BAB II
TINJAUAN PUSTAKA

2.1 Penelitian yang Relevan

Tabel 2.1 Penelitian yang Relevan

PENULIS (TAHUN)	TUJUAN / FOKUS	METODE UTAMA	HASIL KUNCI YANG RELEVAN
IFFA (2025)	Peningkatan Kinerja Klasifikasi Sentimen dengan Dataset Terbatas	SVM + BERT	Kombinasi BERT dan SVM berhasil meningkatkan F1- Score sebesar 3%, membuktikan BERT dapat meningkatkan performa SVM pada dataset terbatas.
MERDIANSAH DKK. (2024)	Analisis Sentimen Kendaraan Listrik	IndoBERT	Model yang dilatih dengan data IndoNLU memberikan kinerja yang lebih baik dan konsisten dalam memahami konteks teks berbahasa Indonesia.

RAO DKK. (2023)	Klasifikasi Disastrous Tweets	BERT Model	BERT berhasil mengklasifikasi tweet bencana secara akurat karena kemampuannya memahami konteks kata yang ambigu, mengungguli metode tradisional.
ALHARM & NAIM (2024)	Analisis Sentimen Bencana Alam	BERT dan LSTM (Deep Learning)	Model hibrid BERT-LSTM mencapai akurasi tertinggi (85.43%), melampaui SVM (76.80%) dan model BERT- FF (82.11%), menunjukkan kekuatan model hibrid dalam konteks bencana.

AMRIZA & SUPRIYADI (2021)	Komparasi ML & DL untuk Deteksi Emosi	Naïve Bayes, SVM, LSTM, CNN	Metode Deep Learning (LSTM, CNN, dll.) secara umum memiliki performa lebih baik daripada Machine Learning klasik karena mampu menangkap informasi kontekstual.

Penelitian mengenai analisis teks dari media sosial untuk berbagai tujuan, termasuk analisis sentimen dan klasifikasi bencana, telah banyak dilakukan. Tinjauan ini berfokus pada studi-studi yang menggunakan metode *machine learning* klasik, *deep learning*, hingga pendekatan hibrid yang relevan dengan penelitian ini.

Secara umum, metode *deep learning* telah menunjukkan keunggulan dibandingkan *machine learning* klasik untuk tugas pemrosesan bahasa alami (NLP). Sebuah studi komparasi oleh **Amriza dan Supriyadi (2021)** menemukan bahwa model *deep learning* seperti LSTM dan CNN memiliki performa akurasi yang lebih tinggi untuk deteksi emosi pada teks media sosial. Hal ini disebabkan kemampuan model *deep learning* untuk menangkap informasi kontekstual dan hubungan antar kata, sesuatu yang sulit dilakukan oleh model *machine learning* klasik yang seringkali memperlakukan kalimat hanya sebagai sekumpulan kata (*bag-of-words*).

Dalam konteks klasifikasi tweet bencana secara spesifik, keunggulan model berbasis *Transformer* seperti BERT semakin terlihat jelas. Rao dkk. (2023) dalam penelitiannya untuk mengklasifikasi *disastrous tweets* menunjukkan bahwa BERT secara signifikan mengungguli model tradisional seperti SVM yang menggunakan fitur TF-IDF. Penelitian tersebut menekankan pentingnya pemahaman konteks kata yang ambigu dalam situasi bencana, di mana BERT mampu membedakan makna kata "*accident*" sebagai

bencana atau kejadian biasa tergantung pada kalimatnya.

2.2 Landasan Teori

2.2.1 Data Media Sosial

Media sosial pada era revolusi industri digital mengalami perkembangan yang sangat pesat, baik dari segi jumlah pengguna maupun keragaman fungsinya. Salah satu platform yang paling banyak digunakan adalah **X (sebelumnya dikenal sebagai Twitter)**. Platform ini memiliki karakteristik unik berupa penyebaran informasi yang cepat, terbuka, dan berbasis teks singkat. Media sosial X sering dijadikan sebagai salah satu sumber utama dalam berbagai penelitian, khususnya dalam bidang **analisis kebencanaan**, karena mampu menyediakan data *real-time* yang berasal langsung dari masyarakat di lokasi kejadian.(Purna Chandrarao et al., 2022)

Pada saat terjadi bencana, pengguna media sosial cenderung membagikan informasi aktual mengenai situasi di lapangan, kondisi lingkungan, serta kebutuhan mendesak seperti makanan, obat-obatan, tempat tinggal, atau bantuan listrik. Informasi yang diunggah oleh masyarakat ini menjadi sumber data yang sangat berharga bagi lembaga kemanusiaan maupun peneliti, karena memberikan gambaran langsung tentang kondisi pascabencana tanpa harus menunggu laporan resmi dari instansi terkait.(Farzindar & Inkpen, 2015)

Dalam penelitian ini, data media sosial yang digunakan sebagai sampel difokuskan pada unggahan teks publik yang berkaitan dengan situasi pascabencana. Data tersebut menggunakan dataset sekunder untuk mengatasi keterbatasan akses API, serta dengan kata kunci tertentu seperti “butuh bantuan”, “listrik padam”, “pengungsian”, atau “kehilangan tempat tinggal”. Data yang berhasil dikumpulkan selanjutnya akan dianalisis untuk mengidentifikasi jenis kebutuhan mendesak korban pascabencana. Dengan demikian, media sosial X berperan penting sebagai sumber data dinamis yang dapat membantu pengembangan sistem prediktif dalam mendukung respons bencana yang lebih cepat dan efektif.

2.2.2 Platform X (Twitter)

Platform **X** (sebelumnya dikenal sebagai **Twitter**) merupakan salah satu media sosial berbasis mikroblogging yang memungkinkan penggunanya untuk berbagi dan berinteraksi melalui pesan singkat yang disebut **“tweet”**.(Farzindar & Inkpen, 2015)

Setiap tweet dapat berisi teks, foto, video, tautan, atau kombinasi dari semuanya. Platform ini dikenal dengan karakteristiknya yang cepat, real-time, dan terbuka, sehingga sering digunakan untuk menyebarkan informasi, berdiskusi, maupun mengomentari peristiwa yang sedang terjadi di dunia.(Purna Chandrarao et al., 2022)

Menurut Merdiansah, Siska, dan Azhari Ali Ridha (2021), X adalah platform yang memungkinkan teman, keluarga, dan rekan kerja untuk berinteraksi dan tetap terhubung melalui pertukaran pesan yang cepat dan sering. Pengguna dapat membagikan konten berupa foto, video, tautan, dan teks dalam posting yang dapat dilihat di profil mereka, dikirim ke pengikut, serta dapat dicari di dalam platform tersebut. Dengan sifatnya yang terbuka, X menjadi salah satu sumber data yang sangat potensial dalam penelitian berbasis analisis teks.(Merdiansah & Ali Ridha, 2024a)

Karakteristik Utama Platform X:

1 Mikroblogging *Real-Time*

X memungkinkan pengguna untuk mempublikasikan pesan singkat dengan batasan karakter tertentu (sebelumnya 280 karakter). Batasan ini mendorong pengguna untuk menulis pesan secara singkat, padat, dan langsung pada inti informasi. Sifat *real-time* menjadikan X sangat efektif dalam menyebarkan berita dan informasi terkini, terutama dalam situasi darurat seperti bencana alam.

2. Tagar (#Hashtag)

Fitur hashtag memungkinkan pengguna untuk mengelompokkan topik tertentu dan mempermudah pencarian informasi yang relevan. Misalnya, dalam konteks bencana, tagar seperti #banjir, #gempa, atau #longsor sering digunakan oleh masyarakat untuk berbagi informasi kondisi terkini, bantuan, atau permintaan pertolongan. Hashtag ini juga berfungsi sebagai *metadata* penting yang dapat membantu proses *data filtering* dalam penelitian berbasis teks.

