****

**SKRIPSI**

**PENGEMBANGAN MODEL NLP BERBASIS EMOSI UNTUK ANALISIS SENTIMEN TEKS BAHASA INDONESIA MENGGUNAKAN ALGORITMA HYBRID CNN-LSTM**

**RANGGA WIDIASMARA**

NPM 21081010085

**DOSEN PEMBIMBING**

-

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**SURABAYA**

**2024**

*Halaman ini sengaja dikosongkan*

# LEMBAR PENGESAHAN

**PENGEMBANGAN MODEL NLP BERBASIS EMOSI UNTUK ANALISIS SENTIMEN TEKS BAHASA INDONESIA MENGGUNAKAN ALGORITMA HYBRID CNN-LSTM**

Oleh :

Rangga Widiasmara

NPM. 21081010085

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi xxxxxxx Fakulktas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal...............................................

|  |  |  |
| --- | --- | --- |
| Nama Dosen  NIP. xxxxxxxx xxxxxx x xxx | …………………… | (Pembimbing I) |
| Nama Dosen  NIP. xxxxxxxx xxxxxx x xxx | …………………… | (Pembimbing II) |
| Nama Dosen  NIP/NPT | …………………… | (Pembimbing III)  *(Opsional/Tambahan)* |
| Nama Dosen  NIP/NPT | …………………… | (Ketua Penguji) |
| Nama Dosen  NIP/NPT | …………………… | (Penguji I) |
|  | Mengetahui,  Dekan Fakultas Ilmu Komputer  Prof. Dr. Ir. Novirina Hendrasarie, MT  NIP. 19681126 199403 2 001 | |

*Halaman ini sengaja dikosongkan*

# SURAT PERNYATAAN ORISINALITAS

Yang bertandatangan di bawah ini:

Nama Mahasiswa : RANGGA WIDIASMARA

Program Studi : Informatika

Dosen Pembimbing : -

dengan ini menyatakan bahwa isi sebagian maupun keseluruhan disertasi dengan judul:

**PENGEMBANGAN MODEL NLP BERBASIS EMOSI UNTUK ANALISIS SENTIMEN TEKS BAHASA INDONESIA MENGGUNAKAN ALGORITMA HYBRID CNN-LSTM**

adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diizinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri. Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

|  |
| --- |
| Surabaya, ….………………  Yang Membuat Pernyataan,  RANGGA WIDIASMARA  NPM. 21081010085 |

*Halaman ini sengaja dikosongkan*

# ABSTRAK

|  |  |
| --- | --- |
| Nama Mahasiswa / NPM : | Rangga Widiasmara / 21081010085 |
| Judul Skripsi : | Pengembangan Model NLP Berbasis Emosi untuk Analisis Sentimen Teks Bahasa Indonesia menggunakan Algoritma Hybrid CNN-LSTM |
| Dosen Pembimbing : | - |
|  | - |

Penelitian ini bertujuan untuk membandingkan performa model Long Short Term Memory (LSTM) dan Gated Recurrent Unit (GRU) dalam memprediksi pergerakan harga Bitcoin. Metode penelitian yang digunakan melibatkan enam tahapan utama yaitu pengumpulan data dari Kaggle, pembacaan dataset, eksplorasi data sederhana, preprocessing data, pembangunan model prediksi, serta evaluasi model menggunakan data pengujian. Dataset harga Bitcoin diunduh dari Kaggle dan kemudian melalui tahap preprocessing untuk membersihkan serta menyiapkan data untuk analisis lebih lanjut. Model LSTM dan GRU dibangun menggunakan data yang telah diproses dan pertimbangan akan kesetaraan kompleksitas masing-masing model dievaluasi untuk menentukan akurasi prediksinya. Hasil penelitian ini menunjukkan perbandingan yang signifikan antara kedua model. Dan setelah dilakukan evaluasi model dan analisis perbandingan performa didapatkan model terbaik pada penelitian ini adalah model LSTM dengan arsitektur 4 layer, dua layer LSTM dengan jumlah unit (neuron) 50, dua Layer Dense dengan jumlah neuron 25 dan 1, parameter learning rate sebesar 0.001, kemudian Time steps sebanyak 20 yang dioptimasi dengan Adam optimizer. Dengan konfigurasi LSTM berikut, didapatkan hasil MAE proses training sebesar 712,11 sedangkan MAE proses pengujian sebesar 507,74. Kemudian untuk MAPE proses training sebesar 5% dan MAPE proses pengujian sebesar 2%. Dengan hasil tersebut penelitian ini menganggap model LSTM lebih baik daripada GRU untuk memprediksi harga Bitcoin.

**Kata kunci** : Bitcoin, Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Time Steps, Learning rate

*Halaman ini sengaja dikosongkan*

**ABSTRACT**

|  |  |
| --- | --- |
| Student Name / NPM : | Bambang Jati Negara / 20212300020000 |
| Thesis Title : | Pdf Document Classification For Course Conversion In Past Learning Recognition Program Using NLP Approach |
| Advisor : | 1. Dr. Ir. I Gede Susrama Mas Diyasa, ST. MT. IPU |
|  | 2. Eva Yulia Puspaningrum, S.Kom., M.Kom |

**ABSTRACT**

This research aims to compare the performance of Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) models in predicting Bitcoin price movements. The research method used involves six main stages, namely data collection from Kaggle, dataset reading, simple data exploration, data preprocessing, prediction model building, and model evaluation using test data. The Bitcoin price dataset was downloaded from Kaggle and then went through a preprocessing stage to clean and prepare the data for further analysis. LSTM and GRU models were built using the preprocessed data and consideration of the equal complexity of each model was evaluated to determine its prediction accuracy. The results of this study show a significant comparison between the two models. And after model evaluation and performance comparison analysis, the best model in this study is the LSTM model with a 4-layer architecture, two LSTM layers with the number of units (neurons) 50, two Dense layers with the number of neurons 25 and 1, a learning rate parameter of 0.001, then 20 time steps optimized with Adam optimizer. With the following LSTM configuration, the MAE result of the training process is 712.11 while the MAE of the testing process is 507.74. Then for the MAPE of the training process of 5% and the MAPE of the testing process of 2%. With these results, this study considers the LSTM model to be better than GRU for predicting the performance of the LSTM model.

**Keywords**: Bitcoin, Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Time Steps, Learning rate

*Halaman ini sengaja dikosongkan*

# KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“** **Pengembangan Model NLP Berbasis Emosi untuk Analisis Sentimen Teks Bahasa Indonesia menggunakan Algoritma Hybrid CNN-LSTM”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Bapak/Ibu - selaku Dosen Pembimbing utama yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Dan penulis juga banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Ibu/Bapak..................... selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu/Bapak ……………. selaku Ketua Program Studi xxxx Fakultas Ilmu Sosial Dan Ilmu Komputer Universitas Pembangunan Nasional “ Veteran “ Jawa Timur.
3. Dosen-dosen Program Studi … dst..

