Klasifikasi Akun Bot dan Human di Twitter dalam Domain Crypto Menggunakan Algoritma Machine Learning

Ahmad Sirajuddin Zaki[1], Endang Wahyu Pamungkas[2]

[1]Universitas Muhammadiyah Surakarta, Jl. A. Yani, Mendungan, Pabelan, Kec. Kartasura, Kabupaten Sukoharjo, Jawa Tengah, Indonesia

[2] Universitas Muhammadiyah Surakarta, Jl. A. Yani, Mendungan, Pabelan, Kec. Kartasura, Kabupaten Sukoharjo, Jawa Tengah, Indonesia

Email korespondensi: l200210081@student.ums.ac.id

Abstrak. Mata uang kripto telah mengalami peningkatan signifikan dalam dekade terakhir, dengan Twitter menjadi platform utama komunikasi antara penerbit aset dan investor. Namun, prevalensi bot otomatis yang terlibat dalam manipulasi pasar menimbulkan risiko bagi investor. Penelitian ini bertujuan mengembangkan model machine learning untuk mengklasifikasikan akun bot dan human di Twitter dalam domain cryptocurrency. Dataset dikumpulkan dari 10 akun Twitter (5 human expert dan 5 AI agents) dengan total 19.798 tweet setelah deduplikasi. Metodologi meliputi preprocessing modular, ekstraksi 15 fitur linguistik dan metrik interaksi, serta implementasi arsitektur hybrid CNN-LSTM dengan dynamic thresholding. Hasil menunjukkan bahwa model XGBoost dengan fitur lengkap mencapai akurasi tertinggi 97.22% dan F1-score 0.9722, sementara CNN-LSTM mencapai F1-score 91.59% pada konfigurasi linguistik. Fitur seperti capital_token_ratio, punctuation_ratio terbukti menjadi indikator kuat pembeda bot dan human. Penelitian ini berkontribusi pada pengembangan sistem deteksi bot yang dapat membantu investor mengidentifikasi manipulasi pasar dalam ekosistem cryptocurrency.

Kata kunci: *deteksi bot; machine learning; cryptocurrency; Twitter; CNN-LSTM*

PENDAHULUAN

Mata uang kripto telah mengalami peningkatan signifikan dalam dekade terakhir, menarik perhatian akademisi dan industri. Pasar kripto sangat bergantung pada platform media sosial seperti Twitter untuk komunikasi antara penerbit aset dan investor. Twitter

menjadi saluran utama untuk pengumuman dan pembaruan, menjadikannya sumber informasi krusial bagi investor [1]. Namun, sifat pasar yang sangat spekulatif membuatnya rentan terhadap berbagai aktivitas penipuan.

Media sosial tidak hanya memperkuat visibilitas diskusi yang sah, tetapi juga menjadi sarana penyebaran penipuan seperti skema pump-and-dump dan Ponzi. Penelitian [2] terhadap 50 juta pesan di Twitter, Telegram, dan Discord mengungkap bahwa 56% akun promosi merupakan bot, dengan 93% tautannya mengarah ke skema pump-and-dump. Fenomena ini semakin kompleks karena karakteristik media sosial yang anonim, memungkinkan bot menyamar sebagai analis finansial berpengalaman [3].

Platform media sosial mengalami lonjakan akun bot otomatis yang terlibat dalam manipulasi pasar. Meskipun berbagai strategi deteksi telah dikembangkan, meningkatnya kecanggihan bot menimbulkan tantangan baru. Sistem deteksi umumnya menggunakan metrik akun, pola keterlibatan, dan analisis konten sebagai indikator utama aktivitas bot. Studi [4] menunjukkan bahwa bot jarang berinteraksi secara bermakna dengan pengguna manusia, dan fitur leksikal secara signifikan meningkatkan akurasi klasifikasi. Berbagai model pembelajaran mesin seperti Decision Tree, Logistic Regression, dan Neural Networks telah dievaluasi untuk deteksi bot, dengan jaringan saraf menunjukkan efektivitas tinggi dalam mengklasifikasikan perilaku bot yang menyerupai manusia [5].

Prevalensi misinformasi dan skema penipuan dalam investasi daring menambah risiko bagi investor. Faktor psikologis seperti sentimen, herd behavior, dan overreaction memainkan peran penting dalam pengambilan keputusan investasi [6]. Algoritme seperti Support Vector Machines (SVM) dan Random Forest telah menunjukkan efektivitas dalam deteksi bot dan penyaringan misinformasi [7]. Namun, variasi kinerja antara pelatihan dan implementasi dunia nyata menunjukkan perlunya perbaikan lebih lanjut [8].

