PENGENALAN AKTIVITAS MANUSIA DI SUPERMARKET DENGAN METODE LONG SHORT TERM MEMORY

Kristian Davidson Runtu [1)] , Lina [2)]

[1,2)] Teknik Informatika, FTI, Universitas Tarumanagara
Jl. Letjen S Parman no. 1, Jakarta 11440 Indonesia
email : kristian.535180135@stu.untar.ac.id[1)], email : lina@fti.untar.ac.id[2)]

ABSTRACT

Since a long time ago, supermarkets have become people's destinations for shopping for various things such as food, cooking ingredients, cleaning products and others. Supermarkets are known for their very large and crowded places, making it difficult to monitor. Therefore, supermarkets need a system to help monitoring. With the development of technology, monitoring systems are increasingly advanced and one of the results of these technological developments is a system for recognizing human activities. By using OpenPose to obtain human skeleton data on the image and using the Long Short Term Memory method to perform recognition, testing of the training data was carried out so as to produce a precision value of 99%, recall 99%, and f1-score 99%. And real-time testing using a camera resulted in an accuracy value of 73% for the picking class, 87% for the standing class and 81% for the walking class.

Key words

Supermarket, Long Short Term Memory, OpenPose, Activity, Human Activity Recognition

1. Pendahuluan

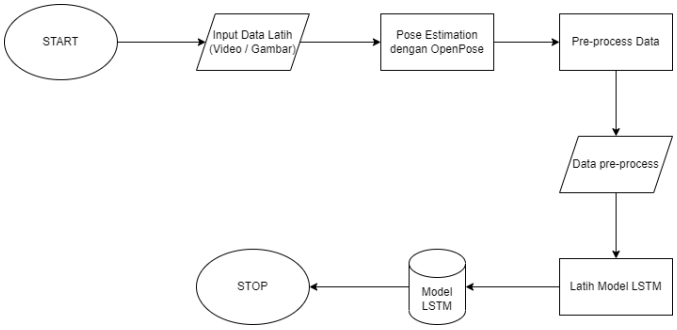
Sejak dari dahulu supermarket sudah menjadi tempat tujuan masyarakat untuk berbelanja berbagai hal seperti makanan, bahan untuk memasak, produk kebersihan dan lain-lain. Supermarket sudah dikenal dengan tempatnya yang sangat besar dan ramai sehingga susah untuk melakukan pemantauan. Maka dari itu dibutuhkan sebuah sistem untuk mempermudah pemantauan. Dengan perkembangan teknologi sistem pemantauan sudah semakin maju dan salah satu hasil perkembangan teknologi tersebut adalah sistem pengenalan aktivitas manusia[1].

Pengenalan aktivitas manusia adalah proses mengidentifikasi dan merekam perubahan posisi dalam postur atau gerak tubuh[2]. Pengenalan aktivitas manusia melalui citra menjadi bidang penelitian yang trendi karena aplikasinya yang luas seperti komunikasi antara komputer dan manusia maupun untuk keamanan dan ada berbagai banyak hal juga[3]. Dalam teknologi jaman sekarang sudah ada beberapa cara untuk melakukan pengolahan data aktivitas manusia seperti CNN, LSTM, BLSTM, MLP and SVM[4].

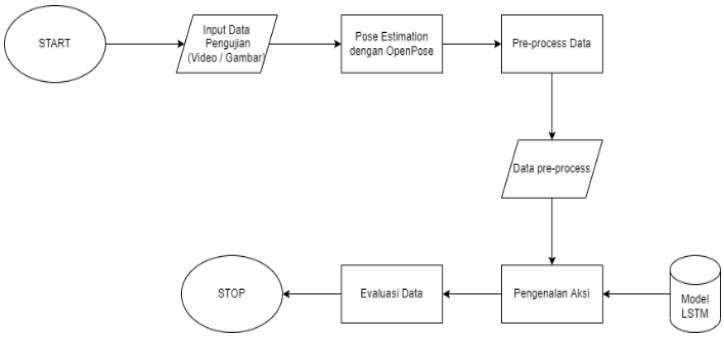
Metode yang akan digunakan adalah Metode *Long Short Term Memory* (LSTM). LSTM adalah salah satu bagian dari *Recurrent Neural Network* (RNN). LSTM memiliki kemampuan untuk menyimpan informasi penting dalam urutan waktu , ini cocok untuk tugas klasifikasi urutan seperti pengenalan aktivitas manusia. LSTM dilatih untuk mengenali aktivitas mana yang dilakukan dengan mempelajari urutannya fitur gerak yang terkait dengan setiap aktivitas [5].

