

Komparasi Algoritma Machine Learning dan Deep Learning untuk Named Entity Recognition : Studi Kasus Data Kebencanaan

Nuli Giarsyani*, Ridho Rahmadi, Ahmad Fathan Hidayatullah

Program Studi Informatika Program Magister, Fakultas Teknologi Industri, Universitas Islam Indonesia

*Email: nuli.giarsyani@students.uui.ac.id

Info Artikel	Abstrak
<p>Kata Kunci : named entity recognition, disaster, machine learning, deep learning</p> <p>Keywords : named entity recognition, disaster, machine learning, deep learning</p> <p>Tanggal Artikel Dikirim : 24 April 2020 Direvisi : 20 Mei 2020 Diterima : 25 Mei 2020</p>	<p>Indonesia merupakan negara yang rawan akan bencana. Beberapa bencana alam yang sering tercatat sering terjadi di Indonesia adalah mulai dari gempa, tsunami, banjir, tanah longsor, gunung meletus dan masih banyak lagi yang lainnya. Adanya kejadian ini mendorong pengguna Twitter atau lembaga penanggulangan kebencanaan untuk mengunggah informasi tentang kondisi bencana dari tempat terjadinya bencana. Namun tidak semua tweet yang diunggah tersebut memuat informasi tentang bencana. Untuk dapat memanfaatkan informasi spesifik berkaitan dengan jenis jenis bencana, lokasi bencana, dan waktu kejadian, maka seseorang harus membaca secara keseluruhan isi tweet tersebut. Apabila jumlahnya banyak, maka membutuhkan waktu yang cukup lama untuk mendapatkan informasi bencana pada tweet tersebut. Oleh sebab itu, dibutuhkan <i>named entity recognition</i> (NER) yang dapat digunakan untuk mendapatkan informasi penting secara otomatis dalam waktu yang singkat. Penelitian ini bertujuan untuk melakukan NER pada data tweet dan melakukan analisa terhadap hasil dari dua pendekatan yaitu machine learning dan deep learning. Secara spesifik, kami membandingkan akurasi beberapa metode dari machine learning dan deep learning dan melakukan klasifikasi, dan menemukan bahwa pendekatan deep learning lebih unggul daripada pendekatan machine learning.</p>
	<p>Abstarct</p> <p><i>Indonesia is a country that is prone to disasters. Some of natural disasters that are often recorded in Indonesia are earthquakes, tsunamis, floods, landslides, volcanic eruptions and others. This incident prompted Twitter users or disaster management agencies to upload information about the condition of the disaster from the place of the disaster. However, not all tweets contain information about disasters. But not all tweets posted contain information about the disaster. To be able to utilize specific information relating to the type of disaster, the location of the disaster, and the time of the incident, one must read the entire contents of the tweet. If there are a lot of tweets, it takes a long time to get disaster information on the tweet. Therefore, we need named entity recognition (NER) which can be used to get important information automatically in a short time. This study aims to classify entity on tweet and analysis the results of two approaches namely machine learning and deep learning. Specifically, we compare the accuracy of several methods of machine learning and deep learning and classification. We find that the deep learning approach is superior to the machine learning approach.</i></p>

1. PENDAHULUAN

Saat ini media sosial telah menjadi bagian penting dari kehidupan seseorang. Berbagai kalangan, usia, dan hampir semua lapisan masyarakat Indonesia telah memiliki dan menggunakan media sosial sebagai salah satu sarana guna memperoleh dan menyampaikan informasi kepada publik [1]. Pada prakteknya, media sosial sering mengambil peran media tradisional dalam melaporkan peristiwa atau kejadian terkini [2], seperti bencana alam, di mana laporan atau informasi yang hadir bisa lebih cepat dan berasal langsung dari lokasi bencana [3]. Oleh sebab itu, media sosial banyak dimanfaatkan oleh organisasi atau lembaga yang bergerak pada bidang penanggulangan bencana untuk memberikan informasi terkait bencana.