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya,\_\_\_\_\_\_\_\_\_\_\_\_

Penulis

# 

*Halaman ini sengaja dikosongkan*

# DAFTAR ISI

|  |  |  |  |
| --- | --- | --- | --- |
| **LEMBAR JUDUL SKRIPSI** .……………………………………..…….… **i** | | | |
| **LEMBAR PENGESAHAN SKRIPSI** ........................……………………. **v** | | | |
| **LEMBAR PERSETUJUAN SKRIPSI** …….....…………………………... **vii** | | | |
| **ABSTRAK** ……………...………………………………………...………... **xi** | | | |
| **KATA PENGANTAR** ……………...……………………………………... **xi** | | | |
| **DAFTAR ISI** …………………………...………………………...………... **xv** | | | |
| **DAFTAR GAMBAR** …………………….…………………………...……. **xviii** | | | |
| **DAFTAR TABEL** …..…………………………………………………..….. **xxiii** | | | |
| **DAFTAR NOTASI** ………………………………………..……………….. **xxv** | | | |
| **BAB 1 PENDAHULUAN**…………………………………………...……… **1** | | | |
|  | 1.1. | Latar Belakang……………………...……………...….................. | 1 |
|  | 1.2. | Rumusan Masalah………………………….…………………….. | 5 |
|  | 1.3. | Tujuan Penelitian………………………………………………… | 7 |
|  | 1.4. | Manfaat Penelitian…………………………………..…………… | 9 |
| **BAB 2 TINJAUN PUSTAKA**…………………………..…..……………… **17** | | | |
|  | 2.1. | Penelitian Terdahulu…...………………………………………… | 17 |
|  | 2.2. | Landasan Teori……...…………………………………………… | 21 |
|  | 2.3. | Pemrosesan Data Akusisi………………………………………… | 23 |
|  | 2.3.1 | Spermatozoa Manusia……………………………………………. | 23 |
|  | 2.3.2 | Analisis Semen Manusia…………………………………………. | 24 |
|  | 2.3.3 | Pengamatan Semen Secara Makroskospis……………………….. | 25 |
|  | 2.4. | Dst…………………………………………….………………...... | 34 |
| **BAB 3 DESAIN DAN IMPLEMNTASI SISTEM..**................................... **71** | | | |
|  | 3.1. | Metode Penelitian…………………..……………………………. | 71 |
|  | 3.2. | Desain Sistem……….……...……………………………………. | 72 |
|  | 3.3. | Pelacakan Pergerakan Kepala Spermatozoa………....................... | 74 |
|  | 3.3.1. | *Preprocessing*…………………………………………….............. | 74 |
|  | 3.4. | Dst……........................................................................................... | 92 |
| **BAB 4 PNGUJIAN DAN ANALISA** ………………………………………  **94** | | | |
|  | 4.1. | Metode Pengujian.......................................…………………......... | 94 |
|  | 4.2. | Hasil Pengujian…….……………………………………………... | 94 |
|  | 4.3. | Dst……………..…………………………………………………. | 114 |
| **BAB 5 PENUTUP** ………………………………………………………… **116** | | | |
|  | 5.1. | Kesimpulan............................................................…………….... | 116 |
|  | 5.1.1. | Saran Pengembangan....…………………………………………. | 118 |
| **DAFTAR PUSTAKA………………………………………………………. 140** | | | |
| **LAMPIRAN 1 ................................................................................................ 144** | | | |

# DAFTAR GAMBAR

|  |  |  |  |
| --- | --- | --- | --- |
|  | Gambar 1.1 | Gambaran Permasalahan Dengan Analisis Spermatozoa Manusia………………………………………………….. | 4 |
|  | Gambar 1.2 | Perangkat yang digunakan untuk mengambil citra dan video spermatozoa, di laboratorium mikrobiologi Poltekes Surabaya – 20 spermatozoa…………………… | 9 |
|  | Gambar 1.3. | Diagram Tulang Ikan Penelitian………………………… | 12 |
|  | Gambar 1.4. | Alur Penentuan Abnormalitas Bentuk dan Pergerakan Spermatozoa ..............................………………………… | 13 |
|  | Gambar 2.1. | Kerangka Konsep Untuk Klasifikasi Hasil Pemeriksaan Spermatozoa……………………………………………... | 22 |
|  | Gambar 2.2. | Struktur Morfologi Sperma……………………………… | 25 |
|  | Gambar 2.3. | *Bright field microscope*: (a) Prinsip kerja *bright field microscope*, (b) Irisan *bright field microscope*………….. | 31 |
|  | Gambar 2.4 | *Phase contrast microscope* …………...………………… | 32 |
|  | Gambar 2.5 | Perbandingan kontras image sel hidup dari dua jenis mikroskop : (a) *bright field microscope* , (b) *phase contrast microscope*……………………………………. | 32 |
|  | Gambar 2.6. | Prosedur pengambilan data citra dan video sperma, (a) *Bright field microscope* yang digunakan, (b) Cairan sperma yang sudah diteteskan di atas kaca preparat…….. | 33 |
|  | Gambar 2.7. | Pemrosesan Awal Ketidaknormalan Sperma Berdasarkan Morfologi……...………………………………………… | 34 |
|  | Gambar 2.8. | Konversi *RGB* ke *Grey scale* pada Citra Spermatozoa. (a) Citra *RGB*, (b) Citra *Grey Scale*…………………….. | 36 |
|  | Gambar 2.9. | Distribusi *Gaussian* 1D…………………………………. | 38 |
|  | Gambar 2.10. | Distribusi 2D *Gaussian*………………………………….. | 38 |
|  | Gambar 2.11. | Prosess *background subtraction*………………………… | 39 |
|  | Gambar 2.12. | Alur proses dari basic model *background subtraction*….. | 40 |
|  | Gambar 2.13. | Alur diagram dari algoritma *Frame Difference*…………. | 41 |
|  | Gambar 2.14. | Alur diagram dari algoritma *Weighted Moving Mean*…... | 42 |

*Halaman ini sengaja dikosongkan*

# DAFTAR TABEL

|  |  |  |  |
| --- | --- | --- | --- |
|  | Tabel 1.1 | Matriks Posisi Penelitian pada Penelitian Terkait………….. | 6 |
|  | Tabel 2.1. | Gambaran Makroskopik Analisis Semen (Standart WHO, 2010)………………………………..………………………. | 28 |
|  | Tabel 2.2 | Klasifikasi Morfologi Sperma (Wein et al., 2012)…………. | 29 |
|  | Tabel 2.3 | Hasil *review background subtraction* (Li, Q 2012) dan Penelitian (Basuki, 2016)…………. | 39 |
|  | Tabel 3.1. | Hasil Ekstraksi Fitur Kelas Spermatozoa (Valid) dan Bukan Spermatozoa (Tidak Valid) untuk Data *Training*…………... | 85 |
|  | Tabel 3.2. | Hasil Pengujian Klasifikasi Sperma Dengan Metode *Support Vector Machine (SVM)*…………………………... | 88 |
|  | Tabel 3.3. | Hasil Pengujian Klasifikasi Sperma Dengan Metode *K-Nearest Neighbour (K-NN)*………………………………… | 90 |
|  | Tabel 4.1. | Contoh perbandingan hasil pelacakan spermatozoa setiap algoritma *Basic background subtraction* pada *frame* ke 120 | 109 |
|  | Tabel 4.2. | Contoh perbandingan hasil pelacakan spermatozoa setiap algoritma *statistical background subtraction* pada *frame* ke 120 …………………………………………………………. | 112 |
|  | Tabel 4.3. | Hasil dari *precision, recall*, dan *f*-*measure* dari setiap algoritma *background subtraction* ………………………… | 114 |
|  | Tabel 5.1. | Identifikasi Spermatozoa (J. Elia, 2010)…………………… | 120 |
|  | Tabel 5.2. | Posisi Sperma Data Uji Selama Penjejakan………………... | 132 |
|  | Tabel 5.3. | Posisi Data Sperma Manusia Selama Penjejakan…………. | 133 |
|  | Tabel 5.4. | Regresi Linear dan Nilai *RMS* Data Sperma Uji Selama Penjejakan………………..…………………………………. | 134 |
|  | Tabel 5.5. | Regresi Linear dan Nilai *RMS* Data Sperma Manusia Selama Penjejakan………………………………………...... | 135 |
|  | Tabel 5.6. | Jumlah Dan Prosentase Dari Kelompok Spermatozoa…....... | 135 |

