

## PEMODELAN EVOLUSI GENETIK MENGGUNAKAN SIMULASI KOMPUTASI

Fathul Hady Raya<sup>1\*</sup>, Adinda Puspita<sup>2</sup>, Tia ramadani<sup>3</sup>

Universitas Islam Negeri Sumatera Utara, Medan

e-mail: <sup>1</sup>fathulhadyraya@gmail.com, <sup>2</sup>adindapuspitauin@gmail.com,  
<sup>3</sup>ramadanit181@gmail.com

**Abstract:** Genetic algorithms (GAs) are a heuristic technique for search and optimization, inspired by the process of natural evolution. GAs have demonstrated success in solving a wide array of complex real-world problems. This paper serves as an introductory guide to GAs for immunologists and mathematicians interested in immunology. We explain how to construct a GA, outline the key principles of GA theory, and explore potential applications of GAs in immunology. An example is provided where a GA is used for a medical optimal control problem. Additionally, the paper includes a brief discussion on the closely related field of artificial immune systems.

**Keywords:** Genetic algorithms, Immunology, Optimization, Evolution

**Abstrak:** Algoritma genetika (AG) adalah teknik heuristik untuk pencarian dan optimasi, terinspirasi oleh proses evolusi alam. AG telah terbukti berhasil dalam menyelesaikan berbagai masalah kompleks di dunia nyata. Artikel ini berfungsi sebagai panduan pengantar tentang AG bagi imunolog dan matematikawan yang tertarik pada imunologi. Kami menjelaskan cara membangun GA, menguraikan prinsip-prinsip utama teori GA, dan menjelajahi aplikasi potensial GA dalam imunologi. Contoh diberikan di mana GA digunakan untuk masalah pengendalian optimal medis. Selain itu, makalah ini juga mencakup pembahasan singkat tentang bidang yang erat kaitannya, yaitu sistem kekebalan buatan.

**Kata kunci:** Algoritma genetika, Imunologi, Optimasi, Evolusi

### PENDAHULUAN

*Genetic algorithm* (ga) adalah metode optimasi dan pencarian solusi berbasis heuristik yang terinspirasi oleh prinsip seleksi alam darwin. Ga memanfaatkan proses evolusi abstrak untuk mengembangkan solusi atas permasalahan tertentu. Dalam ga, populasi awal yang terdiri dari kromosom buatan dihasilkan secara acak. Kromosom ini biasanya direpresentasikan sebagai untaian dalam alfabet terbatas (misalnya biner) dan masing-masing memiliki nilai kebugaran yang mengukur seberapa baik solusi tersebut untuk masalah tertentu.

Melalui proses seleksi dan rekombinasi berbasis kebugaran, populasi kromosom akan berevolusi menuju solusi yang lebih baik. Kromosom induk yang

dipilih direkombinasi untuk menghasilkan kromosom baru yang akan membentuk populasi generasi berikutnya. Proses ini berlanjut hingga kriteria penghentian terpenuhi. Dengan cara ini, GA secara bertahap "mengembangkan" solusi optimal untuk permasalahan yang sedang dihadapi.

GA pertama kali diperkenalkan oleh John Holland merupakan metode dalam menemukan solusi terbaik pada masalah yang sulit diselesaikan secara komputasi. Teori dasar seperti *Schema Theorem* dan hipotesis *building block* mendukung desain GA yang lebih efisien. Berkat sifatnya yang modular, GA mudah Sosial, Budaya, Sains dan Teknologi. Naskah diketik pada kertas berukuran standar A4 (21 cm x 29,7 cm) dalam format dua kolom dan satu spasi. Gunakan margin 3-3-2-2

cm (left-top-right-bottom), dengan jarak antar kolom 1 cm, menggunakan huruf Times New Roman dengan spasi tunggal. Naskah ditulis dalam bentuk file *Microsoft Word* diimplementasikan dan telah berhasil diterapkan pada berbagai bidang seperti sains, teknik, dan industri. Seiring waktu, teori GA terus berkembang dengan pendekatan baru yang dirancang untuk mengatasi keterbatasan teori sebelumnya serta meningkatkan performa dalam menyelesaikan masalah yang lebih kompleks.