3. Retweet dan Mention

Fitur retweet (RT) memungkinkan pengguna menyebarkan ulang tweet orang lain, sehingga memperluas jangkauan penyebaran informasi. Sementara itu, fitur mention (@username) digunakan untuk menandai atau menyebut

pengguna lain, yang meningkatkan interaksi antar pengguna. Kedua fitur ini membuat penyebaran informasi di X bersifat viral dan cepat menyebar di antara jaringan pengguna.

4. Keterbukaan Data (Open API0)

Salah satu keunggulan utama X dibandingkan platform media sosial lainnya adalah ketersediaan Application Programming Interface (API) yang memungkinkan peneliti dan pengembang mengakses data publik secara legal dan sistematis. Melalui API ini, peneliti dapat mengumpulkan tweet berdasarkan kata kunci, tagar, lokasi geografis, atau waktu tertentu, sehingga sangat mendukung kegiatan penelitian yang berbasis *data-driven*.

Karena sifatnya yang terbuka dan real-time, X banyak digunakan dalam penelitian yang berkaitan dengan analisis opini publik, deteksi emosi, pemantauan bencana, dan prediksi tren sosial. Menurut Aulia et al. (2021), data dari media sosial seperti X dapat mencerminkan persepsi dan perilaku masyarakat terhadap suatu peristiwa karena pengguna cenderung mengekspresikan pendapat mereka secara spontan.

Dalam konteks kebencanaan, X sering menjadi sumber utama informasi yang dikirim langsung oleh masyarakat di lokasi terdampak. Postingan seperti “butuh makanan”, “rumah roboh”, atau “jalan terputus” dapat dijadikan indikator awal untuk mendeteksi kebutuhan mendesak atau tingkat keparahan situasi pascabencana. Oleh karena itu, data dari X dapat diolah menggunakan metode *Natural Langguage Processing* dan *machine learning* untuk membangun model prediktif klasifikasi kebutuhan pascabencana.

Penelitian-penelitian sebelumnya juga menunjukkan bahwa data X efektif digunakan dalam analisis bencana. Misalnya, penelitian oleh Imran et al. (2016) menggunakan data Twitter untuk mengklasifikasikan pesan menjadi kategori seperti permintaan bantuan, informasi situasional, atau laporan kerusakan.

Dalam penelitian ini, X digunakan sebagai sumber data utama untuk mengumpulkan teks dari unggahan pengguna yang berkaitan dengan peristiwa pasca bencana seperti banjir, tanah longsor, dan gempa bumi. Data yang dikumpulkan berupa tweet publik dengan kata kunci atau tagar tertentu, yang kemudian dianotasi berdasarkan kategori kebutuhan seperti kebutuhan korban bencana seperti pangan, pelayanan

kesehatan, kebutuhan air bersih dan sanitasi, sandang, penampungan dan tempat hunian sementara, pelayanan psiko-sosial

Setiap tweet akan melalui tahapan pembersihan teks (text preprocessing), representasi teks menggunakan BERT, dan pemodelan klasifikasi berbasis Support Vector Machine (SVM) untuk menghasilkan sistem yang mampu memprediksi kebutuhan mendesak secara otomatis. Dengan pendekatan ini, diharapkan penelitian dapat berkontribusi dalam upaya percepatan respons bencana dan pengambilan keputusan berbasis data.

2.2.3 Kebutuhan Mendesak Korban Bencana

Kebutuhan mendesak korban bencana merupakan kebutuhan dasar yang harus segera dipenuhi pada masa tanggap darurat untuk menjaga keselamatan, kesehatan, dan kelangsungan hidup masyarakat terdampak. Dalam penyelenggaraan penanggulangan bencana, pemerintah daerah memiliki tanggung jawab untuk melaksanakan tanggap darurat melalui kaji cepat, penentuan tingkatan bencana, penyelamatan dan evakuasi, penanganan kelompok rentan, serta menjamin pemenuhan hak dasar masyarakat korban bencana. Dalam Peraturan Kepala Badan Nasional Penanggulangan Bencana Nomor 3 Tahun 2008, pemenuhan hak dasar tersebut mencakup enam jenis kebutuhan utama, yaitu pangan, pelayanan kesehatan, kebutuhan air bersih dan sanitasi, sandang, penampungan dan tempat hunian sementara, serta pelayanan psiko-sosial.

Secara konseptual, kebutuhan mendesak korban bencana dapat dipahami sebagai bentuk kebutuhan prioritas yang muncul segera setelah terjadinya bencana. Kebutuhan ini bersifat darurat karena berhubungan langsung dengan perlindungan jiwa, kondisi fisik, kesehatan, kenyamanan dasar, dan kestabilan psikologis korban. Oleh karena itu, identifikasi kebutuhan mendesak menjadi bagian penting dalam proses penanggulangan bencana, terutama untuk mendukung distribusi bantuan yang cepat, tepat sasaran, dan sesuai kondisi lapangan. Dalam konteks penelitian ini, klasifikasi kebutuhan mendesak dilakukan untuk mengenali jenis kebutuhan yang disampaikan korban atau masyarakat melalui teks media sosial, sehingga informasi yang tersebar dapat diolah menjadi masukan bagi pengambilan keputusan.

2.2.3.1 Pangan

Pangan adalah kebutuhan dasar berupa makanan dan minuman yang dibutuhkan

korban bencana untuk mempertahankan kondisi fisik dan energi selama masa darurat. Dalam situasi bencana, korban sering mengalami keterbatasan akses terhadap bahan makanan akibat kerusakan infrastruktur, gangguan distribusi logistik, atau kehilangan sumber daya rumah tangga. Oleh sebab itu, kebutuhan pangan menjadi salah satu prioritas utama dalam penyaluran bantuan. Pada penelitian ini, kategori pangan digunakan untuk mengelompokkan teks yang memuat permintaan atau informasi terkait makanan siap saji, bahan makanan pokok, susu, atau kebutuhan konsumsi lainnya.

2.2.3.2 Pelayanan Kesehatan

Pelayanan kesehatan adalah kebutuhan yang berkaitan dengan penanganan medis bagi korban bencana, baik untuk cedera fisik, penyakit, maupun kondisi kesehatan umum selama masa tanggap darurat. Bencana dapat menyebabkan luka, infeksi, kelelahan, hingga munculnya penyakit akibat lingkungan yang tidak sehat. Karena itu, pelayanan kesehatan mencakup kebutuhan seperti obat-obatan, tenaga medis, pemeriksaan kesehatan, layanan pertolongan pertama, dan fasilitas kesehatan darurat. Dalam penelitian ini, kategori ini digunakan untuk mengklasifikasikan teks yang berkaitan dengan kebutuhan pengobatan, perawatan medis, atau permintaan tenaga kesehatan.

2.2.3.3Kebutuhan Air Bersih dan Sanitasi

Kebutuhan air bersih dan sanitasi merupakan kebutuhan mendasar yang berkaitan dengan ketersediaan air layak pakai untuk minum, memasak, mandi, dan menjaga kebersihan lingkungan. Dalam kondisi pascabencana, akses terhadap air bersih sering terganggu akibat kerusakan sumber air, pencemaran, atau keterbatasan fasilitas sanitasi. Jika tidak ditangani, kondisi ini dapat meningkatkan risiko penyebaran penyakit. Oleh karena itu, kebutuhan air bersih dan sanitasi mencakup distribusi air bersih, toilet darurat, perlengkapan kebersihan, serta sarana sanitasi yang memadai. Dalam penelitian ini, kategori ini digunakan untuk mendeteksi teks yang menunjukkan kebutuhan air minum, sanitasi, kebersihan, atau fasilitas mandi dan cuci.

2.2.3.4 Sandang

Sandang adalah kebutuhan berupa pakaian dan perlengkapan pribadi yang diperlukan korban untuk menjaga kenyamanan, kesehatan, dan kelayakan hidup selama masa darurat. Korban bencana sering kehilangan pakaian, selimut, alas tidur, dan perlengkapan pribadi akibat kerusakan rumah atau proses evakuasi mendadak. Karena

itu, bantuan sandang menjadi penting terutama bagi kelompok rentan seperti anak-anak, lansia, dan perempuan. Dalam penelitian ini, kategori sandang digunakan untuk mengelompokkan teks yang berisi kebutuhan pakaian, selimut, perlengkapan bayi, atau perlengkapan pribadi lainnya.

