

# Perbandingan Model KNN dan Regresi Linear untuk Prediksi Biaya EC2 AWS

Muhammad Sulthon<sup>1</sup>, Maulana Aditdya<sup>2</sup>, Maulana Izaki<sup>3</sup>

<sup>1, 2</sup>Teknik Informatika, STMIK YMI Tegal, Tegal

<sup>3</sup>Magister Teknik Informatika, Universitas Dian Nuswantoro, Semarang

<sup>1, 2</sup>Jalan Pendidikan No. 1, Kelurahan Pesurungan Lor, Kota Tegal, 52142, Indonesia

<sup>3</sup>Jalan Imam Bonjol No.207, Pendrikan Kidul, Kota Semarang, 50131, Indonesia

email: [1muhammadsulthon0102@gmail.com](mailto:1muhammadsulthon0102@gmail.com), [2aditdyamaulana@gmail.com](mailto:2aditdyamaulana@gmail.com), [3maulanaizaki@gmail.com](mailto:3maulanaizaki@gmail.com)

**Abstract** – This study aims to compare the performance of K-Nearest Neighbors (KNN) and Linear Regression algorithms in predicting the cost of Amazon EC2 services based on performance data such as the number of vCPUs, memory capacity, and bandwidth. The dataset used was obtained from Kaggle and has gone through a process of cleaning, feature selection, and normalization. The KNN algorithm is optimized with the parameter  $k=3$ , while Linear Regression utilizes the Least Squares method. The evaluation results show that KNN has the best performance with an R-squared ( $R^2$ ) value of 1.0, MAE close to zero, and prediction errors that are barely detected. In contrast, Linear Regression produces an R-squared value of 0.8759 with a higher error than KNN. These results show that KNN is superior to data with complex and non-linear relationships. This research provides new insights in choosing a cloud service cost prediction algorithm, specifically for AWS EC2 services.

**Abstrak** – Penelitian ini bertujuan membandingkan kinerja algoritma K-Nearest Neighbors (KNN) dan Regresi Linear dalam memprediksi biaya layanan Amazon EC2 berdasarkan data performa seperti jumlah vCPUs, kapasitas memori, dan bandwidth. Dataset yang digunakan diperoleh dari Kaggle dan telah melalui proses pembersihan, seleksi fitur, serta normalisasi. Algoritma KNN dioptimalkan dengan parameter  $k=3$ , sementara Regresi Linear memanfaatkan metode Least Squares. Hasil evaluasi menunjukkan bahwa KNN memiliki performa terbaik dengan nilai R-squared ( $R^2$ ) sebesar 1.0, MAE mendekati nol, dan kesalahan prediksi yang nyaris tidak terdeteksi. Sebaliknya, Regresi Linear menghasilkan nilai R-squared sebesar 0.8759 dengan kesalahan yang lebih tinggi dibandingkan KNN. Hasil ini menunjukkan bahwa KNN lebih unggul untuk data dengan hubungan kompleks dan non-linier. Penelitian ini memberikan wawasan baru dalam memilih algoritma prediksi biaya layanan cloud, khususnya untuk layanan AWS EC2.

**Kata Kunci** – Amazon EC2, K-Nearest Neighbors, Pembelajaran Mesin, Prediksi Biaya, Regresi Linear, Sumber Daya Cloud.

## I. PENDAHULUAN

Dalam perkembangan teknologi yang pesat, komputasi awan (cloud computing) telah menjadi solusi utama bagi banyak organisasi dalam pengelolaan dan pengolahan data [1]. Teknologi ini menawarkan berbagai keuntungan seperti fleksibilitas, efisiensi biaya, dan skalabilitas, sehingga menjadi pilihan utama untuk mendukung transformasi digital di berbagai sektor [2]. Salah satu penyedia layanan cloud terbesar

adalah Amazon Web Services (AWS), yang menawarkan berbagai layanan berbasis cloud untuk memenuhi berbagai kebutuhan teknologi informasi [3]. Salah satu layanan unggulan AWS adalah Amazon EC2 (Elastic Compute Cloud), yang memungkinkan pengguna menjalankan aplikasi dan layanan dengan fleksibilitas tinggi [4]. Namun, biaya yang dibebankan sangat bergantung pada pemanfaatan sumber daya seperti CPU, memori, dan bandwidth [5].