Salah satu media sosial yang dimanfaatkan dalam melaporkan informasi bencana alam adalah Twitter. Dalam beberapa tahun terakhir, Twitter telah menjadi saluran utama untuk komunikasi selama bencana alam [4]. Hal tersebut dikarenakan Twitter merupakan salah satu media jejaring sosial dengan pengguna terbanyak diantara beberapa situs jejaring sosial yang ada [5]. Data menunjukkan bahwa Twitter memiliki sekitar 288 juta pengguna aktif, memposting lebih dari 500 juta tweet per hari, dan memiliki jaringan yang paling cepat berkembang [6]. Sedangkan untuk di negara Indonesia sendiri jumlah pengguna Twitter menempati peringkat ke 5 terbesar di dunia.

Indonesia merupakan negara yang rawan akan bencana. Beberapa bencana alam yang sering terjadi di Indonesia mulai dari gempa, tsunami, banjir, tanah longsor, gunung meletus dan masih banyak lagi yang lainnya. Adanya kejadian ini mendorong pengguna Twitter atau lembaga penanggulangan kebencanaan untuk mengunggah informasi tentang kondisi bencana dari tempat terjadinya bencana. Namun tidak semua tweet yang diunggah tersebut memuat informasi tentang bencana. Untuk dapat memanfaatkan informasi tersebut seperti mendapatkan informasi jenis bencana, lokasi bencana, dan waktu kejadian, maka seseorang harus membaca secara keseluruhan isi tweet tersebut. Apabila jumlahnya banyak, maka seseorang akan membutuhkan waktu yang cukup lama untuk mendapatkan informasi bencana pada tweet tersebut. Oleh sebab itu, dibutuhkan *named entity recognition* (NER) yang dapat digunakan untuk mendapatkan informasi penting secara otomatis dari tweet-tweet tersebut dalam waktu yang relatif singkat.

Named entity recognition adalah proses otomatis mengekstraksi entitas bernama yang dianggap penting di dalam sebuah teks dan menentukan kategorinya ke dalam kategori terdefinisi [7]. Sebagai contoh dalam tweet informasi bencana, NER dapat mengekstraksi jenis bencana, lokasi bencana, dan waktu kejadian. NER pada tweet informasi bencana biasanya dibutuhkan oleh pihak tertentu untuk mengidentifikasi bencana, dan memantau keadaan ketika terjadi bencana.

Penelitian ini bertujuan untuk melakukan *named entity recognition* guna mengidentifikasi dan mengklasifikasikan kata pada tweet yang memuat informasi bencana ke dalam entitas-entitas yang telah ditentukan. Selain itu, penelitian ini juga akan memberikan kontribusi terhadap NER pada domain kebencanaan dalam bahasa Indonesia, karena NER pada domain tersebut masih sangat terbatas. NER pada domain kebencanaan ini merupakan langkah awal untuk topik penelitian terkait *information extraction*, *question answering system*, dan sistem monitoring kebencanaan.

Pendekatan yang digunakan dalam penelitian ini adalah *machine learning* dan *deep learning*. Algoritma *machine learning* yang digunakan pada penelitian adalah *naive bayes*, *support vector machines*, *decision tree*, dan *random forest*. Sedangkan algoritma *deep learning* yang digunakan adalah *long short term memory* (LSTM), *convolutional neural network* (CNN), dan *gated recurrent units* (GRU). Pendekatan dengan berbagai macam metode ini bertujuan untuk memberikan gambaran secara komprehensif terhadap performa masing-masing metode yang harapannya dapat dijadikan landasan saintifik bagi kasus-kasus yang serupa.

2. METODE PENELITIAN

Tahapan yang akan dilakukan pada penelitian ini adalah sebagai berikut.

