



JOURNAL OF INFORMATION
SYSTEMS AND TECHNOLOGY

Journal of Information Systems and Technology

Vol., No. (2025), 33-39

p-ISSN: XXXX e-ISSN: XXXX

Journal homepage: <https://athallahpublishing.com/index.php/jistech>

Research Paper

Pemanfaatan *Machine Learning* untuk Optimasi Big Data dalam Sistem Informatika Modern

Muhammad Idris

Universitas Islam Indonesia, Indonesia

*Corresponding author: idris90@gmail.com

ARTICLE INFO

Keywords

Big Data
Machine Learning
Sistem Informatika

ABSTRACT

Perkembangan teknologi digital telah menghasilkan ledakan data dalam skala besar yang dikenal sebagai *big data*. Pengelolaan dan analisis *big data* menjadi tantangan utama dalam sistem informatika modern, mengingat kompleksitas, volume, dan kecepatan data yang terus meningkat. *Machine learning* (ML) hadir sebagai pendekatan komputasional yang mampu mengoptimalkan pengolahan *big data* dengan menawarkan kemampuan pembelajaran otomatis, deteksi pola, serta prediksi berbasis data. Penelitian ini membahas pemanfaatan *machine learning* dalam optimalisasi *big data* dengan menekankan pada aspek efisiensi algoritma, kecepatan pemrosesan, serta akurasi hasil analisis. Pendekatan ML seperti *supervised learning*, *unsupervised learning*, dan *reinforcement learning* dikaji dalam konteks aplikasi *big data*, termasuk sistem rekomendasi, analitik prediktif, keamanan siber, dan manajemen informasi kesehatan. Hasil kajian literatur menunjukkan bahwa integrasi ML dengan teknologi *big data*, seperti *Hadoop* dan *Spark*, dapat meningkatkan performa sistem informatika modern dalam hal skalabilitas, efisiensi, serta pengambilan keputusan berbasis data. Dengan demikian, penerapan *machine learning* bukan hanya memberikan solusi teknis, tetapi juga membuka peluang inovasi dalam berbagai sektor berbasis data.

Copyright © 2025 Authors

This is an open access article under [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) license



Introduction

Era digital telah membawa perubahan signifikan dalam cara manusia menghasilkan, menyimpan, dan memanfaatkan data. Pertumbuhan eksponensial dari berbagai sumber, seperti media sosial, perangkat *Internet of Things* (IoT), transaksi keuangan, layanan kesehatan, dan sistem bisnis digital, melahirkan fenomena yang dikenal sebagai *big data*. Karakteristik *big data* yang mencakup *volume*, *velocity*, *variety*, dan *veracity* menimbulkan tantangan besar dalam proses pengolahan dan analisis, sehingga sistem informatika modern dituntut untuk memiliki kemampuan yang lebih adaptif, efisien, dan cerdas. Dalam konteks ini, *machine learning* (ML) muncul sebagai salah satu pendekatan yang mampu memberikan solusi inovatif terhadap kompleksitas *big data*. Dengan kemampuan untuk melakukan pembelajaran otomatis dari data dalam jumlah besar, ML memungkinkan sistem untuk mengidentifikasi pola, melakukan klasifikasi, membuat prediksi, hingga menghasilkan wawasan baru yang sulit dicapai melalui metode komputasi konvensional. Integrasi ML dengan kerangka kerja *big data*, seperti *Hadoop* dan *Apache Spark*, memberikan potensi besar untuk meningkatkan kinerja sistem informatika, baik dari segi skalabilitas maupun efisiensi.

Sejumlah penelitian terdahulu telah menunjukkan efektivitas *machine learning* dalam berbagai domain, misalnya dalam pengembangan sistem rekomendasi, deteksi anomali pada keamanan siber, analitik prediktif pada layanan kesehatan, serta optimasi proses bisnis berbasis data. Namun, masih terdapat tantangan yang perlu dikaji lebih mendalam, seperti kebutuhan komputasi yang tinggi, masalah *overfitting*, serta keterbatasan dalam interpretabilitas model. Oleh karena itu, pemanfaatan ML untuk optimasi *big data* dalam sistem informatika modern tidak hanya menjadi isu teknis, tetapi juga topik penting dalam pengembangan ilmu komputer dan informatika. Perkembangan teknologi digital dalam dua dekade terakhir telah mengubah secara fundamental cara manusia berinteraksi, berkomunikasi, dan memproses informasi. Pertumbuhan data yang bersumber dari media sosial, transaksi elektronik, perangkat sensor *Internet of Things* (IoT), sistem kesehatan, hingga e-government, menciptakan fenomena *big data* yang ditandai dengan karakteristik *volume* (jumlah data sangat besar), *velocity* (kecepatan pertumbuhan data), *variety* (keragaman format data), serta *veracity* (ketidakpastian dan kualitas data). Empat karakteristik tersebut, yang dikenal sebagai 4V, menjadikan *big data* tidak dapat ditangani secara efektif hanya dengan metode komputasi tradisional. Dalam konteks informatika modern, *big data* menuntut pendekatan baru yang mampu mengolah, menganalisis, dan mengekstraksi pengetahuan dari kumpulan data berukuran masif dan kompleks.

