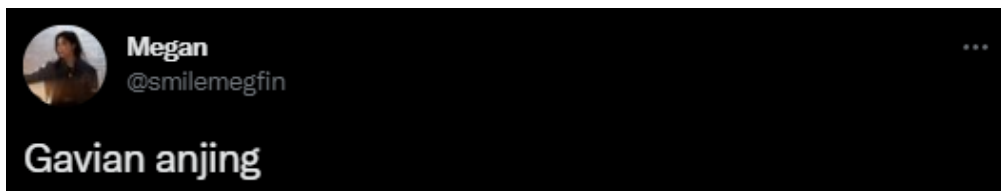


BAB III

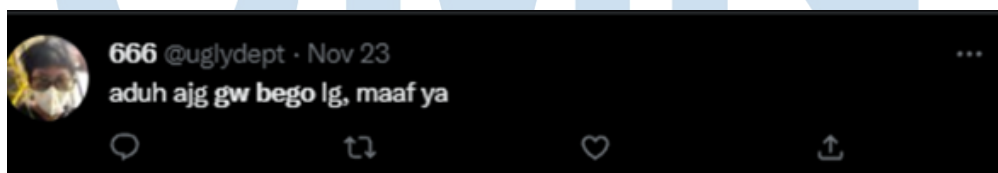
METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Penelitian ini bertujuan untuk mengidentifikasi dan melakukan pendeteksian teks yang mengandung *cyberbullying* di media sosial platform X dengan bantuan metode *deep learning* LSTM dengan *feature selection* Particle Swarm Optimization dan *Salp Swarm Algorithm* untuk membedakan teks yang terdapat unsur *cyberbullying* dan teks yang tidak terdapat *cyberbullying*. Hal ini dikarenakan terkadang ketika user melakukan pengetikan teks *tweets* di platform X user menggunakan kata-kata kasar yang terkadang digunakan juga sebagai imbuhan tidak baku dan sebagai kata sambung dalam menyampaikan kata. Contoh ini dapat dilihat pada Gambar 3.1 terdapat teks *tweets* yang mengarah ke *cyberbullying* kepada orang lain. Tetapi, pada Gambar 3.2 teks *tweets* yang dibuat oleh user mengarah kepada kata sambung atau kata kata kasar yang tidak mengarah kepada tindakan *cyberbullying* kepada orang lain



Gambar 3. 1 Contoh Tweet Yang Mengandung *cyberbullying*
Sumber: platform X



Gambar 3. 2 Contoh Tweet Yang tidak Mengandung *cyberbullying*
Sumber: platform X

Objek penelitian ini akan dilakukan dengan dataset primer dikarenakan kurangnya sumber data yang relevan yang tersedia dan dapat diakses secara publik mengenai data *cyberbullying* di media sosial platform X. Selain itu penggunaan dataset primer ini diharapkan dapat memenuhi kebutuhan dataset

