

OPTIMASI DETEKSI ENTITAS KEBAKARAN HUTAN DENGAN MENGGUNAKAN ALGORITMA GENETIKA PADA PARAMETER *LEARNING RATE* DAN *DROPOUT* DALAM MODEL JARINGAN SARAF REKUREN

Dwi Ahmad Dzulhijjah^{1*}, Kusriani,¹ Kumara Ari Yuana¹

¹Magister Teknik Informatika, Universitas AMIKOM, Yogyakarta

*Penulis korespondensi: dwiamadz@students.amikom.ac.id

ABSTRAK

Kebakaran hutan merupakan bencana yang memiliki dampak merugikan secara signifikan untuk lingkungan dan ekonomi, teknologi berperan penting dalam mitigasi bencana misalnya teknologi kecerdasan buatan. Penelitian ini fokus pada pengoptimalkan deteksi entitas kebakaran hutan dengan menggunakan algoritma genetika pada dua hyperparameter utama, yaitu dropout dan learning rate, dalam model jaringan saraf rekuren, khususnya LSTM. Dengan mengintegrasikan konsep-konsep ini, penelitian bertujuan meningkatkan akurasi dan ketepatan model dalam mengidentifikasi waktu dan lokasi kebakaran. Metodologi penelitian melibatkan pengumpulan dan pemrosesan data teks tweet, pelabelan entitas, dan implementasi algoritma genetika pada hyperparameter model. Hasil menunjukkan bahwa konfigurasi terbaik, dengan dropout 0.247384 dan learning rate 0.019447, menghasilkan akurasi tertinggi sebesar 75.91%. Kesimpulan menegaskan bahwa pengaturan hyperparameter memainkan peran kritis dalam meningkatkan performa model dan relevansinya terhadap mitigasi bencana kebakaran di Indonesia. Saran untuk penelitian masa depan melibatkan eksplorasi metode optimasi tambahan dan peningkatan kompleksitas model untuk menangani variasi data yang lebih luas. Temuan ini memberikan kontribusi pada pengembangan solusi cerdas dalam pencegahan dan respons terhadap bencana kebakaran hutan.

Kata kunci: LSTM, NER, kebakaran hutan, entitas bernama, hyperparameter.

1 PENDAHULUAN

Kebakaran hutan di Indonesia menjadi bencana yang memiliki dampak signifikan, tidak hanya terhadap lingkungan tetapi juga terhadap kerugian ekonomi (Pasai, 2020). Untuk mengatasi tantangan ini, analisis spatio-temporal telah diidentifikasi sebagai suatu kebutuhan penting dalam mitigasi bencana kebakaran (Salsabila dkk., 2020). Penelitian terkini menyoroti potensi pemanfaatan teknologi kecerdasan buatan (AI), khususnya Natural Language Processing (NLP), sebagai alat dalam proses mitigasi dan pencegahan kebakaran (Nuruddin, 2022). Dalam konteks NLP, Long Short-Term Memory (LSTM) adalah jenis model jaringan saraf rekuren (RNN) yang telah terbukti efektif dalam mendeteksi entitas bernama seperti waktu dan lokasi (Pallavi, 2019). Studi terkait dengan penelitian ini juga mencatat bahwa LSTM dapat diaplikasikan secara spesifik untuk mendeteksi entitas lokasi terkait kebakaran (Putra & Kurniawan, 2021). Namun, perlu dicatat bahwa performa model LSTM sangat dipengaruhi oleh hyperparameter yang digunakan seperti *activation function* dan *dropout* (Li, dkk., 2021), sehingga dalam konteks NER juga diperlukan pengaturan hyperparameter untuk pengoptimalan model LSTM untuk NER (Reimers & Gurevych, 2021).

