

Pengaruh Penyeimbangan Data Pada Klasifikasi Terjemahan Al-Quran Dengan Metode Naïve Bayes dan Long Short Term Memory

Sulistia Ningsih, Nazruddin Safaat H*, Surya Agustian, Yusra, Eka Pandu Cynthia

Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Email: ¹12050126567@students.uin-suska.ac.id, ^{2,*}nazruddin.safaat@uin-suska.ac.id, ³surya.agustian@uin-suska.ac.id,

⁴yusra@uin-suska.ac.id, ⁵eka.pandu.cynthia@uin-suska.ac.id

Email Penulis Korespondensi: nazruddin.safaat@uin-suska.ac.id

Submitted: 15/05/2024; Accepted: 31/05/2024; Published: 31/05/2024

Abstrak—Al-Quran adalah sebuah kitab suci umat islam yang merupakan pedoman hidup bagi seluruh umat manusia. Mempelajari dan memahami terjemahan Al-Quran tidaklah mudah, salah satu cara yang dapat dilakukan adalah dengan mengklasifikasikan terjemahan ayat Al-Quran kedalam topik-topik yang ada. Penelitian ini menggunakan metode *Naïve Bayes* dan LSTM dalam proses klasifikasi. Data yang digunakan bersumber dari data terjemahan Al-Quran dalam bahasa Indonesia yang telah dilabeli berdasarkan klasifikasi *multi-class*. Salah satu masalah utama yang dihadapi adalah ketidakseimbangan data. Untuk mengatasi masalah tersebut maka dilakukan proses *balancing data*, teks *preprocessing*, *feature construction* dan *feature extraction* dengan menggunakan teknik *bag of words* dan *TF.IDF*. Hasil penelitian menunjukkan bahwa model Naïve Bayes yang paling optimal mendapatkan nilai rata-rata 55.39% pada data test juz 30, 61.59% pada data test juz 10-20, dan 59.53% pada data test juz 25-28. Sedangkan model LSTM yang paling optimal adalah 58.02% pada data test juz 30, 59.64% pada data test juz 10-20, dan 58.59% pada data test juz 25-28. Berdasarkan nilai tersebut metode Naïve Bayes dan LSTM dapat menjadi pilihan yang baik untuk tugas klasifikasi. Tujuan utama dari penelitian ini adalah untuk meningkatkan performa klasifikasi serta membandingkan akurasi antara *Naïve Bayes* dan LSTM.

Kata Kunci: Al-Quran; Klasifikasi; Naïve Bayes; LSTM; Terjemahan Al-Quran

Abstract—The Al Qur'an is a holy book of Muslims which is a guide to life for all mankind. Studying and understanding the translation of the Al-Quran is not easy, one way that can be done is to classify the translation of Al-Quran verses into existing topics. This research uses Naïve Bayes and LSTM methods in the classification process. The data used comes from translation data of the Al-Quran in Indonesian which has been labeled based on multi-class classification. One of the main problems faced is data imbalance. To overcome this problem, data balancing, text preprocessing, feature construction and feature extraction processes were carried out using the Bag of Words (BoW) and TF.IDF techniques. The research results indicate that the most optimal Naïve Bayes model achieved an average accuracy of 55.39% on test data from juz 30, 61.59% on test data from juz 10-20, and 59.53% on test data from juz 25-28. Meanwhile, the most optimal LSTM model yielded an accuracy of 58.02% on test data from juz 30, 59.64% on test data from juz 10-20, and 58.59% on test data from juz 25-28. The main aim of this research is to improve classification performance and compare the accuracy between naïve Bayes and lstm.

Keywords: Al-Quran; Classification; Naïve Bayes; LSTM; Quran Translation

1. PENDAHULUAN

Al-Quran adalah sebuah kitab suci dalam islam yang dianggap sebagai wahyu Allah yang disampaikan kepada Nabi Muhammad SAW. Umat islam meyakini bahwa Al-Quran adalah kitab yang dijamin keasliannya oleh Allah yang berfungsi sebagai sumber petunjuk dan hikmah bagi umat muslim [1]. Al-Quran berasal dari kata kerja bahasa Arab yaitu "*qara'a-yagrauQur'an*," yang artinya bacaan [2]. Mempelajari Al-Quran merupakan kewajiban utama setiap muslim. Mempelajari dan memahami terjemahan Al-Quran tidaklah mudah, salah satu cara yang dapat dilakukan adalah dengan mengklasifikasikan terjemahan ayat Al-Quran kedalam topik-topik yang ada.

Setiap ayat dalam Al-Qur'an memiliki signifikansi yang beragam dan seringkali mencakup lebih dari satu topik pembahasan. Berdasarkan penelitian [3] topik yang dibahas ayat-ayat Al-Qur'an dapat diklasifikasikan ke dalam 15 kelas, yaitu Alkanul Islam, Iman, Al-Quran, Ilmu dan Cabang-Cabangnya, Amal, Dakwah, Jihad, Manusia dan Hubungan Kemasyarakatan, Akhlak, Peraturan yang Berhubungan Dengan Harta, Hal-Hal yang Berkaitan dengan Hukum, Negara dan Masyarakat, Pertanian dan Perdagangan, Sejarah dan Kisah-Kisah, dan Agama-Agama. Hasil penelitian tersebut mendapatkan nilai *hamming loss* terbaik yaitu 0,1247 dan menggunakan metode *multinomial Naïve Bayes*.

Untuk mempermudah mempelajari serta mengetahui topik-topik pada Al-Quran diperlukan sebuah sistem klasifikasi yang dapat mengkategorikan dan mengidentifikasi Al-Quran ke dalam suatu kelas yang disebut dengan klasifikasi. Klasifikasi teks bertujuan untuk mengelompokkan teks ke dalam satu atau lebih kategori yang telah ditetapkan sebelumnya [4]. Ada banyak jenis teks klasifikasi tetapi pada penelitian ini metode yang digunakan adalah *Naïve bayes* dan LSTM. *Naïve Bayes* adalah metode klasifikasi yang sederhana dan cepat dalam menghitung probabilitas [5]. Sedangkan LSTM merupakan metode yang memanfaatkan klasifikasi data dalam jangka waktu yang panjang dan menyimpan informasinya dalam sel memori [6]. Kedua algoritma tersebut

dipilih berdasarkan berbagai sumber, keduanya terbukti mampu menyelesaikan klasifikasi dengan mencapai tingkat akurasi yang tinggi.