Meskipun memberikan fleksibilitas yang besar, banyak pengguna AWS EC2 mengalami kesulitan dalam merencanakan dan mengontrol biaya operasional [6]. Biaya ini dipengaruhi oleh berbagai faktor seperti penggunaan CPU, Disk I/O, dan lalu lintas jaringan [7]. Ketidakpastian dalam memprediksi biaya sering kali mengakibatkan pemborosan anggaran, terutama bagi organisasi yang mengelola beban kerja dengan skala besar [8]. Selain itu, layanan EC2 yang menawarkan berbagai konfigurasi instance seringkali menyulitkan pengguna untuk memahami dan mengoptimalkan biaya berdasarkan kebutuhan spesifik mereka [9]. Oleh karena itu, diperlukan solusi prediksi biaya yang dapat membantu pengguna membuat keputusan yang lebih tepat dalam mengalokasikan sumber daya dan mengelola anggaran [10].

Berbagai pendekatan telah dilakukan untuk memprediksi biaya layanan cloud, termasuk penggunaan model statistik seperti regresi hingga algoritma pembelajaran mesin [11]. Namun, banyak penelitian sebelumnya hanya berfokus pada satu algoritma tanpa membandingkan efektivitas berbagai model dalam kondisi yang berbeda [12]. Sebagai contoh, model regresi sering digunakan untuk memodelkan hubungan linier antara parameter performa dan biaya, namun model ini tidak selalu efektif untuk menangani data dengan hubungan yang kompleks dan tidak linier [13]. Di sisi lain, algoritma seperti K-Nearest Neighbors (KNN) memiliki kemampuan untuk memodelkan data yang lebih kompleks, tetapi kinerjanya belum banyak dibandingkan secara langsung dengan regresi linier dalam konteks prediksi biaya EC2 [14].

Untuk mengisi celah tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi biaya EC2 yang lebih efektif dengan menggunakan dua algoritma prediksi yang umum digunakan, yaitu K-Nearest Neighbors (KNN) dan Regresi Linear [15]. Penelitian ini tidak hanya membandingkan performa kedua algoritma berdasarkan metrik-metrik evaluasi seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error

(RMSE), dan R-squared, tetapi juga mengeksplorasi pengaruh parameter model terhadap akurasi prediksi [16]. Dengan pendekatan ini, diharapkan penelitian ini dapat memberikan gambaran yang lebih jelas mengenai algoritma mana yang lebih sesuai untuk memprediksi biaya EC2 berdasarkan metrik performa seperti vCPUs, memori, dan bandwidth.

\*) **penulis korespondensi:** Muhammad Sulthon  
Email: muhammadsulthon0102@gmail

## II. PENELITIAN YANG TERKAIT

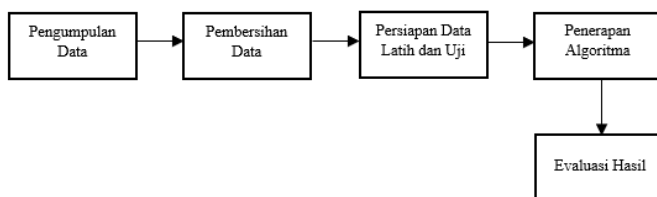
Penelitian terdahulu mengkaji tentang penggunaan pembelajaran mesin untuk manajemen sumber daya cloud, dengan hasil bahwa solusi berbasis pembelajaran mesin dapat meningkatkan efisiensi dalam pengelolaan beban kerja dan optimasi energi [17]. Penelitian selanjutnya, melakukan pengembangan model offloading tugas menggunakan fog computing untuk mengatasi latensi pada aplikasi sensitif waktu, yang menunjukkan peningkatan kinerja dengan memanfaatkan sumber daya edge dan cloud [18]. Penelitian terakhir menggunakan pembelajaran mesin untuk memprediksi peringkat kualitas layanan (QoS) dalam ekosistem cloud hyperconverged, yang berhasil meningkatkan pemilihan layanan berdasarkan parameter QoS seperti waktu henti dan gangguan layanan [19].

Berbeda dengan penelitian terdahulu, penelitian ini fokus pada perbandingan dua algoritma sederhana, yaitu K-Nearest Neighbors (KNN) dan Regresi Linear, untuk memprediksi biaya EC2 AWS. Penelitian ini bertujuan memberikan solusi praktis yang lebih mudah diterapkan dalam memprediksi biaya berdasarkan metrik performa seperti vCPUs dan memori, dibandingkan dengan pendekatan yang lebih kompleks seperti fog computing atau analisis QoS [20].

