BAB II
TINJAUAN PUSTAKA

2.1. Pengertian Corpus

Corpus linguistik mempelajari tentang Bahasa dengan bantuan suatu teknologi komputer modern di pendataan bahasa. Ini hanyalah koleksi secara luaspidato yang dijelaskan. Ada metode dan teknik yang digunakandalam mengolah dan mengembangkan bahasa database. Bidang linguistik terapan menganalisis koleksi besar tertulis dan teks lisan, yang telah hati-hati dirancang untuk mewakili domain tertentu dari penggunaan bahasa, seperti tuturan informal atau tulisan akademik.[8]

Menurut sampai (S Dash, 2010) ada beberapa bagian penting dari korpus, yaitu:

1. Kuantitas, apa yang dimaksud dengan Kuantitas di sini adalah bahwa itu harus dalam ukuran besar mengandung banyak data baik secara lisan maupun bentuk tertulis.
2. Kualitas, semua teks harus diperoleh dari sampel pidato yang sebenarnya dan menulis. Peran seorang linguis sangat besar penting di sini.
3. Representasi, ini harus mencakup contoh berbagai teks
4. Kesederhanaan, harus berisi teks biasa dalam format sederhana.
5. Kesetaraan, sampel yang digunakan dalam korpus ukurannya harus genap.
6. Retrievability adalah Data, informasi, contoh, dan referensi harus mudah diambil dari korpus oleh pengguna akhir. Dia berkaitan dengan data bahasa teknik pengawetan secara elektronik memformat di komputer.
7. Verifikasi, korpus harus terbuka untuk setiap jenis verifikasi empiris. Kita dapat menggunakan data dari korpus untuk apa saja semacam verifikasi. Ini

menempatkan korpus linguistik selangkah lebih maju dari intuitif pendekatan studi bahasa.

8. Augmentasi, ini seharusnya ditingkatkan secara teratur. Ini akan menempatkan korpus untuk merekam perubahan linguistik yang terjadi dalam bahasa dari waktu ke waktu.
9. Dokumentasi, lengkap informasi dari komponen harus dipisahkan dari teks itu sendiri. Ini selalu lebih baik menyimpan informasi dokumentasi terpisah dari teks dan hanya menyertakan header minimal yang berisi referensi ke dokumentasi

2.2. Ujaran Kebencian (Hate Speech)

Menurut Al Hamzi et al (2020), deteksi ujaran kebencian adalah tugas klasifikasi teks yang terkait dengan analisis sentimen. Istilah 'ujaran kebencian' secara formal didefinisikan sebagai 'komunitas apa pun kation yang meremehkan seseorang atau kelompok atas dasar beberapa karakteristik seperti ras, warna kulit, etnis, jenis kelamin, orientasi seksual tion, kebangsaan, agama, atau karakteristik lain. [5]. Di Inggris, terjadi peningkatan ujaran kebencian yang signifikan terhadap mihibah dan komunitas Muslim mengikuti acara terbaru termasuk meninggalkan Uni Eropa, pembunuhan anggota parlemen Jo Cox (dalam hal ini terjadi gelombang percakapan meneriakkan si pembunuh sebagai 'pahlawan' di Twitter, serangan Manchester dan London Ini berkorelasi untuk mencatat lonjakan kejahatan rasial , dan kasus ancaman terhadap keamanan publik karena sifatnya menghasut kejahatan rasial, seperti itu mengikuti serangan van Finsbury Di UE, survei dan repelabuhan yang berfokus pada kaum muda di wilayah EEA menunjukkan

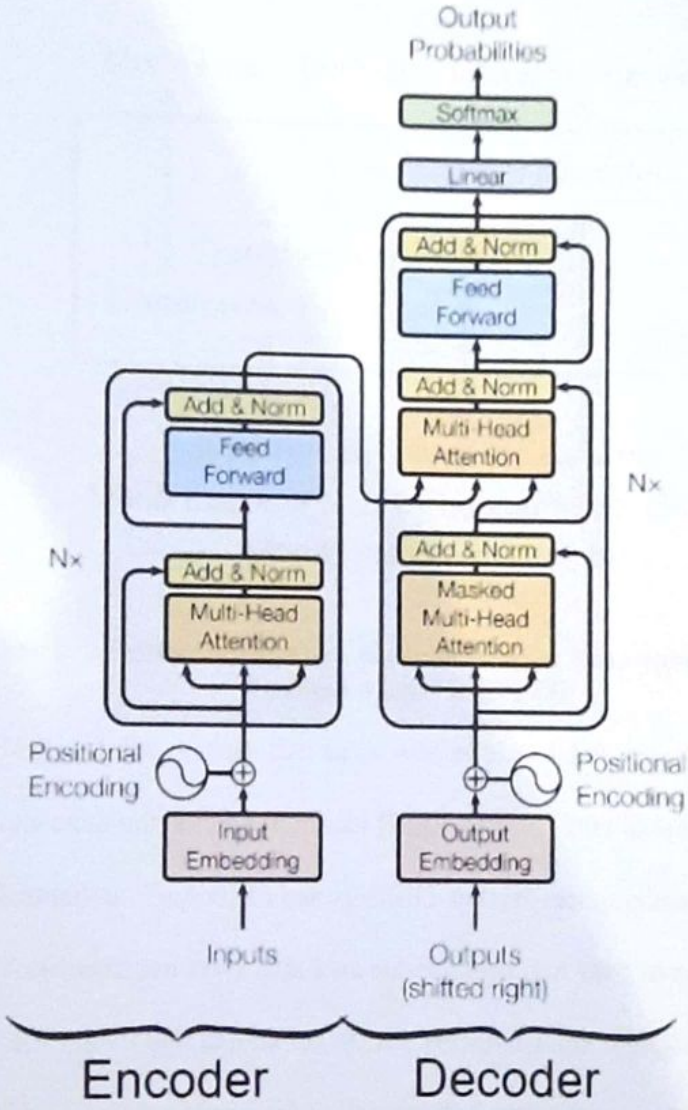
kebencian yang meningkat pidato dan kejahatan terkait berdasarkan keyakinan agama, etnis, seksual orientasi atau jenis kelamin, karena 80% responden mengalami kebencian berbicara online dan 40% merasa diserang atau terancam. [9]

Statistik juga menunjukkan bahwa di AS, ujaran kebencian dan kejahatan terus meningkat sejak saat itu pemilihan Trump Selama bertahun-tahun, perusahaan media sosial seperti Twitter, Facebook, dan YouTube telah memerangi masalah ini dan berhasil diperkirakan bahwa ratusan juta euro diinvestasikan setiap tahun pada langkah-langkah penanggulangan termasuk tenaga kerja. Namun, mereka masih dikritik karena tidak berbuat cukup. Ini sebagian besar karena tindakan tersebut melibatkan peninjauan konten online secara manual untuk mengidentifikasi dan menghapus materi ofensif. Prosesnya adalah tenaga kerja tegang, memakan waktu, dan tidak berkelanjutan atau terukur dalam kenyataan.

