

BAB II

TINJAUAN PUSTAKA

Agar penulis dapat mengkaji dan mengembangkan penelitian ini, diperlukan beberapa penelitian terdahulu sebagai referensi dan pemahaman dalam proses penelitian yang akan dilakukan. Oleh karena itu, penulis mengumpulkan beberapa penelitian terdahulu yang berkaitan dengan permasalahan yang diangkat dalam tinjauan ini.

II.1. Penelitian Terdahulu

Dengan membandingkan algoritma Nave Bayes dengan PSO, SVM, Random Forest, dan penambahan Partical Swarm Optimisation, hasil klasifikasi text mining dan pengujian model menunjukkan bahwa Nave Bayes dengan PSO memiliki akurasi paling tinggi dari metode-metode yang dibandingkan, dengan AUC sebesar 78.7%. Hal ini berarti metode Nave Bayes dengan PSO merupakan metode yang cukup memadai untuk memprediksi spam email dari Gmail. (Ghani & Sulaiman, 2023)

Naive Bayes adalah algoritma *machine learning (ML)* yang dapat digunakan untuk mengatasi masalah pesanan. Karena menerapkan Teorema Bayes, ia disebut sebagai Naive Bayes. Ini adalah salah satu algoritma pembelajaran mesin yang paling sederhana namun paling efektif yang digunakan untuk menemukan aplikasi di berbagai sektor. Sering kali, fitur-fiturnya bergantung pada beberapa hal. Satu lagi kekurangannya terletak pada masalah kemungkinan. *Naive Bayes* memiliki masalah kemungkinan nol, terutama saat menemukan kata-kata dalam informasi pengujian yang tidak ada dalam informasi persiapan. (Ghaleb et al., 2022)

Metode Naive Bayes Classifier memiliki sejumlah kekurangan yang membuatnya kurang efektif. Akibatnya, dokumen dapat ditempatkan di kelas tertentu meskipun mereka jelas tidak pantas berada di sana. Dalam hal penyaringan spam, kekurangan ini umumnya dimanfaatkan oleh spammer berpengalaman untuk memasukkan spam ke dalam kelas non-spam. Ada banyak cara untuk melakukan hal ini, misalnya dengan menyematkan kata-kata baru sehingga produk tidak dapat memeriksanya. Cara lain adalah dengan menggunakan media gambar untuk menyampaikan spam. Hal ini bergantung pada teknik Gullible Bayes Classifier yang dimaksudkan untuk membedakan kata-kata dan bukan gambar. Selanjutnya, produk tidak dapat memeriksa gambar dan pada akhirnya memasukkan spam ke dalam kelas non-spam. (Rodrigues et al., 2022)

Email Spam adalah salah satu bahaya terbesar saat ini dan telah menyebabkan kemalangan anggaran yang sangat besar. Inovasi pengenalan *phishing* lebih layak diharapkan dapat mengendalikan bahaya *email phishing*. Pada penelitian ini, awalnya memeriksa struktur email. pada saat itu, mengingat model kerangka kerja otak konvolusi tidak beraturan (RCNN) yang ditingkatkan dengan vektor dan instrumen yang terhuyung-huyung, mengusulkan model pengenalan email spam lain yang diberi nama THEMIS, yang digunakan untuk menampilkan email di *header* email, email tubuh, level karakter, dan level kata sepanjang waktu. Cara untuk menilai kecukupan THEMIS menggunakan kumpulan data miring yang memiliki proporsi yang wajar dari phishing dan email asli. Hasil eksplorasi menunjukkan bahwa ketepatan umum THEMIS mencapai 99,848%. Kemudian, tingkat positif palsu (FPR) adalah 0,043%.

Hasil yang menjanjikan ini lebih baik daripada teknik pengenalan saat ini dan menegaskan kecukupan THEMIS di membedakan email spam. (Soni, 2019)

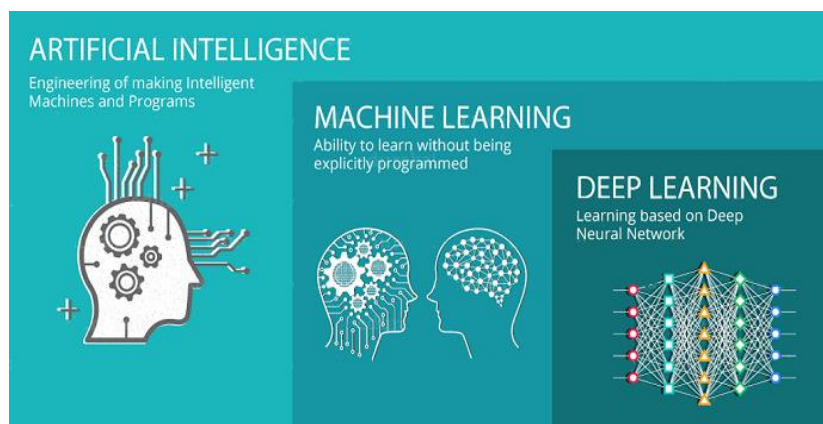
Komunikasi sosial telah berkembang, Email terus menjadi salah satu bentuk komunikasi yang paling banyak digunakan, baik formal maupun informal. Dengan banyaknya bahasa yang didigitalkan untuk dunia elektronik, penggunaan bahasa Inggris masih melimpah. Pesan spam adalah pesan yang tidak pantas dan tidak diinginkan yang biasanya dikirim dengan melanggar keamanan. Email spam ini mencakup *URL phishing*, iklan, *segmen* komersial, dan sejumlah besar penerima sembarangan. Dengan demikian, konten semacam itu selalu berbahaya bagi pengguna, dan banyak penelitian telah dilakukan untuk mendeteksi konten spam tersebut. Usulan tersebut berkonsentrasi pada pemanfaatan perhitungan AI yang ada termasuk Gullible Bayes, CNN, SVM, dan LSTM untuk membedakan dan mengatur konten email. Seperti yang ditunjukkan oleh penemuan ini, model LSTM mengalahkan model-model yang berbeda dengan skor ketepatan yang paling penting 98,4%. (Siddique et al., 2021)