## III. METODE PENELITIAN

### A. Desain Penelitian

Penelitian ini menggunakan desain kuantitatif eksperimental dengan pendekatan komparatif untuk membandingkan kinerja algoritma K-Nearest Neighbors (KNN) dan Regresi Linear dalam memprediksi biaya layanan EC2 AWS.



Gambar 1. Desain Penelitian

Gambar 1. menggambarkan alur desain penelitian yang dimulai dengan pengumpulan data dari sumber yang relevan, diikuti oleh pembersihan data untuk memastikan hanya data yang valid dan relevan yang digunakan. Selanjutnya, data dibagi menjadi data latih dan data uji untuk proses pelatihan model. Setelah itu, dilakukan penerapan algoritma K-Nearest Neighbors (KNN) dan Regresi Linear pada data yang telah diproses. Langkah terakhir adalah evaluasi hasil, di mana kinerja model diukur menggunakan metrik seperti MAE, MSE,

RMSE, dan R-squared untuk menentukan algoritma yang paling akurat. Alur ini dirancang untuk memastikan setiap tahapan dilakukan secara sistematis dan terukur.

### B. Pengumpulan Data

Tahapan pengumpulan data bertujuan untuk mengumpulkan dataset awal yang relevan untuk penelitian. Data dalam penelitian ini diperoleh dari platform publik Kaggle, yang menyediakan dataset terkait berbagai instance Amazon EC2. Dataset ini mencakup atribut penting seperti jumlah vCPUs, kapasitas memori, bandwidth maksimum yang dioptimalkan untuk EBS, serta biaya operasional untuk sistem operasi Linux On-Demand. Pada tahap ini, semua data yang tersedia, yaitu sebanyak 506 baris, dikumpulkan tanpa melakukan seleksi atau pembersihan. Pengumpulan data ini dilakukan untuk memastikan bahwa seluruh informasi yang tersedia dapat digunakan pada tahap pra-pemrosesan berikutnya.

### C. Pra-Pemrosesan Data

Pada tahap pra-pemrosesan, data yang telah dikumpulkan dibersihkan dengan menghapus nilai kosong (missing values) dan data yang tidak relevan. Kemudian, dilakukan seleksi untuk memilih baris dengan atribut lengkap dan fitur penting seperti jumlah vCPUs, kapasitas memori, bandwidth maksimum EBS, dan biaya operasional Linux. Data numerik dinormalisasi untuk menyamakan skala antar fitur, menghindari bias pada model. Terakhir, dataset dibagi menjadi 80% untuk data latih dan 20% untuk data uji, memastikan kualitas data optimal untuk prediksi yang akurat dan andal.

### D. Pemilihan Variabel dan Optimasi Model

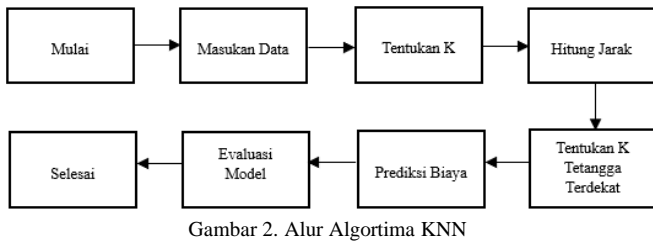
Pada tahap pra-pemrosesan, data yang telah dikumpulkan dibersihkan dengan menghapus nilai kosong (missing values) dan data yang tidak relevan. Kemudian, dilakukan seleksi untuk memilih baris dengan atribut lengkap dan fitur penting seperti jumlah vCPUs, kapasitas memori, bandwidth maksimum EBS, dan biaya operasional Linux. Data numerik dinormalisasi untuk menyamakan skala antar fitur, menghindari bias pada model. Terakhir, dataset dibagi menjadi 80% untuk data latih dan 20% untuk data uji, memastikan kualitas data optimal untuk prediksi yang akurat dan andal.