2.1 Studi Pustaka

Studi pustaka dilakukan untuk mengkaji dan mempelajari sumber literatur dan teori-teori yang berhubungan atau berkaitan dengan penelitian yang sedang dilakukan. Landasan yang diambil atau digunakan dalam studi literatur ini bersumber dari penelitian ilmiah, buku, artikel, jurnal, dan sumber pembelajaran lainnya.

2.2 Persiapan Data

Data yang digunakan pada penelitian ini bersumber dari Twitter. Data tersebut merupakan data mentah dari tweet-tweet yang berisi informasi tentang bencana di Indonesia. Jumlah data yang dikumpulkan sebanyak 378.602 tweet. Data tersebut diambil dari akun Twitter BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) dan BNPB (Badan Nasional Penanggulangan Bencana) baik yang berada di pemerintah pusat sampai dengan pemerintah daerah.

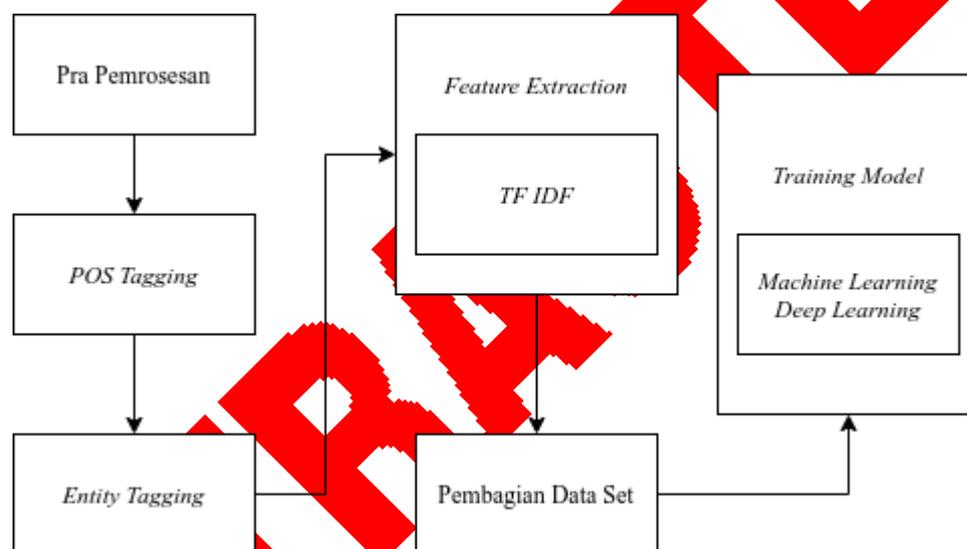
Data tweet diambil dengan menggunakan *library tweepy* dalam bahasa *python*. Proses pengumpulan data menggunakan *tweepy* dimulai dari melakukan proses autentikasi ke Twitter dengan menggunakan *method* atau *function* yang telah disediakan oleh *tweepy*. Proses autentikasi membutuhkan beberapa data diantaranya *consumer key*, *consumer secret*, *access token*, dan *access token secret*. Data untuk proses autentikasi diperoleh dari layanan *Twitter Developer*. Layanan tersebut disediakan oleh Twitter agar pengguna bisa menggunakan fitur-fitur yang ada pada Twitter.

Setelah melalui proses autentikasi, selanjutnya akan dilakukan proses pengambilan data tweet dengan memanfaatkan *method* atau *function* pada *library tweepy*. Kemudian data tweet tersebut dikumpulkan menjadi satu kedalam satu dokumen.

Pada penelitian ini, data tweet akan dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Data *training* nantinya akan digunakan untuk melatih dengan menggunakan beberapa algoritma klasifikasi dalam *machine learning* maupun *deep learning* sehingga menghasilkan sebuah model. Data *testing* akan digunakan untuk mengetes performa dari model yang telah dibuat.