2.2.3.5 Penampungan dan Tempat Hunian Sementara

Penampungan dan tempat hunian sementara merupakan kebutuhan akan tempat tinggal darurat yang aman bagi korban yang kehilangan rumah atau tidak dapat kembali ke tempat tinggalnya. Bencana dapat menyebabkan rumah rusak, terendam, atau tidak layak huni, sehingga korban membutuhkan lokasi pengungsian yang dapat memberikan perlindungan sementara. Kebutuhan ini mencakup tenda pengungsian, posko darurat, tempat istirahat sementara, serta fasilitas dasar di lokasi penampungan. Dalam penelitian ini, kategori ini digunakan untuk mengidentifikasi teks yang memuat informasi atau permintaan terkait tempat pengungsian, tenda, lokasi evakuasi, atau hunian sementara.

2.2.3.6 Pelayanan Psiko-sosial

Pelayanan psiko-sosial adalah kebutuhan yang berkaitan dengan dukungan psikologis dan sosial bagi korban bencana untuk membantu mereka menghadapi trauma, stres, kecemasan, dan tekanan emosional akibat bencana. Dampak bencana tidak hanya bersifat fisik, tetapi juga dapat memengaruhi kondisi mental korban, terutama anak-anak, lansia, dan individu yang kehilangan keluarga atau tempat tinggal. Oleh karena itu, pelayanan psiko-sosial dibutuhkan untuk membantu pemulihan emosional dan menjaga stabilitas psikologis korban selama masa tanggap darurat. Dalam penelitian ini, kategori ini digunakan untuk mengelompokkan teks yang menunjukkan kebutuhan pendampingan psikologis, dukungan mental, atau bantuan sosial bagi korban terdampak.

2.2.3.7 Klasifikasi Kebutuhan Mendesak dalam Penelitian

Dalam penelitian ini, enam kebutuhan mendesak tersebut dijadikan sebagai kelas kategori dalam proses klasifikasi teks. Setiap teks dari media sosial dianalisis untuk menentukan apakah isi informasinya berkaitan dengan salah satu dari enam kebutuhan dasar korban bencana. Pendekatan ini digunakan agar data teks yang awalnya tidak terstruktur dapat diubah menjadi informasi yang terkelompok secara sistematis. Dengan demikian, hasil klasifikasi dapat membantu proses identifikasi kebutuhan korban secara lebih cepat dan mendukung distribusi bantuan yang lebih tepat sasaran. Dasar

pengelompokan ini mengacu pada klasifikasi kebutuhan dasar korban bencana yang telah disebutkan dalam Peraturan Kepala BNPB Nomor 3 Tahun 2008.

2.2.4 Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada interaksi antara komputer dan bahasa manusia. NLP memungkinkan sistem komputer untuk memahami, menganalisis, dan memproses teks atau bahasa alami sehingga dapat digunakan untuk berbagai keperluan seperti klasifikasi teks, analisis sentimen, dan ekstraksi informasi.

NLP menggabungkan konsep linguistik dan pembelajaran mesin untuk memungkinkan komputer memahami makna dari teks yang diproses. Dalam perkembangannya, NLP telah banyak diterapkan dalam berbagai bidang seperti sistem rekomendasi, chatbot, deteksi spam, dan klasifikasi dokumen.

Pada penelitian ini, NLP digunakan untuk mengolah data teks yang berasal dari media sosial melalui tahapan *preprocessing* dan pembentukan representasi fitur menggunakan model BERT sebelum dilakukan proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM).

2.2.5 Text Preprocessing

Text preprocessing merupakan tahap awal yang sangat penting dalam sistem klasifikasi teks, khususnya pada penelitian yang menggunakan data media sosial untuk mengidentifikasi kebutuhan korban bencana. Data teks yang diperoleh dari platform seperti X (Twitter) atau TikTok umumnya masih berupa teks mentah yang tidak terstruktur, mengandung banyak noise, singkatan, simbol, tautan, serta kata-kata yang tidak relevan dengan tujuan klasifikasi. Kondisi tersebut dapat mengganggu proses ekstraksi fitur dan menurunkan kinerja model klasifikasi. Oleh karena itu, text preprocessing dilakukan untuk membersihkan, menyederhanakan, dan menyiapkan data agar informasi yang terkandung di dalam teks lebih mudah dipahami oleh model machine learning maupun deep learning.

Dalam penelitian klasifikasi kebutuhan korban bencana, tahapan preprocessing bertujuan agar teks lebih fokus pada kata-kata yang merepresentasikan kebutuhan mendesak, seperti makanan, obat, air bersih, selimut, atau tempat pengungsian. Dengan preprocessing yang baik, model dapat lebih

mudah mengenali pola bahasa yang berkaitan dengan kategori kebutuhan bantuan, sehingga proses klasifikasi menjadi lebih akurat.Berikut adalah tahapan-tahapan utama dalam *text preprocessing*:

1. *Cleaning* (Pembersihan Teks)

   Tahap cleaning merupakan proses pembersihan elemen-elemen yang tidak diperlukan dari teks mentah. Pada data media sosial, teks sering kali mengandung URL, mention, hashtag, angka, emotikon, karakter berulang, tanda baca berlebihan, dan simbol lain yang tidak memberikan kontribusi langsung terhadap makna kebutuhan bantuan. Dalam konteks klasifikasi kebutuhan korban bencana, pembersihan teks bertujuan untuk mempertahankan kata-kata inti yang menunjukkan jenis kebutuhan, sehingga isi pesan menjadi lebih jelas dan relevan untuk dianalisis.Contoh:

Tabel 2.2 Pembersihan Teks

Kalimat awal	Hasil
“Butuh makanan cepatttttttttt, udah laper banget #korbanbencana https://. ”	“Butuh makanan cepat”

2. *Tokenization*

Tokenization adalah proses memecah kalimat atau dokumen menjadi unit-unit kecil yang disebut token, umumnya berupa kata. Tahap ini penting karena model klasifikasi tidak memproses teks sebagai kalimat utuh, melainkan sebagai kumpulan token yang akan diubah menjadi representasi numerik. Pada penelitian ini, tokenisasi membantu memisahkan kata-kata penting dalam laporan atau keluhan masyarakat, seperti *butuh*, *makanan*, *obat*, *air*, *selimut*, atau *pengungsian*, sehingga setiap kata yang berpotensi menjadi penanda kebutuhan dapat dikenali secara lebih terstruktur. Dalam konteks klasifikasi kebutuhan mendesak korban bencana, tokenisasi berperan dalam mengidentifikasi unsur-unsur linguistik yang menjadi dasar untuk menentukan kategori bantuan yang dibutuhkan.Contoh:

Tabel 2.3 Tokenisasi

Kalimat awal	Hasil
"Butuh makanan cepat"	["butuh", "makanan", "cepat"]

3. *Stopword Removal*

Stopword removal merupakan proses penghapusan kata-kata umum yang sering muncul dalam kalimat tetapi memiliki kontribusi makna yang rendah terhadap proses klasifikasi. Kata-kata seperti *yang*, *dan*, *di*, *ke*, *dari*, atau *itu* umumnya tidak secara langsung menunjukkan jenis kebutuhan bantuan. Dalam penelitian klasifikasi kebutuhan korban bencana, penghapusan stopword bertujuan agar model lebih fokus pada kata-kata inti yang mengandung informasi penting, seperti *makanan*, *obat*, *air bersih*, *baju*, atau *trauma*. Dengan demikian, noise dalam teks dapat dikurangi dan representasi teks menjadi lebih relevan terhadap konteks kebutuhan mendesak.Contoh:

Tabel 2.4 *Stopword Removal*

Kalimat awal	Hasil
"Yang di sana butuh makanan segera"	["butuh", "makanan", "segera"]