*Halaman ini sengaja dikosongkan*

# DAFTAR NOTASI

|  |  |  |
| --- | --- | --- |
| I | : | Intensitas |
| WR | : | *weight factor* |
| H | : | *hue* |
| S | : | *saturation* |
| V | : | *value* |
| *dst* | : | Gambar akumulator |
| *scr* | : | Gambar Input |
| *F* | : | *Foreground* |
| *B* | : | *Background* |
| *f* | : | *Frame* |
| *SE* | : | *Structuring Element* |
| *ECD* | : | *Equivalent Circular Diameter* |
| b | : | bias |
|  | : | bobot *euclidian distance* antara vektor fitur |
|  | : | Fungsi Vektor Masukan |
|  | : | jarak di antara data uji z ke setiap vector data latih |
|  | : | fungsi kernel linear |
|  | : | fungsi Gaussian satu dimensi |
|  | : | standard deviasi dari distribusi |
|  | : | fungsi Gaussian dua dimensi |

*Halaman ini sengaja dikosongkan*

# BAB I PENDAHULUAN

## Latar Belakang

Analisis sentimen telah menjadi salah satu aspek penting dalam Natural Language Processing (NLP) karena kemampuannya untuk memahami opini dan emosi yang terkandung dalam teks. Analisis ini memiliki peran strategis dalam berbagai bidang, seperti layanan pelanggan untuk menganalisis kepuasan pengguna, kesehatan mental untuk mendeteksi potensi masalah emosional, serta media sosial untuk memahami dinamika emosi masyarakat secara real-time (Singgalen, 2024; Madan & Kumar, 2024). Selain itu, dalam konteks pendidikan, analisis emosi dapat digunakan untuk meningkatkan kualitas pembelajaran dengan memahami kebutuhan emosional siswa (Sorour et al., 2024).

Dalam pengembangannya, analisis sentimen telah berkembang dari sekadar klasifikasi polaritas (positif, negatif, netral) menjadi analisis emosi mendalam, seperti deteksi emosi marah, senang, atau sedih (Singh et al., 2024). Namun, pengembangan model NLP berbasis emosi dalam Bahasa Indonesia masih menghadapi sejumlah tantangan, termasuk keterbatasan dataset yang representatif, kompleksitas struktur bahasa, dan kurangnya model yang dirancang khusus untuk Bahasa Indonesia (Phan et al., 2024).

Salah satu pendekatan yang menjanjikan dalam analisis sentimen emosi adalah penggunaan algoritma hybrid CNN-LSTM. Algoritma ini memadukan kekuatan Convolutional Neural Networks (CNN) untuk menangkap fitur lokal dalam teks dan Long Short-Term Memory (LSTM) untuk memahami hubungan temporal dan konteks kata. Pendekatan ini telah terbukti efektif untuk meningkatkan akurasi dalam berbagai aplikasi analisis sentimen (Tran et al., 2024; Parra-Gallego & Arias-Vergara, 2025). Namun, penerapan algoritma hybrid ini untuk teks dalam Bahasa Indonesia masih memerlukan eksplorasi lebih lanjut untuk memastikan efektivitasnya (Das & Pradhan, 2024; Kukreja & Kundra, 2024).

Penelitian ini bertujuan untuk mengembangkan model NLP berbasis emosi untuk analisis sentimen teks Bahasa Indonesia menggunakan algoritma hybrid CNN-LSTM. Melalui penelitian ini, diharapkan dapat dihasilkan model yang tidak hanya mampu meningkatkan akurasi analisis sentimen emosi, tetapi juga memberikan kontribusi pada pengembangan model NLP yang lebih baik untuk Bahasa Indonesia. Selain itu, hasil penelitian ini juga diharapkan dapat menjadi acuan dalam pengembangan aplikasi berbasis NLP untuk berbagai bidang seperti pendidikan, layanan pelanggan, dan kesehatan mental (Chowdhury & Huq, 2024).

## Rumusan Masalah

Analisis sentimen merupakan salah satu cabang penting dalam Natural Language Processing (NLP) yang mampu mengungkap opini dan emosi dari teks. Dalam beberapa tahun terakhir, analisis ini telah berkembang dari sekadar pengklasifikasian polaritas (positif, negatif, netral) menjadi analisis emosi mendalam yang mencakup berbagai kategori emosi seperti marah, sedih, dan bahagia. Namun, pengembangan model NLP berbasis emosi untuk Bahasa Indonesia menghadapi berbagai kendala, seperti keterbatasan dataset representatif yang sesuai dengan struktur unik bahasa ini, kompleksitas pola linguistik, dan minimnya model yang dirancang khusus untuk Bahasa Indonesia. Hal ini menjadikan penerapan analisis sentimen berbasis emosi dalam Bahasa Indonesia belum optimal jika dibandingkan dengan bahasa global seperti Inggris.

Di sisi lain, algoritma hybrid CNN-LSTM menawarkan solusi yang menjanjikan dengan memadukan kemampuan Convolutional Neural Networks (CNN) untuk menangkap pola lokal dalam teks dan Long Short-Term Memory (LSTM) untuk memahami hubungan temporal antar kata. Pendekatan ini telah terbukti meningkatkan akurasi analisis sentimen dalam berbagai bahasa dan domain. Namun, penerapannya untuk teks Bahasa Indonesia masih kurang tereksplorasi. Pertanyaan yang muncul adalah bagaimana pendekatan ini dapat diadaptasi secara efektif untuk Bahasa Indonesia agar dapat mengatasi keterbatasan yang ada.

Berdasarkan permasalahan tersebut, diperlukan penelitian yang mampu menjawab tantangan ini dengan mengembangkan model analisis sentimen berbasis emosi menggunakan algoritma hybrid CNN-LSTM. Penelitian ini juga perlu memastikan bahwa model yang dihasilkan mampu memberikan kinerja yang lebih akurat, relevan, dan aplikatif di berbagai bidang seperti layanan pelanggan, pendidikan, dan kesehatan mental.

## Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan model analisis sentimen berbasis emosi untuk teks dalam Bahasa Indonesia menggunakan algoritma hybrid CNN-LSTM. Tujuan utama dari penelitian ini adalah untuk meningkatkan akurasi analisis sentimen emosi dengan memanfaatkan kemampuan CNN dalam mengenali pola-pola lokal dalam teks dan LSTM dalam memahami hubungan temporal dan konteks antar kata.

Secara lebih spesifik, penelitian ini memiliki beberapa tujuan sebagai berikut:

1. Mengeksplorasi kemampuan algoritma hybrid CNN-LSTM dalam mengatasi tantangan unik struktur bahasa Indonesia, seperti kekayaan morfologi, sintaksis, dan semantik.
2. Mengidentifikasi dan memanfaatkan dataset yang representatif untuk Bahasa Indonesia, sekaligus mengatasi keterbatasan data dengan teknik augmentasi atau metode lain yang relevan.
3. Menguji dan mengevaluasi performa model yang diusulkan pada berbagai kategori emosi, sehingga model ini dapat digunakan untuk berbagai aplikasi, mulai dari layanan pelanggan hingga analisis sosial dan pendidikan.
4. Memberikan kontribusi pada pengembangan teknologi pemrosesan bahasa alami (NLP) yang lebih adaptif terhadap konteks dan kebutuhan lokal.

Melalui pencapaian tujuan-tujuan tersebut, penelitian ini diharapkan mampu menjadi solusi yang efektif dalam mendorong inovasi NLP untuk Bahasa Indonesia.

## Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat signifikan baik dari segi teoretis maupun praktis, yang dapat dirinci sebagai berikut:

1. **Manfaat Teoretis:**

Penelitian ini berkontribusi pada literatur ilmiah mengenai analisis sentimen berbasis emosi, terutama dalam pengembangan model NLP yang dirancang khusus untuk Bahasa Indonesia. Temuan dari penelitian ini akan memperkaya wawasan mengenai efektivitas pendekatan hybrid CNN-LSTM dalam menangani bahasa dengan struktur kompleks seperti Bahasa Indonesia. Model yang dihasilkan dapat menjadi referensi untuk penelitian lanjutan di bidang NLP, khususnya dalam pengembangan metode analisis sentimen berbasis emosi yang lebih presisi dan aplikatif.