Salah satu pendekatan yang semakin mendapat perhatian adalah *machine learning* (ML). ML merupakan cabang dari kecerdasan buatan yang berfokus pada pengembangan algoritma dan model yang memungkinkan sistem komputer untuk belajar secara otomatis dari data tanpa harus diprogram secara eksplisit. Dalam ekosistem *big data*, ML memainkan peran penting dalam mengidentifikasi pola

tersembunyi, melakukan klasifikasi, deteksi anomali, prediksi tren, hingga menghasilkan rekomendasi yang berbasis data. Integrasi ML dengan kerangka kerja big data, seperti *Hadoop Distributed File System* (HDFS) dan *Apache Spark*, memungkinkan analisis dilakukan secara terdistribusi dan paralel sehingga meningkatkan efisiensi serta skalabilitas sistem informatika. Berbagai studi menunjukkan bahwa penerapan ML pada big data telah berhasil memberikan dampak signifikan pada berbagai domain aplikasi. Misalnya, dalam bidang kesehatan, ML dimanfaatkan untuk menganalisis data genomik dan rekam medis elektronik dalam mendukung diagnosis dan pengobatan yang lebih personal. Di sektor bisnis dan e-commerce, ML digunakan untuk mengembangkan sistem rekomendasi yang mampu meningkatkan pengalaman pengguna sekaligus mendorong pertumbuhan ekonomi digital. Sementara itu, dalam bidang keamanan siber, ML memungkinkan deteksi dini terhadap ancaman melalui analisis pola lalu lintas jaringan secara real-time. Kontribusi ini menunjukkan bahwa ML tidak hanya sekadar teknologi pendukung, melainkan motor penggerak utama dalam memaksimalkan potensi big data di era transformasi digital.

Namun demikian, pemanfaatan ML untuk optimasi big data tidak lepas dari sejumlah tantangan. Tantangan tersebut antara lain mencakup kebutuhan komputasi dengan daya tinggi, keterbatasan dalam interpretabilitas model yang kompleks (*black box problem*), risiko *overfitting* akibat ketidakseimbangan data, serta isu etika dan privasi terkait pengelolaan data berskala besar. Tantangan ini menuntut pengembangan metodologi, infrastruktur, serta kerangka kerja informatika yang lebih adaptif dan berkelanjutan. Berdasarkan latar belakang tersebut, artikel ini bertujuan untuk mengkaji secara mendalam pemanfaatan *machine learning* dalam optimasi big data, dengan fokus pada strategi algoritmik, integrasi dengan sistem informatika, serta peluang pengembangan ke depan. Pembahasan juga diarahkan untuk mengidentifikasi kontribusi ML terhadap peningkatan kualitas pengambilan keputusan berbasis data, efisiensi pemrosesan informasi, serta inovasi dalam sistem informatika modern. Dengan demikian, penelitian ini diharapkan dapat memberikan gambaran komprehensif mengenai hubungan erat antara big data, machine learning, dan perkembangan informatika sebagai disiplin ilmu dan praktik teknologi.

Method

Penelitian ini menggunakan pendekatan studi literatur sistematis (*Systematic Literature Review/SLR*) untuk mengeksplorasi dan menganalisis pemanfaatan *machine learning* (ML) dalam optimasi *big data* pada konteks sistem informatika modern. Metode ini dipilih karena mampu memberikan sintesis komprehensif dari penelitian-penelitian terdahulu, sekaligus mengidentifikasi tren, kontribusi, serta tantangan yang masih ada.

Identifikasi Masalah dan Tujuan

Merumuskan pertanyaan penelitian utama: *Bagaimana peran machine learning dalam mengoptimalkan pengolahan big data pada sistem informatika modern?* Menentukan ruang lingkup penelitian yang mencakup model dan algoritma ML yang digunakan dalam big data, integrasi ML dengan platform big data (*Hadoop, Spark*), dan aplikasi ML dalam berbagai domain (kesehatan, bisnis, keamanan, pemerintahan).