4. *Stemming*

Stemming adalah proses mengubah kata berimbuhan menjadi bentuk dasarnya. Tahap ini dilakukan untuk menyatukan berbagai variasi kata yang memiliki akar makna yang sama agar model dapat mengenali pola dengan lebih konsisten. Dalam data kebencanaan, satu jenis kebutuhan dapat dinyatakan dalam beberapa bentuk kata yang berbeda, misalnya *makanan*, *memakan*, atau *dimakan*, yang memiliki akar kata *makan*. Contoh lain, kata *membutuhkan*, *dibutuhkan*, dan *kebutuhan* dapat direduksi ke bentuk dasar *butuh*. Dalam klasifikasi kebutuhan mendesak korban bencana, stemming membantu menyederhanakan variasi bahasa yang digunakan masyarakat, sehingga model lebih mudah mengelompokkan teks ke kategori yang sesuai berdasarkan makna inti dari kata-kata tersebut.Contoh:

Tabel 2.5 *Stemming* atau *Lemmatization*

Kalimat awal	Hasil
[“memakan”, “makanan”, “dimakan”]	[“makan”, “makan”, “makan”]

2.2.6 Representasi Teks

Representasi teks merupakan proses penting dalam *text mining* dan *natural language processing (NLP)* untuk mengubah data teks mentah menjadi bentuk numerik agar dapat diproses oleh algoritma pembelajaran mesin atau *deep learning(Vijay Gaikwad, 2014)*. Karena komputer tidak dapat memahami teks secara langsung, setiap kata, frasa, atau dokumen perlu diubah menjadi vektor angka yang menggambarkan makna semantik dan struktur sintaksisnya.(*Jurnal Zonasi Marwika Rifattul Iffa - MARWIKA RIFATTUL IFFA Teknik Informatika*, n.d.)

Berbagai pendekatan telah dikembangkan untuk merepresentasikan teks, mulai dari metode sederhana berbasis frekuensi kata hingga model kontekstual berbasis *transformer*. Berikut beberapa teknik representasi teks yang umum digunakan:

2.2.5.1 Bag of Words (BoW)

Metode *Bag of Words* (BoW) merupakan salah satu pendekatan paling dasar dalam representasi teks. Konsep dasarnya adalah merepresentasikan setiap dokumen sebagai sekumpulan kata tanpa memperhatikan urutan atau konteks kemunculannya. Setiap kata yang muncul dalam korpus akan dianggap sebagai fitur, dan nilai pada fitur tersebut merepresentasikan frekuensi kemunculan kata tersebut dalam dokumen.

Sebagai contoh, jika terdapat dua kalimat "Butuh makanan cepat" dan "Butuh bantuan medis", maka *vocabulary*-nya terdiri dari kata: [butuh, makanan, cepat, bantuan, medis]. Setiap kalimat kemudian direpresentasikan dalam bentuk vektor berdimensi lima berdasarkan frekuensi kemunculan kata.

Kelebihan dari BoW adalah kesederhanaannya dan kemudahan implementasinya. Namun, metode ini memiliki keterbatasan karena tidak mempertimbangkan konteks dan urutan kata, sehingga dua kalimat dengan struktur berbeda tetapi makna serupa dapat dianggap tidak sama secara matematis.

2.2.5.2 TF-IDF (Term Frequency – Inverse Document Frequency)

Metode TF-IDF merupakan pengembangan dari BoW yang memberikan bobot lebih bermakna terhadap kata dalam dokumen. TF-IDF tidak hanya memperhatikan frekuensi kemunculan kata dalam satu dokumen (term frequency), tetapi juga mempertimbangkan seberapa jarang kata tersebut muncul di

seluruh kumpulan dokumen (inverse document frequency).

Secara umum, nilai TF-IDF dari suatu kata akan tinggi apabila kata tersebut sering muncul dalam dokumen tertentu, tetapi jarang muncul dalam dokumen lainnya. Dengan demikian, TF-IDF mampu memberikan bobot lebih besar pada kata-kata yang dianggap spesifik dan penting, serta mengurangi pengaruh kata umum seperti "dan", "yang", atau "itu".

Pendekatan ini sering digunakan dalam sistem information retrieval dan klasifikasi teks klasik seperti Naïve Bayes, Support Vector Machine (SVM), atau Logistic Regression, karena mampu menghasilkan fitur yang cukup representatif dengan komputasi yang efisien.

2.2.5.3 Word Embedding

Metode *Word Embedding* merupakan pendekatan representasi teks yang lebih canggih karena mampu menangkap makna semantik antar kata. Alih-alih menggunakan representasi berbasis frekuensi, *word embedding* merepresentasikan setiap kata dalam bentuk vektor kontinu berdimensi rendah, di mana kedekatan antar vektor mencerminkan kesamaan makna antar kata.

Misalnya, kata "dokter" dan "perawat" akan memiliki vektor yang berdekatan karena keduanya sering muncul dalam konteks yang mirip, sementara kata "dokter" dan "makanan" akan memiliki jarak vektor yang lebih jauh.

Beberapa metode populer yang digunakan untuk menghasilkan *word embedding,* antara lain:

a. Word2Vec, yang menggunakan arsitektur *Continuous Bag of Words (CBOW)* dan *Skip-gram* untuk mempelajari hubungan antar kata.
b. GloVe (*Global Vectors for Word Representation*), yang memanfaatkan informasi statistik global dari korpus untuk membangun representasi kata.
c. FastText, yang memperluas Word2Vec dengan mempertimbangkan struktur morfologis kata melalui *subword embedding*.

Menurut Rona Nisa Sofia Amriza dan Didi Supriyadi (2021), word embedding bertugas memetakan setiap kata ke dalam ruang vektor berdimensi tinggi yang mampu menangkap informasi semantik dan sintaksis antar kata. Setiap kolom pada matriks representasi menyimpan vektor kata yang mewakili makna kata tersebut dalam

konteks tertentu.

2.2.5.4 Transformer Embedding (BERT)

Perkembangan terbaru dalam representasi teks adalah munculnya model berbasis Transformer, salah satunya BERT (Bidirectional Encoder Representations from Transformers). BERT menghasilkan representasi kata yang bersifat kontekstual, artinya setiap kata direpresentasikan berdasarkan kata-kata di sekitarnya, bukan secara terpisah seperti pada metode embedding tradisional.(Purna Chandrarao et al., 2022)

BERT menggunakan mekanisme self-attention yang memungkinkan model memahami hubungan antar kata dalam kalimat secara dua arah (bidirectional), baik dari kiri ke kanan maupun sebaliknya. Dengan demikian, model dapat menangkap makna kata yang berbeda tergantung pada konteks penggunaannya.

Sebagai contoh, kata “bank” pada kalimat “Saya menabung di bank” dan “Bank tanah longsor di sungai” memiliki makna yang berbeda, dan BERT mampu membedakan kedua konteks tersebut.(Wu et al., n.d.)

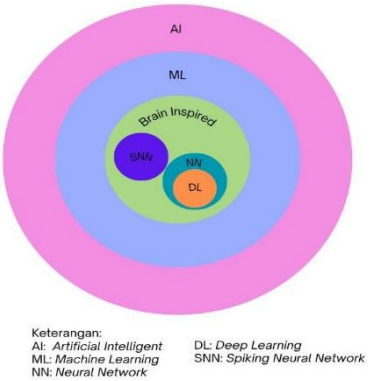
Keunggulan BERT dibandingkan metode representasi teks lain terletak pada kemampuannya menangkap konteks secara mendalam, sehingga embedding yang dihasilkan lebih akurat untuk berbagai tugas NLP seperti klasifikasi teks, analisis sentimen, ekstraksi entitas, dan deteksi topik.

Dalam penelitian ini, BERT digunakan sebagai metode representasi teks untuk mendukung proses klasifikasi kebutuhan mendesak pascabencana berdasarkan teks media sosial. Representasi embedding yang dihasilkan BERT selanjutnya akan digunakan sebagai input untuk algoritma klasifikasi, sehingga model dapat memprediksi kategori kebutuhan secara lebih tepat.