2.3. BERT: Bidirectional Encoder Representations from Transformers

BERT (Devlin et al., 2018) adalah salah satu metode NLP canggih yang merupakan tumpukan pembuat enkode transformator seperti pada gambar 2.1. Dengan kata lain, itu dibangun menggunakan blok encoder transformator. Meskipun beberapa metode NLP seperti XLNet sedikit mengunggulinya, BERT masih menjadi salah satu model terbaik untuk berbagai tugas NLP seperti menjawab pertanyaan (Qu et al., 2019), pemahaman bahasa alami (Dong et al., 2019), analisis sentimen, dan inferensi bahasa (Song et al., 2020). [10]

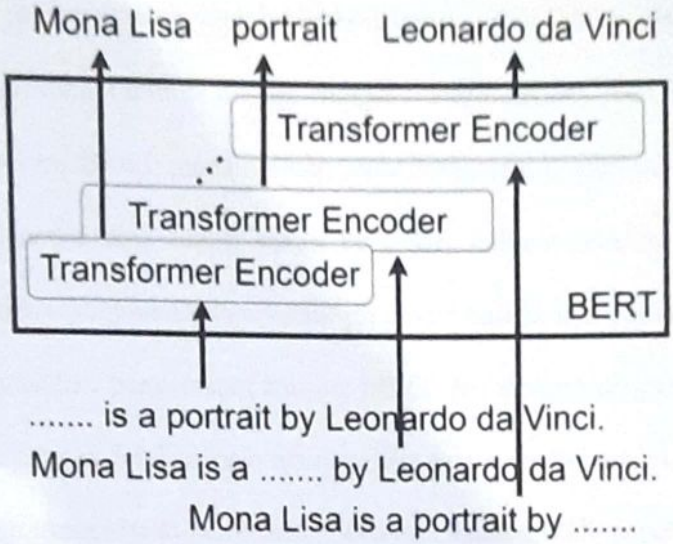
BERT menggunakan teknik pemodelan bahasa bertopeng. Ini menutupi 15% kata dalam dokumen/korpus masukan dan meminta model untuk memprediksi kata yang hilang.



Gambar 2. 1 Encoder dan Decoder Sebagai Bagian dari Sebuah transformer [7]

Seperti yang digambarkan Gambar 2.2, sebuah kalimat dengan kata yang hilang diberikan ke setiap blok pengkodean transformator di tumpukan dan blok tersebut seharusnya memprediksi kata yang hilang. 15% kata hilang dalam

kalimat dan setiap kata yang hilang ditugaskan ke setiap blok encoder di tumpukan. Ini adalah cara yang tidak diawasi karena kata apa pun dapat disamarkan dalam sebuah kalimat dan outputnya seharusnya adalah kata itu.



Gambar 2. 2 Pemberian Umpan Kalimat dengan Kata yang Hilang untuk Latihan Model BERT [8]

Karena tidak diawasi dan tidak memerlukan label, data teks besar Internet dapat digunakan untuk melatih model BERT di mana kata-kata dipilih secara acak untuk disamarkan. Perhatikan bahwa BERT belajar memprediksi kata yang hilang berdasarkan perhatian pada kata-kata sebelumnya dan yang akan datang sehingga bersifat dua arah. Oleh karena itu, BERT bersama-sama mengkondisikan konteks kiri (sebelumnya) dan kanan (akan datang) dari setiap kata. Selain itu, karena kata yang hilang diprediksi berdasarkan kata lain dari kalimat, penyematan BERT untuk kata-kata adalah penyematan sadar konteks. Oleh karena itu, berbeda dengan word2vec (Mikolov et al., 2013a;b; Goldberg & Levy, 2014; Mikolov et al., 2015) dan GloVe (Pennington et al., 2014) yang menyediakan penyematan

tunggal per setiap kata, setiap kata memiliki embedding BERT yang berbeda dalam berbagai kalimat.[10]

Penyematan kata-kata BERT berbeda dalam kalimat yang berbeda berdasarkan konteksnya. Misalnya, kata "bank" memiliki arti yang berbeda dan oleh karena itu penyematan yang berbeda dalam kalimat "Uang ada di bank" dan "Beberapa tanaman tumbuh di tepi sungai". Patut dicatat juga bahwa, untuk kalimat masukan, BERT menampilkan embedding untuk seluruh kalimat selain memberikan embedding untuk setiap kata dari kalimat tersebut. Penyematan kalimat ini tidak sempurna tetapi berfungsi cukup baik dalam aplikasi. Seseorang dapat menggunakan penyematan kalimat BERT dan melatih pengklasifikasi pada mereka untuk tugas deteksi spam atau analisis sentimen. Selama pelatihan model BERT, selain mempelajari penyematan kata dan kalimat utuh, paper (Devlin et al., 2018) juga mempelajari tugas tambahan. [11]

Tugas ini diberikan dua kalimat A dan B, apakah B kemungkinan merupakan kalimat yang mengikuti A atau tidak? Model BERT biasanya tidak dilatih dari awal karena pelatihannya telah dilakukan dalam waktu lama pada data Internet dalam jumlah besar. Untuk menggunakannya dalam aplikasi NLP yang berbeda, seperti analisis sentimen, peneliti biasanya melakukan pembelajaran transfer dan menambahkan satu atau beberapa lapisan jaringan saraf di atas model BERT yang telah dilatih sebelumnya dan melatih jaringan untuk tugas mereka sendiri.