1. **Manfaat Praktis:**

Penelitian ini juga memiliki manfaat praktis yang luas dalam berbagai bidang:

* 1. **Layanan Pelanggan:** Model yang diusulkan dapat digunakan untuk menganalisis kepuasan pelanggan secara otomatis, sehingga perusahaan dapat mengambil keputusan strategis berdasarkan opini pelanggan.
  2. **Pendidikan:** Dalam dunia pendidikan, model ini dapat membantu pengajar memahami kondisi emosional siswa melalui analisis respon atau umpan balik, sehingga proses pembelajaran dapat lebih personal dan efektif.
  3. **Kesehatan Mental:** Dalam bidang kesehatan mental, model ini dapat diterapkan untuk mendeteksi emosi negatif seperti stres atau depresi dari teks yang dihasilkan oleh pengguna, misalnya di media sosial atau survei.
  4. **Media Sosial dan Analisis Sosial:** Model ini dapat digunakan untuk memahami dinamika emosi masyarakat secara real-time, yang bermanfaat dalam penelitian sosial, pengelolaan krisis, atau bahkan pengembangan kebijakan publik.

1. **Manfaat Lokal:**

Dengan berfokus pada Bahasa Indonesia, penelitian ini akan memberikan dampak langsung pada pengembangan teknologi yang lebih inklusif, relevan, dan sesuai dengan kebutuhan masyarakat Indonesia. Hal ini diharapkan dapat mempercepat adopsi teknologi NLP dalam berbagai sektor di Indonesia.

Secara keseluruhan, penelitian ini tidak hanya diharapkan meningkatkan akurasi analisis sentimen berbasis emosi, tetapi juga menjadi pijakan penting dalam pengembangan aplikasi berbasis NLP yang mendukung berbagai inovasi di Indonesia.

**Daftar Referensi**

1. Singgalen, Y. A. (2024). *A Hybrid CNN-LSTM Model with SMOTE for Enhanced Sentiment Analysis of Hotel Reviews*. ResearchGate. [Link PDF](https://www.researchgate.net/publication/386378606_A_Hybrid_CNN-LSTM_Model_with_SMOTE_for_Enhanced_Sentiment_Analysis_of_Hotel_Reviews)
2. Madan, A., & Kumar, D. (2024). *Real-time topic-based sentiment analysis for movie tweets using hybrid approach*. Knowledge and Information Systems, Springer. [DOI Link](https://link.springer.com/article/10.1007/s10115-024-02298-x)
3. Sorour, S. E., Alojail, A., El-Shora, A., & Amin, A. E. (2024). *A Hybrid Deep Learning Approach for Enhanced Sentiment Classification and Consistency Analysis in Customer Reviews*. ProQuest. [Full Text](https://search.proquest.com/openview/e894b27e123a40e845721ef3787e1eef/1)
4. Singh, A. K., Bhushan, A., & Dwivedi, D. (2024). *Analyzing Sentiments on Twitter Using Deep Learning Techniques*. International Journal of Modern Education and Computer Science. [PDF Full Text](https://www.mecs-press.org/ijmecs/ijmecs-v16-n6/IJMECS-V16-N6-2.pdf)
5. Phan, H. T., Seo, Y. S., & Nguyen, N. T. (2024). *Fuzzy Hybrid CNN-LSTM Model for Sentence-Level Sentiment Analysis*. SSRN. [PDF Full Text](https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=4994871)
6. Tran, X. T., Nguyen, D. M., & Nguyen, M. T. (2024). *A Comprehensive Framework for Sentiment Analysis and Cold-Start Recommendations in Vietnam Hospitality Sector*. Springer. [DOI Link](https://link.springer.com/chapter/10.1007/978-981-96-0178-3_19)
7. Parra-Gallego, L. F., & Arias-Vergara, T. (2025). *Multimodal Evaluation of Customer Satisfaction from Voicemails Using Speech and Language Representations*. Digital Signal Processing, Elsevier. [DOI Link](https://www.sciencedirect.com/science/article/pii/S1051200424004457)
8. Das, S., & Pradhan, B. (2024). *Advanced Deep Learning Models for Automatic Detection of Driver's Facial Expressions, Movements, and Alertness in Varied Lighting Conditions: A Comparative Analysis*. Springer. [DOI Link](https://link.springer.com/article/10.1007/s11042-024-20428-z)
9. Kukreja, V., & Kundra, D. (2024). *Deep Learning for Sentiment Analysis: Combining CNN and LSTM for Enhanced Performance on Facebook Data*. IEEE Xplore. [DOI Link](https://ieeexplore.ieee.org/abstract/document/10760800)
10. Chowdhury, T. A., & Huq, M. R. (2024). *An Optimized Bangla Speech Emotion Recognition System Leveraging CNN-LSTM and Boosting Classifiers*. IEEE Xplore. [DOI Link](https://ieeexplore.ieee.org/abstract/document/10756428)

*Halaman ini sengaja dikosongkan*

# BAB II TINJAUN PUSTAKA

**2.1. Penelitian Terdahulu**

Beberapa penelitian telah dilakukan terkait analisis sentimen berbasis emosi menggunakan pendekatan hybrid CNN-LSTM, termasuk untuk Bahasa Indonesia. Berikut ini adalah tinjauan terhadap beberapa penelitian terdahulu yang relevan:

1. **Singgalen, Y. A. (2024)**: Dalam penelitian ini, dikembangkan model hybrid CNN-LSTM yang diperkuat dengan teknik SMOTE untuk analisis sentimen ulasan hotel. Hasilnya menunjukkan peningkatan akurasi dalam menangani data tidak seimbang, meskipun penelitian ini lebih terfokus pada bahasa Inggris. [Tautan PDF](https://www.researchgate.net/profile/Yerik-Singgalen/publication/386378606_A_Hybrid_CNN-LSTM_Model_with_SMOTE_for_Enhanced_Sentiment_Analysis_of_Hotel_Reviews/links/674f3c04f309a268c022b26d/A-Hybrid-CNN-LSTM-Model-with-SMOTE-for-Enhanced-Sentiment-Analysis-of-Hotel-Reviews.pdf)
2. **Ma'aly, A. N., et al. (2024)**: Penelitian ini mengeksplorasi analisis sentimen untuk Pemilu Presiden Indonesia 2024 menggunakan multi-label deep learning, termasuk model hybrid CNN-BiLSTM. Hasilnya menunjukkan efektivitas model hybrid dalam klasifikasi emosi di media sosial. [Tautan HTML](https://www.mdpi.com/2078-2489/15/11/705)
3. **Rozie, A. F., & Arisal, A. (2021)**: Penelitian ini memanfaatkan pendekatan hybrid neural network (CNN+LSTM+MLP) untuk klasifikasi sentimen pesan pendek dalam Bahasa Indonesia. Penelitian ini menyoroti keunggulan arsitektur hybrid dalam menangkap pola linguistik kompleks. [Tautan PDF](https://www.beei.org/index.php/EEI/article/download/2790/2280)
4. **Riyadi, S., et al. (2024)**: Penelitian ini menggunakan model CNN-BiLSTM hybrid untuk deteksi ujaran kebencian di media sosial Twitter dalam Bahasa Indonesia. Pendekatan ini meningkatkan akurasi klasifikasi dengan menangkap konteks temporal. [Tautan PDF](https://ieeexplore.ieee.org/iel8/6287639/6514899/10737089.pdf)
5. **Nuha, U., & Lin, C. H. (2023)**: Penelitian ini memperkenalkan model hybrid deep learning untuk analisis sentimen teks Bahasa Indonesia, menggunakan kombinasi CNN dan LSTM. Hasil menunjukkan bahwa model ini memberikan akurasi tinggi pada dataset emosi tidak seimbang. [Tautan PDF](https://link.springer.com/content/pdf/10.1186/s40537-023-00782-9.pdf" \t "_new)
6. **Murfi, H., et al. (2021)**: Studi ini membandingkan berbagai arsitektur hybrid deep learning untuk analisis sentimen Bahasa Indonesia, seperti CNN-LSTM dan LSTM-CNN. Penelitian ini menggarisbawahi pentingnya struktur hybrid untuk menangkap hubungan semantik dalam teks. [Tautan PDF](http://repository.uin-malang.ac.id/11888/7/11888.pdf)
7. **Putrada, A. G., et al. (2022)**: Penelitian ini memanfaatkan model hybrid CNN-LSTM dengan embedding kata-emoji untuk analisis sentimen kebijakan PPKM Indonesia di Twitter. Model ini menunjukkan keunggulan dalam memahami nuansa emosi yang kompleks. [Tautan PDF](https://www.researchgate.net/profile/Aji-Gautama-Putrada/publication/369155575_A_Hybrid_CNN-LSTM_Model_With_Word-Emoji_Embedding_For_Improving_The_Twitter_Sentiment_Analysis_on_Indonesia's_PPKM_Policy/links/6450ba475762c95ac3678b6f/A-Hybrid-CNN-LSTM-Model-With-Word-Emoji-Embedding-For-Improving-The-Twitter-Sentiment-Analysis-on-Indonesias-PPKM-Policy.pdf)