### E. Implementasi Algoritma

Pada tahap implementasi algoritma, dua algoritma pembelajaran mesin, yaitu K-Nearest Neighbors (KNN) dan Regresi Linear, digunakan secara terpisah untuk memprediksi biaya layanan Amazon EC2. Berikut adalah penjelasan setiap algoritma:

- **K-Nearest Neighbors (KNN)**

Algoritma KNN bekerja dengan mencari kedekatan data uji terhadap data latih menggunakan metrik jarak seperti Euclidean Distance [21]. Pada tahap ini, algoritma dilatih menggunakan data latih yang telah diproses, dan parameter k (jumlah tetangga terdekat) yang telah dioptimalkan sebelumnya diterapkan untuk akurasi maksimal [22]. Prediksi dilakukan dengan mengambil rata-rata nilai target dari k tetangga terdekat, cocok untuk data dengan pola hubungan yang kompleks atau tidak linier [23].

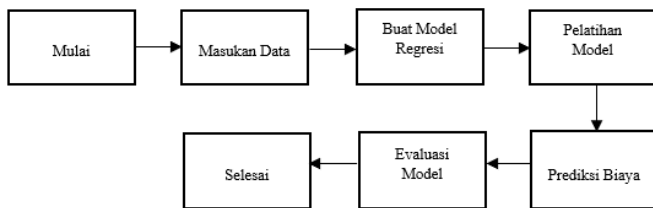


Gambar 2. Alur Algoritma KNN

Gambar 2. menunjukkan alur kerja algoritma KNN. Proses dimulai dengan input data, dilanjutkan dengan penentuan parameter k. Selanjutnya, sistem menghitung jarak antara data uji dan semua data latih, lalu menemukan k tetangga terdekat. Prediksi biaya dilakukan dengan rata-rata nilai target dari tetangga terdekat tersebut. Hasilnya dievaluasi menggunakan metrik seperti MAE, MSE, RMSE, atau R-squared untuk mengukur akurasi prediksi.

• Regresi Linear

Algoritma Regresi Linear memodelkan hubungan linier antara variabel input (vCPUs, memori, bandwidth) dan target prediksi (biaya EC2) [24]. Model ini dilatih menggunakan metode Least Squares untuk meminimalkan selisih kuadrat antara nilai prediksi dan aktual. Setelah model dilatih, data uji digunakan untuk menghasilkan prediksi biaya berdasarkan persamaan linier yang diperoleh [25].



Gambar 3. Alur Regresi Linear

Gambar 3. menunjukkan alur kerja algoritma Regresi Linear. Proses dimulai dengan input data, dilanjutkan dengan pembuatan model regresi yang merumuskan hubungan linier antara variabel input dan target. Setelah pelatihan selesai, prediksi dilakukan pada data uji menggunakan persamaan linier yang dihasilkan. Model dievaluasi dengan metrik seperti MAE, MSE, RMSE, dan R-squared untuk mengukur performanya.

F. Evaluasi Model

Evaluasi model dilakukan untuk mengukur akurasi dan kinerja algoritma yang digunakan dalam penelitian ini. Beberapa metrik evaluasi yang digunakan meliputi:

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

dimana :

- $n$  : Jumlah data atau observasi.
- $y_i$ : Nilai aktual pada observasi ke- $i$ .
- $\hat{y}_i$  : Nilai prediksi pada observasi ke- $i$ .

Mean Squared Error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

dimana :

- $n$  : Jumlah data atau observasi.
- $y_i$ : Nilai aktual pada observasi ke- $i$ .
- $\hat{y}_i$  : Nilai prediksi pada observasi ke- $i$ .

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

dimana :

- $n$  : Jumlah data atau observasi.
- $y_i$ : Nilai aktual pada observasi ke- $i$ .
- $\hat{y}_i$  : Nilai prediksi pada observasi ke- $i$ .

R-squared ( $R^2$ )

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

dimana :

- $n$  : Jumlah data atau observasi.
- $y_i$ : Nilai aktual pada observasi ke- $i$ .
- $\hat{y}_i$  : Nilai prediksi pada observasi ke- $i$ .
- $\bar{y}$  : Nilai rata-rata dari semua  $y_i$

IV. HASIL DAN PEMBAHASAN

Data yang digunakan untuk prediksi biaya layanan Amazon EC2 terdiri dari atribut seperti jumlah vCPUs, kapasitas memori, kinerja jaringan, bandwidth maksimum EBS yang dioptimalkan, dan biaya Linux On Demand. Berikut adalah tabel data awal yang digunakan dalam penelitian ini:

