

BAB V

KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan terkait penerapan *Genetic Algorithm* terhadap kasus *feature selection*. Saran juga diberikan untuk penelitian selanjutnya untuk pengembangan dalam penerapan *Genetic Algorithm*. Penjelasan terkait kesimpulan dan saran akan dipaparkan lebih rinci pada subbab berikut.

V.1 Kesimpulan

Subbab ini menjelaskan tentang kesimpulan penerapan *Genetic Algorithm* dalam kasus *feature selection*, khususnya dalam melakukan prediksi *repurchase intention*. Kesimpulan yang didapat adalah sebagai berikut.

1. *Genetic Algorithm* yang dikembangkan berhasil digunakan untuk melakukan prediksi *repurchase intention*. *Genetic Algorithm* digunakan untuk mencari *feature subset*, dimana dalam menghitung *fitness value* dari *feature subset* menggunakan algoritma *machine learning*, yaitu *decision tree*. Terdapat 2 ukuran yang diusulkan peneliti pada penelitian ini, yaitu *feature importance* dan nilai proporsi kecenderungan. Penerapan *feature importance* digunakan dalam membantu proses *crossover*, dan penerapan proporsi kecenderungan digunakan dalam membantu proses *mutation*.
2. Terdapat 10 *features* yang paling mempengaruhi keputusan *repurchase* konsumen terhadap produk kosmetik, untuk kasus “cleanser” dan kasus “treatment”. *Features* yang paling mempengaruhi untuk kasus “cleanser” adalah review, warna, facial, repurchase, sociolla, botol, cocok, product, pokok, dan bekas. *Features* yang paling mempengaruhi untuk kasus “treatment” adalah pokok, terpilih, bekas, kering, panas, cocok, serum, bagus, asa, dan lembab,
3. Parameter yang digunakan untuk membandingkan performansi *Genetic Algorithm* adalah jumlah kromosom dalam populasi (kromosom), besar *mutation rate* (*mutation_rate*), dan besar *crossover rate* (*crossover_rate*).

Nilai yang digunakan untuk parameter kromosom adalah 15 dan 30; nilai pada parameter *mutation_rate* adalah 0,01 , 0,05, dan 0,1 ; nilai pada parameter *crossover_rate* untuk kasus “cleanser” adalah 0,5, 0,6, 0,7, 0,85 dan nilai parameter *crossover_rate* untuk kasus “treatment” adalah 0,6, 0,7, 0,85,dan 0,9. Semakin besar nilai parameter yang digunakan, maka performansi GA akan semakin baik. Nilai parameter yang terpilih dalam untuk kedua kasus merupakan nilai parameter tertinggi. Kombinasi parameter terpilih untuk kasus “cleanser” adalah kromosom bernilai 30, *mutation_rate* bernilai 0,1, dan *crossover_rate* bernilai 0,85 , sedangkan kombinasi parameter terpilih untuk kasus “treatment” adalah kromsom bernilai 30, *mutation_rate* berniali 0,1, dan *crossover_rate* bernilai 0,9.

V.2 Saran

Saran diberikan berdasarkan penelitian yang telah dilakukan. Saran yang diusulkan dapat digunakan untuk penelitian selanjutnya terkait *Genetic Algorithm*, terutama penyelesaian kasus *feature selection*. Beberapa saran yang diusulkan adalah sebagai berikut.

1. Mengkombinasikan GA dengan algoritma *machine learning* lain untuk mendapatkan solusi yang lebih baik dan pencarian tidak terjebak dalam solusi prematur.
2. Mencoba kombinasi semua nilai yang memungkinkan dalam parameter GA, untuk mendapatkan kombinasi terbaik, agar solusi yang diperoleh lebih baik.

DAFTAR PUSTAKA

- Blum, A., Hopcroft, J., & Kannan, R. (2018). *Foundations of Data Science*. Cambridge: Cambridge University Press
- Brid, R. S. (2018). "Decision Trees—A simple way to visualize a decision", <https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb>, diakses 17 Juni 2020 pukul 12.44
- Brownlee, J. (2019). "An Introduction to Feature Selection", <https://machinelearningmastery.com/an-introduction-to-feature-selection>, diakses 15 Juni 2020 pukul 18.24.
- Brownlee, J. (2020). "Tour of Evaluation Metrics for Imbalanced Classification", <https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification>, diakses 4 Juni 2020 pukul 10.12
- Ciortan, M. (2019). "Overview of feature selection methods - Towards Data Science", <https://towardsdatascience.com/overview-of-feature-selection-methods-a2d115c7a8f7>, diakses pada 24 Mei 2020 pukul 13.33
- De Jong, K. A., & Spears, W. M. (1991). An analysis of the interacting roles of population size and crossover in genetic algorithms. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 496 LNCS, 38–47. doi : 10.1007/BFb0029729
- Gad, A. (2018). "Introduction to Optimization with Genetic Algorithm", <https://towardsdatascience.com/introduction-to-optimization-with-genetic-algorithm-2f5001d9964b>, diakses pada 24 Mei 2020 pukul 13.37
- Gen, M., & Cheng, R. (1996). Genetic Algorithms and Engineering Design. In *Genetic Algorithms and Engineering Design*. doi : 10.1002/9780470172254
- Grefenstette, J. J. (1986). Optimization of Control Parameters for Genetic Algorithms. *IEEE Transactions on Systems, Man and Cybernetics*, 16(1), 122–128. doi : 10.1109/TSMC.1986.289288
- Harvey, H. B., & Sotardi, S. T. (2018). The Pareto Principle. *Journal of the American*