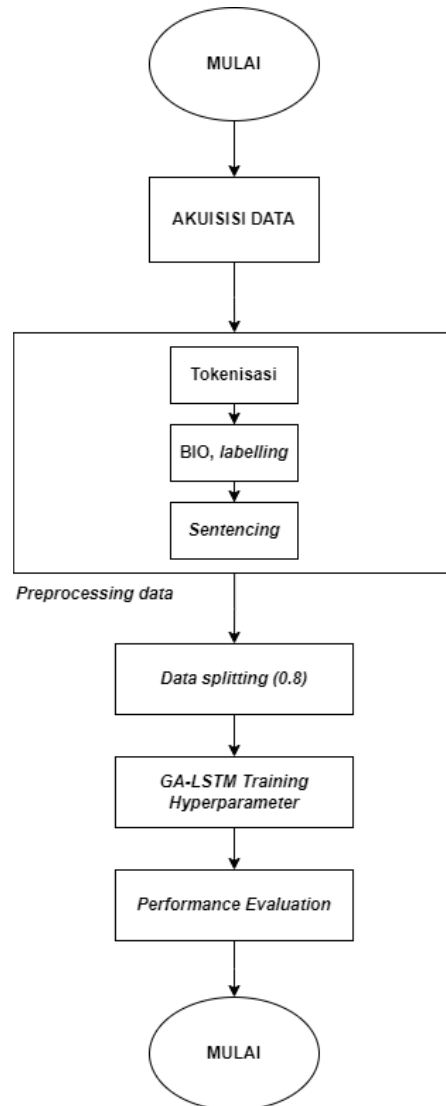
Sejumlah penelitian sebelumnya telah mengeksplorasi pengoptimalkan performa model LSTM dengan hyperparameter dalam berbagai bidang, seperti penggunaan Wolf Optimizer untuk model bahasa (Aufa dkk., 2020), prediksi pasar saham (Rokhsatyazdi dkk., 2020), dan peramalan harga Bitcoin (Kervanci & Akay, 2023). Penerapan algoritma ini juga telah ditemui dalam komputasi umum (Parizad & Hatziadoniou, 2022) serta prediksi penyakit Parkinson dengan CNN-LSTM (Lilhore dkk., 2023). Dalam konteks lain, penelitian telah menyelidiki optimasi hyperparameter dalam kebutuhan panas dan listrik gedung (Morteza, 2023), sentimen analisis (Kandhro, 2020), peramalan permintaan (Abbasimehr dkk., 2020), dan perbandingan normalisasi batch dan dropout dalam deep learning (Garbin dkk., 2020). Menyatukan wawasan dari beragam penelitian tersebut, penelitian ini memfokuskan pada pengoptimalkan hyperparameter Dropout dan Learning Rate pada model LSTM untuk mendeteksi entitas waktu dan lokasi kebakaran hutan, mengintegrasikan pengetahuan lintas-disiplin untuk meningkatkan efektivitas deteksi kebakaran dan upaya mitigasi di Indonesia.

Penelitian ini bertujuan untuk mengoptimalkan hyperparameter learning rate dan dropout, dengan menggunakan algoritma genetika dalam konteks deteksi entitas kebakaran hutan. Dengan mengintegrasikan konsep-konsep ini, penelitian ini berusaha untuk meningkatkan akurasi dan ketepatan model dalam mengidentifikasi waktu dan lokasi kebakaran, memberikan kontribusi positif terhadap upaya mitigasi dan pencegahan bencana kebakaran di Indonesia.

2 METODE

Metodologi penelitian ini dimulai dengan akuisisi data, yang melibatkan pengumpulan sebanyak 19,845 tweet sebagai sumber informasi utama. Selanjutnya, dilakukan proses tokenisasi untuk memecah teks menjadi unit data yang lebih kecil, membentuk dasar pemrosesan lanjutan. Proses pelabelan selanjutnya menggunakan skema BIO (Begin-Interior-Out), di mana kata-kata dalam teks diberi label sesuai peran dan hubungannya dalam entitas yang dimaksud. Sebagai contoh, kata-kata umum seperti "Terjadi," "Kebakaran," dan "Hutan" diberi label "O," menandakan bahwa kata-kata tersebut berada di luar entitas bernama. Sementara itu, kata-kata yang berkaitan dengan waktu seperti "17:00" diberi label "B-Time" (Beginning), dan "Sore" diberi label "I-Time" (Inside) (Lin, dkk., 2020). Setelah pelabelan, dilakukan sentensasi untuk mengembalikan struktur kalimat yang utuh, menghasilkan total 1103 kalimat.

