<i>ArrivalDateYear</i>	<i>Integer</i>
<i>AssignedRoomType</i>	<i>Categorical</i>
<i>Babies</i>	<i>Integer</i>
<i>BookingChanges</i>	<i>Integer</i>



Tabel 2. Nama Kolom dan Tipe Data dari *Independent Variables* (lanjutan).

Nama Kolom	Tipe Data
<i>Children</i>	<i>Integer</i>
<i>Company</i>	<i>Categorical</i>
<i>Country</i>	<i>Categorical</i>
<i>CustomerType</i>	<i>Categorical</i>
<i>DaysInWaitingList</i>	<i>Integer</i>
<i>DepositType</i>	<i>Categorical</i>
<i>DistributionChannel</i>	<i>Categorical</i>
<i>Hotel</i>	<i>Categorical</i>
<i>IsRepeatedGuest</i>	<i>Categorical</i>
<i>LeadTime</i>	<i>Integer</i>
<i>MarketSegment</i>	<i>Categorical</i>
<i>Meal</i>	<i>Categorical</i>
<i>PreviousBookingsNotCanceled</i>	<i>Integer</i>
<i>PreviousCancellations</i>	<i>Integer</i>
<i>RequiredCardParkingSpaces</i>	<i>Integer</i>
<i>ReservationStatus</i>	<i>Categorical</i>
<i>ReservationStatusDate</i>	<i>Date</i>
<i>ReservedRoomType</i>	<i>Categorical</i>
<i>StaysInWeekendNights</i>	<i>Integer</i>
<i>StaysInWeekNights</i>	<i>Integer</i>
<i>TotalOfSpecialRequests</i>	<i>Integer</i>

Dan untuk nama kolom dan tipe data untuk *dependent variable* adalah seperti pada Tabel 3.

Tabel 3. Nama Kolom dan Tipe Data dari *Dependant Variable*.

Nama Kolom	Tipe Data
<i>IsCanceled</i>	<i>Categorical</i>

2.2. Pre-processing

Dataset yang sudah diperoleh akan dilakukan proses *pre-processing*. Proses yang dilakukan adalah menghapus variabel atau kolom yang memiliki nilai *NaN* terlalu banyak, merubah tipe data yang bersifat objek menjadi data yang numerik agar bisa diproses dengan model yang diusulkan. Kolom yang dihapus atau yang tidak digunakan adalah kolom *Country*, *Company*, *ReservationStatus*, dan *ReservationStatusDate*. Setelah dilakukan proses *pre-processing*, akan didapatkan *dataset* yang memiliki 28 *independent variables* dan 1 *dependent variable*. Hasil dari *pre-processing* yang telah dilakukan ditampilkan pada Tabel 4 dan 5.

Tabel 4. Hasil *Pre-processing*.

Nama Kolom	Tipe Data
<i>ADR</i>	<i>Integer</i>
<i>Adults</i>	<i>Integer</i>
<i>Agent</i>	<i>Float</i>
<i>ArrivalDateDayOfMonth</i>	<i>Integer</i>
<i>ArrivalDateMonth</i>	<i>Integer</i>
<i>ArrivalDateWeekNumber</i>	<i>Integer</i>
<i>ArrivalDateYear</i>	<i>Integer</i>
<i>AssignedRoomType</i>	<i>Integer</i>
<i>Babies</i>	<i>Integer</i>
<i>BookingChanges</i>	<i>Integer</i>
<i>Children</i>	<i>Integer</i>
<i>CustomerType</i>	<i>Integer</i>
<i>DaysInWaitingList</i>	<i>Integer</i>
<i>DepositType</i>	<i>Integer</i>