**2.2. Landasan Teori**

**2.2.1 Analisis Sentimen dan Emosi**

Analisis sentimen merupakan salah satu bidang utama dalam pemrosesan bahasa alami (NLP) yang bertujuan untuk memahami opini, emosi, dan sentimen dalam teks. Analisis ini telah berkembang dari sekadar mengidentifikasi polaritas (positif, negatif, atau netral) menjadi analisis emosi yang lebih mendalam, seperti mendeteksi emosi spesifik seperti bahagia, marah, atau sedih. Analisis ini penting dalam berbagai bidang, termasuk pengelolaan hubungan pelanggan, penelitian sosial, hingga aplikasi kesehatan mental.

Dalam konteks emosi, analisis sentimen bertujuan untuk menangkap nuansa perasaan yang diungkapkan dalam teks. Pendekatan ini tidak hanya bermanfaat bagi perusahaan untuk memahami pelanggan mereka, tetapi juga memberikan wawasan tentang dinamika emosi masyarakat. Gogula et al. (2023) menjelaskan pentingnya model yang mampu menangkap emosi dalam ulasan buku atau media sosial untuk meningkatkan pengalaman pengguna. Dengan memanfaatkan pendekatan CNN-LSTM, penelitian ini berhasil meningkatkan akurasi prediksi emosi melalui kombinasi fitur lokal dan temporal.  
*Referensi: Gogula, S. D., et al. (2023). An emotion-based rating system for books using sentiment analysis and machine learning in the cloud.* [*Unduh PDF*](https://www.mdpi.com/2076-3417/13/2/773/pdf)*.*

**2.2.2 Hybrid CNN-LSTM**

Model hybrid CNN-LSTM telah menjadi salah satu arsitektur yang banyak digunakan dalam NLP untuk analisis sentimen berbasis emosi. Model ini menggabungkan dua kekuatan utama dari Convolutional Neural Networks (CNN) dan Long Short-Term Memory (LSTM). CNN memiliki keunggulan dalam menangkap pola lokal dalam data teks, seperti n-gram atau kombinasi kata yang sering muncul bersama. Di sisi lain, LSTM dirancang untuk memahami hubungan temporal dan mempertahankan konteks dalam data urutan panjang.

Penelitian oleh Kim et al. (2023) menunjukkan bahwa kombinasi CNN-LSTM memberikan keakuratan lebih tinggi dibandingkan model yang hanya menggunakan CNN atau LSTM saja. Model ini mampu menangani data teks yang panjang dan kompleks, termasuk yang berasal dari bahasa dengan struktur sintaksis yang berbeda seperti Bahasa Indonesia. Dalam penelitian ini, arsitektur CNN digunakan untuk mengekstraksi fitur lokal dari teks, sementara LSTM bertugas memahami hubungan antar-konteks temporal. Hasilnya, model hybrid ini berhasil mengklasifikasikan emosi dengan lebih akurat.  
*Referensi: Kim, D. H., et al. (2023). A Hybrid Deep Learning Emotion Classification System Using Multimodal Data.* [*Unduh PDF*](https://www.mdpi.com/1424-8220/23/23/9333/pdf)*.*

**2.2.3 Implementasi pada Analisis Sentimen Berbasis Emosi**

Implementasi CNN-LSTM dalam analisis emosi memiliki beberapa langkah utama, mulai dari preprocessing data hingga pelatihan model. Langkah preprocessing mencakup tokenisasi, pembersihan teks dari karakter yang tidak relevan, hingga stemming. Proses ini penting untuk memastikan bahwa teks yang dimasukkan ke dalam model dapat dipahami dengan baik. Dalam penelitian oleh Chowdhury et al. (2021), preprocessing dilakukan untuk mengurangi kebisingan dalam data teks dan meningkatkan akurasi klasifikasi.

Setelah preprocessing, data akan melalui lapisan CNN untuk ekstraksi fitur lokal, yang kemudian diteruskan ke lapisan LSTM untuk analisis hubungan temporal antar kata. Pendekatan ini terbukti efektif dalam analisis dataset media sosial untuk mendeteksi emosi. Penelitian ini menunjukkan bahwa kombinasi model dapat menghasilkan prediksi yang lebih akurat dibandingkan model individual.  
*Referensi: Chowdhury, I. A., et al. (2021). A hybrid based model on LSTM-CNN to multi-class emotion analysis.* [*Unduh PDF*](https://dspace.bracu.ac.bd:8443/xmlui/bitstream/handle/10361/24786/16101302,%2016101226,%2016101006,%2016301041,%2016101292_CSE.pdf?sequence=1)*.*

**2.2.4 Evaluasi Model Hybrid CNN-LSTM**

Evaluasi model hybrid CNN-LSTM dilakukan menggunakan berbagai metrik, seperti akurasi, presisi, recall, dan F1-score. Metrik ini digunakan untuk menilai kemampuan model dalam mengklasifikasikan emosi dengan benar. Penelitian oleh Huang et al. (2021) mengintegrasikan mekanisme perhatian (attention mechanism) ke dalam model CNN-LSTM untuk meningkatkan akurasi prediksi. Dengan menambahkan lapisan perhatian, model dapat lebih fokus pada kata-kata penting yang menentukan emosi dalam teks.

Hasil evaluasi menunjukkan bahwa model dengan perhatian tambahan secara konsisten mengungguli model tradisional dalam analisis sentimen berbasis emosi. Model ini sangat berguna dalam aplikasi dunia nyata, seperti analisis data ulasan pelanggan atau pengelolaan hubungan sosial di media.  
*Referensi: Huang, F., et al. (2021). Attention-emotion-enhanced convolutional LSTM for sentiment analysis.* [*Unduh PDF*](https://drive.google.com/file/d/1sLtqFo3Tbu4jyOBjFBtcL40ZAx8B1kcF/view)*.*

Dengan landasan teori ini, penelitian difokuskan untuk mengatasi keterbatasan dataset dan meningkatkan performa analisis sentimen berbasis emosi menggunakan algoritma hybrid CNN-LSTM yang dioptimalkan untuk Bahasa Indonesia.

# BAB III DESAIN DAN IMPLEMENTASI SISTEM

* 1. **Metode Penelitian**

Penelitian ini menggunakan pendekatan eksperimen untuk mengembangkan model analisis sentimen berbasis emosi menggunakan algoritma hybrid CNN-LSTM. Metode penelitian terdiri dari beberapa tahap, yaitu pengumpulan data, preprocessing data, pengembangan arsitektur model, pelatihan, evaluasi, dan analisis hasil.