Selama pelatihan, seseorang dapat membekukan bobot model BERT dan hanya melatih lapisan tambahan atau juga menyempurnakan bobot BERT dengan backpropagation. Parameter encoder pada transformator (Vaswani et al., 2017)

adalah 6 lapisan encoder, 512 unit lapisan tersembunyi pada jaringan yang terhubung penuh dan 8 kepala perhatian (h = 8). Sedangkan BERT (Devlin et al., 2018) memiliki 24 lapisan encoder, 1024 unit lapisan tersembunyi dalam jaringan yang terhubung penuh dan 16 kepala perhatian (h = 16). Biasanya, ketika kami mengatakan BERT, yang kami maksud adalah BERT besar (Devlin et al., 2018) dengan parameter yang disebutkan di atas. Karena model BERT berukuran besar dan membutuhkan banyak memori untuk menyimpan model, model BERT tidak dapat dengan mudah digunakan dalam sistem tertanam. Oleh karena itu, banyak versi BERT komersial yang lebih kecil diusulkan dengan jumlah parameter dan jumlah tumpukan yang lebih sedikit.[11]

2.4. Pengertian Long Short Term Memory

Salah satu algoritma deep learning yang dapat memprediksi dan klasifikasi waktu dari jenis *Recurrent Neural Network* adalah *Long Short Term Memory*. Output dari langkah terakhir diumpankan kembali sebagai input pada langkah yang sedang aktifDalam algoritma RNN, Namun, algoritma RNN memiliki kekurangan yaitu tidak dapat memprediksi kata yang disimpan dalam memori jangka panjang.[12]

LSTM memiliki beberapa gerbang yang memiliki fungsi dan tugasnya masing-masing. Berikut penjelasan singkat mengenai struktur LSTM.[7]

1. Forget Gate

Gerbang pertama dalam LSTM disebut dengan *forget gate*. Mudahnya, gerbang ini bertugas untuk melupakan beberapa informasi yang tidak relevan dan sudah

tidak diperlukan oleh sebuah sistem. Alhasil, LSTM dapat menyajikan kumpulan informasi yang lengkap, tetapi tetap aktual sesuai dengan kebutuhan.

2. Input Gate

Gerbang kedua, yakni *input gate* yang bertugas untuk memasukkan informasi yang berguna untuk mendukung keakuratan data. Tugas *input gate* adalah untuk menambahkan informasi yang sebelumnya telah diseleksi terlebih dahulu melalui gerbang *forget gate*. Gerbang ini tidak dimiliki oleh RNN yang dapat memungkinkan satu *input* data untuk satu *output* data.Dalam *input gate* kemudian dikenal istilah *input modulation gate* yang sering tidak ditulis dalam beberapa ulasan tentang LSTM. Sesuai namanya, *input modulation gate* berfungsi dalam melakukan modulasi informasi yang ada, sehingga dapat mengurangi kecepatan konvergensi dari data *zero-mean*.

3. Output Gate

Output yang ketiga adalah *output gate* yang menjadi gerbang terakhir untuk menghasilkan informasi data yang komplet dan aktual. Gerbang ini bisa menjadi terakhir atas sebuah informasi atau hanya menjadi bagian dari tahap pertama saja, sebelum akhirnya informasi akan diproses lewat *input gate* di sel berikutnya.

Data *corpus* yang digunakan terdiri dari bermacam data yang bertotal 3 juta artikel. Artikel-artikel diambil dari teks pada *web*, berita *online*, dan lain-lain, kemudian artikel yang diambil dari *website* atau situs-situs berita dan data dari *website* yang dipilih secara acak pada setiap 1 juta artikel. Contoh data yang diambil dari *corpus* dapat dilihat pada Tabel 1.

Tabel 2. 1 Contoh Data Corpus

No	Sumber	Contoh Data
1	Data campuran	Sementara ketiga kelompok kecil itu sedang berusaha menelusuri gerak pasukan Pati, maka para prajurit dan pengawal yang berada di perkemahan tetap bersiaga sepenuhnya untuk menghadapi segala ke?
2	Website atau situs berita	Mungkin bom rakit-an, tapi masih dalam penelitian laboratorium forensik (labfor)," kata Kabid Humas Polda Sulsel, Joko Subroto, Senin (03/07).
3	Website dipilih secara acak	Misalnya ditahun ini ada anggaran Program Pemberdayaan Masyarakat Kelurahan (PPMK) Rp540 juta, kita bisa meningkatkan anggaran itu, tetapi tidak juga menguap, dan harus ada pengawasan dan kerjasama dengan pihak yang sukses dalam memberdayakan masyarakat.

Penelitian ini menggunakan d*ataset* yang diambil dari komentar media sosial Twitter API. *Training data* yang digunakan terdiri dari 3 kolom yaitu kolom id, kolom label dan kolom *comment* yang berisi komentar-komentar yang berasal dari Facebook. Kolom label berisi 2 buah kategori yang didefinisikan yaitu *Hate Speech* (HS) dan *Non Hate Speech* (Non_HS). *Training data* berjumlah total 5000 komentar. Contoh t*raining data* dapat dilihat di Tabel 2.

Tabel 2. 2 Contoh Dataset

No	Label	Comment
1	HS	Selamat ya pak, sudah d peralat oleh antek2 politik bapak..Sudah memberikan warisan kebodohan dan kecurangan kepada para penerus bangsa ini..Ini akan menjadi catatan bagi kami masyarakat yg ingin kejujuran dan keadilan

No	Label	Comment
2	HS	Sy sebagai rakyat indonesia, malu punya presiden hasil dari pemilihan yg curang dg merampok dan memanipulasi suara rakyat.
3	HS	jangan pencitraan ,rakyat sudah jenuh dan bosan dengan pencitraan yang tak sesuai dengan citra
4	Non_HS	Nilai2.pancasila yg perlu di aplikasikan dlm kehidupan di negri ini namun seiringny waktu hanya sebagai slogan, pengakuan, gambaran, dan teriakan saja, entah ada letak salah dimn, ??
5	Non_HS	Kita boleh membanggakan Panutan kita. Tapi gak boleh merendahkan yg lain. Krn semua manusia pasti ada kelebihan dan kekurangan.

2.5. Aplikasi LSTM

LSTM sebagai bagian dari RNN banyak digunakan untuk mengoperasikan sebuah sistem atau aplikasi. Dengan keunggulan yang dimilikinya, LSTM dapat menyajikan informasi dari riwayat penyimpanan yang telah berlangsung cukup lama. Bahkan, LSTM juga mampu untuk mengklasifikasikan dan menghapus informasi yang telah usang.LSTM banyak digunakan dalam aplikasi-aplikasi seperti penerjemahan, pemodelan bahasa, hingga memberikan keterangan pada gambar. Selain itu, LSTM juga dapat digunakan dalam mesin penjawab pertanyaan otomatis lewat *chatbots*. [13]

Sebelum mengetahui beberapa perbedaan yang terdapat dalam LSTM dan RNN, perlu diketahui dahulu bahwa LSTM merupakan salah satu bentuk modifikasi dari bentuk asalnya, yaitu *recurrent neural network* atau RNN. LSTM juga bukanlah bentuk modifikasi satu-satunya dari RNN, tetapi menjadi salah satu yang populer di antara lainnya.[13]