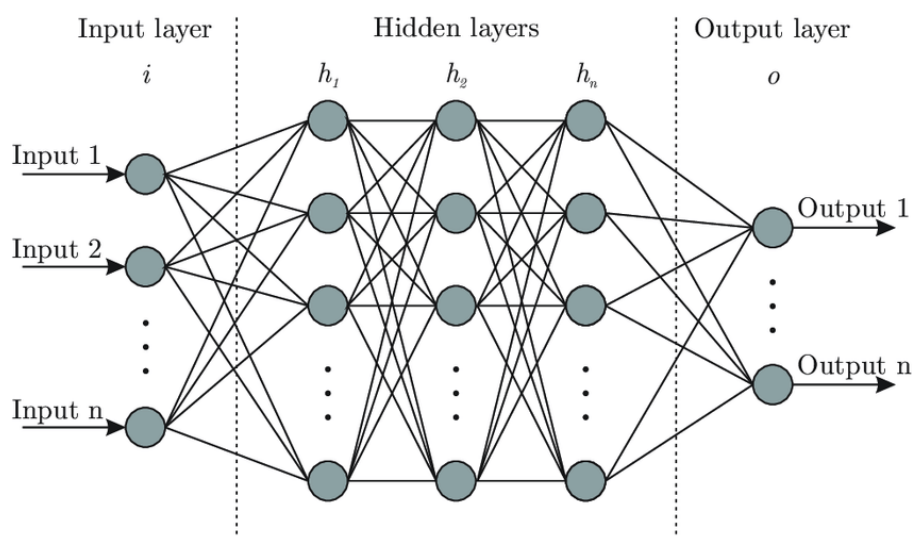
Tabel 5. Hasil *Pre-processing* (lanjutan).

Nama Kolom	Tipe Data
<i>DistributionChannel</i>	<i>Integer</i>
<i>Hotel</i>	<i>Integer</i>
<i>IsRepeatedGuest</i>	<i>Integer</i>
<i>LeadTime</i>	<i>Integer</i>
<i>MarketSegment</i>	<i>Integer</i>
<i>Meal</i>	<i>Integer</i>
<i>PreviousBookingsNotCanceled</i>	<i>Integer</i>
<i>PreviousCancellations</i>	<i>Integer</i>
<i>RequiredCardParkingSpaces</i>	<i>Integer</i>
<i>ReservedRoomType</i>	<i>Integer</i>
<i>StaysInWeekendNights</i>	<i>Integer</i>
<i>StaysInWeekNights</i>	<i>Integer</i>
<i>TotalOfSpecialRequests</i>	<i>Integer</i>

2.3. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) merupakan model matematika yang cocok digunakan dalam *data mining*, terutama dalam kasus prediktif *data mining* (Bishop, 1995). ANN adalah sebuah sistem komputasi komputer yang terinspirasi dari jaringan syaraf otak manusia. Jutaan neuron akan melakukan komunikasi dengan sinapsis yang akan menciptakan jaringan syaraf yang nantinya manusia bisa beraktivitas seperti berbicara, membaca, bernafas, *face detection*, *voice recognition*, menyelesaikan sebuah masalah, bahkan menyimpan sebuah data (Dunne, 2007).

ANN merupakan *computational learning system* yang merupakan fungsi jaringan atau *network* untuk memahami dan menerjemahkan input dari *user* untuk mengeluarkan *output* yang *user* inginkan. ANN bukanlah sebuah algoritma, melainkan sebuah *framework* yang digunakan *machine learning* untuk menyelesaikan input data yang kompleks (Santos et al., 2018). Secara umum bentuk arsitektur dari ANN adalah *input layer*, *hidden layer*, *output layer*.



Gambar 2. Arsitektur *Artificial Neural Network* (ANN) (Bre et al., 2018).

2.4. Logistic Regression

Logistic regression adalah sebuah metode klasifikasi di dalam *statistical machine learning*, dan juga termasuk metode *supervised learning*. Metode ini memiliki performa yang unggul ketika menangani data dengan skala yang besar dan merupakan metode yang paling umum digunakan di *data mining* (Ran et al., 2018; Shamsaei & Gao, 2016; Z. Yang & Li, 2019). Selain digunakan



di *data mining*, juga bisa digunakan dalam ilmu kedokteran dan ilmu sosial. Contohnya dalam dunia ilmu kedokteran *logistic regression* bisa digunakan untuk membantu dalam pengambilan keputusan yang penting (Vono et al., 2018).

