**SISTEM DETEKSI *HATE SPEECH* BERBAHASA INDONESIA MENGGUNAKAN METODE *INDONESIAN BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS* (INDOBERT)**

**PROPOSAL SKRIPSI**

diajukan untuk menempuh ujian sarjana

pada Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Padjadjaran

RIZKY ANUGERAH

NPM 140810180049



UNIVERSITAS PADJADJARAN  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
PROGRAM STUDI TEKNIK INFORMATIKA  
SUMEDANG  
2022

# DAFTAR ISI

[DAFTAR ISI ii](#_Toc94314050)

[DAFTAR TABEL iii](#_Toc94314051)

[DAFTAR GAMBAR iv](#_Toc94314052)

[BAB I PENDAHULUAN 1](#_Toc94314053)

[1.1 Latar Belakang 1](#_Toc94314054)

[1.2 Identifikasi Masalah 2](#_Toc94314055)

[1.3 Batasan Masalah 3](#_Toc94314056)

[1.4 Tujuan dan Manfaat Penelitian 3](#_Toc94314057)

[BAB II TINJAUAN PUSTAKA 5](#_Toc94314058)

[2.1 Landasan Teori 5](#_Toc94314059)

[2.1.1 *Hate Speech* 5](#_Toc94314060)

[2.1.2 *Deep Learning* 6](#_Toc94314061)

[2.1.3 *Text Preprocessing* 7](#_Toc94314062)

[2.1.4 *Bidirectional Encoder Representations from Transformers* (BERT) 7](#_Toc94314063)

[2.1.5 *Indonesian Bidirectional Encoder Representations from Transformers (*IndoBERT*)* 8](#_Toc94314064)

[2.1.6 *Confusion Matrix* 8](#_Toc94314065)

[2.2 Penelitian Sebelumnya 9](#_Toc94314066)

[BAB III METODOLOGI PENELITIAN 11](#_Toc94314067)

[3.1 Metode Penelitian 11](#_Toc94314069)

[3.2 Metode Pengumpulan Data 13](#_Toc94314070)

[3.3 Metode Pengembangan Sistem 14](#_Toc94314071)

[BAB IV JADWAL PENELITIAN 16](#_Toc94314072)

[4.1 *Timeline* Penjadwalan Penelitian 16](#_Toc94314074)

[DAFTAR PUSTAKA 17](#_Toc94314075)

# DAFTAR TABEL

[Tabel 4.1 *Timeline* Penjadwalan Penelitian 16](#_Toc93969438)

# DAFTAR GAMBAR

[Gambar 2.1 Hubungan antara *Artificial Intelligence*, *Machine Learning*, dan *Deep Learning* (Chollet, 2018) 6](#_Toc94313909)

[Gambar 2.2 Arsitektur BERT 8](#_Toc94313910)

[Gambar 3.1 Diagram Alur (*Flowchart*) Pengembangan Sistem 11](#_Toc94313913)

[Gambar 3.2 Alur Pengembangan Sistem dengan Metode *Waterfall* 15](#_Toc94313914)

# BAB I

**PENDAHULUAN**

## Latar Belakang

Saat ini penggunaan internet paling banyak digunakan untuk mengakses media sosial. Dengan adanya media sosial, setiap orang bisa saling berbagi informasi terhadap orang lain tanpa harus bertemu langsung satu dengan yang lainnya dan setiap orang juga memiliki kebebasan untuk mengemukakan pendapatnya di media sosial. Namun pada kenyataannya banyak juga masyarakat yang menyalahgunakan media sosial dengan memposting mengenai kalimat bernada ujaran kebencian atau *hate speech* seperti tuduhan, fitnah, berita *hoax*, ancaman, hinaan, maupun SARA, dimana postingan seperti ini tentunya sangat meresahkan bagi masyarakat.