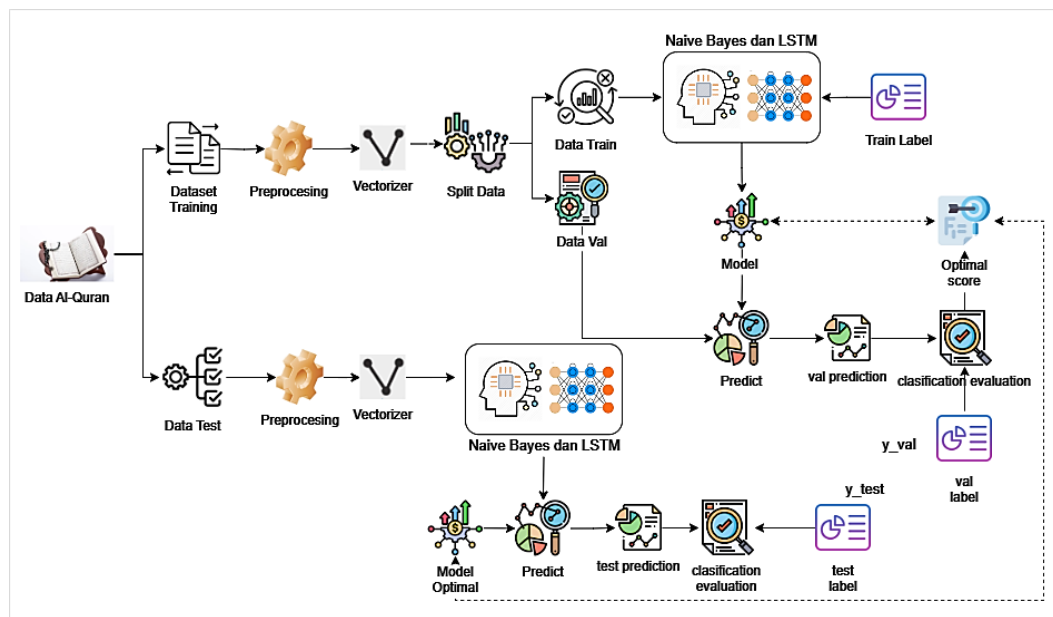
Salah satu penelitian yang dilakukan sebelumnya adalah analisis sentimen perbandingan LSTM dan *Naïve Bayes* yang hasilnya menunjukkan bahwa metode LSTM mencapai tingkat akurasi, *presisi*, dan *recall* sebesar 83.33%, sementara metode *Naïve Bayes* hanya mencapai 82% [7]. Selain itu metode *Naïve Bayes* juga digunakan pada pada klasifikasi teks dimana menghasilkan performa yang baik yaitu pada *bernoulli Naïve Bayes* sebesar 90,19% sedangkan *multinomial Naïve Bayes* sebesar 93,45% [8]. Selanjutnya pada penelitian [9], [10], [11], [12] juga telah menguji analisis sentimen menggunakan metode *Naïve Bayes* dan telah berhasil mencapai hasil akurasi rata-rata yaitu >80%.

Selanjutnya penelitian sebelumnya mengenai penggunaan LSTM yaitu pada klasifikasi Sentimen Vaksin COVID-19 di Twitter. Hasil penelitian menunjukan bahwa tingkat akurasi yang dicapai adalah sebesar 64% pada data validasi dan 66% pada data uji [13]. Selain itu pada penelitian [14], [15], [16], [17] juga melakukan penelitian dengan metode LSTM dan mendapatkan hasil yang lebih baik daripada metode konvensional lainnya yaitu rata-rata akurasi sebesar >85%.

Penelitian ini membandingkan metode LSTM dan *Naïve Bayes* untuk klasifikasi terjemahan ayat Al-Quran bahasa Indonesia kedalam kelas-kelas yang sudah ditentukan di dalam penelitian [18]. Adapun dari 15 kelas yang menjadi fokus penelitian adalah “Arkanul Islam”, “Iman”, “Al-Quran”, “Ilmu dan Cabang-Cabangnya”, “Amal”, dan “Undefined class” (kelas untuk ayat yang tidak termasuk kedalam salah satu topik) dengan menggunakan metode *Naïve Bayes* dan LSTM. Penelitian ini menggunakan data dalam bahasa Indonesia dan memiliki tujuan yaitu meningkatkan performa klasifikasi serta membandingkan hasil akurasi dari *Naïve Bayes* dan LSTM.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Desain Penelitian

Pada gambar 1 dapat dijelaskan bahwa penelitian ini menggunakan data dari terjemahan Al-Quran dimana data tersebut dibagi menjadi dua yaitu *data train* dan *data test*, kedua data tersebut melakukan proses *preprocessing*, *vectorizer* dan *split data*. Selanjutnya diterapkan metode yaitu *Naïve Bayes* atau LSTM dimana kedua metode tersebut menghasilkan model optimal yang akan diujikan pada data testing untuk mendapatkan hasil performa klasifikasi.

2.2 Dataset

Dataset yang dipakai dalam penelitian ini adalah terjemahan Al-Quran yang mulia dalam bahasa Indonesia [19]. Data terjemahan ini disediakan oleh Kementerian Agama Republik Indonesia. Setiap ayat pada data telah diberi label topik oleh [18] yang terdiri dari 15 kategori topik. Namun, pada penelitian ini hanya menggunakan 6 kategori topik saja, meliputi: “Arkanul Islam”, “Iman”, “Al-Quran”, “Ilmu dan Cabang-Cabangnya”, “Amal”, dan “Undefined class” (kelas untuk ayat yang tidak termasuk kedalam salah satu topik). Data yang digunakan bersifat multi-kelas (*multi-class*). Representasi dataset multi-kelas dari data tersebut dapat dilihat dari tabel berikut:

Tabel 1. Representasi data *multi-class*

Terjemahan	Kelas					
	Arkanul Islam	Iman	Al-Quran	Ilmu Dan Cabang-Cabangnya	Amal	Undefined Class
Pemilik hari pembalasan. (1:4)	1	1	0	0	0	0
Kitab (Al-Quran) ini tidak ada keraguan padanya; petunjuk bagi mereka yang bertakwa. (2:2)	0	1	1	0	1	0

Keterangan:

0 : terjemahan tersebut tidak ada di dalam kelas; 1 : terjemahan tersebut ada di dalam kelas.