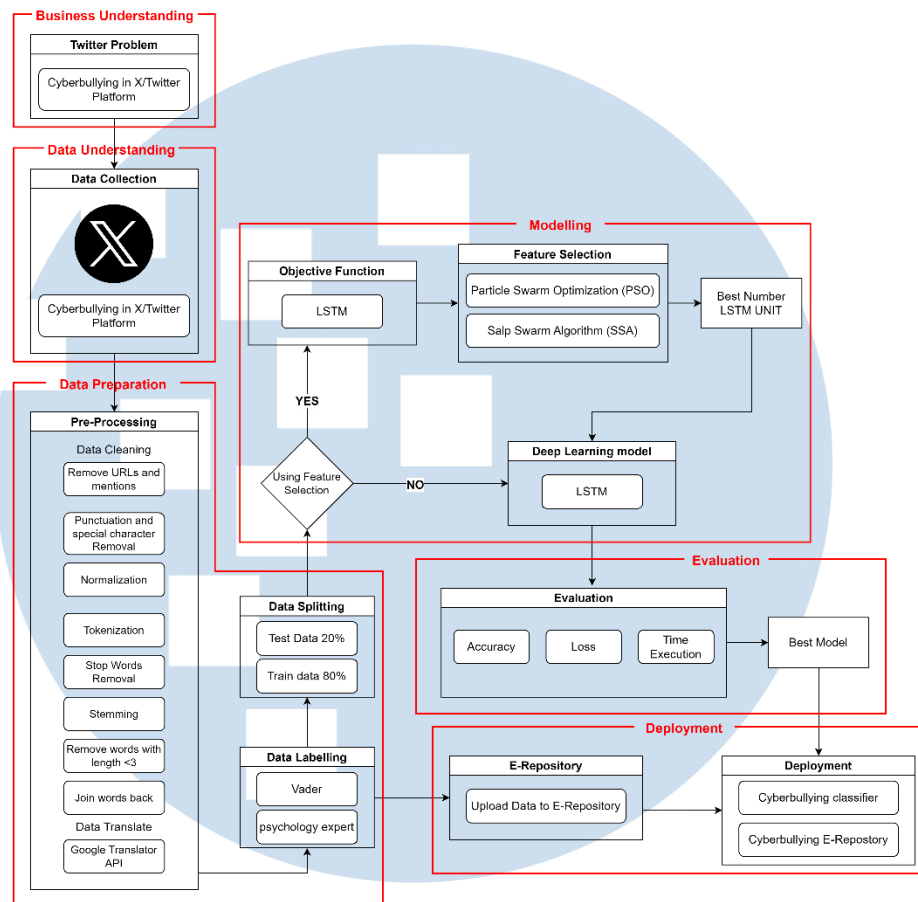
dengan menggunakan kata kunci *cyberbullying* yang ditentukan dengan penelitian terdahulu [98] dan dibantu diterjemahkan ke dalam bahasa Indonesia dari bahasa Inggris untuk mendapatkan kata *cyberbullying* berbahasa Indonesia yang sesuai. Metode pengambilan data akan dilakukan menggunakan data scraping dengan kata kunci “bajingan, anjing, tai, jablay, dan goblok” beserta beberapa *thread* platform X yang mengandung *cyberbullying* dari periode 31 Desember 2023 – 1 Februari 2024 yang hasilnya didapatkan 2000 data.

3.2 Metode Penelitian

3.2.1 Alur Penelitian

Alur penelitian merupakan gambaran yang berperan sebagai acuan untuk memastikan bahwa proses dilakukan penelitian berjalan dengan teratur dan sistematis sesuai dengan alur yang sudah ditentukan. Pada penelitian ini alur penelitian digambarkan seperti pada Gambar 3.3. penelitian diawali dengan melakukan pengambilan data yang akan dilakukan di platform X menggunakan metode data scraping dengan menggunakan *twint library*. Data yang sudah diambil nantinya akan diproses melalui tahap *pre-processing* didalam tahap ini dikarenakan data masih dalam keadaan mentah yang dapat mengintervensi hasil dan jalannya penelitian data akan dilakukan berbagai macam *pra-pemrosesan* seperti penghapusan url dan hashtag, melakukan *stemming*, penghapusan kata-kata dibawah 3 kata. Setelah data selesai dilakukan pemrosesan data akan ditranslate dari bahasa Indonesia ke bahasa Inggris agar dapat dilakukan *labeling* otomatis menggunakan *VADER*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A



Gambar 3. 3 Alur Penelitian

Setelah dilakukan *labelling* menggunakan vader data yang sudah dilabel akan diperiksa kembali oleh mahasiswa lulusan psikologi untuk menentukan dan membenarkan apabila terdapat kesalahan pemberian label yang dilakukan oleh vader. Setelah verifikasi data selesai data akan diupload ke dalam E-repository *cyberbullying* yang akan menjadi *database* penyimpanan data yang dapat diakses oleh peneliti lain sesuai dengan kebutuhan penelitian mendatang.

