

Deteksi Berita Palsu Menggunakan CNN-RNN dalam Penentuan Penyebaran Berita

Khoirinisa Umi Rahmawati¹

¹Jurusan Ilmu Komputer, FMIPA, Universitas Negeri Semarang
Email: ¹khairinisaumi61@students.unnes.ac.id

Abstrak

Berkembangnya zaman maka berkembang juga teknologi yang ada, didalam teknologi terdapat adanya komunikasi yang melibatkan beberapa orang melalui media sosial online. Media sosial saat ini tidak hanya dapat digunakan untuk berkomunikasi, tetapi juga untuk menyebarkan berita, dan sebagainya, dalam media sosial ini tentu saja tidak semua orang berniat baik, ada juga yang berniat jahat dengan cara menyebarkan berita palsu. Beberapa orang yang membaca suatu berita tidak memeriksa silang referensi dari sumber berita yang ada dan memastikan kredibilitas berita. Oleh karena itu, dengan mempertimbangkan skala pengguna dan kontributor media sosial, diperlukan adanya deteksi berita palsu ini yang memungkinkan untuk mengambil tindakan perbaikan, serta dapat memfilter mana berita benar dan berita palsu. Dalam beberapa penelitian sebelumnya belum terdapat penggolongan berdasarkan sumber dan tahun berita, topik dari data yang ada, cara pencegahan penyebaran berita palsu dan mengkonfirmasi keaslian akun berita. Dalam penelitian ini bertujuan untuk menggolongkan berita menurut sumber dan tahun berita, klasifikasi berdasarkan topik, pencegahan, dan deteksi akun dari berita palsu. Dalam penelitian ini akan menggunakan metode CNN-RNN yang dimana metode tersebut diharapkan dapat menyempurnakan penelitian sebelumnya. Hasil penelitian ini yaitu menghasilkan akurasi menggunakan CNN-RNN dalam penentuan penyebaran berita sebesar 50.1244 % dan 52.9851 %

Kata Kunci: Data Mining, CNN-RNN, Berita Palsu, Media Sosial

Abstract

With the development of the times, the existing technology also develops, in technology there is communication that involves several people through online social media. Today's social media can not only be used to communicate, but also to spread news, and so on, in this social media of course not everyone has good intentions, there are also those who have bad intentions by spreading fake news. Some people who read a story do not cross-check the references from existing news sources and ensure the credibility of the news. Therefore, taking into account the scale of social media users and contributors, it is necessary to have this fake news detection that allows for corrective action, and can filter out which news is true and fake news. In some previous studies, there has been no classification based on the source and year of news, topics from existing data, ways to prevent the spread of fake news and confirm the authenticity of news accounts. This study aims to classify news by source and year of news, classification by topic, prevention, and account detection of fake news. In this study, the CNN-RNN method will be used which is expected to be able to complete previous research. These aspects such as source, year, topic, and prevention of fake news contribute to improving which methods are most suitable for detecting fake news on social media.

Keyword: Data Mining, CNN-RNN, Fake News, Social Media

1. PENDAHULUAN

Berkembangnya zaman maka berkembang pula teknologi yang ada, didalam teknologi tersebut terdapat adanya komunikasi yang melibatkan beberapa orang melalui media sosial online, media sosial merupakan media yang digunakan oleh konsumen atau pengguna untuk menyebarkan teks, gambar, suara, dan video informasi baik dengan orang lain maupun perusahaan serta *vice versa*[1]. Media sosial saat ini tidak hanya dapat digunakan untuk berkomunikasi, namun juga untuk berbagi berita, liputan dan sebagainya, didalam media sosial ini tentu saja tidak semua orang berniat baik, terdapat pula yang berniat jahat dengan cara menyebarkan berita palsu.