Dengan merebaknya postingan pengguna media sosial di Indonesia yang bernada ujaran kebencian atau *hate speech*, membuat suatu sistem yang dapat mendeteksi *hate speech* pada teks berbahasa Indonesia merupakan langkah yang sangat tepat agar masyarakat tidak lagi sembarangan dalam memposting kalimat bernada ujaran kebencian atau *hate speech*. Alasan perlu dibuatnya sistem ini agar mempermudah masyarakat dalam menentukan kalimat mana yang termasuk *hate speech* atau bukan secara otomatis karena dalam setiap waktu tentunya banyak pengguna media sosial di Indonesia yang memposting kalimat bernada *hate speech* sehingga masyarakat tidak perlu mengklasifikasikan setiap postingan yang ada satu per satu secara manual.

Untuk itu pada penelitian akan merancang sistem deteksi *hate speech* pada teks berbahasa Indonesia dengan metode IndoBERT yang digunakan untuk mengklasifikasikan teks mana yang bermakna *hate speech* dan *non-hate speech*. Sistem ini akan berfokus pada teks berbahasa Indonesia karena banyak postingan pengguna media sosial di Indonesia yang bernada *hate speech*. Alasan penggunaan IndoBERT karena terdapat model *pre-trained* berbahasa Indonesia yang disediakan jadi tidak perlu melakukan *train* dari awal dan tidak membutuhkan *data training* yang besar sehingga dapat mempersingkat waktu *train*. Berdasarkan hasil *benchmark* dengan metode lain yang dilakukan (Koto, Rahim, Lau, & Baldwin, 2020) didapatkan hasil bahwa penggunaan IndoBERT mendapatkan hasil yang lebih baik dibanding dengan metode lain seperti Bi-LSTM, mBERT, dan MalayBERT.

## Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka dapat dibuat identifikasi masalah sebagai berikut:

Bagaimana cara mengimplementasikan metode IndoBERT dalam membuat suatu sistem yang dapat mendeteksi *hate speech* pada teks berbahasa Indonesia?

Bagaimana cara membuat sebuah aplikasi berbasis *website* yang dapat mendeteksi *hate speech* pada teks berbahasa Indonesia?

## Batasan Masalah

Berdasarkan identifikasi masalah yang telah dijelaskan sebelumnya dan untuk mendapatkan hasil sesuai yang diharapkan, maka dapat dibuat batasan masalah sebagai berikut:

1. Penelitian ini mencoba untuk membuat sistem berbasis *website* untuk mendeteksi kalimat bernada *hate speech* pada teks berbahasa Indonesia dengan metode yang digunakan yaitu IndoBERT.
2. Dataset yang digunakan didapat dari penelitian yang dilakukan oleh (Alfina, Mulia, Fanany, & Ekanata, 2017), (Putri, 2018), dan (Ibrohim & Budi, 2019).
3. Hanya membagi label menjadi 2 jenis klasifikasi yaitu *hate speech* dan *non-hate speech*.

## Tujuan dan Manfaat Penelitian

Maksud dari penelitian ini adalah melihat hasil dari sistem deteksi *hate speech* pada teks berbahasa Indonesia menggunakan metode IndoBERT.

Tujuan dari penelitian ini adalah:

1. Untuk mengimplementasikan metode IndoBERT dalam membuat suatu sistem berbasis *website* yang dapat mendeteksi *hate speech* pada teks berbahasa Indonesia.
2. Dapat mengklasifikasikan teks berbahasa Indonesia menjadi 2 jenis klasifikasi yaitu *hate speech* dan *non-hate speech*.

Manfaat dari penelitian ini adalah:

1. Menghasilkan aplikasi berbasis *website* yang dapat mendeteksi *hate speech* pada teks berbahasa Indonesia.
2. Mengurangi jumlah pengguna yang menyalahgunakan media sosial dengan memposting kalimat bernada ujaran kebencian atau *hate speech*.
3. Membantu masyarakat dan aparat hukum agar dapat lebih mudah untuk mengklasifikasikan postingan yang bermakna *hate speech* dan *non-hate speech* di media sosial.

# BAB II

**TINJAUAN PUSTAKA**

## Landasan Teori

Pada bab ini dibahas tentang seluruh landasan teori yang berhubungan dan menunjang dalam proses penelitian ini. Landasan teori tersebut berisi metode yang digunakan, penjelasan teoritis mengenai algoritma yang digunakan dalam proses pengembangan sistem, dan teori lainnya yan berguna untuk memahami permasalahan yang dibahas.