Data yang sudah diberi label sebelumnya akan *displitting* menjadi 20% test data dan 80% train data. Data yang sudah di *split* akan di *run* menggunakan model *deep learning* LSTM untuk mendapatkan hasil tanpa optimisasi lalu dievaluasi dari akurasi, *loss*, dan waktu eksekusi. Setelah mendapatkan hasil tanpa feature selection akan dilakukan optimisasi *hyperparameter* optimisasi *lstm_unit* dengan menggunakan algoritma

optimisasi yaitu *Particle Swarm Optimization* dan *Salp Swarm algorithm* yang dilakukan dengan bantuan *objective function* untuk mengevaluasi kinerja model LSTM selama pelatihan dan memberikan umpan balik kepada model tentang seberapa baik model tersebut melakukan prediksi terhadap data latih. Dari hasil optimisasi akan didapat number lstm unit yang paling optimal digunakan untuk lstm yang akan dirun ulang menggunakan LSTM dan dievaluasi dan dibandingkan akurasi, *loss*, dan waktu pengerjaan atau eksekusi untuk mendapatkan hasil yang paling baik.

3.2.2 Metode Pengembangan Data Mining

Framework yang akan dipakai untuk melakukan pengembangan data mining untuk penelitian ini akan dibandingkan 2 tipe framework yang cukup terkenal didalam melakukan penelitian yaitu *data mining* yaitu *crisp-dm* dan *kdd*. Tabel 3.2 menggambarkan perbandingan framerowk data minig CRISP-DM dan KDD.

Tabel 3. 1 perbandingan framework data mining
Sumber: [56], [99]

Perbedaan	CRISP-DM	KDD
Langkah	6 (<i>Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.</i>)	5 (<i>selection, preprocessing, transformation, data mining, and interpretation/evaluation.</i>)
Focus	Menekankan pengertian dan pemahaman terhadap pehaman bisnis dan aspek data yang berkaitan dengan penelitian sebelum dilakukan permodelan dan evaluasi untuk memastikan hasil dari data mining sesuai dengan tujuan bisnis	Menekankan Melakukan pencarian pengetahuan dari dataset yang besar dengan fokus terhadap <i>pre-procssing, preparation, and evaluation.</i>
Flexibility	Dikenal dengan proses pengulangan, proses yang akan mendorong peninjauan kembali langkah-langkah sebelumnya ketika wawasan baru ditemukan atau ketika fokus proyek berubah dengan tujuan untuk membantu menyempurnakan hasil proyek sesuai dengan kebutuhan bisnis.	memungkinkan untuk melakukan <i>backtracking</i> dan mengulangi langkah-langkah seiring dengan diperolehnya wawasan baru dan berkembangnya kebutuhan.

Perbedaan	CRISP-DM	KDD
Application	Dapat diaplikasikan dalam proyek data mining industri bisnis maupun perusahaan.	Dapat diaplikasikan diberbagai kebutuhan yang membutuhkan eksplorasi dan pengambilan pengetahuan dari kumpulan dataset yang besar.
Goal	Untuk menyediakan model proses yang komprehensif untuk melaksanakan proyek penggalian data	Untuk mengekstrak pola yang dapat dimengerti dan berguna dari volume data yang besar.

Pada penelitian ini framework yang dipilih untuk melakukan penelitian adalah CRISP-DM. Pemilihan framework CRISP-DM. ini sesuai dengan tujuan penelitian yaitu melakukan pembentukan model klasifikasi *cyberbullying*. Berikut merupakan tahapan CRISP-DM yang akan digunakan didalam penelitian.

a) *Business Understanding*

Proses pertama didalam tahapan CRISP-DM adalah business understanding untuk menentukan tujuan penelitian. Tujuan dari penelitian ini adalah untuk membantu media sosial platform X dalam melakukan klasifikasi teks *cyberbullying* berbahasa indonesia di media sosial platform X dengan menggunakan algoritma LSTM yang digabungkan dengan *Swarm optimization Particle Swarm Optimization* dan *Salp Swarm Algorithm* dan dibandingkan hasilnya untuk mendapatkan model yang paling baik. Model yang paling baik akan digunakan didalam klasifikasi *cyberbullying* didalam website untuk mengklasifikasi teks yang mengandung *cyberbullying*