2.3 Data Training dan Validasi

Data yang digunakan pada penelitian ini dibagi menjadi dua, yaitu *data training* dan *data test*. Adapun data yang digunakan untuk data training yaitu berjumlah 1383 ayat yang terdiri dari 22% yaitu surat ke-2 (Surah Al-Baqarah) sampai surat ke-6 (Surah Al-An'am) dan terjemahan ayat yang termasuk kedalam juz 29 (Surah Al-Mulk sampai Surah Al-Mursalat) pada Al-Quran. *Data training* ini mewakili profil surah-surah dengan ayat yang panjang-panjang (surah Al-Baqarah – Al-Maidah), sedang (surah Al-An'am), dan pendek (surah-surah pada juz 29). Sedangkan untuk data validasi diambil dari 10% data training.

2.4 Data Testing

Skenario pengujian dilakukan dengan pemilihan surah-surah yang mewakili profil *data testing*, yaitu:

1. Profil data dalam bentuk ayat Panjang, dipilih surah-surah pada juz 10-20
2. Profil data dalam bentuk ayat-ayat yang sedang, dipilih surah-surah pada juz 25-28.
3. Profil data uji dalam bentuk ayat-ayat pendek, dipilih surah-surah pada juz 30.

Dari scenario ini, kita bisa menilai apakah pemilihan profil *data testing* sudah cukup baik untuk mengklasifikasikan ayat-ayat Al-Qur'an lainnya.

2.5 Teks Preprocessing

Teks preprocessing penting untuk peringkasan dan pengelompokan dokumen. Meliputi *Tokenisasi*, *case folding*, *remove of punctuation*, *Stopwords*, dan *Stemming* yang dapat membantu mempersiapkan data agar lebih mudah dipahami oleh model.

- a. *Tokenisasi*: proses membagi teks kalimat menjadi bagian tertentu atau memecah kalimat menjadi kata-kata [20].
- b. *Remove of punctuation*: merupakan cara membersihkan Data yang mengandung angka, tanda baca, tautan, tanda pagar, dan penanda pengguna [21].
- c. *Stopwords*: Menghilangkan kata-kata umum yang dianggap tidak memberikan makna, seperti kata sambung dan kata keterangan [22].
- d. *Stemming*: Proses mengubah kata berimbuhan menjadi kata dasar [23].

Teks *preprocessing* ini diterapkan pada *data train* maupun *data test*. Penelitian ini mengikuti teknis klasifikasi optimasi teks *preprocessing* seperti pada penelitian [24], [25] untuk mencari nilai yang paling optimal. Tetapi ada beberapa kelas yang tidak menggunakan proses *stemming* dan *stopwords* untuk mendapatkan nilai *f1-score* dan akurasi yang baik dan dapat dilihat pada tabel 4 dan 8.

2.6 Vectorizer

Feature construction atau pembentukan fitur baru dari data mentah, membutuhkan waktu dan observasi yang cermat karena dilakukan secara manual untuk memastikan fitur yang dihasilkan memberikan manfaat optimal [26]. Pada penelitian ini *feature construction* dilakukan menggunakan proses *vectorizer* dan dibagi menjadi dua tahapan yaitu *feature extraction* dan *feature selection*. *Feature extraction* adalah di mana sistem mengubah kata-kata menjadi *vector* yang terdiri dari dua fase: ekstraksi primer yang mengonversi data menjadi format standar, dan kemudian mengubahnya menjadi *vector*. Ekstraksi fitur dapat dilakukan menggunakan representasi *bag of words*, metode yang umum digunakan NLP [27]. *Feature selection* adalah tahap dalam preprocessing data untuk menyeleksi fitur penting dalam analisis, meningkatkan akurasi model, dan mengurangi waktu komputasi. Ini dilakukan melalui *TF.IDF* dan *word count*, teknik untuk mengukur nilai frekuensi kata dalam dokumen [28]

2.7 Naïve Bayes

Naïve Bayes merupakan salah satu metode klasifikasi yang dipilih karena kesederhanaan dan efisiensinya. Metode ini efektif digunakan pada dataset dalam jumlah besar dan mampu mengatasi nilai data yang hilang [29]. Pada penelitian ini metode *Naïve Bayes* menggunakan *library scikit-learn (sklearn)* dan tidak ada penyetalan

parameter tuning karena dalam banyak kasus, parameter *default* dapat memberikan performa yang cukup baik pada data yang umum digunakan.

2.8 Long Short Term Memory

Metode LSTM adalah salah satu metode *deep learning* yang bisa diterapkan dalam NLP, termasuk terjemahan teks, pengenalan ucapan, serta klasifikasi baik gambar maupun teks [30]. LSTM adalah pengembangan dari RNN dengan tambahan sel memori untuk menyimpan informasi dalam rentang waktu yang lebih lama. Setiap sel memori dilengkapi dengan tiga gerbang: *input*, *forget*, dan *output*. Ini memungkinkan pengaturan, pengendalian, dan perhitungan nilai yang digunakan sebagai masukan pada langkah berikutnya [31]. LSTM memiliki *parameter tuning* yang dapat diatur sesuai dengan kebutuhan, seperti pada penelitian ini untuk meningkatkan nilai *f1-score* dan akurasi maka menggunakan *parameter tune* yang beragam seperti *dense layer* = 286 dan 320; *dropout* = 0.2 dan 0.5; *epoch* = 50; *batch size* = 32,64,128, dan 256.

2.9 Evaluasi

Pengujian model yang telah dilatih dilakukan dengan menggunakan data pengujian (*testing*). Data pengujian berisi kumpulan data yang digunakan untuk mengevaluasi kinerja model setelah proses pelatihan selesai. Setelah pengujian dilakukan, maka tahap selanjutnya adalah evaluasi kinerja model berdasarkan nilai *precision*, *recall*, *accuracy*, dan *f1-score* yang dihasilkan.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan total 6236 data. Data latih yang berisi kumpulan data digunakan untuk melatih atau membangun model, yang dalam penelitian ini terdiri dari surat ke-2 hingga ke-6 dan terjemahan ayat pada juz 29 Al-Qur'an, dengan total keseluruhan 1383 ayat. Data pengujian penelitian ini terdiri dari 16 juz yang meliputi juz 10 hingga 20 yang merupakan ayat-ayat panjang, juz 25 hingga 28 yang merupakan ayat-ayat sedang dan pendek, serta juz 30 yang merupakan ayat-ayat pendek. Total keseluruhan data pengujian adalah 3551 ayat. Data tersebut dipilih berdasarkan profil ayat pada Al-Quran yang berbeda-beda, yaitu profil ayat panjang, sedang, dan pendek, serta mengikuti pada data penelitian [3] yang telah dimodifikasi untuk mencapai hasil yang lebih optimal dan berbeda.

Jumlah data pada terjemahan Al-Quran ini tidak seimbang. Maka, dilakukan penanganan terhadap ketidakseimbangan jumlah data dengan melakukan proses *balancing data* berupa *oversampling* dan *undersampling* seperti yang dilakukan pada penelitian [32] untuk mendapatkan nilai optimal, sebelum melakukan proses tersebut penelitian ini memiliki data *baseline* yaitu data yang tidak melakukan proses *balancing data* dan tidak melakukan penyesuaian *parameter tune* pada kedua metode. Hasil pengujian *baseline Naïve Bayes* dapat dilihat pada tabel 2.

Tabel 2. Hasil *Baseline Data Training Naïve Bayes*

Kelas	Text Prep		Train Score		Val Score	
	Stop	Stem	F1	Acc	F1	Acc
Arkanul islam	P	P	94.00%	94.13%	64.93%	66.91%
Iman	P	P	91.59%	91.60%	51.79%	67.47%
Al-Quran	P	P	94.16%	94.17%	62.15%	76.98%
Ilmu dan cabang-cabangnya	O	O	97.57%	97.57%	66.37%	87.05%
Amal	P	P	91.94%	91.97%	59.33%	69.06%
Undefined class	P	P	96.73%	96.73%	58.25%	86.33%

Berdasarkan tabel 2 hasil dari *baseline data train* metode *Naïve Bayes* dimana metode tersebut melakukan proses *preprocessing* tetapi tidak melakukan *balancing data*. Selanjutnya dilakukan pengujian *data test* dan menghasilkan performa klasifikasi yang dapat dilihat pada tabel 3. Untuk menghasilkan nilai yang optimal maka perlu dilakukan optimasi data yang hasilnya dapat dilihat pada tabel 4.

Tabel 3. Hasil pengujian *baseline data test* metode *naïve bayes*

Kelas	Data Test Juz 30		Data Test Juz 10-20		Data Test Juz 25-28	
	F1	Acc	F1	Acc	F1	Acc
Arkanul islam	36.65%	45.88%	38.60%	52.67%	40.81%	59.57%
Iman	61.47%	67.78%	60.75%	74.44%	56.82%	64.99%
Al-Quran	48.51%	94.22%	48.05%	92.51%	47.67%	91.10%
Ilmu dan cabang-cabangnya	48.00%	92.29%	48.05%	92.51%	48.20%	93.04%
Amal	47.95%	92.12%	47.59%	90.81%	47.07%	88.95%

Kelas	Data Test Juz 30		Data Test Juz 10-20		Data Test Juz 25-28	
	F1	Acc	F1	Acc	F1	Acc
Undefined class	47.18%	89.32%	44.73%	80.93%	47.78%	91.50%
Rata-rata	48.29%	80.27%	47.96%	80.65%	48.06%	81.53%

Berdasarkan tabel 3 yang merupakan hasil pengujian *data test* dari semua kelas dengan menggunakan metode *Naïve Bayes*, yang terdiri dari nilai *f1-score* dan *accuracy* data test juz 30, juz 10 sampai 20, dan juz 25 sampai 28. Semua kelas berhasil mendapatkan nilai yang bervariasi pada ketiga *data test*. Pada tahap selanjutnya perlu melakukan proses optimasi untuk mendapatkan nilai optimal. Hasil optimal *data train* dapat dilihat pada tabel 4.

Tabel 4. Hasil optimal data train metode naïve bayes

Kelas	Balancing*		Text Prep		Train Score		Val Score	
	1	0	Stop	Stem	F1	Acc	F1	Acc
Arkanul islam	758	758	P	P	94.00%	94.13%	64.93%	66.91%
	486 530	758 714						
Iman	714	714	O	O	91.59%	91.60%	51.79%	67.47%
	146 1098	714 1098						
Al-Quran	1098	1098	P	P	94.16%	94.17%	62.15%	76.98%
	90 1154	1098 1154						
Ilmu dan cabang-cabangnya	1154	1154	O	O	97.57%	97.57%	66.37%	87.05%
	229 1015	1154 1015						
Amal	1015	1015	O	O	91.94%	91.97%	59.33%	69.06%
	66 1178	1015 1178						
Undefined class	1178	1178	O	O	96.73%	96.73%	58.25%	86.33%

)* balancing : Melakukan *undersampling* atau *oversampling* untuk kelas yang tidak seimbang. Dengan jumlah data yang berbeda-beda. Untuk mendapatkan nilai yang optimal maka dilakukan *oversampling* dari 486 data menjadi 758, 530 menjadi 714, 146 menjadi 1098, 90 menjadi 1154, 229 menjadi 1015, dan 66 menjadi 1178.

Text Prep : Menerapkan semua proses teks *preprocessing* dan tidak melakukan proses *stopword* dan *stemming*.

Train dan Val Score : Nilai *f1-score* dan *accuracy* dari data train dan validasi.

Tabel 4 memberikan gambaran hasil dari penerapan teks *preprocessing* dan *balancing data* sangat mempengaruhi performa klasifikasi pada setiap kelas dengan data yang berbeda. Dengan demikian hasil dari *f1-score* dan *accuracy* menjadi lebih baik dibandingkan hasil *baseline*. Kemudian dilakukan pengujian *data test* seperti pada tabel 5.