Landasan teori yang ada didapatkan dari studi pada penelitian sebelumnya, buku, artikel internet, dan jurnal terkait yang dapat mendukung pemecahan masalah dengan topik yang akan dibahas dalam penelitian ini.

### *Hate Speech*

Ujaran kebencian atau *hate speech* adalah tindakan komunikasi yang dilakukan oleh suatu individu atau kelompok dalam bentuk provokasi, hasutan, ataupun hinaan kepada individu atau kelompok lain dalam hal berbagai aspek seperti ras, warna kulit, jenis kelamin, cacat, orientasi seksual, kewarganegaraan, agama dan lain-lain (Zulkarnain, 2020).

Kejahatan ujaran kebencian dapat dilakukan melalui berbagai media, antara lain dalam orasi kegiatan kampanye, spanduk atau *banner*, media sosial, penyampaian pendapat, ceramah keagamaan, media masa, cetak, ataupun elektronik (Prawira, 2016).

### *Deep Learning*

*Deep Learning* adalah salah satu bidang *Machine Learning* yang memanfaatkan banyak *layer* pengolahan informasi nonlinier untuk melakukan ekstraksi fitur, pengenalan pola, dan klasifikasi (Deng & Yu, 2014)*. Deep Learning* bekerja untuk mempelajari sehingga tidak hanya dapat memprediksi tetapi juga merepresentasikan data dengan benar, sehingga cocok untuk melakukan prediksi (Goldberg, 2017).

Pendekatan *Deep Learning* memungkinkan komputer untuk mempelajari suatu model yang rumit dengan cara membangun suatu model yang sederhana (Goodfellow, Bengio, & Courvil, 2016). Berdasarkan hal tersebut, *Deep Learning* dapat dikatakan sebagai turunan dari *Machine Learning*, dimana metode ini terdiri dari banyak tingkatan proses informasi *non-linear* dan abstraksi untuk dapat melakukan *supervised*, *unsupervised learning* dan *representation*, klasifikasi, dan pengenalan pola (Deng & Yu, 2014).

Diagram, venn diagram

Description automatically generated

Gambar 2.1 Hubungan antara *Artificial Intelligence*, *Machine Learning*, dan *Deep Learning* (Chollet, 2018)

### *Text Preprocessing*

*Text preprocessing* merupakan tahapan proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah selanjutnya (Feldman & Sanger, 2007). *Text preprocessing* digunakan untuk mempersiapkan teks sebelum digunakan dalam pengujian atau pelatihan dengan tujuan untuk mengurangi *noise* yang ada pada data sehingga dapat meningkatkan kinerja *classifier* dan mempercepat proses klasifikasi (Sivakumar & Gunasundari, 2017). Langkah-langkah pada *text* *preprocessing* berupa *cleaning*, *case folding*, *tokenizing, normalization*, *filtering*, dan *stemming*.

### *Bidirectional Encoder Representations from Transformers* (BERT)

*Bidirectional Encoder Representations from Transformers* atau BERT adalah algoritma *Deep Learning* yang dirancang untuk mengolah *Natural Language Processing* atau NLP yang dikembangkan oleh para peneliti di Google AI Language pada tahun 2018.

*Transformer* adalah sebuah mekanisme yang mempelajari hubungan kontekstual antara kata-kata dalam teks (Vaswani, et al., 2017). Dalam bentuk, *transformer* menyertakan dua mekanisme terpisah, yaitu *encoder* yang membaca input teks dan *decoder* yang menghasilkan prediksi untuk tugas tersebut. Karena tujuan BERT adalah untuk menghasilkan model bahasa, hanya diperlukan mekanisme *encoder* (Putra, Bijaksana, & Romadhony, 2021). Arsitektur pada metode BERT dapat dilihat pada Gambar 2.2.