b) *Data Understanding*

Pada proses data understanding ini, data akan dikumpulkan dari platform X menggunakan metode *data crawling*. Proses ini akan melakukan pengumpulan data dengan cara memasukkan kata kunci yang akan digunakan didalam penelitian dengan menggunakan *scrapint tools* berupa *tweet-harvest*. Data yang akan diambil merupakan data yang mengandung kata kunci “bajingan, anjing, tai, jablay, dan goblok”. Pengambilan data akan dimulai dalam periode 1 bulan yang

dimulai dari 31 desember 2023 hingga 1 febuari 2024. Hal ini dikarenakan pada periode tersebut sedang dilalukannya masa pemilu yang menimbulkan banyak kata-kata ujaran kebencian dan *cyberbullying* yang dapat terjadi di media sosial platform X. Data yang sudah diambil dan sudah selesai di *scraping* akan disimpan didalam format .csv.

c) *Data Preparation*

Tahap ketiga dari tahapan didalam CRISP-DM adalah *data preparation*. Data yang sudah dikumpulkan sebelumnya akan dibersihkan dan dipersiapkan agar sesuai dengan kebutuhan *modelling deep learning* pada untuk tahap selanjutnya. Berikut merupakan tahapan preparasi data yang dilewati:

(1) Remove character yang tidak sesuai punktuasi kata dan *special character*

Didalam media sosial platform X terdapat beberapa text yang tidak dapat diolah atau tidak memberikan makna tertentu yang bisa membuat noise yang dapat mempengaruhi hasil dari data yang akan digunakan didalam *modelling*. Terdapat beberapa teks yang harus di cleaning didalam tweets yang sudah di *crawling* seperti penghapusan URL dan mention, hashtag, *special character* dan angka, dan punktuasi kata.

(2) *Normalization*

Normalization adalah proses untuk mengkonversi berbagai bentuk kata yang berbeda atau kata yang salah menjadi bentuk standar yang ditentukan sebelumnya. Proses ini membantu konsistensi kata dalam data untuk menghasilkan pemahaman dan analisis yang lebih akurat.

(3) *Tokenize twets into the words*

Tokenikasi adalah tahap dimana setiap batas-batas didalam teks akan diidentifikasi seperti spasi dan tanda baca untuk melakukan pemecahan kalimat menjadi bagian-bagian yang lebih kecil yang bermakna untuk mengidentifikasi entitas kata didalamnya.

(4) *Remove Stop Words*

Penghapusan *stop words* adalah tahapan pemrosesan yang akan melakukan penghapusan kata-kata yang lebih sering muncul didalam kata-kata namun tidak membawa banyak informasi atau makna signifikan didalam analisis data. Kata-kata yang tidak membawa banyak makna didalam data dapat menimbulkan noise atau gangguan dalam analisis teks sehingga harus dihilangkan.

(5) *Stemming words*

Stemming adalah proses untuk menghilangkan infleksi dari kata yang ada untuk menghasilkan bentuk kata dasar. *Stemming* akan dapat membantu untuk mensesederhanakan kata dari variasi kata yang ada menjadi kata dasar untuk membantu mengurangi variasi data sehingga dapat mereduksi kompleksitas teks menjadi hasil yang lebih sederhana.

(6) *Remove words with length <3*

Tahap Kelima adalah melakukan penghapusan kata yang kurang dari 3 karakter. Penghapusan kata ini dilakukan untuk mengurangi kata-kata yang kurang informatif atau tidak bermakna secara individual didalam teks.

(7) *Translate Text*

Tahap terakhir yang dilakukan dalam proses persiapan data untuk dipakai untuk modeling adalah melakukan *translate data* agar dapat diberikan *auto labeling* menggunakan VADER. *Data translation* akan dilakukan menggunakan google cloud translation API yang akan *mentraslate* seluruh text bahasa indonesia yang sudah di *pre-processing* menjadi bahasa inggris dan disimpan didalam column bahasa inggris.