Tabel 5. Hasil pengujian model *Naïve Bayes* optimal terhadap data test

Kelas	Data Test Juz 30		Data Test Juz 10-20		Data Test Juz 25-28	
	F1	Acc	F1	Acc	F1	Acc
Arkanul islam	58.10%	58.14%	65.22%	65.70%	60.15%	61.31%
Iman	59.99%	60.07%	64.01%	67.20%	67.41%	67.55%
Al-Quran	53.81%	73.20%	57.19%	77.98%	58.26%	77.48%
Ilmu dan cabang-cabangnya	54.94%	71.10%	63.75%	85.82%	59.36%	81.68%
Amal	52.94%	70.93%	60.11%	77.38%	58.32%	72.16%
Undefined class	52.54%	75.48%	59.23%	75.34%	53.65%	78.40%
Rata-rata	55.39%	68.15%	61.59%	74.90%	59.53%	73.10%

Tabel ini menunjukkan bahwa setelah melakukan proses optimasi maka hasil dari nilai data test diatas lebih baik dibandingkan sebelumnya. Dapat dilihat bahwa terdapat banyak peningkatan antara *f1-score* dan *accuracy* pada semua *data test*. Pada ketiga *data test* terbukti bahwa pada profil ayat yang pendek mendapatkan nilai *f1-score* dan *accuracy* yang rendah, dan meningkat pada profil ayat yang sedang dan panjang. Oleh karena itu, proses *balancing data* sangat penting dalam meningkatkan performa klasifikasi. Selanjutnya melakukan pengujian *baseline* pada metode LSTM seperti pada tabel 6.

Tabel 6. Hasil *baseline* data train metode LSTM

Kelas	Text Prep		Parameter Tuning					Train Score		Val Score	
	Stop	Stem	DL	DO	EP	BS	EPO	F1	Acc	F1	Acc
Arkanul islam	P	P	286	0.2	50	64	6	99.67%	99.67%	62.35%	62.59%
Iman	P	P	286	0.2	50	64	11	99.77%	99.77%	74.83%	74.79%

Kelas	Text Prep		Parameter Tuning					Train Score		Val Score	
	Stop	Stem	DL	DO	EP	BS	EPO	F1	Acc	F1	Acc
Al-Quran	P	P	286	0.2	50	64	12	100%	100%	68.08%	89.93%
Ilmu dan cabang-cabangnya	O	O	286	0.2	50	64	14	100%	100%	99.13%	99.13%
Amal	P	P	286	0.2	50	64	20	100%	100%	97.04%	97.04%
Undefined class	P	P	286	0.2	50	64	5	99.67%	99.67%	98.72%	98.73%

Text Prep : Menerapkan semua proses teks *preprocessing* dan tidak melakukan proses *stopword* dan *stemming*.

Parameter Tuning : Melakukan penyesuaian *Dense Layer* (DL), *Drop Out* (DO), *Epoch* (EP), *Batch Size* (BS) dan *Epoch Optimal* (EPO).

Train dan Val Score : Nilai *f1-score* dan *accuracy* dari data train dan validasi.

Berdasarkan tabel 6 hasil dari *baseline* metode LSTM dimana metode tersebut berhasil mendapatkan nilai yang cukup baik dan melakukan proses *preprocessing* tetapi tidak melakukan *balancing data*. Selanjutnya dilakukan pengujian data *test* dan menghasilkan performa klasifikasi yang dapat dilihat pada tabel 7.

Tabel 7. Hasil pengujian *baseline* data test metode LSTM

Kelas	Data Test Juz 30		Data Test Juz 10-20		Data Test Juz 25-28	
	F1	Acc	F1	Acc	F1	Acc
Arkanul islam	53.59%	53.59%	61.22%	61.41%	59.05%	59.77%
Iman	57.11%	59.54%	60.52%	65.35%	60.79%	62.64%
Al-Quran	54.37%	81.96%	61.01%	88.77%	60.62%	86.69%
Ilmu dan cabang-cabangnya	59.34%	90.89%	61.34%	92.11%	57.38%	90.79%
Amal	61.62%	88.97%	59.28%	84.87%	59.81%	82.80%
Undefined class	50.77%	87.92%	49.90%	80.68%	49.31%	89.97%
Rata-rata	56.13%	77.15%	58.88%	78.87%	57.83%	78.78%

Berdasarkan Tabel 7 yang merupakan hasil pengujian *data test* dari semua kelas dengan menggunakan metode LSTM, yang terdiri dari nilai *f1-score* dan *accuracy* data test juz 30, juz 10 sampai 20, dan juz 25 sampai 28. Untuk mendapatkan nilai optimal dan peningkatan nilai *data test* maka dapat dilakukan semua proses optimasi dan penyesuaian *parameter tune* seperti pada tabel 8.

Tabel 8. Skor optimal untuk proses training LSTM

Kelas	Balancing*		Text Prep		Parameter Tuning					Train Score		Val Score	
	1	0	Stop	Stem	DL	DO	EP	BS	EPO	F1	Acc	F1	Acc
Arkanul islam	758	758	P	P	320	0.5	50	256	6	99.67%	99.67%	62.35%	62.59%
	486	758								%	%	%	%
	530	714								99.77%	99.77%	76.22%	76.22%
Iman	714	714	O	O	286	0.2	50	128	11	%	%	%	%
	146	1098								100%	100%	68.08%	89.93%
	1098	1098								%	%	%	%
Ilmu dan cabang-cabangnya	90	1154	O	O	286	0.2	50	128	4	100%	100%	99.13%	99.13%
	1154	1154								%	%	%	%
	229	1015								100%	100%	97.04%	97.04%
Amal	1015	1015	O	O	28	0.2	5	12	8	100%	100%	97.04%	97.04%
	66	1178								%	%	%	%
	1178	1178								99.67%	99.67%	98.72%	98.73%

)* balancing : Melakukan *undersampling* atau *oversampling* untuk kelas yang tidak seimbang. Dengan jumlah data yang berbeda-beda. Untuk mendapatkan nilai yang optimal maka dilakukan *oversampling* dari 486 data menjadi 758, 530 menjadi 714, 146 menjadi 1098, 90 menjadi 1154, 229 menjadi 1015, dan 66 menjadi 1178.

Text Prep : Menerapkan semua proses teks *preprocessing* dan tidak melakukan proses *stopword* dan *stemming*.

Parameter Tuning : Melakukan penyesuaian *Dense Layer* (DL), *Drop Out* (DO), *Epoch* (EP), *Batch Size* (BS) dan *Epoch Optimal* (EPO).

Train dan Val Score : Nilai *f1-score* dan *accuracy* dari *data train* dan validasi.