1. **Pengumpulan Data**

Dataset yang digunakan adalah data dari Twitter yang mencakup sekitar 4.400 entri. Dataset ini berisi teks pendek dengan label emosi yang telah dikategorikan menjadi lima kelas emosi utama, yaitu:

* + **Happy**
  + **Love**
  + **Sadness**
  + **Anger**
  + **Fear**

Data akan dibagi menjadi tiga subset:

* + **Data Training:** 70% dari total dataset untuk melatih model.
  + **Data Validasi:** 15% dari dataset untuk memvalidasi performa selama pelatihan.
  + **Data Uji:** 15% dari dataset untuk mengevaluasi performa akhir model.

1. **Preprocessing Data**

Proses preprocessing dilakukan untuk memastikan data dalam format yang dapat diproses oleh model. Langkah-langkahnya meliputi:

* + **Tokenisasi:** Memecah teks menjadi kata-kata individu atau token.
  + **Stopword Removal:** Menghapus kata-kata yang tidak relevan seperti "di", "ke", "dan".
  + **Stemming:** Mengembalikan kata ke bentuk dasar.
  + **Padding:** Menyelaraskan panjang teks agar konsisten.
  + **Encoding:** Mengonversi teks menjadi representasi numerik berbasis indeks.

1. **Pengembangan Model**

Model yang dirancang adalah arsitektur hybrid CNN-LSTM. CNN digunakan untuk menangkap pola lokal dalam teks, sedangkan LSTM digunakan untuk memahami hubungan temporal antar kata. Desain model mencakup tahap berikut:

* + Penambahan lapisan embedding untuk representasi vektor kata.
  + Lapisan CNN untuk mengekstraksi fitur lokal.
  + Lapisan LSTM bertingkat untuk menangkap konteks temporal dalam teks.

1. **Pelatihan dan Evaluasi**

Model dilatih menggunakan algoritma optimasi Adam dengan learning rate 0.005. Evaluasi dilakukan menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Data validasi digunakan untuk memastikan bahwa model tidak overfitting.

* 1. **Desain Sistem**

Desain sistem ini memanfaatkan kombinasi Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM) untuk menghasilkan model analisis sentimen berbasis emosi yang mampu menangkap pola lokal (lokalitas) dan hubungan temporal dalam teks. Berikut adalah deskripsi terperinci setiap komponen dalam desain sistem.

**3.2.1 Input Data**

Input data untuk sistem ini berasal dari dataset Twitter yang mencakup 4.400 entri teks dengan label emosi. Setiap entri telah dikategorikan menjadi salah satu dari lima kelas emosi: **happy**, **love**, **sadness**, **anger**, dan **fear**. Teks mentah diproses melalui tahap-tahap preprocessing, termasuk tokenisasi, penghapusan stopword, dan stemming, untuk memastikan teks siap diolah oleh model. Panjang teks diseragamkan menggunakan padding agar semua data memiliki dimensi input yang sama.

**3.2.2 Lapisan Embedding**

Lapisan pertama dalam model ini adalah embedding. Lapisan ini digunakan untuk mengonversi kata-kata dalam teks menjadi representasi vektor berdimensi tetap. Representasi ini didasarkan pada embedding pretrained (seperti GloVe atau FastText) untuk memanfaatkan pengetahuan semantik dan sintaksis yang telah dilatih sebelumnya.

Setiap kata dalam teks direpresentasikan sebagai vektor berdimensi 200. Embedding matrix yang digunakan tidak dapat di-train ulang (non-trainable) agar dapat memanfaatkan keunggulan representasi pretrained secara optimal. Output dari lapisan ini adalah tensor dengan dimensi (batch size, sequence length, embedding dimension)\text{(batch size, sequence length, embedding dimension)}(batch size, sequence length, embedding dimension).

**3.2.3 Lapisan Convolutional Neural Network (CNN)**

Setelah lapisan embedding, data diteruskan ke lapisan CNN untuk mengekstraksi fitur lokal dari teks. CNN dirancang untuk mendeteksi pola-pola penting dalam teks, seperti n-gram (kombinasi kata yang sering muncul bersama). CNN menggunakan beberapa ukuran filter untuk menangkap berbagai tingkat granularitas informasi.

* **Konfigurasi CNN:**
  + Jumlah filter: 128 untuk setiap ukuran filter.
  + Ukuran filter: 3, 4, dan 5 (menggambarkan panjang kombinasi kata yang berbeda).
  + Fungsi aktivasi: ReLU (Rectified Linear Unit) untuk menghasilkan non-linearitas.
  + Global Max Pooling: Setelah fitur dihasilkan oleh CNN, pooling dilakukan untuk mereduksi dimensi fitur menjadi ukuran tetap sambil mempertahankan informasi penting dari setiap filter.

Output dari lapisan ini adalah representasi fitur lokal teks yang akan diteruskan ke lapisan berikutnya.

**3.2.4 Lapisan Long Short-Term Memory (LSTM)**

Setelah fitur lokal diekstraksi oleh CNN, data diteruskan ke lapisan LSTM untuk memahami hubungan temporal antar kata dalam teks. LSTM digunakan karena kemampuannya untuk mempertahankan konteks kata dalam urutan panjang.

Model ini menggunakan arsitektur bidirectional LSTM untuk menangkap konteks dari kedua arah teks (maju dan mundur). Dengan demikian, model dapat lebih akurat memahami konteks emosi yang terkandung dalam teks.

* **Konfigurasi LSTM:**
  + **Layer 1:** 256 unit dengan dropout 0.2 dan recurrent dropout 0.2. Lapisan ini dirancang untuk menangkap pola temporal tingkat tinggi.
  + **Layer 2:** 128 unit dengan dropout 0.1 dan recurrent dropout 0.2. Lapisan ini mendeteksi pola yang lebih spesifik dari data.
  + **Layer 3:** 128 unit dengan dropout 0.2 dan recurrent dropout 0.2. Lapisan ini menyelesaikan proses ekstraksi pola temporal dan menghasilkan output akhir untuk lapisan dense.

Output dari lapisan LSTM adalah tensor berdimensi tetap yang merepresentasikan hubungan temporal antar kata dalam teks.

**3.2.5 Lapisan Dense dan Softmax**

Output dari lapisan LSTM diteruskan ke lapisan dense (fully connected layer) dengan aktivasi softmax. Lapisan ini bertugas menghasilkan probabilitas untuk setiap kelas emosi, sesuai dengan jumlah kategori dalam dataset (5 emosi: happy, love, sadness, anger, fear).

* **Konfigurasi Lapisan Dense:**
  + Jumlah neuron: 5 (sesuai dengan jumlah kelas emosi).
  + Fungsi aktivasi: Softmax, untuk memastikan output adalah probabilitas yang dijumlahkan hingga 1.

Probabilitas yang dihasilkan akan digunakan untuk menentukan kelas emosi akhir dari teks input.

**3.2.6 Optimasi dan Kompilasi**

Model dikompilasi menggunakan fungsi loss categorical crossentropy karena ini adalah tugas klasifikasi multikelas. Algoritma optimasi Adam digunakan untuk mempercepat konvergensi, dengan learning rate 0.005 yang dipilih berdasarkan eksperimen awal.

**3.2.7 Evaluasi Model**

Evaluasi model dilakukan dengan menggunakan subset data validasi dan uji. Metrik yang digunakan mencakup:

* **Akurasi:** Untuk mengukur persentase prediksi benar.
* **Presisi:** Untuk mengevaluasi keakuratan prediksi positif.
* **Recall:** Untuk mengukur sensitivitas terhadap kelas yang benar.
* **F1-Score:** Kombinasi harmonis antara presisi dan recall.

Evaluasi dilakukan pada data uji untuk memastikan bahwa model tidak overfitting pada data pelatihan.