Pesan spam adalah pesan yang dikirim dari banyak penerima yang pada umumnya berisi substansi yang tidak aman dan penuh dendam. Spam akan diklasifikasikan menggunakan mesin vektor pendukung dan algoritme memori jangka pendek, sehingga pengguna email tidak perlu khawatir. Struktur lanjutan dari Recurrent Neural Network (RNN) yang dikenal sebagai Long Short-Term Memory (LSTM) mampu menangani masalah klasifikasi teks. Support Vector Machine (SVM) adalah salah satu teknik yang memberikan hasil terbaik untuk karakterisasi pada informasi dua dimensi. Hasil akhir dari penelitian karakterisasi spam menggunakan

Long Short-Term Memory (LSTM) dengan nilai ketepatan sebesar 96.2% sedangkan Support Vector Machine (SVM) memiliki nilai ketepatan sebesar 88.6%. Pengujian ini menunjukkan bahwa Long Short-Term Memory (LSTM) lebih unggul dibandingkan dengan Vector Machine (SVM) dalam mengenali spam email.. (Mubarikah et al., 2021)

II.2 Konsep Pengetahuan

Perbedaan antara *Artificial Intelligent*, *Machine Learning* dan *Deep Learning* (Gambar II.1) Pembelajaran mendalam adalah bagian dari AI, dan AI adalah bagian dari kecerdasan buatan manusia, yang merupakan istilah umum untuk program PC apa pun yang mencapai sesuatu yang cerdas. (Xia, 2020)



Gambar II.1 perbedaan AI, ML dan DL

[\(https://ids.ac.id/apa-itu-artificial-intelligence-machine-learning-dan-deep-learning/\)](https://ids.ac.id/apa-itu-artificial-intelligence-machine-learning-dan-deep-learning/)

II.2.1 Artificial Intelligence

Studi tentang bagaimana membuat komputer melakukan tugas-tugas cerdas disebut kecerdasan buatan yang di masa lalu hanya dapat dilakukan oleh manusia. Dalam beberapa tahun terakhir, AI telah berkembang pesat, dan telah mengubah gaya hidup masyarakat. Perkembangan AI telah menjadi strategi pembangunan yang penting bagi negara-negara di seluruh dunia, meningkatkan daya saing nasional dan menjaga keamanan. Banyak negara telah memperkenalkan kebijakan preferensial dan memperkuat penerapan kunci. (Zhang et al., 2023)

Dalam sepuluh tahun terakhir, kecerdasan buatan (AI) telah menjadi kemajuan yang signifikan dalam teknologi informasi dan komunikasi. Penggunaan kecerdasan buatan oleh industri tidak hanya terbatas pada industri komunikasi media, namun juga di bidang perbankan, perakitan, administrasi, dan bahkan pemerintahan. Di beberapa negara tertentu, pelaksanaan intelijen berbasis komputer telah mencapai hampir 56%, terutama di area modern. (Shi et al., 2021)

II.2.2 Machine Learning

Pembelajaran mesin (ML) telah matang menjadi teknologi yang akurat dan populer, yang banyak diterapkan dalam perangkat lunak industri. Pendekatan populer untuk *machine learning* adalah bidang yang disebut *Deep Learning (DL)*, yang telah melihat kemajuan yang signifikan selama dekade terakhir. Pada saat yang sama, komputasi awan, seringkali ditawarkan oleh raksasa teknologi seperti Amazon, Google, dan Microsoft. (Roihan et al., 2020)

Machine learning AI menggambarkan batas kerangka kerja yang dapat diperoleh

dari informasi persiapan eksplisit untuk merobotkan metode yang terlibat dalam membangun model yang berwawasan luas dan menyelesaikan tanggung jawab terkait. Model pembelajaran mendalam berkinerja lebih baik daripada model pembelajaran mesin yang dangkal dan metode analisis data konvensional dalam beberapa aplikasi. (Díaz-Ramírez, 2021)