Diagram

Description automatically generated

Gambar 2.2 Arsitektur BERT

### *Indonesian Bidirectional Encoder Representations from Transformers (*IndoBERT*)*

IndoBERT merupakan salah satu nonlingual model BERT untuk bahasa Indonesia. IndoBERT sendiri merupakan modifikasi dari BERT-Base yang sudah ada dengan mengikuti kofigurasi dari BERT-Base (*uncased*) yang memiliki 12 *hidden layers* masing masing memiliki 768d, 12 *attention heads*, dan *feed-forward hidden layers* 3,072d (Koto, Rahim, Lau, & Baldwin, 2020).

Jika di total IndoBERT di-*train* dengan lebih dari 220 juta kata. Terdiri dan 3 sumber utama, yaitu Wikipedia Indonesia (74 juta kata), artikel Kompas, Tempo, dan Liputan6 (total 55 juta), dan Web Corpus Indonesia (90 juta kata) (Koto, Rahim, Lau, & Baldwin, 2020).

### *Confusion Matrix*

*Confusion Matrix* merupakan sebuah metode untuk mengukur performa dan melakukan evaluasi terhadap model klasifikasi pada *Machine Learning*. *Confusion Matrix* berupa tabel yang menyatakan klasifikasi dan jumlah data uji benar dan data uji salah yang memuat informasi hasil perbandingan data prediksi dan data aktual (Terrada, Cherradi, Raihani, & Bouattane, 2019). Evaluasi dengan menggunakan metode confusion matrix menghasilkan nilai *accuracy*, *precision*, *recall*, *F1-score*, dan *support*. Model *Confusion Matrix* dapat dilihat pada Tabel 2.1.

Tabel 2.1 *Confusion Matrix*

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Actual Class | |
| Positive (P) | Negative (N) |
| Predicted Class | Positive (P) | True Positive (TP) | False Positive (FP) |
| Negative (N) | False Negative (FN) | True Negative (TN) |

## Penelitian Sebelumnya

Penelitian yang dilakukan oleh penulis ini dibangun berdasarkan beberapa referensi dari penelitian yang sudah dilakukan sebelumnya. Referensi yang didapat dari penelitian sebelumnya sangat penting untuk dilakukan agar penelitian ini terhindar dari kesamaan atau duplikasi dari penelitian sebelumnya. Hal ini dilakukan sebagai bentuk kontribusi penelitian bagi penulis agar penelitian mengenai topik ini terus berkembang.

Pada penelitian yang dilakukan oleh (Ibrohim & Budi, 2019), peneliti mencoba untuk mengklasifikasi teks multi-label untuk kalimat *abusive* dan deteksi *hate speech* menggunakan pendekatan *Support Vector Machine* (SVM), *Naive Bayes* (NB), dan *Random Forest Decision Tree* (RFDT) *classifier* dan *Binary Relevance* (BR), *Label Power-set* (LP), dan *Classifier Chains* (CC) sebagai metode transformasi data. Hasil percobaan menunjukkan bahwa secara umum RFDT *classifier* menggunakan LP sebagai metode transformasi memberikan akurasi terbaik dengan waktu komputasi yang paling cepat dibanding dengan pendekatan lainnya.

Pada penelitian yang dilakukan oleh (Wilie, et al., 2020), peneliti mengenalkan IndoNLU, yaitu sebuah koleksi sumber untuk riset dalam topik *Natural Language Understanding* (NLU) untuk Bahasa Indonesia. Dari penelitian ini dihasilkanlah 4 model IndoBERT, yaitu IndoBERT-base, IndoBERT-large, IndoBERT-lite-base, dan IndoBERT-lite-large.

Pada penelitian yang dilakukan oleh (Koto, Rahim, Lau, & Baldwin, 2020), peneliti memaparkan tentang IndoBERT dan IndoLEM. IndoBERT adalah model BERT *pre-trained* berbahasa Indonesia yang dilatih dengan *Transformer*, sebuah algoritma yang diadaptasi dari algoritma CNN. Sedangkan IndoLEM adalah dataset NLP berbahasa Indonesia yang dapat digunakan untuk berbagai tugas seperti tugas *labelling*, analisis sentimen, *text summarization*, dan prediksi *tweet*.

