

HASIL_Sheraton Pawestri

by Universitas Ahmad Dahlan Yogyakarta 22

Submission date: 17-Nov-2023 10:40AM (UTC+0700)

Submission ID: 2186319553

File name: Turnitin_-_Proposal_Sheraton_Pawestri.docx (189.98K)

Word count: 2621

Character count: 16480

Analisis Perbandingan Metode Recurrent Neural Networks dan Convolutional Neural Networks pada Deteksi Sarkasme Judul Berita

RINGKASAN

Era globalisasi dan teknologi yang semakin mumpuni, terciptanya sosial media yang bisa menjadi wadah bagi masyarakat untuk mampu mengutarakan sesuatu dan bisa dilihat oleh banyak orang di seluruh penjuru dunia. Tidak hanya memberikan informasi sehari-hari, pada sosial media juga bisa mengutarakan pendapat-pendapat dan saling balas mengenai kasus tertentu. Kadang pendapat tidak dilontarkan secara lugas, namun dengan kiasan-kiasan tertentu seperti sarkasme sehingga sulit untuk dipahami makna sesungguhnya. Oleh karena itu, penelitian ini membuat deteksi sarkasme yang bertujuan untuk membantu dalam menganalisis apakah kalimat tersebut sebuah sarkasme atau bukan.

Banyak metode yang digunakan dalam deteksi sarkasme, salah satunya adalah metode-metode sederhana seperti Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), dan lain-lain. Penelitian terdahulu menyarankan untuk menggunakan berbagai macam metode yang lebih modern seperti deep learning [10]. Neural Network telah menjadi trend pada dunia penelitian termasuk pada Natural Language Processing (NLP) dan dapat diulas lebih lanjut. Oleh karena itu, penelitian ini bertujuan untuk menganalisis performa RNN dan CNN dalam deteksi sarkasme sehingga dapat dengan optimum membantu manusia mengetahui apakah kalimat tersebut sarkasme atau bukan secara otomatis.

Metode ini dipilih karena sampai lima tahun terakhir CNN dan RNN masih banyak dipakai terutama dengan berbagai macam kombinasi antara CNN dengan metode lainnya. Beberapa contoh yang menggunakan CNN adalah [11], [12], [13] sedangkan penelitian dari Hazarika, D., dkk. [11] menyarankan untuk menggunakan RNN dalam deteksi sarkasme. Untuk transformasi data, metode yang akan dilakukan adalah RoBERTa. RoBERTa juga metode yang populer yang digunakan dalam penelitian [14] dan [15], dan memiliki tingkat akurasi yang lebih tinggi dibandingkan metode-metode dasar. Dataset yang digunakan dalam penelitian ini adalah dataset yang berisi judul berita yang telah dilabeli ke dalam sarkasme atau tidak. Judul berita biasanya menggunakan bahasa yang formal. Oleh sebab itu, hal ini menarik untuk dikaji seperti bagaimana sebuah model bisa menebak apakah judul berita ini sarkasme atau tidak, padahal sarkasme biasanya ada di kalimat-kalimat tidak formal. Dataset ini diambil dari TheOnion dan HuffPost yang terdiri dari 26,709 judul berita. Dataset ini dibuat oleh Misra, R. and Arora, P. [12] dan sudah memiliki lisensi sendiri.

Tahapan penelitian yang akan dilakukan adalah studi literatur terlebih dahulu mengenai metode CNN dan RNN serta pengembangannya, mengenai RoBERTa, dan hal-hal yang menunjang penelitian. Step selanjutnya adalah pembuatan model. Model yang dibuat adalah dua model. Pertama RoBERTa dengan CNN lalu kemudian RoBERTa dengan RNN yang kemudian dianalisis serta dibandingkan performa modelnya. Kemudian implementasi model. Implementasi akan menggunakan python. Tahapan awal implementasi adalah mengambil data dan lakukan prapemrosesan pada data. Data kemudian akan diubah bentuknya menggunakan RoBERTa agar mampu diolah oleh mesin. Setelah prapemrosesan akan dilakukan

pembentukan model, yang kemudian data tersebut dimasukkan dalam model tersebut. Terakhir adalah evaluasi dan analisis model. Hasil analisis akan berbentuk paper. Output yang didapat dari penelitian ini adalah sebuah model yang dapat mendeteksi secara otomatis apakah sebuah judul berita tersebut termasuk sarkasme atau bukan dan sebuah paper yang membahas analisis perbandingan metode CNN dan RNN yang mendukung penelitian ini.