Perbedaan mendasar dari LSTM dan RNN adalah bahwa LSTM melengkapi kekurangan-kekurangan yang dimiliki oleh pendahulunya, *recurrent*

neural network, yang tidak dapat memprediksi data berdasarkan informasi yang telah disimpan dalam waktu cukup lama. Dengan kata lain, persoalan jangka waktu penyimpanan tidak menjadi permasalahan dalam LSTM.Sistem yang menerapkan LSTM dapat memproses, memprediksi, dan mengklasifikasikan informasi berdasarkan data deret waktu. Sesuai dengan konsepnya, LSTM dapat mengingat dan menghapus data-data lawas yang sudah tidak relevan lagi. Dengan demikian, manajemen informasi akan lebih komplet sekaligus aktual.[13]

Perbedaan berikutnya dari LSTM dan RNN terletak pada lapisan rantainya. Jika RNN memungkinkan satu neuron untuk memproses satu input data pada satu output data, LSTM tidak berlaku demikian. LSTM memiliki berbagai gerbang yang dapat menambah kumpulan informasi dan menggabungkannya.Ada empat gerbang dalam sistem LSTM, yakni *forget gate, input gate, input modulation gate,* serta *output gate*. Keempat gerbang tersebut mempunyai fungsi dan tugasnya masing-masing dalam mengumpulkan, mengklasifikasi, dan memproses data. Tak hanya mempunyai empat gerbang tersebut, LSTM juga memiliki *internal cell state* yang berfungsi untuk menyimpan informasi pilihan dari unit sebelumnya.

2.6. Pembelajaran Mendalam

Pembelajaran mendalam merupakan jenis pembelajaran mesin dan *Artificial Intelligence (AI)* yang meniru cara manusia memperoleh jenis pengetahuan tertentu. Pembelajaran mendalam adalah elemen penting dari ilmu data, yang mencakup statistik dan pemodelan prediktif. Ini sangat bermanfaat bagi*programming* data yang bertugas mengumpulkan, menganalisis, dan

menafsirkan data dalam jumlah besar; pembelajaran mendalam membuat proses ini lebih cepat dan lebih mudah.[14]

Paling sederhana, deep learning dapat dianggap sebagai cara untuk mengotomatiskan analitik prediktif. Meskipun algoritme machine learning tradisional bersifat linier, algoritme pembelajaran mendalam ditumpuk dalam hierarki dengan kompleksitas dan abstraksi yang semakin meningkat. Untuk memahami pembelajaran mendalam, bayangkan seorang balita yang kata pertamanya adalah anjing. Balita belajar apa itu anjing -- dan bukan -- dengan menunjuk ke objek dan mengucapkan kata anjing. Orang tua berkata, "Ya, itu anjing," atau, "Tidak, itu bukan anjing." Saat balita terus menunjuk ke objek, ia menjadi lebih sadar akan fitur yang dimiliki semua anjing. Apa yang dilakukan balita, tanpa menyadarinya, adalah memperjelas abstraksi yang kompleks -- konsep anjing -- dengan membangun hierarki di mana setiap level abstraksi dibuat dengan pengetahuan yang diperoleh dari lapisan hierarki sebelumnya.[14]

2.6.1. Cara Kerja Pembelajaran Mendalam

Program komputer yang menggunakan pembelajaran mendalam melalui proses yang hampir sama dengan pembelajaran balita untuk mengidentifikasi anjing. Setiap algoritma dalam hierarki menerapkan transformasi nonlinier ke inputnya dan menggunakan apa yang dipelajarinya untuk membuat model statistik sebagai output. Iterasi berlanjut hingga output mencapai tingkat akurasi yang dapat diterima. Jumlah lapisan pemrosesan yang harus dilalui data adalah yang menginspirasi dalam label.[14]

Dalam pembelajaran mesin tradisional, proses pembelajaran diawasi, dan pemrogram harus sangat spesifik ketika memberi tahu komputer jenis hal apa yang harus dicari untuk memutuskan apakah gambar berisi anjing atau tidak. Ini adalah proses yang melelahkan yang disebut ekstraksi fitur, dan tingkat keberhasilan komputer sepenuhnya bergantung pada kemampuan pemrogram untuk secara akurat menentukan kumpulan fitur untuk anjing. Keuntungan dari deep learning adalah program membangun set fitur dengan sendirinya tanpa pengawasan. Pembelajaran tanpa pengawasan tidak hanya lebih cepat, tetapi biasanya lebih akurat.[14]

Awalnya, program komputer mungkin diberikan data pelatihan -- sekumpulan gambar yang diberi label oleh manusia untuk setiap gambar anjing atau anjing dengan metatag. Program menggunakan informasi yang diterimanya dari data pelatihan untuk membuat kumpulan fitur untuk anjing dan membuat model prediktif. Dalam hal ini, model yang pertama kali dibuat komputer dapat memprediksi bahwa apa pun dalam gambar yang memiliki empat kaki dan ekor harus diberi label anjing. Tentu saja, program tidak mengetahui label empat kaki atau ekor. Ini hanya akan mencari pola piksel dalam data digital. Dengan setiap iterasi, model prediksi menjadi lebih kompleks dan lebih akurat. Tidak seperti balita, yang akan membutuhkan waktu berminggu-minggu atau bahkan berbulan-bulan untuk memahami konsep anjing, program komputer yang menggunakan algoritme pembelajaran mendalam dapat menampilkan kumpulan pelatihan dan memilah jutaan gambar, mengidentifikasi secara akurat gambar mana yang berisi anjing dalam beberapa menit.

Untuk mencapai tingkat akurasi yang dapat diterima, program deep learning memerlukan akses ke sejumlah besar data pelatihan dan kekuatan pemrosesan, yang keduanya tidak tersedia dengan mudah bagi programmer hingga era big data dan komputasi awan. Karena pemrograman pembelajaran mendalam dapat membuat model statistik kompleks langsung dari keluaran iteratifnya sendiri, ia mampu membuat model prediktif yang akurat dari sejumlah besar data tidak berlabel dan tidak terstruktur. Ini penting karena internet of things (IoT) terus menjadi lebih luas karena sebagian besar data yang dibuat manusia dan mesin tidak terstruktur dan tidak diberi label.[15]

2.6.2. Metode *Deep Learning*

Berbagai metode dapat digunakan untuk menciptakan model deep learning yang kuat. Teknik ini mencakup penurunan kecepatan pembelajaran, transfer pembelajaran, pelatihan dari awal, dan putus sekolah.[16]

1. Tingkat peluruhan pembelajaran.
2. Laju pembelajaran adalah hyperparameter

Faktor yang menentukan sistem atau menetapkan kondisi untuk operasinya sebelum proses pembelajaran -- yang mengontrol seberapa banyak perubahan yang dialami model sebagai respons terhadap perkiraan kesalahan setiap kali bobot model diubah. Tingkat pembelajaran yang terlalu tinggi dapat mengakibatkan proses pelatihan yang tidak stabil atau pembelajaran serangkaian bobot yang tidak optimal. Tingkat pembelajaran yang terlalu kecil dapat menghasilkan proses pelatihan yang panjang yang berpotensi macet.

