JSCE (Journal of System and Computer Engineering)
Vol.6, No.1, January 2025, pp. 144-150
ISSN (online): 2723-1240
DOI : https://doi.org/10.61628/jsce.v6i1.1636

Parameter Optimization Supports Vector Machine Using Genetic Algorithms to Improve the Efficiency of Data Transfer Prediction on Google Cloud

Respaty Namruddin [1,a]; Ricky Mahendra [2,b]; Aang Kunaefi [3,c,*]; Ramlah Abu Bakar [4,d];
[1,2,3,4] Universitas Handayani, Adiyaksa Baru No.1 , Makassar and 90213, Indonesia
[a] respatynamruddin@handayani.ac.id; [b] rickymahendra465@gmail.com; [c] aan.mks@gmail.com ;[d] ramlahabubakar77@gmail.com
** Corresponding author*

Abstract

*Data transfer efficiency is a key element in cloud infrastructure like Google Cloud. This study aims to improve the accuracy of data transfer efficiency prediction using a Support Vector Machine (SVM) optimized with Genetic Algorithm (GA). The dataset contains information about file size, network latency, server utilization, and data transfer time. The Genetics algorithm is applied to find the optimal parameters, namely the **C** and **gamma** values. The results show that parameter optimization using GA can increase prediction accuracy by up to 90%, compared to the traditional Grid Search method which achieves a maximum accuracy of 88%.*

***Keywords**— Support Vector Machine (SVM), Genetic Algorithm (GA), Parameter Optimization, Data Transfer Efficiency, Google Cloud*

1. Introduction

In the era of rapidly developing digital technology, cloud computing has become the main foundation for supporting various modern computing needs. Services like Google Cloud offer a wide range of storage, data processing, and accessibility capabilities. However, one of the challenges that is still faced is the efficiency of data transfer. Slow or inefficient data transfer not only impacts service quality but can also lower user satisfaction levels, ultimately affecting the reputation and competitiveness of cloud service providers (Safii et al., 2023) therefore, improving data transfer efficiency is a priority in optimizing cloud computing services.

Some Factors that affect the efficiency of data transfer in a cloud environment include file size, network latency, and server utilization rates. Large file sizes often extend transfer times, while network latency, which is determined by infrastructure conditions and geographical distance, can add complexity in maintaining transfer stability and speed. On the other hand, high server utilization rates can lead to bottlenecks, thus hindering the data transfer process. Taking these factors into account, a solution that can predict and improve transfer efficiency effectively is needed.

To overcome these problems, machine learning-based approaches have shown great potential. One of the algorithms is the support vector machine (SVM). SVM can classify data precisely, so it is often used in various fields including prediction of data transfer efficiency. However, SVM performance is highly dependent on parameters such as C (regularization) and gamma (kernel parameters). Improper selection of parameters can lead to a decrease in model accuracy, thus posing challenges in the implementation of SVM for real cases. (MARGARITA, 2024).

To overcome these obstacles, this study adopts the Genetic Algorithm (GA) as the SVM parameter optimization method. GA is an evolution-based algorithm inspired by the natural selection process. Iterating on the selection, crossover, and mutation processes, GA can effectively explore the search space to find the best combination of parameters. This method has been widely used in various optimization applications due to its ability to handle problems with many variables and complex search spaces. (Yunita et al., 2023).

In addition to offering parameter optimization solutions, the use of GA in this study aims to improve the accuracy of prediction and overall system efficiency. By integrating GA into the SVM modeling process, it is hoped that a prediction model that is more adaptive to changes in network conditions can be produced so that it can provide consistent performance in various scenarios. This approach is also relevant to the needs of cloud computing services, which continue to evolve as the number of users and data complexity increases (Sulartopo et al., 2023).

Related research also shows that combining SVM with optimization algorithms, such as Genetic Algorithms, is able to provide better results than conventional SVM use. This provides a strong foundation for adopting this approach in the Google Cloud case study. By utilizing GA to set SVM parameters, the prediction model can be dynamically adapted for different dataset sizes and network conditions

2. Method

Genetic Algorithm is one of the metaheuristic-based computational methods that mimics the process of biological evolution in solving various optimization problems or finding solutions. This method is inspired by the principle of natural selection from Darwin's theory of evolution, where the best individuals in a population have a greater chance of surviving and producing offspring (Arka, 2021).

In Genetic Algorithm, the solution to a problem is represented as an individual in the form of chromosomes. This individual population undergoes an evolutionary process through operations such as selection, crossover, and mutation to produce a new generation with better solutions. This process continues iteratively until the criteria stop being met, such as achieving an optimal solution of a certain number of iterations (Isai & Nugroho, 2024).