Tabel 1. DATA EC2

vCPU	Memor y	Network Performanc e	EBS Optimize d: Max Bandwid th	Linux On Demand cost
96	384	25	20000	4,1472
4	16	10	4750	0,272
32	64	25	6666,664	1,224
⋮	⋮	⋮	⋮	⋮
128	1024	25	30000	5,5
2	8	5	2000	0,2
48	128	20	15000	3

Tabel 1. Data EC2 menunjukkan konfigurasi instance Amazon EC2 berdasarkan lima atribut utama, yaitu vCPUs (2 hingga 128 unit), memori (8 hingga 1024 GB), Network Performance (5 hingga 25), EBS Optimized: Max Bandwidth (2000 hingga 30000 MBps), dan Linux On Demand cost (0,2 hingga 5,5 USD). Data ini mencakup variasi spesifikasi dari

rendah hingga tinggi, yang digunakan untuk analisis dan prediksi biaya layanan berbasis performa instance.

Setelah menerapkan algoritma K-Nearest Neighbors (KNN) dengan parameter  $k=3$  dan Regresi Linear pada dataset, diperoleh hasil evaluasi kinerja kedua model. Algoritma KNN menunjukkan performa yang sangat baik dengan nilai R-squared ( $R^2$ ) mencapai 1.0, yang menunjukkan bahwa model mampu memprediksi biaya EC2 secara sempurna. Di sisi lain, algoritma Regresi Linear juga memberikan hasil yang cukup baik dengan R-squared ( $R^2$ ) sebesar 0.8759, meskipun kesalahan prediksinya lebih tinggi dibandingkan KNN. Berikut adalah tabel hasil evaluasi kedua model:

Tabel 2. Hasil

Metrik Evaluasi	KNN ( $k=3$ )	Regresi Linear
MAE	6.938893903907228e-18	0.5126806639459506
MSE	6.548161810916602e-34	0.3361950195744885
RMSE	2.5589376332604517e-17	0.5798232658099264
R-squared ( $R^2$ )	1.0	0.8759249541328842

Tabel 2 menunjukkan hasil evaluasi kinerja model K-Nearest Neighbors (KNN) dengan  $k=3$  dan Regresi Linear untuk prediksi biaya layanan EC2. Algoritma KNN memberikan performa yang sangat baik dengan nilai R-squared ( $R^2$ ) sebesar 1.0, yang menunjukkan akurasi prediksi yang sempurna. Selain itu, nilai MAE, MSE, dan RMSE dari KNN sangat kecil, menunjukkan kesalahan prediksi yang nyaris tidak terdeteksi. Sebaliknya, Regresi Linear menunjukkan kinerja yang cukup baik dengan R-squared ( $R^2$ ) sebesar 0.8759, namun memiliki nilai kesalahan yang lebih tinggi dibandingkan KNN, seperti MAE sebesar 0.5127 dan RMSE sebesar 0.5798. Hasil ini menunjukkan bahwa KNN lebih unggul dalam menangani prediksi pada dataset ini dibandingkan dengan Regresi Linear.

Berdasarkan tabel 2, KNN memberikan hasil evaluasi yang nyaris sempurna dengan nilai MAE, MSE, dan RMSE yang sangat kecil. Hal ini menunjukkan bahwa KNN memiliki kemampuan prediksi yang sangat akurat, dengan kesalahan yang hampir tidak terdeteksi. Nilai R-squared ( $R^2$ ) sebesar 1.0 memperkuat kesimpulan bahwa model KNN sangat baik dalam memodelkan data ini.

Sebaliknya, Regresi Linear menunjukkan kesalahan prediksi yang lebih besar dibandingkan KNN, dengan MAE sebesar 0.5127 dan RMSE sebesar 0.5798. Meskipun nilai R-squared ( $R^2$ ) sebesar 0.8759 menunjukkan bahwa model ini cukup baik dalam menjelaskan variasi data, kinerjanya tetap lebih rendah dibandingkan KNN. Hal ini kemungkinan disebabkan oleh keterbatasan Regresi Linear dalam menangani hubungan non-linier antara variabel input dan output.

Secara keseluruhan, hasil ini menunjukkan bahwa algoritma KNN lebih unggul dibandingkan Regresi Linear untuk tugas prediksi biaya layanan EC2 berdasarkan dataset ini. Keunggulan KNN disebabkan oleh kemampuannya dalam menangani pola data yang kompleks dan non-linier, menjadikannya pilihan yang lebih baik untuk tugas prediksi biaya layanan cloud. Dengan demikian, KNN dapat direkomendasikan untuk digunakan pada data dengan

hubungan yang kompleks, sementara Regresi Linear lebih cocok untuk data dengan pola linier.