2.3 Komputasi Model

Komputasi model yang akan dilakukan dalam penelitian ini terbagi menjadi beberapa bagian diantaranya pra pemrosesan, *part of speech tagging*, *entity tagging*, *feature extraction*, pembagian data set, dan *training model* seperti ditunjukkan gambar 1.



Gambar 1. Komputasi Model

2.4 Evaluasi

Evaluasi terhadap model yang telah dibangun dari proses *training* merupakan hal yang sangat penting. Tujuan evaluasi adalah untuk mengukur apakah model mampu melakukan klasifikasi dengan baik atau tidak. Hasil dari proses evaluasi ini digunakan untuk mengetahui kinerja dari model. Sehingga memperoleh mana model terbaik yang dihasilkan dari setiap algoritma yang telah disebutkan di atas. Metode evaluasi yang digunakan pada penelitian ini adalah perbandingan akurasi. Akurasi sebuah algoritma mengindikasikan persentase dari prediksi yang benar. Formula dari akurasi adalah

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Benar}}{\text{Jumlah Total Prediksi}} \quad (1)$$

atau lebih detailnya,

$$\text{Akurasi} = \frac{PB+NB}{PB+NB+PP+NP} \quad (2)$$

di mana *PB*= Positif Benar, *NB*= Negatif Benar, *PP*= Positif Palsu, dan *NP*= Negatif Palsu.

Keterangan :

PB: Jumlah data positif yang terklasifikasi dengan benar.

NB: Jumlah data negatif yang terklasifikasi dengan benar.

PP: Jumlah data negatif namun terklasifikasi sebagai data positif.

NP: Jumlah data positif namun terklasifikasi sebagai data negatif.

3. HASIL DAN PEMBAHASAN

Hasil penelitian *named entity recognition* ini berfokus pada analisis dari variasi ukuran sampel data set terhadap kinerja algoritma machine learning dan deep learning. Analisis performa kinerja dari algoritma diperoleh berdasarkan akurasi.

3.1 Hasil Akurasi

Hasil akurasi dibagi menjadi dua bagian, yaitu akurasi pada rasio 70%-30% dan akurasi pada rasio 80%-20%. Setiap algoritma pada kedua rasio menggunakan sample size yang berbeda-beda. Sample size yang digunakan mulai dari 50%, 60%, 70%, 80%, 90% dan 100% dari data tweet yang berjumlah 378.602.