Berita palsu adalah suatu jenis jurnaslisme kuning atau propaganda yang terdiri dari adanya misinformasi (salah menafsirkan suatu informasi) yang dilakukan secara disengaja atau penyebaran berita palsu (*hoax deliberate*) melalui media berita cetak, siaran radio serta televisi, maupun media sosial online[2]. Ciri-ciri berita palsu yaitu terdapat ajakan untuk menyebarkan supaya berita tersebut viral dan menimbulkan perkumpulan orang-orang yang merasa memiliki pemahaman yang sama, sumber informasi atau mediana tidak jelas identitasnya, tidak mengandung unsur berita 5W+1H lengkap.[3]

Dengan berkembangnya berita online portal, situs jejaring sosial, dan media sosial online lainnya, berita palsu ini menjadi perhatian utama saat ini. Namun beberapa orang atau sebagian besar yang membaca suatu berita tidak mengusut silang referensi dari sumber berita yang ada dan memastikan kredibilitas berita tersebut. Pesatnya penyebaran berita palsu disebabkan oleh meluasnya penggunaan media sosial yang memberikan lahan subur untuk menyebarkan secara instan dan beredar berita dengan pengguna tidak mempunyai sarana kualitas untuk mengusut konten yang dibagikan. Oleh sebab itu, dengan mempertimbangkan skala pengguna dan kontributor media sosial, dibutuhkan adanya deteksi otomatis berita palsu ini yang memungkinkan satu-satunya cara untuk mengambil tindakan perbaikan, serta dapat memfilter mana berita benar dan mana berita palsu.

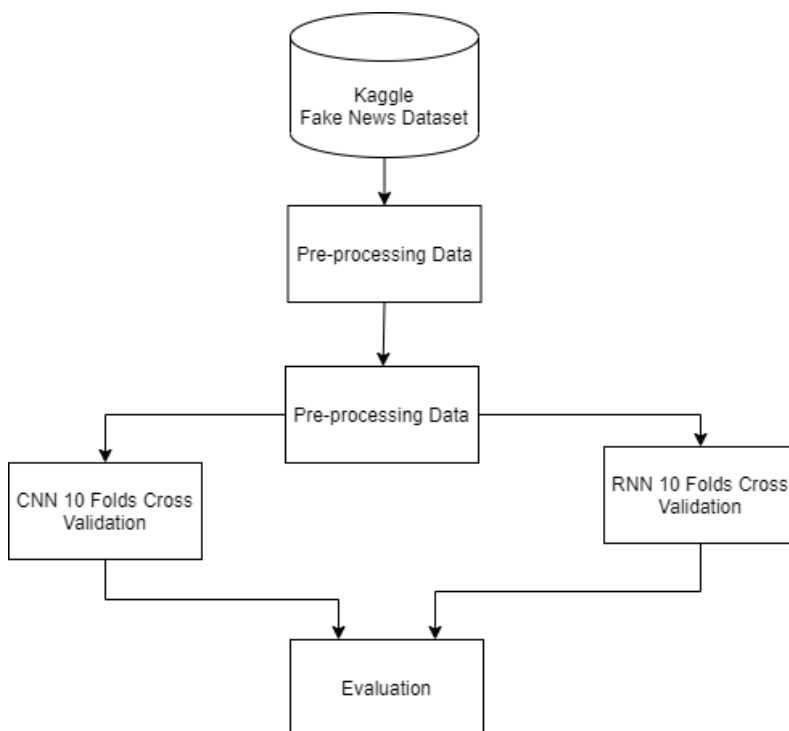
Data mining merupakan salah satu teknik yang digunakan untuk mengolah data yang telah diambil. *Convolutional Neural Network* (CNN) dan *Recurrent Neural Network* (RNN) adalah salah satu bentuk teknologi *deep learning* yang melibatkan *feature extraction layer* dan *full connected layer* yang akan digunakan untuk mendeteksi dan mengenali object pada data. Metode CNN memiliki cara kerja yang sama dengan *Multi-Layer Perceptron* (MLP), namun dalam CNN setiap *neuron* dipresentasikan dalam bentuk dua dimensi, tidak seperti MLP yang setiap *neuron* hanya berukuran satu dimensi. Sedangkan dalam metode RNN melibatkan pemrosesan sekuensial berasal data learning. Proses ini berurutan dibenarkan oleh kemampuannya untuk mempertahankan ingatan tentang apa yang terjadi sebelum urutan saat ini diproses, disebut berulang (*recurrent*) karena hasil pada setiap langkah waktu digunakan pada langkah waktu sebelumnya. Hal ini pada gilirannya memungkinkan kita untuk mengenal istilah dalam data latihan (*data training/preprocessing*).[4] Dalam kasus berita palsu, banyak artikel berita dapat dipertimbangkan untuk dipelajari secara relative satu sama lain, alih-alih mempelajari setiap artikel berita secara terpisah. RNN terdiri dari lapisan dengan sel memori. Terdapat berbagai jenis sel memori yang

digunakan pada RNN. Salah satu jenis tersebut merupakan *Long Short Term Memory* (LSTM) unit atau sel. RNN bertanggung jawab untuk memastikan bahwa tidak terdapat kehilangan informasi selama proses sekuensial.