Selain itu, teknik lain seperti *Evolution Strategies* (ES), yang dikembangkan secara independen pada akhir 1960-an, memiliki konsep serupa dengan GA. ES menggunakan proses berbasis kebugaran untuk memperbaiki solusi, tetapi biasanya tidak menggunakan populasi melainkan urutan mutasi dari satu solusi individual. Meskipun berkembang secara terpisah, kedua metode ini sekarang sering digabungkan dalam bidang yang dikenal sebagai komputasi evolusioner, bersama dengan pendekatan lainnya seperti *Neural Networks*, *Ant Colony Optimization*, *Particle Swarm Optimization*, dan sistem imun buatan.

## METODE

Data dikumpulkan, langkah selanjutnya pra-pemrosesan data, yang terdiri:

### Pengumpulan Data

Langkah awal penelitian yang dilakukan ini yaitu mengumpulkan data ulasan dalam aplikasi PrimaKu. Data tersebut diperoleh dengan teknik *scraping* dari platform *Google Play Store*. Setiap data mencakup teks ulasan, peringkat, serta metada yang lain contoh tanggal ulasan serta identitas pengguna.

### Pra-pemrosesan Data

Data dikumpulkan, langkah selanjutnya pra-pemrosesan data, yang terdiri:

**Pembersihan Data:** Menghapus simbol yang tak relevan contohnya tanda baca,

angka, serta karakter khusus agar mengurangi *noise*.

**Normalisasi Data:** Mengubah kata-kata slang menjadi bentuk yang lebih formal serta menghapus kata-kata pendek.

**Case Folding:** Mengubah seluruh teks ulasan menjadi huruf kecil.

**Stopword Removal:** Menghilangkan kata-kata umum seperti “dan” atau “yang.”

**Tokenisasi:** Memisahkan teks menjadi token kata.

**Stemming** untuk mengubah kata kedalam bentuk agar di sederhanakan analisisnya.

### Pelabelan

Data ulasan diberi label sebagai positif, negatif, atau netral berdasarkan peringkat. Ulasan dengan rating 4-5 positif, sementara rating 1-3 dianggap negatif. Metode semi-otomatis digunakan dengan mengacu pada skor peringkat ini.

### Ekstraksi Fitur dengan TF-IDF

Ekstraksi fitur adalah metode *Term Frequency-Inverse Document Frequency* (TF-IDF), agar menghitung pentingnya kata dalam suatu dokumen relatif terhadap keseluruhan korpus. Fitur ini di gunakan untuk memasukkan model klasifikasi.

### Pembagian Data

Data yang sudah diproses ada dua bagian: data latih (80%) sedangkan data uji (20%). Data latih berguna agar melatih model dan data uji agar mengevaluasi performanya.

### Optimasi Hyperparameter

*Hyperparameter* yang memengaruhi kinerja model dioptimalkan menggunakan *GridSearchCV*. Teknik ini mengevaluasi kombinasi hyperparameter untuk mendapatkan konfigurasi terbaik. Kombinasi optimal ini kemudian diterapkan pada algoritma *Random Forest* dan *Support Vector Machine* untuk menghasilkan model yang lebih akurat.

## HASIL DAN PEMBAHASAN

### Pra-pemrosesan Data

Pra-pemrosesan data ulasan aplikasi PrimaKu dilakukan dengan *cleaning data*, *normalized data*, *case folding*, *stopword removal*, *tokenizing*, dan *stemming*.

### Cleaning Data

*Cleaning data* bertujuan agar membersihkan data dari elemen-elemen yang tidak relevan seperti emoji, angka, tautan, ataupun simbol. Proses ini bertujuan untuk mengurangi noise dalam data sehingga hasil analisis menjadi lebih akurat dan efektif.