Tabel 8 memberikan gambaran hasil dari penyesuaian *parameter tuning*, teks *preprocessing* dan *balancing data* sangat mempengaruhi performa klasifikasi pada setiap kelas dengan data yang berbeda. Dengan demikian hasil dari *f1-score* dan *accuracy* menjadi lebih baik dibandingkan hasil *baseline*. Kemudian dilakukan pengujian *data test* seperti pada tabel 9.

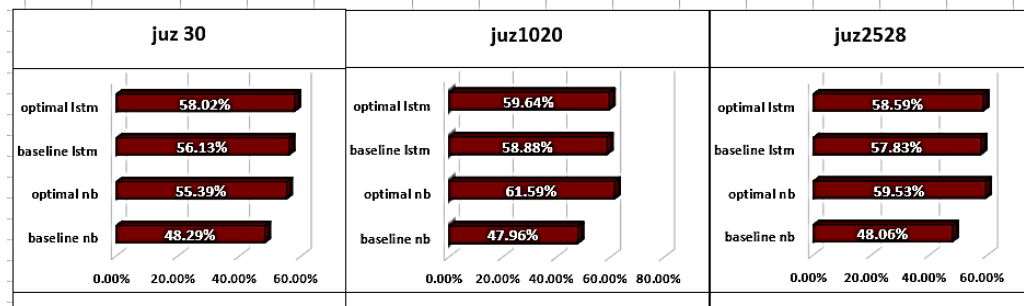
Tabel 9. Hasil pengujian model LSTM optimal terhadap data test

Kelas	Data Test Juz 30		Data Test Juz 10-20		Data Test Juz 25-28	
	<i>F1</i>	<i>Acc</i>	<i>F1</i>	<i>Acc</i>	<i>F1</i>	<i>Acc</i>
Arkanul islam	56.35%	56.39%	62.20%	62.96%	59.67%	60.90%
Iman	58.52%	61.82%	61.73%	66.15%	62.78%	63.77%
Al-Quran	57.05%	85.29%	61.87%	88.22%	59.31%	86.59%
Ilmu dan cabang-cabangnya	59.96%	88.97%	61.45%	91.96%	57.98%	91.30%
Amal	62.60%	90.37%	59.38%	84.27%	61.35%	83.52%
<i>Undefined class</i>	53.61%	87.22%	51.22%	79.98%	50.45%	88.84%
Rata-rata	58.02%	78.34%	59.64%	78.92%	58.59%	79.15%

Setelah mendapatkan hasil pengujian *data test* pada tabel 9, maka didapatkan nilai rata-rata perbandingan data test *baseline* dan optimal metode *Naïve Bayes* dan LSTM yang dapat dilihat pada tabel 10.

Tabel 10. Perbandingan nilai rata-rata data test *baseline* dan optimal

Model	Juz 30	Juz 1020	Juz 2528
<i>Baseline Naïve Bayes</i>	48.29%	47.96%	48.06%
<i>Optimal Naïve Bayes</i>	55.39%	61.59%	59.53%
<i>Baseline lstm</i>	56.13%	58.88%	57.83%
<i>Optimal lstm</i>	58.02%	59.64%	58.59%

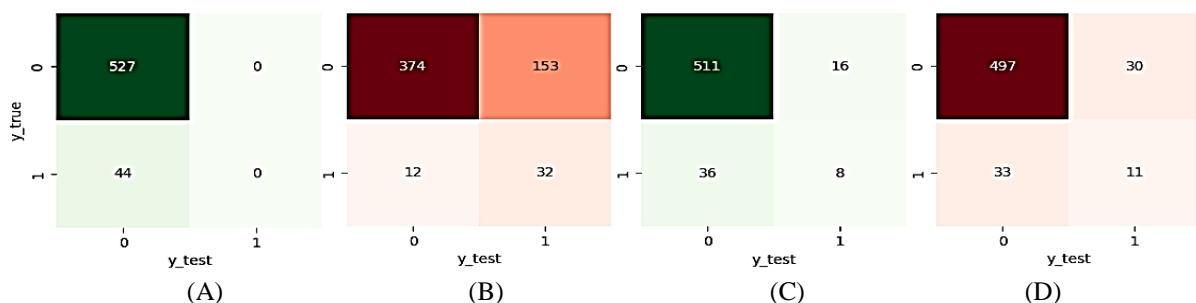
**Gambar 1.** Grafik 1 Perbandingan model baseline dan optimal

Berdasarkan tabel 10 dan grafik 1 setelah melakukan proses optimasi maka nilai dari kedua metode tersebut yaitu *Naïve Bayes* dan LSTM memiliki kinerja yang lebih baik dibanding sebelum melakukan optimasi atau *baseline*.

Tabel 11. *Confusion matrix* data test juz 30

Kelas	Model	<i>F1-score</i>
Ilmu dan cabang-cabangnya	<i>Baseline Naïve Bayes</i>	48.00%
	<i>Optimal Naïve Bayes</i>	54.94%
	<i>Baseline lstm</i>	59.34%
	<i>Optimal lstm</i>	59.96%

Berdasarkan tabel 11 *confusion matrix* yang ditampilkan adalah pada *data test* juz 30. Mempunyai nilai beragam serta mengalami peningkatan pada bagian *f1-score* setelah melakukan *balancing data*. *Confusion matrix* dapat dilihat pada gambar 2 dibawah:

**Gambar 2.** *Confusion Matriks* (A) *Naïve Bayes* Baseline (B) *Naïve Bayes* Optimal (C) *LSTM* Baseline (D) *LSTM* Optimal

Berdasarkan gambar 2 *Confusion matrix* pada *Naïve Bayes* terhadap *data test* juz 30 maka ilmu dan cabang-cabangnya tidak dapat diprediksi pada *baseline* (A). Tetapi setelah melakukan optimasi, sistem bisa

memprediksi ilmu dan cabang-cabangnya sebanyak 32 data dari 374 data dan masih ada 12 data yg masih terprediksi sebagai yang bukan ilmu dan cabang-cabangnya. kemudian terprediksi sebanyak 153 data yang seharusnya tidak mengandung ilmu dan cabang-cabangnya (B). Sedangkan pada LSTM hanya sedikit data yang bisa diprediksi yang berkaitan dengan ilmu dan cabang-cabangnya (C). Setelah dilakukannya optimasi *balancing data* dengan *parameter tune* yaitu *dense layer* = 286; *dropout* = 0.2; *epoch* = 50; *batch size* = 128. Maka sistem memprediksi ilmu dan cabang-cabangnya sebanyak 11 data dari 497 data dan 33 data yang terprediksi sebagai yang bukan ilmu dan cabang-cabangnya, dan terprediksi 30 data yang seharusnya tidak mengandung ilmu dan cabang-cabangnya (D). Kesalahan prediksi klasifikasi metode *Naïve Bayes* dan LSTM dapat dilihat pada tabel 12 dan 13.