The main advantage of Genetic Algorithms is their ability to efficiently explore large and complex solution search spaces, so they are often used to solve non-linear, discrete, or multivariate problems. This method is widely applied in various fields such as optimization, planning, scheduling, network design, and machine learning (Mukhlis, 2024).

2.1 Dataset

The Dataset consists of 10 sample data transfers in Google Cloud, with the following variables:

- **File_Size_MB**: File size in megabytes (Wijaya et al., 2024)
- **Network_Latency_ms**: Network in miliseconds (Farhan & Setiaji, 2023)
- **Server_Utilization_%**: Server utillization rate in percent (Hatta, 2011)
- **Data_Transfer_Time_s**: Data transfer time in seconds (A. Antu et al., 2020)

Predicted_Class: Classification of data transfer efficiency (Qulub & Agustin, 2024), Is ”Efficient” and ”Inefficient”.

Table 1. Google Cloud Data Transfer Sample Data

No.	File_Size_ MB	Network_ Latency_ms	Server_ Utilization_%	Data_ Transfer_Time_s	Predicted_ Class
1.	8500	180.50	70.00	20.00	Efficient
2.	3500	120.75	65.00	10.50	Efficient
3.	2200	50.25	85.00	30.25	Inefficient
4.	9500	160.00	60.00	18.00	Efficient
5.	4500	130.50	75.00	22.50	Inefficient
6.	7500	140.25	68.00	15.75	Efficient
7.	3200	110.50	72.00	12.50	Efficient
8.	5000	90.75	80.00	25.00	Inefficient

DOI : https://doi.org/10.61628/jsce.v6i1.1636

9.	4200	115.00	65.50	13.75	Efficient
10.	6100	150.50	70.50	20.75	Efficient

2.2 Data Proccesing

- Data Cleanup: Empty data is imputed using an average
- Data Normalization: All numerical variables are normalized within the range [0,1] to avoid bias in the model training process.

2.3 Support Vector Machine (SVM)

SVM is used as a prediction model with the Radial Basis Funtion (RBF) kernel. The functions of the RBF kernel are:

$$K(x_i,x_j) = \exp\left(-\gamma\left|\left|x_i - x_j\right|\right|^2\right)$$

Calculations with Sample Data :
Suppose two data have data :

- $x_i = [0.85, 0.7, 0.6]$
- $x_j = [0.9, 0.65, 0.55]$

With γ is 0,05

$$||x_i - x_j||^2 = (0.85 - 0.9)^2 + (0.7 - 0.65)^2 + (0.6 - 0.55)^2 = 0.0025$$

$$K_{(x_i,x_j)} = \exp(-0.05 \times 0.0075) = \exp(-0.000375) \approx 0.999625$$

2.4 Parameter Optimization With Genetic Algorithm (GA)

GA is used to determine the optimal parameters:

- Initial Population: Randomly form a parameter population (Fikriansyah et al., 2023).
- Fitness evaluation: using the following functions:

$$Fitness = \frac{Jumlah\ Prediksi\ benar}{Total\ Data}$$ (Feronica et al., 2022)

- Selection, crossover, and mutation: the best individual selection process and generate new parameters (Yunita et al., 2023).

- Iteration: Repeated until optimal parameters are found (Sutomo et al., 2022).

3. Results And Discussion

3.1 Parameter Optimization Result

3.1.1 Use of Genetic Algorithm (GA) :

The optimization process begins by initializing the Support Vector Machine (SVM) parameters using the Genetic Algorithm. The purpose of this optimization is to find the values of the C and gamma parameters that produce the model with the best prediction accuracy.

- **Initial Step:**

The process begins by initializing the initial population, which is a randomly selected set of solution candidates in the C and gamma parameter spaces. The value of C is set in the range [0.1, 10], while the gamma is set in the range [0.001, 1].

- **Evaluasi Fitness:**
  Each individual (combination of C and gamma) was evaluated using the fitness function based on SVM prediction accuracy on the validation data. The higher the accuracy, the better the individual's fitness score.

3.1.2 Optimization iteration :

In this study, 50 iterations were carried out. Genetic algorithms process the initial population, and perform selection, crossover, and mutation to produce individuals with the best parameters in each generation (Saputro et al., 2015).

- **Selection:** Individuals with the best fitness scores are selected to produce offspring in the next generation (Yusron Mubarok & Chotijah, 2021).
- **Crossover**: The best combination of two individuals is used to form a new individual, in the hope of producing a better combination of parameters (R. R. Pratama, 2020).
- **Mutation**: New individuals undergo slight random changes to maintain population diversity and avoid being trapped in optimal localities (Farhan & Setiaji, 2023).