Kata kunci: Recurrent Neural Network; Convolutional Neural Network; Deteksi Sarkasme; Judul Berita

LATAR BELAKANG

Era globalisasi dan teknologi yang semakin mumpuni, terciptanya sosial media dan internet yang bisa menjadi wadah bagi masyarakat untuk mampu mengutarakan sesuatu dan bisa dilihat oleh banyak orang di seluruh penjuru dunia. Di sosial media dan internet, kita bisa memberikan informasi apapun yang kita inginkan karena kita bisa menuliskan apapun yang kita inginkan dan mengutarakan pendapat-pendapat mengenai kasus tertentu. Misal pada berita yang sedang trend, kita bisa memberikan komentar kita mengenai kasus tersebut. Kadang komentar atau pendapat dari pengguna tidak dilontarkan secara lugas, namun dengan kiasan-kiasan tertentu seperti sarkasme sehingga lebih sulit untuk dipahami makna sesungguhnya daripada komentar secara lugas. Selain pada media sosial, sarkasme bisa ditemukan pada judul berita. Hal ini dapat membuat kebingungan dan menyebabkan misinterpretasi yang bisa menyebabkan kesalahpahaman [1].

Sarkasme sendiri adalah gaya bahasa atau cara mengungkapkan pikiran melalui bahasa yang biasanya menunjukkan semangat dan kepribadian penulis. Salah satu gaya bahasa yang biasa dipakai adalah Ironi dan sarkasme[2]. Menurut Kreuz dan Glucksberg pada tahun 1989 sarkasme adalah ironi verbal yang mengungkapkan sikap negatif dan kritis terhadap orang atau peristiwa [3]. Sarkasme merupakan bahasa kiasan yang mempunyai makna yang tergolong kasar dan menyakitkan. Menurut McDonald's dalam buku Dauphin halaman 486, sarkasme adalah suatu bentuk ekspresi tidak langsung yang sengaja digunakan untuk menghasilkan efek dramatis tertentu pada pendengarnya [4]. Sarkasme diartikan jenis gaya bahasa yang mengandung celaan dan hinaan sehingga kurang enak didengar oleh orang yang membacanya [5]. Sarkasme sendiri sebenarnya cukup mudah dianalisis jika dilihat secara langsung dengan cara mengamati ekspresi dan gerak tubuh seseorang, namun sarkasme akan sangat sulit dikenali jika dalam berbentuk text karena dalam teks seseorang akan sulit menebak ekspresi serta gerak tubuh orang lain [6].

Banyak penelitian terdahulu mengenai deteksi sarkasme, terutama dalam lima tahun terakhir. Beberapa metode telah dilakukan, namun performa masih bisa ditingkatkan serta banyak metode yang masih dapat ditelusuri seperti Neural Network. Salah satu metode yang digunakan untuk metode-metode sederhana adalah Naïve Bayes (NB), Support Vector Machine (SVM) [7], Random Forest (RF) [8], atau bahkan ketiganya [9], dan lain-lain. Penelitian terdahulu menyarankan untuk menggunakan berbagai macam metode yang lebih modern seperti Deep Learning (ANN, CNN, dan RNN) [10]. CNN dan RNN masih banyak dipakai