Pada penelitian ini, dataset yang digunakan yaitu *Fake News Dataset* yang diperoleh dari *website* kaggle yang terdiri dari 804 artikel dan 7 atribut, sedangkan tools yang digunakan untuk menerapkan model yaitu menggunakan aplikasi Weka 3.8.5 dengan tujuan dari penelitian ini adalah untuk mengetahui tingkat akurasi metode CNN-RNN untuk mendeteksi dalam penentuan penyebaran berita menurut sumber berita ditulis.

2. METODE

Metode penelitian yang akan digunakan dalam penelitian ini yaitu menggunakan algoritma CNN-RNN menggunakan pilihan tes *Cross-validation* dengan *10 Folds*. Adapula fungsi yang digunakan yaitu *wekadeeplearning4j* yang terlebih dahulu melakukan instalasi package tambahan pada aplikasi Weka 3.8.5. Hasil akhir yang didapat yaitu diketahuinya tingkat akurasi deteksi dalam penentuan penyebaran berita menurut sumber berita ditulis menggunakan algoritma CNN-RNN. Gambar 1. merupakan gambaran kerangka penelitian yang akan dilakukan.



Gambar 1. Kerangka Penelitian

Berdasarkan kerangka penelitian yang telah disajikan dalam Gambar 1. maka metode penelitian tersebut dapat dijabarkan menjadi seperti berikut: a) Dataset yang digunakan dalam penelitian ini yaitu *Fake News Dataset* yang berasal dari *website* kaggle yang

memiliki 804 artikel dan 7 atribut. b) *Pre-processing data* merupakan teknik pengolahan data awal yang digunakan dalam penelitian ini dengan cara melakukan perubahan *StringToNominal* yang dilakukan pada tahap *Preprocess* di aplikasi Weka yang berguna untuk merubah tipe data string ke nominal. c) Proses CNN-RNN 10 *Folds Cross Validation* ini terbagi menjadi dua proses yang digunakan yaitu CNN dan RNN, kedua fungsi tersebut merupakan fungsi yang menjalankan algoritma CNN-RNN, adapun perbedaannya yaitu pada CNN dapat digunakan untuk mengenali aspek yang informatif pada data regional tertentu (*local*) dan RNN dapat mempresentasikan data sekuensial (*sequential*) dengan bantuan fitur LSTM. Pada proses ini validasi akan dilakukan dengan cara membagi dataset menjadi 10 bagian dengan rincian satu bagian sebagai data uji dan bagian lain sebagai data latih. Proses ini akan dilakukan secara berulang dimulai dari data uji hingga data latih yang terakhir, hal ini menjadikan semua data dalam dataset telah teruji. Proses ini bertujuan agar menghasilkan akurasi dalam mendeteksi penentuan penyebaran berita menurut sumber berita yang ditulis. d) Evaluation berfungsi untuk melihat kinerja penentuan penyebaran berita menurut sumber berita yang ditulis yang terkalkulasi dalam bentuk presentase akurasi serta nilai *Area Under ROC Curve (AUC)* yang digunakan untuk memberikan matriks numerik single sehingga dapat digunakan untuk membandingkan hasil kinerja. Nilai yang dihasilkan oleh AUC dapat dijadikan ukuran untuk melihat prediksi yang berbentuk angka dengan rentang 0 sampai dengan 1, dengan perkiraan angka yang mendekati satu menjadikan nilai dengan tingkat akurasi yang lebih baik karena meningkatkan efektivitas klasifikasi.