1. Rasio 70%-30%

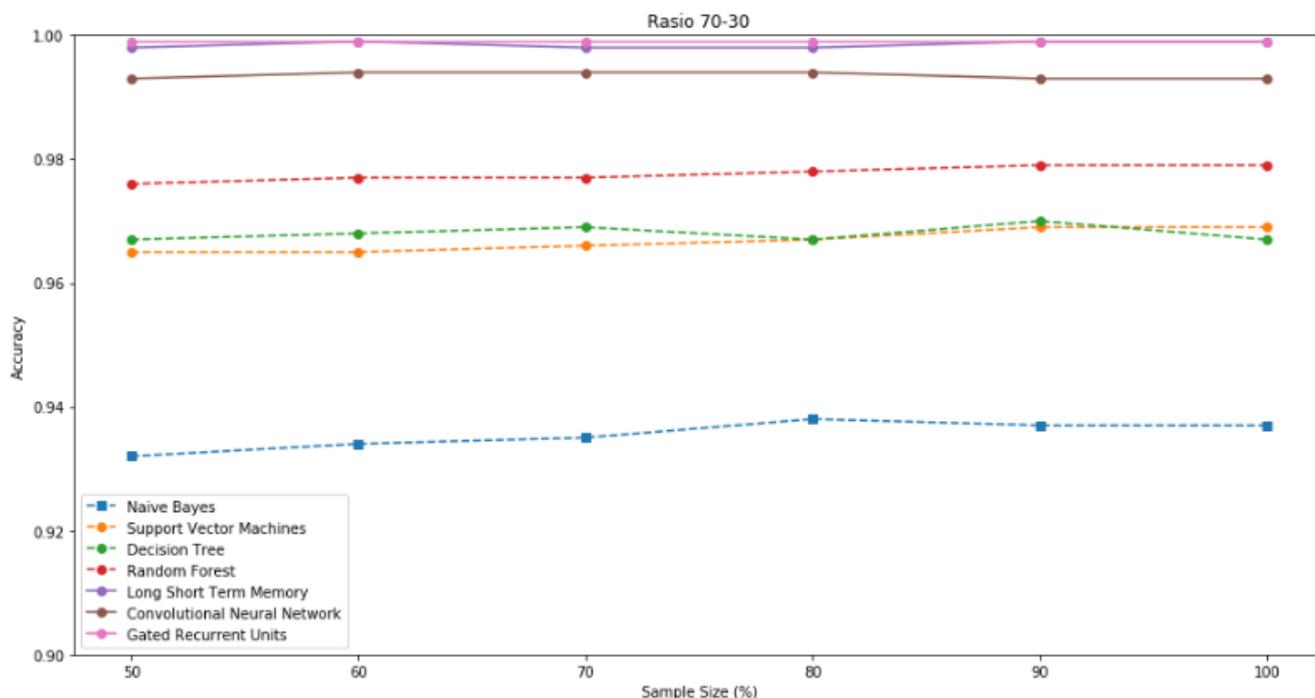
Tabel 1. Akurasi Rasio 70%-30%

Algoritma	Sample Size (%)						
	50	60	70	80	90	100	
Machine Learning	Naive Bayes	0.932	0.934	0.935	0.938	0.937	0.937
	SVM	0.965	0.965	0.966	0.967	0.969	0.969
	Decision Tree	0.967	0.968	0.969	0.967	0.97	0.967
	Random Forest	0.976	0.977	0.977	0.978	0.979	0.979
Deep Learning	LSTM	0.998	0.999	0.998	0.998	0.999	0.999
	CNN	0.993	0.994	0.994	0.994	0.993	0.993
	GRU	0.999	0.999	0.999	0.999	0.999	0.999

Berdasarkan dari hasil akurasi yang ditampilkan pada Tabel 1 di atas dapat disimpulkan bahwa pada *sample size* 50% akurasi terbaik dihasilkan oleh algoritma GRU sebesar 0.999, pada *sample size* 60% akurasi terbaik dihasilkan oleh algoritma LSTM dan GRU sebesar 0.999, pada *sample size* 70% dan 80% akurasi terbaik dihasilkan oleh algoritma GRU, pada *sample size* 90% dan 100% akurasi terbaik dihasilkan oleh algoritma LSTM dan GRU. Setiap algoritma pada setiap *sample size* memiliki selisih yang berbeda-beda, seperti pada *sample size* 50% rentang perbedaan akurasinya ada diantara 0.001- 0.067, pada *sample size* 60% selisihnya antara 0.005-0.065, pada *sample size* 70% antara 0.001-0.064, pada *sample size* 80% antara 0.001-0.061, pada *sample size* 90% dan 100% berada diantara 0.006-0.062.

Selain dilihat berdasarkan *sample size*, juga dapat disimpulkan berdasarkan algoritma, seperti pada algoritma *naive bayes* mendapatkan akurasi terbaik dengan nilai 0.938 pada *sample size* 80%, dan akurasi terburuk dengan nilai 0.932 pada *sample size* 50%. Algoritma SVM memperoleh akurasi terbaik pada *sample size* 90% dan 100% dengan nilai 0.969. Algoritma *Decision Tree* memperoleh hasil akurasi terbaik sebesar 0.97 pada *sample size* 90%. Algoritma *random forest* memperoleh hasil akurasi terbaik pada *sample size* 90% dan 100% dengan nilai 0.979. Algoritma LSTM memperoleh akurasi terbaik dengan nilai 0.999 pada *sample size* 60%, 90% dan 100%. Algoritma CNN memperoleh akurasi terbaik dengan nilai 0.994 pada *sample size* 60%, 70% dan 80%. Algoritma GRU memperoleh akurasi terbaik dengan nilai 0.999 pada semua *sample size*.