Setelah keseluruhan data selesai dilakukan *pre-processing* akan dilakukan *labeling data* yang akan digunakan untuk melakukan klasifikasi dan mengkategorikan data. *Labeling* akan dilakukan menggunakan VADER yang akan melakukan labeling teks yang

sudah diubah ke bahasa inggris sebelumnya menjadi 2 klasifikasi yaitu positif dan negatif berdasarkan scor lexicon yang sudah ditentukan. Data yang sudah dilabel oleh vader akan diverifikasi oleh pakar yang berkaitan dengan psikologi untuk menentukan apakah label yang diberikan oleh vader sudah sesuai dengan makna teks sesungguhnya.

Setelah dilakukan *labelling* dan sudah diverifikasi kembali oleh pakar data sudah siap digunakan dan di upload ke dalam e-repository sebagai *database* kumpulan dataset teks yang mengandung *cyberbullying*.

d) *Modelling*

Data yang sudah dibersihkan dan dilakukan *labelling* dengan bantuan VADER dan pakar akan dapat dilanjutkan ke dalam tahap *modelling*. Pada tahap *modelling* ini akan digunakan metode deep learning dengan menggunakan algoritma LSTM. LSTM digunakan untuk mendeteksi teks karena kemampuannya dalam mengatasi masalah temporalitas dan ketergantungan konteks didalam teks. Selain itu, LSTM mempunyai kemampuan memori jangka panjang yang dapat memahami teks yang kompleks yang memerlukan pemahaman yang lebih mendalam. Proses *modelling* akan dilakukan dengan menggunakan jupyterlabs dengan menggunakan bahasa pemrograman python.

Pada tahap *modelling* ini akan di run dua tipe model yang pertama adalah model yang tanpa menggunakan optimisasi swarm yang berupa model LSTM saja yang nantinya hasil dari model akan dianalisa hasilnya menggunakan *Confusion matrix* dan dibandingkan hasilnya dari akurasi, presisi, dan waktu eksekusi . Setelah model LSTM tanpa optimisasi swarm selesai dibuat, akan dilakukan pemodelan menggunakan optimisasi swarm dengan bantuan *Swarm Intelligence optimization* dengan menggunakan tiga algoritma *Swarm Intelligence* yaitu *Particle Swarm Optimization* (PSO) dan *Swarm Salp Algoritm* (SSA)

Pada model LSTM yang dibuat akan dibuat untuk menggunakan *Swarm Intelligence* akan dibuat *objective function* yang bertujuan untuk menemukan jumlah unit LSTM tertentu yang dapat menghasilkan kinerja model yang paling optimal yang dibuat berdasarkan kriteria yang ditentukan. *Objective function* akan melakukan evaluasi kinerja model LSTM dengan berbagai unit LSTM dengan bantuan PSO dan SSA yang hasilnya akan menghasilkan jumlah unit LSTM yang terbaik berdasarkan *objective function* yang ditentukan.

Dua model *swarm intelligence* akan mendapatkan hasil number of LSTM unit yang paling baik, setiap LSTM unit tersebut akan di run kembali dengan menggunakan LSTM model yang ada dan akan dibandingkan akurasi, *loss*, dan waktu eksekusi dari model untuk mendapatkan model yang paling baik dari seluruhnya.

e) Evaluation

Dari setiap model *machine learning* yang dibuat akan *evaluation* akan dilakukan dengan menggunakan *confusion matrix* untuk mendapatkan jumlah prediksi dan memungkinkan peneliti untuk menghitung berbagai metrik seperti akurasi, presisi, dan recall. Selain *confusion matrix* akan dibandingkan juga *time execution* dari ketiga model untuk menentukan hasil model yang paling cepat melakukan klasifikasi hingga yang paling lama. Dari ketiga model yang sudah dibuat akan dilakukan perbandingan ketiga model untuk melihat model mana yang paling baik diantara LSTM tanpa optimisasi, PSO-LSTM, dan SSA LSTM. Perbandingan model ini akan dilakukan berdasarkan 3 metrik yaitu akurasi yang paling tinggi, *loss* yang paling rendah, dan waktu eksekusi yang paling cepat. Untuk membuktikan salah satu model adalah model yang paling baik, akan dilakukan *statistical test* menggunakan Shapiro dan Anova untuk menguji normalitas dan signifikansi diantara satu model dengan model lainnya.

f) Deployment

Pada tahap terakhir didalam tahapan CRISP-DM akan dilakukan *deployment* atau penggunaan model kepada user akhir yang dapat digunakan oleh user. Tahap ini dilakukan ketika evaluasi model selesai dilakukan dan mendapatkan hasil best model yang paling baik untuk digunakan. Dalam melakukan *deployment*, terdapat berbagai macam metode yang dapat dilakukan seperti, dashboard, website, aplikasi mobile.

Website yang dibuat didalam penelitian hanya akan digunakan untuk melakukan pengetesan klasifikasi model dengan fitur tambahan E-repository data yang digunakan untuk menyebarkan data untuk penelitian lain. Oleh karena itu, website tidak akan memiliki fungsi secara lengkap.

3.2.3 Metode Pengembangan Sistem Website

Metode Pengembangan sistem yang dilakukan untuk penelitian untuk pembuatan website klasifikasi *cyberbullying* dengan fitur tambahan E-repository. Terdapat 2 metode pengembangan yang cukup umum digunakan dalam melakukan pengembangan aplikasi berbasis website untuk yaitu *Prototype* dan *Waterfall*. Tabel 3.4 memberikan Gambar perbandingan pengembangan sistem dari *Prototype* dan *Waterfall*.

Tabel 3. 2 Tabel Perbandingan Pengembangan Sistem
Sumber: [47]

Aspek	<i>Prototype</i>	<i>WaterFall</i>
<i>Suitability</i>	Cocok digunakan untuk proyek yang membutuhkan keterlibatan pengguna dan persyaratannya tidak dipahami dengan baik.	Cocok untuk proyek dengan persyaratan yang jelas dan tidak ada perubahan selama proses pengembangan
<i>Flexibility</i>	Memiliki fleksibilitas yang tinggi karena memungkinkan pengujian berulang dan dimodifikasi sesuai kebutuhan pengguna.	Memiliki fleksibilitas yang rendah karena perubahan yang sulit dilakukan dan memakan biaya ketika
<i>Scalability</i>	Cocok untuk proyek yang memiliki skala kecil atau menengah dikarenakan dilakukan pengujian	Cocok untuk proyek dengan skala besar yang mempunyai tujuan yang jelas dan tidak berubah-ubah.

Aspek	<i>Prototype</i>	<i>WaterFall</i>
	berulang yang memiliki perubahan cepat.	

Berdasarkan perbandingan kedua metode dalam Tabel 3.1, ditentukan penelitian akan menggunakan metode Prototyping sebagai pendekatan penelitian. Metode ini dipilih agar proses perancangan sistem dapat dimulai dengan pembuatan prototype yang memungkinkan pengguna untuk memberikan *Feedback* secara langsung. Hal ini membuat hasil akhir dari website dapat lebih jelas dan terGambarkan sesuai dengan kebutuhan user dan penelitian. Fokus dari proses pembuatan sistem adalah pengembangan prototype *website* yang dapat menjalankan fungsi-fungsi sesuai dengan kebutuhan pengguna. Oleh Karena hal tersebut, tahapan penelitian dapat dijelaskan sebagai berikut:

1. Identifikasi kebutuhan penelitian

Pada tahap identifikasi kebutuhan penelitian akan dilakukan analisa kebutuhan website penelitian dan fitur apa saja yang diinginkan didalam website.

2. Perancangan sistem desain awal penelitian

Melakukan perancangan sistem awal penelitian sebagai garis dasar dilakukan pembuatan *wesbite*. Membangun desain awal penelitian meliputi beberapa tahapan seperi *perancangan Use Case, Activity Diagram, class diagram*, dan ERD.