Aplikasi yang digunakan untuk menerapkan penelitian ini yaitu aplikasi Weka versi 3.8.5. tinjauan Pustaka yang digunakan dalam algoritma CNN-RNN yaitu Model yang digunakan yaitu fitur berbasis konten dan algoritma Machine Learning (ML). Dari penelitian ini menunjukkan bahwa fitur yang diusulkan dikombinasikan dengan algoritma ML diperoleh akurasi hingga 95%[5]; Penelitian ini menghasilkan model berbasis BERT telah mencapai kinerja yang lebih baik daripada semua model lain di semua kumpulan data dan dapat berkinerja lebih baik secara signifikan pada ukuran sampel yang sangat kecil dengan informasi yang memadai yang disediakan dalam sebuah berita artikel, model berbasis LSTM memiliki kemungkinan lebih tinggi untuk mengatasi *overfitting*[6]; Penelitian ini menghasilkan bentuk arsitektur CNN dinamis dengan menggunakan lapisan pooling WalkPool untuk mengoptimalkan komputasi CNN dan meningkatkan akurasi yang lebih tinggi dan kerugian yang lebih kecil daripada pengoptimal dinamis, termasuk Adam, Adagrad, dan RMSProp [7]. Metode yang digunakan yaitu CNN-RNN berbasis deep learning. Penelitian ini menghasilkan cenderung berfungsi dengan baik pada dataset tertentu, tetapi tidak menggeneralisasi dengan baik. Cakrawala baru bisa dieksplorasi dengan mempertimbangkan generalisasi model deteksi berita palsu [8].

Adapun hasil pengujian dihitung menggunakan *Confusion Matrix* untuk mencari akurasi dan AUC. *Confusion matrix* adalah metode yang digunakan untuk mengukur kinerja model klasifikasi berdasarkan perhitungan objek, dengan data hasil prediksi diantara dua kelas yang menghasilkan kelas positif dan negatif [9]. *Confusion matrix* yang disajikan berisi informasi aktual dan prediksi klasifikasi [10]. Proses evaluasi *confusion matrix* dapat memperoleh nilai *precision*, *recall*, dan *accuracy* yang didapat dari rumus berikut:

$$Precision = \frac{TP}{(TP+FP_)} \quad (1)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (2)$$

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (3)$$

$$F-Measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (4)$$

Keterangan:

TP : Jumlah berita benar yang diklasifikasikan sebagai benar (True = 1).

FP : Jumlah berita palsu yang diklasifikasikan sebagai benar.

TN ; Jumlah berita palsu yang diklasifikasikan sebagai palsu (False = 0).

FN : Jumlah berita benar yang diklasifikasikan sebagai palsu (False = 0).

Adapun Tabel 1. menjelaskan contoh tabel *confusion matrix* yang menunjukkan klasifikasi dua kelas.

Tabel 1. Model *confusion matrix*.

Classification	Predicted Class	
	Class = TRUE (1)	Class = FALSE(0)
Class = TRUE (1)	True Positives (TP)	False Negatives (FN)
Class = FALSE (0)	False Positive (FP)	True Negatives (TN)

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini menggunakan dataset yang diambil dari *website* Kaggle dengan judul *Fake News Dataset* dengan 804 artikel dan 7 atribut yang disajikan seperti terlampir di Tabel 2.

Tabel 2. Dataset

unit_id	article_title	article_cont				
		ent	source	date	location	labels
		Wed 05				
	Syria	Apr 2017				
	attack	Syria attack				
	symptoms	symptoms				
	consistent	consistent				
	with nerve	with nerve				
	agent use	agent use		2017-		
1914947530	WHO	WHO...	nna	05-04	idlib	0

1914947532	Homs governor says U.S. attack caused deaths but doesnt see big human losses	Fri 07 Apr 2017 at 0914 Homs governor says U.S. attack caused deaths but doesnt see big human losses...	nna	2017- 07-04	homs	0
1914947533	Death toll from Aleppo bomb attack at least 112	Sun 16 Apr 2017 Death toll from Aleppo bomb attack at least 112...	nna	2017- 04-16	aleppo	0
1914947534	Aleppo bomb blast kills six Syrian state TV	Wed 19 Apr 2017 Aleppo bomb blast kills six Syrian state TV...	nna	2017- 04-19	aleppo	0
1914947535	29 Syria Rebels Dead in Fighting for Key Aleppo Road	Sun 10 Jul 2016 29 Syria Rebels Dead in Fighting for Key Aleppo Road...	nna	2016- 10-07	aleppo	0
1914947536	Suicide bombing kills at least 16 in northeast Syria	Tue 05 Jul 2016 Suicide bombing kills at least 16 in northeast Syria...	nna	2016- 05-07	hasakeh	0