**3.2.8 Arsitektur Sistem**

Berikut adalah deskripsi arsitektur sistem secara keseluruhan:

1. **Input Data:** Dataset Twitter yang telah diproses (preprocessed).
2. **Embedding Layer:** Mengonversi kata menjadi vektor.
3. **CNN Layer:** Mengekstraksi pola lokal dengan beberapa ukuran filter.
4. **Global Max Pooling:** Mengurangi dimensi fitur tanpa kehilangan informasi penting.
5. **Bidirectional LSTM:** Menganalisis hubungan temporal dalam teks.
6. **Dense Layer dengan Softmax:** Menghasilkan prediksi probabilitas untuk lima kelas emosi.

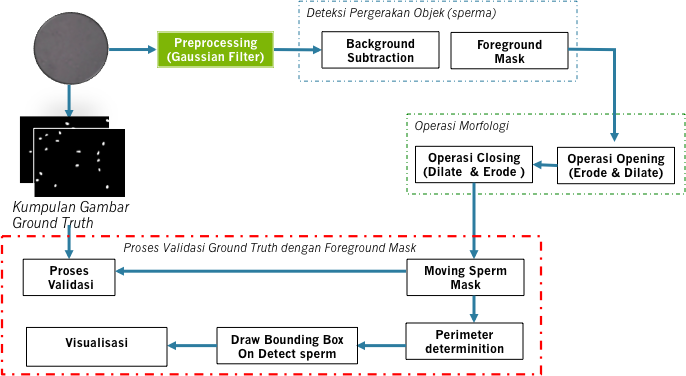
# BAB IV PENGUJIAN DAN ANALISA

**4.1. Metode Pengujian**

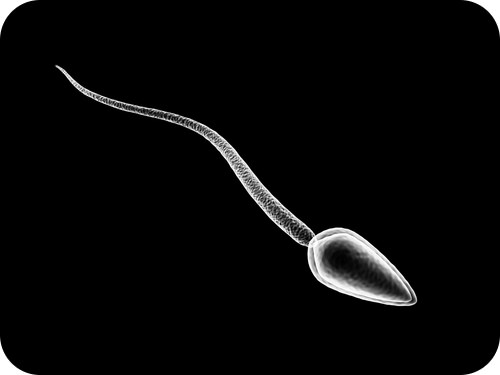
Ada 4 prosess yang dilakukan untuk dapat mendeteksi dan menghitung sperma manusia (Blok diagram pada Gambar 4.1). Prosess yang pertama adalah *preprocessing* menggunakan *gaussian filter*, yang dilakukan pada setiap *frame* yang dibaca dari video sperma. Selanjutnya adalah prosess *background subtraction*, hasil dari proses ini adalah *binary image* yang merepresentasikan daerah objek yang bergerak pada *frame*. Dilanjutkan dengan operasi morfologi yang terdiri dari operasi *opening* dan operasi *closing*, yang bertujuan untuk menghilangkan *noise* dan menyempurnakan bentuk sperma (bergerak) yang berhasil terekstraksi dari pada prosess sebelumnya. Untuk menguji atau memvalidasi hasil deteksi dari setiap algoritma *background subtraction* yang digunakan, hasil *foreground mask* dari operasi morfologi akan dibandingkan dengan gambar *ground truth* sperma yang bergerak hasil dari pengamatan secara manual. Dan untuk keperluan visualisasi, setiap daerah BLOB (objek putih pada *binary image*) pada *foreground mask* akan diberi *bounding box* pada *frame* aslinya dan sekaligus akan dihitung jumlah dari objek BLOB yang ada pada *foreground mask*, sehingga terlihat bahwa sistem telah mampu untuk melakukan deteksi dan perhitungan terhadap sperma yang bergerak. Pembahasan lebih rinci dari setiap prosess akan dibahas pada sub-bab selanjutnya

### 4.2. Hasil Pengujian

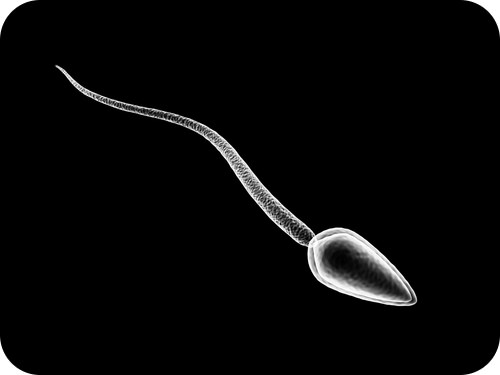
Gambar *ground truth* dalam penelitian ini merupakan gambar yang berisi informasi tentang daerah sesungguhnya dari objek sperma yang bergerak pada *frame* tertentu dari video data sperma. Gambar *ground truth* diperoleh dengan cara mengamati secara manual daerah-daerah pada *frame* video yang terdapat objek sperma yang bergerak. Cara dalam pembentukan gambar *ground truth*, dapat dilihat pada Gambar 4.2.



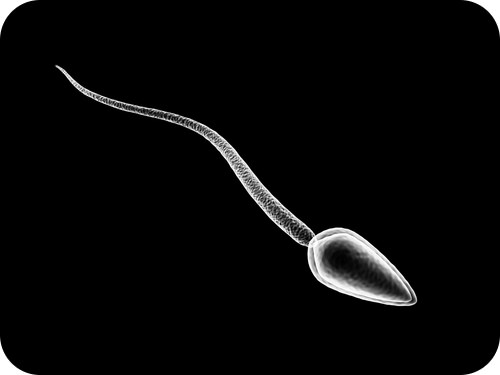
**Gambar 4.1. Blok Diagram Prosess Deteksi dan Menghitung spermatozoa**