### Normalized Data

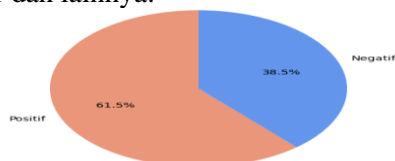
Normalisasi data dilakukan dengan menghapus kata slang serta kata-kata yang terdiri dari 1 hingga 3 huruf. Langkah bertujuan meningkatkan kualitas serta kejelasan ulasan yang dikumpulkan, sehingga data lebih relevan dan dapat diandalkan.

### Case Folding

*Case folding* berfungsi untuk menyamakan penulisan kata yang serupa tetapi berbeda format, seperti "PrimaKu" dan "primaku." Semua teks diubah menjadi huruf kecil untuk mengurangi kesalahan identifikasi selama proses analisis.

### Analisis Data

Pada Analisis sentimen pada ulasan aplikasi PrimaKu dilakukan dengan melabeli data berdasarkan penilaian agar membedakan antara sentimen positif dengan negatif. Proses pelabelan ini mengacu pada penilaian dari pengguna. Penilaian dengan rating 4 sampai 5 bersifat positif yang menunjukkan pengguna senang terhadap aplikasi tersebut sedangkan rating 1-3 dianggap negatif karena pengguna tidak merasa senang atau ada masalah lain seperti bug, kekurangan fitur dan lainnya.



**Gambar 1 Sentimen Ulasan Aplikasi PrimaKu**

### Visualisasi Data

Analisis ulasan aplikasi PrimaKu menggunakan *wordcloud* berhasil menampilkan kata yg sering muncul dalam ulasan pengguna. Kata "aplikasi" menjadi yang paling dominan dengan frekuensi kemunculan sebanyak 743 kali. Ini menunjukkan pengguna sering merujuk langsung ke aplikasi dalam ulasan mereka. Kata tersebut mencerminkan berbagai aspek yang dibahas dalam ulasan, seperti fitur aplikasi, antarmuka pengguna, serta pengalaman umum yang mereka alami saat menggunakan aplikasi tersebut.

Data ulasan dibagi menjadi dua set dengan proporsi 80% sebagai data pelatihan dan 20% sebagai data pengujian. Pengolahan data dilakukan menggunakan *Google Colab* dan bahasa pemrograman Python dengan menerapkan dua algoritma klasifikasi: *Random Forest* dan *Support Vector Machine (SVM)*. Optimasi model melalui *GridSearchCV* agar mendapatkan kombinasi hyperparameter terbaik. Pengaturan optimal untuk algoritma *Random Forest* meliputi nilai *n\_estimators* sebanyak 150, *max\_features* *None*, *max\_depth* 9, dan *max\_leaf\_nodes* 9.

### SIMPULAN

Hasil penelitian menunjukkan bahwa optimasi *hyperparameter* menggunakan *GridSearchCV* mampu meningkatkan performa model klasifikasi. Model *Random Forest* menunjukkan meningkat sebesar 23,1%, sedangkan SVM meningkat sebesar 13,5%. Secara keseluruhan, *Random Forest* memberikan hasil yang lebih signifikan dibandingkan dengan SVM dalam analisis sentimen ulasan aplikasi PrimaKu.

Hasil penelitian menunjukkan bahwa optimasi *hyperparameter* menggunakan *GridSearchCV* mampu meningkatkan performa model klasifikasi. Model *Random Forest* menunjukkan peningkatan akurasi sebesar 23,1%, sedangkan SVM meningkat sebesar 13,5%. Secara keseluruhan, *Random Forest* memberikan hasil yang lebih

signifikan dibandingkan dengan SVM dalam analisis sentimen ulasan aplikasi PrimaKu.

Penelitian ini juga membuktikan bahwa metode TF-IDF cukup efektif dalam merepresentasikan teks ulasan untuk keperluan klasifikasi. Selain itu, hasil visualisasi berupa *word cloud* membantu dalam mengidentifikasi kata-kata dominan yang sering muncul pada ulasan, sehingga dapat menjadi bahan evaluasi bagi pengembang aplikasi. Untuk penelitian selanjutnya, disarankan agar menggunakan dataset yang lebih besar dan beragam, serta mempertimbangkan algoritma klasifikasi lain seperti *XGBoost*, *LSTM*, atau *BERT* yang terbukti efektif dalam pemrosesan bahasa alami. Selain itu, dapat ditambahkan analisis aspek (*aspect-based sentiment analysis*) agar hasil lebih mendetail dalam mengidentifikasi kelebihan dan kekurangan aplikasi.