Selama beberapa dekade terakhir, bidang ML telah menghasilkan berbagai kemajuan luar biasa dalam algoritme pembelajaran yang canggih dan teknik pra-pemrosesan yang efisien. Salah satu kemajuan ini adalah evolusi jaringan saraf tiruan (JST) menuju arsitektur jaringan saraf yang semakin dalam dengan peningkatan kemampuan pembelajaran yang dirangkum sebagai pembelajaran mendalam (DL) untuk aplikasi khusus di lingkungan tertutup, DL sudah menunjukkan kinerja manusia super dengan mengungguli kemampuan manusia. Namun, manfaat seperti itu juga ada harganya karena ada beberapa tantangan yang harus diatasi agar berhasil menerapkan model analitik dalam pengaturan bisnis nyata. Hal ini termasuk pilihan yang sesuai dari opsi implementasi *manifold*, bias dan penyimpangan data, mitigasi properti kotak hitam, dan penggunaan kembali model yang telah dikonfigurasi sebelumnya (sebagai layanan). (Studi et al., 2020)

II.2.3 Deep Learning

Deep Learning adalah cabang dari pembelajaran mesin, ansambel teknik dan algoritma pembelajaran yang biasanya menggunakan unit transformasi *non-linear* (seperti sebagai neuron buatan atau *Restricted Boltzmann Machines*) dalam lapisan hierarkis skema. Melalui algoritme pembelajaran, abstraksi yang berbeda dari data

masuk dipelajari dari abstraksi yang lebih rendah ke yang lebih tinggi, lapisan yang lebih rendah di dekat data input, lebih tinggi di lapisan lebih jauh. Penelitian tentang *deep learning* terutama dimotivasi oleh intuisi, argumen teoretis dari teori sirkuit, hasil empiris, dan pengetahuan terkini tentang ilmu saraf. Ini adalah tren yang berkembang dalam pembelajaran mesin karena hasil yang menguntungkan di aplikasi di mana fungsi targetnya *kompleks* dan kumpulan datanya besar. Berbicara secara informal, *metode deep learning* mengekstraksi beberapa struktur yang signifikan dari kumpulan data alih-alih melakukan cukup dangkal atau pada kesimpulan *statistic* representasi data tanpa struktur seperti vektor dan matriks nilai. (Klemetti et al., 2023)

Istilah "deep learning" (DL) dapat dipahami sebagai metode pembelajaran mesin yang mengajarkan komputer untuk berperilaku dengan cara yang mirip dengan manusia, khususnya dengan mempelajari situasi tertentu melalui pembelajaran atau pemrograman khusus. Pembelajaran Mendalam sangat penting untuk kemajuan inovasi berbasis kesadaran buatan (kecerdasan buatan). Komputer mempelajari berbagai model dan menggunakan data yang dikumpulkan untuk melakukan tugas klasifikasi dalam DL. Informasi dapat berupa teks, gambar, atau suara. Dengan banyak informasi, ketepatan dapat diperluas ke tingkat yang lebih signifikan. Secara lebih eksplisit, membangun DL terkait dengan memajukan pameran informasi yang tidak terstruktur di web. (Rudianto et al., 2022)

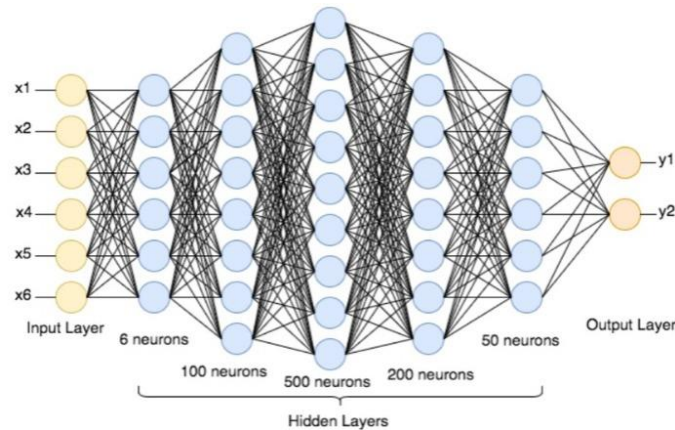
Selain Pembelajaran Mendalam dapat membuat strategi manipulatif dan strategi rencana menjadi lebih kuat. Sebagian besar DL bekerja dengan memanfaatkan strategi jaringan otak. Akibatnya, mereka juga disebut sebagai jaringan saraf dalam karena

berfungsi pada berbagai tingkat atau lapisan. Misalnya, jaringan otak konvensional hanya memiliki 2-3 lapisan yang dapat diakses. Pada saat yang sama, ada lebih dari 150 lapisan dalam jaringan dalam.(Zhou et al., 2020)

II.3 Deep Neural Network (DNN)