1914947537	Syria raids on IS stronghold	Sun 05 Jul 2015 22 dead in heavy U.S. raids on IS Syria stronghold ...	nna	2015- 05-07	raqqa	0
1914947538	Assad clans hometown	Sun 22 Feb 2015 Suicide bomber kills 4 in Assad clans hometown ...	nna	2015- 02-22	lattakia	0
1914947539	Damascus	Sun 01 Feb 2015 Explosion rocks down town Damascus ...	nna	2015- 01-02	damascus	1
1914947540	Damascus explosion due to rocket bomb	Sat 24 Aug 2013 Damascus explosion due to rocket bomb ...	nna	2013- 08-24	damascus	0
...
1965511231	Syrian Army Kills 48 ISIL Terrorists in Deir Ezzor	April 6 2017 Syrian Army Kills 48 ISIL Terrorists in Deir Ezzor...	mannar	2017- 04-04	deir ezzor	1

Adapula deskripsi dari *attribut dataset* tersebut akan dijelaskan seperti yang tertera di **Tabel 3.**

Tabel 3. *Attribute Dataset*

Attribute	Type	Description
Unit_id	Numeric	Id number start in 1914947530 - 1965511231
Article_title	String	Title of the article news
Article_content	String	Content of article description about news
source	string	Source of the article written : nna, alaraby, asharqalawsat, dailysabah, trt, ahram, jordantimes, tass, sana, etilaf, manar, arabiya, reuters, alalam, sputnik
date	date	Date of article realease in format "yyyy-mm-dd"
location	string	Location of article made : idlib, homs, aleppo, hasakeh, raqqa, lattakia, damascus, hama, deir ezzor, quneitra, daraa, tartous
labels	Binary(0,1)	Whether the article is true or false with an example of 0 and 1

Tahap awal penggunaan metode ini yaitu dimulai dari tahap *pre-processing data* yang ada dengan menggunakan aplikasi Weka. *Pre-processing data* yang dilakukan yaitu dengan menggunakan fungsi *StringToNominal* yang berfungsi untuk merubah tipe data string menjadi nominal. Setelah itu maka akan dilakukan proses klasifikasi menggunakan algoritma CNN-RNN dengan pilihan test 10 *Folds Cross Validation* dimana dataset akan dibagi menjadi 10 bagian dengan rincian satu bagian sebagai data uji dan bagian lain sebagai data latih. Proses ini akan dilakukan secara berulang dimulai dari data uji hingga data latih yang terakhir, hal ini menjadikan semua data dalam dataset telah teruji. Setelah semua proses selesai maka akan didapatkan nilai akurasi bersama dengan waktu prosesnya yang akan dijabarkan pada Tabel 4.

Tabel 4. Hasil akurasi

Algotima	Test Options	Akurasi		Waktu	
		CNN	RNN	CNN	RNN
CNN-RNN	10 Folds				
	Cross Validation	50.1244 %	52.9851 %	29.41 s	0.18 s

Adapun beberapa perbedaan hasil perhitungan menggunakan dua fungsi yang telah diuji akan dipaparkan pada **Tabel 5**.

Tabel 5. Perbedaan dua fungsi

	CNN	RNN
Correctly Classified Instances	403	426
Incorrectly Classified Instances	401	378
Kappa statistic	0.005	0
Mean absolute error	0.5004	0.4977
Root mean squared error	0.7022	0.4994
Relative absolute error	100.4379%	99.9038 %
Root relative squared error	140.6857 %	100.0658 %

Correctly classified instances menunjukkan jumlah data yang benar sedangkan *incorrectly classified instances* menunjukkan jumlah data yang diklasifikasikan salah. *Kappa statistic* menunjukkan ukuran yang menyatakan konsistensi pengukuran dan ada rata-rata selisih mutlak dengan nilai prediksi serta hasil dari pengakarannya dan nilai error absolut relatif dan hasil pengakarannya. Dari hasil tersebut juga didapatkan *confusion matrix* yang berfungsi untuk mencari nilai akurasi AUC yang dijelaskan dalam Tabel 6.