3. Metode peluruhan kecepatan pembelajaran –

Peningkatan kecepatan pembelajaran atau kecepatan pembelajaran adaptif -- adalah proses mengadaptasi kecepatan pembelajaran untuk meningkatkan performa dan mengurangi waktu pelatihan. Adaptasi kecepatan belajar yang paling mudah dan paling umum selama pelatihan mencakup teknik untuk mengurangi kecepatan belajar dari waktu ke waktu.

4. Transfer pembelajaran

Proses ini melibatkan penyempurnaan model yang telah dilatih sebelumnya; itu membutuhkan antarmuka ke internal jaringan yang sudah ada sebelumnya. Pertama, pengguna memberi makan data baru jaringan yang ada yang berisi klasifikasi yang sebelumnya tidak diketahui. Setelah penyesuaian dilakukan pada jaringan, tugas baru dapat dilakukan dengan kemampuan pengkategorian yang lebih spesifik. Metode ini memiliki keuntungan membutuhkan data yang jauh lebih sedikit daripada yang lain, sehingga mengurangi waktu komputasi menjadi menit atau jam.

5. Pelatihan dari awal

Metode ini mengharuskan pengembang untuk mengumpulkan kumpulan data berlabel besar dan mengonfigurasi arsitektur jaringan yang dapat mempelajari fitur dan model. Teknik ini sangat berguna untuk aplikasi baru, serta aplikasi dengan banyak kategori keluaran. Namun, secara keseluruhan, ini adalah pendekatan yang kurang umum, karena membutuhkan banyak sekali data, menyebabkan pelatihan memakan waktu berhari-hari atau berminggu-minggu.

6. Putus sekolah

Metode ini mencoba memecahkan masalah overfitting di jaringan dengan parameter dalam jumlah besar dengan menjatuhkan unit dan koneksinya secara acak dari jaringan saraf selama pelatihan. Telah terbukti bahwa metode putus sekolah dapat meningkatkan kinerja jaringan saraf pada tugas pembelajaran yang diawasi di berbagai bidang seperti pengenalan ucapan, klasifikasi dokumen, dan biologi komputasi.

2.6.3. Jaringan Syaraf Deep learning

Jenis algoritme machine learning tingkat lanjut, yang dikenal sebagai jaringan neural artifisial, mendukung sebagian besar model deep learning. Akibatnya, pembelajaran mendalam terkadang disebut sebagai pembelajaran saraf yang dalam atau jaringan saraf yang dalam. Jaringan saraf datang dalam beberapa bentuk yang berbeda, termasuk jaringan saraf berulang, jaringan saraf convolutional, jaringan saraf tiruan dan jaringan saraf feedforward, dan masing-masing memiliki manfaat untuk kasus penggunaan tertentu. Namun, mereka semua berfungsi dengan cara yang agak mirip -- dengan memasukkan data dan membiarkan model mencari tahu sendiri apakah model tersebut telah membuat interpretasi atau keputusan yang tepat tentang elemen data yang diberikan.[17]

Jaringan saraf melibatkan proses coba-coba, sehingga mereka membutuhkan sejumlah besar data untuk dilatih. Bukan kebetulan bahwa jaringan saraf menjadi populer hanya setelah sebagian besar perusahaan menggunakan analitik data besar dan mengumpulkan penyimpanan data yang besar. Karena beberapa iterasi pertama model melibatkan tebakan yang agak terdidik tentang isi

gambar atau bagian ucapan, data yang digunakan selama tahap pelatihan harus diberi label sehingga model dapat melihat apakah tebakannya akurat. Ini berarti, meskipun banyak perusahaan yang menggunakan data besar memiliki data dalam jumlah besar, data yang tidak terstruktur kurang bermanfaat. Data tidak terstruktur hanya dapat dianalisis oleh model pembelajaran mendalam setelah dilatih dan mencapai tingkat akurasi yang dapat diterima, tetapi model pembelajaran mendalam tidak dapat melatih data tidak terstruktur.[17]

2.6.4. Batasan dan Tantangan Deep Learning

Keterbatasan terbesar model pembelajaran mendalam adalah mereka belajar melalui observasi. Ini berarti mereka hanya tahu apa yang ada dalam data yang mereka latih. Jika pengguna memiliki sejumlah kecil data atau berasal dari satu sumber spesifik yang belum tentu mewakili area fungsional yang lebih luas, model tidak akan belajar dengan cara yang dapat digeneralisasikan. Masalah bias juga merupakan masalah utama bagi model pembelajaran yang mendalam. [10]Jika sebuah model melatih data yang mengandung bias, model akan mereproduksi bias tersebut dalam prediksinya. Ini telah menjadi masalah yang menjengkelkan bagi pemrogram pembelajaran mendalam karena model belajar untuk membedakan berdasarkan variasi halus dalam elemen data. Seringkali, faktor-faktor yang dianggap penting tidak dijelaskan secara eksplisit kepada programmer. Artinya, misalnya, model pengenalan wajah mungkin membuat penentuan tentang karakteristik orang berdasarkan hal-hal seperti ras atau jenis kelamin tanpa sepengetahuan programmer.[18]

Tingkat pembelajaran juga dapat menjadi tantangan besar bagi model pembelajaran yang mendalam. Jika rate terlalu tinggi, maka model akan konvergen terlalu cepat, menghasilkan solusi yang kurang optimal. Jika kecepatannya terlalu rendah, maka prosesnya mungkin macet, dan akan lebih sulit untuk mencapai solusi. Persyaratan perangkat keras untuk model pembelajaran mendalam juga dapat membuat batasan. Unit pemrosesan grafis (GPU) berperforma tinggi multicore dan unit pemrosesan serupa lainnya diperlukan untuk memastikan peningkatan efisiensi dan penurunan konsumsi waktu. Namun, unit ini mahal dan menggunakan energi dalam jumlah besar. Persyaratan hardware lainnya mencakup memori akses acak dan hard disk drive (HDD) atau solid-state drive (SSD) berbasis RAM. Keterbatasan dan tantangan lainnya adalah sebagai berikut:[18]