3.1.3 Optimal Parameter :

After the 50th iteration, the optimal parameters are found:

- C:2.0 This value controls the hyperplane margin and the penalty for classification errors.
- Gamma: 0.05 This parameter affects the effect of a data on the hyperplane. Small values on gamma indicate a wider influence

The results of the accuracy of the model with these parameters are:

$$Accuracy = \frac{Prediksi\ Benar}{Total\ Data} \times 100 = \frac{9}{10} \times 100 = 90\%$$

3.1.4 Model Accuracy :

With these parameters, the SVM model produces an accuracy of 90%, indicating an increase in prediction efficiency compared to other methods.

3.2 Result Visualization

3.2.1 Graph of File Size and Transfer Time Relationship :

The graph shows the relationship between the size and transfer time variables, which are classified based on their level of efficiency. The results of the SVM model classification are depicted through colors that indicate the level of transfer efficiency:

- Green: Efficient Transfer
- Red: Inefficient Transfer
  This classification reflects the model's ability to predict whether or not data transfer is efficient.

3.2.2 Convergence of Genetic Algorithms :

Convergence graphs show an increase in accuracy with each iteration. After 30 iterations, the accuracy value starts to stabilize until it reaches a maximum value of 90% on the 50th iteration. This ensures the effectiveness of the Genetic Algorithm in finding the optimal solution with an efficient number of iterations.

3.2.3 Comparison with Grid Search

The following table is a comparison of the results of the Genetic Algorithm and Grid Search methods:

Table 2. Method comparison

Method	**Accuracy (%)**	**Compute Time (seconds)**	**Optimal Parameters**
Genetic Algorithm	90	45	*C*=2.0, gamma=0.05
Grid Search	88	120	*C*=1.5, gamma=0.1

- Accuracy:
  The Grid Search method yielded an accuracy of 88%, slightly lower than optimization using the Genetic Algorithm (90%).
- Compute Time:
  Genetic algorithms take less time than Grid Search to find optimal parameters. This shows GA's superiority in terms of time efficiency.

From the table, it can be seen that the Genetic Algorithm not only provides higher accuracy but is also more efficient in terms of computational time.

3.2.4 Optimal Parameter Validation

To ensure the robustness of the results, the optimal parameters are tested on independent test data. The results show that the prediction accuracy remains consistent at a value of 90%, indicating that the model does not overfit during the optimization process.

3.3 Interpretation and Discussion

3.3.1 Advantages of Genetic Algorithms :

Genetic algorithms can efficiently explore a wide parameter space (Y. Pratama, 2021). With a population-based approach, GA avoids the risk of being trapped in local optimization, which is often the case with traditional optimization methods such as Grid Search.

3.3.2 for Data Transfer on Google Cloud :

The optimal parameters found show that the SVM model with this configuration can predict data transfer efficiency based on file size and time (Qisthan, 2023). This is important for managing network resources on Google Cloud, especially in large-scale data transfer scenarios.

3.3.3 Limitations and Development Potential :

- The model can be further tested on datasets with more complex variations to improve generalization.

- The use of hybrid methods, such as the combination of GA with other algorithms (e.g. PSO), can also be an alternative for further exploration.

4. Conclusions

This study analyzed data transfer efficiency in Google Cloud using a dataset with 10 samples that included file size, network latency, server utilization, and data transfer time. The prediction model used is a **Support Vector Machine (SVM)** with a Radial Basis Function (RBF) kernel, which is parameterally optimized using the **Genetic Algorithm (GA)**. The process involves data cleaning, normalization, and iteration to find optimal parameters.

The optimization results show that the best parameters for the SVM model are **C=2.0** and **Gamma=0.05**, resulting in a prediction accuracy of **90%**. Compared to the Grid Search method, the Genetic Algorithm provides higher accuracy and more efficient computing time. The visualization also shows the relationship between file size, transfer time, and data transfer efficiency classification.

References

A. Antu, A. W., Abdussamad, S., & Z. Nasibu, I. (2020). Rancang Bangun Running Text pada Dot Matrix 16X160 Berbasis Arduino Uno Dengan Update Data System Menggunakan Perangkat Android Via Bluetooth. *Jambura Journal of Electrical and Electronics Engineering*, *2*(1), 8–13. https://doi.org/10.37905/jjeee.v2i1.4321

Arka, I. G. P. (2021). *Oleh : I Gede Suputra Widharma I Gusti Putu Arka Team Mikrokontroler VATO*. *September 2020*.

Farhan, N. M., & Setiaji, B. (2023). Indonesian Journal of Computer Science. *Indonesian Journal of Computer Science*, *12*(2), 284–301. http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135

Feronica, E., Nasution, Y. N., & Purnamasari, I. (2022). Optimasi Algoritma Naïve Bayes Menggunakan Algoritma Genetika Untuk Memprediksi Kelulusan. *Eksponensial*, *13*(2), 147. https://doi.org/10.30872/eksponensial.v13i2.1057

Fikriansyah, M., Wahyu Wiriasto, G., & Rachman, A. S. (2023). Non-Player Character in Fire Fighter Games Using Genetic Algorithm. *Jurnal Mnemonic*, *6*(1), 11–19. https://doi.org/10.36040/mnemonic.v6i1.5757

Hatta, P. (2011). Optimalisasi Utilisasi Server Menggunakan Virtualisasi Server (Studi Kasus Di Ict Center Fkip Uns). *Jurusan Informatika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret*.