2.2.7 Machine Learning

Machine Learning adalah sub-bidang dari kecerdasan buatan (AI) yang berfokus pada pengembangan algoritma yang memungkinkan komputer "belajar" dari data tanpa diprogram secara eksplisit. ML memiliki berbagai aplikasi, termasuk klasifikasi, regresi, dan pengelompokan. *Support Vector Machine* (SVM), yang disinggung oleh Ifa (2023), adalah salah satu algoritma ML klasik yang kuat untuk tugas klasifikasi dengan menemukan *hyperplane* optimal yang memisahkan kelas-kelas data. Amriza dan Supriyadi (2023) juga membahas komparasi metode ML dan DL untuk deteksi emosi

pada teks.(Farman Ali Baig et al., 2024b)



Gambar 2. 1 Struktur Kecerdasan Buatan

Machine Learning (Pembelajaran Mesin) merupakan salah satu cabang dari kecerdasan buatan (*Artificial Intelligence / AI*) yang berfokus pada pengembangan algoritma dan model matematis untuk memungkinkan sistem komputer belajar dari data tanpa perlu diprogram secara eksplisit. Konsep dasar Machine Learning adalah penggunaan data sampel atau *training data* untuk membangun model yang dapat mengenali pola, membuat prediksi, atau mengambil keputusan secara otomatis.

Dalam Machine Learning, proses pembelajaran dilakukan dengan memanfaatkan hubungan antara fitur-fitur input dan target output yang tersedia pada data berlabel. Algoritma belajar untuk mengekstraksi pola dan aturan dari data tersebut, sehingga ketika diberikan data baru yang belum pernah ditemui sebelumnya, model mampu menghasilkan prediksi atau keputusan yang sesuai dengan pola yang telah dipelajari. Dengan kata lain, Machine Learning memungkinkan komputer untuk menyimpulkan aturan atau strategi dari pengalaman data secara sistematis dan adaptif.

Machine Learning dapat dikategorikan ke dalam beberapa jenis utama berdasarkan metode pembelajaran dan sifat data yang digunakan:

1. *Supervised Learning*

   Pada pendekatan ini, model dilatih menggunakan data berlabel, yaitu setiap sampel data memiliki jawaban atau kategori yang diketahui. Algoritma belajar memetakan input ke output dengan meminimalkan kesalahan prediksi. Contoh penerapan supervised learning antara lain klasifikasi teks, prediksi harga saham, deteksi penyakit, dan sistem rekomendasi.

2. *Unsupervised Learning*

   Berbeda dengan supervised learning, metode ini bekerja dengan data yang tidak berlabel. Tujuan utamanya adalah menemukan pola, struktur, atau hubungan tersembunyi dalam data. Teknik unsupervised learning yang umum digunakan meliputi *clustering*, *dimensionality reduction*, dan deteksi anomali.

3. *Reinforcement Learning*

   Pendekatan ini mengandalkan interaksi agen dengan lingkungan dan memanfaatkan umpan balik berupa *reward* atau *penalty* untuk mengoptimalkan perilaku atau strategi tertentu.

2.2.8 Deep Learning

Deep learning adalah kecerdasan buatan yang meniru cara manusia memperoleh jenis pengetahuan tertentu. (RonaNisaSofiaAmriza, DidiSupriyadi). Deep Learning merupakan cabang dari Machine Learning yang menggunakan arsitektur jaringan saraf tiruan (neural network) kelebihan dari Deep Learning berada pada kamampuanya menangkap pola non-Linear dan interaksi yang kompleks antar variable dalam data.(Nisa Sofia Amriza et al., n.d.)

Sejak tahun 1950-an, cabang kecerdasan buatan (*Artificial Intelligence* / AI) yang dikenal dengan *Machine Learning* (ML) telah mengalami perkembangan signifikan dan diterapkan di berbagai bidang. Salah satu bentuk implementasi Machine Learning adalah *Neural Network* (NN), sedangkan *Deep Learning* (DL) merupakan pengembangan lanjutan dari Neural Network dengan arsitektur yang lebih kompleks dan mendalam (García-Tapia-Mateo et al., n.d.)

Deep Learning memanfaatkan arsitektur pembelajaran berlapis (*deep architecture of learning*) atau pendekatan *hierarchical learning*, di mana proses pembelajaran dilakukan melalui estimasi parameter model agar dapat menyelesaikan tugas tertentu secara optimal. Setiap lapisan dalam *Deep Learning* melakukan pemrosesan nonlinier bertahap yang memungkinkan model mengekstraksi fitur dan melakukan klasifikasi pola dengan lebih efektif. Jumlah lapisan yang bervariasi memungkinkan tingkat abstraksi yang berbeda, menjadikan *Deep Learning* unggul dalam menangkap representasi kompleks dari data (Wu et al., n.d.)

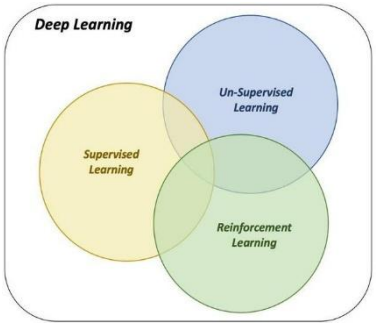
Secara umum, *Deep Learning* dapat dianggap sebagai kelas algoritma *Machine*

Learning yang menggunakan beberapa lapisan nonlinier secara berurutan untuk melakukan *feature extraction* dan *transformation*. Setiap lapisan menggunakan hasil dari lapisan sebelumnya sebagai masukan. Pendekatan ini dapat diterapkan baik dalam *supervised learning* maupun *unsupervised learning*, tergantung pada jenis data dan tujuan analisis (García- Tapia-Mateo et al., n.d.)

Lebih lanjut, metode *Deep Learning* dapat diklasifikasikan ke dalam beberapa kategori utama, yaitu:

1) Deep Supervised Learning – menggunakan data berlabel (labeled data) dalam proses pelatihan. Contoh populer dari kategori ini meliputi Deep Neural Network (DNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), dan Gated Recurrent Unit (GRU).
2) Deep Semi-Supervised Learning – menggunakan sebagian data berlabel (partially labeled data). Beberapa model seperti Generative Adversarial Networks (GAN), Deep Reinforcement Learning (DRL), serta varian RNN, LSTM, dan GRU juga dapat memanfaatkan pendekatan ini.
3) Deep Unsupervised Learning – memanfaatkan data tanpa label (unlabeled data), dengan contoh model seperti Autoencoder (AE), Restricted Boltzmann Machine (RBM), dan generasi baru dari GAN.
4) Deep Reinforcement Learning (DRL) – diterapkan pada lingkungan yang tidak diketahui (unknown environments) dan menjadi populer sejak diperkenalkan oleh Google DeepMind pada tahun 2013 (Diponegoro et al., 2023).

Dengan demikian, DL menjadi salah satu pendekatan yang paling fleksibel dalam domain pembelajaran mesin karena kemampuannya untuk menyesuaikan diri dengan berbagai jenis data dan tujuan pemodelan, baik dalam konteks klasifikasi, prediksi, maupun analisis pola.(García-Tapia-Mateo et al., n.d.)



Gambar 2. 2 Deep Learning

2.2.8.1 Deep Learning untuk NLP

Deep Learning adalah pendekatan yang meniru cara kerja jaringan saraf otak manusia. Dalam NLP, beberapa arsitektur popular:

1) RNN (Recurrent Neural Network): menangkap urutan kata
2) LSTM (Long Short – Term Memory): mampu mengatasi masalah long term 31ependency
3) Transformer: memanfaatkan mekanisme self-attention, lebih efisen dan unggul dalam memahami konteks Panjang.

2.2.9 Bidirectional Encoder Representations from Transformers (BERT)

BERT adalah model bahasa yang dikembangkan oleh Google AI yang menggunakan arsitektur *Transformer Encoder*. Nama BERT merupakan singkatan dari Bidirectional Encoder Representations from Transformers. Inovasi utamanya adalah kemampuannya untuk memahami konteks sebuah kata dari dua arah (kiri dan kanan) secara bersamaan, yang disebut sebagai *bidirectionality*. Kemampuan ini memungkinkan BERT menangkap informasi semantik dan hubungan antar kata dalam sebuah kalimat secara lebih mendalam dibandingkan model-model sebelumnya yang hanya memproses teks secara sekuensial (satu arah). (Wu et al., n.d.)