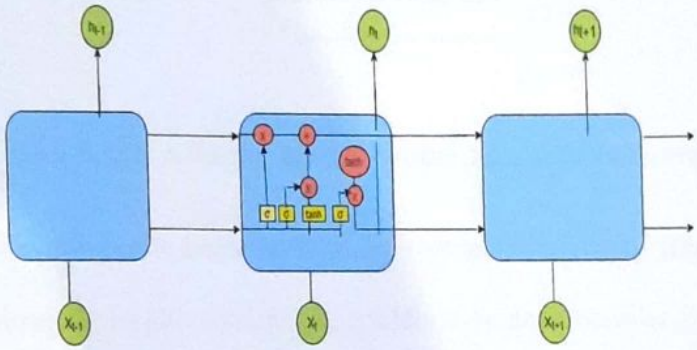
1. Pembelajaran mendalam membutuhkan data dalam jumlah besar. Lebih jauh lagi, model yang lebih kuat dan akurat akan membutuhkan lebih banyak parameter, yang, pada gilirannya, membutuhkan lebih banyak data.
2. Setelah dilatih, model pembelajaran mendalam menjadi tidak fleksibel dan tidak dapat menangani multitasking. Mereka dapat memberikan solusi yang efisien dan akurat tetapi hanya untuk satu masalah tertentu. Bahkan memecahkan masalah serupa akan membutuhkan pelatihan ulang sistem.
3. Aplikasi apa pun yang memerlukan penalaran -- seperti pemrograman atau penerapan metode ilmiah -- perencanaan jangka panjang dan manipulasi data seperti algoritme benar-benar melampaui apa yang dapat dilakukan teknik pembelajaran mendalam saat ini, bahkan dengan data besar.

Tingkat pembelajaran juga dapat menjadi tantangan besar bagi model pembelajaran yang mendalam. Jika rate terlalu tinggi, maka model akan konvergen terlalu cepat, menghasilkan solusi yang kurang optimal. Jika kecepatannya terlalu rendah, maka prosesnya mungkin macet, dan akan lebih sulit untuk mencapai solusi. Persyaratan perangkat keras untuk model pembelajaran mendalam juga dapat membuat batasan. Unit pemrosesan grafis (GPU) berperforma tinggi multicore dan unit pemrosesan serupa lainnya diperlukan untuk memastikan peningkatan efisiensi dan penurunan konsumsi waktu. Namun, unit ini mahal dan menggunakan energi dalam jumlah besar. Persyaratan hardware lainnya mencakup memori akses acak dan hard disk drive (HDD) atau solid-state drive (SSD) berbasis RAM. Keterbatasan dan tantangan lainnya adalah sebagai berikut:[18]

1. Pembelajaran mendalam membutuhkan data dalam jumlah besar. Lebih jauh lagi, model yang lebih kuat dan akurat akan membutuhkan lebih banyak parameter, yang, pada gilirannya, membutuhkan lebih banyak data.
2. Setelah dilatih, model pembelajaran mendalam menjadi tidak fleksibel dan tidak dapat menangani multitasking. Mereka dapat memberikan solusi yang efisien dan akurat tetapi hanya untuk satu masalah tertentu. Bahkan memecahkan masalah serupa akan membutuhkan pelatihan ulang sistem.
3. Aplikasi apa pun yang memerlukan penalaran -- seperti pemrograman atau penerapan metode ilmiah -- perencanaan jangka panjang dan manipulasi data seperti algoritme benar-benar melampaui apa yang dapat dilakukan teknik pembelajaran mendalam saat ini, bahkan dengan data besar.

2.7. Arsitektur Algoritma LSTM

Struktur algoritma LSTM terdiri atas neural network dan beberapa blok memori yang berbeda. Blok memori ini disebut sebagai **cell**. State dari cell dan hidden state akan diteruskan ke cell berikutnya.Seperti yang ditunjukkan pada gambar di bawah, bangun berbentuk persegi panjang berwarna biru adalah ilustrasi cell pada LSTM. [8]



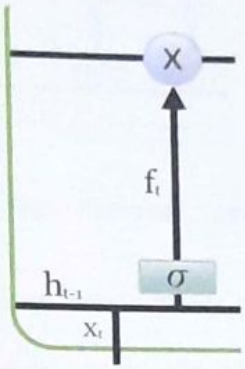
Gambar 2. 3 Arsitektur Algoritma (Sumber : projectpro.io)]8]

Informasi yang dikumpulkan oleh algoritma LSTM kemudian akan disimpan oleh *cell* dan manipulasi memori dilakukan oleh komponen yang disebut dengan ***gate***. Ada tiga jenis gate pada algoritma LSTM, di antaranya *Forget gate, Input gate, dan Output gate.*

1. Forget gate

Forget gate berfungsi untuk menghapus Informasi yang tidak lagi digunakan pada cell. Caranya adalah dengan mengevaluasi output biner dari dua input x(t) dan

output cell sebelumnya h(t-1) dikalikan dengan matriks bobot kemudian ditambahkan dengan nilai bias. Nilai yang didapat kemudian dilewatkan melalui fungsi aktivasi dan menghasilkan output biner. [8]

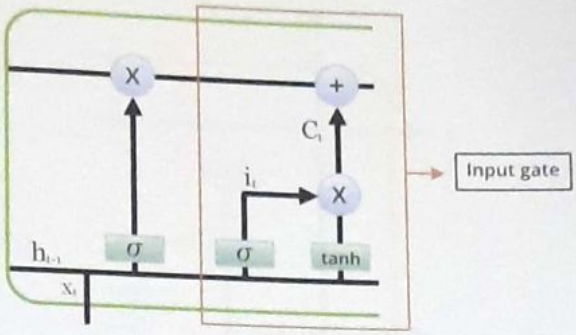


Gambar 2. 4 Forget Gate (Sumber : geeksforeeks.org) [8]

Apabila outputnya bernilai 0, maka informasi dianggap tidak lagi berguna dan bisa dihapus. Begitu sebaliknya, apabila outputnya bernilai 1 maka informasi tersebut disimpan untuk penggunaan di masa mendatang.

2. Input gate

Penambahan informasi yang berguna ke cell state dilakukan oleh input gate. Pertama, informasi diatur menggunakan fungsi sigmoid dan menyaring nilai yang akan disimpan, prosesnya mirip dengan forget gate yang menggunakan input h(t-1) dan x(t). [8]

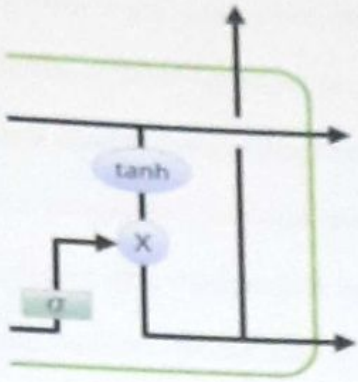


Gambar 2. 5 Input Gate (Sumber : geeksforgeeks.org) [8]

Kemudian, setelah itu sebuah vektor dibuat menggunakan fungsi tanh yang memberikan output dari -1 hingga +1, yang berisi semua kemungkinan nilai dari h(t-1) dan x(t). Terakhir, nilai-nilai vektor dan nilai-nilai yang diatur dikalikan untuk mendapatkan informasi yang berguna.