Logistic regression menentukan hubungan antara *output* yang berupa *binary classification* dengan *independent variables* menggunakan probabilitas dengan cara memprediksi nilai untuk *dependant variable* (Fujii et al., 2015; Rushin et al., 2017; Zou et al., 2019). Bentuk matematika dari model *logistic regression* ditunjukkan pada Pers. 1, dimana $\sigma(\cdot)$ merupakan *logistic function* yang mengadopsi dari *sigmoid activation*, seperti yang dituliskan pada Pers. 2. Dari kedua persamaan tersebut nantinya akan menghasilkan nilai *binary prediction* seperti yang ditunjukkan pada Pers. 3.

$$\hat{p} = h\theta(x) = \sigma(\theta^T \cdot x) \quad (1)$$

$$\sigma(t) = \frac{1}{1 + \exp(-t)} \quad (2)$$

$$\hat{y} = f(x) = f(x) = \begin{cases} 0, & \text{jika } \hat{y} < 0.5 \\ 1, & \text{jika } \hat{y} \geq 0.5 \end{cases} \quad (3)$$

Selain digunakan untuk melakukan *binary classification*, *logistic regression* juga bisa melakukan *multinomial regression*, *multinomial regression*, *multi-class logistic regression* atau *the maximum entropy classifier* (B. Yang et al., 2019).

3. HASIL DAN PEMBAHASAN

Untuk mengetahui performa model yang paling bagus untuk melakukan prediksi pembatalan pemesanan hotel, akan dilakukan lima kali percobaan dengan melakukan proses *hyperparameter tuning*. Untuk *dataset* akan dibagi menjadi 80% untuk *data training* dan 20% untuk *data test*. Pada penelitian ini, untuk mengetahui performa dari model yang diusulkan bisa diukur dengan beberapa nilai. Untuk mengetahui apakah model yang diusulkan ini bisa melakukan prediksi dengan benar, bisa diukur dengan nilai *accuracy*. Jika ingin mendapatkan nilai proporsi, berapa banyak data yang bisa diprediksi oleh model dengan nilai asli dari data yang tersedia bisa dilihat dengan nilai *confusion matrix* (Gambar. 3).

		Model Prediction	
		Positive	Negative
True Value	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Gambar 3. *Confusion Matrix*.

Menghasilkan nilai *True Positive* (TP), jika model mengeluarkan nilai positif dan *dataset* yang dituju juga mengeluarkan nilai positif. Sedangkan menghasilkan nilai *True Negative* (TN), jika model mengeluarkan nilai negatif dan *dataset* yang dituju mengeluarkan nilai negatif. Jika model mengeluarkan nilai positif tetapi *dataset* yang dituju mengeluarkan nilai negatif, maka nilai tersebut dinamakan *False Positive* (FP) (juga dikenal dengan *error tipe I*). Dan, jika model mengeluarkan nilai negatif tetapi *dataset* yang dituju mengeluarkan nilai positif, maka nilai tersebut dinamakan *False Negative* (FN) (*error tipe II*) (Strandberg & Låås, 2019). Berikut rumus untuk mengukur nilai *accuracy*, *precision*, dan *recall*.

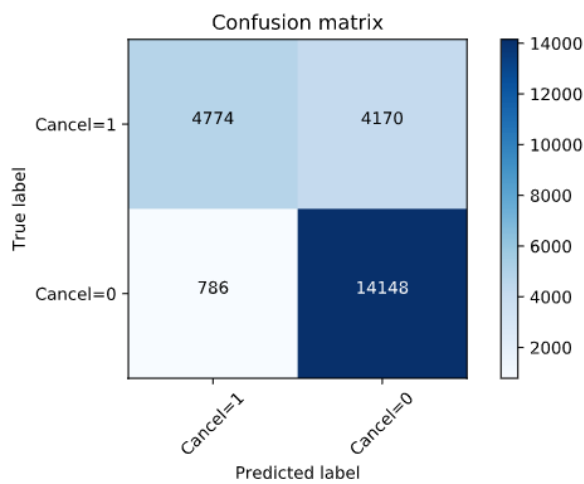
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$