Pada penelitian yang dilakukan oleh (Putra, Bijaksana, & Romadhony, 2021), peneliti mencoba untuk mengimplementasikan metode IndoBERT untuk merancang sistem deteksi penggunaan kalimat *abusive* pada teks bahasa Indonesia. Data terdiri dari tiga label, yaitu *Not Abusive*, *Abusive Not Offensive*, dan *Abusive And Offensive*. Hasil pengujian menunjukkan IndoBERT dapat lebih baik dalam mengklasifikasikan kalimat *abusive* pada teks bahasa Indonesia. Di model BERT-base, berhasil menghasilkan nilai *F1 Score* untuk semua kelas sebesar 0.6842.

# BAB III

**METODOLOGI PENELITIAN**



## Metode Penelitian

Sistem yang akan dirancang merupakan sistem berbasis *website* yang dapat mendeteksi kalimat *hate speech* pada teks bahasa Indonesia. Sistem yang akan dikembangkan pada penelitian ini dapat dilihat pada diagram alur (*flowchart*) di bawah ini.



Gambar 3.1 Diagram Alur (*Flowchart*) Pengembangan Sistem

Berdasarkan diagram alur (*flowchart*) pengembangan sistem di atas, rincian tahapan metode penelitian ini adalah sebagai berikut:

1. Pengumpulan Data

Data pada penelitian ini didapatkan dari dataset yang sudah dibangun dari penelitian (Alfina, Mulia, Fanany, & Ekanata, 2017), (Putri, 2018), dan (Ibrohim & Budi, 2019). Dataset didapatkan dengan cara melakukan *crawling* terhadap *tweet* menggunakan API Twitter yang diimplementasikan menggunakan *library* Tweepy. Kata kunci yang digunakan untuk melakukan *crawling* data adalah kata berbahasa Indonesia yang sering digunakan oleh masyarakat ketika memposting kalimat *hate speech* dan *abusive* di media sosial Twitter. Peneliti tersebut melakukan *crawling* terhadap data Twitter selama kurang lebih 7 bulan, dimulai dari tanggal 20 Maret 2018 sampai dengan tanggal 10 September 2018. Total data yang didapatkan sebanyak 13169 data yang terdiri dari 2 label, yaitu *Hate Speech* sebanyak 5561 data dan *Non-Hate Speech* sebanyak 7608 data.

1. *Preprocessing* Data

*Preprocessing* data dilakukan agar data yang didapatkan sebelumnya dapat menjadi data yang lebih efisien, bermanfaat, dan memberikan hasil yang lebih baik untuk dapat diolah pada tahap selanjutnya. Langkah-langkah yang dilakukan pada *preprocessing* data adalah *data* *cleaning*, *case folding*, *tokenizing, normalization*, dan *stemming*.

1. Klasifikasi Label

Data yang sudah diproses pada tahap sebelumnya akan diklasifikasikan berdasarkan label yang sudah diberikan pada setiap datanya. Untuk label *Hate Speech* akan diberi nilai 0 sedangkan label *Non-Hate Speech* akan diberi nilai 1. Setelah itu data disimpan dalam *file* *comma-separated values* (CSV) dan di-*split* menjadi 2 data, yaitu dataset untuk *training* dan *testing*.

1. Klasifikasi dan Pemodelan dengan *Pre-Trained* IndoBERT

Pada tahap ini dilakukan proses klasifikasi teks bahasa Indonesia dengan menggunakan model IndoBERT-Base yang didapat dari penelitian yang dilakukan oleh (Wilie, et al., 2020). Terdapat 2 langkah yang dilakukan pada penggunaan metode IndoBERT yaitu *pre-training* dan *fine tuning*.

1. Evaluasi Hasil

Evaluasi hasil klasifikasi yang didapat dari tahap sebelumnya dengan menggunakan *confusion matrix* dan menghitung nilai *accuracy*, *precision*, *recall*, *F1-score*, dan *support*. Kemudian simpan hasil *trained model* yang sudah dibuat.