3. Output gate

Tugas mengekstraksi informasi yang berguna dari cell state saat ini untuk disajikan sebagai nilai keluaran dilakukan oleh output gate. Pertama, sebuah vektor dibangkitkan dengan menerapkan fungsi tanh pada sel. Kemudian, informasi tersebut diatur menggunakan fungsi sigmoid dan menyaring nilai-nilai yang akan disimpan menggunakan input h_t-1 dan x_t. Terakhir, nilai vektor dan nilai yang diatur dikalikan untuk dikirim sebagai output dan input ke sel berikutnya.



Gambar 2. 6 Output Gate (Sumber : geeksforgeeks.org) [8]

2.7.1. Cara Kerja Algoritma LSTM

Secara sederhana, cara kerja algoritma LSTM dapat dijabarkan dalam langkah-langkah berikut:[19]

Langkah 1: LSTM memutuskan informasi apa yang harus tetap utuh dan apa yang harus dibuang dari cell state. Lapisan sigmoid bertanggung jawab untuk membuat keputusan ini.

Langkah 2: LSTM menentukan informasi baru apa yang harus disimpan dan menggantikan yang tidak relevan yang berhasil diidentifikasi pada langkah 1. Fungsi tanh dan sigmoid memainkan peran penting dalam mengidentifikasi informasi yang relevan.

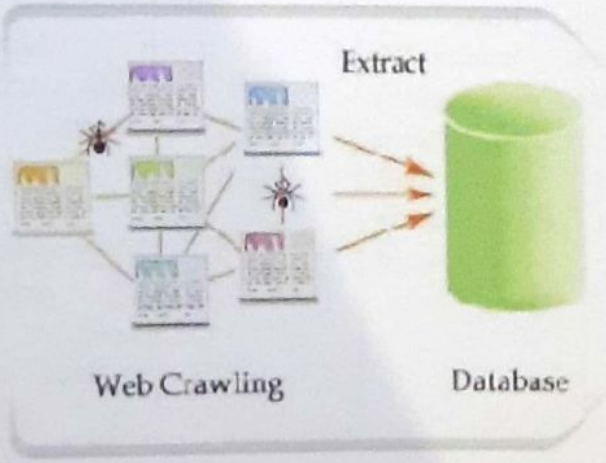
Langkah 3: Output ditentukan dengan bantuan cell state yang sekarang akan menjadi versi yang difilter karena fungsi sigmoid dan tanh yang diterapkan.

2.8. Metode Web Crawling

Aplikasi data retrieval berbasis Web dengan metode crawler atau yang dikenal juga web spider atau web robot adalah suatu program yang yang dibangun dan dirancang dengan metode tertentu yang secara otomatis mengumpulkan semua

data informasi yang diinginkan yang ada dalam bermacam sumber website. Sebuah program yang melintasi struktur hypertext dari web, dimulai dari sebuah alamat awal (seed) dan secara sekursif mengunjungi alamat web di dalam halaman web. Aplikasi Web crawler mengambil informasi pada website yang diberikan kepadanya, kemudian menyerap dan menyimpan semua data informasi yang terkandung didalam website tersebut. Setiap kali aplikasi mengunjungi sebuah website, maka secara otomatis akan merekam semua link yang ada di halaman yang dikunjunginya itu untuk kemudian dikunjungi lagi satu persatu. [19]

Selanjutnya aplikasi akan melakukan data retrieval dan menyimpannya ke dalam suatu media penyimpanan *(harddisk)* dengan kapasitas yang cukup besar. Data-data yang disimpan dalam hardisk ini kemudian, nantinya akan di ambil atau diakses pada saat dilakukan query (mining data). Dalam prosesnya Crawling data yang berhasil dihimpun dapat mencapai milyaran sementara penyajiannya dapat dilakukan secara real time.



Gambar 2. 7 Pencarian Konsep *Web Crawling [11]*

Proses mengunjungi website yang dilakukan crawler dimulai dari mengambil data dan informasi di semua seluruh URL dari website, menelusurinya

satu-persatu, kemudian memasukkannya dalam daftar halaman pada indeks *search engine*, sehingga setiap kali ada perubahan pada website tersebut, akan terupdate secara otomatis. Aplikasi *Web crawling* juga melakukan proses mengambil dan mengumpulkan hyperlink yang ada dalam sebuah web yang dikunjunginya kemudian simpan dalam media berkapasitas besar. *Indexing data* yang berupa *list hyperlink* dilakukan demi memperlancar kegiatan penelusuran oleh *search engine*.

2.9. Road Map Penelitian

Tabel 2. 3 Road Map Penelitian

No	Judul Penelitian	Tahun dan Studi Kasus	Metode Penelitian	Bahasan dan Simpulan
1	Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga	2018, Studi Kasus Pada *Harga Bitcoin*	Sistem yang dibangun pada penelitian ini adalah menggunakan metode	asil analisis menunjukkan bahwa sistem yang dibangun mampu memprediksi
2	Bitcoin Muhammad Wildan Putra Aldi, Jondri, Annisa Aditsania. e-Proceeding of Engineering : Vol.5, No.2 Agustus 2018 | Page 3548		jaringan syaraf tiruan yaitu dengan menggunakan arsitektur *Long Short Term Memory Neural Networks*.	harga Bitcoin dengan baik, dengan rata-rata tingkat akurasi sebesar 93.5% terhadap data testing.

No	Judul peneitiam	Tahun dan Studi Kasus	Metode Penelitian	Bahasan dan Simpulan
3	Analisis Peramalan dengan *Long Short Term Memory* pada Data Kasus Covid-19 di Provinsi Jawa Tengah	2022. Data Kasus Covid-19 di Jawa Tengah.	Jaringan Long ShortTerm Memory atau jaringan LSTM adalah jenis jaringan saraf berulang yang digunakan dalam deep learning dalam kerangka peramalan data time series	Berdasarkan hasil pembahasan maka diperoleh metode terbaik untuk meramalan data kasus Covid-19 di Provinsi Jawa Tengah adalah metode LSTM untuk regresi menggunakan metode window. Hal ini dikarenakan ketepatan nilai RMSE terendah sebesar 715,62 dibandingkan dengan metode yang lainnya. Dalam peramalan ini terdapat kemungkinan bahwa sewaktu-waktu berubah dikarenakan faktor tertentu.