terutama dengan berbagai macam kombinasi antara CNN dengan metode lainnya. Beberapa contoh yang menggunakan CNN dan pengembangannya adalah penelitian dari Hazarika, D., dkk. [11], Misra, R. and Arora, P. [12], dan penelitian dari Wijaya, A.C., dan Wibawa, I.G.A. [13]. Penelitian dari Hazarika, D., dkk. menyarankan untuk menggunakan RNN dalam deteksi sarkasme. Untuk transformasi data, metode yang akan dilakukan adalah RoBERTa. RoBERTa juga metode yang populer yang digunakan dalam penelitian [14] dan [15] memiliki tingkat akurasi yang lebih tinggi dibandingkan metode-metode dasar. Dataset yang digunakan dalam penelitian ini adalah dataset yang berisi judul berita yang telah dilabeli ke dalam sarkasme atau tidak. Judul berita biasanya menggunakan bahasa yang formal sehingga lebih sulit lagi untuk dikenali. Oleh karena itu, penelitian ini membuat deteksi sarkasme yang bertujuan untuk membantu dalam menganalisis apakah kalimat judul berita tersebut sebuah sarkasme atau bukan menggunakan membandingkan metode CNN dan RNN dengan prapemrosesan RoBERTa.

TINJAUAN PUSTAKA

Tinjauan pustaka dari penelitian ini diambil lima tahun sebelum penelitian ini dan dalam lima tahun terakhir, acuan dari penelitian ini antara lain penelitian di tahun 2018 yang menggunakan modifikasi dari CNN dan menghasilkan CASCADE untuk deteksi sarkasme[11]. Penelitian ini menggunakan berbagai macam CNN untuk dibandingkan dengan metode miliknya yaitu CASCADE dengan kerangka utama menggunakan CNN. Dataset yang digunakan adalah dataset publik SARC. Hasil menunjukkan akurasi tertinggi pada penelitian ini adalah metode CASCADE dengan akurasi 0.74 dan F1-score 0.75. Penelitian ini juga mengemukakan jika penelitian selanjutnya bisa menggunakan RNN untuk performa yang lebih baik. Penelitian mengenai deteksi sarkasme selanjutnya pada tahun 2020 menggunakan metode RoBERTa yang dipadukan dengan deep neural network[14]. Sebagai pembanding, terdapat beberapa metode yang digunakan yaitu ELMO, USE, dan BERT. Dataset yang digunakan adalah dataset twitter dengan cara crawling. Hasil menunjukkan performa tertinggi didapat dari RoBERTa dengan akurasi Precision sebesar 0.77, Recall sebesar 0.78, dan F1-Score sebesar 0.77. Penelitian ini mengatakan jika untuk penelitian kedepannya diharapkan melakukan tuning parameter untuk hasil yang lebih baik. Penelitian lainnya pada tahun 2021 yaitu tentang studi komparasi dari beberapa metode. Metode yang digunakan dalam komparasi ini adalah ELMO, MBSVM, XLnet, BERT, RoBERTa, CASCADE, dan yang terakhir adalah gabungan RNN dan CNN dengan RoBERTa [15]. Dataset yang digunakan adalah Twitter, Reddit, IAC, transcript and dialog. Hasil pada penelitian ini menunjukkan jika metode terakhir tersebut adalah yang paling baik performanya dengan akurasi sebesar 0.78 dan F1-Score sebesar 0.79.

Penelitian selanjutnya adalah penelitian mengenai deteksi sarkasme dengan metode hybrid neural network pada tahun 2022 dibarengi dengan pembuatan dataset News Headlines Dataset yang akan digunakan pada penelitian ini. Dataset berdasarkan judul berita yang ada pada TheOnion dan HuffPost[12]. Metode yang digunakan pada penelitian ini adalah Hybrid LSTM digabungkan dengan CNN. Hasil menunjukkan akurasi yang baik yaitu 0.90 dengan saran pengembangan adalah memperbaiki *tuning parameters*, pengembangan lebih baik buat

arsitektur, dan pengetahuan yang lebih mumpuni sehingga sistem bisa lebih baik performanya. Masih di tahun yang sama, penelitian berikutnya berfokus pada deteksi sarkasme dan ironi menggunakan CNN dengan tujuh layer yang berbeda[13]. Dataset yang digunakan adalah dataset dari Twitter yang sudah dilabeli sarkasme atau ironi. Hasil penelitian ini menunjukkan rata-rata akurasi sebesar 0.75.

