

Klasifikasi Multikelas pada Teks Judul Berita Menggunakan Algoritma Recurrent Neural Network

Bagus Bramantyo^{1*}, Muhammad Pajar Kharisma Putra², Nirwana Hendrastuty³

^{1,2}Informatika, Universitas Teknokrat Indonesia, Indonesia

³Sistem Informasi, Universitas Teknokrat Indonesia, Indonesia

Email: ^{1*} bagus_bramantyo@teknokrat.ac.id, ² pajarkharisma@teknokrat.ac.id, ³ nirwanahendrastuty@teknokrat.ac.id

Abstrak: Saat ini, informasi dapat dengan mudah diperoleh melalui internet, salah satunya melalui platform portal berita online yang memberikan akses informasi sesuai permintaan. Namun terkadang kita kesulitan menemukan konten berita yang diinginkan karena jumlahnya yang sangat banyak. Hal ini terjadi karena proses pengkategorian konten berita secara manual oleh author, yang bisa menyebabkan kesalahan seperti topik berita yang tercampur. Penelitian ini bertujuan untuk mengetahui kinerja metode *Deep Learning* menggunakan *Recurrent Neural Network* (RNN) dalam tugas multiklasifikasi pada judul berita dengan topik Ekonomi, Kesehatan, Olahraga, dan Politik. Judul berita untuk data latih dan uji diperoleh menggunakan *Web Scraping* dan setelah itu melalui tahap *Text Preprocessing* yang meliputi case folding, tokenization, stopwords removal, dan stemming. Selanjutnya, dilakukan ekstraksi fitur menggunakan TF-IDF untuk memberikan bobot pada setiap kata. Hasil pengujian kinerja model menggunakan *Confusion Matrix* menunjukkan akurasi mencapai 97%, sehingga metode RNN dapat digunakan dengan baik dalam tugas multiklasifikasi pada judul berita dan dapat diaplikasikan pada sistem pengklasifikasian judul berita.

Kata Kunci: *Confusion Matrix; Deep Learning; Multiklasifikasi; Recurrent Neural Network; Text Preprocessing;*

Abstract: Currently, obtaining information has become easier with the advent of internet technology. Online news portals provide on-demand access to desired information. However, the abundance of news content can make it difficult to find specific articles due to manual categorization errors. This research focuses on evaluating the performance of the Deep Learning method using a Recurrent Neural Network (RNN) for multi-classification tasks on news headlines related to Economics, Health, Sports, and Politics. Training and testing data were collected from news portals using Web Scraping, followed by Text Preprocessing stages such as case folding, tokenization, stopwords removal, and stemming. TF-IDF was then used for feature extraction to assign weights to each word. Testing the model's performance using the Confusion Matrix showed an accuracy of 97%, indicating that the RNN method effectively handles news headline classification and can be applied in news classification systems.

Keywords: Confusion Matrix; Deep Learning; Multiclassification; Recurrent Neural Network; Text Preprocessing;

1. PENDAHULUAN

Saat ini informasi dapat diperoleh dengan cara yang relatif mudah dengan adanya bantuan teknologi internet, salah satu cara yang saat ini banyak digunakan yaitu dengan mengakses platform portal berita online terpercaya yang tersebar luas diinternet. Dengan adanya platform tersebut dapat mempermudah dalam mengakses sumber berita sesuai dengan konten yang diinginkan, karena bersifat on demand sehingga dapat membantu dalam penyebaran distribusi informasi yang lebih merata ke semua lapisan masyarakat dengan lebih informatif, cepat, dan praktis mengingat memperoleh informasi merupakan hak semua orang tanpa terkecuali[1]. Meskipun saat ini kita diberikan kemudahan dalam mengakses sumber berita secara online, namun terkadang kita merasa kesulitan dalam mencari berita yang sesuai dengan konten yang diinginkan, hal ini disebabkan karena banyaknya konten berita yang harus dikategorikan oleh author secara manual sehingga dapat terjadi kesalahan yang mengakibatkan berita tercampur antara topik satu dengan topik yang lainnya. Hal ini didukung oleh pernyataan yang diberikan oleh Kementerian Komunikasi dan Informatika (Kominfo) di mana pada tahun 2018 di Indonesia terdapat kurang lebihnya 43.000 portal berita online (Kominfo, 2018).

Klasifikasi adalah metode untuk mengelompokkan data berdasarkan karakteristik yang sama pada data tersebut, prosesnya seperti membedakan dan mendistribusikan data ke dalam kelas yang sama [2]. klasifikasi merupakan salah satu bidang yang sangat penting dalam melakukan pengolahan data, terutama dalam dunia teknologi informasi. Salah satu penerapan klasifikasi yaitu pada sistem pengelompokan konten berita yang membantu dalam mengelompokkan berita ke dalam kategori yang sesuai, contohnya seperti berita politik, ekonomi, kesehatan, dan olahraga.

Natural Language Processing (NLP) merupakan salah satu cabang dari kecerdasan buatan, yang berfungsi untuk memberikan kemampuan pada mesin untuk dapat membaca dan memahami bahasa manusia sehingga dapat melakukan proses klasifikasi pada teks sesuai dengan kategorinya[3]. Pada NLP terdapat proses text preprocessing yang meliputi beberapa tahapan seperti case folding, tokenization, stopwords removal, dan stemming. Text Preprocessing berfungsi untuk mempersiapkan teks yang akan digunakan dalam pengujian dan pelatihan dengan tujuan untuk mengurangi noise yang terdapat pada data teks sehingga dapat meningkatkan kinerja classifier yang berpengaruh pada kecepatan proses klasifikasi[4].

Recurrent Neural Network (RNN) merupakan salah satu algoritma deep learning yang dapat melakukan prediksi pada data bertipe numerik time series, yang cara kerjanya yaitu dengan melakukan pemerosesan yang dilakukan secara berulang-ulang. RNN adalah jaringan saraf tiruan yang mengakomodasi output jaringan untuk menjadi input pada jaringan tersebut yang kemudian digunakan untuk menghasilkan output yang baru[5].

Berdasarkan referensi tersebut Penulis pada penelitian ini akan melakukan multiklasifikasi pada judul berita yang dibatasi pada topik Ekonomi, Kesehatan, Olahraga, dan Politik menggunakan pendekatan jaringan saraf tiruan RNN (Recurrent Neural Network). Untuk mengetahui kinerja yang dihasilkan oleh RNN dalam melakukan tugas multiklasifikasi pada judul berita.

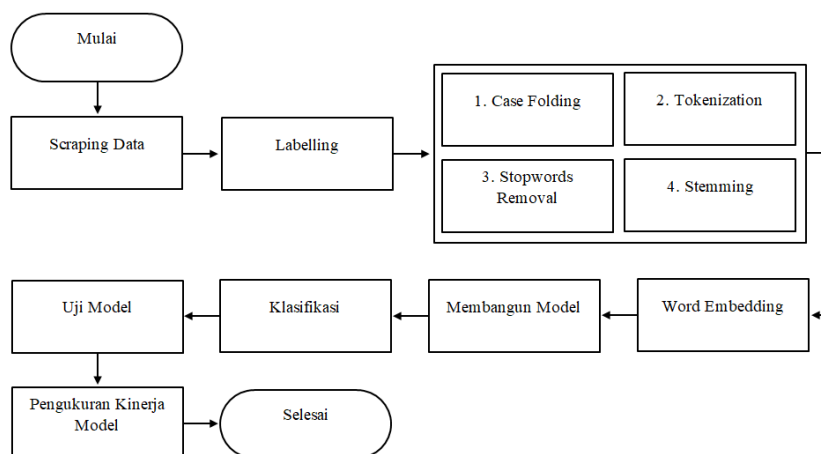
2. METODE PENELITIAN

Data yang digunakan pada penelitian ini merupakan judul berita yang diperoleh dari portal berita online menggunakan metode *Web Scraping*, yaitu metode yang digunakan untuk mengekstrak data dari World Wide Web (WWW) dan data tersebut disimpan ke dalam file sistem (database) untuk dilakukan pengolahan dan analisis. Metode ini umumnya digunakan saat melakukan pengunduhan data dalam jumlah yang sangat besar (big data) karena dianggap lebih efisien dan kuat[6]. Data yang dikumpulkan berjumlah 3000 judul berita yang diambil dari bulan maret tahun 2021 sampai dengan bulan maret 2023 dengan jumlah data pada setiap class berjumlah 750 data.

Untuk dapat melakukan multiklasifikasi, data yang berbentuk dokumen teks harus dilakukan tahap *Text Preprocessing* terlebih dahulu agar lebih terstruktur yang bertujuan untuk mempersiapkan dokumen teks sehingga dapat meningkatkan kinerja model classifier dan mempercepat proses klasifikasi[7] Tahapan yang dilakukan pada text preprocessing adalah sebagai berikut:

1. **Case Folding**, pada tahap ini dokumen teks akan diubah menjadi huruf kecil (lowercase) dan menghilangkan tanda baca (punctuation removal) sehingga dapat memudahkan proses analisis dan pengolahan data teks[8].
2. **Tokenization**, tahap ini bertujuan untuk memecah sebuah kalimat menjadi bentuk bagian-bagian kecil dari setiap kata pada kalimat yang disebut dengan token, sehingga memudahkan dalam mengidentifikasi kata-kata yang paling sering muncul dalam dokumen teks[9].
3. **Stopwords Removal**, Pada tahap ini akan menghilangkan kata penghubung (stopwords) yang tidak memiliki nilai informatif, sehingga informasi penting dalam teks menjadi lebih mudah untuk diidentifikasi dan dapat meringkas waktu proses analisis dan pengolahan data teks[10].
4. **Stemming**, pada tahap ini kata-kata yang memiliki kata kerja akan diubah menjadi kata dasar yang memiliki makna yang sama, tujuannya adalah untuk memperbaiki konsistensi data, meningkatkan akurasi model prediksi, menghemat ruang penyimpanan, dan mempercepat proses analisis saat melakukan pengolahan data teks[11].

Setelah tahap text preprocessing selesai dilakukan, tahapan selanjutnya yaitu melakukan ekstraksi fitur pada dokumen teks yang akan digunakan sebagai data latih pada model. *Word embedding* merupakan fungsi parameter yang digunakan untuk memetakan setiap kata ke dalam vektor berdimensi tinggi, yang memiliki keunggulan tidak membutuhkan anotasi yang dapat langsung diturunkan dari korpus tak teranotasi[12]. Metode word embedding yang digunakan untuk ekstraksi fitur yaitu menggunakan TF-IDF, pada tahap ini kata yang terdapat pada dokumen teks akan diberikan pembobotan kata. Pada pembobotan kata menggunakan metode TF-IDF yang menjadi fokus utama yaitu menghitung frekuensi dari setiap kata yang sering muncul pada dokumen teks (TF : Term Frequency) dan menghitung bagaimana kata didistribusikan keseluruhan dokumen teks (IDF : Inverse Document Frequency). Pengujian kinerja model dilakukan menggunakan metode *Confusion Matrix*, untuk menghitung nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN)[13]. Dengan menggunakan confusion matrix kinerja dari model untuk tugas multiklasifikasi dapat dihitung nilai akurasi, presisi, recall, dan f1-score. Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

Proses scraping melibatkan pencarian elemen HTML dengan menggunakan selektor atau atribut tertentu yang tugasnya untuk mengidentifikasi judul berita sehingga hanya mengambil konten yang sesuai. Judul berita yang sudah berhasil dikumpulkan dapat dilihat pada Tabel 1 berikut.

Tabel 1. Judul Berita Hasil Scrapping

No	Judul Berita	URL Berita
1	Tembus 46 Derajat Celcius, Ini Alasan di India Bisa Sangat Panas	https://www.detik.com/edu/detikpedia/d-6698332/tembus-46-derajat-celcius-ini-alasan-di-india-bisa-sangat-panas
2	Bukti Krisis Utang dan Perbankan AS Makin Menakutkan!	https://finance.detik.com/moneter/d-6699516/bukti-krisis-utang-dan-perbankan-as-makin-menakutkan
3	Kualitas Udara Jakarta Hari Ini 'Tidak Sehat' Versi Situs IQIAR	https://news.detik.com/berita/d-6699499/kualitas-udara-jakarta-hari-ini-tidak-sehat-versi-situs-iqiar
4	Pendaftaran PKN STAN 2023 Diperpanjang hingga 5 Mei, Buruan Daftar Sebelum Ditutup!	https://www.detik.com/edu/seleksi-masuk-pt/d-6699337/pendaftaran-pkn-stan-2023-diperpanjang-hingga-5-mei-buruan-daftar-sebelum-ditutup

Judul berita yang sudah selesai dikumpulkan selanjutnya akan dilakukan proses labelling yang bertujuan untuk memisahkan berita sesuai dengan kateorinya, berita dibagi menjadi 4 class yaitu ekonomi, kesehatan, olahraga, dan politik. Judul berita yang sudah diberi label dapat dilihat pada Tabel 2 berikut.

Tabel 2. Judul Berita yang Sudah Dilabelling

No	Judul Berita	Kategori
1	Total Aset Keuangan Syariah RI Tembus Rp.2.200 T per Agustus	EKONOMI
2	Antisipasi Gagal Ginjal Akut, Ini yang Perlu Dilakukan Orang Tua	KESEHATAN
3	Jelang MotoGP Valencia 2022, Marc Marquez Siap Ambil Risiko	OLAHRAGA
4	Partai Berkarya Gugat KPU ke PN Jakpus, Minta Tahapan Pemilu Ditunda	POLITIK

Hasil akhir dari judul berita yang sudah diperoleh dan dilakukan proses labelling, menghasilkan data sebanyak 3004 judul berita. Jumlah total keseluruhan judul berita dapat dilihat pada Tabel 3 berikut.

Tabel 3. Jumlah Judul Berita

No	Kategori Berita	Jumlah
1	EKONOMI	751
2	KESEHATAN	751
3	OLAHRAGA	751
4	POLITIK	751

Pada tahap preprocessing bertujuan untuk membersihkan, merapikan, dan mempersiapkan data sebelum dilakukan analisis lebih lanjut, tahap ini wajib dilakukan karena data judul berita yang sudah berhasil diperoleh masih terdapat beberapa kata yang tidak relevan dan juga tidak konsisten seperti adanya huruf kapital maupun tanda baca

pada judul berita. Pada tahap preprocessing judul berita akan melewati beberapa proses seperti case folding, tokenization, stopwords removal, dan stemming. Judul berita yang sudah dilakukan case folding dapat dilihat pada Tabel 4 berikut.

Tabel 4. Hasil Case Folding Judul Berita

No	Judul Berita	Hasil Case Folding
1	Total Aset Keuangan Syariah RI Tembus Rp2.200 T per Agustus	total aset keuangan syariah ri tembus rp2200 t per agustus
2	Antisipasi Gagal Ginjal Akut, Ini yang Perlu Dilakukan Orang Tua	antisipasi gagal ginjal akut ini yang perlu dilakukan orang tua
3	Jelang MotoGP Valencia 2022, Marc Marquez Siap Ambil Risiko	jelang motogp valencia 2022 marc marquez siap ambil risiko
4	Partai Berkarya Gugat KPU ke PN Jakpus, Minta Tahapan Pemilu	partai berkarya gugat kpu ke pn jakpus minta tahapan pemilu ditunda

Setelah selesai melakukan case folding pada judul berita, proses dilanjutkan dengan melakukan tahap tokenization yang bertujuan untuk memecah kalimat menjadi unit satuan terkecil yang berbentuk kata. Judul berita yang sudah dilakukan tokenization dapat dilihat pada Tabel 5 berikut.

Tabel 5. Hasil Tokenization Judul Berita

No	Judul Berita	Hasil Tokenization
1	total aset keuangan syariah ri tembus rp2200 t per agustus	[total, aset, keuangan, syariah, ri, tembus, rp2200, t, per, agustus]
2	antisipasi gagal ginjal akut ini yang perlu dilakukan orang tua	[antisipasi, gagal, ginjal, akut, ini, yang, perlu, dilakukan, orang, tua]
3	jelang motogp valencia 2022 marc marquez siap ambil risiko	[jelang, motogp, valencia, 2022, marc, marquez, siap, ambil, risiko]
4	partai berkarya gugat kpu ke pn jakpus minta tahapan pemilu ditunda	[partai, berkarya, gugat, kpu, ke, pn, jakpus, minta, tahapan, pemilu, ditunda]

Setelah selesai melakukan tokenization pada judul berita, proses selanjutnya yaitu melakukan tahap stopwords removal untuk menghapus kata hubung yang ada pada dokumen teks. Judul berita yang sudah dilakukan stopwords removal dapat dilihat pada Tabel 6 berikut.

Tabel 6. Hasil Stopwords Removal Judul Berita

No	Judul Berita	Hasil Stopwords Removal
1	[total, aset, keuangan, syariah, ri, tembus, rp2200, t, per, agustus]	[total, aset, keuangan, syariah, ri, tembus, rp2200, t, agustus]
2	[antisipasi, gagal, ginjal, akut, ini, yang, perlu, dilakukan, orang, tua]	[antisipasi, gagal, ginjal, akut, orang, tua]

3	[jelang, motogp, valencia, 2022, marc, marquez, siap, ambil, risiko]	[jelang, motogp, valencia, 2022, marc, marquez, ambil, risiko]
4	[partai, berkarya, gugat, kpu, ke, pn, jakpus, minta, tahapan, pemilu, ditunda]	[partai, berkarya, gugat, kpu, pn, jakpus, tahapan, pemilu, ditunda]

Setelah selesai melakukan stopwords removal pada judul berita, proses selanjutnya yaitu melakukan tahap stemming untuk menghapus imbuhan yang terdapat pada kata. Judul berita yang sudah dilakukan stemming dapat dilihat pada Tabel 7 berikut.

Tabel 7. Hasil Stemming Judul Berita

No	Judul Berita	Hasil Stemming
1	[total, aset, keuangan, syariah, ri, tembus, rp2200, t, agustus]	[total, aset, uang, syariah, ri, tembus, rp2200, t, agustus]
2	[antisipasi, gagal, ginjal, akut, orang, tua]	[antisipasi, gagal, ginjal, akut, orang, tua]
3	[jelang, motogp, valencia, 2022, marc, marquez, ambil, risiko]	[jelang, motogp, valencia, 2022, marc, marquez, ambil, risiko]
4	[partai, berkarya, gugat, kpu, pn, jakpus, tahapan, pemilu, ditunda]	[partai, karya, gugat, kpu, pn, jakpus, tahap, pemilu, tunda]

Setelah judul berita melewati semua proses yang terdapat pada tahap preprocessing, judul berita yang sudah bersih selanjutnya akan dilakukan ekstraksi fitur. Judul berita yang sudah dilakukan preprocessing dapat dilihat pada Tabel 8 berikut.

Tabel 8. Hasil Preprocessing Judul Berita

No	Judul Berita	Hasil Preprocessing
1	Total Aset Keuangan Syariah RI Tembus Rp2.200 T per Agustus	[total, aset, uang, syariah, ri, tembus, rp2200, t, agustus]
2	Antisipasi Gagal Ginjal Akut, Ini yang Perlu Dilakukan Orang Tua	[antisipasi, gagal, ginjal, akut, orang, tua]
3	Jelang MotoGP Valencia 2022, Marc Marquez Siap Ambil Risiko	[jelang, motogp, valencia, 2022, marc, marquez, ambil, risiko]
4	Partai Berkarya Gugat KPU ke PN Jakpus, Minta Tahapan Pemilu	[partai, karya, gugat, kpu, pn, jakpus, tahap, pemilu, tunda]

Ekstraksi fitur bertujuan untuk memberikan pembobotan pada setiap kata yang terdapat pada dokumen teks, sehingga mengubah kata-kata dalam teks menjadi representasi vektor numerik yang memiliki makna dan konteks semantik. Contoh kata yang sudah dilakukan ekstraksi fitur dapat dilihat pada Tabel 9 berikut.

Tabel 9. Pembobotan Kata Menggunakan TF-IDF

No	Kata	Hasil TF-IDF
1	fakta	[-0.01965572 0.04603744 0.0344272]

2	dan	[0.02622599 0.01017879 - 0.02631575]
3	ahistoris	[0.02143338 0.00393938 0.04414273]
4	pisahkan	[0.00245609 0.03545865 0.03473778]
5	olahraga	[-0.00143487 -0.0071972 - 0.01739461]
6	dari	[-0.04753917 0.03042506 0.00705056]
7	politik	[0.03799437 -0.01213868 0.00370164]

Training model bertujuan untuk melatih RNN dalam mempelajari pola dan mengekstraksi fitur dari data latih yang digunakan sehingga model dapat melakukan tugas multiklasifikasi[14]. Arsitektur model RNN yang digunakan pada penelitian dapat dilihat pada Gambar 2 berikut.

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, None, 16)	160016
dropout (Dropout)	(None, None, 16)	0
global_average_pooling1d (GlobalAveragePooling1D)	(None, 16)	0
dropout_1 (Dropout)	(None, 16)	0
dense (Dense)	(None, 4)	68

```

=====
Total params: 160,084
Trainable params: 160,084
Non-trainable params: 0
    
```

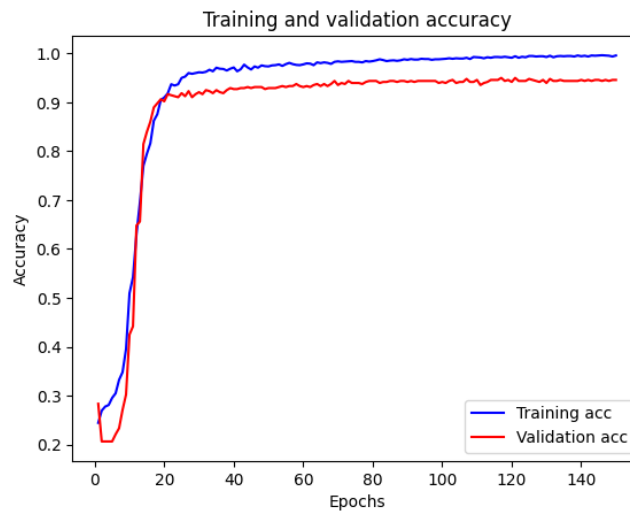
Gambar 2. Arsitektur Model RNN

Proses training model dilakukan sampai mendapatkan nilai accuracy yang semakin besar dan nilai loss yang semakin kecil[15], sehingga akurasi model dalam memprediksi data uji semakin baik. Hasil training dari model dapat dilihat pada tabel 10 berikut.

Tabel 10. Learning Rate Training Model

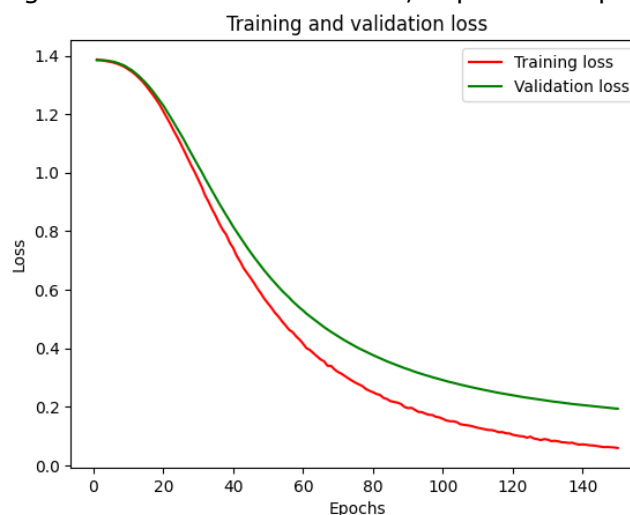
No	epoch_set	epoch_real	accuracy	loss	accuracy correction	loss correction
1	150	150	0.9305	0.2626	0	0
2	150	91	0.9288	0.2156	0.0017	0.0469
3	150	10	0.9288	0.2191	0	0.0034
4	150	8	0.9321	0.2140	0.0033	0.0050

Nilai accuracy hasil training model dalam bentuk kurva, dapat dilihat pada gambar 3 berikut.



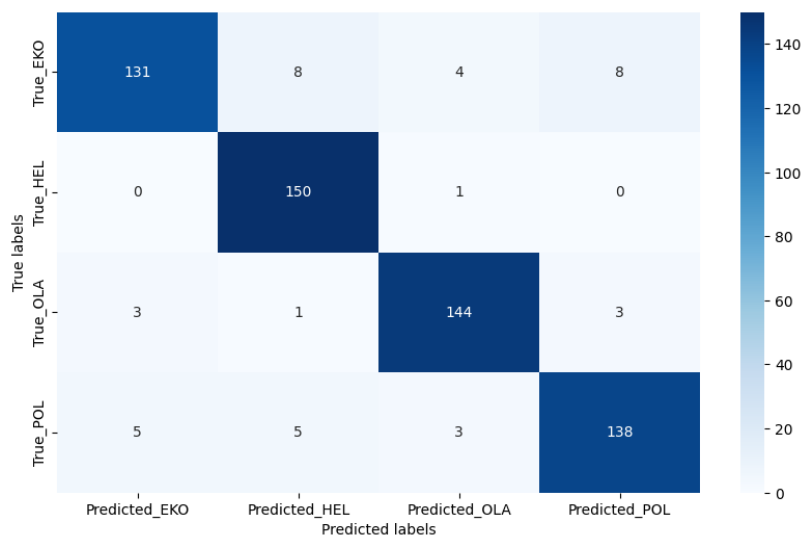
Gambar 3. Kurva Nilai Accuracy Training Model

Nilai loss hasil training model dalam bentuk kurva, dapat dilihat pada gambar 4 berikut.



Gambar 4. Kurva Nilai Loss Training Model

Hasil pengujian kinerja model menggunakan metode confusion matrix dengan jumlah data uji sebanyak 604 yang dibagi menjadi 4 class (ekonomi, kesehatan, olahraga, dan politik) dengan jumlah data pada masing-masing class sebanyak 151 judul berita, diperoleh nilai akurasi, presisi, recall, dan f1-score pada setiap class dan juga total keseluruhan dari keempat class. Hasil confusion matrix dapat dilihat pada gambar 5 berikut.



Gambar 5. Hasil Confusion Matrix

Untuk dapat mengetahui nilai akurasi, presisi, recall, dan f1-score dari model yang sudah dilatih, maka harus dilakukan perhitungan pada setiap class. Berikut ini hasil perhitungan dari setiap class.

1. Class Ekonomi, pada class ekonomi diperoleh nilai TP sebanyak 131, nilai FN sebanyak 20, nilai FP sebanyak 8, dan nilai TN sebanyak 445. Untuk lebih jelasnya dapat dilihat pada Tabel 11 berikut.

Tabel 11. Hasil Prediksi Pada Class Ekonomi

TP	FN	FP	TN
131	20	8	445

$$\text{Akurasi} = \frac{131 + 445}{131 + 445 + 8 + 20} * 100\% = 95\%$$

$$\text{Presisi} = \frac{131}{131 + 8} * 100\% = 94\%$$

$$\text{Recall} = \frac{131}{131 + 20} * 100\% = 87\%$$

$$\text{F1 - Score} = 2 * \frac{94\% * 87\%}{94\% + 87\%} * 100\% = 90\%$$

Perhitungan yang telah dilakukan pada class ekonomi menggunakan metode confusion matrix, diperoleh nilai akurasi sebesar 95%, nilai presisi sebesar 94%, nilai recall sebesar 87%, dan nilai f1-score sebesar 90%.

2. Class Kesehatan, pada class kesehatan diperoleh nilai TP sebanyak 150, nilai FN sebanyak 1, nilai FP sebanyak 14, dan nilai TN sebanyak 439. Untuk lebih jelasnya dapat dilihat pada Tabel 12 berikut.

Tabel 12. Hasil Prediksi Pada Class Kesehatan

TP	FN	FP	TN
150	1	14	439

439

$$\text{Akurasi} = \frac{150 + 439}{150 + 439 + 14 + 1} * 100\% = 98\%$$

$$\text{Presisi} = \frac{150}{150 + 14} * 100\% = 91\%$$

$$\text{Recall} = \frac{150}{150 + 1} * 100\% = 99\%$$

$$\text{F1 - Score} = 2 * \frac{91\% * 99\%}{91\% + 99\%} * 100\% = 95\%$$

Perhitungan yang telah dilakukan pada class kesehatan menggunakan metode confusion matrix, diperoleh nilai akurasi sebesar 98%, nilai presisi sebesar 91%, nilai recall sebesar 99%, dan nilai f1-score sebesar 95%.

3. Class Olahraga, class olahraga diperoleh nilai TP sebanyak 144, nilai FN sebanyak 7, nilai FP sebanyak 8, dan nilai TN sebanyak 445. Untuk lebih jelasnya dapat dilihat pada Tabel 13 berikut.

Tabel 13. Hasil Prediksi Pada Class Olahraga

TP	FN	FP	TN
144	7	8	445

$$\text{Akurasi} = \frac{144 + 445}{144 + 445 + 8 + 7} * 100\% = 98\%$$

$$\text{Presisi} = \frac{144}{144 + 8} * 100\% = 95\%$$

$$\text{Recall} = \frac{144}{144 + 7} * 100\% = 95\%$$

$$\text{F1 - Score} = 2 * \frac{95\% * 95\%}{95\% + 95\%} * 100\% = 95\%$$

Perhitungan yang telah dilakukan pada class olahraga menggunakan metode confusion matrix, diperoleh nilai akurasi sebesar 98%, nilai presisi sebesar 95%, nilai recall sebesar 95%, dan nilai f1-score sebesar 95%.

4. Class Politik, class politik diperoleh nilai TP sebanyak 138, nilai FN sebanyak 13, nilai FP sebanyak 11, dan nilai TN sebanyak 442. Untuk lebih jelasnya dapat dilihat pada Tabel 14 berikut.

Tabel 14. Hasil Prediksi Pada Class Politik

TP	FN	FP	TN
138	13	11	442

$$\text{Akurasi} = \frac{138 + 442}{138 + 442 + 11 + 13} * 100\% = 96\%$$

$$\text{Presisi} = \frac{138}{138 + 11} * 100\% = 93\%$$

$$\text{Recall} = \frac{138}{138+13} * 100\% = 91\%$$

$$\text{F1 - Score} = 2 * \frac{93\% * 91\%}{93\% + 91\%} * 100\% = 92\%$$

Perhitungan yang telah dilakukan pada class politik menggunakan metode confusion matrix, diperoleh nilai akurasi sebesar 96%, nilai presisi sebesar 93%, nilai recall sebesar 91%, dan nilai f1-score sebesar 92%.

Nilai akurasi, presisi, recall, dan f1-score pada masing-masing class dapat dilihat pada Tabel 15 berikut.

Tabel 15. Nilai Akurasi, Presisi, Recall, dan F1-score setiap Class

No	Class	Akurasi	Presisi	Recall	F1-Score
1	Ekonomi	95%	94%	87%	90%
2	Kesehatan	98%	91%	99%	95%
3	Olahraga	98%	95%	95%	95%
4	Politik	96%	93%	91%	92%

Dari keseluruhan class diperoleh nilai TP sebanyak 563, nilai FN sebanyak 41, nilai FP sebanyak 41, dan nilai TN sebanyak 1771. Untuk lebih jelasnya dapat dilihat pada Tabel 16 berikut.

Tabel 16. Hasil Prediksi Pada Keseluruhan Class

TP	FN	FP	TN
563	41	41	1771

selanjutnya akan dilakukan perhitungan nilai dari keseluruhan class untuk mengetahui kinerja yang dihasilkan model. Hasil kinerja model dapat dilihat pada Tabel 17 berikut.

Tabel 17. Hasil Kinerja Model

Akurasi	Presisi	Recall	F1-Score
97%	93%	93%	93%

$$\text{Akurasi} = \frac{563+1771}{563+1771+41+41} * 100\% = 97\%$$

$$\text{Presisi} = \frac{563}{563+41} * 100\% = 93\%$$

$$\text{Recall} = \frac{563}{563+41} * 100\% = 93\%$$

$$\text{F1 - Score} = 2 * \frac{93\% * 93\%}{93\% + 93\%} * 100\% = 93\%$$

4. KESIMPULAN

Dari hasil pengujian menggunakan confusion matrix dengan data uji sebanyak 604 judul berita yang dibagi menjadi 4 class (ekonomi, kesehatan, olahraga, dan politik) untuk melakukan tugas multiklasifikasi diperoleh nilai akurasi pada class ekonomi sebesar 95%, class kesehatan sebesar 98%, class olahraga sebesar 95%, class politik sebesar 93% dan

nilai akurasi untuk keseluruhan class diperoleh nilai sebesar 97%. Jumlah dataset yang digunakan yaitu sebanyak 3004 judul berita dengan perbandingan 80% digunakan sebagai data latih sebanyak 2400 dan 20% sebagai data uji sebanyak 604 judul berita. Dari hasil pengujian dapat ditarik kesimpulan bahwa metode Recurrent Neural Network dapat melakukan tugas multiklasifikasi pada judul berita dengan cukup baik sehingga dapat diterapkan pada sistem pengklasifikasian judul berita.

5. REFERENCES

- [1] L. Lindawati, "Pola Akses Berita Online Kaum Muda," *J. Stud. Pemuda*, vol. 4, no. 1, pp. 241–259, 2015.
- [2] I. P. Ninditama, W. Cholil, M. Akbar, and D. Antoni, "Klasifikasi keluarga sejahtera study kasus : Kecamatan Kota Palembang," *J. TEKNO KOMPAK*, vol. 15, no. 2, pp. 37–49, 2020.
- [3] I. Diah, "ARTIFICIAL INTELLIGENCE (AI)," no. 187038039, pp. 1–17, 2018.
- [4] A. Sivakumar and R. Gunasundari, "A Survey on Data Preprocessing Techniques for Bioinformatics and Web Usage Mining," *Int. J. Pure Appl. Math.*, vol. 117, no. 20, pp. 785–794, 2017.
- [5] D. Tarkus, S. R. U. A. Sompie, and A. Jacobus, "Implementasi Metode Recurrent Neural Network pada Pengklasifikasian Kualitas Telur Puyuh," *J. Tek. Inform.*, vol. 15, no. 2, pp. 137–144, 2020.
- [6] S. J. Mooney, D. J. Westreich, and A. M. El-Sayed, "Epidemiology in the era of big data," *Epidemiology*, vol. 26, no. 3, pp. 390–394, 2015, doi: 10.1097/eDe.0000000000000274.
- [7] J. S. Ronen Feldman, "The Text Mining Handbook: Advanced Approaches to Analyzing Unstructured Data," *Crit. Sociol.*, vol. 37, no. 4, pp. 493–497, 2007, doi: 10.1177/0261018311403863.
- [8] A. S. Alamsah, "IMPLEMENTASI SISTEM TEMU KEMBALI INFORMASI UNTUK PENCARIAN BUKU PADA TOKO BUKU ONLINE MENGGUNAKAN METODE VECTOR SPACE MODEL," pp. 6–20, 2018.
- [9] G. Ngurah, M. Nata, P. Yudiastra,) Stmik, S. Bali, and J. Raya Puputan, "Preprocessing Text Mining Pada Email Box Berbahasa Indonesia," *E-Proceedings KNS&I STIKOM Bali*, pp. 479–483, 2017, [Online]. Available: <http://www.knsi.stikom-bali.ac.id/index.php/e-proceedings/article/view/88>
- [10] F. Rahutomo and A. R. T. H. Ririd, "Evaluasi Daftar Stopword Bahasa Indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 1, p. 41, 2019, doi: 10.25126/jtiik.2019611226.
- [11] J. Wibowo, "Aplikasi penentuan kata dasar dari kata berimbuhan pada kalimat bahasa Indonesia dengan algoritma Stemming," *J. Ris. Komput.*, vol. 3, no. 5, pp. 346–350, 2016.
- [12] C. J. Yoshua Bengio, Réjean Ducharme, Pascal Vincent, "A Neural Probabilistic Language Model," *Fullerenes Nanotub. Carbon Nanostructures*, vol. 26, no. 8, pp. 465–470, 2018, doi: 10.1080/1536383X.2018.1448388.
- [13] Karsito and S. Susanti, "Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura Residencia," *J. Teknol. Pelita Bangsa*, vol. 9, pp. 43–48, 2019.
- [14] F. A. D. Aji Prasetya Wibawa, Muhammad Guntur Aji Purnama, Muhammad Fathony Akbar, "Metode-metode Klasifikasi," *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 1, p. 134, 2018.
- [15] M. R. Firmansyah, R. Ilyas, and F. Kasyidi, "Klasifikasi Kalimat Ilmiah Menggunakan Recurrent Neural Network," *Pros. 11th Ind. Res. Work. Natl. Semin.*, vol. 11, no. 1, pp. 488–495, 2020.