Iput Kasianto. (2021). *Disain Pengendali Prediktif pada Sistem HVAC Berbasis Model Hammerstein Bilinear (Design of Hammerstein-Bilinear Model Based Predictive Controller for HVAC Systems)*. *August*. https://doi.org/10.13140/RG.2.2.21656.88325

Isai, A., & Nugroho, A. (2024). Pencarian Rute Terpendek menggunakan Algoritma Genetika (Studi Kasus : Pengoptimalan Mobilitas Kota Salatiga Terhadap Kota-Kota Tetangga). *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, *5*(1), 681–692. https://doi.org/10.35870/jimik.v5i1.538

MARGARITA, D. (2024). *Klasifikasi Penyakit Daun Padi Menggunakan Support Vector Machine Berdasarkan Fitur Mendalam (Deep Feature)*. *9*(4), 1–21.

Mukhlis, I. R. (2024). *ALGORITMA PEMBELAJARAN MESIN (Dasar , Teknik , dan Aplikasi)* (Issue April).

Pratama, R. R. (2020). Penyelesaian Travelling Salesman Problem Menggunakan Algoritma Genetika. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, *2*(1), 10–18. https://doi.org/10.52005/jursistekni.v2i1.35

Pratama, Y. (2021). Optimalisasi Penjadwalan Karyawan Paruh Waktu Berdasarkan Nilai Fitness Terbaik Menggunakan Algoritma Genetika. *Jurnal Nasional Informatika*, *2*(2), 114–142.

Qisthan, A. H. (2023). Analisis performa metode convolutional neural network dengan arsitektur convnext dalam klasifikasi spesies ular berbisa dan tidak berbisa di indonesia. *Repository.Uinjkt.Ac.Id*. https://repository.uinjkt.ac.id/dspace/handle/123456789/77034%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/77034/1/ALIFIAR HAZAZI QISTHAN-FST.pdf

Qulub, M. S., & Agustin, S. (2024). *INDENTIFIKASI PENYAKIT MATA DENGAN KLASIFIKASI CITRA FOTO FUNDUS MENGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN)*. *8*(5), 11034–11039.

Safii, M., S, W., Pasaribu, J. S., Bau, R. T. R. ., Munawar, Z., A, H., Harto, B., Joshua, S. R., Putri, N. I., Safii, M., Amna, Sophian, S., Rukmana, A. Y., & Hariyadi. (2023). *Layanan Digital di Era 5.0* (Issue June).

Saputro, H. A., Mahmudy, W. F., & Dewi, C. (2015). Implementasi Algoritma Genetika Untuk Optimasi Penggunaan Lahan Pertanian. *Jurnal Mahasiswa PTIIK*, *5*(12), 12.

Sulartopo, S., Kholifah, S., Danang, D., & Santoso, J. T. (2023). Transformasi Proyek Melalui Keajaiban Kecerdasan Buatan: Mengeksplorasi Potensi AI Dalam Project Management. *Jurnal Publikasi Ilmu Manajemen (JUPIMAN)*, *2*(2), 363–392. https://doi.org/10.55606/jupiman.v2i2.2477

Sutomo, F., Hidayat, T. P., & Bachri, K. O. (2022). Perbaikan Kondisi Keseimbangan Lintasan Produksi Dengan Penerapan Metode Genetic Algorithm Pada Lintasan Produksi Muffler K-59J PT XYZ. *Jurnal METRIS*, *23*(01), 15–22. https://doi.org/10.25170/metris.v23i01.2961

Wijaya, W., Panjaitan, F., Rizal, S., & Ulfa, M. (2024). Perbandingan Kinerja Sistem Operasi Network Attached Storage: Studi Kasus Truenas dan Xigmanas. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, *5*(1), 255–261. https://doi.org/10.35870/jimik.v5i1.468

Yunita, Y., Miraswan, K. J., & Andini, D. (2023). Penerapan Algoritma Genetika Sistem Penjadwalan Mata Pelajaran pada SMA Bina Jaya Palembang. *Teknomatika*, *13*(02). http://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/617%0Ahttp://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/617/434

Yusron Mubarok, A., & Chotijah, U. (2021). Penerapan Algoritma Genetika Untuk Mencari Optimasi Kombinasi Jalur Terpendek Dalam Kasus Travelling Salesman Problem. *Jurnal Teknologi Terpadu*, *7*(2), 77–82. https://doi.org/10.54914/jtt.v7i2.424