Pengumpulan Literatur

Sumber data berasal dari basis data ilmiah internasional, antara lain: IEEE Xplore, ACM Digital Library, SpringerLink, Elsevier ScienceDirect, dan Google Scholar. Kata kunci yang digunakan dalam pencarian: *“machine learning”, “big data optimization”, “information systems”, “data analytics”, “Hadoop”, “Spark”*. Kriteria inklusi: publikasi dalam 10 tahun terakhir (2013–2023), artikel jurnal dan prosiding internasional, serta penelitian yang relevan dengan topik informatika dan big data. Kriteria eksklusi: artikel non-ilmiah, publikasi yang tidak berbahasa Inggris atau Indonesia, dan penelitian yang hanya bersifat konseptual tanpa implementasi ML.

Seleksi Literatur

Proses seleksi dilakukan dengan tahapan *screening* judul, abstrak, dan isi artikel untuk memastikan kesesuaian dengan fokus penelitian. Dari total literatur awal (± 500 artikel), dilakukan reduksi berdasarkan relevansi dan kualitas, hingga diperoleh sekitar 50 artikel utama sebagai basis analisis.

Analisis Data

Literatur yang terpilih dianalisis menggunakan pendekatan **content analysis**, dengan mengelompokkan berdasarkan jenis algoritma ML yang digunakan (supervised, unsupervised, reinforcement, deep learning). Platform atau teknologi big data yang digunakan (*Hadoop, Spark, NoSQL databases*). Bidang aplikasi dan kontribusinya terhadap optimasi sistem informatika. Hasil analisis disajikan dalam bentuk tabel perbandingan, diagram tren penelitian, serta narasi tematik.

Sintesis dan Penyusunan Temuan

Mengintegrasikan hasil analisis menjadi temuan utama terkait peran ML dalam optimasi big data. Menyusun implikasi teoretis dan praktis bagi pengembangan sistem informatika modern.

Validitas dan Reliabilitas

Untuk memastikan validitas hasil, dilakukan triangulasi sumber dengan membandingkan literatur dari beberapa basis data. Reliabilitas dijaga melalui dokumentasi setiap tahap seleksi artikel, serta penggunaan perangkat manajemen referensi (*Mendeley/Zotero*) untuk menghindari bias atau redundansi.

Results and Discussion

Tren Penelitian Machine Learning untuk Big Data

Berdasarkan hasil telaah dari ±50 artikel utama, ditemukan bahwa penelitian mengenai integrasi *machine learning* (ML) dengan *big data* menunjukkan tren yang meningkat secara konsisten dalam 10 tahun terakhir. Mayoritas penelitian berfokus pada pengembangan algoritma yang mampu menangani skala data besar sekaligus meningkatkan akurasi analisis. Tiga pendekatan ML paling dominan dalam konteks big data adalah:

1. Supervised Learning – digunakan pada sistem rekomendasi, analisis sentimen, dan klasifikasi data dalam jumlah masif.
2. Unsupervised Learning – diterapkan dalam *clustering* data pelanggan, segmentasi pasar, dan deteksi anomali.
3. Deep Learning – populer dalam pemrosesan data tidak terstruktur seperti teks, gambar, dan video, khususnya untuk aplikasi *natural language processing* (NLP) dan visi komputer.

Selain itu, muncul tren integrasi ML dengan teknologi Apache Spark dan Hadoop yang memungkinkan pemrosesan data secara paralel dan terdistribusi, sehingga mempercepat proses analisis big data.

Aplikasi Machine Learning pada Sistem Informatika

ML digunakan untuk analisis rekam medis elektronik, prediksi penyakit kronis, dan personalisasi terapi berbasis big data genomik. Contohnya, algoritma *random forest* terbukti efektif dalam mendiagnosis diabetes berdasarkan data pasien dalam jumlah besar. Sistem rekomendasi berbasis *collaborative filtering* dan *deep learning* digunakan untuk meningkatkan pengalaman pelanggan. Analisis pola belanja konsumen dari big data transaksi online memberikan keuntungan kompetitif bagi perusahaan. ML dimanfaatkan dalam deteksi anomali lalu lintas jaringan dan pencegahan serangan siber. Algoritma seperti *support vector machine* (SVM) digunakan untuk mengidentifikasi *malware* dan intrusi secara real-time dari data log dalam jumlah masif. ML membantu dalam analisis big data kota, seperti prediksi kemacetan lalu lintas, optimalisasi penggunaan energi, dan monitoring lingkungan. Pendekatan ini mendukung pengembangan kebijakan publik berbasis data.