Pada Gambar 2 dapat dilihat secara keseluruhan bahwa pendekatan *deep learning* menunjukkan nilai akurasi lebih tinggi dibandingkan dengan pendekatan *machine learning*. Pada pendekatan *machine learning*, nilai akurasi tertinggi dihasilkan oleh algoritma *random forest* dan terburuk dihasilkan oleh algoritma *naive bayes*. Sedangkan pada pendekatan *deep learning*, nilai akurasi tertinggi dihasilkan oleh algoritma GRU dan nilai akurasi terburuk dihasilkan oleh algoritma CNN.



Gambar 2. Grafik Akurasi Rasio 70%-30%

Keterangan Gambar 2 :

- Garis lurus menunjukkan algoritma *machine learning*.
- Garis putus-putus menunjukkan algoritma *deep learning*.

2. Rasio 80%-20%

Tabel 2. Akurasi Rasio 80%-20%

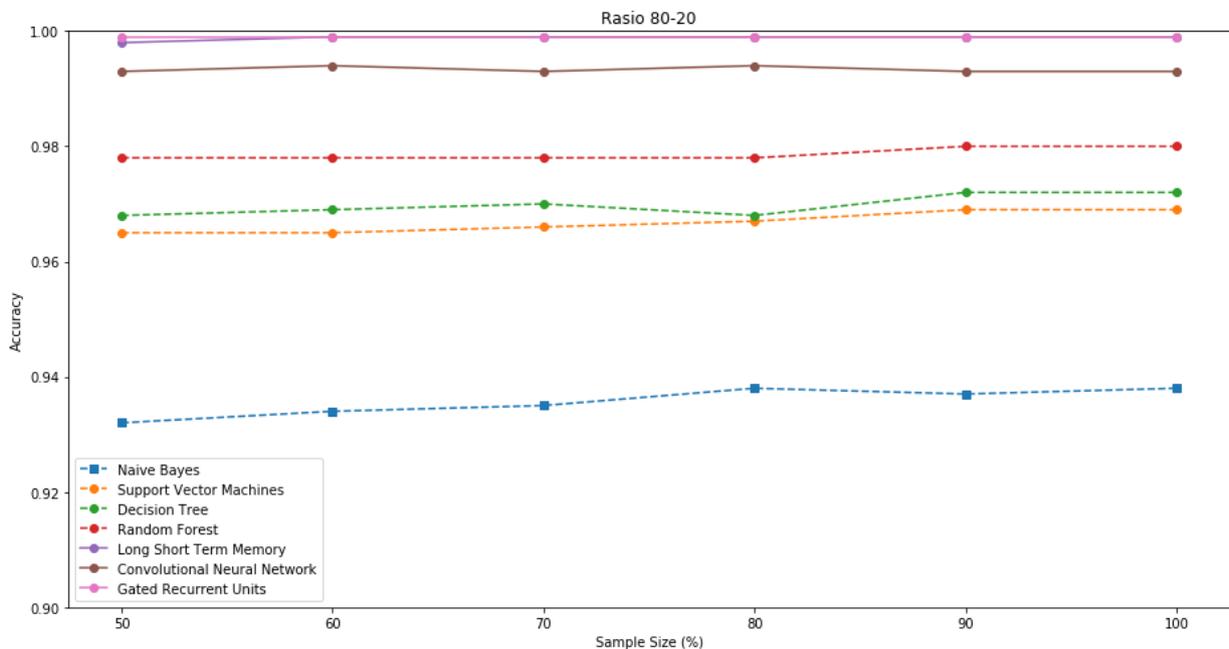
Algoritma	Sample Size (%)						
	50	60	70	80	90	100	
Machine Learning	Naive Bayes	0.932	0.934	0.935	0.938	0.937	0.938
	SVM	0.965	0.965	0.966	0.967	0.969	0.969
	Decision Tree	0.968	0.969	0.970	0.968	0.972	0.972
	Random Forest	0.978	0.978	0.978	0.978	0.980	0.980
Deep Learning	LSTM	0.998	0.999	0.999	0.999	0.999	0.999
	CNN	0.993	0.994	0.993	0.994	0.993	0.993
	GRU	0.999	0.999	0.999	0.999	0.999	0.999

Berdasarkan dari hasil akurasi yang ditampilkan pada Tabel 2 di atas dapat disimpulkan bahwa pada *sample size* 50% akurasi terbaik dihasilkan oleh algoritma GRU sebesar 0.999, pada *sample size* 60% sampai dengan 100% akurasi terbaik dihasilkan oleh algoritma LSTM dan GRU dengan nilai 0.999.

Setiap *sample size* memiliki selisih yang berbeda-beda, seperti pada *sample size* 50%, rentang perbedaan akurasi setiap algoritmanya ada diantara 0.001-0.067, pada *sample size* 60% selisihnya antara 0.005-0.065, pada *sample size* 70% antara 0.006-0.064, pada *sample size* 80% antara 0.005-0.061, pada *sample size* 90% antara 0.005-0.062 dan 100% berada diantara 0.005-0.061.

Selain dilihat berdasarkan *sample size*, juga dapat dilihat berdasarkan algoritma, seperti pada algoritma *naive bayes* mendapatkan akurasi terbaik dengan nilai 0.938 pada *sample size* 100%, dan akurasi terburuk dengan nilai 0.932 pada *sample size* 50%. Algoritma SVM memperoleh akurasi terbaik pada *sample size* 90% dan 100% dengan nilai 0.969. Algoritma *Decision Tree* memperoleh hasil akurasi terbaik sebesar 0.972 pada *sample size* 90% dan 100%. Algoritma *random forest* memperoleh hasil akurasi terbaik pada *sample size* 90% dan 100% dengan nilai 0.98. Algoritma LSTM memperoleh akurasi terbaik dengan nilai 0.999 pada *sample size* 60%, sampai 100%. Algoritma CNN memperoleh akurasi terbaik dengan nilai 0.994 pada *sample size* 60% dan 80%. Algoritma GRU memperoleh akurasi terbaik dengan nilai 0.999 pada semua *sample size*.

Pada Gambar 3 dapat dilihat secara keseluruhan bahwa pendekatan *deep learning* menunjukkan nilai akurasi lebih tinggi dibandingkan dengan pendekatan *machine learning*. Pada pendekatan *machine learning*, nilai akurasi tertinggi dihasilkan oleh algoritma *random forest* dan terburuk dihasilkan oleh algoritma *naive bayes*. Sedangkan pada pendekatan *deep learning*, nilai akurasi tertinggi dihasilkan oleh algoritma GRU dan nilai akurasi terburuk dihasilkan oleh algoritma CNN.



Gambar 3. Grafik Akurasi Rasio 80%-20%

Keterangan Gambar 3 :

- Garis lurus merupakan algoritma *machine learning*.
- Garis putus-putus merupakan algoritma *deep learning*.