II.3.1 Pengertian Deep Neural Network

Deep neural network biasanya terdiri dari lebih dari satu lapisan tersembunyi, diatur dalam arsitektur jaringan yang sangat bersarang. Selain itu, mereka biasanya mengandung neuron tingkat lanjut berbeda dengan jaringan saraf tiruan sederhana. Artinya, *deep neural network* memungkinkan menggunakan lanjutan operasi (misal konvolusi) atau beberapa aktivasi dalam satu neuron daripada menggunakan fungsi aktivasi sederhana. (Amrouche & Stipanovic, 2022)



Gambar II.2. Konsep Deep Neural Network

https://www.researchgate.net/figure/Deep-Neural-Network-architecture_fig1_330120030

Arsitektur canggih pada DNN memberi kemampuan pembelajaran fitur otomatis

ekstrak representasi fitur diskriminatif dengan minimal upaya dari manusia, untuk alasan ini *deep learning* mengatasi data berskala besar dan tidak terstruktur dengan lebih baik. Proses fitur pembelajaran umumnya berlangsung secara hirarkis, dengan fitur abstrak tingkat tinggi sedang dirakit oleh yang lebih sederhana. Namun demikian, tergantung pada jenis data dan pilihan arsitektur, ada mekanisme fitur yang berbeda pembelajaran bersamaan dengan membangun model. (Kwon & Kim, 2023)

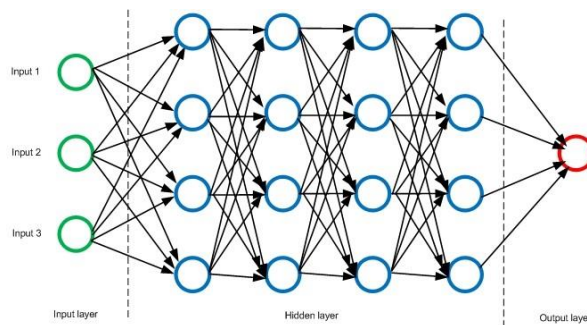
Deep neural network dapat digunakan sebagai *regressor* dan *classifier*. *Regressors* mengeluarkan prediksi numerik tunggal, dan pengklasifikasi menampilkan kelas yang diprediksi. Inferensi *deep neural network* adalah urutan operasi matriks, atau lebih istilah umum, operasi tensor. Ketika sifat spesifik *deep neural network* dari operasi diketahui maka perhitungan dan akses memori dapat dikurangi. Sebagai sisi negatifnya, inferensi bukan lagi rangkaian operasi matriks yang langsung. Semua metode di bagian ini bekerja secara efisien hanya dengan perangkat keras sadar algoritme khusus. ReLU adalah fungsi aktivasi yang umum digunakan di *deep neural network*. ReLU memetakan semua input negatif ke nol. Mengalikan aktivasi bernilai nol dengan bobot tidak berguna karena hasilnya juga demikian nol. Tanda operasi MAC dapat diprediksi menggunakan bit orde tinggi dari aktivasi input dan bobot. Jika tanda operasi MAC diperkirakan negatif, yaitu hasil ReLU akan menjadi nol. (Samek et al., 2021)

II.3.2 Cara Kerja *Deep Neural Network*

JST bekerja dengan menghasilkan fungsi nonlinier dari fitur-fitur yang diekstrak dari kombinasi linier dalam data masukan. Pada umumnya, desain jaringan otak yang

mendalam terdiri dari lapisan informasi, lapisan penyimpanan, dan lapisan hasil di mana setiap lapisan memiliki banyak neuron yang terkoordinasi seperti yang ditampilkan pada Gambar II.2.(Rudianto et al., 2022)

Setiap neuron di setiap lapisan membentuk koneksi yang menghubungkan lapisan-lapisan tersebut. Setiap koneksi memiliki nilai bobot yang digunakan oleh perceptron untuk melakukan peningkatan informasi. Fungsi aktivasi akan menggunakan hasil komputasi untuk melakukan transformasi nonlinier seperti softmax pada lapisan output dan ReLU pada lapisan tersembunyi. Kemampuan softmax akan menghasilkan nilai di antara 0 dan 1 untuk setiap kelas, kualitas ini membahas kemungkinan model meramalkan setiap kelas.. (Adhiatma & Qoiriah, 2022)

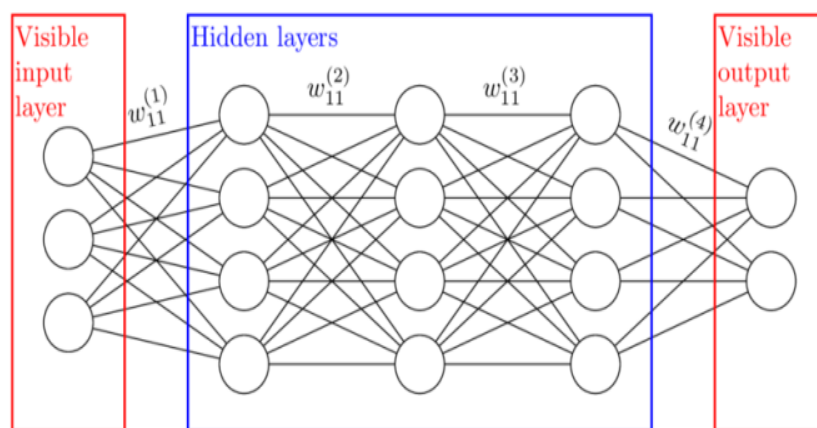


Gambar II.3 contoh arsitektur deep neural network

[\(https://www.megabagus.id/deep-learning/2/\)](https://www.megabagus.id/deep-learning/2/)