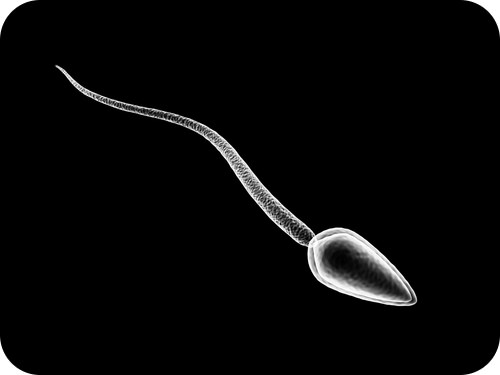
S1b



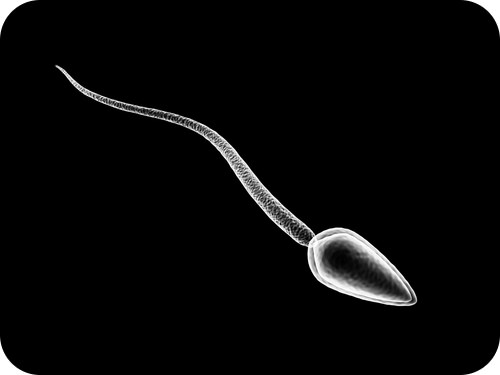
S1a



S3b



S3a



S2a

**Frame Video**

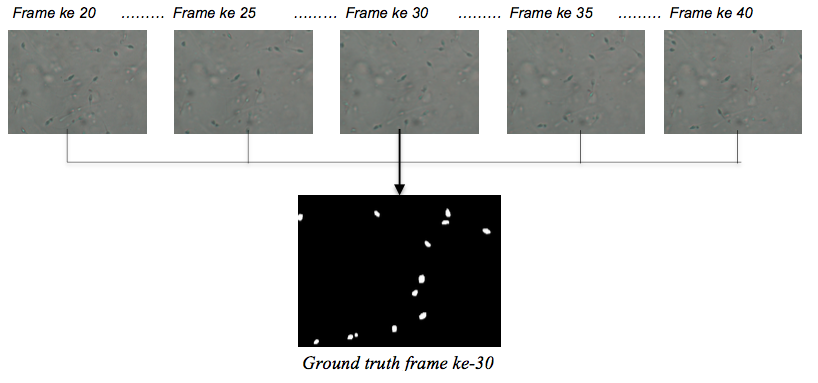
**Ground Truth**

Segmentasi Manual

###### 

###### Gambar 4.2. Cara Untuk Membuat *Ground Truth Image*

Untuk memastikan daerah tersebut memang terdapat objek sperma yang bergerak, maka diamati 10 *frame* sebelum dan 10 *frame* sesudah dari *frame* yang akan dibuat gambar *ground truth*-nya. Misal: gambar *ground truth* frame ke-30 dibuat dengan cara mengamati pergerakan sperma mulai dari *frame* ke-20 sampai dengan *frame* ke-40 dari video. Daerah dimana terdapat objek sperma yang bergerak, ditandai dengan memberikan nilai piksel 255 (putih) dan daerah yang tidak memiliki objek sperma yang bergerak ditandai dengan memberikan nilai piksel 0 (hitam). Dengan cara ini maka terbentuk gambar *ground truth* yang akan menjadi acuan pada proses pengujian dari hasil deteksi dan perhitungan sperma. Ilustrasi dari pembuatan gambar *ground truth* pada *frame* 30 dapat dilihat pada Gambar 4.3.



###### Gambar 4.3. Ilustrasi Pembuatan *Ground Truth Image*

*Halaman ini sengaja dikosongkan*

**BAB V**

**PENUTUP**

**5.1. Kesimpulan**

Dari keseluruhan pengujian sistem, hasil dari penelitian ini dapat disimpulkan antara lain sebagai berikut:

1. Berdasarkan pengujian, identifikasi dan penghitungan spermatozoa bisa dilakukan secara automatis. Dalam penelitian ini, Blob Analysis mendeteksi ROI (Region of Interest) berdasarkan nilai treshold citra dari H dan S. Selanjutnya dilakukan klasifikasi dengan *Support Vector Machine (SVM)* dan *K-NN*, menggunakan fitur *Area, Eccentricity dan ECD*. Berdasarkan pengujian SVM didapatkan akurasi yang signifikan mencapai tingkat akurasi 91,92%. Selanjutnya dilakukan klasifikasi dengan metode K-NN yang digunakan sebagai metode pembanding dengan hasil akurasi mencapai 91%.
2. Dari 21 algoritma yang digunakan untuk mendeteksi dan melakukan perhitungan terhadap objek sperma bergerak pada video (*Basic* dan *Statistical background subtraction),* algoritma *Mixture Of Gaussian V2* pada proses *background subtraction* mampu digunakan untuk mendeteksi dan melakukan perhitungan terhadap objek sperma bergerak pada video, dengan hasil *foreground* yang memiliki sedikit noise, objek *background* bergerak pada video tidak terdeteksi sebagai *foreground* objek, dan bentuk sperma dapat terekstraksi lebih sempurna. Dari hasil pengujian yang telah dilakukan, dalam melakukan deteksi dan perhitungan terhadap sperma bergerak algoritma *Mixture Of Gaussian V2* memiliki nilai *f-measure* sebesar 0.9449. Hasil ini merupakan nilai yang paling tinggi dibandingkan dengan algoritma *background subtraction* lain yang dicoba. Ini menunjukkan bahwa algoritma *Mixture Of Gaussian V2* sesuai untuk digunakan pada kasus deteksi dan perhitungan sperma bergerak dengan berhasil mengatasi tantangan dan keuntungan yang ada pada kasus ini. Nilai *f-measure* dari algoritma *Adaptive Background Learning* sebesar 0.9205. Beda nilai antara algoritma *Mixture Of Gaussian V2* dengan *Adaptive Background Learning* hanya 0.0244, hal ini memperlihatkan bahwa *basic model background subtraction* juga mampu untuk digunakan pada kasus deteksi dan perhitungan sperma yang bergerak.
3. Dengan perangkat yang dikembangkan, penentuan abnormalitas pergerakan *spermatozoa* manusia pada file video dapat dilakukan. Posisi pergerakan *spermatozoa* hasil penjejakan dikenali bentuk lintasannya berdasarkan rata-rata jarak posisinya terhadap garis regresi linier, dengan *threshold* 10 terdapat 4 *spermatozoa* progresif dan 4 *spermatozoa* non progresif untuk data uji, sedangkan untuk video data spermatozoa manusia terdapat 10 *spermatozoa* progresif dan 4 *spermatozoa* non progresif. Metode yang digunakan berhasil menentukan 8 (delapan) *spermatozoa* data UNSW *Embryology*, dan 14 *spermatozoa* manusia. Dari 8 *spermatozoa* data uji yang dijejaki, terdapat 50% progresifdan 50% non progresif. Sedangkan untuk 14 *spermatozoa* manusia yang dijejaki, terdapat 71% progresif dan 29% non progresif. Menurut *WHO laboratory manual for the examination and processing of human semen* tahun 2010 dengan nilai 71% progresif berarti pergerakan *spermatozoa* manusia normal

**5.2. Saran Pengembangan**

1. Dalam penelitian ini sperma yang bergerak telah dapat dideteksi dan dihitung jumlahnya. Hal ini dapat dikembangkan untuk penelitian selanjutnya yang berkatian tentang analisa tingkat infertilitas sperma. Hasil dari deteksi sperma dapat dikembangkan dengan melakukan penjejakan terhadap sperma yang terdeteksi sehingga dapat diketahui pola pergerakan dari sperma. Hasil dari deteksi sperma juga dapat digunakan untuk menganalisa bentuk (morfologi) dari sperma, karena bentuk dari sperma juga termasuk parameter dalam menentukan tingkat infertilitas sperma.
2. Dalam penelitian ini hanya digunakan tahap *prepocessing* berupa *image enhancement* dengan beberapa metode. Dalam penelitian selanjutnya dapat ditambahkan tahapan *prepocessing* yang lebih banyak untuk mengurangi *noise* pada video masukan. tahap *prepocessing* berupa *image enhancement* yang dapat ditambahkan, misalnya menambah kecerahan video atau *histogram equalization* pada video masukan sebelum melakukan tahapan *background substraction*.
3. Algortima *matching-based* yang diterapkan dalam penelitian ini masih belum optimal, masih ditemukan beberapa faktor yang tidak dapat diatasi atau menjadi penyebab kesalahan dalam penjejakan dan penentuan abnormalitas pergerakan. Dalam penelitian selanjutnya dapat menerapkan metode lain atau mengoptimalkan algoritma *matching-based* agar dapat mengatasi faktor-faktor yang menjadi penyebab kesalahan penjejakan dan penentuan abnormalitas pergerakan.

*Halaman ini sengaja dikosongkan*

# DAFTAR PUSTAKA

1. D. S. Alex, and A. Wahi, (2014) “Background subtraction frame difference algorithm for moving object detection and extraction,” *Journal of Theoretical and Applied Information Technology*, vol. 60, no. 3, pp. 623–628.
2. H. B. Basoeki, A.D. Wibawa and I. K. E. Purnama, (2016), “Improving sperms detection and counting using single Gaussian background subtraction,” in *International Seminar on Application for Technology of Information and Communication (ISemantic)*. IEEE, pp. 295–299.
3. Borges Jr, E., Setti, A., Braga, D., Figueira, R. and Iaconelli Jr, A. (2016), “Total motile sperm count has a superior predictive value over the who 2010 cut-off values for the outcomes of intracytoplasmic sperm injection cycles,” *Andrology Journal*, vol. 4, no. 5, pp. 880–886.
4. Djurayev, A. and Primora, G., (1990), “Automatic interesting object extraction from images based on edge information and texture analysis,” *Interna- tional Journal of Scientific and Research Publications*, vol. 6, no. 12, pp. 395–399.
5. Dst…………