Selanjutnya pada tahun 2019 sebagai pembandingan deteksi sarkasme dengan menggunakan metode lainnya yaitu menggunakan metode sederhana seperti Support Vector Machine, Naive Bayes, dan Random Forest untuk klasifikasinya[9]. Dataset yang digunakan dalam penelitian ini adalah dataset Twitter hasil *crawling*. Penelitian ini juga menganalisis dampak dari sarkasme terhadap sentimen pada Twitter di Indonesia. Hasil menunjukkan precision dan recall dimenangkan oleh Naive Bayes dengan nilai 0.63 dan 0.64, akurasi dimenangkan oleh Random Forest sebesar 0.61, dan F1-Score tertinggi diperoleh saat menggunakan SVM sebesar 0.58. Masih di tahun 2019, terdapat penelitian dengan menggabungkan beberapa metode seperti ekstraksi fitur menggunakan unigram dan empat set fitur Boazizi dengan deteksi sarkasme Random Forest kemudian dan ekstraksi fitur dengan TF-IDF dengan deteksi sarkasme menggunakan Naive Bayes [16]. Dataset yang digunakan adalah dataset Twitter hasil *crawling* dan label manual menggunakan ahli. Hasil menunjukkan rata-rata akurasi terbaik 0.80, *presisi* 0.83, dan *recall* 0.91. Penelitian tersebut menyebutkan hal yang perlu diperhatikan adalah kata-kata yang tidak baku serta konteks kalimat.

Penelitian lainnya pada tahun 2022 menganalisis deteksi sarkasme menggunakan LSTM dan BiLSTM yang kemudian dibandingkan [17]. Dataset yang digunakan pada penelitian tersebut sama dengan penelitian ini, yaitu dataset News Headlines Dataset. Hasil terbaik yang didapat pada penelitian tersebut adalah dengan metode BiLSTM dengan akurasi, *precision*, dan *recall* sebesar 0.83. Penelitian lain untuk deteksi sarkasme pada tahun 2020 dengan menggunakan Support Vector Machine yang dibandingkan dengan Naive Bayes[7]. Dataset yang digunakan adalah dataset Twitter karena penelitian tersebut berisi tentang deteksi sarkasme pada politik saat itu di Indonesia. Hasil menunjukkan terbaik didapat saat menggunakan SVM dengan nilai akurasi 0.86. Penelitian tersebut menyarankan untuk menggunakan algoritma lain selain SVM dan NB. Penelitian terbaru pada tahun 2023 terdapat deteksi sarkasme dengan menggunakan Random Forest sebagai metode klasifikasinya[8]. Dataset yang digunakan pada penelitian tersebut adalah dataset Facebook dengan akurasi 0.53. Saran yang tertera adalah untuk penggunaan kata serapan dan kata-kata tidak baku masih belum ada sehingga masih sangat bisa untuk dikembangkan lebih lanjut. *State of the art* pada penelitian ini dapat dilihat pada Tabel 1 dan peta jalan dapat dilihat pada Gambar 1 sebagai berikut.