1. Analisis Hasil Klasifikasi

Agar tujuan pada penelitian ini tercapai, maka dibuatlah sistem berbasis *website* untuk menganalisis hasil dari model yang sudah dibuat sebelumnya. Skenario pengujiannya dibagi menjadi dua, yang pertama yaitu pengguna nantinya bisa memasukkan teks atau kalimat yang ingin dicek, lalu *output* yang dihasilkan adalah klasifikasi label apakah teks atau kalimat yang dimasukkan termasuk *hate speech* atau *non-hate speech*.

Selanjutnya yang kedua yaitu klasifikasi *hate speech* dengan memasukkan kata kunci yang ingin dicari, lalu *output* yang dihasilkan yaitu hasil klasifikasi yang didapatkan dengan melakukan *crawling* pada media sosial Twitter dan *output*-nya dapat dilihat dalam format CSV.

## Metode Pengumpulan Data

Metode pengumpulan data adalah cara yang dilakukan peneliti untuk memperoleh data yang akan digunakan pada penelitian ini. Pengumpulan data memegang peranan yang sangat penting untuk mendapatkan informasi dari penelitian yang dilakukan. Langkah-langkah dalam pengumpulan data yang digunakan pada penelitian ini adalah sebagai berikut:

1. Studi Literatur

Pada bagian ini penulis melakukan pengumpulan data yang berhubungan dengan permasalahan penelitian yang bersumber dari studi pada penelitian sebelumnya, buku, artikel internet, dan jurnal terkait yang dapat mendukung pemecahan masalah dengan topik yang akan dibahas dalam penelitian ini.

1. Observasi

Pada bagian ini penulis melakukan pengamatan atau observasi secara langsung untuk mengumpulkan data yang dapat mendukung pemecahan masalah dengan topik yang akan dibahas dalam penelitian ini.

Observasi yang dilakukan oleh penulis berupa pengumpulan dataset yang didapat dari penelitian sebelumnya, yaitu penelitian yang dilakukan oleh (Alfina, Mulia, Fanany, & Ekanata, 2017), (Putri, 2018), dan (Ibrohim & Budi, 2019). Total data yang didapatkan sebanyak 13169 data yang terdiri dari 2 label, yaitu *Hate Speech* sebanyak 5561 data dan *Non-Hate Speech* sebanyak 7608 data.

## Metode Pengembangan Sistem

Pada proses pengembangan suatu sistem, terdapat sebuah pemodelan yang disebut sebagai *Software Development Life Cycle* (SDLC). SDLC adalah siklus yang digunakan dalam pembuatan atau pengembangan sistem informasi yang bertujuan untuk menyelesaikan masalah secara efektif agar dapat menghasilkan sistem berkualitas tinggi yang sesuai dengan keinginan atau tujuan dibuatnya sistem tersebut. Dalam rekayasa sistem dan rekayasa sebuah *software*, SDLC juga digunakan untuk membangun suatu sistem informasi agar dapat berjalan sesuai dengan apa yang diharapkan (Gestama, 2020). Terdapat banyak sekali jenis pemodelan dari SDLC ini, salah satunya adalah model *Waterfall*.

Metode yang digunakan untuk mengembangkan sistem ini adalah metode *Waterfall*.Metode *Waterfall* dapat dikatakan sebagai model yang tertua dan tersingkat dalam penerapannya. Pada model ini ketika satu fase selesai maka akan lanjut ke fase berikutnya. Proses pengembangan sistem dengan menggunakan model ini dilakukan secara bertahap dan berurutan mulai dari tahap perencanaan, konsep, pemodelan (*design*), implementasi, pengujian, dan pemeliharaan.

Metode *Waterfall* memiliki tahapan-tahapan sebagai berikut (Sommerville, 2011):

1. *Requirement analysis and definition*
2. *System and software design*
3. *Implementation and unit testing*
4. *Integration and system testing*
5. *Operation and maintenance*

Diagram

Description automatically generated

Gambar 3.2 Alur Pengembangan Sistem dengan Metode *Waterfall*

# BAB IV

**JADWAL PENELITIAN**



## *Timeline* Penjadwalan Penelitian

Penelitian ini membutuhkan perkiraan waktu sekitar 4 bulan dengan detail jadwal penelitian dapat dilihat pada Tabel 4.1.