Dengan memperkirakan nilai bobot pada setiap koneksi, proses pembelajaran atau pelatihan pada deep neural network bertujuan untuk meminimalisir nilai loss atau error. Ada dua tahap dalam proses iteratif ini: forward pass dan backward pass. Forward pass bertujuan untuk menghitung nilai kesalahan antara hasil ekspektasi model dengan nama pertama dengan menggunakan kemampuan misfortune, sedangkan

backward pass bertujuan untuk memperbaharui nilai bobot melalui teknik peningkatan. Berdasarkan parameter model (bobot dan bias), metode optimasi akan menghitung nilai gradien dari kesalahan model. Nilai gradien ini akan digunakan untuk menentukan besarnya nilai bobot dan bias yang harus diubah untuk mengurangi error. (Ojugo & Yoro, 2021)



Gambar II.4 bentuk hidden layer dari deep neural network

[\(https://rifqimulyawan.com/kamus/hidden-layer/\)](https://rifqimulyawan.com/kamus/hidden-layer/)

Karena jumlah neuron di setiap lapisan dan jumlah lapisan dapat bervariasi, jaringan syaraf tiruan memiliki arsitektur yang fleksibel. Semakin banyak jumlah lapisan, semakin membingungkan contoh-contoh yang dapat diberikan oleh model. Oleh karena itu, penting untuk melakukan pengujian untuk menentukan jumlah lapisan yang tepat atau lapisan rahasia. Pengalaman yang berkembang dalam jaringan otak yang mendalam terjadi secara berulang-ulang dan diharapkan dapat membatasi harga kemalangan. Nilai laju pembelajaran memiliki efek pada kecepatan di mana jaringan saraf dalam menemukan titik kerugian terendah. Jika nilai laju pembelajaran yang

dipilih terlalu sedikit, organisasi otak dalam akan menghabiskan banyak waktu untuk sampai pada kondisi yang sama (paling sedikit kemalangan), terlepas dari kenyataan bahwa ada jaminan bahwa tidak ada titik paling kecil yang terlewatkan. Jika nilai tingkat pembelajaran dipilih terlalu besar, perubahan nilai kemalangan selama pengalaman pendidikan akan sangat besar sehingga organisasi otak yang mendalam akan mengalami masalah dalam mencapai kondisi yang sama.(Assael et al., 2022)

Model komputasi yang disebut Deep Neural Network (DNN) didasarkan pada cara kerja otak manusia. DNN terdiri dari banyak lapisan neuron buatan atau unit pemrosesan yang saling terhubung. Langkah-langkah berikut ini dapat digunakan untuk menjelaskan cara kerja DNN:

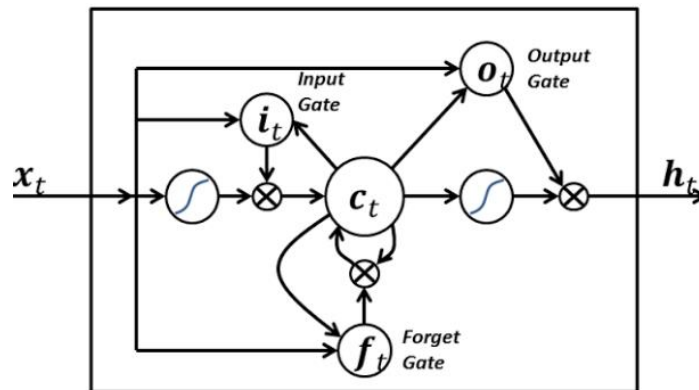
1. Inisialisasi Bobot dan Bias: Setiap neuron dalam DNN memiliki bobot (*weight*) yang menghubungkannya dengan neuron di lapisan sebelumnya. Pada awalnya, bobot dan bias (nilai tambahan) diinisialisasi secara acak atau dengan pendekatan lain yang sesuai.
2. Propagasi Maju (*Forward Propagation*): Data masukan diberikan ke neuron di lapisan input. Nilai masukan dikalikan dengan bobot dan ditambahkan dengan bias di setiap neuron. Kemudian, fungsi aktivasi diterapkan pada hasil tersebut untuk menghasilkan output atau aktivasi neuron.
3. Perhitungan Galat (*Error Calculation*): Setelah propagasi maju, dilakukan perhitungan galat atau perbedaan antara nilai yang diharapkan dan keluaran jaringan. Metode pengukuran kesalahan meliputi *Mean Squared Error (MSE)* atau *Cross-Entropy Loss*.