Langkah berikutnya adalah pembagian data menjadi dua bagian, di mana 80% digunakan sebagai data pelatihan dan 20% sebagai data pengujian. Implementasi model menggunakan metode LSTM (Staudemeyer & Morris, 2019) dan Genetic Algorithm (Mirjalili, 2019) melibatkan 10 epoch dengan ukuran batch sebesar 32. Pilihan metode ini bertujuan untuk meningkatkan efisiensi dan akurasi pembelajaran jaringan saraf. Evaluasi kinerja dilakukan menggunakan metrik akurasi dan loss, yang dipantau dengan bantuan library PlotLossesCallback. Metodologi ini menggabungkan langkah-langkah cermat dan efisien, mulai dari tokenisasi hingga implementasi algoritma yang dipilih, seperti terlihat pada Gambar 1. Keseluruhan proses ini membentuk dasar untuk memahami data tweet dengan mendalam, khususnya terkait dengan deteksi entitas waktu dan lokasi kebakaran hutan. Pengaturan parameter algoritma genetika yakni dengan jumlah populasi sebanyak 100 dan jumlah generasi sebanyak 50. Adapun *dropout range* antara 0.1 hingga 0.3 dengan *learning rate range* antara 0.001 hingga 0.1.



Gambar 1. Alur Metodologi LSTM dan GA untuk Hyperparameter

Metodologi penelitian ini juga mencakup langkah kritis dalam mengoptimalkan performa model LSTM melalui algoritma genetika (GA). Setelah pembagian data, fokus utama adalah pada pengaturan hyperparameter yang krusial dalam model. GA diterapkan untuk mengoptimalkan dua hyperparameter, yaitu Dropout dan Learning Rate, dengan tujuan meningkatkan adaptabilitas dan ketepatan model terhadap variasi data serta kompleksitas masalah.

Proses optimasi dimulai dengan inialisasi populasi parameter, di mana individu dalam populasi mewakili kombinasi nilai Dropout dan Learning Rate. Melalui mekanisme seleksi genetika, individu yang memiliki kinerja lebih baik dalam menangani data pelatihan dipertahankan dan direproduksi untuk menciptakan generasi baru. Crossover dan mutasi diterapkan untuk menciptakan variasi dalam populasi, mirip dengan evolusi alamiah.

Pemilihan individu terbaik dari setiap generasi menjadi parameter optimal yang digunakan dalam model LSTM. Proses ini diulang selama beberapa generasi untuk mencapai konvergensi terhadap

kombinasi hyperparameter yang menghasilkan performa terbaik. Dengan menerapkan GA pada hyperparameter Dropout dan Learning Rate, metodologi ini tidak hanya menggabungkan langkah-langkah cermat dalam pemrosesan teks dan implementasi model, tetapi juga memastikan bahwa model LSTM dapat belajar dengan efisien dan optimal untuk mendeteksi entitas waktu dan lokasi kebakaran hutan. Gambar 1 memberikan visualisasi proses metodologi ini secara keseluruhan.

3 HASIL DAN PEMBAHASAN

Tabel 3 menunjukkan lima konfigurasi terbaik dari kombinasi hyperparameter Dropout dan Learning Rate yang dihasilkan oleh algoritma genetika selama proses optimasi. Evaluasi performa dilakukan berdasarkan metrik akurasi dan loss pada data pengujian. Berikut adalah pembahasan komprehensif hasil yang diperoleh:

Tabel 3. Lima Konfigurasi Terbaik

No	Dropout	Learning Rate	Akurasi	Loss
1	0.103758	0.057275	0.750317	1.449127
2	0.182932	0.052663	0.743801	1.361733
3	0.247384	0.019447	0.759095	1.356593
4	0.223387	0.061597	0.748597	1.308936
2	0.104044	0.009626	0.757104	1.279542

Konfigurasi terbaik ditemukan pada baris ke-3, dengan nilai Dropout sebesar 0.247384 dan Learning Rate sebesar 0.019447. Model yang dihasilkan dengan konfigurasi ini mencapai akurasi sebesar 75.91% dengan loss sebesar 1.356593. Dalam hal ini, akurasi mencerminkan sejauh mana model mampu mengenali entitas waktu dan lokasi kebakaran hutan dengan benar, sedangkan loss menunjukkan seberapa baik model dalam mengukur kesalahan selama proses pembelajaran.

Perlu diperhatikan bahwa konfigurasi lainnya juga menunjukkan hasil yang cukup baik. Konfigurasi ke-4, dengan Dropout 0.223387 dan Learning Rate 0.061597, memberikan akurasi 74.86% dan loss 1.308936. Konfigurasi ini menunjukkan bahwa variasi kecil dalam hyperparameter dapat tetap menghasilkan model yang handal. Meskipun konfigurasi terbaik ditemukan pada baris ke-3, variasi dalam performa antar konfigurasi tidak terlalu signifikan. Hasil ini menunjukkan kestabilan algoritma genetika dalam menemukan solusi yang baik di sekitar ruang parameter yang kompleks.

Selanjutnya, hasil ini memberikan gambaran bahwa pengaturan hyperparameter sangat mempengaruhi performa model LSTM dalam mendeteksi entitas waktu dan lokasi kebakaran hutan. Dropout, yang mengontrol jumlah neuron yang diabaikan selama pelatihan, dan Learning Rate, yang menentukan seberapa besar model beradaptasi terhadap data, memainkan peran krusial dalam memastikan model dapat belajar dengan efisien dan menghasilkan prediksi yang akurat. Pembahasan ini memberikan pemahaman mendalam tentang kontribusi konfigurasi hyperparameter terhadap kinerja model, dan implikasinya terhadap deteksi kebakaran hutan. Hasil ini membuka potensi untuk penelitian lebih lanjut dalam pengembangan model yang lebih canggih dan optimal dalam mengatasi tantangan deteksi kebakaran hutan.

4 KESIMPULAN

Kesimpulan dari penelitian ini menegaskan bahwa pengoptimalkan hyperparameter dropout dan learning rate pada model jaringan saraf rekuren, khususnya LSTM, dengan menggunakan algoritma genetika, menghasilkan peningkatan signifikan dalam deteksi entitas kebakaran hutan. Temuan utama adalah bahwa konfigurasi terbaik, dengan dropout 0.247384 dan learning rate 0.019447, memberikan akurasi tertinggi sebesar 75.91%, menunjukkan kemampuan model untuk mengidentifikasi waktu dan lokasi kebakaran dengan tingkat kepercayaan yang baik. Hasil ini tidak hanya mengonfirmasi keberhasilan algoritma genetika dalam mencapai solusi optimal, tetapi juga menyoroti pentingnya pengaturan hyperparameter dalam meningkatkan kinerja model. Relevansi temuan ini terhadap mitigasi bencana kebakaran di Indonesia sangat besar, memperkuat peran teknologi kecerdasan buatan dalam upaya pencegahan dan respons terhadap bencana alam. Untuk pengembangan lebih lanjut, disarankan untuk mengeksplorasi metode optimasi lainnya dan meningkatkan kompleksitas model guna menangani variasi data yang lebih luas serta meningkatkan ketepatan deteksi.

UCAPAN TERIMAKASIH

Ucapan terimakasih disampaikan oleh penulis kepada Universitas AMIKOM Yogyakarta dan Konsorsium SILVANUS yang telah memberikan fasilitas komputasi dan kesempatan dalam pengolahan data kebakaran hutan.