No	Judul peneitiam	Tahun dan Studi Kasus	Metode Penelitian	Bahasan dan Simpulan
4.	Implementasi Model *Long-Short Term Memory* (LSTM) pada Klasifikasi Teks Data SMS Spam Berbahasa Indonesia Erico Dwi Pratama *The Journal on Machine Learning and Computational Intelligence (JMLCI)* Received: June 2, 2022; Revised: July 10, 2022; Accepted: July 15, 2022	2022, SMS Berbahasa Indonesia	Dalam penelitian ini, untuk melakukan klasifikasi teks perlu melibatkan beberapa proses, yaitu: *Data Collection*, *Data Preprocessing*, *Word Representation*, *Classification*, dan *Evaluation/Testing*.	Berdasarkan hasil penelitian yang dilakukan klasifikasi teks pada SMS spam berbahasa indonesia, metode LSTM menghasilkan nilai *accuracy* sebesar 94%. Metode LSTM menghasilkan hasil yang lebih baik dari kedua metode pembandingnya. Model dengan metode LSTM menghasilkan nilai *accuracy* 27% lebih baik dibandingkan dengan metode *Naive Bayes* dan 24% lebih baik dari metode KNN.
4.	Implementasi Algoritma Long Short-Term Memory (LSTM) untuk Mendeteksi Penggunaan Kalimat *Abusive* Pada Teks Bahasa Indonesia	2021. Kalimat abusive pada Bahasa Indonesia	Pada penelitian ini menggunakan salah satu arsitektur dari RNN yaitu Long Short Term Memory (LSTM) yang	Pengujian dilakukan terhadap arsitektur LSTM dan didapatkan hasil bahwa arsitektur ini hanya dapat memprediksi

No	Judul penelitiam	Tahun dan Studi Kasus	Metode Penelitian	Bahasan dan Simpulan
4.	Rizka Dwi Wulandari Santosa , Moch. Arif Bijaksana, Ade Romadhony. e-Proceeding of Engineering : Vol.8, No.1 Februari 2021 | Page 691		biasa digunakan untuk masalah deep learning.	terhadap kelas mayoritas sehingga dilakukan penambahan penggunaan arsitektur yaitu Bidirectional LSTM (BiLSTM). Hasil uji coba menunjukkan BiLSTM lebih baik dalam mengklasifikasi kan kalimat karena terdapat forward dan backward layer yang membuat proses pembelajaran model lebih kompleks dalam mengenal konteks kalimat dan hal ini akan meningkatkan keakuratan hasil klasifikasi pada setiap label. Pada LSTM hanya menghasilkan nilai F1 Score untuk kelas mayoritas saja sebesar 0.812

No	Judul penelitiam	Tahun dan Studi Kasus	Metode Penelitian	Bahasan dan Simpulan
4.				sedangkan pada BiLSTM sudah dapat menghasilkan nilai F1 Score untuk semua kelas.
5.	Implementasi Algoritma Long Short-Term Memory (LSTM) Untuk Mendeteksi Ujaran Kebencian (*Hate Speech*) Pada Kasus Pilpres 2019 Aini Suri Talita , Aristiawan Wiguna Jurnal Matrik Vol.19 No.1 (November) 2019, Hal 37-44 Doi : Https://Doi.Org/10.30812/Matrik.v19i1.495	2019. Kasus Ujaran Kebencian Pilpres 2019.	Penelitian Dengan Menggunakan Jaringan Syaraf Tiruan (Artificial Neural Network/Ann) Maupun Turunannya Telah Banyak Dilakukan, Khususnya Dalam Masalah Praktis *Data Mining*, Klasifikasi, *Clustering*, Ataupun Kasus Khusus Pendeteksian Suatu Objek. Terdapat Beberapa Jenis Metode Yang Merupakan Tipe Khusus Dari Ann, Misalnya Recurrent	Berdasarkan Hasil Penelitian Yang Dilakukan Dengan Menggunakan *Data Testing* 190 Kalimat Dari 950 Kalimat Dari *Dataset*, Algoritma *Long Short Term Memory* Sudah Cukup Baik Dalam Mendeteksi Kalimat Ujaran Kebencian Dengan Nilai Parameter *Recall* Mencapai 0.7021. Nilai Parameter Lainnya Masih Cenderung Rendah Yang Dapat Disebabkan Oleh Kalimat Ujaran

No	Judul peneitiam	Tahun dan Studi Kasus	Metode Penelitian	Bahasan dan Simpulan
5.			Neural Network (Rnn). Pada Penelitian Ini, Salah Satu Arsitektur Dari Rnn Yang Biasa Digunakan Pada Masalah *Deep Learning* Yaitu Long Short Term Memory (Lstm) Akan Diimplementasikan Untuk Mendeteksi Ujaran Kebencian (*Hate Speech*) Berkaitan Dengan Pemilihan Presiden (Pilpres) 2019.	Kebencian Yang Digunakan Mencakup Bahasa Informal, Dan Diambil Langsung Dari Media Sosial Dimana Penulisan Dari Kata-Kata Seringkali Berubah Karena Disingkat, Banyaknya Ejaan Yang Tidak Benar Atau Tidak Konsisten, Juga Kecenderungan Mengganti Huruf Tertentu Sebagai Angka Sehingga Tidak Dapat Terdeteksi Sebagai Ujaran Kebencian Sesuai Dengan Kriteria Pada Model Yang Telah Dibangun.
6	Implementasi *Deep Learning* Menggunakan Metode *CNN* dan *LSTM* untuk Menentukan Berita	2020. Berita Palsu dalam Bahasa Indonesia.	Metode yang digunakan adalah *Convolutional Neural Network* (CNN) dan	Hasil dari penelitian ini menunjukkan bahwa metode CNN dan LSTM

No	Judul peneitiam	Tahun dan Studi Kasus	Metode Penelitian	Bahasan dan Simpulan
6	Palsu dalam Bahasa Indonesia Antonius Angga Kurniawan, Metty Mustikasari Jurnal Informatika Universitas Pamulang Penerbit: Program Studi Teknik Informatika Universitas Pamulang Vol. 5, No. 4, Desember 2020 (544-552)		*Long Short Term Memory* (LSTM). Tahapan penelitian terdiri dari pengumpulan data, *labeling data, preprocessing data, word embedding, splitting data,* proses pembentukan model CNN dan LSTM, evaluasi, pengujian data *input* baru dan perbandingan evaluasi dari model CNN dan LSTM yang sudah terbentuk. Pengumpulan data diambil dari situs penyedia berita-berita hoaks dan berita fakta yang sudah valid, yaitu TurnbackHoax.id.	berhasil diterapkan untuk menentukan berita fakta dan berita palsu dalam bahasa Indonesia dengan baik. CNN memiliki tingkat *accuracy test, precision* dan *recall* sebesar 0.88, sedangkan model LSTM memiliki tingkat *accuracy test, precision* sebesar 0.84 dan *recall* sebesar 0.83.