4. Propagasi Mundur (*Backpropagation*): Langkah ini adalah inti dari pelatihan DNN. Galat dihitung dari lapisan output dikembalikan melalui jaringan, dan bobot dan bias diperbarui berdasarkan galat tersebut. *Backpropagation* memanfaatkan aturan rantai (*chain rule*) dalam kalkulus untuk menentukan kontribusi masing-masing bobot terhadap galat.
5. Pembaruan Bobot dan Bias: Bobot dan bias diperbarui menggunakan algoritma optimisasi, seperti *Stochastic Gradient Descent (SGD)* atau varian lainnya. Tujuan dari pembaruan ini adalah mengurangi galat pada setiap iterasi sehingga jaringan dapat belajar dan meningkatkan kinerjanya.
6. Iterasi dan Pelatihan: Proses propagasi maju, perhitungan galat, propagasi mundur, dan pembaruan bobot dan bias diulang untuk setiap contoh data pelatihan. Jumlah iterasi atau *epoch* dapat ditentukan sebelumnya atau berhenti ketika kriteria penghentian tertentu terpenuhi.
7. Prediksi: DNN dapat digunakan untuk membuat prediksi atau mengklasifikasikan data baru setelah pelatihan. Input baru digunakan untuk menghasilkan output melalui proses propagasi maju.

Langkah-langkah ini diulang selama pelatihan hingga jaringan mencapai tingkat kinerja yang memadai. Dalam DNN yang lebih dalam (*deep*), jumlah lapisan dan neuron dapat sangat besar, memungkinkan jaringan untuk mempelajari fitur-fitur kompleks dan menjalankan tugas yang lebih rumit seperti pengenalan gambar atau pemrosesan bahasa alami. (Aljofey et al., 2020)

II.4 Long Short-Term Memory (LSTM)

Penelitian ini menggunakan Long Short-Term Memory (LSTM) untuk meningkatkan akurasi klasifikasi spam dan rekayasa fitur. Penerapan LSTM sangat penting untuk keberhasilan penelitian ini. LSTM adalah bentuk unggul dari jaringan otak yang terputus-putus di mana terdapat banyak usaha dan bekerja lebih baik dibandingkan dengan organisasi konvensional. Hasilnya, metode LSTM pada data set email untuk klasifikasi spam cocok untuk digunakan. (Supriyadi, 2021)

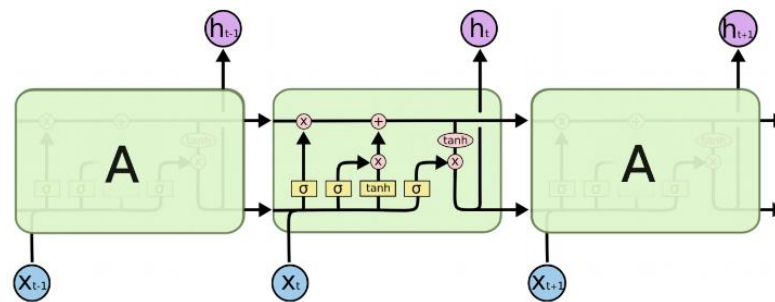


Gambar II.5 bentuk algoritma LSTM

https://www.academia.edu/34438387/PENGENALAN_LSTM_LONG_SHORT_TERM_MEMORY

Hochreiter dan Schmidhuber menciptakan LSTM pada tahun 1997. Algoritma RNN dasar telah ditingkatkan dengan menyertakan LSTM, yang menambahkan status sel untuk mengingat atau melupakan data untuk memecahkan masalah yang hilang. Gerbang sel adalah struktur yang ditemukan dalam keadaan sel. Pintu masuk sel terdiri dari empat bagian: pintu masuk input, pintu pengabaian, pintu masuk keadaan sel

memori, dan pintu hasil. Pintu masukan adalah pintu yang digunakan untuk mengontrol informasi masukan yang layak disimpan atau tidak. Keadaan tersembunyi sebelumnya yang akan disimpan dalam sel memori keadaan tersembunyi saat ini dikendalikan oleh pintu lupa. Berdasarkan informasi dari gerbang lupa dan gerbang masukan, gerbang keadaan sel memori digunakan untuk memperbarui data. Pintu hasil digunakan untuk mengetahui informasi hasil dari organisasi berdasarkan keadaan sel memori. (Hartono et al., 2021)



Gambar II.6 struktur algoritma LSTM

(<https://meydina-rhma.medium.com/implementasi-algoritma-long-short-term-memory-lstm-untuk-prediksi-penutupan-harga-saham-pt-telkom-1d3997ddb81c>)

Sel-sel LSTM dikategorikan dengan menggunakan persamaan berikut:

1. Memutuskan data baru mana yang akan dimasukkan ke dalam CT. Ada dua bagian dalam proses ini. Pertama, sebuah "gerbang input", sebuah gerbang sigmoid yang menentukan nilai mana yang akan diperbarui. Kemudian lapisan $\tanh$ "c_int" yang menghasilkan vektor pengaturan yang baru dan kemudian menggabungkan keduanya untuk melakukan pembaruan pada pengaturan nanti.
2. Menggunakan gerbang sigmoid yang kita sebut sebagai "gerbang lupa",

memutuskan informasi apa yang akan dibuang dari konteks C_{t-1} . Pintu masuk ini membaca dengan teliti sisi atas dari h_{t-1} dan x_t , dan menghasilkan sebuah angka di suatu tempat di antara 0 dan 1 untuk setiap komponen di C_{t-1} . Nilai 1 menandakan "simpan komponen tersebut" sementara 0 menandakan "gagal mengingat komponen tersebut".