DAFTAR PUSTAKA

- Pasai, M. (2020). Dampak kebakaran hutan dan penegakan hukum. *Jurnal pahlawan*, 3(1), 36-46.
- Salsabila, H. N., Sahitya, A. F., & Mahyatar, P. (2020, July). Spatio-temporal pattern analysis of forest fire event in South Kalimantan using integration remote sensing data and GIS for forest fire disaster mitigation. *In IOP Conference Series: Earth and Environmental Science* (Vol. 540, No. 1, p. 012011). IOP Publishing.
- Nuruddin, A. A. (2022). Mempersiapkan Pengendalian Kebakaran Hutan di Masa Datang: Teknologi dan Inovasi Baru. *Journal of Tropical Silviculture*, 13(01), 29-36.
- Pallavi, B. G., Kumar, E. R., Karnati, R., & Kumar, R. A. (2022, March). LSTM Based Named Entity Chunking and Entity Extraction. *In 2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)* (pp. 1-4). IEEE.
- Putra, A. A., & Kurniawan, R. (2020). Bidirectional LSTM-CNNs Untuk Ekstraksi Entity Lokasi Kebakaran Pada Berita Online Berbahasa Indonesia. *In Seminar Nasional Official Statistics* (Vol. 2020, No. 1, pp. 319-327).
- Li, W., Ng, W. W., Wang, T., Pelillo, M., & Kwong, S. (2021). HELP: An LSTM-based approach to hyperparameter exploration in neural network learning. *Neurocomputing*, 442, 161-172.
- Reimers, N., & Gurevych, I. (2017). Optimal hyperparameters for deep lstm-networks for sequence labeling tasks. *arXiv preprint arXiv:1707.06799*.
- Aufa, B. Z., Suyanto, S., & Arifianto, A. (2020, August). Hyperparameter setting of LSTM-based language model using grey wolf optimizer. *In 2020 international conference on data science and its applications (ICoDSA)* (pp. 1-5). IEEE.
- Rokhsatyazdi, E., Rahnamayan, S., Amirinia, H., & Ahmed, S. (2020, July). Optimizing LSTM based network for forecasting stock market. *In 2020 IEEE congress on evolutionary computation (CEC)* (pp. 1-7). IEEE.
- sibel KERVANCI, I., & Fatih, A. K. A. Y. (2023). LSTM Hyperparameters optimization with Hparam parameters for Bitcoin Price Prediction. *Sakarya University Journal of Computer*

- and Information Sciences*, 6(1), 1-9.
- Parizad, A., & Hatziaodniu, C. (2021). Deep learning algorithms and parallel distributed computing techniques for high-resolution load forecasting applying Hyperparameter optimization. *IEEE Systems Journal*, 16(3), 3758-3769.
- Lilhore, U. K., Dalal, S., Faujdar, N., Margala, M., Chakrabarti, P., Chakrabarti, T., ... & Velmurugan, H. (2023). Hybrid CNN-LSTM model with efficient hyperparameter tuning for prediction of Parkinson's disease. *Scientific Reports*, 13(1), 14605.
- Morteza, A., Yahyaieian, A. A., Mirzaeibonehkhater, M., Sadeghi, S., Mohaimeni, A., & Taheri, S. (2023). Deep learning hyperparameter optimization: Application to electricity and heat demand prediction for buildings. *Energy and Buildings*, 289, 113036.
- Kandhro, I. A., Jumani, S. Z., Ali, F., Shaikh, Z. U., Arain, M. A., & Shaikh, A. A. (2020). Performance analysis of hyperparameters on a sentiment analysis model. *Engineering, Technology & Applied Science Research*, 10(4), 6016-6020.
- Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers & industrial engineering*, 143, 106435.
- Garbin, C., Zhu, X., & Marques, O. (2020). Dropout vs. batch normalization: an empirical study of their impact to deep learning. *Multimedia Tools and Applications*, 79, 12777-12815.
- Lin, J. C. W., Shao, Y., Zhang, J., & Yun, U. (2020). Enhanced sequence labeling based on latent variable conditional random fields. *Neurocomputing*, 403, 431-440.
- Mirjalili, S., & Mirjalili, S. (2019). Genetic algorithm. *Evolutionary Algorithms and Neural Networks: Theory and Applications*, 43-55.
- Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM--a tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv:1909.09586*.