Arsitektur BERT sepenuhnya berbasis pada mekanisme *Transformer* yang mengandalkan *self-attention* untuk memahami hubungan kontekstual antara semua kata dalam sebuah teks. Tidak seperti model lain, BERT hanya menggunakan bagian *encoder* dari arsitektur Transformer. (Merdiansah & Ali Ridha, 2024b)

Ada dua ukuran model utama yang tersedia:

1) BERT-Base: Terdiri dari 12 lapisan *transformer*, 768 unit tersembunyi, dan 110 juta parameter.

2) BERT-Large: Lebih besar dengan 24 lapisan *transformer*, 1024 unit tersembunyi, dan 340 juta parameter.

Untuk bahasa Indonesia, telah dikembangkan model spesifik bernama IndoBERT yang dilatih menggunakan dataset besar berbahasa Indonesia.

1. Proses *Pre-training (Pra-pelatihan) BERT*

Keunggulan utama BERT berasal dari proses pra-pelatihannya pada data teks dalam jumlah masif (misalnya dari Wikipedia dan BookCorpus) tanpa memerlukan label. Proses ini bertujuan agar model dapat "memahami" bahasa sebelum diterapkan pada tugas spesifik. Ada dua tugas utama dalam pra-pelatihan BERT:

1) Masked Language Model (MLM): Dalam tugas ini, beberapa kata dalam sebuah kalimat secara acak "ditutupi" atau diganti dengan token [MASK]. Model kemudian dilatih untuk memprediksi kata asli yang ditutupi tersebut berdasarkan konteks dari kata-kata di sekitarnya. Ini memaksa model untuk belajar representasi kontekstual yang mendalam.
2) Next Sentence Prediction (NSP): Model diberi dua kalimat (Kalimat A dan Kalimat B) dan dilatih untuk memprediksi apakah Kalimat B adalah kalimat yang sebenarnya mengikuti Kalimat A dalam teks asli atau hanya kalimat acak. Tugas ini membantu model memahami hubungan antar kalimat.

2. Klasifikasi Sentimen dan Tugas NLP Lainnya

Setelah melalui tahap pra-pelatihan, BERT dapat diadaptasi untuk berbagai tugas *Natural Language Processing* (NLP) spesifik melalui proses yang disebut *fine-tuning* (penyesuaian). Dalam proses ini, lapisan output tambahan yang sesuai dengan tugas baru ditambahkan ke model BERT, dan seluruh model dilatih kembali dalam jumlah iterasi yang lebih sedikit menggunakan dataset berlabel yang lebih kecil.

Dalam konteks analisis sentimen, BERT terbukti sangat efektif. Model ini mampu menghasilkan representasi teks (disebut *embedding*) yang bersifat kontekstual. Artinya, makna sebuah kata direpresentasikan secara berbeda tergantung pada kalimat tempat kata itu digunakan. Contohnya, kata "bank" akan memiliki representasi yang berbeda dalam kalimat "Saya menabung di bank" dan

"Tebing itu longsor membentuk bank tanah".

Kemampuan kontekstual ini memberikan keunggulan signifikan dibandingkan metode representasi teks tradisional seperti TF-IDF atau Word2Vec yang bersifat statis.

3. Kombinasi BERT dengan Model Lain (Hibrida Model)

Penelitian menunjukkan bahwa menggabungkan BERT dengan algoritma *machine learning* klasik seperti Support Vector Machine (SVM) dapat menghasilkan performa yang sangat baik, terutama pada dataset yang terbatas. Dalam pendekatan *hibrida* ini:

1) BERT bertugas sebagai pengekstrak fitur (*feature extractor*). Ia mengubah teks mentah menjadi vektor numerik (*embedding*) yang kaya akan informasi kontekstual dan semantik.
2) SVM kemudian bertindak sebagai pengklasifikasi (*classifier*). Algoritma ini menerima input berupa vektor dari BERT dan bertugas menemukan *hyperplane* optimal untuk memisahkan data ke dalam kategori sentimen yang berbeda (misalnya, positif, negatif, netral).

Kombinasi ini terbukti mampu meningkatkan kinerja klasifikasi secara signifikan karena SVM dapat bekerja secara efektif dengan fitur berkualitas tinggi yang disediakan oleh BERT. Pendekatan ini menawarkan efisiensi komputasi yang lebih baik dibandingkan menggunakan arsitektur *deep learning* yang lebih kompleks, namun tetap mempertahankan performa yang kompetitif

2.2.10 Feature Extraction (EKSTRAKSI FITUR)

Setelah teks dibersihkan melalui proses preprocessing, tahap berikutnya adalah mengubah teks menjadi bentuk representasi numerik agar dapat dipahami oleh komputer. Komputer tidak dapat memproses teks secara langsung, sehingga setiap kata atau kalimat perlu direpresentasikan ke dalam vektor angka.

Dalam penelitian ini digunakan representasi berbasis BERT (Bidirectional Encoder Representations from Transformers). BERT merupakan model *transformer-based embedding* yang menghasilkan representasi kontekstual, artinya setiap kata direpresentasikan berdasarkan konteks kalimat di sekitarnya. Pendekatan ini lebih unggul dibandingkan metode tradisional seperti *Bag of Words* (BoW) atau *TF-IDF*, karena

mampu memahami hubungan semantik dan sintaksis antar kata dalam kalimat.

Proses ekstraksi fitur dengan BERT melibatkan tahapan *tokenization* khusus berbasis *WordPiece*, di mana setiap kata diubah menjadi token sub-kata, kemudian diproses melalui lapisan *transformer encoder* untuk menghasilkan vektor representasi berdimensi tinggi. Hasil dari tahap ini berupa *embedding vector* yang merepresentasikan makna kontekstual setiap teks.

2.2. 11 Konsep Dasar Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma klasifikasi yang bekerja dengan mencari batas pemisah (*decision boundary*) optimal antar kelas. Berbeda dengan metode klasifikasi biasa yang hanya mencari pemisah, SVM berfokus pada pencarian pemisah dengan margin maksimum, sehingga memiliki kemampuan generalisasi yang baik terhadap data baru. Dalam SVM terdapat beberapa komponen utama, yaitu:

1. *Hyperplane*

   Hyperplane adalah batas pemisah antar kelas dalam ruang fitur. Pada data dua dimensi, hyperplane berbentuk garis lurus. Pada data tiga dimensi, berbentuk bidang. Pada dimensi yang lebih tinggi, disebut sebagai hyperplane dalam ruang berdimensi -n. Secara matematis, hyperplane dapat dituliskan sebagai:

   $$w \cdot x + b = 0 \qquad (2.1)$$

   di mana:

   w adalah vektor bobot

   x adalah vektor fitur

   b adalah bias

   Hyperplane inilah yang berfungsi memisahkan data ke dalam dua kelas yang berbeda.

2. *Margin*

   Margin adalah jarak antara hyperplane dengan titik data terdekat dari masing-masing kelas.

   SVM tidak hanya mencari hyperplane yang memisahkan data, tetapi mencari hyperplane dengan margin terbesar. Semakin besar margin, maka semakin baik kemampuan model dalam melakukan generalisasi terhadap data

yang belum pernah dilihat sebelumnya.

Konsep margin maksimum inilah yang menjadi ciri khas utama SVM dibandingkan algoritma klasifikasi lainnya.

3. *Support Vectors*

Support vectors adalah titik data yang berada paling dekat dengan hyperplane. Titik-titik inilah yang:

a. Menentukan posisi hyperplane
b. Menentukan lebar margin
c. Paling berpengaruh dalam proses pelatihan model

Menariknya, hanya sebagian kecil data (support vectors) yang benar-benar menentukan model akhir. Data lain yang berada jauh dari hyperplane tidak terlalu memengaruhi posisi batas keputusan. Hal ini membuat SVM relatif efisien dan stabil terhadap perubahan kecil pada data.