3. Pembuatan prototype

Ketika pembangunan desain awal dan kerangka *website* sudah selesai maka akan dilakukan pembuatan prototype berdasarkan desain awal penelitian.

4. Evaluasi prototype

Pada tahap ini akan dilakukan prototype yang sudah dibuat untuk melakukan evaluasi apakah prototype sudah sesuai dengan kebutuhan yang dirincikan di awal.

3.3 Teknik Pengumpulan Data

Penelitian ini akan melakukan pengambilan data berupa teks kata-kata media sosial platform X yang akan disimpan didalam spreadsheet menggunakan metode *data scrapping* dengan bantuan *tools* tweet-harvest.

3.3.1 Populasi dan Sampel

Pengambilan data *tweets* akan dilakukan dengan menggunakan teknik *Purposive sampling*. difokuskan pada pengambilan *tweets* yang mengandung atau memiliki kata *Cyberbullying* dan kata yang tidak memiliki *cyberbullying*. Kata kunci *cyberbullying* terdiri dari kata “Bajingan njing, bangsat tolol, bego”. Kata kunci ini diambil dari penelitian terdahulu yang sudah dilakukan sebelumnya dalam pendeteksian kata *cyberbullying* dalam media social dalam bahasa inggris [98]. Kata-kata yang berbahasa inggris diolah kembali ke dalam bahasa indonesia dan disesuaikan dengan kata-kata *bullying* bahasa indonesia yang dibantu oleh pakar bahasa inggris yang berkerja sebagai *text translator* bahasa inggris bapak Muhammad Damarezky Joedohadi Perdana.

3.3.2 Periode Pengambilan Data

Periode pengumpulan data menggunakan metode *data scrapping* di platform X diambil dari periode 31 desember 2023 sampai dengan 1 Febuari 2024 dan dilakukan dengan metode *Purposive sampling*. Alasan dari pengambilan data pada periode ini adalah pada masa ini sedang berlangsung masa kampanye pemilu di platform X dan sedang terjadi beberapa fenomena yang viral di platform X yang dapat menimbulkan ujaran kata-kata yang mengandung *cyberbullying* kepada pihak-pihak tertentu yang dilontarkan oleh pengguna platform X kepada pihak lain atau kepada diri sendiri. Dataset yang dikumpulkan ini akan disimpan di dalam E-repository Database terkait dengan *cyberbullying* yang disimpan dan dapat digunakan untuk kebutuhan penelitian lain.

3.4 Variabel Penelitian

Variabel penelitian yang ada pada data *cyberbullying* berbahasa indonesia dapat dibagi menjadi variabel independen dan variabel dependen variabel

dependen adalah variabel yang dipengaruhi yang menjadi fokus penelitian dan variabel independen adalah variabel yang mempengaruhi variabel dependen.

3.4.1 Variabel Independen

Variabel Independen merupakan sebuah variabel yang dapat mempengaruhi klasifikasi teks berbahasa Indonesia apakah terdeteksi sebagai *cyberbullying* yaitu variabel *sentiment_score*. *Sentiment score* akan mempengaruhi variabel dependen dengan mempengaruhi dan mengklasifikasi apakah variabel text positif atau negatif *cyberbullying*.

3.4.2 Variabel Dipenden

Variabel Dependen merupakan sebuah variabel yang dipengaruhi oleh variabel lain dan merupakan variabel yang ingin diteliti didalam penelitian. Dalam konteks penelitian, variabel yang dipenden adalah variabel teks *cyberbullying* berbahasa Indonesia yang dikumpulkan dari platform X yang akan diklasifikasi dari variabel independen yaitu *sentiment_score*.

3.5 Teknik Analisis Data dan Implementasi Metode

Pada penelitian ini diperlukan *tools* untuk mengimplementasikan metode *sentimen analysis* kedalam algoritma LSTM yang akan digabungkan dengan *optimization feature selection* berbasis *Swarm Intelligence*. Proses analisa dan pengembangan model dapat dilakukan dengan bahasa pemrograman yang sudah sering dilakukan, yaitu python dan R yang dapat digunakan sebagai *tools* untuk melakukan data preparation, training/modelling, dan melakukan perbandingan model. Tabel 3.4 merupakan perbandingan dari bahasa pemrograman Python dan R.