2. Metode Penelitian

Rancangan sistem yang dibuat adalah sebuah program pengenalan aktivitas manusia di supermarket melalui sebuah video yang didapatkan lewat kamera yang dilatih menggunakan metode *Long Short Term Memory* (LSTM).



Gambar 1. *Flowchart* Pelatihan.



Gambar 2. *Flowchart* Pengenalan.

2.1 Dataset

Data dikumpulkan sesuai dengan kelas aktivitas yang ditentukan yaitu mengambil barang (*picking*), berdiri (*standing*), dan berjalan (*walking*). Sumber data yang digunakan berasal dari berbagai *website* penyedia gambar dan *website* penyedia video, untuk data yang diambil secara individu menggunakan kamera *smartphone* untuk mengambil video dan gambar. Seluruh data yang diambil disatukan dan dijadikan dataset.

Tabel 1. Jumlah Data Yang Digunakan.

Kelas Aktivitas	Jumlah Data
Picking	4.512
Standing	3.982
Walking	3.984
TOTAL	12.478

Setelah dikumpulkan dataset yang dipakai, dilakukan proses ekstraksi 25 titik *skeleton* yang didapatkan menggunakan *OpenPose* dimana tiap titik diidentifikasi dengan 3 nilai koordinat yaitu x,y dan z.. Dari data yang terkumpul dilakukan pembagian data menjadi dua kategori yaitu data pelatihan dan data pengujian/validasi.

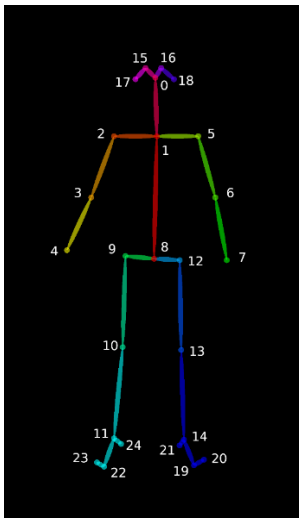
2.2 *OpenPose*

OpenPose adalah sebuah program estimasi bentuk pose manusia yang menghasilkan titik-titik kerangka manusia secara *realtime* yang berbasis *Convolutional Neural Network* dan dibangun pada *framework caffe* [6]. *OpenPose* dikembangkan oleh Universitas Carnegie Mellon (CMU) yang dirilis ke publik sebagai proyek *open source*.

OpenPose menggunakan pendekatan *bottom-up* untuk melakukan estimasi, yang berarti *OpenPose* mendeteksi semua bagian anggota tubuh dari orang-orang yang ada di gambar, kemudian dikelompokkan bagian tubuh mana yang dimiliki orang di gambar.

OpenPose dapat menghasilkan model dengan format *BODY_25*, di mana *OpenPose* mengekstrak 25 titik koordinat pada setiap orang yang ada pada citra dengan pada tiap titik terdapat nilai koordinat x,y dan z.

Gambar 3. Visualisasi Hasil dari *OpenPose*



Gambar 4. Model Format *BODY_25*.

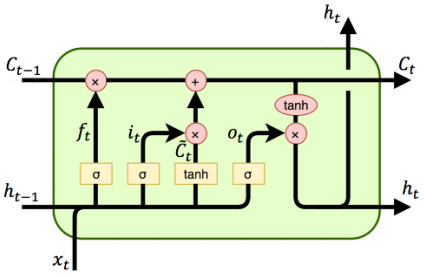
2.3 *Recurrent Neural Network* (RNN)

Jaringan saraf berulang atau *Recurrent Neural Network* (RNN) adalah jenis arsitektur jaringan saraf tiruan yang pemrosesannya dipanggil berulang-ulang untuk memroses masukan yang biasanya adalah data sekuensial [7].