Tabel 6. Hasil *confusion matrix*

CNN	RNN
a b <-- classified as	a b <-- classified as
198 180 a = 0	0 378 a = 0
221 205 b = 1	0 426 b = 1

Berdasarkan tabel diatas maka fungsi CNN menunjukkan ada 198 data yang dikelompokkan benar dan 180 data salah untuk *class 0* dan 221 data benar dan 205 data salah untuk *class 1*, sedangkan untuk fungsi RNN menunjukkan hasil 0 data benar dan 378 data salah untuk *class 0* dan 0 data benar dan 426 data salah untuk *class 1*. Adapun perhitungan nilai AUC yang telah didapatkan dari CNN-RNN akan dijelaskan pada Tabel 7. dan Tabel 8.

Tabel 7. Nilai AUC CNN

	TP Rate	FP Rate	Precision	Recall	F-Measure	MC C	ROC Area	PRC Area	Class
	0.524	0.519	0.473	0.524	0.497	0.005	0.495	0.465	0
	0.481	0.476	0.532	0.481	0.506	0.005	0.494	0.523	1
Weighted Avg.	0.501	0.496	0.504	0.501	0.501	0.005	0.495	0.495	

Tabel 8. Nilai AUC RNN

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,000	0,000	?	0,000	?	?	0,479	0,454	0
	1,000	1,000	0,530	1,000	0,693	?	0,479	0,523	1
Weighted Avg.	0.530	0.530	?	0,530	?	?	0,479	0,491	

Dengan TP menunjukkan jumlah kasus positif yang diklasifikasikan sebagai positif, FP menunjukkan jumlah kasus negatif yang diklasifikasikan sebagai positif dilengkapi dengan proses evaluasi *confusion matrix* masing-masing kelas serta berat rata-rata atribut yang dihasilkan.

4. SIMPULAN

Berdasarkan penelitian yang telah dilakukan, teknik data mining berupa klasifikasi menggunakan metode CNN-RNN dengan dataset yang digunakan yaitu *Fake News Dataset* yang diperoleh dari *website* Kaggle dengan judul *Fake News Dataset* dengan 804 artikel dan 7 atribut yang diimplementasikan mampu mendeteksi penentuan penyebaran berita menurut sumber berita ditulis dengan akurasi sebesar 50.1244 % dan 52.9851 %. Dua fungsi yang digunakan yaitu CNN dan RNN tidak memiliki perbedaan terlalu jauh terhadap data yang digunakan. Dengan nilai akurasi tersebut maka disimpulkan bahwa algoritma CNN-RNN memiliki tingkat akurasi yang cukup tinggi dengan diaplikasikan pada dataset yang telah didapatkan. Algoritma ini dapat memprediksi penentuan penyebaran berita dengan akurasi yang cukup tinggi sehingga

dapat dijadikan acuan bagi khalayak umum di media sosial yang sedang membutuhkan keakuratan berita untuk mengetahui berita mana yang benar dan berita mana yang palsu dan melakukan deteksi berita berdasarkan sumber yang beredar.

5. REFERENSI

- [1] Kotler, Philip and Kevin Lane Keller, 2016. *Marketing Managemen*, 15th Edition, Pearson Education, Inc.
- [2] Leonhardt, D., & Thompson, S. A. (2017). *Trump's Lies*.
- [3] Kementerian Komunikasi dan Informatika. (2021). Retrieved July 11, 2021, from <https://www.kominfo.go.id/content/detail/15945/permainan-politik-era-digital-itu-bernama-hoaks/0/artikel>
- [4] Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*, 1(1), 100007.
- [5] Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications*, 4(March), 100032.
- [6] Gravanis, G., Vakali, A., Diamantaras, K., & Karadais, P. (2019). Behind the cues: A benchmarking study for fake news detection. *Expert Systems with Applications*, 128, 201–213.
- [7] Wanda, P., & Jie, H. J. (2020). DeepProfile: Finding fake profile in online social network using dynamic CNN. *Journal of Information Security and Applications*, 52, 102465.
- [8] Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*, 1(1), 100007.
- [9] Kurniawan I. 2020. *Prediksi Gejala Autism Spectrum Disorders pada Remaja Menggunakan Optimasi Particle Swarm Optimization dan Algoritma Support Vector Machine. Informatics for Educators and Professionals*. 113-122.
- [10] M. Y. Kurniawan dan M. E. Rosadi. 2017. *Optimasi Decision Tree Menggunakan Particle Swarm Optimization Pada Data Siswa Putus Sekolah. Jtiulm Vol. 2 No. 1*.