$$Precision = \frac{TP}{TP+FP} \tag{5}$$

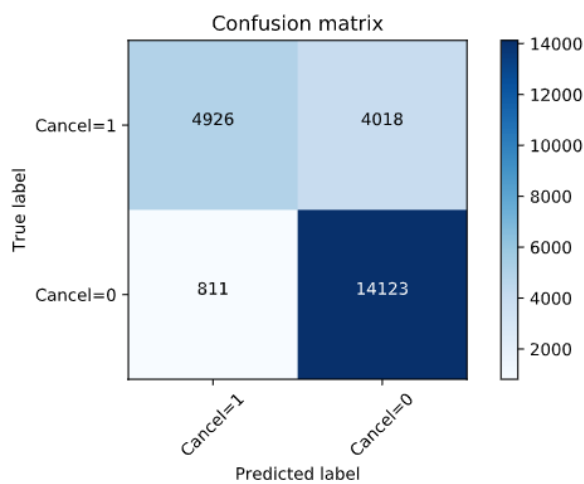
$$Recall = \frac{TP}{TP+FN} \tag{6}$$

Pada percobaan pertama, menggunakan metode ANN. Untuk arsitektur dari metode ANN yang diusulkan adalah seperti berikut: satu *input layer*, tiga *hidden layer* yang masing-masing terdiri dari 256 neuron yang menggunakan *relu activation*, *dropout layer* yang bernilai 0.5, dan satu *output layer* yang menggunakan *sigmoid activation* karena pada penelitian ini akan melakukan proses klasifikasi. Lalu menggunakan *optimizer Adam* dengan *learning rate* 0.0001, 50 *epochs*, menggunakan *loss function binary cross entropy*, dan menggunakan *early stop callbacks* yang berguna untuk menghentikan proses *training* apabila performa dari model tidak melakukan improvisasi. Dari percobaan pertama menghasilkan nilai 79.24% untuk *accuracy*, 85.86% untuk *precision* dan 53% untuk *recall*.



Gambar 4. Hasil Confusion Matrix ANN (Percobaan Pertama).

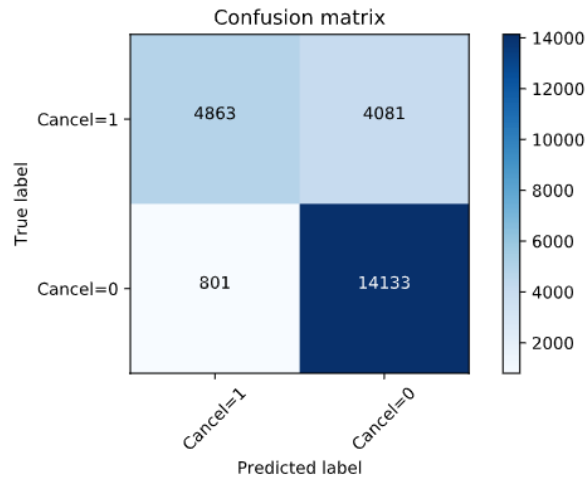
Pada percobaan kedua, menggunakan *logistic regression* dan menggunakan *default parameter* dari *library* bawaan. Dari percobaan kedua menghasilkan nilai 79.77% untuk *accuracy*, 85.86% untuk *precision* dan 55.07% untuk *recall*.



Gambar 5. Hasil Confusion Matrix Logistic Regression dengan Default Parameter (Percobaan Kedua).

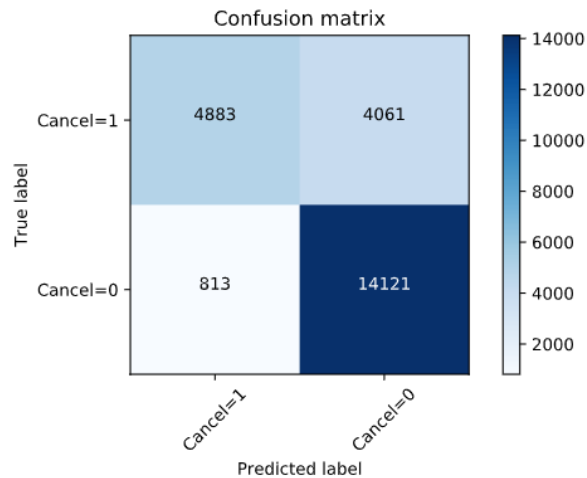


Pada percobaan ketiga, menggunakan *logistic regression* dan menggunakan *L1 penalty*. Dari percobaan ketiga menghasilkan nilai 79.55% untuk *accuracy*, 85.84% untuk *precision* dan 54.37% untuk *recall*.



Gambar 6. Hasil Confusion Matrix dengan L1 penalty (Percobaan Ketiga).

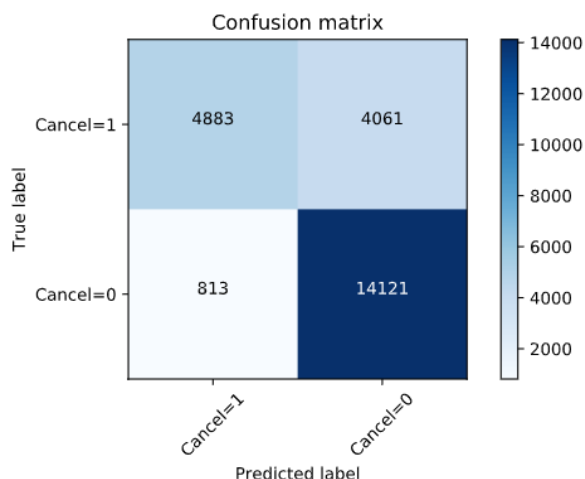
Pada percobaan keempat, menggunakan *logistic regression* dan menggunakan *L2 penalty* dan menggunakan '*newton-cg*' solver. Dari percobaan keempat menghasilkan nilai 79.58% untuk *accuracy*, 85.72% untuk *precision* dan 54.59% untuk *recall*.



Gambar 7. Hasil Confusion Matrix dengan L2 penalty (Percobaan Keempat).

Pada percobaan kelima, menggunakan *logistic regression* dan menggunakan *GridSearchCV*. Pada percobaan ini akan menggunakan *cross-validation* sebanyak 10 kali. Dari percobaan kelima menghasilkan nilai 79.78% untuk *accuracy*, 85.84% untuk *precision* dan 55.13% untuk *recall*.





Gambar 8. Hasil *Confusion Matrix* dengan *GridSearchCV* (Percobaan Kelima).

Untuk mengambil metode mana yang terbaik untuk melakukan prediksi pembatalan hotel, berikut tabel untuk melakukan perbandingan hasil dari percobaan pertama sampai percobaan kelima.

Tabel 6. Perbandingan Hasil dari Kelima Percobaan.

Model	Accuracy (%)	Precision (%)	Recall (%)
ANN (Percobaan Pertama)	79.24	85.86	53
<i>Logistic Regression</i> (Percobaan Kedua)	79.77	85.86	55.07
<i>Logistic Regression with L1 Penalty</i> (Percobaan Ketiga)	79.55	85.84	54.37
<i>Logistic Regression with L2 Penalty</i> (Percobaan Keempat)	79.58	85.72	54.59
<i>Logistic Regression with GridSearchCV</i> (Percobaan Kelima)	79.78	85.54	55.13

4. KESIMPULAN

Dari lima kali percobaan yang dilakukan telah didapatkan hasil yang paling optimal yaitu percobaan yang kelima, karena memiliki nilai *accuracy* yang paling tinggi. Tetapi jika diperhatikan lebih lanjut bahwa semua percobaan yang dilakukan menghasilkan nilai yang tidak terlalu signifikan perbedaannya di nilai *accuracy*, *precision* dan *recall*. Saran untuk peneliti selanjutnya bisa lebih ditekankan lagi pada proses *pre-processing* atau menggunakan metode klasifikasi yang lain, seperti SVM, *Naïve Bayes*, *Decision Tree*, dan sebagainya.

DAFTAR PUSTAKA

- Antonio, N., Almeida, A. de, & Nunes, L. (2017). Predicting Hotel Booking Cancellation to Decrease Uncertainty and Increase Revenue. *Tourism & Management Studies*, 13(2), 25–39. <https://doi.org/10.18089/tms.2017.13203>
- Antonio, N., de Almeida, A., & Nunes, L. (2019). Hotel booking demand datasets. *Data in Brief*, 22, 41–49. <https://doi.org/10.1016/j.dib.2018.11.126>
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.
- Bre, F., Gimenez, J. M., & Fachinotti, V. D. (2018). Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, 158, 1429–1441. <https://doi.org/10.1016/j.enbuild.2017.11.045>
- Chiang, W.-C., Chen, J. C. H., & Xu, X. (2007). An overview of research on revenue management: current issues and future research. *International Journal of Revenue Management (IJRM)*, 1(1), 97–128.
- Dunne, R. A. (2007). *A Statistical Approach to Neural Networks for Pattern Recognition*. Wiley-