- College of Radiology*, 15(6), 931. doi : 10.1016/j.jacr.2018.02.026
- Helal, M. Al, Haydar, M. S., & Mostafa, S. A. M. (2017). Algorithms efficiency measurement on imbalanced data using geometric mean and cross validation. *IWCI 2016 - 2016 International Workshop on Computational Intelligence*, 110–114. doi : 10.1109 /IWCI.2016.7860349
- Hellier, P. K., Geursen, G. M., Carr, R. A., & Rickard, J. A. (2003). Customer repurchase intention. *European Journal of Marketing*, 37(11/12), 1762–1800. doi : 10.1108/0309 0560310495456
- Homsapaya, K., & Sornil, O. (2017). Improving floating search feature selection using genetic algorithm. *Journal of ICT Research and Applications*, 11(3), 299–317. doi : 10.5614/itbj.ict.res.appl.2017.11.3.6
- Ibzan, E., Balarabe, F., & Jakada, B. (2016). *Consumer Satisfaction and Repurchase Intentions*. 6(2).
- Kadhim, A. I. (2018). An Evaluation of Preprocessing Techniques for Text Classification. *International Journal of Computer Science and Information Security*, 16(6), 22–32.
- Kamal, P., & Ahuja, S. (2019). Data Mining Techniques : Identification of Influential Factors Effecting the Academic Performance in Undergrad. *Harmony Search and Nature Inspired Optimization Algorithms*, 741(January), 835–843. doi : 10.1007/978-981-13-0761-4
- Kaytan, M., & Aydilek, I. B. (2017). *A review on machine learning tools*. October, 1–4. doi : 10.1109/idap.2017.8090257
- M, H., & M.N, S. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2), 01–11. doi : 10.5121/ijdkp.2015.5201
- Ma, B., & Xia, Y. (2017). A Tribe Competition-Based Genetic Algorithm for Feature Selection in Pattern Classification. *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*
- Nithya, V. I. (2016). *Preprocessing Techniques for Text Mining Preprocessing Techniques for Text Mining*. 5(October 2014), 7–16.

- Patel, B. N. (2012). Efficient Classification of Data Using Decision Tree. *Bonfring International Journal of Data Mining*, 2(1), 06–12. doi : 10.9756/bijdm.1098
- Pavlenko, T. (2003). On feature selection, curse-of-dimensionality and error probability in discriminant analysis. *Journal of Statistical Planning and Inference*, 115(2), 565–584. doi : 10.1016/S0378-3758(02)00166-0
- Refaeilzadeh, P., Tang, L., & Liu, H. (2008). *C Cross-Validation*. Arizona : Arizona State University
- Remesan, R., & Mathew, J. (2015). Hydrological data driven modelling: A case study approach. In *Hydrological Data Driven Modelling: A Case Study Approach*. Springer International Publishing. doi : 10.1007/978-3-319-09235-5
- Shaharane, I. N. M., & Hadzic, F. (2015). Feature Selection for Data and Pattern Recognition. In U. Stańczyk & L. C. Jain (Eds.), *Studies in Computational Intelligence* (Vol. 584). Springer Berlin Heidelberg. doi : doi.org/10.1007/978-3-662-45620-0
- Shahiri, A. M., Husain, W., & Rashid, N. A. (2017). A proposed framework on hybrid feature selection techniques for handling high dimensional educational data. *AIP Conference Proceedings*, 1891, 020130. doi : 10.1063/1.5005463
- Shao, J. (1993). Linear Model Selection by Cross-Validation. In *Source: Journal of the American Statistical Association* (Vol. 88, Issue 422).
- Sugiyono. (2007). *Statistik Untuk Penelitian.pdf* (pp. 1–370). doi : 10.1016/S0969-4765(04)00066-9
- Wang, L., Ni, H., Yang, R., Pappu, V., Fenn, M. B., & Pardalos, P. M. (2014). Feature selection based on meta-heuristics for biomedicine. *Optimization Methods and Software*, 29(4), 703–719. doi : 10.1080/10556788.2013.834900
- Wei, Q., & Dunbrack, R. L. (2013). The Role of Balanced Training and Testing Data Sets for Binary Classifiers in Bioinformatics. *PLoS ONE*, 8(7), e67863. doi : 10.1371/journal.pone.0067863
- Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*. doi : 10.1007/BF00175354
- Xu, Q. S., & Liang, Y. Z. (2001). Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems*, 56(1), 1–11. doi : 10.1016/S0169-

7439(00)00122-2

Yaghini, M. (2009). *What is a Metaheuristic?*. Diunduh dari http://webpages.iust.ac.ir/yaghini/Courses/AOR_872/What%20is%20a%20Metaheuristic.pdf

Yusta, S. C. (2009). Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters*, 30(5), 525–534.