3.2 Perbandingan Hasil Pengujian

Hasil komputasi menggunakan algoritma *machine learning* dan *deep learning* dengan rasio 70-30 dan 80-20 pada Tabel 3.1 dan Tabel 3.2 menghasilkan beberapa perbedaan akurasi, diantara :

1. Hasil akurasi menggunakan algoritma *decision tree* pada rasio 80-20 untuk *sample size* 50%, 60%, 70% dan 80% sedikit lebih baik daripada rasio 70-30, selisih antara keduanya adalah sebesar 0.001. Pada *sample size* 90% selisihnya sebesar 0.002 dan *sample size* 100% sebesar 0.005.
2. Hasil akurasi algoritma *random forest* pada rasio 80-20 untuk *sample size* 50% sedikit lebih baik dari rasio 70-30, dengan perbedaan sebesar 0.002. Sedangkan pada *sample size* 60%, 70%, 90% dan 100% selisihnya sebesar 0.001.
3. Pada algoritma *naive bayes* terlihat lebih baik pada rasio 80-20 untuk *sample size* 100% dengan selisih sebesar 0.001.
4. Pada algoritma *long short term memory*, akurasi yang lebih baik dihasilkan oleh rasio 80-20 untuk *sample size* 70% dan 80% dengan selisih sebesar 0.001.
5. Pada algoritma *convolutional neural network*, akurasi yang lebih baik dihasilkan oleh rasio 70-30 pada *sample size* 70% dengan selisih sebesar 0.001.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat diambil beberapa kesimpulan yaitu guna mendapatkan model terbaik, dilakukan proses *training* menggunakan algoritma *machine learning* dan *deep learning*. Algoritma *machine learning* yang digunakan pada penelitian adalah *naive bayes*, *support vector machines*, *decision tree*, dan *random forest*. Sedangkan algoritma *deep learning* yang digunakan adalah *long short term memory*, *convolutional neural network*, dan *gated recurrent units*. Berdasarkan eksperimen, metode *deep learning* menghasilkan akurasi yang lebih baik daripada metode *machine learning* dengan akurasi terbaik dihasilkan dari algoritma *deep learning* yaitu *gated recurrent units* dan *long short term memory* sebesar 0.999. Adapun hasil akurasi terbaik pada metode *machine learning* dihasilkan oleh algoritma *random forest* dengan akurasi 0.98. Pada proses *training*, diketahui

bahwa semakin besar ukuran *sample size*, maka akurasi semakin tinggi. Namun selisih akurasi antara *sample size* terkecil hingga terbesar tidak terlalu jauh dan pada algoritma *deep learning* menunjukkan bahwa ukuran *sample size* tidak terlalu mempengaruhi akurasi. Akurasi yang dihasilkan oleh algoritma *deep learning* cenderung stabil, baik pada *sample size* terkecil maupun *sample size* terbesar.

DAFTAR PUSTAKA

- [1] A. S. Cahyono, "Pengaruh Media Sosial Terhadap Perubahan Sosial Masyarakat di Indonesia," J. Ilmu Sos. dan Ilmu Polit., pp. 140–157, 2016.
- [2] J. An, M. Cha, K. Gummadi, and J. Crowcroft, "Media landscape in Twitter: A world of new conventions and political diversity," Assoc. Adv. Artificial Intell., pp. 18–25, 2011.
- [3] H. Februariyanti, E. Zuliarso, D. Bulan, and L. Suryati, Rancang bangun sistem layanan informasi bencana melalui twitter menggunakan basis data xml, vol. 035, no. November. 2013.
- [4] Z. Ashktorab, C. Brown, M. Nandi, and A. Culotta, "Tweed: Mining Twitter to Inform Disaster Response," Proc. 11th Int. ISCRAM Conf., vol. 12, no. 4, pp. 354–358, 2014.
- [5] A. F. Hidayatullah, "Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter," TESIS, vol. 2014, no. semnasIF, p. A-1, 2014.
- [6] L. Derczynski et al., "Analysis of named entity recognition and linking for tweets," Inf. Process. Manag., vol. 51, no. 2, pp. 32–49, 2015.
- [7] Y. Wibisono and M. L. Khodra, "Pengenalan Entitas Bernama Otomatis untuk Bahasa Indonesia dengan Pendekatan Pembelajaran Mesin," pp. 1–5, 2018.

RETRACTED