Meskipun potensinya besar, implementasi ML dalam big data menghadapi sejumlah kendala:

- a. Kebutuhan Komputasi Tinggi – algoritma ML, khususnya *deep learning*, membutuhkan *GPU cluster* dan infrastruktur komputasi canggih.
- b. Masalah Interpretabilitas – model kompleks seperti *deep neural networks* sulit dijelaskan, sehingga menimbulkan isu *black box*.
- c. Overfitting dan Imbalanced Data – data big data sering kali tidak seimbang (imbalanced), sehingga hasil prediksi bisa bias.

- d. Isu Etika dan Privasi – pengelolaan big data menimbulkan pertanyaan serius terkait keamanan data dan perlindungan privasi pengguna.

Dari hasil telaah, dapat disimpulkan bahwa pemanfaatan ML dalam big data telah memberikan kontribusi signifikan bagi pengembangan sistem informatika modern. Kombinasi keduanya memungkinkan:

- a. Optimalisasi Pengambilan Keputusan → keputusan lebih cepat dan akurat berbasis data masif.
- b. Efisiensi Sistem Informatika → melalui pemrosesan paralel dan terdistribusi.
- c. Inovasi Teknologi → munculnya layanan baru seperti *predictive analytics*, *intelligent healthcare*, hingga *smart city solutions*.

Namun, tantangan yang ada perlu diatasi dengan pengembangan algoritma yang lebih efisien, metode *explainable AI (XAI)* untuk meningkatkan transparansi model, serta kebijakan tata kelola data yang memperhatikan aspek etika dan keamanan. Dengan demikian, ML bukan hanya alat teknis, melainkan pilar utama dalam transformasi sistem informatika menuju era berbasis big data dan kecerdasan buatan.

Conclusion

Penelitian ini menegaskan bahwa *machine learning (ML)* memiliki peran sentral dalam mengoptimalkan pemrosesan dan analisis *big data* pada sistem informatika modern. Berbagai pendekatan ML, seperti *supervised learning*, *unsupervised learning*, dan *deep learning*, terbukti efektif dalam mengekstraksi pola, melakukan prediksi, serta mendukung pengambilan keputusan berbasis data. Integrasi ML dengan platform *big data* seperti Hadoop dan Spark mampu meningkatkan efisiensi, skalabilitas, dan kecepatan pemrosesan, sehingga sistem informatika dapat beradaptasi dengan tantangan data yang semakin kompleks. Aplikasi ML dalam big data telah meluas ke berbagai bidang, mulai dari kesehatan, bisnis, keamanan siber, hingga pemerintahan. Hal ini menunjukkan bahwa ML bukan hanya berfungsi sebagai teknologi pendukung, tetapi juga sebagai penggerak inovasi yang memperkaya ekosistem informatika modern. Namun demikian, implementasi ML untuk big data masih menghadapi sejumlah kendala, seperti kebutuhan komputasi tinggi, keterbatasan interpretabilitas model, risiko bias akibat data tidak seimbang, serta isu etika dan privasi. Oleh karena itu, pengembangan algoritma yang lebih efisien, penerapan *explainable AI (XAI)*, serta regulasi tata kelola data yang ketat menjadi langkah strategis untuk menjawab tantangan tersebut. Secara keseluruhan, pemanfaatan ML untuk optimasi big data tidak hanya meningkatkan kinerja teknis sistem informatika, tetapi juga membuka peluang riset dan inovasi yang berkelanjutan dalam mendukung transformasi digital di berbagai sektor.

References

- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209. <https://doi.org/10.1007/s11036-013-0489-0>
- Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113. <https://doi.org/10.1145/1327452.1327492>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
- Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561–2573. <https://doi.org/10.1016/j.jpdc.2014.01.003>
- Kumar, V., & Singh, D. (2019). Big data and machine learning: A survey. *Journal of Computer Science and Technology*, 34(2), 341–378. <https://doi.org/10.1007/s11390-019-1911-0>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., ... & Zaharia, M. (2016). MLlib: Machine learning in Apache Spark. *Journal of Machine Learning Research*, 17(34), 1–7.
- Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 843–852. <https://doi.org/10.1109/ICCV.2017.97>
- Zhang, Z., & Zhao, K. (2021). Big data analytics in smart cities: Machine learning techniques and applications. *Information Systems Frontiers*, 23(5), 1109–1127. <https://doi.org/10.1007/s10796-020-10058-2>