## V. KESIMPULAN

Penelitian ini membandingkan kinerja algoritma K-Nearest Neighbors (KNN) dan Regresi Linear dalam memprediksi biaya layanan Amazon EC2 berdasarkan metrik performa. Hasil evaluasi menunjukkan bahwa KNN dengan  $k=3$  memberikan performa terbaik dengan nilai R-squared ( $R^2$ ) sebesar 1.0, yang menunjukkan akurasi prediksi sempurna, sementara Regresi Linear memiliki R-squared sebesar 0.8759 dengan kesalahan prediksi yang lebih tinggi. Kesimpulannya, KNN lebih unggul dalam menangani pola data yang kompleks dan non-linier, sehingga lebih direkomendasikan untuk prediksi biaya layanan EC2 dibandingkan Regresi Linear.

## DAFTAR PUSTAKA

- [1] A. Latifian, "How does cloud computing help businesses to manage big data issues," *Kybernetes*, vol. 51, no. 6, pp. 1917–1948, Jan. 2022, doi: 10.1108/K-05-2021-0432.
- [2] G. Dutta, R. Kumar, R. Sindhwani, and R. K. Singh, "Digital transformation priorities of India's discrete manufacturing SMEs – a conceptual study in perspective of Industry 4.0," *Compet. Rev. An Int. Bus. J.*, vol. 30, no. 3, pp. 289–314, Jan. 2020, doi: 10.1108/CR-03-2019-0031.
- [3] P. Kaushik, A. M. Rao, D. P. Singh, S. Vashisht, and S. Gupta, "Cloud Computing and Comparison based on Service and Performance between Amazon AWS, Microsoft Azure, and Google Cloud," in *2021 International Conference on Technological Advancements and Innovations (ICTAI)*, 2021, pp. 268–273. doi: 10.1109/ICTAI53825.2021.9673425.
- [4] S. Park *et al.*, "FaaSr: Cross-Platform Function-as-a-Service Serverless Scientific Workflows in R," in *2024 IEEE 20th International Conference on e-Science (e-Science)*, 2024, pp. 1–10. doi: 10.1109/e-Science62913.2024.10678660.
- [5] D. Bortolini and R. R. Obelheiro, "Investigating Performance and Cost in Function-as-a-Service Platforms BT - Advances on P2P, Parallel, Grid, Cloud and Internet Computing," L. Barolli, P. Hellinckx, and J. Natwichai, Eds., Cham: Springer International Publishing, 2020, pp. 174–185. doi: 10.1007/978-3-030-33509-0\_16.
- [6] G. Terrazas, N. Ferry, and S. Ratchev, "A cloud-based framework for shop floor big data management and elastic computing analytics," *Comput. Ind.*, vol. 109, pp. 204–214, Aug. 2019, doi: 10.1016/j.compind.2019.03.005.
- [7] A. A. Khan and M. Zakarya, "Energy, performance and cost efficient cloud datacentres: A survey," *Comput. Sci. Rev.*, vol. 40, p. 100390, 2021, doi: <https://doi.org/10.1016/j.cosrev.2021.100390>.
- [8] P. Elyar, S. Diana, H. Farook, and A. Simaan, "The Need for a Multidimensional Project Control Perspective," *J. Constr. Eng. Manag.*, vol. 150, no. 12, p. 4024169, Dec. 2024, doi:

- 10.1061/JCEMD4.COENG-14540.
- [9] M. Liu, L. Pan, and S. Liu, "Cost Optimization for Cloud Storage from User Perspectives: Recent Advances, Taxonomy, and Survey," *ACM Comput. Surv.*, vol. 55, no. 13s, Jul. 2023, doi: 10.1145/3582883.
- [10] M. J. F. Rosa, C. G. Ralha, M. Holanda, and A. P. F. Araujo, "Computational resource and cost prediction service for scientific workflows in federated clouds," *Futur. Gener. Comput. Syst.*, vol. 125, pp. 844–858, 2021, doi: <https://doi.org/10.1016/j.future.2021.07.030>.
- [11] A. M. S. Kheir, K. A. Ammar, A. Amer, M. G. M. Ali, Z. Ding, and A. Elnashar, "Machine learning-based cloud computing improved wheat yield simulation in arid regions," *Comput. Electron. Agric.*, vol. 203, p. 107457, 2022, doi: <https://doi.org/10.1016/j.compag.2022.107457>.
- [12] A. Merghadi *et al.*, "Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance," *Earth-Science Rev.*, vol. 207, p. 103225, Aug. 2020, doi: 10.1016/j.earscirev.2020.103225.
- [13] D. A. Otchere, T. O. Arbi Ganat, R. Gholami, and S. Ridha, "Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models," *J. Pet. Sci. Eng.*, vol. 200, p. 108182, May 2021, doi: 10.1016/j.petrol.2020.108182.
- [14] S. Wang, F. Zhu, Y. Yao, W. Tang, Y. Xiao, and S. Xiong, "A computing resources prediction approach based on ensemble learning for complex system simulation in cloud environment," *Simul. Model. Pract. Theory*, vol. 107, p. 102202, Feb. 2021, doi: 10.1016/j.simpat.2020.102202.
- [15] A. Mazidi, M. Golsorkhtabamiri, and M. Y. Tabari, "Autonomic resource provisioning for multilayer cloud applications with K-nearest neighbor resource scaling and priority-based resource allocation," *Softw. Pract. Exp.*, vol. 50, no. 8, pp. 1600–1625, Aug. 2020, doi: 10.1002/spe.2837.
- [16] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification," *Augment. Hum. Res.*, vol. 5, no. 1, p. 12, Dec. 2020, doi: 10.1007/s41133-020-00032-0.
- [17] T. Khan, W. Tian, G. Zhou, S. Ilager, M. Gong, and R. Buyya, "Machine learning (ML)-centric resource management in cloud computing: A review and future directions," *J. Netw. Comput. Appl.*, vol. 204, p. 103405, Aug. 2022, doi: 10.1016/j.jnca.2022.103405.
- [18] M. M. Bukhari *et al.*, "An Intelligent Proposed Model for Task Offloading in Fog-Cloud Collaboration Using Logistics Regression," *Comput. Intell. Neurosci.*, vol. 2022, no. 1, p. 3606068, Jan. 2022, doi: <https://doi.org/10.1155/2022/3606068>.
- [19] N. Tabassum *et al.*, "Prediction of Cloud Ranking in a Hyperconverged Cloud Ecosystem Using Machine Learning," *Comput. Mater. Contin.*, vol. 67, no. 3, pp. 3129–3141, 2021, doi: 10.32604/cmc.2021.014729.
- [20] M. Zakarya *et al.*, "epcAware: A Game-Based, Energy, Performance and Cost-Efficient Resource Management Technique for Multi-Access Edge Computing," *IEEE Trans. Serv. Comput.*, vol. 15, no. 3, pp. 1634–1648, 2022, doi: 10.1109/TSC.2020.3005347.
- [21] R. Ehsani and F. Drabløs, "Robust Distance Measures for kNN Classification of Cancer Data," *Cancer Inform.*, vol. 19, p. 1176935120965542, Jan. 2020, doi: 10.1177/1176935120965542.
- [22] T. S. Ruprah, B. Regmi, S. B. Jadhav, and S. Singh, "Early Stage Lung Cancer Detection Using Deep Learning," in *2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon)*, 2024, pp. 1–6. doi: 10.1109/MITADTSoCiCon60330.2024.10575345.
- [23] A. Ali *et al.*, "A k-Nearest Neighbours Based Ensemble via Optimal Model Selection for Regression," *IEEE Access*, vol. 8, pp. 132095–132105, 2020, doi: 10.1109/ACCESS.2020.3010099.
- [24] E. Yildirim and A. Akon, "Predicting Short-Term Variations in End-to-End Cloud Data Transfer Throughput Using Neural Networks," *IEEE Access*, vol. 11, pp. 78656–78670, 2023, doi: 10.1109/ACCESS.2023.3299311.
- [25] P. N. Sharma, G. Shmueli, M. Sarstedt, N. Danks, and S. Ray, "Prediction-Oriented Model Selection in Partial Least Squares Path Modeling," *Decis. Sci.*, vol. 52, no. 3, pp. 567–607, Jun. 2021, doi: <https://doi.org/10.1111/dec.12329>.