

## ANALISIS PERFORMA ALGORITMA CNN-LSTM PADA IMAGE CAPTIONING UNTUK MENDESKRIPSIKAN KESEGARAN BUAH

Yosua Calvin<sup>1</sup>, Herry Sujaini<sup>2</sup>, Khairul Hafidh<sup>3</sup>

[yosuacalvin@student.untan.ac.id](mailto:yosuacalvin@student.untan.ac.id)<sup>1</sup>, [hs@untan.ac.id](mailto:hs@untan.ac.id)<sup>2</sup>, [hafidh@informatika.untan.ac.id](mailto:hafidh@informatika.untan.ac.id)<sup>3</sup>

Universitas Tanjungpura

### ABSTRAK

Penelitian mengenai image captioning semakin berkembang, dengan berbagai objek penelitian dan algoritma yang diterapkan untuk melatih model image captioning guna menghasilkan deskripsi yang baik. Algoritma populer dalam image captioning adalah CNN-LSTM, di mana CNN digunakan sebagai ekstraktor fitur gambar dan LSTM sebagai penghasil deskripsi otomatis. Objek penelitian dalam penelitian ini adalah mendeskripsikan kesegaran buah. Buah banyak dikonsumsi oleh masyarakat karena kandungan gizi yang baik untuk kesehatan. Namun, informasi mengenai kesegaran buah masih dinilai secara manual. Penelitian ini bertujuan untuk menganalisis performa algoritma CNN-LSTM dalam mendeskripsikan kesegaran buah menggunakan image captioning. Dataset gambar yang digunakan terbagi menjadi dua kategori, yaitu original dan crop, masing-masing dengan 900 deskripsi referensi yang sama. CNN dengan arsitektur ResNet50 dari Keras digunakan untuk mengekstraksi fitur gambar, sementara LSTM digunakan untuk menghasilkan urutan teks deskripsi prediksi. Pengujian dilakukan pada model original dan model crop dengan dua jenis data uji, yaitu data uji original dan data uji crop. Analisis hasil pengujian dilakukan dengan dua cara yaitu kuantitatif dan kualitatif. Analisis kuantitatif menggunakan metrik BLEU dengan dua n-gram yaitu BLEU-3 dan BLEU-4. Hasil pengujian menunjukkan bahwa skenario model crop terhadap data uji crop memberikan nilai BLEU tertinggi dengan nilai BLEU-3 dan BLEU-4 berturut-turut sebesar 0.4565 dan 0.3177. Analisis kualitatif dengan membandingkan deskripsi prediksi dengan gambar data uji. Hasil pengujian menunjukkan bahwa skenario model crop terhadap data uji crop mencapai persentase nilai Tepat tertinggi yaitu sebesar 70.00%. Penelitian ini menunjukkan bahwa skenario model crop terhadap data uji crop lebih unggul dalam mendeskripsikan kesegaran buah. Dengan melakukan cropping 10% pada setiap sisi gambar dataset, performa algoritma CNN-LSTM dalam mendeskripsikan kesegaran buah pada image captioning dapat meningkat.

**Kata Kunci:** Image Captioning, Cnn, Lstm, Kesegaran Buah, Analisis Kuantitatif, Analisis Kualitatif.

### ABSTRACT

Research on image captioning is growing, with various research objects and algorithms applied to train image captioning models to produce good descriptions. A popular algorithm in image captioning is CNN-LSTM, where CNN is used as an image feature extractor and LSTM as an automatic description generator. The research object in this study is to describe the freshness of fruit. Fruit is widely consumed by the public because of its nutritional content that is good for health. However, information about fruit freshness is still assessed manually. This research aims to analyze the performance of CNN-LSTM algorithm in describing fruit freshness using image captioning. The image dataset used is divided into two categories, namely original and crop, each with the same 900 reference descriptions. CNN with Keras ResNet50 architecture is used to extract image features, while LSTM is used to generate a sequence of predictive description texts. Tests were conducted on the original and crop models with two types of test data, namely original test data and crop test data. Analysis of the test results is done in two ways, namely quantitative and qualitative. Quantitative analysis uses BLEU metrics with two n-grams, namely BLEU-3 and BLEU-4. The test results show that the crop model scenario on the crop test data provides the highest BLEU value with BLEU-3 and BLEU-4 values of 0.4565 and 0.3177, respectively. Qualitative analysis by comparing the prediction description with the test data image. The test results show that the crop model scenario on the crop test data achieves the highest percentage of correct value which is

70.00%. This research shows that the crop model scenario on the crop test data is superior in describing the freshness of the fruit. By cropping 10% on each side of the dataset image, the performance of the CNN-LSTM algorithm in describing fruit freshness in image captioning can be improved.

**Keywords:** Image Captioning, Cnn, Lstm, Fruit Freshness, Quantitative Analysis, Qualitative Analysis.

## PENDAHULUAN

Dalam beberapa tahun terakhir, computer vision di bidang pemrosesan gambar telah membuat kemajuan yang signifikan seperti klasifikasi gambar dan deteksi objek. Manfaat dari kemajuan pada bidang klasifikasi objek dan deteksi objek memungkinkan untuk secara otomatis menghasilkan satu kalimat atau lebih untuk menjelaskan konten visual dari suatu gambar, yang dikenal dengan Image Captioning. Membuat deskripsi gambar yang lengkap dan alami secara otomatis memiliki manfaat yang besar, seperti pada pembuatan judul yang dilampirkan pada gambar berita, deskripsi yang terkait dengan gambar medis, pengambilan gambar berbasis teks, informasi yang diakses oleh pengguna tunanetra, interaksi manusia-robot (Liu et al., 2018). Bagi penyandang tunanetra, image captioning mampu memberikan manfaat yang besar dalam menjelaskan gambaran atas apa yang sedang terjadi dan apa yang ada di sekitar mereka. Image captioning juga memiliki beberapa peran penting untuk berbagai macam tujuan serta dapat diaplikasikan dalam berbagai skenario seperti menambahkan takarir pada video, video tanya jawab, pencarian gambar, pencarian informasi penting menurut konten gambar, dan pencarian menurut kata kunci (Wang et al., 2021).

Menghasilkan deskripsi dengan bahasa alami yang memiliki makna dari sebuah gambar memerlukan tingkat pemahaman yang lebih kompleks dibandingkan dengan deteksi dan mengklasifikasikan objek dalam gambar. Permasalahan tersebut sangat menarik karena menghubungkan antara Computer vision dan Natural Language Processing (NLP) yang merupakan dua gabungan bidang utama Artificial Intelligence (You et al., 2016). Kedua bidang tersebut memiliki peran masing-masing dalam proses pendeskripsian suatu gambar. Computer vision digunakan komputer agar dapat mengenali objek yang terdapat pada gambar, sedangkan NLP untuk menghasilkan deskripsi gambar berupa bahasa alami suatu model bahasa (Devlin et al., 2015).

Salah satu teknologi Computer vision yang paling banyak digunakan untuk pengolahan gambar adalah Convolutional Neural Network (CNN). CNN merupakan metode yang banyak digunakan untuk pekerjaan seperti deteksi gambar, klasifikasi gambar, dan segmentasi gambar dengan akurasi baik. Selain itu fitur gambar yang dipelajari oleh CNN dapat digunakan untuk pembuatan model deskripsi gambar (Pu et al., 2016). Setelah proses pengenalan fitur gambar oleh CNN dilakukan, maka dilanjutkan dengan proses pendeskripsian dalam teknologi NLP menggunakan model Recurrent Neural Network (RNN). Namun, penggunaan RNN terbatas karena tidak mampu memperkirakan kata yang disimpan dalam memori jangka panjang sehingga diperlukan suatu model jaringan yang sanggup melakukannya yaitu Long Short-Term Memory (LSTM).

Teknologi kecerdasan buatan yang selalu mengalami perkembangan ini merambah ke dalam berbagai bidang salah satunya dalam bidang pendeskripsian otomatis (image captioning) dalam mendeskripsikan kesegaran buah. Pada tahun 2022, produksi buah-buahan mengalami peningkatan dibandingkan pada tahun 2021. Dalam perkembangannya kontribusi komoditas buah-buahan cenderung mengalami peningkatan setiap tahunnya (Badan Pusat Statistik, 2022). Kesadaran masyarakat akan manfaat dari komoditas ini menyebabkan permintaan buah-buahan nasional mengalami peningkatan.

Dalam jurnal sebuah jurnal yang ditulis oleh (Paraijun et al., 2022) mengutip dari Seminar Hasil-Hasil Penelitian IPB Tahun 2013 mengatakan bahwa di Indonesia sendiri kualitas buah nasional dilihat dari sisi konsistensi ukuran, kematangan petik, warna, rasa, dan kesegaran masih jauh dari harapan konsumen karena banyaknya buah yang tak dapat diklasifikasikan kesegarannya secara keseluruhan. Kesegaran buah merupakan faktor penting dalam menentukan kualitas buah dan daya tahan buah. Namun, dalam industri hortikultura, sering kali sulit untuk secara konsisten mengklasifikasikan kesegaran buah secara akurat. Variasi dalam warna, kematangan, ukuran, dan tekstur buah menyebabkan kesulitan dalam memberikan deskripsi yang objektif dan konsisten. Permasalahan ini memerlukan pengembangan teknologi yang dapat secara otomatis mendeskripsikan kesegaran buah dengan akurasi yang baik.

Dalam industri buah, informasi yang akurat dan deskriptif tentang tingkat kesegaran buah sangat penting dalam proses penjualan, distribusi, dan manajemen persediaan. Dengan adanya teknologi image captioning, deskripsi gambar yang otomatis dan alami dapat menggambarkan secara detail tingkat kesegaran buah berdasarkan fitur visual yang diidentifikasi oleh model komputer. Hal ini dapat mempermudah produsen, distributor, maupun konsumen untuk membuat keputusan lebih baik berdasarkan informasi yang akurat dan objektif.

Sebelumnya terdapat penelitian yang berkaitan dengan deskripsi gambar (image captioning), yaitu penelitian yang dilakukan oleh (Artyani, 2019) yang menggunakan algoritma CNN dengan membandingkan dua model CNN yaitu VGG-16 dan ResNet50 untuk menghasilkan deskripsi gambar lalu lintas kendaraan. Dari penelitian tersebut dihasilkan sebagai berikut. Pada model ke-1 yang menggunakan VGG-16 pada model ekstraksi fitur gambar menghasilkan hasil skor BLEU {1, 2, 3, 4} masing-masing sebesar 0,66; 0,50; 0,43; 0,30. Sementara untuk model ke-2 yang menggunakan ResNet50 pada model ekstraksi fitur gambar menghasilkan skor BLEU {1, 2, 3, 4} sebesar 0,67; 0,51; 0,42; 0,27.

Berdasarkan latar belakang di atas, akan dilakukan eksperimen dalam mendeskripsikan kesegaran buah menggunakan mesin image captioning dengan algoritma CNN-LSTM yang kemudian akan dianalisis performa menggunakan metrik evaluasi BLEU (Bilingual Evaluation Understudy).

## **METODOLOGI**

Penelitian ini mengadopsi pendekatan deep learning yang memanfaatkan arsitektur jaringan syaraf tiruan, khususnya menggunakan Convolutional Neural Network (CNN) dengan model ResNet50 untuk ekstraksi fitur gambar dan Long Short-Term Memory (LSTM) untuk menghasilkan deskripsi otomatis. Data yang digunakan terdiri dari gabungan data primer yang dikumpulkan langsung dari beberapa toko buah dan data sekunder yang diperoleh dari internet. Kedua jenis data ini diproses dan dianalisis menggunakan perangkat keras dan lunak yang memadai, termasuk komputer dengan spesifikasi tertentu dan berbagai perangkat lunak pengembangan seperti Google Colaboratory dan TensorFlow.

Metodologi penelitian terstruktur dalam beberapa tahapan utama yang meliputi pengumpulan data, praproses data, pelatihan (training), pengujian (testing), dan analisis hasil. Proses pengumpulan data melibatkan akuisisi gambar buah yang segar dan tidak segar dari sumber primer dan sekunder, yang kemudian diatur distribusinya dalam dataset. Setiap gambar dilengkapi dengan deskripsi yang sesuai untuk membantu dalam proses pelatihan dan pengujian model. Praproses data melibatkan pengaturan ulang ukuran gambar dan pembersihan dataset deskripsi dari kebisingan data.

Fase pelatihan data menggunakan teknik encoder-decoder dimana encoder gambar dilengkapi dengan arsitektur ResNet50 untuk ekstraksi fitur, dan decoder menggunakan LSTM untuk memproses deskripsi. Model ini dilatih untuk mengoptimalkan pengenalan fitur dari gambar dan generasi teks yang relevan. Proses training diikuti dengan fase testing dimana model yang telah dilatih diuji untuk menghasilkan deskripsi pada data baru dan mengukur keefektifan model menggunakan metrik BLEU, yang menilai kesesuaian prediksi teks dengan referensi.

Hasil dari pelatihan dan pengujian kemudian dianalisis untuk mengevaluasi kinerja model, meliputi evaluasi kuantitatif melalui skor BLEU dan analisis kualitatif dari kesesuaian prediksi dengan data asli. Analisis ini membantu dalam mengidentifikasi keberhasilan model dalam konteks aplikasi nyata. Tahap analisis diakhiri dengan penarikan kesimpulan berdasarkan kinerja model dalam menghasilkan deskripsi yang akurat dan relevan dari gambar buah yang diuji.

Keseluruhan penelitian ini mencerminkan penggunaan canggih teknologi deep learning dalam pengolahan gambar dan natural language processing, yang menunjukkan potensi besar dalam aplikasi industri pengolahan data visual dan produksi teks otomatis. Dengan pendekatan yang sistematis dan alat yang tepat, penelitian ini memberikan wawasan baru dalam pengembangan model pembelajaran mesin untuk aplikasi praktis.

## **HASIL DAN PEMBAHASAN**

Penelitian ini mengusung judul "Analisis Performa Algoritma CNN-LSTM Pada Image Captioning Untuk Mendeskripsikan Kesegaran Buah", di mana CNN berfungsi sebagai ekstraktor fitur gambar dan LSTM bertugas menghasilkan deskripsi. Sebanyak 300 gambar buah segar dan busuk telah dikumpulkan untuk dataset, yang kemudian disimpan dalam google drive dan didistribusikan ke dalam subkategori untuk analisis lebih lanjut.

Dalam proses pembuatan dataset, gambar dipisah menjadi buah segar dan busuk dengan proporsi tertentu dan dideskripsikan sesuai subbab metodologi. Gambar yang sudah di-crop dan di-resize menjadi 224x224 piksel untuk memenuhi syarat input model ResNet50, digunakan sebagai dasar untuk ekstraksi fitur menggunakan model ini. Selanjutnya, gambar diubah menjadi array numpy dan diproses melalui fungsi `preprocess_input` dari model ResNet50 untuk normalisasi.

Deskripsi yang sesuai dengan gambar buah segar dan busuk dihasilkan, dipetakan, dan dibersihkan dari karakter non-alfabetik serta spasi ekstra untuk memastikan keseragaman data. Proses tokenizing mengonversi teks deskripsi menjadi kata-kata individu dan menyimpannya dalam sebuah vocabular internal menggunakan Tokenizer dari Keras.

Setelah dataset dibagi menjadi data latih, validasi, dan uji dengan proporsi 80:10:10, pembuatan model melibatkan pembangunan fungsi data generator yang berfungsi menghasilkan data dalam batch. Model yang dibangun menggunakan arsitektur encoder dan decoder, di mana encoder menangani ekstraksi fitur dan urutan teks sementara decoder menggabungkan kedua output tersebut untuk prediksi.

Model dilatih dengan parameter yang telah ditetapkan, menggunakan teknik EarlyStopping untuk mencegah overfitting. Hasil pelatihan menunjukkan bahwa model berhasil menghentikan pelatihan pada epoch ke-17 sebagai epoch dengan performa terbaik berdasarkan `val_loss`.

Pengujian model menggunakan dataset uji memperlihatkan bahwa fungsi prediksi yang dibuat mampu menghasilkan deskripsi yang relatif akurat, yang ditunjukkan melalui kemampuannya menghasilkan deskripsi yang relevan dengan kondisi buah yang dianalisis. BLEU scores yang diperoleh dari pengujian memberikan indikasi efektivitas model dalam menghasilkan deskripsi yang semakin mendekati deskripsi sebenarnya.

Analisis kesesuaian antara deskripsi yang dihasilkan dengan gambar yang sesungguhnya menunjukkan bahwa model mampu mengidentifikasi kesegaran buah dengan akurat dalam sebagian besar kasus, namun masih ada ruang untuk perbaikan, terutama dalam konsistensi output yang dihasilkan oleh model.

Dalam kajian ini, penggunaan model CNN-LSTM untuk image captioning menunjukkan potensi yang baik dalam aplikasi praktis seperti mengidentifikasi kesegaran buah, yang bisa sangat berguna dalam konteks industri pangan dan ritel. Namun, diperlukan lebih banyak penelitian untuk meningkatkan akurasi dan mengurangi variabilitas hasil yang diberikan oleh model terutama dalam kondisi buah yang kurang standar.

## **KESIMPULAN**

Berdasarkan hasil dari pengumpulan data, praproses data, training, testing, hingga analisis hasil yang telah dilakukan pada penelitian ini yang berjudul, “Analisis Performa Algoritma CNN-LSTM Pada Image Captioning Untuk Mendeskripsikan Kesegaran Buah” dapat ditarik kesimpulan sebagai berikut.

1. Analisis kuantitatif dengan nilai BLEU menunjukkan nilai tertinggi pada skenario model crop terhadap data uji crop dengan BLEU-3 dan BLEU-4 berturut sebesar 0.4565 dan 0.3177. Sementara tertinggi kedua yaitu pada skenario model original terhadap data uji original dengan nilai BLEU-3 dan BLEU-4 berturut sebesar 0.4532 dan 0.2937. Skenario model original terhadap model crop memiliki performa lebih rendah yakni dengan nilai BLEU-3 dan BLEU-4 berturut sebesar 0.4094 dan 0.2746. Sementara nilai performa terendah dialami pada skenario model crop terhadap data uji original dengan nilai BLEU-3 dan BLEU-4 sebesar 0.4058 dan 0.2618.
2. Analisis kualitatif dengan membandingkan antara deskripsi prediksi dengan gambar uji. Nilai tertinggi ada pada skenario model crop terhadap data uji crop dengan persentase Tepat sebesar 70.00%, Kurang Tepat sebesar 26.57%, dan 3.33%. Sementara tertinggi kedua terjadi pada skenario model original terhadap data uji original dengan persentase Tepat sebesar 66.67%, Kurang Tepat sebesar 30.00%, dan Tidak Tepat sebesar 3.33%. Skenario model original terhadap data uji crop memiliki performa lebih rendah dengan persentase Tepat sebesar 63.33%, Kurang Tepat sebesar 3.33%, dan Tidak Tepat sebesar 3.33%. Sementara nilai performa terendah dialami oleh skenario model crop terhadap data uji original dengan persentase Tepat sebesar 43.33%, Kurang Tepat sebesar 53.33%, dan Tidak Tepat sebesar 3.33%.
3. Secara keseluruhan, performa terbaik ditunjukkan pada skenario model crop terhadap data uji crop dengan nilai analisis kuantitatif dan analisis kualitatif tertinggi. Hal ini menunjukkan bahwa dengan melakukan modifikasi pada dataset dengan cropping sebesar 10% setiap sisinya dapat meningkatkan performa algoritma CNN-LSTM dalam tugas image captioning.

## **DAFTAR PUSTAKA**

- Aini, Q., & Amrizal, V. (2013). Kecerdasan Buatan. Halaman Moeka Publishing. Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a Convolutional Neural Network. 2017 International Conference on Engineering and Technology (ICET), 1–6. <https://doi.org/10.1109/ICEngTechnol.2017.8308186>
- Aldi, M. W. P., Jondri, & Aditsania, A. (2018). Analisis dan Implementasi Long Short- Term Memory Neural Network untuk Prediksi Harga Bitcoin.
- Al-Faruq, U. A. A. (2021). Implementasi Arsitektur Transformer Pada Image Captioning Dengan Bahasa Indonesia.
- Artyani, I. (2019). Simulasi Metode Convolutional Neural Network Dan Long Short-Term Memory Untuk Generate Image Captioning Pada Gambar Lalu Lintas Kendaraan Berbahasa Indonesia.

- Badan Pusat Statistik. (2022). Statistik Hortikultura 2022 (pp. 33–53). Badan Pusat Statistik Republik Indonesia.
- Brownlee, J. (2017). Deep Learning for Natural Language Processing: Develop Deep Learning Model for Natural Language in Python. 271–275.
- Devlin, J., Cheng, H., Fang, H., Gupta, S., Deng, L., He, X., Zweig, G., & Mitchell, M. (2015). Language Models for Image Captioning: The Quirks and What Works. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 100–105. <https://doi.org/10.3115/v1/P15-2017>
- Geron, A. (2019). Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems SECOND EDITION. O'Reilly Media.
- Goldberg, D. E., & Holland, J. H. (1988). Genetic Algorithms and Machine Learning. 3, 95–99.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition (arXiv:1512.03385). arXiv. <http://arxiv.org/abs/1512.03385>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hossain, M. Z., Sohel, F., Shiratuddin, M. F., & Laga, H. (2018). A Comprehensive Survey of Deep Learning for Image Captioning (arXiv:1810.04020). arXiv. <http://arxiv.org/abs/1810.04020>
- Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1–3), 489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>
- KBBI. (2016). Kamus Besar Bahasa Indonesia (KBBI Online) (VI). Badan Pengembangan dan Pembinaan Bahasa. <https://kbbi.kemdikbud.go.id/>
- Liu, S., Bai, L., Hu, Y., & Wang, H. (2018). Image Captioning Based on Deep Neural Networks. *MATEC Web of Conferences*, 232, 01052. <https://doi.org/10.1051/mateconf/201823201052>
- Manaswi, N. K. (2018). Deep Learning with Applications Using Python. Apress. <https://doi.org/10.1007/978-1-4842-3516-4>
- Nugraha, A. A., Arifianto, A., & Suyanto. (2019). Generating Image Description on Indonesian Language using Convolutional Neural Network and Gated Recurrent Unit. 2019 7th International Conference on Information and Communication Technology (ICoICT), 1–6. <https://doi.org/10.1109/ICoICT.2019.8835370>
- O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks (arXiv:1511.08458). arXiv. <http://arxiv.org/abs/1511.08458>
- Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02, 311. <https://doi.org/10.3115/1073083.1073135>
- Paraijun, F., Aziza, R. N., & Kuswardani, D. (2022). Implementasi Algoritma Convolutional Neural Network Dalam Mengklasifikasi Kesegaran Buah Berdasarkan Citra Buah. *KILAT*, 11(1), 1–9. <https://doi.org/10.33322/kilat.v10i2.1458>
- Parlika, R., Pradika, S. I., Hakim, A. M., & N M, K. R. (2020). Analisis Sentimen Twitter Terhadap Bitcoin Dan Cryptocurrency Berbasis Python Textblob. *Jurnal Ilmiah Teknologi Informasi dan Robotika*, 2(2), 33–37. <https://doi.org/10.33005/jifti.v2i2.22>
- Pramestya, R. A. (2018). Deteksi Dan Klasifikasi Kerusakan Jalan Aspal Menggunakan Metode YOLO Berbasis Citra Digital. Institut Teknologi Sepuluh Nopember.
- Prijono, B. (2018). Pengenalan Long Short-Term Memory (LSTM) dan Gate Recurrent Unit (GRU)—RNN Bagian 2. Belajar Pembelajaran Mesin Indonesia. <https://indoml.com/2018/04/13/pengenalan-long-short-term-memory-lstm-dan-gated-recurrent-unit-gru-rnn-bagian-2/>
- Pu, Y., Gan, Z., Henao, R., Yuan, X., Li, C., Stevens, A., & Carin, L. (2016). Variational Autoencoder for Deep Learning of Images, Labels and Captions (arXiv:1609.08976). arXiv. <http://arxiv.org/abs/1609.08976>
- Putra, I. W. S. E., Wijaya, A. Y., & Sulaiman, R. (2016). Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101. *Jurnal Teknik ITS*, 5(1).

- <https://doi.org/10.12962/j23373539.v5i1.15696>
- Satria, M. R., & Pardede, J. (2022). Image Captioning Menggunakan Metode ResNet50 dan Long Short-Term Memory. *Jurnal Tera*, 2(2), 84–94.
- Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016). A review of machine learning techniques using decision tree and support vector machine. 2016 International Conference on Computing Communication Control and Automation (ICCUBEA), 1–7. <https://doi.org/10.1109/ICCUBEA.2016.7860040>
- Sukhdeve, Dr. S. R., & Sukhdeve, S. S. (2023). *Google Cloud Platform for Data Science: A Crash Course on Big Data, Machine Learning, and Data Analytics Services*. Apress. <https://doi.org/10.1007/978-1-4842-9688-2>
- Susatyono, J. D. (2021). *Kecerdasan Buatan Kajian Konsep dan Penerapan*. Yayasan Prima Agus Teknik.
- Taufiq, I. (2018). *Deep Learning Untuk Deteksi Tanda Nomor Kendaraan Bermotor Menggunakan Algoritma Convolutional Neural Network Dengan Python Dan Tensorflow*.
- Vashee, K. (2019). Understanding Machine Translation Quality: BLEU Scores. In *Machine Translation*. <https://www.rws.com/blog/understanding-mt-quality-bleu-scores/>
- Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3156–3164. <https://doi.org/10.1109/CVPR.2015.7298935>
- Wang, C., Zhou, Z., & Xu, L. (2021). An Integrative Review of Image Captioning Research. *Journal of Physics: Conference Series*, 1748(4), 042060. <https://doi.org/10.1088/1742-6596/1748/4/042060>
- You, Q., Jin, H., Wang, Z., Fang, C., & Luo, J. (2016). Image Captioning with Semantic Attention (arXiv:1603.03925). arXiv. <http://arxiv.org/abs/1603.03925>.