3. Perbarui pengaturan lama C_{t-1} ke pengaturan baru C_t . Kita menduplikasi pengaturan lama dengan f_t agar tidak lupa dengan hal-hal yang kita abaikan. Duplikasi pengaturan baru kita c_{int} dengan itu untuk menyimpulkan jumlah yang akan kita masukkan ke dalam bilah pengaturan yang baru dan kemudian tambahkan keduanya.
4. Pilih apa yang akan dihasilkan. Untuk memilih bagian mana dari konteks yang akan dihasilkan, jalankan gerbang sigmoid "gerbang keluaran". Konteks kemudian diproses oleh $\tanh(c_t)$ untuk menghasilkan nilai antara -1 dan 1, dan nilai yang dihasilkan dikalikan dengan output gerbang sigmoid untuk menghasilkan hanya bagian yang ditentukan.

Lima parameter hiper, yaitu learning rate, ukuran batch, epoch, tingkat putus sekolah, dan algoritma optimasi, dipilih dengan menggunakan algoritma optimasi grid search untuk parameter hiper model LSTM. Jangkauan dan sisi atas ideal dari batas-batas hiper yang dipilih oleh model LSTM. (Karim et al., 2021)

II.5 Electronic-Mail (E-Mail)

Email adalah salah satu elemen tingkat tinggi yang terhubung dengan korespondensi elektronik dengan menggunakan inovasi web. Salah satu alasan mengapa email digunakan oleh banyak orang adalah karena email memberikan metode yang sederhana dan cepat untuk mengirim data. Bagaimanapun, email juga tidak dapat dibedakan dari masalah keamanan, misalnya, disadap, dibuat, ditembus oleh infeksi, spamming, mailbomb, transfer email. Namun, administrator sistem dan juga peretas merupakan ancaman bagi kerahasiaan email. Salah satu metode untuk menaklukkan hal ini adalah dengan menyandikan email. (AISYAH et al., 2020)

Email memiliki akar teknologinya di era *pra-Internet* dengan ratifikasi RFC 821, yaitu *Simple Mail Transfer Protokol (SMTP)* di awal 80-an. Kemudian standarisasi HTTP dan munculnya Internet di tahun 90-an, memimpin SMTP menjadi salah satu landasan protokol pendukung infrastruktur layanan *Internet modern* di seluruh dunia. Dalam Sementara itu, Internet terbukti menjadi lahan subur bagi proliferasi layanan digital yang tak ada habisnya yang telah merevolusi hampir setiap aspek publik dan kehidupan pribadi. Sejalan dengan evolusi tersebut, saat ini email masih digunakan secara *massif*. (Reguig, 2022)

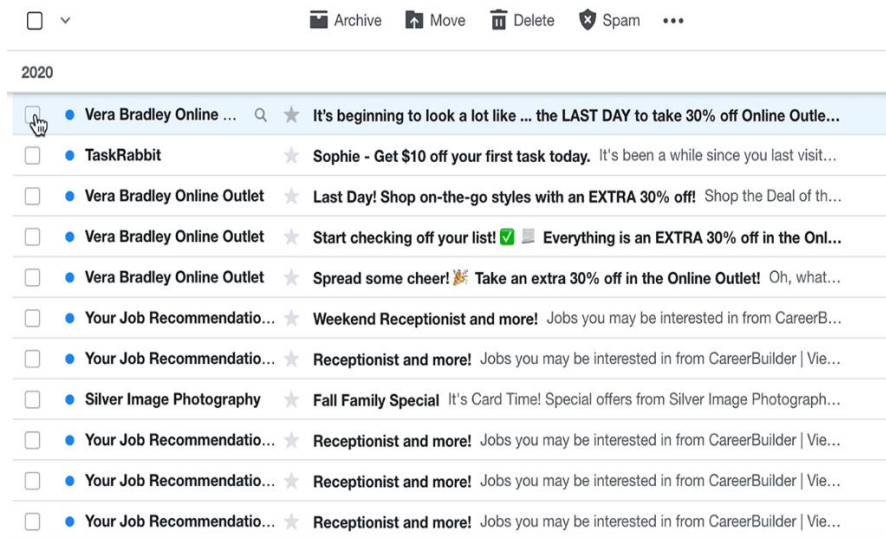
Istilah "e-mail" mengacu pada metode pengiriman pesan secara elektronik. Biasanya dikirim dengan bantuan web. Konten, alamat pengirim, dan alamat tujuan membentuk sebuah pesan elektronik. Model penyimpanan dan penerusan berfungsi sebagai fondasi untuk sistem email jaringan. Kerangka kerja ini menerapkan kerangka kerja server email yang menerima, memajukan, mengirim, dan menyimpan pesan

klien, di mana klien hanya perlu menghubungkan PC mereka ke organisasi. (Iqbal & Khan, 2022)

Email dapat dianalogikan sebagai kotak surat di kantor POS, sementara server email dapat dibandingkan dengan kantor POS. Dalam RFC, ejaan email adalah mail, dan email dikenal sebagai pesan. RFC yang baru dan kelompok IETF menuntut penggunaan huruf besar yang konsisten dalam pengejaan email.. (Mubarikah et al., 2021)

II.6 Spam Email

Spam adalah email spontan yang dikirim dari banyak orang. Email spam adalah istilah yang digunakan untuk menggambarkan pesan yang dikirim dalam email massal atau pesan yang masuk tanpa persetujuan. Sebagian besar waktu, email spam datang dalam bentuk penawaran obat, proposal bisnis palsu, pesan tentang produk, atau tipuan. Karena hasil prediksi yang akurat sangat penting untuk mengurangi spam, maka email harus diprediksi secara akurat. (AISYAH et al., 2020)



Gambar II.7 Contoh Spam E-mail

(<https://dqlab.id/macam-algoritma-data-science-pada-email-spam>)

Pemfilteran spam ditujukan untuk mengurangi volume seminimal mungkin dari email yang tidak diminta. Pemfilteran email adalah pemrosesan email untuk mengatur ulang sesuai dengan beberapa standar yang pasti. Filter surat umumnya digunakan untuk mengelola email masuk, memfilter email spam, mendeteksi dan hilangkan email yang berisi kode berbahaya seperti virus, *trojan* atau *malware*. (Abuzaid & Abuhammad, 2022)

Cara efektif menangani ancaman yang ditimbulkan oleh spam email, pimpinan email penyedia seperti Gmail, Yahoo mail dan Outlook telah menggunakan kombinasi berbagai teknik pembelajaran mesin (ML) seperti *Neural Networks* dalam *filter spamnya*. Teknik ML ini memiliki kemampuan untuk mempelajari dan mengidentifikasi email spam dan pesan *phishing* menganalisis banyak pesan semacam

itu di seluruh koleksi komputer yang sangat banyak. Karena pembelajaran mesin memiliki kapasitas untuk beradaptasi dengan berbagai syaratnya, *filter* spam email Gmail dan Yahoo melakukan lebih dari sekadar memeriksa email sampah menggunakan aturan yang sudah ada sebelumnya. Mereka menghasilkan aturan baru diri mereka sendiri berdasarkan apa yang telah mereka pelajari saat mereka melanjutkan spam mereka operasi penyaringan. Model pembelajaran mesin yang digunakan oleh Google memiliki sekarang maju ke titik yang dapat mendeteksi dan *memfilter spam* dan *email phishing* dengan akurasi sekitar 99,9 persen. Implikasi dari ini adalah bahwa satu dari seribu pesan berhasil menghindari email mereka penyaring spam. Statistik dari Google mengungkapkan bahwa antara 50-70 persen email yang diterima Gmail adalah email yang tidak diminta. deteksi Google model juga memasukkan alat yang disebut *Google Safe Browsing for* mengidentifikasi situs web yang memiliki URL jahat. (Karim et al., 2021)

Rumus pemfilteran spam yang berbeda telah digunakan oleh *Gmail*, *Outlook.com* dan *Yahoo Mail* hanya mengirimkan email yang *valid* ke alamat email mereka pengguna dan menyaring pesan yang tidak sah. Sebaliknya, *filter* ini juga terkadang keliru memblokir pesan otentik. Telah dilaporkan bahwa sekitar 20 persen email berbasis otorisasi biasanya gagal masuk ke kotak masuk penerima yang diharapkan. Penyedia email memiliki merancang berbagai mekanisme untuk digunakan dalam *filter* anti spam email untuk membatasi bahaya yang ditimbulkan oleh *phishing*, *email malware* dan *ransomware* pengguna email. Mekanisme tersebut digunakan untuk menentukan tingkat risiko masing-masing email masuk. Contoh

mekanisme tersebut termasuk spam yang memuaskan batasan, kerangka kerja kebijakan pengirim, daftar putih dan daftar hitam, dan penerima alat verifikasi. Mekanisme ini dapat digunakan oleh satu atau beberapa pengguna. Ketika ambang spam yang memuaskan terlalu rendah, hal itu dapat menyebabkan lebih banyak spam yang menghindari *filter spam* dan memasuki kotak masuk pengguna. Sedangkan memiliki ambang batas yang sangat tinggi dapat menyebabkan beberapa hal penting email diisolasi kecuali *administrator* mengalihkannya. Bagian ini membahas pengoperasian filter anti spam email Gmail, Yahoo, dan Outlook.