4. *Kernel Trick*

Dalam praktiknya, tidak semua data dapat dipisahkan secara linier. Untuk mengatasi hal tersebut, SVM menggunakan pendekatan yang disebut kernel trick.

Kernel trick memungkinkan data dipetakan ke ruang berdimensi lebih tinggi sehingga data yang sebelumnya tidak terpisahkan secara linier menjadi dapat dipisahkan oleh hyperplane.

Beberapa jenis kernel yang umum digunakan adalah:

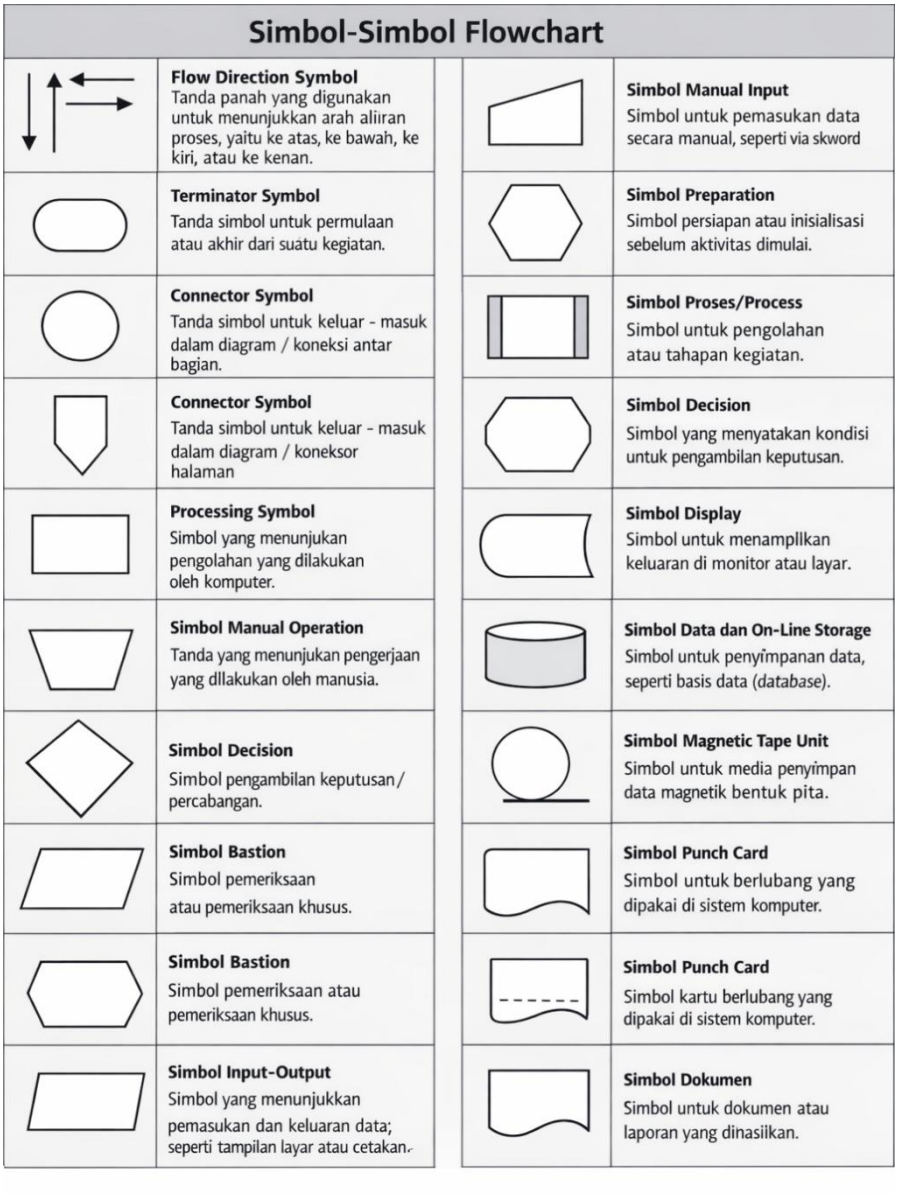
1) Linear Kernel
2) Polynomial Kernel
3) Radial Basis Function (RBF)

Dalam penelitian ini digunakan kernel linear, karena representasi fitur yang dihasilkan oleh IndoBERT telah berada pada ruang berdimensi tinggi dan kaya informasi semantik, sehingga pemisahan linier sudah memadai.

2.2.12 Flowchart

Flowchart atau diagram alir merupakan representasi visual yang digunakan

untuk menggambarkan urutan langkah, aliran proses, serta hubungan antar tahapan dalam suatu sistem atau penelitian. Flowchart disusun menggunakan simbol-simbol standar yang masing-masing memiliki fungsi tertentu, seperti simbol terminator untuk menandai awal dan akhir proses, simbol process untuk menunjukkan tahapan pengerjaan, simbol input/output untuk menunjukkan data yang masuk atau keluar, serta garis alir untuk menunjukkan arah urutan proses. Dengan adanya flowchart, suatu proses dapat dipahami secara lebih sistematis, terstruktur, dan mudah dianalisis.



Gambar 2. 3 Teori Flowchart

Flowchart menggunakan berbagai simbol standar untuk menggambarkan alur proses dalam suatu sistem atau program. Setiap simbol memiliki fungsi tertentu untuk menunjukkan jenis aktivitas yang dilakukan pada suatu tahap proses. Penggunaan simbol-simbol ini bertujuan untuk mempermudah pemahaman terhadap alur kerja suatu sistem secara visual.

Beberapa simbol flowchart yang umum digunakan antara lain sebagai berikut:

1. **Flow Direction Symbol (Panah Alur)**

   Simbol panah digunakan untuk menunjukkan arah aliran proses dari satu langkah ke langkah berikutnya. Panah ini menghubungkan setiap simbol sehingga membentuk urutan proses yang jelas.
2. **Terminator Symbol**

   Simbol terminator digunakan untuk menunjukkan awal dan akhir suatu proses dalam flowchart. Biasanya simbol ini digunakan pada bagian **mulai (start)** dan **selesai (end)**.
3. **Process Symbol**

   Simbol proses berbentuk persegi panjang digunakan untuk menunjukkan langkah atau aktivitas yang dilakukan dalam suatu sistem atau program.
4. **Input/Output Symbol**

   Simbol ini digunakan untuk menunjukkan proses pemasukan data (input) atau keluaran data (output) dalam sistem.
5. **Decision Symbol**

   Simbol decision berbentuk belah ketupat digunakan untuk menunjukkan proses pengambilan keputusan atau percabangan berdasarkan kondisi tertentu.
6. **Connector Symbol**

   Simbol connector digunakan untuk menghubungkan bagian flowchart yang terpisah, baik dalam halaman yang sama maupun halaman yang berbeda.
7. **Manual Input Symbol**

   Simbol ini digunakan untuk menunjukkan proses input data secara manual oleh pengguna, misalnya melalui keyboard.

8. **Document Symbol**

   Simbol dokumen digunakan untuk menunjukkan dokumen atau laporan yang dihasilkan dari suatu proses.

9. **Data Storage Symbol**

   Simbol ini digunakan untuk menunjukkan tempat penyimpanan data seperti database atau media penyimpanan lainnya.

Dengan menggunakan simbol-simbol tersebut, flowchart dapat menggambarkan alur proses suatu sistem secara terstruktur sehingga memudahkan dalam memahami tahapan kerja sistem yang dirancang.

Dalam konteks penelitian, flowchart berfungsi sebagai alat bantu untuk memvisualisasikan tahapan penelitian secara ringkas dan logis. Melalui flowchart, peneliti dapat menunjukkan urutan proses mulai dari pengumpulan data, prapemrosesan, ekstraksi fitur, pembagian dataset, pelatihan model, pengujian, hingga evaluasi hasil. Penyajian dalam bentuk diagram alir juga memudahkan pembaca dalam memahami alur penelitian tanpa harus membaca seluruh uraian teknis secara rinci terlebih dahulu. Oleh karena itu, flowchart sering digunakan untuk mendukung penjelasan metodologi maupun kerangka proses dalam penelitian berbasis sistem dan komputasi.