Tabel 4.1 *Timeline* Penjadwalan Penelitian

| **Kegiatan** | **Bulan ke-1**  **(Februari)** | | | | **Bulan ke-2**  **(Maret)** | | | | **Bulan ke-3**  **(April)** | | | | **Bulan ke-4**  **(Mei)** | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** |
| Permbuatan Proposal |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Pengumpulan dan Pengolahan Data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Perancangan Model |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Implementasi dan Pengujian Model |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Dokumentasi dan Laporan Akhir |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# DAFTAR PUSTAKA

Alfina, I., Mulia, R., Fanany, M. I., & Ekanata, Y. (2017). Hate Speech Detection in the Indonesian Language: A Dataset and Preliminary Study. *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 233-237. doi:10.1109/icacsis.2017.8355039

Chollet, F. (2018). *Deep Learning with Python.* Shelter Island: Manning Publications Co.

Deng, L., & Yu, D. (2014, June). Deep Learning: Methods and Applications. *Foundations and Trends in Signal Processing, 7*, 197-387. doi:https://doi.org/10.1561/2000000039

Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data.* Cambridge Univ. Press.

Gestama, P. (2020, February 1). *PENGERTIAN SDLC adalah: Fungsi, Metode dan Tahapan SDLC*. Retrieved from Salamadian: https://salamadian.com/sdlc-system-development-life-cycle/

Goldberg, Y. (2017, April). Neural Network Methods for Natural Language Processing. (G. Hirst, Ed.) *Synthesis Lectures on Human Language Technologies, 10*(1), 1-309. doi:https://doi.org/10.2200/S00762ED1V01Y201703HLT037

Goodfellow, I., Bengio, Y., & Courvil, A. (2016). *Deep Learning.* Massachusetts: MIT Press.

Ibrohim, M. O., & Budi, I. (2019, August). Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter. *Proceedings of the Third Workshop on Abusive Language Online*, 46-57. doi:10.18653/v1/W19-3506

Koto, F., Rahim, A., Lau, J. H., & Baldwin, T. (2020, November 2). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *In Proceedings of the 28th COLING*. Retrieved from https://arxiv.org/pdf/2011.00677.pdf

Prawira, Y. (2016). Upaya Kepolisian Dalam Menanggulangi Kejahatan Ujaran Kebencian (Hate Speech) Berdasarkan Surat Edaran Kapolri NO.SE/06/X/2015. *Skripsi*.

Putra, H. K., Bijaksana, M. A., & Romadhony, A. (2021, April). Deteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia Menggunakan Metode IndoBERT. *e-Proceeding of Engineering, 8*(2), 3028-3038.

Putri, T. T. (2018). Analisis dan deteksi hate speech pada sosial twitter berbahasa indonesia. *Master's thesis*, Master's thesis, Faculty of Computer Science, Universitas Indonesia.

Sivakumar, A., & Gunasundari, R. (2017). A Survey on Data Preprocessing Techniques for Bioinformatics and Web Usage Mining. *International Journal of Pure and Applied Mathematics, 117*(20), 785-794.

Sommerville, I. (2011). *Software Engineering (9th Edition).* New York: Addison-Wesley.

Terrada, O., Cherradi, B., Raihani, A., & Bouattane, O. (2019). Classification and Prediction of atherosclerosis diseases using machine learning algorithms. *2019 5th International Conference on Optimization and Applications (ICOA)* (pp. 1-5). Kenitra, Morocco: IEEE. doi:10.1109/ICOA.2019.8727688

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017, December 6). Attention Is All You Need. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000-6010. Retrieved from https://arxiv.org/abs/1706.03762

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., . . . Purwarianti, A. (2020, December). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 843-857. Retrieved from https://aclanthology.org/2020.aacl-main.85

Zulkarnain. (2020). Ujaran Kebencian (Hate Speech) Di Masyarakat Dalam Kajian Teologi. *Studia Sosia Religia*, 70-82.