Dalam algoritma RNN terdapat lapisan *Neural Network* seperti biasa yaitu lapisan *input*, lapisan *hidden*, lapisan *output*. RNN diartikan sebagai mekanisme untuk menahan memori yang disimpan dalam lapisan *hidden* [8]. Dalam RNN mempunyai banyak tipe yang salah satunya adalah *Long Short Term Memory* (LSTM).

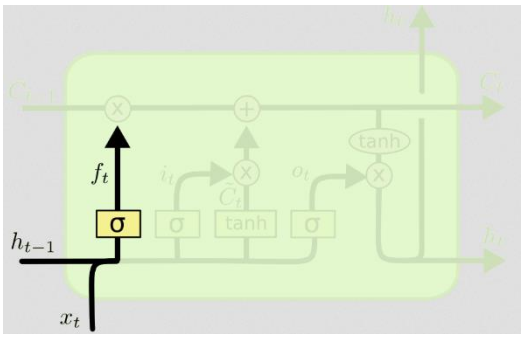
2.4 *Long Short Term Memory* (LSTM)

Long Short Term Memory (LSTM) merupakan salah satu tipe dari *Recurrent Neural Network* (RNN) . Dampak jaringan LSTM sangat menonjol dalam pemodelan bahasa, ucapan-ke-teks transkripsi, terjemahan mesin, dan aplikasi lainnya [9].



Gambar 5. Sel LSTM

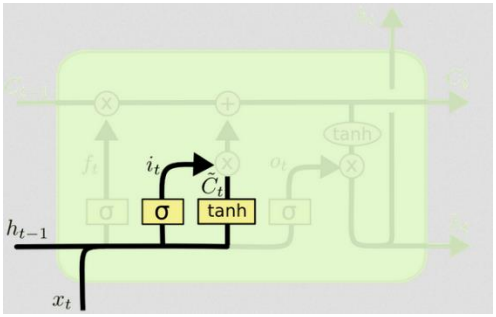
Forget Gate bertugas untuk membuang informasi yang sudah tidak relevan. Bobot dari *state unit* diatur oleh *forget gate*.



Gambar 6. *Forget Gate*

$$f_t = \sigma(W_f\, x_t + U_f\, h_{t-1} + b_f) \quad (1)$$

Nilai dari suatu *input* hanya dapat disimpan ke dalam *cell state* hanya jika diijinkan oleh *input gate*.

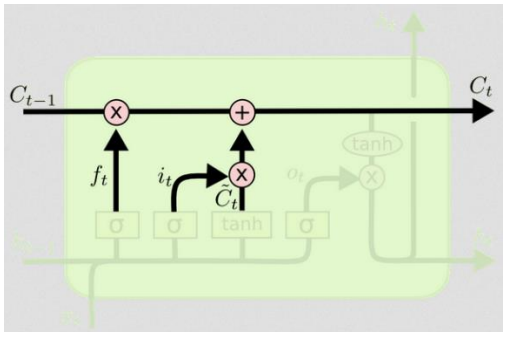


Gambar 7. *Input Gate*

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

Dalam perhitungan *input gate* selanjutnya dilakukan untuk mengetahui informasi baru yang harus disimpan.

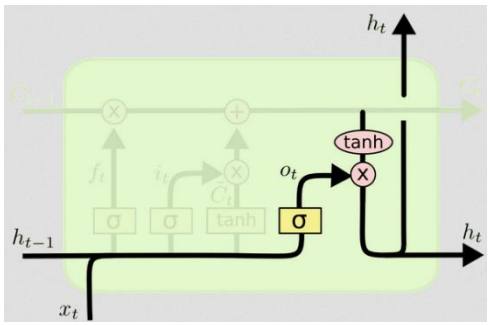
$$\tilde{C}_t = \tan h\,(W_c x_t + U_c h_{t-1} + b_c) \quad (3)$$



Gambar 8. *Cell State*

$$C_t = i_t * \tilde{C}_t + f_t * C_{t-1} \quad (4)$$

Setelah dihasilkan *memory cell state* yang baru, nilai dari *output gate* dapat dihitung dengan menggunakan persamaan.