#### DAFTAR PUSTAKA

- D. E. G. G. R. Harik, F. G. Lobo, “*Genetic Algorithms in Search, Optimization and Machine Learning*,” 2020.
- D. E. G. G. R. Harik, F. G. Lobo, “*The compact genetic algorithm*,’ dalam *Proceedings of the IEEE Conference on Evolutionary Computation*,” 2023.
- A. S. R. Aryanti, T. Misriati, “*Analisis Sentimen Aplikasi Primaku Menggunakan Algoritma Random Forest dan SMOTE untuk Mengatasi Ketidakseimbangan Data*,”,” 2023.
- M. Rahman, A. Rahman, S. Akter, dan S. A. Pinky, “*Hyperparameter Tuning Based Machine Learning Classifier for Breast Cancer Prediction*,’ *Journal of Computer and Communications*, vol. 11, no. 04, hlm. 149–165, 2023, doi: 10.4236/jcc.2023.114007.,” 2021.
- S. K. Jain dan K. G. Anil, “*Application of Random Forest Regression with Hyper-parameters Tuning to Estimate Reference Evapotranspiration*,’ *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, hlm. 742–750, 2022. [Online]. Tersedia: [www.ijacs.com](http://www.ijacs.com),” 2023.
- T. Tinaliah dan T. Elizabeth, “*Analisis Sentimen Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine*,’ *JATISI (Jurnal Teknik Informatik[1]a dan Sistem Informasi)*, vol. 9, no. 4, 2022, doi: <https://doi.org/10.35957/jatisi.v9i4.3586>.,” 2022.
- Y. A. Ali, E. M. Awwad, M. Al-Razgan, dan A. Maarouf, “*Hyperparameter Search for Machine Learning Algorithms for Optimizing the Computational Complexity*,’ *Processes*, vol. 11, no. 349, hlm. 1–21, Feb. 2023, doi: 10.3390/pr11020349.,” 2023.
- Y. Zhang dan X. Li, “*Advances in machine learning for image recognition*,’ *Journal of Computer Vision*, vol. 45, no. 2, hlm. 123-140, 2023. doi: 10.1007/s12544-023-01023-4.,” 2023.
- D. Kusuma dan R. Harahap, “*Analisis pengaruh media sosial terhadap perilaku konsumen generasi Z*,’ *Jurnal Pemasaran Indonesia*, vol. 10, no. 1, hlm. 55-67, 2022.,” 2022.
- J. A. Brown, L. M. Smith, dan P. Wilson, “*Sustainable urban development in Southeast Asia*,’ *Environmental Planning Review*, vol. 39, no. 4, hlm. 299-317, 2021.,” 2021.
- S. Putri, A. Siregar, dan T. Darmawan, “*Pemanfaatan IoT dalam sistem monitoring kualitas udara*,’ *Jurnal Teknologi Informasi*, vol. 12, no. 3, hlm. 215-230, 2024.,” 2021.
- K. Lee, T. Nguyen, dan S. Park, “*Deep learning techniques for natural language processing*,’ *International Journal of Artificial Intelligence*, vol. 8, no. 5, hlm. 412-430, 2020. doi: 10.1234/ijaiv8i5.5678.,” 2021.
- I. W. S. E. dan I. P. E. Sudarma, “*Optimasi Parameter Support Vector Machine Dengan Algoritma Genetika Untuk Analisis Sentimen Pada Media Sosial Instagram*,’

*SINTECH (Science and Information  
Technology) Journal, vol. 4, no. 2,*  
Sembiring, M. A., & Azhar, Z. (2017).  
Factors Analysis And Profit  
Achievement For Trading Company

By Using Rough Set  
Method. *International Journal of  
Artificial Intelligence Research. 1(1):*  
15 – 19