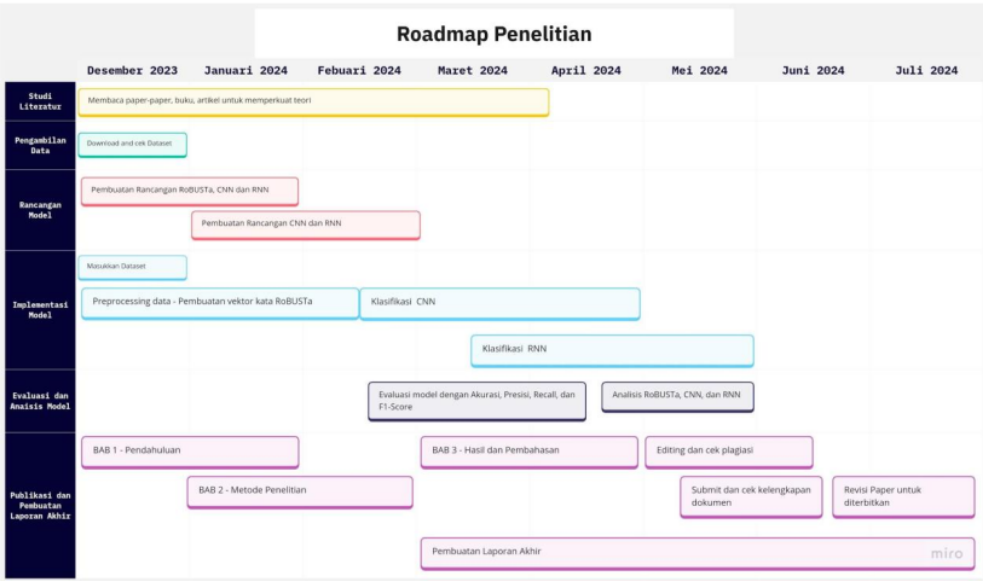
Tabel 1. State of The Art Penelitian

No	Penulis	Tahun	Judul	Isi	Metode	Hasil
1	Hazarika, D., dkk.	2018	CASCADE: Contextual	Bahasan: Pembuatan model deteksi sarkasme dengan	Hybrid Approach:	Tertinggi CASCADE :

			Sarcasm Detection in Online Discussion Forums	metode CASCADE lalu evaluasi performa dan dibandingkan dengan metode lain. Dataset: SARC Future works: gunakan RNN bisa lebih baik	- CNN - CNN SVM - CUE CNN - CASCADE (no personality feature) - CASCADE	akurasi 0.74 f1 0.75
2	Alita, D., dkk.	2019	Analysis of Emoticon and Sarcasm Effect on Sentiment Analysis of Indonesian Language on Twitter	Bahasan: menganalisis emosi dan efek dari sarkasme pada sentimen di Twitter Indonesia Dataset: Twitter (crawl) Future works: -	Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF)	Precision recall 62.82 64.23 (NB) Accuracy 60.61 (RF) F1 Score 57.78 (SVM)
3	Yunitasari, Y., dkk.	2019	Sarcasm Detection For Sentiment Analysis in Indonesian Tweets	Bahasan: sarkasme untuk analisis sentimen di Twitter Indonesia Dataset: Twitter (crawl) Future works: slang word perlu diperhatikan, konteks juga perlu diperhatikan	Ekstraksi fitur --> unigram dan 4 set fitur Boazizi Deteksi sarkasme RF Ekstraksi Fitur sentimen TF-IDF klasifikasi NB	Akurasi 80.4, presisi 83.2, recall 91.3
4	Ardiansyah, F. dan Kartika, D.	2020	Sarcasm Detection on Indonesian Politics Tweet Using Multi Labeling Method and Support Vector Machine	Bahasan: Deteksi sarkasme pada politik di Indonesia menggunakan Multi Labeling Method dan Support Vector Machine Dataset: Twitter (crawl) Future works: untuk labeling coba menggunakan linguistic expert (human) dan metode lain perlu dicoba	Support Vector Machine (SVM), Naive Bayes (NB)	Tertinggi SVM Akurasi: 0.86 Sensitivitas 0.86 Specificity 0.87 Precision 0.87
5	Jaiswal, Nikhil	2020	Neural Sarcasm Detection using Conversation Context	Bahasan: pembuatan model deteksi sarkasme dengan deep neural architecture dan membandingkan performa dari pre-trained language representation models (PLTM) Dataset: Twitter (crawl)	ELMo USE BERT RoBERTa	Tertinggi RoBERTa - Precision 0.77 - Recall 0.78 - F1 Score 0.77