Gambar 9. *Output Gate*

$$o_t = \sigma(W_o\, x_t + U_o\, h_{t-1} + b_o) \quad (5)$$
$$h_t = o_t * tanh\,(C_t) \quad (6)$$

3. Hasil Percobaan

Model dievaluasi dengan menggunakan akurasi, presisi, *recall*, dan *F1-Score* untuk masing-masing model. Akurasi, merupakan indikator jumlah data yang terklasifikasi benar dari jumlah keseluruhan data.

$$\text{Akurasi} = \frac{TP + TN}{TP + FP + FN + TN} \quad (7)$$

Presisi, adalah rasio data yang terklasifikasi benar dari jumlah keseluruhan data yang diprediksi benar.

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (8)$$

Recall, merupakan rasio data yang terklasifikasi benar dari jumlah keseluruhan data yang benar.

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

F1-score, adalah perbandingan dari median presisi dan *recall* yang dibobotkan.

$$F1\text{-}score = 2 * \frac{(recall * presisi)}{recall + presisi} \quad (10)$$

Dalam pengujian rancangan ini memakai dua model arsitektur *Long Short Term Memory* yang tiap model memiliki konfigurasi lapisan yang berbeda. Konfigurasinya dapat dilihat di **Tabel 2**. Perbedaan dari kedua model adalah jumlah nilai parameter yang digunakan dalam menentukan bentuk nilai *output* yang dihasilkan tiap lapisan dalam model.

Tabel 2. Susunan dua model arsitektur LSTM

Model	Layer	Fungsi Aktivasi	Tipe Parameter	Nilai Parameter	*Output Shape*
Model 1	LSTM	*Sigmoid*	Dimensi *Output*	34	(None, 1, 34)
	LSTM	*Sigmoid*	Dimensi *Output*	34	(None, 34)

Tabel 2. Susunan dua model arsitektur LSTM (Lanjutan)

	Dense	-	Dimensi *Output*	64	(None, 64)
	Dense	*Softmax*	Dimensi *Output*	3	(None, 3)
Model 2	LSTM	*Sigmoid*	Dimensi *Output*	128	(None, 1, 128)
	LSTM	*Sigmoid*	Dimensi *Output*	64	(None, 64)
	Dense	-	Dimensi *Output*	64	(None, 64)
	Dense	*Softmax*	Dimensi *Output*	3	(None, 3)

Untuk konfigurasi *hyperparameter* yang digunakan dalam pengujian model dapat dilihat pada table dibawah.

Tabel 3. Susunan konfigurasi *hyperparameter*

PRESET	*OPTIMIZER*	*LEARNING RATE*	*BATCH SIZE*	*EPOCH*
1	*ADAM*	0.0001	8	100
2	*ADAM*	0.0001	16	100
3	*ADAM*	0.0001	32	100

Dengan menggunakan model-model yang disusun pada Tabel 3, pengujian dimulai dengan percobaan konfigurasi *test size* dan *train size* yang berbeda. Pada percobaan ini, *test size* yang akan dicobakan adalah 0.1, 0.15, 0.2, 0.25, 0.3, 0.33, 0.35, 0.4, 0.45, 0.5. Percobaan ini dilakukan untuk melihat pembagian data mana yang mengeluarkan hasil yang terbaik. Hasil percobaan dengan berbagai konfigurasi *test size*, pada Gambar 10.

OPTIMIZER = ADAM, LEARNING_RATE = 0.0001, BATCH_SIZE = 32, EPOCH = 100								
Test/Train	Model 1				Model 2			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
10%/90%	95%	95%	95%	95%	95%	95%	95%	95%
15%/85%	93%	93%	93%	93%	94%	94%	94%	94%
20%/80%	93%	93%	93%	93%	95%	95%	95%	95%
25%/75%	93%	93%	93%	93%	94%	94%	94%	94%
30%/70%	93%	93%	93%	93%	94%	94%	94%	94%
35%/65%	92%	92%	92%	92%	93%	93%	93%	93%
40%/60%	93%	93%	93%	93%	94%	94%	94%	94%
45%/55%	92%	92%	92%	92%	94%	94%	94%	94%
50%/50%	92%	92%	92%	92%	93%	93%	93%	93%

Gambar 10. Nilai *Accuracy*, *Presicion*, *Recall*, dan *F1-Score* dari dua model LSTM pada tiap *test size*.