- Interscience. <https://doi.org/10.1002/9780470148150>
- El_Jerjawi, N. S., & Abu-Naser, S. S. (2018). Diabetes Prediction Using Artificial Neural Network. *International Journal of Advanced Science and Technology*, 121, 55–64. <https://doi.org/10.14257/ijast.2018.121.05>
- Fujii, A., Tanaka, M., Yabushita, H., Mori, T., & Odashima, T. (2015). Detection of localization failure using logistic regression. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4313–4318. <https://doi.org/10.1109/IROS.2015.7353988>
- Lee, M. (2018). Modeling and forecasting hotel room demand based on advance booking information. *Tourism Management*, 66, 62–71. <https://doi.org/10.1016/j.tourman.2017.11.004>
- Rajopadhye, M., Ben Ghalia, M., Wang, P. P., Baker, T., & Eister, C. V. (2001). Forecasting uncertain hotel room demand. *Information Sciences*, 132(1–4), 1–11. [https://doi.org/10.1016/S0020-0255\(00\)00082-7](https://doi.org/10.1016/S0020-0255(00)00082-7)
- Ran, J., Zhang, G., Zheng, T., & Wang, W. (2018). Logistic Regression Analysis on Learning Behavior and Learning Effect Based on SPOC Data. *2018 13th International Conference on Computer Science & Education (ICCSE)*, 1–5. <https://doi.org/10.1109/ICCSE.2018.8468834>
- Rushin, G., Stancil, C., Sun, M., Adams, S., & Beling, P. (2017). Horse race analysis in credit card fraud—deep learning, logistic regression, and Gradient Boosted Tree. *2017 Systems and Information Engineering Design Symposium (SIEDS)*, 117–121. <https://doi.org/10.1109/SIEDS.2017.7937700>
- Santos, C. N. dos, Melnyk, I., & Padhi, I. (2018). Fighting Offensive Language on Social Media with Unsupervised Text Style Transfer. *56th Annual Meeting of the Association for Computational Linguistics*, 2, 189–194. <https://doi.org/10.18653/v1/p18-2031>
- Shamsaei, B., & Gao, C. (2016). Comparison of some machine learning and statistical algorithms for classification and prediction of human cancer type. *3rd IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2016*, 296–299. <https://doi.org/10.1109/BHI.2016.7455893>
- Strandberg, R., & Låås, J. (2019). *A comparison between Neural networks, Lasso regularized Logistic regression, and Gradient boosted trees in modeling binary sales.*
- Vono, M., Dobigeon, N., & Chainais, P. (2018). Sparse Bayesian binary logistic regression using the split-and-augmented gibbs sampler. *IEEE International Workshop on Machine Learning for Signal Processing, MLSP, 2018-Septe*, 1–6. <https://doi.org/10.1109/MLSP.2018.8516963>
- Wang, Y. H., Ou, Y., Deng, X. D., Zhao, L. R., & Zhang, C. Y. (2019). The Ship Collision Accidents Based on Logistic Regression and Big Data. *Proceedings of the 31st Chinese Control and Decision Conference, CCDC 2019*, 4438–4440. <https://doi.org/10.1109/CCDC.2019.8832686>
- Weatherford, L. R., Kimes, S. E., & Scott, D. A. (2001). Forecasting for hotel revenue management: Testing aggregation against disaggregation. *The Cornell Hotel and Restaurant Administration Quarterly*, 42(4), 53–64. [https://doi.org/https://doi.org/10.1016/S0010-8804\(01\)80045-8](https://doi.org/https://doi.org/10.1016/S0010-8804(01)80045-8)
- Yang, B., Wang, M., Xu, Z., & Zhang, T. (2019). Streaming Algorithm for Big Data Logistic Regression. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 2940–2950. <https://doi.org/10.1109/BigData.2018.8622392>
- Yang, Z., & Li, D. (2019). Application of Logistic Regression with Filter in Data Classification. *2019 Chinese Control Conference (CCC)*, 3755–3759. <https://doi.org/10.23919/ChiCC.2019.8865281>
- Zhang, Y. (2019). *FORECASTING HOTEL DEMAND USING MACHINE LEARNING APPROACHES.*
- Zou, X., Hu, Y., Tian, Z., & Shen, K. (2019). Logistic Regression Model Optimization and Case Analysis. *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, 135–139. <https://doi.org/10.1109/ICCSNT47585.2019.8962457>