Tabel 12. Kesalahan klasifikasi pada data test juz 30 kelas ilmu dan cabang-cabangnya metode *Naïve Bayes*

Terjemahan	Data Label	Hasil Prediksi
Bacalah, dan Tuhanmulah Yang Mahamulia, (96:3)	1	0
Yang mengajar (manusia) dengan pena. (96:4)	1	0

Keterangan:

0: terjemahan tersebut tidak ada di dalam kelas; 1: terjemahan tersebut ada di dalam kelas. Tabel diatas merupakan hasil dari kesalahan klasifikasi pada *data test* juz 30 pada kelas ilmu dan cabang-cabangnya dengan metode *Naïve Bayes* yang telah melakukan proses *balancing data*. Hasil prediksi menunjukkan bahwa pada prediksi pertama, model memprediksi ayat sebagai kategori yang salah, sedangkan pada prediksi kedua, hasil memprediksi ayat tidak sesuai dengan label seharusnya. Hal ini menunjukkan bahwa terdapat kesalahan dalam proses klasifikasi terhadap *data test* yang telah diberikan. Kesalahan klasifikasi dengan metode LSTM dapat dilihat pada tabel 13.

Tabel 13. Kesalahan klasifikasi pada data test juz 30 kelas ilmu dan cabang-cabangnya metode LSTM

Terjemahan	Data Label	Hasil Prediksi
Maka barangsiapa mengerjakan kebaikan seberat zarrah, niscaya dia akan melihat (balasan)nya, (99:7)	1	0
dan barangsiapa mengerjakan kejahatan seberat zarrah, niscaya dia akan melihat (balasan)nya. (99:8)	1	0

Keterangan:

0: terjemahan tersebut tidak ada di dalam kelas; 1: terjemahan tersebut ada di dalam kelas. Tabel diatas merupakan hasil dari kesalahan klasifikasi pada *data test* juz 30 pada kelas ilmu dan cabang-cabangnya dengan metode LSTM yang telah melakukan proses *balancing data*. Hasil prediksi menunjukkan bahwa pada prediksi pertama, model memprediksi ayat sebagai kategori yang salah, sedangkan pada prediksi kedua, hasil memprediksi ayat tidak sesuai dengan label seharusnya. Hal ini menunjukkan bahwa terdapat kesalahan dalam proses klasifikasi terhadap *data test* yang telah diberikan.

4. KESIMPULAN

Hasil pengujian pada data terjemahan Al-Quran menggunakan metode *Naïve Bayes* dan LSTM menunjukkan bahwa kedua metode tersebut mengalami peningkatan kinerja setelah dilakukan proses optimasi data, baik dari segi *f1-score* maupun akurasi. Metode *Naïve Bayes* dan LSTM dapat menjadi pilihan yang baik untuk tugas klasifikasi. Dalam setiap kelas, optimasi *balancing data* terbukti sangat penting untuk meningkatkan kemampuan model untuk memprediksi kelas yang kurang mewakili dalam kumpulan data, seperti penelitian ini menggunakan teknik *oversampling*. Teknik *undersampling* tidak digunakan dalam penelitian ini karena terbukti tidak meningkatkan akurasi, bahkan mengurangi performa *f1-score*nya. Model *Naïve Bayes* yang paling optimal mendapatkan nilai rata-rata 55.39% pada *data test* juz 30, 61.59% pada *data test* juz 10-20, dan 59.53% pada *data test* juz 25-28. Sedangkan model LSTM yang paling optimal adalah 58.02% pada *data test* juz 30, 59.64% pada *data test* juz 10-20, dan 58.59% pada *data test* juz 25-28. Berdasarkan nilai rata-rata tersebut, dapat disimpulkan bahwa pada *data test* juz 30 metode LSTM lebih unggul dibanding metode *Naïve Bayes*. Sedangkan pada juz 10-20 dan juz 25-28 metode *Naïve Bayes* lebih unggul dibandingkan metode LSTM. Untuk pengembangan penelitian selanjutnya, dapat dilakukan eksperimen dengan menggunakan berbagai arsitektur dan teknik pengolahan data lainnya untuk meningkatkan kinerja klasifikasi.

REFERENCES

- [1] S. Hidayatullah, "Classification of Al-Qur'an Arabic Verses Used Naive Bayes," *Jurnal Mantik*, vol. 6, no. 1, pp. 639–648, 2022.
- [2] S. Th. I. M. A. J. Muhammad Yasir, *Studi Al-Qur'an*. 2016.

- [3] R. Ananda Pane and M. Syahrul Mubarak, "Klasifikasi Multi-Label Pada Topik Ayat Al-Quran Terjemahan Bahasa Inggris Menggunakan Multinomial Naive Bayes A Multi-Label Classification On Topics Of Quranic Verses In English Translation Using Multinomial Naive Bayes," *e-Proceeding of Engineering*, vol. 5, no. 1, pp. 1551–1555, 2018.
- [4] E. Supriyati and M. Iqbal, "Pengukuran Similarity Tema Pada Juz 30 Al Qur'an Menggunakan Teks Klasifikasi," *Jurnal SIMETRIS*, vol. 9, no. 1, pp. 361–370, 2018, [Online]. Available: <https://translate.google.co.id/>
- [5] M. R. Choirulfikri, K. M. Lhaksamana, and S. Al Faraby, "A Multi-Label Classification of Al-Quran Verses Using Ensemble Method and Naïve Bayes," *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 4, pp. 473–479, Mar. 2022, doi: 10.47065/bits.v3i4.1287.
- [6] Y. Astari and S. Wahib Rozaqi, "Analisis Sentimen Multi-Class pada Sosial Media menggunakan metode Long Short-Term Memory (LSTM)," *JLK*, vol. 4, no. 1, pp. 8–12, 2021.
- [7] A. Rahman *et al.*, "Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 8, no. 2, pp. 299–303, 2022.
- [8] N. S. Wardani, A. Prahutama, and P. Kartikasari, "Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naïve Bayes Untuk Model Bernoulli Dan Multinomial," *Jurnal Gaussian*, vol. 9, no. 3, pp. 237–246, 2020, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [9] M. Muslimin and V. Lusiana, "Analisis Sentimen Terhadap Kenaikan Harga Bahan Pokok Menggunakan Metode Naive Bayes Classifier," *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, pp. 1200–1209, 2023, doi: 10.30865/mib.v7i3.6418.
- [10] K. A. Nugraha, "Analisis Sentimen Berbasis Emoticon pada Komentar Instagram Bahasa Indonesia Menggunakan Naïve Bayes," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 3, pp. 715–721, Dec. 2021, doi: 10.28932/jutisi.v7i3.4094.
- [11] S. W. Ritonga, . Y., M. Fikry, and E. P. Cynthia, "Klasifikasi Sentimen Masyarakat di Twitter terhadap Ganjar Pranowo dengan Metode Naïve Bayes Classifier," *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 1, pp. 134–143, Jun. 2023, doi: 10.47065/bits.v5i1.3535.
- [12] M. Suhendri and Y. Afrilia, "Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (Nbc)," *Sistemasi: Jurnal Sistem Informasi*, vol. 10, no. 2, pp. 268–279, May 2021, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [13] M. Ihsan, Benny Sukma Negara, and Surya Agustian, "LSTM (Long Short Term Memory) for Sentiment COVID-19 Vaccine Classification on Twitter," *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 13, no. 1, pp. 79–89, May 2022, doi: 10.31849/digitalzone.v13i1.9950.
- [14] B. Arief, H. Kholifatullah, and A. Prihanto, "Penerapan Metode Long Short Term Memory Untuk Klasifikasi Pada Hate Speech," *Journal of Informatics and Computer Science*, vol. 04, 2023.
- [15] F. N. Fajri and S. Syaiful, "Klasifikasi Nama Paket Pengadaan Menggunakan Long Short-Term Memory (LSTM) Pada Data Pengadaan," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, pp. 1625–1633, Dec. 2022, doi: 10.47065/bits.v4i3.2635.
- [16] W. Widayat, "Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning," *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, pp. 1018–1026, Jul. 2021, doi: 10.30865/mib.v5i3.3111.
- [17] P. A. Nabila, S. Soim, and A. S. Handayani, "Klasifikasi Kondisi Kendaraan Berpotensi Kecelakaan Berbasis Android Menggunakan Long Short Term Memory," *Jurnal Media Informatika Budidarma*, vol. 8, no. 1, pp. 30–40, Jan. 2024, doi: 10.30865/mib.v8i1.7005.
- [18] M. Alshammeri, E. Atwell, and M. Alsalka, "Classifying Verses of the Quran using Doc2vec," in *the 18th International Conference on Natural Language Processing*, 2021, pp. 284–288.
- [19] "Al-Quran yang mulia - Quran.com." Accessed: Feb. 14, 2024. [Online]. Available: <https://quran.com/id>
- [20] D. Jacarria Pangestu and A. Kodar, "Implementasi Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Terhadap Pelayanan Perusahaan Otobus Menggunakan Data Facebook (Studi Kasus: Grup Facebook Murni Jaya Lovers)," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 7, no. 3, pp. 156–160, 2022.
- [21] F. Hasibuan, W. Priatna, and T. Sri Lestari, "Analisis Sentimen Terhadap Kementerian Perdagangan Pada Sosial Media Twitter Menggunakan Metode Naïve Bayes Sentiment Analysis Of The Ministry Of Trade On Twitter Social Media Using Naïve Bayes Method," *Techno.COM*, vol. 21, no. 4, pp. 741–752, 2022.
- [22] U. Kurniasih and A. T. Suseno, "Analisis Sentimen Terhadap Bantuan Subsidi Upah (BSU) pada Kenaikan Harga Bahan Bakar Minyak (BBM)," *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, pp. 2335–2340, Oct. 2022, doi: 10.30865/mib.v6i4.4958.
- [23] N. Satya Marga, A. Rahman Isnain, and D. Alita, "Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naïve Bayes," *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 453, no. 4, pp. 453–463, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/informatika>
- [24] M. Sahbuddin and S. Agustian, "Support Vector Machine Method with Word2vec for Covid-19 Vaccine Sentiment Classification on Twitter," *Journal Of Informatics And Telecommunication Engineering*, vol. 6, no. 1, pp. 288–297, Jul. 2022, doi: 10.31289/jite.v6i1.7534.
- [25] I. H. Hasibuan, E. Budianita, S. Agustian, and Pizaini, "Klasifikasi Sentimen Komentar Youtube Tentang Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia U 20 Menggunakan Algoritma Naïve Bayes Classifier Tugas Akhir," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 5, no. 2, pp. 249–257, 2023, doi: 10.30865/json.v5i2.7096.
- [26] Ibnu Daqiqil Id, *Machine Learning*. 2021.
- [27] J. Nurvania and K. Muslim Lhaksamana, "Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short-Term Memory (LSTM)," *e-Proceeding of Engineering*, vol. 8, no. 4, pp. 4124–4134, 2021.
- [28] M. Fauzan, H. Junaedi, and E. Setyati, "Klasifikasi Al-Qur'an Terjemahan Bahasa Indonesia Dengan Menggunakan Algoritma Support Vector Machine (SVM)," *KONVERGENSI*, vol. 18, no. 2, pp. 42–49, 2022.

- [29] N. Widiastuti, A. Hermawan, and D. Avianto, "Implementasi Metode Naïve Bayes Untuk Klasifikasi Data Blogger," *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, pp. 985–994, Aug. 2023, doi: 10.29100/jipi.v8i3.3713.
- [30] Y. Romadhoni, K. Fahmi, and H. Holle, "Analisis Sentimen Terhadap PERMENDIKBUD No.30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 7, no. 2, pp. 118–124, 2022.
- [31] M. N. Farid and S. Ferdiana Kusuma, "Analisis Sentimen pada Media Sosial Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Berbasis Deep Learning," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 8, no. 1, pp. 44–49, 2022.
- [32] S. A. Afri Naldi, "Klasifikasi Sentimen Vaksin Covid-19 Menggunakan KNN Berdasarkan Word Embeddings Fasttext pada Twitter," *ZONAsi: Jurnal Sistem Informasi*, vol. 5, no. 2, pp. 323–333, May 2023.