Dengan mengambil data *test size* yang menghasilkan akurasi tertinggi. Pengujian model akan dilanjutkan dengan percobaan menggunakan konfigurasi hyperparameter berbeda.

PRESET	Model 1				Model 2			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
1	97%	97%	97%	97%	96%	96%	96%	96%
2	96%	96%	96%	96%	97%	97%	97%	97%
3	98%	98%	98%	98%	99%	99%	99%	99%

Gambar 11. Hasil pengujian *Accuracy*, *Presicion*, *Recall*, dan *F1-Score* dari konfigurasi *hyperparameter* yang berbeda.

Dalam pengujian terhadap data uji dilakukan perhitungan untuk mendapatkan nilai akurasi yang dihasilkan oleh model yang dibuat.

Pengujian ini menggunakan data video yang belum pernah dilatih sebelumnya tetapi memiliki aktivitas yang sama dengan durasi antara 10 detik sampai 15 detik yang di dalamnya terdapat rekaman satu orang melakukan satu kelas aktivitas.

Tabel 4. Hasil Prediksi dengan Data Uji

Video	Durasi Video	Jumlah *Frame*	Aktivitas	Hasil Prediksi(%)
1	15 Detik	450 *Frame*	*Picking*	74%
2	13 Detik	390 *Frame*	*Picking*	54%
3	10 Detik	280 *Frame*	*Standing*	82%
4	10 Detik	300 *Frame*	*Standing*	62%
5	15 Detik	400 *Frame*	*Walking*	76%
6	11 Detik	330 *Frame*	*Walking*	89%

Untuk pengujian selanjutnya dilakukan secara langsung atau *real-time* menggunakan kamera *webcam* laptop. Dalam pengujian ini terdapat satu orang melakukan satu aktivitas selama 10 detik dengan 3 fps (*frame* per detik). Jarak antara kamera dan subjek 1 sampai 2 meter dengan sudut pandang kamera setingkat dengan badan.

Tabel 5. Hasil Prediksi Pengujian Langsung atau *Realtime*

No.	Durasi	Jumlah *Frame*	Jarak	Aktivitas	Hasil Prediksi(%)
1	10 Detik	30 *Frame*	1 Meter	*Picking*	28%
2	10 Detik	30 *Frame*	2 Meter	*Picking*	73%
3	10 Detik	30 *Frame*	1 Meter	*Standing*	78%
4	10 Detik	30 *Frame*	2 Meter	*Standing*	87%
5	10 Detik	30 *Frame*	1 Meter	*Walking*	81%
6	10 Detik	30 *Frame*	2 Meter	*Walking*	84%

4. Kesimpulan

Setelah melakukan penelitian tentang pendeteksian pada supermarket, dapat ditarik beberapa kesimpulan yaitu sebagai berikut:

1. Hasil *blackbox* testing pada modul-modul yang dirancang berjalan dengan baik. Semua fungsi dan fitur pada setiap modul telah berjalan sesuai spesifikasi rancangan.
2. Program yang telah dibuat dapat mendeteksi aktivitas seseorang yang termasuk ke dalam kelas target yaitu berdiri, berjalan, dan mengambil.
3. Hasil terbaik yang didapatkan menggunakan *test size* 90%/10% dan menggunakan model LSTM dengan konfigurasi 4 layer yaitu 2 layer LSTM dan 2 *dense* layer. Layer LSTM pertama memiliki nilai parameter 128 dan menggunakan fungsi aktivasi *sigmoid*. Layer LSTM kedua memiliki nilai parameter 64 dan menggunakan fungsi aktivasi *sigmoid*. Layer *Dense* pertama memiliki nilai parameter 64 dan tidak menggunakan fungsi aktivasi. Layer *Dense* kedua memiliki nilai parameter 3 dan menggunakan fungsi aktivasi *softmax*. Menggunakan konfigurasi *hyperparameter* dengan *Optimzer Adam*, *Batch Size* 32, *Epoch* 100 dan *Learning Rate* 0.0001.
4. Pengujian dengan data latih mendapatkan nilai *accuracy* sebesar 99%, *precision* sebesar 99%, *recall* sebesar 99%, dan *F1-score* sebesar 99%.
5. Pengujian dengan data uji yang memiliki jarak 2 meter antara objek dan kamera mendapatkan nilai akurasi prediksi yang lebih tinggi sebesar 74% untuk aktivitas *picking*, 82% untuk *standing*, dan 89% untuk *walking*. Sedangkan untuk pengujian secara langsung menghasilkan nilai akurasi prediksi sebesar 73% untuk aktivitas *picking*, 87% untuk *standing*, dan 84% untuk *walking*.