Pada penelitian ini, flowchart digunakan untuk menggambarkan alur kerja sistem klasifikasi kebutuhan mendesak korban bencana berbasis data media sosial. Diagram alir tersebut menunjukkan hubungan antar tahapan mulai dari data teks yang diperoleh dari media sosial, proses preprocessing, ekstraksi fitur menggunakan IndoBERT, pembentukan model klasifikasi menggunakan Support Vector Machine, hingga evaluasi performa model. Dengan demikian, penggunaan flowchart dalam penelitian ini membantu menjelaskan alur proses secara lebih jelas, terstruktur, dan mudah dipahami.

2.2.13 Modeling Dan Evaluasi Model

Tahapan modeling merupakan proses membangun model prediksi atau klasifikasi berdasarkan representasi teks yang telah diperoleh. Dalam penelitian ini, digunakan pendekatan transformer-based embedding (misalnya BERT atau model serupa) sebagai

feature extractor, di mana setiap teks diubah menjadi representasi numerik yang menangkap konteks dan makna semantik. Hasil embedding tersebut kemudian dihubungkan dengan lapisan klasifikasi (misalnya Dense layer dengan fungsi aktivasi Softmax) atau algoritma klasifikasi lain seperti SVM, untuk memprediksi kategori kebutuhan korban pascabencana, misalnya pangan, sandang, dan kebutuhan air bersih.

Model dilatih menggunakan data teks yang telah diberi label, sehingga termasuk ke dalam pendekatan supervised learning. Selama proses pelatihan, model belajar mengenali pola linguistik dan konteks semantik yang membedakan tiap kategori. Hasil akhir berupa model klasifikasi yang mampu memprediksi kebutuhan mendesak berdasarkan teks media sosial secara otomatis. Pendekatan ini fleksibel karena memungkinkan:

1) Eksperimen menggunakan deep learning murni (transformer + Softmax).
2) Eksperimen menggunakan hibrida (transformer embedding + classic ML seperti SVM).
3) Mendukung dataset multi-bahasa atau general tanpa tergantung pada bahasa tertentu.

Evaluasi model dilakukan menggunakan metrik berikut:

1. *Accuracy*: proporsi prediksi benar dibanding total data.

$$Accuracy = \frac{jumlah\ prediksi\ benar}{jumlah\ total\ data}$$

$$= \frac{TP + TN}{TP + TN + FP + FN} \quad (2.2)$$

Keterangan:

TP (*True Positive*)

TN (*True Negative*)

FP (*False Positive*)

FN (False Negativ)

2. *Precision*: ketepatan model dalam mengklasifikasikan kebutuhan tertentu.

$$Precision = \frac{TP}{TF + FP} \quad (2.3)$$

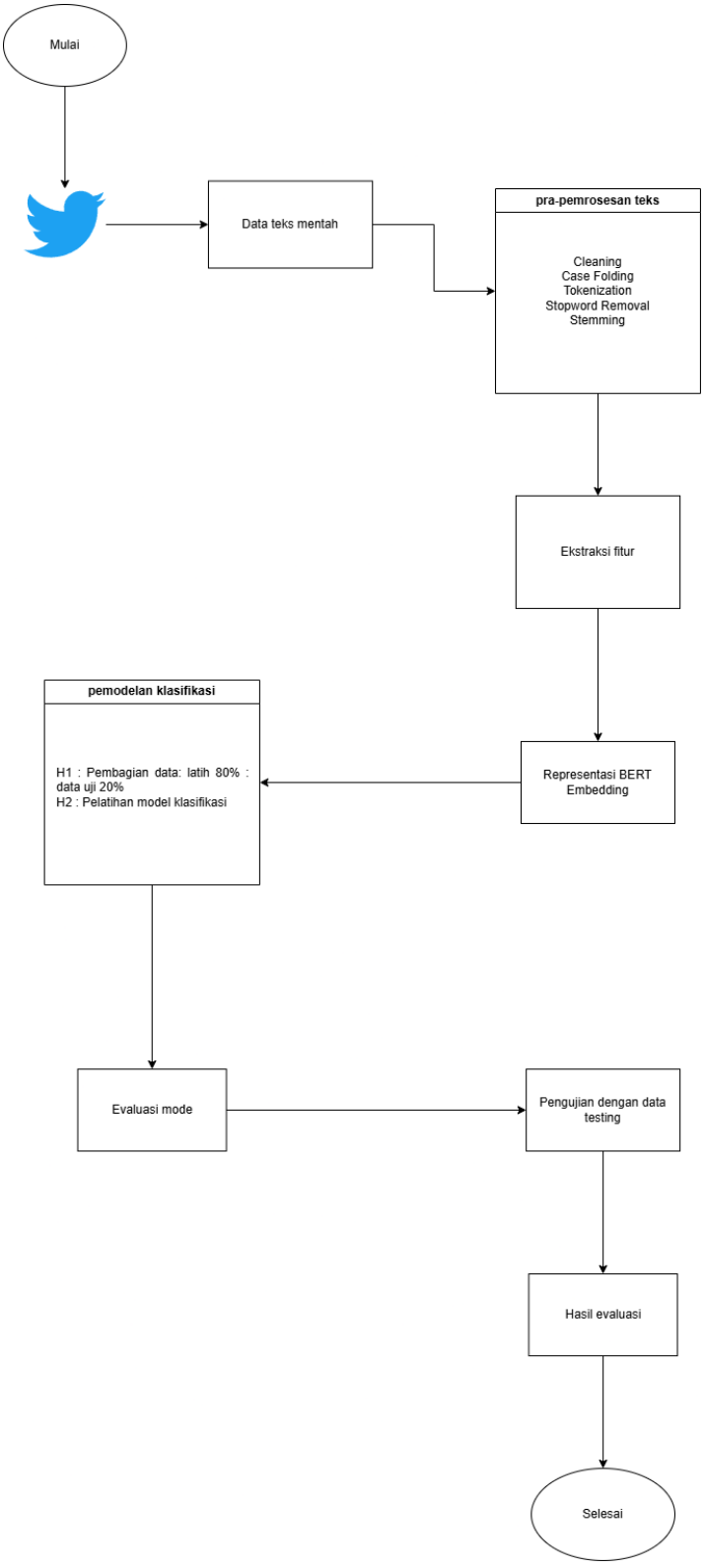
3. *Recall*: sejauh mana model menemukan semua kebutuhan yang relevan.

$$Recall = \frac{TP}{TF + FP} \quad (2.4)$$

4. F1-*Score*: harmonisasi antara precision dan recall.

$$F1 - Score = 2. \frac{Precision\,.\,Recall}{Precision + Recall} \quad (2.5)$$

2.3 Kerangka Pemikiran



Gambar 2. 4 Kerangka pemikiran

Penjelasan Flow Chart:

1. Mulai: Titik awal dari seluruh proses penelitian.
2. Pengumpulan Data: Tahap awal di mana Anda mendapatkan dataset sekunder dari media sosial X.
3. Data teks mentah: Data teks mentah adalah hasil crawling data teks dari twitter yang masih belum melewati tahap pra pemrosesan data
4. Text Preprocessing: Data teks mentah dibersihkan melalui beberapa langkah seperti *cleaning*, *case folding*, *tokenization*, dan *stopword removal*.
5. Ekstraksi Fitur: Teks yang sudah bersih diubah menjadi representasi numerik (vektor) menggunakan model BERT.
6. Pemodelan Klasifikasi: Data dibagi menjadi data latih, validasi, dan uji. Model SVM kemudian dilatih menggunakan data latih.
7. Evaluasi Model: Kinerja model yang telah dilatih diukur menggunakan data uji.
8. Hasil Evaluasi: Hasilnya disajikan dalam bentuk metrik kuantitatif seperti *Accuracy, Precision, Recall*, dan *F1-Score* untuk menjawab rumusan masalah ketiga.
9. Selesai: Titik akhir dari proses penelitian.