				Future works: tuning parameter, gunakan pada dataset publik, gunakan metode lain		
6	Yaghoobian, H., dkk.	2021	Sarcasm Detection: A Comparative Study	Bahasan: membandingkan metode-metode untuk deteksi sarkasme Dataset: Twitter, Reddit, IAC, transcript and dialog Future works: -	ELMo MBSVM XLnet BERT RoBERTa CASCADE RCNN- RoBERTa	Tertinggi RCNN - RoBERTa accuracy 0.79 F1-score 0.78
13 7	Misra, R. and Arora, P.	2022	Sarcasm Detection using Hybrid Neural Network	Bahasan: membuat model deteksi sarkasme dengan menggunakan Hybrid Neural Network (LSTM + CNN) Dataset: News Headlines dataset (TheOnion dan HuffPost) Future works: gunakan arsitektur yang lebih baik, gunakan tuning parameter yang lebih baik, dan gunakan pengetahuan yang lebih baik	Hybrid Approach LSTM and CNN	Akurasi 0.90
8 9	Hilmawan, M.D.	2022	Deteksi Sarkasme Pada Judul Berita Berbahasa Inggris Menggunakan Algoritme Bidirectional LSTM	Bahasan: membuat model deteksi sarkasme menggunakan BiLSTM yang dibandingkan dengan LSTM biasa. Dataset: News Headlines dataset (TheOnion dan HuffPost) Future works: -	LSTM, BiLSTM	Tertinggi BiLSTM Akurasi, Presisi, Recall, F1 Score 0.83
9	Wijaya, A.C., and Wibawa, I.G.A.	2022	Deteksi Sarkasme dan Ironi pada Twitter dengan Menggunakan Metode CNN	Bahasan: Membuat model deteksi sarkasme dan ironi pada Twitter menggunakan metode CNN dengan 7 layer yang berbeda. Dataset: Twitter Future works: -	CNN dengan 7 layer	akurasi 0.75
10	Arsya, C.S. and Elsera, M.	2023	Implementasi Random Forest dalam Melakukan Klasifikasi Kata	Bahasan: Membuat model klasifikasi sarkasme pada media sosial Facebook dengan menggunakan algoritma Random Forest	Random Forest (RF)	Akurasi 0.53

			Sarkasme pada Media Sosial Facebook	Dataset: Facebook (crawl) Future works: Kurang mampu pada kata-kata asing atau serapan dan kata-kata tidak baku		
--	--	--	-------------------------------------	--	--	--



Gambar 1. Roadmap Penelitian

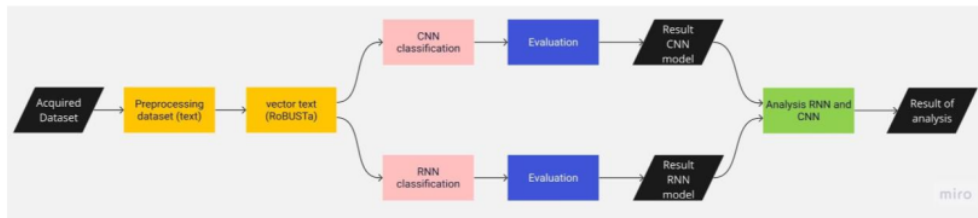
METODE PENELITIAN

Proses awal yang dilakukan adalah pengumpulan data. Pada penelitian ini, data menggunakan dataset yang berisi judul berita yang telah dilabeli ke dalam sarkasme atau tidak. Judul berita biasanya menggunakan bahasa yang formal. Oleh sebab itu, hal ini menarik untuk dikaji seperti bagaimana sebuah model bisa menebak apakah judul berita ini sarkasme atau tidak, padahal sarkasme biasanya ada di kalimat-kalimat tidak formal. Dataset ini diambil dari TheOnion dan HuffPost yang terdiri dari 26,709 judul berita. Dataset ini dibuat oleh Misra, R. and Arora, P.[12] dan sudah memiliki lisensi sendiri yaitu Attribution 4.0 International (CC BY 4.0) yang berarti bisa digunakan untuk penelitian. Dataset ini bisa diakses melalui github <https://github.com/rishabhmisra/News-Headlines-Dataset-For-Sarcasm-Detection>.

Setelah pengumpulan dataset, dataset yang berbentuk kata akan direpresentasikan menjadi vektor-vektor paragraf dengan menggunakan RoBERTa. Metode tersebut baik untuk representasi awal teks dan RoBERTa adalah optimasi dari BERT [18]. Setelah dilakukan representasi awal teks, tahap selanjutnya adalah membentuk model yang bisa memprediksi kalimat tersebut adalah sarkasme atau bukan. Metode yang digunakan untuk model dalam penelitian ini adalah CNN dan RNN. CNN adalah versi tradisional dari ANN. CNN dan ANN sama-

sama mempunyai neuron yang selalu mengoptimalkan diri setiap iterasinya. CNN sangat baik untuk mencari pola pada data gambar, namun dia juga sangat baik dalam proses ekstraksi fitur yang kompleks [19]. Namun pada penelitian yang sama, disebutkan juga penggunaan RNN bisa membuat hasil menjadi lebih baik. Selain itu, pembelajaran mengenai dasar dari RNN dapat membawa keuntungan di bidang pembelajaran mesin [20]. Sehingga pada penelitian ini akan dilakukan evaluasi antara CNN dan RNN dengan representasi awal kata menggunakan RoBERTa.

Setelah mencari pola dengan CNN atau RNN, kedua model yang telah dibentuk mampu memprediksi mana kalimat sarkasme dan mana yang bukan. Model-model tentunya akan dievaluasi dengan akurasi, presisi, recall, dan F1-Score nya. Setelah itu kedua model tersebut akan dianalisis performanya dan dianalisis untuk kelemahan kekurangan kedua model tersebut. Hal ini berguna untuk penelitian-penelitian selanjutnya, sehingga penelitian selanjutnya akan tau bagaimana harus dikembangkan terutama untuk kedua model tersebut. Diagram alir untuk penelitian ini dapat dilihat pada Gambar 2 sebagai berikut.



Gambar 2 Diagram Alir Deteksi Sarkasme

HASIL_Sheraton Pawestri

ORIGINALITY REPORT

13%

SIMILARITY INDEX

12%

INTERNET SOURCES

5%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1

www.researchgate.net

Internet Source

2%

2

ejournal.unesa.ac.id

Internet Source

1%

3

repository.unhas.ac.id

Internet Source

1%

4

ceur-ws.org

Internet Source

1%

5

eprints.umm.ac.id

Internet Source

1%

6

ejnteti.jteti.ugm.ac.id

Internet Source

1%

7

kopiluwakliar01.blogspot.com

Internet Source

1%

8

journal.binadarma.ac.id

Internet Source

1%

9

jurnal.umb.ac.id

Internet Source

1%

10 Submitted to Program Pascasarjana
Universitas Negeri Yogyakarta <1 %
Student Paper

11 Submitted to Universitas Diponegoro <1 %
Student Paper

12 repositori.unsil.ac.id <1 %
Internet Source

13 Saumya Bhardwaj, Manas Ranjan Prusty.
"BERT Pre-processed Deep Learning Model
for Sarcasm Detection", National Academy
Science Letters, 2022 <1 %
Publication

14 core.ac.uk <1 %
Internet Source

15 www.eecs.yorku.ca <1 %
Internet Source

16 ejournal.undiksha.ac.id <1 %
Internet Source

17 journal.fkpt.org <1 %
Internet Source

18 e-journal.hamzanwadi.ac.id <1 %
Internet Source

19 mmt.its.ac.id <1 %
Internet Source

doku.pub

20

Internet Source

<1 %

21

es.scribd.com

Internet Source

<1 %

22

www.coretanzone.id

Internet Source

<1 %

23

id.scribd.com

Internet Source

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On