Setelah melakukan penelitian, beberapa saran yang dapat diterapkan dalam penelitian berikutnya di topik serupa adalah sebagai berikut:

1. Dalam menggunakan *OpenPose* sebagai alat untuk *Human Pose Estimation* sangat disarankan untuk memakai kartu grafis yang mempunyai CUDA karena akan menghasilkan akurasi *skeleton* yang baik.
2. Mencoba model LSTM dengan berbagai macam konfigurasi lapisan untuk mendapatkan hasil latih yang lebih baik.
3. Menambahkan data latih lebih banyak agar dapat menghasilkan akurasi yang baik.
4. Membuat model BiLSTM untuk dibandingkan dengan hasil LSTM biasa.

REFERENSI

[1] Nguyen, Binh; Coelho, Yves; Bastos, Teodiano; and Krishnan, Sridhar. "Trends in human activity recognition with focus on machine learning and power requirements." Machine Learning with Applications. Vol. 5, Hal. 100072. Tahun 2021.

[2] Jain, Deepak Kumar; Mahanti, Aniket; Pourya, Shamsolmoali; and Manikandan, Ramachandran. "Deep neural learning techniques with long short-term memory for gesture recognition." Neural Computing and Applications. Vol. 32, hal. 16073–16089. Tahun 2020.

[3] Singh, Tej; and Vishwakarma, Dinesh Kumar. "Human activity recognition in video benchmarks: A survey." Advances in Signal Processing and Communication. Hal. 247-259. https://doi.org/10.1007/978-981-13-2553-3_24. Tahun 2019.

[4] Wan, Shaohua; Qi, Lianyong; Xu, Xiaolong; Tong, Chao; and Gu, Zonghua. "Deep learning models for real-time human activity recognition with smartphones." Mobile Networks and Applications. Vol. 25, No. 2, Hal. 743-755. Tahun 2020.

[5] Noori, Farzan Majeed; Wallace, Benedikte; Uddin, Md; and Torresen, Jim. "A robust human activity recognition approach using openpose, motion features, and deep recurrent neural network." Scandinavian conference on image analysis. Hal. 299-310. https://doi.org/10.1007/978-3-030-20205-7_25. Tahun 2019.

[6] Cao, Zhe; Hidalgo, Gines; Simon, Tomas; Wei, Shih-En; and Sheikh, Yaser. "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields." IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 43, No. 1, Hal. 172-186. Tahun 2019.

[7] Prijono, Benny. Pengenalan Recurrent Neural Network (RNN) Bagian 1. https://indoml.com/2018/04/04/pengenalan-rnn-bag-1/. Maret 13, 2022.

[8] Ayyadevara, V. Kishore. "Recurrent neural network." Pro Machine Learning Algorithms. Hal. 217-257. https://doi.org/10.1007/978-1-4842-3564-5_10. Tahun 2018.

[9] Chung, Hyejung; and Shin, Kyung-shik. "Genetic algorithm-optimized long short-term memory network for stock market prediction." Sustainability. Vol. 10, No. 10, Hal. 3765. Tahun 2018

Kristian Davidson Runtu, mahasiswa S1, program studi Teknik Informatika, Fakultas Teknologi Informasi Universitas Tarumanagara

Lina S.T., M.Kom., Ph.D., memperoleh gelar Sarjana dari Universitas Tarumanagara, Indonesia tahun 2001 dan gelar Magister dari Universitas Indonesia, Indonesia tahun 2004. Kemudian tahun 2009 memperoleh gelar Ph.D. dari Nagoya University, Jepang. Saat ini sebagai Dosen Tetap Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara.