

Klasifikasi Berita Bahasa Indonesia dengan Recurrent Neural Network (RNN)

Muh. Ridha Agam^{*1}, Christian Sri Kusuma Aditya¹, Galih Wasis Wicaksono¹

Universitas Muhammadiyah Malang

muhridhaagam@webmail.umm.ac.id^{*1}, christianskaditya@umm.ac.id¹, galih.w.w@umm.ac.id¹

Abstrak

Penelitian ini bertujuan untuk mengembangkan model klasifikasi dokumen berita berbahasa Indonesia menggunakan Recurrent Neural Network (RNN). Dengan kemajuan teknologi informasi dalam satu dekade terakhir, media online telah menjadi sumber informasi utama bagi masyarakat, yang mengakibatkan perlunya mekanisme yang efisien untuk mengelola dan mengklasifikasikan konten berita yang terus bertambah setiap hari. Teknologi pembelajaran mesin, khususnya RNN, menjadi sangat penting dalam mengklasifikasikan dokumen berita berbahasa Indonesia karena kemampuannya dalam mempelajari ketergantungan waktu pada data sekuensial. Penelitian ini bertujuan untuk meningkatkan akurasi RNN dengan RNN dan juga melalui augmentasi data. Penelitian ini melibatkan beberapa tahapan utama, yaitu pengumpulan data melalui teknik scraping dari portal berita Kompas.com, preprocessing teks yang mencakup pembersihan, penyaringan, stemming, dan tokenisasi, augmentasi data menggunakan Back Translation, serta pelatihan model RNN. Augmentasi data dilakukan untuk mengatasi ketidakseimbangan kelas dalam dataset dan meningkatkan akurasi model. Hasil penelitian menunjukkan bahwa model RNN yang dilatih dengan augmentasi data menghasilkan peningkatan kinerja yang signifikan dibandingkan dengan metode konvensional seperti Naïve Bayes Classifier. Model RNN tanpa augmentasi mencapai akurasi 92%, sementara model RNN dengan augmentasi data mencapai akurasi 96%. Peningkatan ini juga terlihat pada metrik evaluasi lainnya seperti precision, recall, dan F1-Score

Kata Kunci: Recurrent Neural Network, Back Translation, Augmentasi Data, Klasifikasi Teks, Klasifikasi Berita

Abstract

This study aims to develop a classification model for Indonesian news documents using Recurrent Neural Network (RNN). With the advancement of information technology over the past decade, online media has become the primary source of information for society, necessitating efficient mechanisms to manage and classify the ever-growing content. Machine learning technology, especially RNN, is crucial for classifying Indonesian news documents due to its ability to learn temporal dependencies in sequential data. This study aims to improve RNN accuracy through normal RNN and data augmentation using Back Translation. This research involves several main stages, including data collection through scraping techniques from the Kompas.com news portal, text preprocessing that includes cleaning, filtering, stemming, and tokenization, data augmentation using Back Translation, and RNN model training. Data augmentation is performed to address class imbalance in the dataset and enhance model accuracy. The results show that the RNN model trained with data augmentation significantly improves performance compared to conventional methods like the Naïve Bayes Classifier. The RNN model without augmentation achieved an accuracy of 92%, while the RNN model with data augmentation achieved an accuracy of 96%. This improvement is also reflected in other evaluation metrics such as precision, recall, and F1-Score.

Keywords: Recurrent Neural Network, Back Translation, Data Augmentation, Text Classification, News Classification

1. Pendahuluan

Kemajuan teknologi informasi dalam satu dekade terakhir telah membawa perubahan besar pada industri media, pers dan jurnalisme. Media cetak seperti surat kabar tidak lagi menjadi sumber utama penyebaran informasi. Perkembangan teknologi informasi dan komunikasi telah

mengubah cara kita mengakses, memproses, dan menyebarkan informasi. Industri media, khususnya, telah mengalami pergeseran paradigma dari media tradisional ke platform digital. Adanya internet memungkinkan masyarakat mendapatkan informasi jauh lebih mudah dan cepat, informasi yang ditemukan di internet bisa jadi valid dan bisa jadi tidak valid (*hoax*) [1]

Media online kini menjadi sumber informasi utama bagi sebagian besar masyarakat, mengingat kemampuannya untuk menyediakan berita secara real-time. Di sinilah peran teknologi pembelajaran mesin, khususnya Recurrent Neural Network, menjadi penting dalam mengklasifikasikan dokumen berita berbahasa Indonesia. Metode-metode pada pendekatan deep-learning sangat disarankan karena telah banyak penelitian yang membuktikan bahwa nilai performance dan nilai akurasi menggunakan deep-learning menghasilkan nilai yang baik. Dalam beberapa penelitian, metode Recurrent Neural Network berhasil digunakan untuk masalah klasifikasi dengan data sekuensial

RNN merupakan salah satu arsitektur deep-learning yang efektif untuk pemrosesan bahasa alami dan analisis teks, karena kemampuannya dalam mempelajari ketergantungan waktu pada data sekuensial. Beberapa penelitian di Indonesia telah menunjukkan potensi RNN dalam berbagai aplikasi NLP [2], termasuk klasifikasi teks. berhasil mengaplikasikan RNN dalam klasifikasi kalimat ilmiah, menunjukkan efektivitas RNN dalam mengolah data teks dengan struktur yang kompleks. Sementara itu, pada penelitian lainnya menggaris bawahi keunggulan Stemming dalam implementasi dan dalam memodelkan pada tipe berita, sehingga memberikan hasil dari pemahaman berita dengan jumlah kata yang telah diolah [3]. Selanjutnya peneliti [4] melakukan pelatihan dengan enhanced confix stripping stemmer yang menunjukkan percobaan klasifikasi berita dengan mencoba stemming pada berita yang diolah. Sedangkan pada [5] peneliti mengeksplorasi penerapan RNN untuk text summarization, menegaskan kemampuan RNN dalam menangkap nuansa semantik dari teks berita

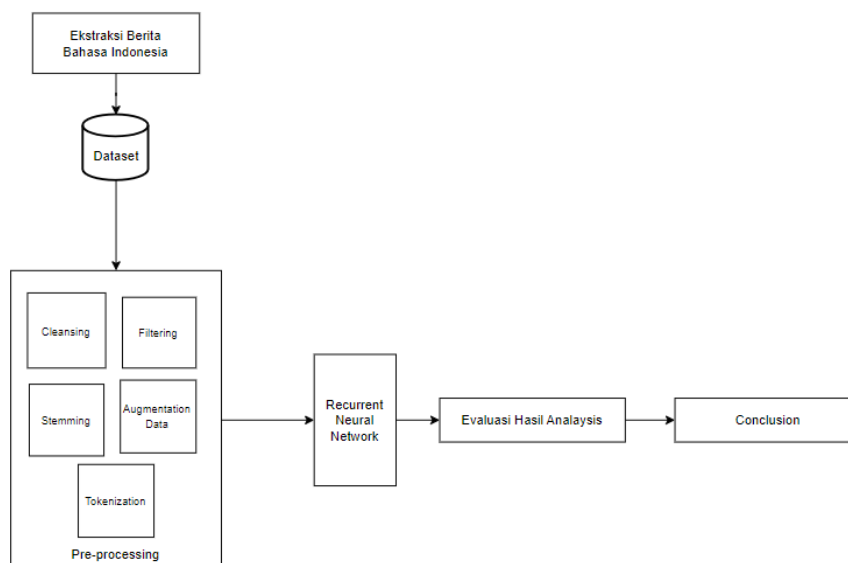
Beberapa studi sebelumnya telah melakukan klasifikasi dokumen berita menggunakan berbagai metode, termasuk penggunaan algoritma Naïve Bayes dengan penambahan Bayesian boosting [6] yang mampu menghasilkan model klasifikasi yang andal dan sering kali menghindari overfitting. Penelitian ini mampu memperoleh hasil akurasi, presisi, dan recall sebesar 72.1%. Lebih lanjut, aplikasi Bayesian boosting untuk label yang bersifat polinomial memiliki efek meningkatkan hasil evaluasi sebesar 1.3%. Namun, dengan demikian belum dapat memperlihatkan hasil yang cukup baik, masih terdapat beberapa keterbatasan yang mana kemudian di evaluasi ulang dengan menggunakan *Deep learning* yakni dengan menggunakan RNN dan juga faktor dari keseimbangan data yang juga ikut dikembangkan.

Selanjutnya, penelitian oleh Setiawan [7] menunjukkan klasifikasi teks berita dengan Naïve Bayes juga dengan menggunakan scraping yang mengklasifikasikan kategori yang ada yang mana juga memperlihatkan fitur N-Gram yang menunjukkan proses dari klasifikasi itu sendiri. Dalam studi yang dilakukan oleh Iftitah [8] menjelaskan mengenai augmentasi teks pada data tidak seimbang dengan melakukan *back translation* guna untuk meningkatkan pada dataset.

Penelitian yang dilakukan oleh [9] menyoroti penggunaan RNN dalam deteksi berita palsu, sebuah isu kritis dalam era informasi saat ini. Studi ini menunjukkan bagaimana teknologi pembelajaran mesin, khususnya RNN, tidak hanya membantu dalam klasifikasi teks tetapi juga dalam memerangi disinformasi. Selain itu [10] memberikan pembuktian pada klasifikasi multi-label dengan menggunakan deep learning dengan variasi sinonim.

2. Metode Penelitian

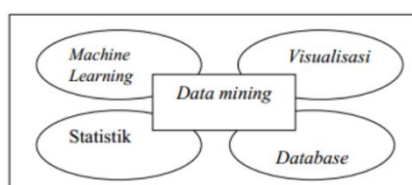
Pada penelitian ini, terdapat proses penggunaan model *Recurrent neural network* untuk mendapatkan hasil Klasifikasi dari Berita Bahasa Indonesia. Diagram alur metode penelitian ada pada Gambar 1.



Gambar 1. Alur penelitian

Untuk memperoleh data yang digunakan dalam penelitian ini digunakan teknik scraping. Dimana pengambilannya menggunakan teknik yang dapat mengambil data indeks yang berisi judul, link, kategori, dan tanggal, nilai dalam XPATH pada website berita kompas.com. Penelitian ini dilakukan dengan menggunakan dataset hasil proses scraping pada salah satu website berita yang ada, dengan jumlah total sekitar 7610 data. Kemudian dilakukan proses seleksi data guna mendapatkan 7 kelas yang paling tepat sasaran yang berjumlah total 1979 data yang mana data tersebut merupakan data paling banyak dalam kelasnya guna untuk dapat memastikan data tersebut dapat seimbang. Total dataset tersebut kemudian akan dibagi menjadi 60% data pelatihan dan 20% data pengujian serta 20% data validasi.

Data mining adalah proses mendapatkan informasi berguna dari data menggunakan berbagai metode, teknik, atau algoritma. Tujuan dan proses KDD harus dipertimbangkan saat memilih metode atau teknik *data mining* yang tepat. Kesimpulan dari proses data mining, Knowledge Discovery in Databases, dan teknik yang terlibat dalamnya menunjukkan pentingnya menggunakan pendekatan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi informasi berharga dari database besar. Proses ini dimulai dengan pengumpulan data yang komprehensif, diikuti dengan tahap pre-processing untuk membersihkan, mengintegrasikan, dan mempersiapkan data untuk analisis lebih lanjut. Pre-processing, sebagai tahap kritis sebelum data mining dimulai, memastikan bahwa data yang diolah berkualitas tinggi dan dapat dipercaya. Langkah-langkah seperti menghapus duplikat, memeriksa konsistensi data, dan memperbaiki kesalahan memungkinkan data menjadi lebih bersih dan siap untuk diinterpretasikan dengan benar. Selain itu, dalam Data Mining bukan hanya sekadar proses teknis, tetapi juga melibatkan aspek kreatif untuk mengubah dan menyusun ulang data dengan cara yang efektif. Hasilnya adalah wawasan yang lebih dalam dan bernilai bagi pengambilan keputusan strategis.



Gambar 2. Proses Data Mining

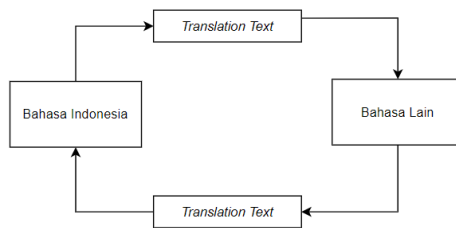
Text preprocessing merupakan tahapan atau proses penyempurnaan teks dimana data yang sebelumnya tidak terstruktur menjadi data terstruktur. Dalam proses pembersihan, berbagai konten yang URL-nya tercantum dalam konten berita akan dihapus. Pada proses ini juga terjadi casefolding yang mengubah huruf dari huruf besar menjadi huruf kecil.

Sedangkan proses filtering atau penghapusan stopwords dilakukan dengan menggunakan library «sastrawi» dari python yang akan mengidentifikasi berbagai macam kata yang tidak diperlukan sehingga akan terhapus dan menjadikan kalimat memiliki nilai klasifikasi yang lebih tinggi karena tidak perlu. mengolah kata-kata yang tidak diperlukan, contoh proses pembersihan dan penyaringan disajikan pada Tabel 1.

Tabel 1. Proses Pembersihan dan Penyaringan

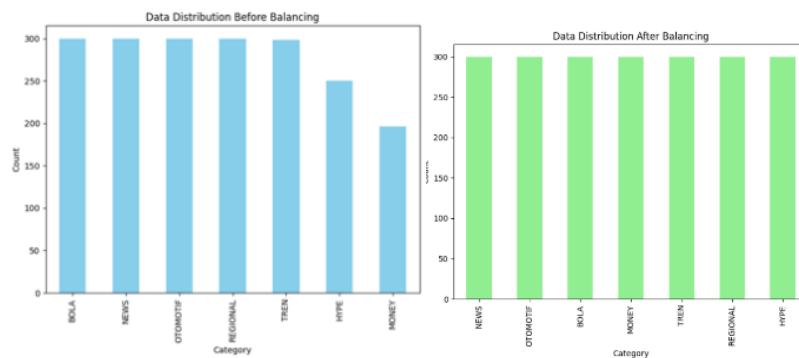
Memasukkan	Hasil Pembersihan	Hasil Penyaringan
www.kompas.com, JakPro secara tegas menyatakan lahan yang dimanfaatkan oleh ruko merupakan aset PT. JakPro	jakpro secara tegas menyatakan lahan yang dimanfaatkan oleh ruko merupakan aset pt jakpro	jakpro secara tegas menyatakan lahan yang dimanfaatkan ruko merupakan aset pt jakpro

Stemming merupakan tahapan dalam pengolahan teks yang bertujuan untuk menemukan kata dasar dari kata asli yang muncul dalam sebuah teks. Pada proses awalnya perlu dilakukan penyeimbangan distribusi jumlah data pada masing-masing kelas kategori. Metode Augmentasi dilakukan dengan Back Translation yang mana memiliki suatu prinsip kerja dengan memperbanyak jumlah data kelas minor dengan melakukan penerjemahan teks dari Bahasa asal menjadi suatu Bahasa lain. Dalam hal ini, penggunaan Bahasa lain digunakan untuk augmentasi data ini sendiri. Pada dasarnya, augmentasi data dengan Back Translation dapat menghasilkan sebuah teks yang baru dengan melakukan terjemahan balik dari Bahasa lain ke dalam Bahasa Indonesia.



Gambar 3. Alur Penelitian

Penggunaan Augmentasi data dengan Back Translation ini dilakukan guna untuk dapat menambah data dan juga mampu memperoleh teks baru dengan makna yang sama tetapi berbeda dari teks aslinya. Pada penelitian sebelumnya juga menyebut bahwa penggunaan beberapa Bahasa untuk translation dapat membantu untuk pengolahan data yang lebih baik dan baru. Untuk mengatasi kekurangan data yang terjadi pada dua kelas setelah dilakukan pemotongan data dengan mengambil kelas dengan jumlah tertinggi terdapat kekurangan pada beberapa kelas yang mana dapat dilakukan balancing kembali. Sehingga dengan augmentasi data hal ini dilakukan guna untuk mendapatkan distribusi data antar kelas yang sama untuk dapat dilakukan klasifikasi pada distribusi data yang ada.



Gambar 4. Hasil Augmentasi Data dengan Back Translation

Proses Tokenisasi adalah proses mengubah kata dalam kalimat menjadi bilangan bulat yang akan digunakan untuk mempelajari data di dalamnya oleh mesin. One hot encoding adalah teknik yang kami gunakan untuk merepresentasikan variabel kategori sebagai nilai numerik dalam model pembelajaran mesin. Dalam satu hot coding, setiap kata yang merupakan bagian dari data teks tertentu ditulis dalam bentuk vektor. Setiap kata ditulis atau dikodekan sebagai satu hot vector, dengan masing-masing hot vector bersifat unik, contoh proses stemming dan tokenisasi disajikan pada Tabel 2.

Tabel 2. Proses Stemming dan Tokenisasi

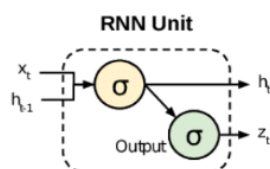
Memasukkan	Hasil Steming	Hasil Tokenisasi
berita tentang pernyataan pt jakarta propertindo jakpro menyebut lahan yang dicaplok ruko	berita tentang nyata pt jakarta propertindo jakpro sebut lahan caplok ruko	[31, 229, 4, 230, 89, 231]

RNN memiliki putaran umpan balik di lapisan berulang. Hal ini memungkinkan mereka menyimpan informasi dalam 'memori' seiring waktu. Mungkin sulit untuk melatih RNN untuk memecahkan masalah yang memerlukan pembelajaran ketergantungan temporal jangka panjang.

$h(t)$ di RNN mewakili keadaan tersembunyi pada waktu dan bertindak sebagai "memori" jaringan. Persamaan 1 dihitung berdasarkan masukan saat ini dan keadaan tersembunyi dari langkah waktu sebelumnya $h(t)$.

$$h(t) = f(Ux_t + Wh_{t-1}) \quad (1)$$

x_t diambil sebagai input ke jaringan pada langkah waktu. Misalnya dapat berupa vektor one-hot yang sesuai dengan kata dalam sebuah kalimat. Fungsi tersebut dianggap sebagai transformasi non-linier seperti, ReLU, \tanh



Matriks konfusi adalah alat analitik prediktif yang menampilkan dan membandingkan nilai aktual atau nilai sebenarnya dengan nilai model prediksi yang dapat digunakan untuk menghasilkan metrik evaluasi seperti Presisi yang dijelaskan pada Persamaan 2, Recall yang dijelaskan pada Persamaan 3, Akurasi yang dijelaskan pada Persamaan 4 dan Skor F1 dijelaskan dalam Persamaan 5. Ada empat nilai yang dihasilkan dalam tabel matriks konfusi, antara lain True Positive (TP), False Positive (FP), False Negative (FN), dan True Negative (TN).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

$$\text{F1 - Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

3. Hasil Penelitian dan Pembahasan

3.1 Evaluasi Perbandingan Data RNN

Pada klasifikasi yang telah dilakukan pada data dokumen berita Bahasa Indonesia menggunakan Recurrent Neural Network, penelitian ini dibagi menjadi dua kelompok, yakni RNN yang tidak dilakukan augmentasi dan balancing data, serta RNN yang telah dilakukan augmentasi menggunakan teknik Back Translation. Hal ini bertujuan untuk memperlihatkan perbedaan antara klasifikasi data menggunakan RNN tanpa augmentasi dan dengan augmentasi, serta untuk menilai sejauh mana peningkatan kinerja model setelah dilakukan augmentasi. Hal ini dapat memperlihatkan perbedaan antara Klasifikasi data menggunakan RNN pada umumnya, dan juga pada data yang telah di Augmentasi dengan Back Translation.

Tabel 3. Kategori Dataset sebelum Augmentasi

Kategori	Total
Bola	300
Hype	250
Uang	231
Berita	300
Otomotif	300
Daerah	300
Tren	298

Sumber: Data Primer diolah secara pribadi, 2024

Tabel 4. Kategori Dataset Setelah Augmentasi

Kategori	Total
Bola	300
Hype	300
Uang	300
Berita	300
Otomotif	300
Daerah	300
Tren	300

Dapat terlihat perbedaan yang signifikan antar dataset sehingga dilakukan pengujian terhadap kedua untuk menentukan hasil evaluasi antara teks klasifikasi sebelum dilakukan Augmentasi data dan teks setelah Augmentasi dengan Back Translation. Kedua ini menjadi perbandingan untuk memperlihatkan hasil output yang menjadi pelatihan pada metode klasifikasi saat ini. Pelatihan disini dilakukan dengan menggunakan:

Tabel 5. Dasar Train dataset

Parameter	Nilai
Batch Size	32
Dropout Rate	0.2
RNN Units	64
Embedding Input Dim	5000
Activation Function	Softmax
Optimizer	Adam

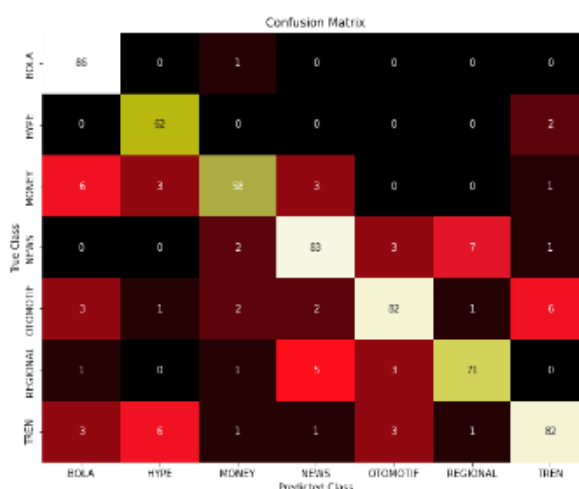
3.2 Perbandingan Hasil Pelatihan Sebelumnya Menggunakan Neural Network

Dari pengambilan dataset yang telah dilakukan Augmentasi dengan Back Translation dan yang sebelum dilakukan Augmentasi kita dapat menghasilkan perbandingan antara kedua dataset yang ada. Hal ini menunjukkan perbedaan yang dihasilkan antara pelatihan model dengan hanya melakukan pemotongan data dan mengambil kelas tertinggi dan data yang telah di seimbangkan Kembali dengan Augmentasi Back Translation. Berikut adalah hasil yang didapatkan antara kedua dataset.

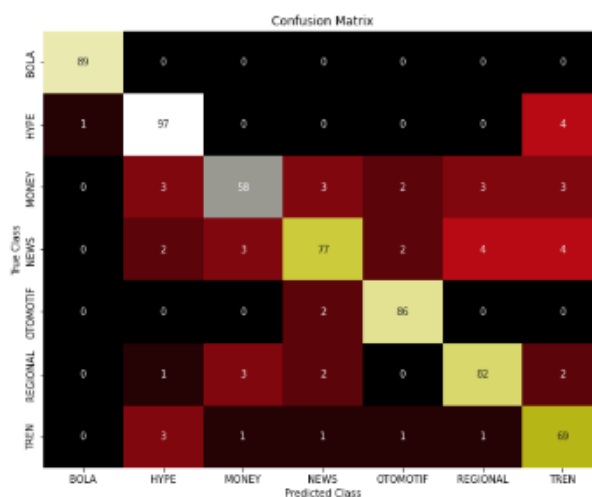
Tabel 6. Perbandingan Hasil Training

Model	Akurasi	Presisi	Recall
Naïve Bayes Bayesian Boosting	73.2%	73.2%	73.2%
RNN	92%	92%	92%
RNN dengan Augmentasi Back Translation	96%	96%	96%

Dari hasil diatas terdapat kenaikan dari hasil training sebelumnya menggunakan Conventional Naïve Bayes dan yang telah dilakukan dengan Neural Network yang mana sebelumnya setinggi 73.2% dapat meningkat signifikan dengan menggunakan Neural Network itu sendiri hingga diatas 90% baik tanpa Augmentasi dan juga dengan Augmentasi dengan menggunakan Back Translation. Berikut merupakan hasil Confusion Matrix pada hasil training RNN.



Gambar 5. Confusion Matrix Dataset Tanpa Augmentasi



Gambar 6. Confusion Matrix Data Augmentasi dengan Back Translation

Dari Confusion Matrix yang tertera pada Gambar 5 dan Gambar 6 terdapat perbedaan pada tiap kelas. Pada Gambar 5, confusion matrix tanpa augmentasi menunjukkan bahwa terdapat banyak kesalahan klasifikasi terutama pada kategori-kategori yang jumlah datanya tidak seimbang. Hal ini terlihat dari jumlah False Positives (FP) dan False Negatives (FN) yang cukup tinggi.

3.3 Hasil Implementasi Neural Network

Dari hasil evaluasi dari kedua model RNN, baik yang tanpa augmentasi maupun yang menggunakan augmentasi data dengan Back Translation. Peningkatan kinerja model setelah dilakukan augmentasi dapat diatribusikan kepada beberapa faktor, antara lain:

3.3.1 Peningkatan Variasi Data

Augmentasi data menggunakan teknik Back Translation menghasilkan variasi data yang lebih kaya. Dengan menerjemahkan teks ke bahasa lain dan kemudian menerjemahkannya kembali ke bahasa asli, data baru yang dihasilkan tetap memiliki makna yang sama tetapi dengan struktur kalimat yang berbeda. Ini memberikan model lebih banyak contoh untuk belajar, sehingga meningkatkan kemampuannya dalam mengenali pola-pola yang lebih beragam.

3.3.2 Penurunan Overfitting

Overfitting terjadi 358 bahasa model terlalu cocok dengan data pelatihan dan tidak dapat melakukan generalisasi dengan baik pada data uji. Dengan lebih banyak data, model dapat belajar untuk mengenali pola umum daripada hanya menghafal data pelatihan. Penurunan overfitting ini sangat penting dalam konteks klasifikasi teks, di mana variasi dalam 358 bahasa dan struktur kalimat dapat menyebabkan model yang overfit tidak dapat mengenali pola-pola baru yang mungkin muncul dalam data uji.

3.3.3 Peningkatan Kualitas Data

Augmentasi data juga membantu dalam menyeimbangkan distribusi data. Dalam dataset asli, mungkin terdapat ketidakseimbangan antara kategori-kategori yang ada. Misalnya, beberapa kategori berita mungkin memiliki lebih banyak contoh daripada yang lain. Ketidakseimbangan ini dapat menyebabkan model menjadi bias terhadap kategori yang lebih banyak diwakili dalam data pelatihan. Dengan augmentasi data, distribusi data dapat dibuat lebih seimbang, sehingga model dapat belajar secara adil dari semua kategori.

3.3.4 Peningkatan Metrik Evaluasi

Hasil evaluasi menunjukkan bahwa model dengan augmentasi data memiliki peningkatan signifikan dalam semua metrik utama: akurasi, precision, recall, dan F1-Score. Berikut adalah perbandingan metrik evaluasi antara model tanpa augmentasi dan model dengan augmentasi:

Secara keseluruhan, penelitian ini menunjukkan bahwa penggunaan augmentasi data menggunakan Back Translation dapat secara signifikan meningkatkan kinerja model RNN dalam klasifikasi dokumen berita berbahasa Indonesia. Hanya saja, perlu perhatian lebih lanjut mengenai Overfitting yang mungkin terjadi pada saat penggunaan *Neural Network* terutama RNN dalam pelatihan karena resiko overfitting dapat terjadi lebih tinggi daripada penggunaan model lainnya.

4. Kesimpulan

Penelitian ini menunjukkan bahwa model Recurrent Neural Network menawarkan peningkatan performa yang signifikan dibandingkan dengan model Naïve Bayes Classifier yang sering dijadikan benchmark dalam studi sebelumnya. Model RNN menunjukkan akurasi yang lebih tinggi serta peningkatan pada metrik precision, recall, dan F1-Score, yang menunjukkan kemampuan yang lebih baik dalam mengklasifikasikan teks berita bahasa Indonesia. Penggunaan teknik augmentasi data dengan *Back Translation* mampu membantu tantangan ketidakseimbangan kelas dalam dataset. Teknik augmentasi ini meningkatkan keberimbangan data dan secara langsung memengaruhi peningkatan akurasi model klasifikasi. Hasil penelitian ini tidak hanya memperkuat literatur dalam bidang klasifikasi teks tetapi juga membuka peluang untuk aplikasi teknik deep learning dalam berbagai aspek penelitian dan pengembangan sistem informasi yang lebih luas.

Referensi

- [1] M. J. Denny and A. Spirling, "Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It," *Politycal Analysis*, pp. vol. 26, no. 2, 168–189, April 2018.

-
- [2] L. Hickman, S. Thapa, . L. Tay, . M. Cao and P. Srin, "Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations," *Organ Res Methods Vol, 25*, p. 114–146, January 2022.
- [3] C. P. Chai, "Comparison of text preprocessing methods," *Nat Lang Eng, vol. 29*, p. 509–553, 2023.
- [4] Muhammad Ichwan Utari and Henny Medyawati, "Classification of News Types by Implementing Enhanced Confix Stripping Stemmer," *International journal of engineering technologies and management research*, vol. 6, no. 5, pp. 135–141, Mar. 2020, doi: <https://doi.org/10.29121/ijetmr.v6.i5.2019.380>
- [5] K. Ivanedra and M. Mustikasari, "Implementasi Metode Reccurrent Neural Network pada Text Summarization dengan Teknik Abstraktif," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 4, p. 377, Jul. 2019, doi: <https://doi.org/10.25126/jtiik.2019641067>.
- [6] D. S.K, in *Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifier Dengan Seleksi Fitur Dan Boosting*, 2017.
- [7] A. Setiawan, L. W. Santoso, and R. Adipranata, "Klasifikasi Artikel Berita Bahasa Indonesia Dengan Naive Bayes Classifier," *Jurnal Infra*, vol. 8, no. 1, pp. 146–151, Apr. 2020.
- [8] Iffitah Athiyyah Rahma and Lya Hulliyyatus Suadaa, "Penerapan Text Augmentation untuk Mengatasi Data yang Tidak Seimbang pada Klasifikasi Teks Berbahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer/Jurnal teknologi informasi dan ilmu komputer*, vol. 10, no. 6, pp. 1329–1340, Dec. 2023.
- [9] E. Prasetya and R. Nuraeni, "Deteksi Berita Palsu dengan Pendekatan RNN," *Jurnal Keamanan Siber*, vol. 4, no. 2, pp. 187-194, 2021.
- [10] A. D. Rendragraha, M. A. Bijaksana, and A. Romadhony, "Pendekatan Metode Transformers untuk Deteksi Bahasa Kasar dalam Komentar Berita Online Indonesia," *e-Proceeding of Engineering*, vol. 8, no. 2, pp. 3385-3395, Apr. 2021.