Tabel 3. 3 Perbandingan bahasa R dan Python
Sumber: [100], [101].

Aspek	R	Python
Fokus	R dibangun sebagai software untuk melakukan statistik dan analisis data yang didukung juga dengan visualisasi data	Python sebagai bahasa pemrograman tingkat tinggi yang umum dan serbaguna digunakan dengan pengaplikasian yang luas dari pengembangan website samapi dengan <i>machine learning</i> .

Library	Menyediakan berbagai <i>library</i> untuk melakukan analisis statistik dan visualisasi data	Memiliki banyak <i>library</i> membantu pengembangan aplikasi, analisis data, dan <i>machine learning</i> .
<i>Machine learning</i>	Menyediakan <i>library</i> seperti <i>caret</i> , <i>nnet</i> namun lebih sering digunakan untuk statistik data.	Menyediakan berbagai macam <i>library</i> untuk <i>machine learning</i> maupun <i>deep learning</i> dengan berbagai macam <i>library</i> seperti Tensor Flow, Keras, Scikit-learn
Integrasi	Dapat diintegrasikan dengan database dan software statistik	Dapat diintegrasikan dengan berbagai api dan software lain

Berdasarkan perbandingan yang sudah dibuat pada Tabel 3.4, dapat disimpulkan bahwa kedua bahasa pemrograman memiliki keunggulan masing-masing di area yang berbeda. R memiliki keunggulan didalam bidang statistik dan visualisasi data, sedangkan Python memiliki keunggulan yang memiliki banyak kegunaan dan sangat fleksibel dalam penggunaannya. Oleh karena itu, bahasa pemrograman yang akan digunakan dalam penelitian ini adalah bahasa python. Pemilihan menggunakan python didasarkan pada kegunaan yang beragam dan fleksibel serta mendukung proses persiapan data, pengembangan model, hingga sampai ke dalam pembuatan website pada proses *deployment*. Dalam melakukan proses *deployment* dari model ke dalam website, penggunaan *Application Programming Interface* (API) dapat menjadi solusi yang lebih efektif yang menjadi komunikator antara model dengan website. Dalam penggunaan python, terdapat beberapa *framework* untuk API yang dapat digunakan, diantaranya Flask dan Django. Tabel 3.5 Adalah perbandingan diantara kedua framework API.

Tabel 3. 4 Perbandingan Flask dan Django
Sumber: [102], [103].

Aspek	Flask	Django
Fokus	<i>framework</i> yang memberikan fleksibilitas tinggi yang dapat dikembangkan sesuai kebutuhan	<i>Framework</i> yang memiliki banyak fitur dan menyediakan seluruh kebutuhan pengembangan dari awal
Skalabilitas	digunakan untuk aplikasi dengan skala kecil hingga menengah.	digunakan untuk aplikasi dengan skala besar
Kompleksitas	Memiliki struktur lebih sederhana dan memiliki pendekatan minimalis	Memiliki struktur lebih kompleks karena fitur dan fungsionalitas yang luas
Penggunaan	Cocok untuk proyek yang dibutuhkan cepat, membutuhkan fleksibilitas, dan struktur yang mudah disesuaikan	Cocok untuk proyek yang memerlukan fitur lebih lengkap

Berdasarkan perbandingan pada Tabel 3.5 Flask akan menjadi pilihan *framework* API yang sesuai untuk penelitian. Hal ini dikarenakan Flask menyediakan tingkat kompleksitas yang lebih rendah dan memiliki tingkat kompleksitas yang lebih sederhana dalam melakukan *deployment* ke dalam website dibandingkan dengan django. Pendekatan minimalis dan cepat yang disediakan oleh Flask juga memungkinkan proses integrasi dan implementasi model ke dalam website dapat berjalan dengan efisien dan cepat yang sejalan dengan tujuan dan keterbatasan penelitian dalam sumber daya dan waktu yang terbatas.



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA