

**KLASIFIKASI PENILAIAN PRODUK NUTRIMAX *FOOD SUPPLEMENT*
MENGUNAKAN METODE *HYBRID CONVOLUTIONAL NEURAL
NETWORK (CNN) - LONG SHORT TERM MEMORY (LSTM)***

(Skripsi)

Oleh

AZZAHRA ZULFA RISWINDA



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRACT

CLASSIFICATION OF NUTRIMAX FOOD SUPPLEMENT PRODUCT ASSESSMENT USING HYBRID CONVOLUTIONAL NEURAL NETWORK (CNN) – LONG SHORT TERM MEMORY (LSTM)

By

AZZAHRA ZULFA RISWINDA

Nutrimax food supplement is one of the vitamin and drug production companies that has a rating of 4.9 on Shopee e-commerce. Product ratings or reviews are a consideration when deciding to buy a product online. Consumers pay attention to reviews of previous buyers to reduce the risk of purchase. Consumer reviews are opinions in the form of text and a rating of 1-5. This study aims to examine the classification performance of Nutrimax food supplement product assessments with accuracy, precision, recall, and f1-score values using the CNN-LSTM hybrid method. CNN and LSTM are two popular methods used to solve text classification problems. CNN has advantages in the process of extracting features or retrieving features inside text documents, meanwhile LSTM has the advantage of processing the relationship between contexts in the text and can store information in the long term. Therefore, combined hybrid methods of CNN and LSTM to obtain the advantages of each method, a CNN-LSTM hybrid accuracy score of 0.994, precision of 0.994, recall of 1.00, and f1-score of 0.996 were obtained.

Keyword : Nutrimax Food Supplement, Classification, CNN, LSTM, Hybrid
CNN-LSTM

ABSTRAK

KLASIFIKASI PENILAIAN PRODUK NUTRIMAX *FOOD SUPPLEMENT* MENGUNAKAN METODE HYBRID *CONVOLUTIONAL NEURAL* *NETWORK (CNN) – LONG SHORT TERM MEMORY (LSTM)*

Oleh

AZZAHRA ZULFA RISWINDA

Nutrimax *food supplement* merupakan salah satu perusahaan produksi vitamin dan obat-obatan yang memiliki rating penilaian 4.9 di *e-commerce Shopee*. Penilaian produk atau ulasan menjadi pertimbangan saat memutuskan untuk membeli produk secara *online*. Konsumen memperhatikan ulasan pembeli sebelumnya untuk mengurangi risiko pembelian. Ulasan konsumen merupakan opini berupa teks dan rating 1-5. Penelitian ini bertujuan untuk mengkaji performa klasifikasi penilaian produk Nutrimax *food supplement* dengan nilai akurasi, presisi, *recall*, dan *f1-score* menggunakan metode *hybrid* CNN-LSTM. CNN dan LSTM adalah dua metode populer yang digunakan untuk menyelesaikan masalah klasifikasi teks. CNN memiliki kelebihan dalam proses ekstraksi fitur atau mengambil fitur di dalam dokumen teks, sedangkan LSTM memiliki kelebihan untuk memproses keterkaitan antar konteks di dalam teks dan dapat menyimpan informasi dalam jangka waktu panjang. Oleh karena itu, dikombinasikan metode *hybrid* CNN dan LSTM untuk memperoleh kelebihan masing-masing metode, diperoleh nilai akurasi *hybrid* CNN-LSTM sebesar 0.994, presisi sebesar 0.994, *recall* 1.00, dan *f1-score* 0.996.

Kata Kunci : Nutrimax Food Supplement, Klasifikasi, CNN, LSTM, Hybrid CNN-LSTM

**KLASIFIKASI PENILAIAN PRODUK NUTRIMAX *FOOD SUPPLEMENT*
MENGUNAKAN METODE *HYBRID CONVOLUTIONAL NEURAL
NETWORK (CNN) - LONG SHORT TERM MEMORY (LSTM)***

Oleh

Azzahra Zulfa Riswinda

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA**

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

Judul Skripsi : **KLASIFIKASI PENILAIAN PRODUK
NUTRIMAX FOOD SUPPLEMENT
MENGUNAKAN METODE *HYBRID
CONVOLUTIONAL NEURAL NETWORK
(CNN) - LONG SHORT TERM MEMORY
(LSTM)***

Nama Mahasiswa : **Azzahra Zulfa Riswinda**


Nomor Pokok Mahasiswa : **1917031058**

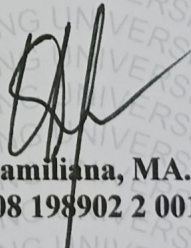
Program Studi : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**




1. **Komisi Pembimbing**


Dian Kurniasari, S.Si., M.Sc
NIP. 19690305 199603 2 001


Prof. Dra. Wamiliana, MA., Ph.D.
NIP. 19631108 198902 2 001

2. **Ketua Jurusan Matematika**


Dr. Aang Nuryaman, S.Si., M.Si
NIP. 19740316 200501 1 001


MENGESAHKAN

1. Tim Penguji

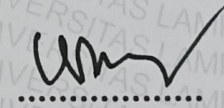
Ketua : **Dian Kurniasari, S.Si., M.Sc.**



Sekretaris : **Prof. Dra. Wamiliana, MA., Ph.D.**



Penguji
Bukan Pembimbing : **Ir. Warsono, M.S., Ph.D.**



Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung



Dr. Eng Heri Satria, S.Si., M.Si.
NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi : **25 Mei 2023**

PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : **Azzahra Zulfa Riswinda**
Nomor Pokok Mahasiswa : **1917031058**
Jurusan : **Matematika**
Judul Skripsi : **KLASIFIKASI PENILAIAN PRODUK
NUTRIMAX *FOOD SUPPLEMENT*
MENGUNAKAN METODE *HYBRID
CONVOLUTIONAL NEURAL NETWORK
(CNN) - LONG SHORT TERM MEMORY
(LSTM)***

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila di kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 25 Mei 2023

Penulis,



Azzahra Zulfa Riswinda
NPM. 1917031058

RIWAYAT HIDUP

Penulis bernama lengkap Azzahra Zulfa Riswinda. Lahir di Bandar Lampung pada tanggal 24 Agustus 2001, merupakan anak pertama dari dua bersaudara, pasangan Bapak Riswanto dan Ibu Sri Indarsih. Penulis mempunyai adik bernama Azkia Salwa Maulia.

Penulis mengawali Pendidikan taman kanak-kanak di TK Amarta Tani HKTI Bandar Lampung pada tahun 2005-2007. Kemudian penulis melanjutkan pendidikan Sekolah Dasar di SD Negeri 1 Labuhan dalam pada tahun 2007-2013. Selanjutnya penulis melanjutkan Pendidikan Sekolah Menengah Pertama di MTs Negeri 2 Bandar Lampung pada tahun 2013-2016 dan melanjutkan pendidikan Sekolah Menengah Atas di MA Negeri 1 Bandar Lampung jurusan MIPA pada tahun 2016-2019.

Pada tahun 2019 penulis melanjutkan pendidikan Strata Satu (S1) di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung, melalui jalur SBMPTN. Selama menjadi mahasiswa penulis aktif di beberapa organisasi yaitu Generasi Muda HIMATIKA (GEMATIKA) 2019, Himpunan Mahasiswa Matematika (HIMATIKA) FMIPA Unila sebagai Anggota Bidang Eksternal periode 2020, Staf Bidang Humas Koperasi Mahasiswa (KOPMA) Universitas Lampung tahun 2021, dan Anggota Aktif Koperasi Mahasiswa (KOPMA) Universitas Lampung sejak tahun 2019.

Pada bulan Januari hingga Februari 2022 penulis melaksanakan Kerja Praktik (KP) di Badan Pendapatan Daerah (BAPENDA) Provinsi Lampung sebagai bentuk pengembangan diri serta menerapkan ilmu yang telah didapat selama perkuliahan.

Selanjutnya pada bulan Juni hingga Agustus 2022 penulis melaksanakan Kuliah Kerja Nyata (KKN) Periode II di Kelurahan Sumberrejo, Kecamatan Kemiling, Kota Bandar Lampung sebagai bentuk pengabdian kepada masyarakat. Selama menjadi mahasiswa penulis juga mengikuti program MBKM, yaitu Kampus Mengajar di SDN 4 Rejomulyo Lampung Selatan pada bulan Agustus hingga Desember 2021, dan juga aktif mengikuti kegiatan *volunteer* online melalui *instagram*.

KATA INSPIRASI

“ Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya...”
(Q.S Al-Baqarah: 286)

“Maka bersabarlah kamu, sesungguhnya janji Allah itu benar”
(Q.S Ar-Rum:60)

“...dan aku belum pernah kecewa dalam berdoa kepada-Mu, ya Tuhanku”
(Q.S Maryam: 4)

“Cukuplah Allah (menjadi penolong) bagi kami dan Dia sebaik-baiknya pelindung”
(Q.S Ali Imran : 173)

PERSEMBAHAN

Dengan mengucapkan puji dan syukur saya haturkan kepada Allah SWT. Yang telah memberikan rahmat, hidayah, dan karunia-Nya kepada saya. Saya persembahkan karya sederhana dengan penuh ketulusan hati sebagai rasa cinta dan sayang saya kepada :

Ibu, Ayah, dan Adik Salwa

Terima kasih telah memberikan do'a yang tiada henti untuk kelancaran setiap langkah baik dan juga dukungan kepada penulis.

Dosen Pembimbing dan Pembahas

Terima kasih kepada dosen pembimbing dan pembahas yang telah membantu, memberikan motivasi, arahan serta ilmu yang berharga kepada penulis.

Almamater Tercinta Universitas Lampung

SANWACANA

Puji dan syukur penulis haturkan kepada Allah SWT., yang telah memberikan rahmat, hidayah, serta karunia-Nya kepada penulis, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Penilaian Produk Nutrimax *Food Supplement* Menggunakan Metode *Hybrid Convolutional Neural Network (CNN) – Long Short Term Memory (LSTM)*”. terselesaikannya skripsi ini tidak lepas dari dukungan, bimbingan, saran, serta do’a dari berbagai pihak. Dengan segala kerendahan hati penulis mengucapkan terima kasih kepada:

1. Ibu Dian Kurniasari, S.Si, M.Sc., selaku Dosen Pembimbing 1 atas kesabaran dan kesediaannya untuk memberikan bimbingan, kritik, dan saran dalam proses penyelesaian skripsi ini serta selalu meluangkan waktunya untuk bimbingan.
2. Ibu Prof. Dra. Wamiliana, M.A., Ph.D., selaku Dosen Pembimbing 2 yang telah memberikan saran serta arahan kepada penulis dan selalu meluangkan waktunya untuk bimbingan.
3. Bapak Ir. Warsono, M.S., Ph.D., selaku Dosen Pembahas skripsi dan Dosen Pembimbing Akademik yang telah memberikan kritik, saran dan masukan yang sangat membantu penulis dalam memperbaiki skripsi dan menyelesaikan perkuliahan dengan baik
4. Bapak Dr. Aang Nuryaman, S.Si, M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Bapak Dr. Eng Heri Satria, S.Si., M.Si, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen, Staf, dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang telah banyak

membantu selama perkuliahan.

7. Ayah, Ibu, Adik Salwa yang selalu memberi doa dan dukungan selama perkuliahan serta penulisan skripsi ini.
8. Sahabat-sahabatku, yaitu Alfira, Melisa, Niken, Widya, dan Silvi yang selalu menemani selama pengerjaan skripsi ini.
9. Teman-teman satu bimbingan, yaitu Zida, Fiqih, Dea, Anin, Mia, Tiara yang selalu memberikan semangat dalam pengerjaan skripsi ini.
10. Teman-teman Matematika 2019, terima kasih atas kebersamaannya.
11. Seluruh pihak yang telah membantu dan terlibat dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa banyak terdapat kekurangan dalam penulisan skripsi ini masih. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari semua pihak.

Bandar Lampung, 25 Mei 2023
Penulis

Azzahra Zulfa Riswinda
NPM. 1917031058

DAFTAR ISI

	Halaman
DAFTAR TABEL	xvi
DAFTAR GAMBAR.....	xvii
I. PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Tujuan Penelitian.....	4
1.3 Manfaat Penelitian.....	4
II. TINJAUAN PUSTAKA.....	5
2.1 <i>Data Mining</i>	5
2.2 Klasifikasi.....	6
2.3 <i>Imbalanced Data</i>	7
2.3.1 <i>Random Oversampling</i>	7
2.3.2 <i>Random Undersampling</i>	8
2.4 <i>Word Embedding</i>	9
2.4.1 <i>Word2Vec</i>	9
2.5 Fungsi Aktivasi.....	10
2.5.1 Fungsi Aktivasi <i>Rectified Linear Unit (ReLU)</i>	10
2.5.2 Fungsi Aktivasi Sigmoid	11
2.5.3 Fungsi Aktivasi <i>Tangent Hyperbolic (TANH)</i>	12
2.5.4 Fungsi Aktivasi <i>Softmax</i>	12
2.6 <i>Deep Learning</i>	13
2.6.1 <i>Convolution Neural Network (CNN)</i>	13
2.6.2 <i>Recurrent Neural Network (RNN)</i>	15
2.6.3 <i>Long-Short Term Memory (LSTM)</i>	15
2.6.4 <i>Hybrid CNN-LSTM</i>	18
2.7 <i>Hyperparameter</i>	19
2.8 Evaluasi Model.....	20
2.8.1 Akurasi	21
2.8.2 Presisi	21
2.8.3 <i>Recall</i>	21
2.8.4 <i>F1-Score</i>	22

III. METODE PENELITIAN	24
3.1 Waktu dan Tempat Penelitian	24
3.2 Spesifikasi Perangkat	24
3.3 Data Penelitian	24
3.4 Metode Penelitian.....	26
IV. HASIL DAN PEMBAHASAN	28
4.1 <i>Input data</i>	28
4.2 <i>Preprocessing Data</i>	28
4.2.1 <i>Case Folding</i>	29
4.2.2 <i>Punctuation Mark Removal</i>	30
4.2.3 <i>Stopword Removal</i>	31
4.2.4 <i>Tokenisasi</i>	31
4.2.5 <i>Stemming</i>	32
4.3 <i>Visualisasi Data</i>	33
4.4 <i>Balancing Data</i>	35
4.5 <i>Word Embedding</i>	36
4.6 <i>Splitting Data</i>	37
4.7 <i>Membangun Model Hybrid CNN-LSTM</i>	37
4.8 <i>Hypertuning</i>	39
4.9 <i>Fitting Model</i>	41
4.9.1 <i>Fitting Model 80% data training dan 20% data testing</i>	41
4.9.2 <i>Fitting Model 90% data training dan 10% data testing</i>	42
4.10 <i>Evaluasi Model</i>	43
4.10.1 <i>Evaluasi Model 80% data training dan 20% data testing</i>	43
4.10.2 <i>Evaluasi Model 90% data training dan 10% data testing</i>	46
V. KESIMPULAN	49
5.1 <i>Kesimpulan</i>	49
DAFTAR PUSTAKA	50

DAFTAR TABEL

Tabel	Halaman
Tabel 1. Data Penilaian Produk Nutrimax <i>Food Supplement</i>	25
Tabel 2. Contoh Hasil Proses <i>Case Folding</i>	29
Tabel 4. Contoh Hasil Proses <i>Stopword Removal</i>	31
Tabel 5. Contoh Hasil Proses Tokenisasi.....	32
Tabel 6. Contoh Hasil Proses <i>Stemming</i>	32
Tabel 7. <i>Splitting</i> Data.....	37
Tabel 8. Nilai Akurasi, Presisi, <i>Recall</i> , dan <i>F1 score</i> 80% data <i>training</i> dan 20% data <i>testing</i>	44
Tabel 9. Nilai Akurasi, Presisi, <i>Recall</i> , dan <i>F1 score</i> 90% data <i>training</i> dan 20% data <i>testing</i>	47

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Grafik <i>Random Oversampling</i>	8
Gambar 2. Grafik <i>Random Undersampling</i>	8
Gambar 3. Grafik Fungsi Aktivasi ReLU	11
Gambar 4. Grafik Fungsi Aktivasi Sigmoid	11
Gambar 5. Grafik Fungsi Aktivasi Tanh.....	12
Gambar 6. Grafik Fungsi Aktivasi <i>Softmax</i>	13
Gambar 7. Arsitektur CNN	14
Gambar 8. Arsitektur RNN	15
Gambar 9. Arsitektur LSTM.....	16
Gambar 14. Alur Metode <i>Hybrid CNN-LSTM</i>	19
Gambar 15. Gambaran <i>Confusion Matrix</i>	20
Gambar 16. <i>Flowchart</i> Penelitian.	27
Gambar 17. Visualisasi Data.....	33
Gambar 18. <i>Wordcloud</i> Rating 5	33
Gambar 19. <i>Wordcloud</i> Rating 4	34
Gambar 20. <i>Wordcloud</i> Rating 3	34
Gambar 21. <i>Wordcloud</i> Rating 2	34
Gambar 22. <i>Wordcloud</i> Rating 1	35
Gambar 23. <i>Syntax Random Oversampling</i>	35
Gambar 24. <i>Balancing Data</i>	36
Gambar 25. <i>Word Embedding</i>	37
Gambar 26. Membangun model <i>hybrid CNN-LSTM</i>	38
Gambar 27. <i>Hypertuning</i>	40
Gambar 28. Parameter terbaik 80% data <i>training</i> dan 20% data <i>testing</i>	40

Gambar 29. Parameter terbaik 90% data <i>training</i> dan 10% data <i>testing</i>	41
Gambar 30. Grafik <i>accuracy</i> dan <i>val accuracy</i> .80% data <i>training</i> dan 20% data <i>testing</i>	42
Gambar 31. Grafik <i>accuracy</i> dan <i>val accuracy</i> 90% data <i>training</i> dan 10% data <i>testing</i>	42
Gambar 32. <i>Confusion Matrix</i> 80% data <i>training</i> dan 20% data <i>testing</i>	43
Gambar 33. <i>Confusion Matrix</i> 90% data <i>training</i> dan 10% data <i>testing</i>	46

I. PENDAHULUAN

1.1 Latar Belakang Masalah

Suryaprana Nutrisindo atau dikenal dengan *Nutrimax food supplement* merupakan salah satu perusahaan produksi vitamin terkenal di Indonesia. Perusahaan ini merupakan perusahaan distribusi nasional yang menjual vitamin dan obat-obatan tradisional. *Nutrimax food supplement* menjual sebanyak 70 jenis produk vitamin dan obat-obatan yang dapat dikonsumsi bagi segala kalangan, mulai dari bayi, anak kecil, orang tua, ibu hamil, sampai dengan pengidap penyakit tertentu. Selain itu *Nutrimax food supplement* berstatus halal untuk dikonsumsi, dan juga sangat mudah untuk dibeli, via online melalui *Shopee*, Tokopedia, serta dapat dibeli langsung di Kimia Farma dan Apotek terdekat.

Nutrimax food supplement memiliki rating penilaian 4.9, serta merupakan produk berkualitas dengan penjualan tertinggi di *e-commerce Shopee*. Penilaian produk atau ulasan menjadi pertimbangan saat memutuskan untuk membeli produk secara *online*. Konsumen memperhatikan ulasan pembeli sebelumnya untuk mengurangi risiko pembelian (Mulyana, 2021). Penilaian yang terdapat pada fitur *shopee* meliputi ulasan konsumen dan juga rating bintang 1-5 yang diberikan oleh konsumen sesuai dengan kepuasan terhadap produk yang dibeli. Bintang 1 diberikan ketika konsumen merasa amat sangat kecewa terhadap produk yang dibeli, sedangkan bintang 5 diberikan ketika konsumen mendapatkan produk yang sangat memuaskan.

Ulasan konsumen merupakan opini publik berupa teks. *Text mining* adalah proses penemuan informasi dalam jumlah besar pada data teks, sehingga bentuknya tidak terstruktur atau kemudian disebut sebagai *unstructured data*. *Text mining* dapat diterapkan pada banyak kasus, salah satunya yaitu klasifikasi. Klasifikasi adalah cara mengelompokkan objek berdasarkan karakteristik dan ciri-ciri objek klasifikasi. Metode klasik yang sering digunakan dalam klasifikasi yaitu *Naïve Bayes*, *Support Vector Machine*, dan *Decission Tree*. Seiring berjalannya waktu dan pengembangan metode, salah satu bidang dari *machine learning* yaitu *deep learning* muncul dengan memanfaatkan jaringan saraf tiruan untuk implementasi permasalahan *dataset* yang besar dan menjadikan komputer untuk dapat berpikir layaknya seperti manusia (Wibawa, dkk., 2018).

CNN merupakan salah satu metode yang mengawali keberhasilan *deep learning*. Perbedaan CNN dari metode *deep learning* lainnya adalah CNN memiliki tambahan arsitektur yang dioptimalkan untuk citra (Rasywir, dkk., 2020). Namun, penggunaan CNN tidak terbatas pada citra, bisa juga digunakan untuk mencapai hasil dalam masalah pemrosesan data teks dan pengenalan suara. CNN memiliki jaringan arsitektur *spatial* yaitu properti data yang mendefinisikan bahwa titik data dalam unit data berhubungan dan tidak dapat dipisahkan (Siddiqui, 2019). Namun, CNN yang digunakan untuk pemrosesan teks seringkali mengabaikan hubungan antar konteks dalam dokumen teks. Hal ini menyebabkan banyak peneliti mengganti metode CNN dengan LSTM. LSTM memiliki arsitektur jaringan *sequential*. *Sequential* adalah sifat data dimana suatu kemunculan data dipengaruhi oleh data sebelumnya. Meskipun LSTM memiliki kelebihan dalam menangani keterkaitan antar konteks di dalam teks, CNN masih lebih baik dari LSTM dalam ekstraksi fitur dari dokumen teks. Oleh karena itu, CNN dan LSTM digabungkan untuk mendapatkan keunggulan masing-masing metode. *Hybrid CNN-LSTM* memberikan solusi untuk menggunakan informasi *spatial* dan karakterisasi *sequential* secara bersamaan. Lapisan CNN pada *hybrid CNN-LSTM* berfungsi untuk ekstraksi fitur pada data *input* dan lapisan LSTM untuk menyimpan informasi jangka waktu lama (Tasdelen, dkk., 2021).

Pada prosesnya, klasifikasi diasumsikan bahwa jumlah dari kelas data adalah seimbang. Namun, sering ditemui permasalahan ketidakseimbangan kelas data pada klasifikasi. Ketidakseimbangan kelas merupakan salah satu faktor yang paling berpengaruh dalam kinerja prediksi suatu klasifikasi. Pengklasifikasi cenderung membuat model pembelajaran yang bias dengan akurasi prediksi yang buruk untuk kelas minoritas dibandingkan dengan kelas mayoritas (Prasetya, 2022). Salah satu metode yang dapat mengatasi ketidakseimbangan data adalah metode *random oversampling*.

Beberapa penelitian bidang kesehatan dengan metode klasifikasi dilakukan oleh Nursyahfitri, dkk pada tahun 2021 menggunakan metode *Decision Tree* terhadap klasifikasi penentuan jenis obat. Penelitian lainnya dilakukan pada tahun 2021 oleh Cahya, dkk. tentang klasifikasi penyakit mata menggunakan metode CNN. Adapun penelitian yang pernah dilakukan peneliti terdahulu menggunakan metode *hybrid CNN-LSTM* adalah penelitian yang dilakukan oleh Zhu, dkk. (2020) dengan menggunakan metode *hybrid CNN-LSTM* pada data aktivitas manusia berdasarkan Radar Mikro-Doppler. Penelitian oleh Rehman, dkk. (2019) melakukan klasifikasi metode *hybrid CNN-LSTM* untuk meningkatkan akurasi analisis sentiment ulasan film, metode *hybrid CNN-LSTM* mendapatkan nilai akurasi terbaik dibandingkan menggunakan metode klasifikasi klasik dan metode klasifikasi lainnya. Penelitian analisis sentimen klasifikasi pada media online dilakukan oleh Hermanto, dkk. (2020) menggunakan algoritma *hybrid LSTM-CNN*, pada penelitian ini dilakukan perbandingan metode LSTM, LSTM-CNN, dan CNN-LSTM. Hasil yang diperoleh menyatakan bahwa metode *hybrid CNN-LSTM* memiliki akurasi paling baik diantara 2 metode lainnya. Penelitian pada tahun 2021 yang dilakukan oleh Tasdelen dan Sen menggunakan metode *hybrid CNN-LSTM* terhadap klasifikasi *pre-miRNA*, hasil penelitian tersebut menunjukkan bahwa metode yang diusulkan berhasil dan mencapai kinerja yang baik untuk klasifikasi *dataset pre-miRNA*.

Berdasarkan pemaparan di atas dan penelitian terdahulu, penulis tertarik untuk melakukan klasifikasi penilaian produk *nutrimax food supplement* menggunakan metode *hybrid CNN-LSTM*.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini yaitu :

1. Membangun model *hybrid* CNN-LSTM.
2. Mengkaji performa klasifikasi *hybrid* CNN-LSTM menggunakan nilai akurasi, presisi, *recall*, dan *f1-score*.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

1. Memberikan wawasan terhadap klasifikasi penilaian produk Nutrimax *food supplement*.
2. Menjadi referensi bagi peneliti selanjutnya dalam menggunakan metode *hybrid* CNN-LSTM maupun tentang klasifikasi terhadap produk-produk lain.
3. Menjadi rujukan bagi masyarakat Indonesia dalam memilih produk vitamin dan obat-obatan yang baik.

II. TINJAUAN PUSTAKA

2.1 Data Mining

Data *mining* merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola statistik, basis data, dan visualisasi untuk penanganan masalah dan pengambilan informasi dari *database* yang besar (Larose, 2005). Berikut merupakan tahapan pemrosesan data *mining* :

1. Seleksi Data

Seleksi data adalah proses pengambilan data yang diseleksi berdasarkan kebutuhan dalam melakukan data *mining*. Seleksi data harus dilakukan sebelum dimulainya tahap data mining. Data terpilih yang digunakan dalam proses penambahan data disimpan dalam penyimpanan yang berbeda dari *database* operasional.

2. *Preprocessing* Data

Preprocessing data merupakan pembersihan data, meliputi membuang duplikasi data, memeriksa data yang hilang, dan membuang atribut yang tidak relevan.

3. Transformasi

Transformasi merupakan teknik mengubah bentuk format data agar sesuai, sebelum melalui proses data *mining*.

4. Data *Mining*

Data *mining* merupakan proses yang dilakukan untuk menemukan pola atau informasi dari data yang telah diseleksi dengan teknik, metode atau algoritma yang sesuai.

5. Evaluasi

Evaluasi merupakan proses menampilkan informasi hasil dari data *mining* dalam bentuk sederhana dan mudah dipahami.

Salah satu jenis data *mining* adalah *text mining*. *Text mining* adalah proses penggalian pola teks berupa informasi yang sebelumnya tidak diketahui dari sejumlah besar sumber data teks (Jo, 2019). Penggalian data teks mirip dengan penggalian data, hanya saja penggalian data dirancang untuk menangani data terstruktur dari *database*, sedangkan penggalian data teks bekerja pada data tidak terstruktur (Choi, dkk., 2020). Penggalian data teks digunakan untuk mengekstrak informasi yang berguna dari dokumen dalam data teks. Informasi berkualitas tinggi didapatkan karena memperhatikan pola dan tren dengan cara mempelajari pola statistik (Deolika, dkk., 2019).

2.2 Klasifikasi

Klasifikasi merupakan suatu proses yang bertujuan untuk menentukan suatu obyek kedalam suatu kelas atau kategori yang sudah ditentukan sebelumnya (Puspitasari, dkk., 2018). Klasifikasi membangun model berdasarkan data pelatihan yang ada dan kemudian menggunakan model tersebut untuk mengklasifikasikan data baru. Klasifikasi dapat didefinisikan sebagai tugas yang melakukan pelatihan/pembelajaran berdasarkan fungsi tujuan yang memetakan setiap kumpulan atribut ke kumpulan label kelas yang tersedia (Utomo dan Mesran, 2020). Terdapat 2 jenis klasifikasi yaitu *binary classification* dan *multiclass classification*. *Binary classification* merupakan klasifikasi yang memisahkan 2 buah kelas, sedangkan *multiclass classification* merupakan klasifikasi lebih dari dua kelas (Tantika, 2022). Beberapa metode yang dapat digunakan dalam melakukan klasifikasi diantaranya, metode *K-Nearest Neighbors*, *Naïve Bayes*, *Support Vector Machine*, *Convolutional Neural Network*, *Long-Short Term Memory*, dan lain-lain.

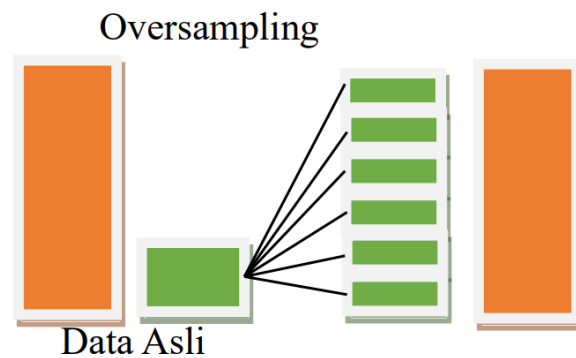
2.3 *Imbalanced Data*

Imbalanced data atau yang biasa dikenal dengan ketidakseimbangan kelas data adalah salah satu permasalahan yang ada pada klasifikasi. Hal ini terjadi pada saat kelas minoritas jauh lebih kecil atau lebih jarang dari kelas mayoritas (Ren et al., 2018). Dalam himpunan data yang tidak seimbang, kelas yang memiliki jumlah data lebih banyak disebut kelas mayoritas, sedangkan kelas yang memiliki jumlah data relatif lebih sedikit disebut kelas minoritas. Ketidakseimbangan kelas dapat berpengaruh pada kualitas data dalam kinerja klasifikasi. Model yang dibuat dengan data yang tidak seimbang memberikan akurasi prediksi yang kurang baik. Informasi dari kelas mayoritas mendominasi kelas minoritas, menyebabkan kendala pengambilan keputusan yang bias dalam klasifikasi (Leevy, dkk., 2018). Metode keseimbangan data dibagi menjadi dua, yaitu metode *random oversampling* dan metode *random undersampling*.

2.3.1 *Random Oversampling*

Random oversampling (ROS) meningkatkan kelas minoritas dengan cara menduplikasi sampel secara acak sehingga jumlah data pada kelas minoritas akan sama dengan kelas mayoritas. *Random oversampling* (ROS) mengacu pada penambahan acak data kelas minoritas ke data pelatihan. Proses penjumlahan ini dilakukan berulang-ulang hingga banyaknya data pada kelas minoritas sama dengan jumlah data pada kelas mayoritas (Fitriani, dkk., 2021). Langkah Pertama, hitung selisih antara kelas mayoritas dan kelas minoritas. Kemudian, lakukan perulangan sebanyak hasil penghitungan selisih sambil membaca data kelas minoritas secara acak dan ditambahkan ke data pelatihan (Syukron dan Subekti, 2018).

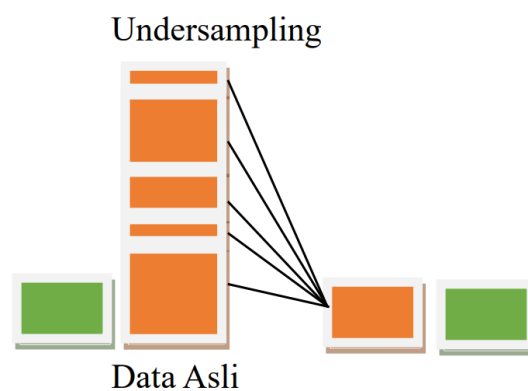
Berikut merupakan grafik dari *Random Oversampling* :



Gambar 1. Grafik *Random Oversampling*.

2.3.2 *Random Undersampling*

Random Undersampling (RUS) mengurangi kelas mayoritas dengan cara menghilangkan data secara acak sehingga jumlah sampel pada kelas mayoritas akan sama dengan kelas minoritas. *Random undersampling* mengacu pada pengurangan acak data kelas mayoritas ke data pelatihan. Menghilangkan sampel pada kelas mayoritas merupakan hal yang baik untuk menghindari kemungkinan terjadinya *noise* yang dapat memperburuk hasil klasifikasi (Prasetya, 2022). Berikut merupakan grafik dari *Random Undersampling* :



Gambar 2. Grafik *Random Undersampling*.

2.4 *Word Embedding*

Ulasan dalam bentuk teks tidak dapat dikenali langsung oleh komputer, maka diperlukan konversi data teks menjadi bentuk numerik seperti vektor. Menurut Sabrila, dkk. (2021), *Word embedding* adalah metode yang digunakan untuk mendeskripsikan kata-kata ke dalam bentuk vektor. *Word embedding* memiliki beberapa keunggulan dibandingkan dengan metode *Bag of Word* (BOW) maupun TF-IDF, termasuk kemampuan untuk mengurangi dimensi fitur dan menangkap hubungan semantik dari kata-kata yang terdapat dalam kumpulan dokumen.

Terdapat beberapa cara yang dapat membentuk representasi kata ke dalam bentuk vektor yaitu, *word2vec*, *glove*, dan *fasttext*. Pada tahun 2013, *Word2vec* diperkenalkan oleh Mikolov dkk. dengan dua metode utamanya yaitu *Skip-gram* dan *Continuous Bag of Words* (CBOW). Kemudian pada tahun berikutnya, Pennington dkk, memperkenalkan *word embedding* baru yaitu *Glove* yang menggunakan rasio *co-occurrence probability* antarkata. Pada tahun 2017, Bojanowski dkk, mengembangkan model *word2vec* dan memperkenalkan *FastText* yang mempelajari informasi *subword* dari kata (Bojanowski, dkk., 2017).

2.4.1 *Word2Vec*

Word2Vec merupakan metode *Word Embedding* yang digunakan untuk merepresentasikan sebuah kata pada bentuk vektor. *Word2Vec* telah terbukti membawa makna semantik dan berguna dalam berbagai tugas untuk mengoptimalkan sebuah data dan menghasilkan vektor kata yang mirip sehingga memudahkan dalam pengenalan analisis data (Naufal, dkk., 2021). Tahapan *word embedding* pada penelitian ini menggunakan *Word2Vec*. Corpus atau kamus dari kumpulan teks besar yang digunakan adalah corpus yang disediakan oleh Wiki untuk *word embedding* dalam bahasa Indonesia. *Word2Vec* adalah model populer yang terdiri dari dua jenis, yaitu *Continuous Bag-of-Words* (CBOW) dan *Skip-*

Gram. Perbedaan antara *Skip-Gram* dan CBOW terletak pada arsitektur jaringan saraf dan bagaimana setiap jenis model dibangun. *Skip-Gram* memasukkan kata dan meminta jaringan saraf untuk menebak kata-kata yang muncul di sekitar kata input. Sedangkan pada arsitektur CBOW, jaringan saraf digunakan untuk menebak kata, dengan input kata di sekitar kata yang ditebak (Santoso, dkk., 2020).

2.5 Fungsi Aktivasi

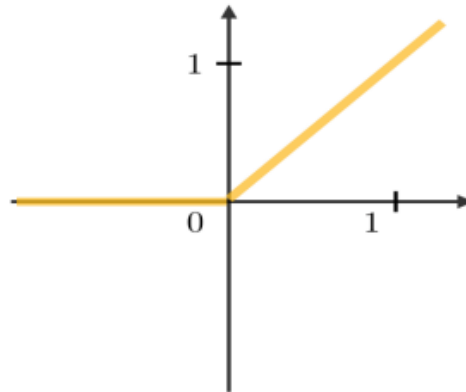
Fungsi aktivasi adalah suatu fungsi yang dapat mentransformasikan suatu *input* menjadi *output* tertentu (Sitepu, dkk., 2021). Fungsi aktivasi diterapkan ke setiap neuron dalam jaringan dan dapat menentukan apakah neuron harus diaktifkan atau tidak berdasarkan pentingnya *input* setiap neuron terhadap prediksi model. Pada jaringan saraf tiruan terdapat beberapa fungsi aktivasi seperti fungsi aktivasi ReLU, sigmoid, *softmax*, tangen hiperbolik (Tanh), dan lain sebagainya.

2.5.1 Fungsi Aktivasi *Rectified Linear Unit* (ReLU)

Fungsi aktivasi *Rectified Linear Unit* (ReLU) sering digunakan oleh para peneliti karena berfungsi lebih baik dan digunakan pada setiap *hidden layer* dari *neural network*. Fungsi aktivasi *Rectified Linear Unit* (ReLU) merupakan lapisan aktivasi dimana ReLU pada intinya hanya membuat pembatas pada bilangan nol, apabila $x \leq 0$ maka $x = 0$ dan apabila $x > 0$ maka $x = x$ (Sitepu, dkk., 2021). Fungsi ReLU ditunjukkan oleh persamaan berikut :

$$f(x) = \max(0, x) \quad (2.1)$$

Berikut merupakan grafik dari fungsi ReLU :



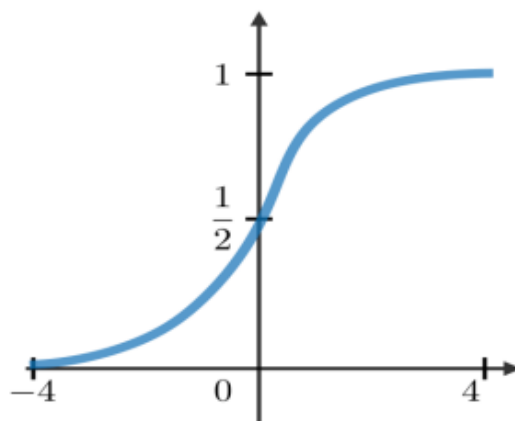
Gambar 3. Grafik Fungsi Aktivasi ReLU. (Sumber: Amidi dan Amidi., 2019)

2.5.2 Fungsi Aktivasi Sigmoid

Fungsi aktivasi sigmoid adalah fungsi *nonlinier* yang menjadikan *output value* dengan *range* antara 0 hingga 1. Fungsi aktivasi sigmoid akan menghasilkan kurva dengan bentuk "S". Fungsi Sigmoid ditunjukkan oleh persamaan berikut :

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (2.2)$$

Berikut merupakan grafik dari fungsi Sigmoid :



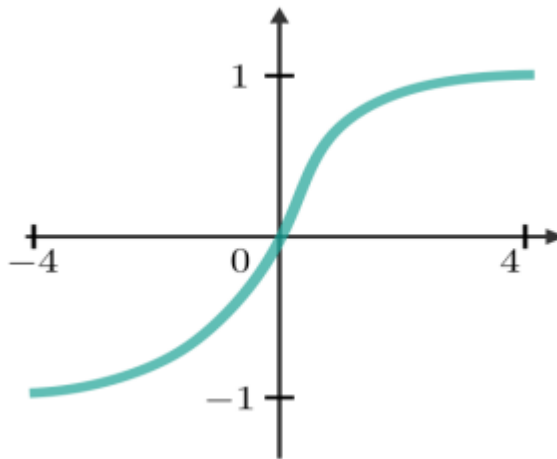
Gambar 4. Grafik Fungsi Aktivasi Sigmoid. (Sumber: Amidi dan Amidi., 2019)

2.5.3 Fungsi Aktivasi *Tangent Hyperbolic* (TANH)

Fungsi *Tangent hyperbolic* nilainya terletak pada kisaran -1 hingga 1. Dibandingkan dengan fungsi sigmoid, gradien fungsi tanh lebih curam. Fungsi tanh ditunjukkan oleh persamaan berikut :

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.3)$$

Berikut merupakan grafik dari fungsi Tanh :



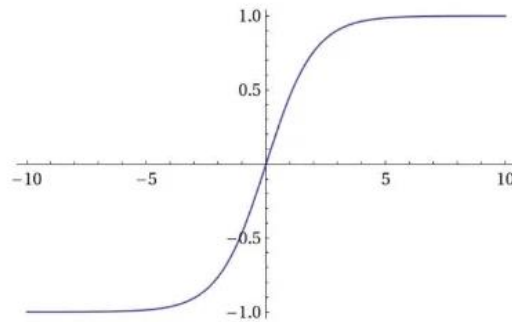
Gambar 5. Grafik Fungsi Aktivasi Tanh. (Sumber: Amidi dan Amidi., 2019)

2.5.4 Fungsi Aktivasi *Softmax*

Fungsi *softmax* mengubah nilai x menjadi sebuah nilai *output* yang memiliki *range* mulai dari 0 sampai 1 dan jumlah semua probabilitasnya akan sama dengan satu. Fungsi *softmax* digunakan untuk masalah klasifikasi *multiclass* yang dapat mengembalikan sebuah peluang dari masing-masing kelas dan kelas target yang memiliki probabilitas yang tinggi. Fungsi *softmax* ditunjukkan oleh persamaan berikut :

$$p(x_i) = \frac{e^{x_i}}{\sum_{k=1}^k e^{x_j}} \quad (2.4)$$

Berikut merupakan grafik dari fungsi softmax :



Gambar 6. Grafik Fungsi Aktivasi *Softmax*. (Sumber: Bag., 2021)

2.6 *Deep Learning*

Deep Learning merupakan cabang dari *Machine Learning* berbasis Jaringan Syaraf Tiruan atau pembaharuan JST dalam melatih komputer untuk dapat berpikir dan bertindak seperti manusia dalam mengambil keputusan (Peryanto, dkk., 2019). *Deep Learning* pada komputer diklasifikasikan sebagai pencarian gambar, suara, teks, atau video. Seperti dalam komputer, kumpulan data dalam jumlah besar digunakan untuk pengujian dan pelatihan, sebagai representasi internal atau *feature vector*. *Deep learning* merupakan pembelajaran representasi untuk membentuk arsitektur jaringan syaraf tiruan dengan *multilayer* yaitu *input layer*, *hidden layer*, dan *output layer*. *Hidden layer* dengan lapisan yang banyak berfungsi untuk membuat komposisi algoritma dalam hal meminimalisir nilai *error* pada *output* (Yanto, dkk., 2021).

2.6.1 *Convolution Neural Network (CNN)*

Convolutional Neural Network (CNN) adalah metode yang paling populer dan banyak diminati oleh peneliti karena CNN merupakan jaringan saraf tiruan yang digunakan untuk mengenali objek pada citra digital. CNN dapat menangani data berdimensi tinggi seperti gambar. Namun, penggunaannya tidak terbatas pada

gambar, tetapi dapat mencapai hasil dalam pemrosesan data teks dan pengenalan suara. (Hermanto, dkk., 2021).

Menurut Hidayat dan Hermawan (2018), tahapan yang akan dilakukan pada proses CNN dari awal masukan hingga akhir yaitu *convolution layer*, *pooling layer* dan *fully connected layer*.

a. *Convolution Layer*

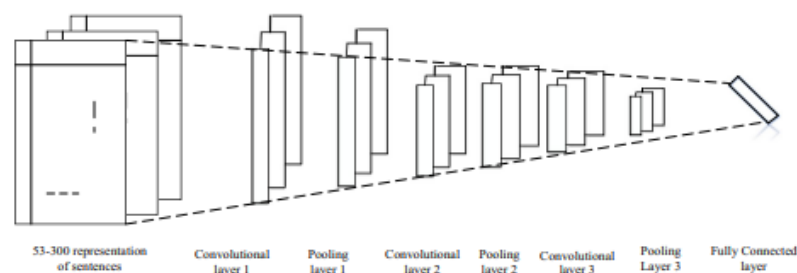
Pada *convolution layer* operasi konvolusi dilakukan antara matriks masukan dengan matriks kernel. Tujuan dari konvolusi adalah untuk mengekstrak fitur dari data. Konvolusi menciptakan transformasi *linier* dari data input sesuai dengan informasi *spatial* data. Bobot dari layer menentukan *convolutional* kernel yang akan digunakan, sehingga kernel konvolusi dapat dilatih pada input CNN.

b. *Pooling Layer*

Pooling layer merupakan layer ekstraksi setelah *convolution layer* dengan prinsipnya yaitu filter yang memiliki ukuran dan *stride* tertentu bergeser pada seluruh area *feature map*. Ide utama *pooling layer* adalah untuk mengurangi kompleksitas pada layer selanjutnya. *Pooling* yang umum digunakan adalah *max pooling*, yaitu mengambil nilai maksimum dari setiap *grid* untuk menyusun matriks baru.

c. *Fully Connected Layer*

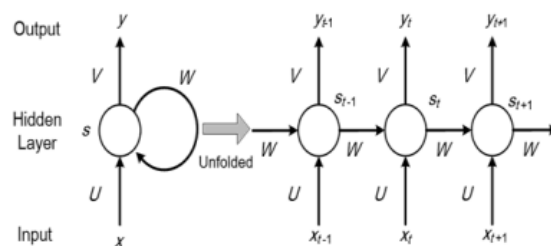
Fully Connected Layer merupakan lapisan dimana semua *neuron* dari lapisan-lapisan sebelumnya dijadikan satu untuk dilakukan klasifikasi menggunakan *neural network*. Ilustrasi CNN secara umum dapat dilihat pada arsitektur berikut:



Gambar 7. Arsitektur CNN (Sumber: Rehman, dkk., 2019)

2.6.2 Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) merupakan arsitektur jaringan saraf tiruan yang dilakukan pemrosesan secara berulang untuk memproses data masukan yang bersifat *sequential* (Afandi, dkk., 2022). Namun, jaringan saraf tiruan tidak dapat menyimpan informasi dalam waktu yang panjang. RNN dapat mengirimkan informasi sebelumnya ke informasi selanjutnya dan sering diterapkan pada klasifikasi gambar, keterangan gambar, pengenalan suara, prediksi saham, dan pemrosesan bahasa alami (Hanifa, dkk., 2021). Ilustrasi RNN secara umum dapat dilihat pada gambaran arsitektur berikut:



Gambar 8. Arsitektur RNN (Sumber: Firmansyah, dkk., 2020)

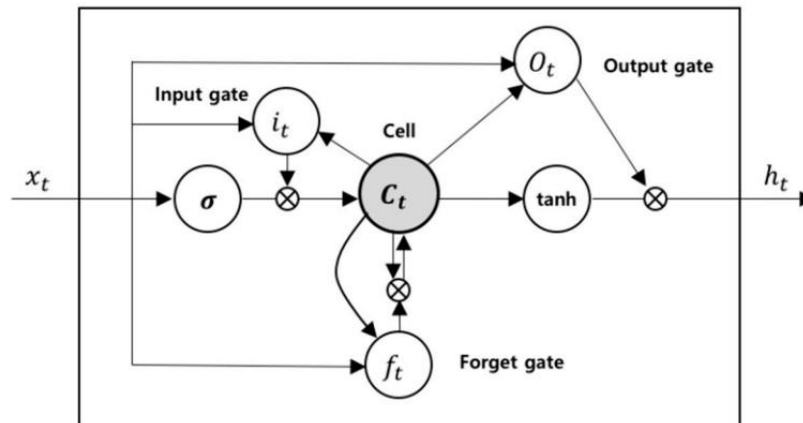
Berdasarkan Gambar 8. x_t adalah input tiap waktu, S_t merupakan *hidden state* yang digunakan untuk menyimpan perhitungan yang dilakukan. Adapun Informasi sebelumnya untuk setiap t dikendalikan oleh bobot W .

2.6.3 Long-Short Term Memory (LSTM)

Long-Short Term Memory (LSTM) pertama kali diperkenalkan pada tahun 1997 oleh Hochreiter dan Schmidhuber, lalu ditingkatkan oleh Graves pada tahun 2013. LSTM merupakan versi terbaru yang lebih efisien dan memiliki keunggulan lebih dari RNN. Dalam LSTM, analisis dapat dilakukan meskipun terdapat jarak antar kata. Pada LSTM, nilai *error* pada proses *backpropagation* dari waktu ke waktu

akan disimpan, sehingga memungkinkan LSTM untuk menyimpan lebih dari 1000 langkah waktu informasi (Antonio, dkk., 2021).

Berikut merupakan Layer Pada Setiap *Cell* LSTM:



Gambar 9. Arsitektur LSTM (Sumber: Chung & Shin, 2018).

LSTM memiliki Jaringan dengan ukuran besar, pada jaringan ini terdapat sel-sel LSTM yang menggantikan letak *hidden layer* pada RNN, yang bertujuan untuk menyimpan teks sebelumnya. Terdapat 3 jenis *gate* pada LSTM diantaranya adalah *forget gate*, *input gate*, dan *output gate*.

a. *Forget Gate* (f_t)

Pada gerbang ini, nilai *output* sebelumnya digabungkan dengan *input* saat ini kemudian melewati fungsi aktivasi sigmoid. Gerbang ini menentukan apakah akan melupakan data sebelumnya atau tidak. Informasi ini kemudian ditransfer ke *memory cell* atau *cell state*. Persamaan *forget gate* ditunjukkan pada persamaan berikut:

$$f_t = \sigma(W_f \times [x_t, h_{t-1}] + b_f) \quad (2.5)$$

dengan:

- f_t : *forget gate*
- W_f : *weights forget gate*
- h_{t-1} : *output cell previously*
- σ : *sigmoid activation function*

x_t : *input cell*
 bf : *bias forget gate*

b. *Input Gate (i_t)*

Pada gerbang ini, nilai *input* sebelumnya digabungkan dengan *input* saat ini, sehingga melewati dua fungsi aktivasi. Baris pertama melewati fungsi aktivasi sigmoid dari nilai *input*, baris kedua melewati fungsi aktivasi tanh dari nilai *candidate memory cell*. Persamaan *input gate* ditunjukkan pada persamaan berikut:

$$i_t = \sigma(W_i \times [x_t, h_{t-1}] + b_i) \quad (2.6)$$

$$C^{\sim}t = \tanh(W_C \times [x_t, h_{t-1}] + b_C) \quad (2.7)$$

dengan:

I_t : *input gate*
 $C^{\sim}t$: *candidate*
 W_C : *weights candidate*
 W_i : *weights input*
 \tanh : *tanh activation function*
 b_C : *bias candidate*
 b_i : *bias input gate*

Selanjutnya akan dieksekusi apa yang telah diputuskan pada tahap sebelumnya. Vektor *candidate* baru dibuat melalui lapisan tanh agar dapat mengontrol berapa banyak informasi baru. Persamaan *cell state* ditunjukkan pada persamaan berikut:

$$C_t = f_t \times C_{t-1} + i_t \times C^{\sim}t \quad (2.8)$$

dengan :

C_t : *cell state*
 f_t : *forget gate*
 C_{t-1} : *cell state previously*
 i_t : *input gate*
 $C^{\sim}t$: *candidate*

c. *Output gate (ot)*

Gerbang ini menghasilkan nilai *output* dimana nilai tersebut dihasilkan dari kombinasi nilai sebelumnya dengan nilai saat ini yang telah dilewatkan melalui fungsi aktivasi sigmoid. Persamaan *output gate* ditunjukkan pada persamaan berikut:

$$ot = \sigma(Wo \times [xt, ht - 1] + bo) \quad (2.9)$$

dengan:

- ot : *output gate*
- σ : *sigmoid activation function*
- Wo : *weights output gate*
- xt : *input cell*
- $ht - 1$: *output cell previously*
- bo : *bias output gate*

Setelah mendapatkan *output* dari *output gate* maka *cell state* dilakukan operasi tanh. Lalu dikalikan *output gate* dengan sigmoid layer. Persamaan nilai *output* orde ke t ditunjukkan pada persamaan berikut:

$$ht = ot \times \tanh(C_t) \quad (2.10)$$

dengan:

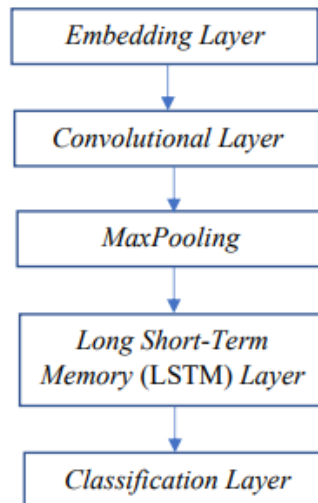
- ht : nilai output orde ke t
- \tanh : *tanh activation function*
- ot : *output gate*
- ct : *cell state*

2.6.4 *Hybrid CNN-LSTM*

Convolutional Neural Network (CNN) dan *Long-Short Term Memory (LSTM)* adalah dua buah metode yang populer digunakan untuk menyelesaikan masalah klasifikasi teks. CNN memiliki kelebihan dalam proses ekstraksi fitur atau mengambil fitur di dalam dokumen teks. Sedangkan LSTM memiliki kelebihan untuk memproses keterkaitan antar konteks di dalam teks dan dapat menyimpan

informasi dalam jangka waktu panjang. Oleh karena itu, dikombinasikan metode *hybrid* CNN dan LSTM untuk memperoleh kelebihan masing-masing metode (Widhiyasana, dkk., 2021).

Berikut merupakan Alur Metode *Hybrid* CNN-LSTM



Gambar 10. Alur Metode *Hybrid* CNN-LSTM.

Hybrid CNN-LSTM merupakan kombinasi antara algoritma CNN dengan algoritma LSTM. Seperti namanya, membangun model CNN-LSTM menempatkan algoritma LSTM pada lapisan terakhir model jaringan CNN, sehingga sebelum masuk ke jaringan LSTM, vektor *input* harus melalui proses algoritma CNN. Pada tahap CNN diterapkan dua tahapan yaitu proses konvolusi dan proses *pooling*, setelah dilakukan konvolusi dan *pooling*, model langsung diteruskan ke lapisan LSTM untuk dilakukan klasifikasi.

2.7 *Hyperparameter*

Memilih parameter yang tepat pada model dapat meningkatkan hasil akurasi pada klasifikasi. *Hyperparameter* merupakan parameter pada model *neural network* yang telah ditentukan di awal proses *training*. Parameter yang digunakan adalah jumlah *neuron*, *batch size*, *dropout*, dan *epochs*. *Batch size* merupakan jumlah total

sampel data *training* dalam 1 iterasi. *Dropout* merupakan teknik memilih beberapa *neuron* secara acak untuk di *nonaktifkan* saat proses *training* untuk mencegah *overfitting*. *Epochs* merupakan parameter yang menentukan berapa kali algoritma *deep learning* bekerja melewati seluruh *dataset*.

Hypertuning merupakan metode yang digunakan untuk mencari parameter yang tepat dan dapat meningkatkan kinerja model dengan mencoba kombinasi *hyperparameter* yang diberikan. *Hypertuning* dapat menentukan parameter *dropout*, *epochs*, dan *batch size* yang tepat untuk model yang diusulkan (Minarno, dkk., 2021).

2.8 Evaluasi Model

Menurut Nurvania, dkk. (2021), *Confusion matrix* adalah tabel atau matriks yang berisi empat nilai yang merupakan ukuran kinerja dari performa klasifikasi yang dilakukan. *Confusion matrix* memiliki empat nilai atau poin yaitu *true positive*, *false positive*, *false negative*, dan *true negative*. Gambaran *Confusion matrix* ditunjukkan pada tabel berikut:

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 11. Gambaran *Confusion Matrix* (Sumber : Hidayatullah, dkk., 2019)

2.8.1 Akurasi

Akurasi adalah perhitungan seberapa akurat pengklasifikasi yang dibangun terhadap target yang ada. Akurasi secara matematis dapat dinyatakan dalam persamaan berikut :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.10)$$

2.8.2 Presisi

Presisi adalah perhitungan keakuratan antara data aktual dengan hasil prediksi dari model. Presisi secara matematis dapat dinyatakan dalam persamaan berikut :

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (2.11)$$

2.8.3 Recall

Recall adalah perhitungan yang menggambarkan keberhasilan model dalam mengingat informasi. *Recall* secara matematis dapat dinyatakan dalam persamaan berikut:

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (2.12)$$

2.8.4 *F1-Score*

F1-Score adalah perhitungan yang menggambarkan perbandingan antara presisi dan *recall*. Jika nilai FN dan FP tidak dekat, *f1-Score* harus digunakan sebagai ganti nilai presisi. *F1-Score* secara matematis dapat dinyatakan dalam persamaan berikut:

$$F1 - Score = \frac{2 \times recall \times presisi}{(recall + presisi)} \times 100\% \quad (2.13)$$

III. METODE PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester genap tahun akademik 2022/2023. Bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Spesifikasi Perangkat

Perangkat yang digunakan untuk membantu penelitian ini adalah laptop dengan merek Dell model Latitude 3400 dengan type x64-based PC. Spesifikasi *hardware* yang digunakan adalah sebagai berikut:

Processor	: Intel® Core™ i3-8145U CPU @2.10GHz
Memory	: SSD 512 GB
RAM	: 8 GB

3.3 Data Penelitian

Data yang digunakan pada penelitian ini diambil dari hasil *scrapping* penilaian produk Nutrimax *food supplement* pada *e-commerce* Shopee <https://shopee.co.id/nutrimaxonline>. Data yang digunakan berupa data ulasan konsumen terhadap produk Nutrimax *food supplement* dan rating penilaian 1-5.

Tabel 1. Data Penilaian Produk Nutrimax *Food Supplement*

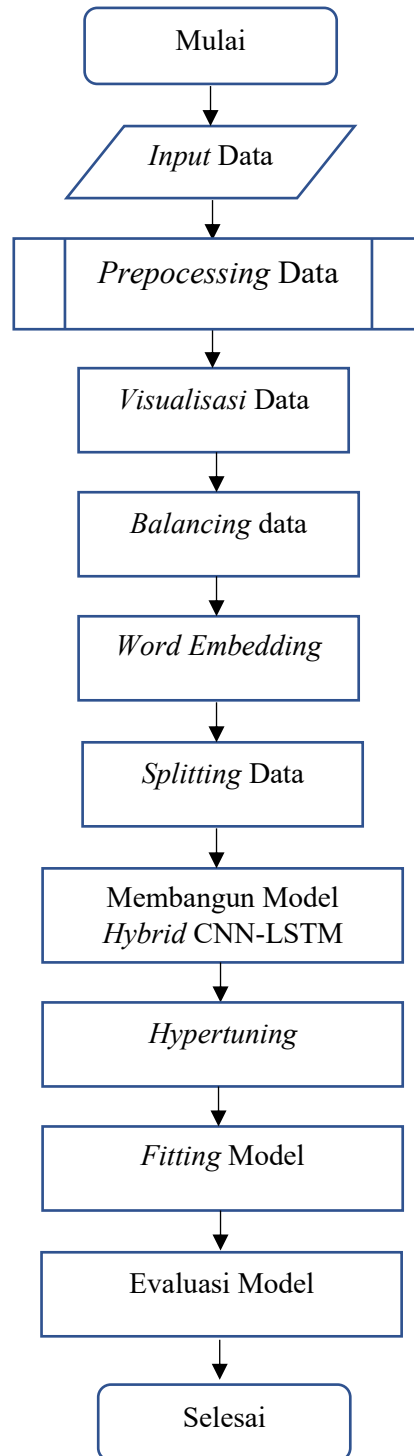
No	Ulasan	Rating
1.	<p>Khasiat: menjaga imun tubuh</p> <p>Rasa: cocok buat penderita asam lambung</p> <p>Harga: oke</p> <p>Vitamin c yg cocok banget buat penderita asam lambung.</p> <p>Pernah beli di drugstore, dan harga ini disaat promo.</p> <p>Ternyata klo normal, jauh banget ya di drugstore</p>	5
2.	<p>Khasiat: Baru mau coba</p> <p>Rasa: Belum tau</p> <p>Harga: Lumayan</p> <p>Alhamdulillah barang udah sampe dan sesuai dengan pesanan termasuk lama banget terimakasih</p>	4
3.	<p>Khasiat: belum tau</p> <p>Rasa: gak ada rasa</p> <p>Harga: standar</p> <p>Maaf ya, kok beda warna sama saya beli di apotek.. yang beli di sini warna lebih pucat terus kok itu pada patah isinya gompal2 gt... Jadi meragukan keasliannya ini huhuhu</p>	3
4.	<p>Pas mau beli lagi ternyata harganya naik, yang nutrimax c kids belum di refund</p>	2
⋮	⋮	⋮
799.	<p>Aku kira softgell, soalnya lagi nyarik vitamin softgell</p> <p>..Jadi kurang memuaskan</p>	1

3.4 Metode Penelitian

Dalam penelitian ini, peneliti akan menampilkan akurasi terbaik metode *hybrid* CNN-LSTM dalam klasifikasi penilaian produk *nutrimax food supplement* dengan bantuan *software Python* pada *Google Colab*. Adapun langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Input data penilaian produk *Nutrimax food supplement* berlabel 1-5, dengan label 1 adalah penilaian yang buruk sampai dengan label 5 penilaian yang baik.
2. *Preprocessing* data dengan *case folding*, *punctuation mark removal*, *stopword removal*, *tokenisasi* dan *stemming*.
3. Visualisasi data dengan membuat diagram batang untuk melihat perbandingan jumlah penilaian rating 1-5.
4. *Balancing* data menggunakan metode *Random Oversampling* (ROS).
5. *Word embedding* menggunakan *Word2Vec*.
6. Melakukan *splitting* data yaitu data *training* dan data *testing* dengan pembagian sebesar 80% data *training* 20% data *testing* dan 90% data *training* 10% data *testing*.
7. Membangun model *hybrid* CNN-LSTM untuk dilakukan *hypertuning*.
8. *Hypertuning* untuk mendapatkan kombinasi parameter terbaik pada model yang dibangun.
9. *Fitting* model *hybrid* CNN-LSTM sesuai dengan hasil *hypertuning*.
10. Mengevaluasi kinerja metode *hybrid* CNN-LSTM menggunakan akurasi, *presisi*, *recall*, dan *f1-score* terhadap model yang telah dibangun.

Di bawah ini merupakan *flowchart* dari metode *hybrid* CNN-LSTM:



Gambar 12. *Flowchart* Penelitian.

V. KESIMPULAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang sudah dipaparkan pada bab sebelumnya, dapat diambil kesimpulan sebagai berikut :

1. Klasifikasi penilaian produk *nutrimax food supplement* menggunakan metode *hybrid* CNN-LSTM dengan bantuan *hypertuning* untuk menentukan parameter terbaik pada model 80% data *training* dan 20% data *testing* didapatkan parameter hasil *hypertuning* yaitu *epoch* 200, *batch size* 64, dan *dropout* 0.3. Sedangkan untuk 90% data *training* dan 10% data *testing* didapatkan parameter hasil *hypertuning* yaitu *epoch* 200, *batch size* 32, dan *dropout* 0.3.
2. Hasil klasifikasi penilaian produk *nutrimax food supplement* menggunakan metode *hybrid* CNN-LSTM dengan 80% data *training* dan 20% data *testing* mendapatkan nilai akurasi sebesar 0.988, presisi sebesar 0.988, *recall* sebesar 0.99 dan *f1-score* sebesar 0.988. Sedangkan untuk 90% data *training* dan 10% data *testing* mendapatkan nilai akurasi sebesar 0.994, presisi sebesar 0.994, *recall* sebesar 0.994, dan *f1-score* sebesar 0.996. Berdasarkan hasil akurasi yang didapatkan, disimpulkan bahwa semakin banyak data yang dilatih, maka akan lebih baik hasil klasifikasi yang didapatkan.

DAFTAR PUSTAKA

- Afandi, A., Saputro, S. N., Kusumaningrum, A. M., Ardiansyah, H., Kafabi, M. H., dan Sudianto. 2022. Klasifikasi Judul Berita Clickbait Menggunakan RNN-LSTM. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*. **7**(2): 85-89.
- Amidi, A. dan Amidi, S. 2019. VIP Cheatsheet : Recurrent Neural Network. *CS 230- Deep Learning*. 1-5
- Antonio, J., dan Rifqi, M. 2021. Analisis Sentimen Terhadap Ulasan pada Aplikasi Provider Internet di Indonesia Menggunakan Algoritma Long-Short Term Memory. *Jurnal Resti*. **5**(2): 1-9.
- Bag, Sukanya. Activation Functions All You Need To Know. <https://medium.com/analytics-vidhya/activation-functions-all-you-need-to-know-355a850d025e>. Diakses pada 06 Maret 2023
- Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. 2017. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*. **10**(5): 135-146.
- Cahya, F.N., ardi, N., Riana, D., dan Hardianti, R. 2021. Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network (CNN). *Sistemasi: Jurnal Sistem Informatika*. **10**(3): 618-626.
- Choi, Y. J., Jeon, B. J., and Kim, H. W. 2021. Identification of Key Cyberbullies: A text Mining and Social Network Analysis Approach. *Telematics and Informatics*. **56**

- Chung, H., and Shin, K. S. 2019. Genetic Algorithm-Optimized Long Short – Term Memory Network for Stock Market Prediction. *Sustainability Journal*. **10**(10): 1-18
- Deolika, A., Kusriani., Luthfi, E. H. 2019. Analisis Pembobotan Kata pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi*. **3**(2): 179-184.
- Firmansyah. M, R., Ilyas. R., Kasyidi. F. 2020. Klasifikasi Kalimat Ilmiah Menggunakan Recurrent Neural Network. *Prosiding The 11th Industrial Research Workshop and National Seminar*. Indonesia, Bandung.
- Fitriani, R. D., Yasin, H., dan Tarno. 2021. Penanganan Klasifikasi Kelas Data Tidak Seimbang dengan Random Oversampling pada Naïve Bayes (Studi Kasus : Status Peserta Kb IUD di Kabupaten Kendal). *Jurnal Gaussian*. **10**(1): 11-20.
- Hanifa, A., Fauzan, S. A., Hikal, M., dan Ashfiya, M. B. 2021. Perbandingan Metode LSTM dan GRU (RNN) untuk Klasifikasi Berita Palsu Berbahasa Indonesia. *Dinamika Rekayasa*. **17**(1): 25-34.
- Hermanto, D. T., Setyanto, A., dan Luthfi, E, T. 2021. Algoritma LSTM-CNN untuk Sentimen Klasifikasi dengan Word2vec pada Media Online. *Citec Journal*. **8**(1) : 65-77.
- Hidayat, B., dan Hermawan. G. 2018. Deteksi Hama Pada Daun Teh Dengan Metode Convolutional Neural Network(CNN). *Jurnal Linguistik Komputasional*. **5**(12): 112-116.
- Hidayatullah, A. F., Yusuf, A. A.F., Juwairi, K. P., dan Nayoan, R. A. N. 2019. Identifikasi Konten Kasar pada tweet Bahasa Indonesia. *Jurnal Linguistik Komputasional*. **2**(1): 1-5.
- Jo, T. 2019. *Text Mining Concepts, Implementastion, and Big Data Challenge*. Hongik University, Seoul, Korea.
- Jovita, N., Jondri., dan Lhaksamana, K. M. 2021. Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short Term Memory (LSTM), pp. 4127-4128. *Proceeding of engineering*. Indonesia, Bandung.

- Larose, D. T. 2005. *Discovering Knowledge in Data : An Introduction to Data mining*. John Willey and Sons, Inc. New York.
- Leevy, J. L., Khoshgoftaar, T. M., Bauder, R. A., and Seliya. N. 2018. A Survey on Addressing High-class Imbalance in Big Data. *Journal of Big Data*. 5(42): 1-30.
- Minarno, A. E., Mandiri, M. H. C., dan Alfarizy, M. R. 2021. Klasifikasi Covid-19 menggunakan Filter Gabor dan CNN dengan Hyperparameter Tuning. *EKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, dan Teknik Elektronika*. 9(3): 493-504.
- Mulyana, S. 2021. Pengaruh Harga dan Ulasan Produk Terhadap Keputusan Pembelian Produk Fashion Secara Online Pada Shopee di Pekanbaru. *Jurnal Saya Saing*. 7(2) : 185-195.
- Naufal, H. F., dan Setiawan, E. R. 2021. Ekspansi Fitur Pada Analisis Sentimen Twitter dengan Pendekatan Metode Word2Vec, hlm. 10339-10349. *E-Proceeding of Engineering*. Indonesia, Bandung.
- Nursyahfitri, R., Maharadja, A.N., Farissa, R.A., dan Umaidah. Y. 2021. Klasifikasi Penentuan Jenis Obat Menggunakan Algoritma Decision Tree. *JIP (Jurnal Informatika Polinema)*. 7(3): 53-59.
- Peryanto, A., Yudhana, A., dan Umar Rusydi. 2019. Rancang Bangun Klasifikasi Citra dengan Teknologi Deep Learning Berbasis Metode Convolutional Neural Network. *Jurnal Format*. 8(2): 138-147.
- Prasetya, J. 2022. Penerapan Klasifikasi Naïve Bayes dengan Algoritma Rando Oversampling dan Random Undersampling pada Data Tidak Seimbang Cervical cancer Risk Factors. *Leibnis : Jurnal Matematika*. 2(2): 13-18.
- PT. Suryaprana Nutrisindo. 2020. *Nutrimax food supplement. Nutrimax Food Supplement: Nature's Way for a Better Living*. Diakses pada 24 Oktober 2022.

- Puspitasari, A. M., Ratnawati, D. E., dan Widodo, A.W. 2018. Klasifikasi Penyakit Gigi dan Mulut menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 2(2): 802-810.
- Rasywir, E., Sinaga, R., dan Pratama, Y. 2020. Analisis dan Implementasi Diagnosis Penyakit Sawit dengan Metode Convolutional Neural Network(CNN). *Jurnal Informatika dan Komputer*. 2(22): 117-123.
- Rehman, A. U., Malik, A. K., Raza. B., and Ali. W. 2019. A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis. *Multimedia Tools and Applications*. 78(18): 26597-26613
- Ren, F., Cao, P., Li, W., Zhao, D., dan Zaiane, O. 2018. Ensemble based adaptive over-sampling method for imbalanced data learning in computer aided detection of microaneurysm. *Computerized Medical Imaging and Graphics*, 55: 54–67.
- Rozi, I. F., Wijyaningrum, V. N., dan Khozin, N. 2020. Klasifikasi Teks Laporan Masyarakat pada Situs Laporkan! Menggunakan Recurrent Neural Network. *Temasi (Jurnal Sistem Informasi)*. 9(3): 635-637.
- Sabrila, T. S., Sari, V. R., dan Minarno, A. E. 2021. Analisis Sentimen pada Tweet tentang Penanganan Covid-19 menggunakan Word Embedding pada Algoritma Support Vector Machine dan K-Nearest Neighbor. *Fountain of Informatics Journal*. 6(2): 69-75.
- Sagala, L. O. A. S. 2022. Klasifikasi Cats dan Dogs dengan Metode CNN dalam Fungsi Aktivasi ReLU, Sigmoid, Softmax, Softplus, Softsign dan SeLU. Institut Teknologi Bandung. https://www.researchgate.net/publication/363924413_Klasifikasi_Cats_dan_Dogs
- Santoso, J., Soetioo, A. D. B., Gunawan., Setyati. E., Yuniarno, E.M., Hariadi. M., dan Purnomo, M.H. 2018. Self-Training Naïve Bayes Berbasis Word2Vec untuk Kategorisasi Berita Bahasa Indonesia. *JNTETI*. 7(2): 158-166.
- Sharma, S., Sharma. S. 2017. Activation Functions in Neural Networks. *International Journal of Engineering Applied Sciences and Technology*. 4(12): 310-314.

- Siddiqui, A. A. 2019. Spatial Relationship in CNN. What does 'spatial relationship' in CNN mean? - Quora. Diakses pada 25 Oktober 2022.
- Sitepu, A. C., dan Sgiro, M. 2021. Analisis Fungsi Aktivasi ReLu dan Sigmoid Menggunakan Optimizer SGD dengan Representasi MSE pada Model Backpropagation. *JUTISAL (Jurnal Teknik Informatika Komputer Universal)*. **1(1)**: 12-25.
- Syukron, A., dan Subkti. A. 2018. Penerapan Metode Random ver-Under Sampling dan Random Forest untuk Klasifikasi Penilaian Kredit. *Jurnal Informatika*. **5(2)**: 175-185.
- Tantika, R. S. 2022. Penggunaan Metode Support Vector Machine Klasifikasi Multiclass pada Data Pasien Penyakit Tiroid. *Bandung Conference Series: Statistics*. **2(2)**: 159-166.
- Tasdelen, A., and Sen, B. 2021. A Hybrid CNN-LSTM Model for Pre-miRNA Classification. *Scientific Reports*. [A hybrid CNN-LSTM model for pre-miRNA classification | Scientific Reports \(nature.com\)](#)
- Utomo, D. P., dan Mesran. 2020. Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*. **4(2)**: 437-444.
- Wibawa, A.P., Purnama, M. G., Akbar, M. F., dan Dwiyanto, F. A. 2018. Metode-metode Klasifikasi, pp 134-138. Prosiding Seminar Ilmu Komputer dan Teknologi Informasi. Indonesia, Malang.
- Widhiyasana, Y., Semiawan. T., Mudzakir, I. G., dan Noor, M. R. 2021. Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*. **10(4)**: 355-356.
- Yanto, B., Fimawahin, L., Supriyanto, A., Hayadi, B. H., dan Pratama, R. R. 2021. Klasifikasi Tekstur Kematangan Buah Jeruk Manis Berdasarkan Tingkat Kecerahan Warna dengan Metode Deep Learning Convolutional Neural Network. *Jurnal Invotek Polbeng – Seri Informatika*. **6(2)**: 260-261.

Zhu, J., Chen, H., and Ye, W. 2020. A Hybrid CNN-LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar. *IEEE Access*. 8: 24713-2471