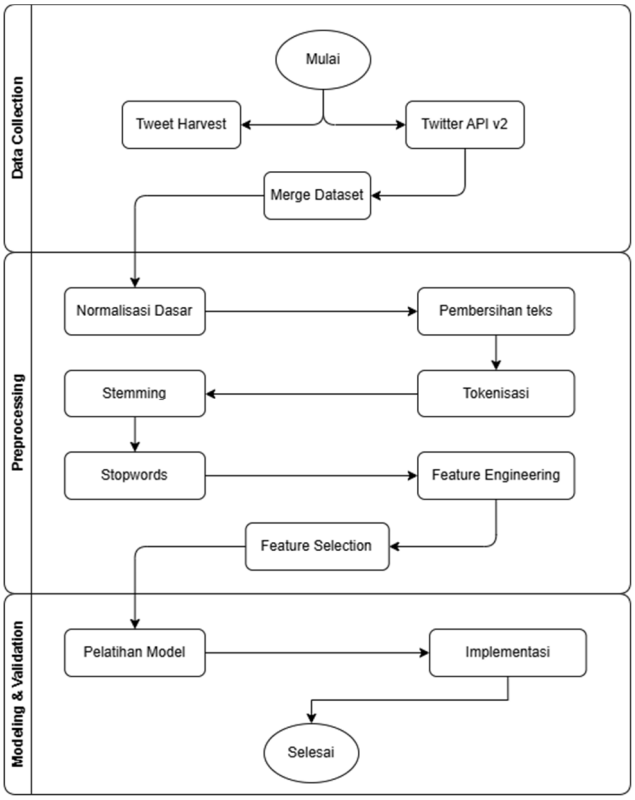
Penelitian terkini berfokus pada peningkatan model deteksi bot untuk menganalisis aktivitas Twitter terkait mata uang kripto. Jaringan saraf dan penyematan linguistik telah menunjukkan hasil menjanjikan dalam mendeteksi bot tanpa memerlukan fitur buatan atau asumsi sebelumnya [9]. Beberapa pendekatan mengutamakan skalabilitas dengan menggunakan metadata akun minimal dan pemilihan data strategis [10]. Fitur yang tidak bergantung pada bahasa juga telah dieksplorasi untuk meningkatkan portabilitas dalam berbagai konteks linguistic [11]. Upaya terbaru dalam deteksi bot multimodal seperti BotRGCN [12] dan BotSAI [13] telah membuktikan efektivitas integrasi fitur tekstual-jaringan. Penelitian [14] menemukan bahwa model Random Forest mengalami penurunan akurasi 12-15% ketika diterapkan pada tweet berbahasa Indonesia, menekankan pentingnya pengembangan sistem deteksi yang tidak

bergantung pada bahasa tertentu.

Berdasarkan identifikasi celah penelitian sebelumnya, studi ini merancang solusi integratif melalui tiga pendekatan inovatif. Pertama, pengembangan rekayasa fitur fusi yang mengombinasikan 15 fitur linguistik dengan metrik interaksi, mengadaptasi temuan [15] tentang analisis teks anomali dan [16] mengenai pola perilaku dengan modifikasi spesifik untuk konteks kripto. Kedua, arsitektur hybrid CNN-LSTM yang memadukan convolutional neural networks untuk mengekstraksi pola leksikal terinspirasi dari [5] dengan long short-term memory networks dalam menganalisis urutan temporal, dengan target capaian 85% F1-score pada dataset TwiBot-20 [12]. Ketiga, implementasi dynamic thresholding berbasis cost-sensitive learning [17] yang secara adaptif menyesuaikan klasifikasi untuk mengatasi ketidakseimbangan data.

Tujuan penelitian ini adalah: (1) Mengembangkan model machine learning untuk mengklasifikasikan akun bot dan human di Twitter dalam domain cryptocurrency; (2) Menganalisis fitur linguistik dan metadata yang dapat membedakan akun bot dan human; (3) Mengevaluasi performa model klasifikasi menggunakan metrik akurasi, *precision*, *recall*, dan F1-score.

METODE



Gambar 1. Sistematika Alur

Pengumpulan Data

Pengumpulan data dilakukan menggunakan Tweet Harvest, pustaka berbasis Node.js yang memungkinkan web crawling terhadap tweet publik tanpa memerlukan akses penuh ke API Twitter. Dataset dikumpulkan dari 10 akun Twitter yang dikategorikan menjadi 5 akun Human Expert analis kripto terverifikasi dan 5 akun AI Agents dengan konten kripto. Rentang waktu pengambilan menggunakan filter lang:en untuk memastikan hanya tweet berbahasa Inggris yang dikumpulkan, dengan maksimum 2000 tweet per akun.

Selain web crawling, penelitian ini memanfaatkan Twitter API v2 untuk memperkaya dataset dengan metadata akun menggunakan library Tweepy. Metadata meliputi jumlah pengikut, jumlah akun yang diikuti, total tweet, deskripsi profil, dan tanggal pembuatan akun. Data akhir disimpan dalam format CSV dan digunakan untuk proses *feature engineering*.

Preprocessing Data

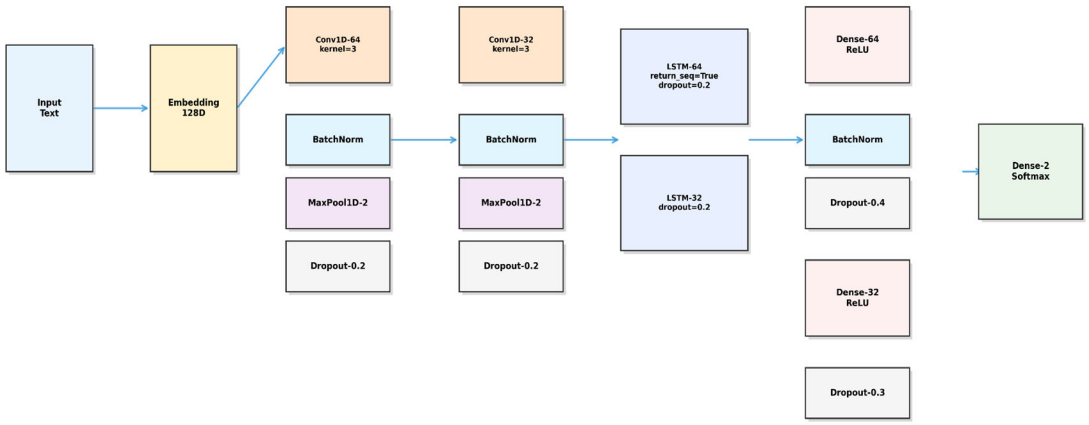
Tahapan *preprocessing* dilakukan secara modular menggunakan beberapa skrip Python yang masing-masing berfokus pada tugas spesifik:

1. **Normalisasi Data**: Pembersihan elemen seperti URL, mention, hashtag, serta ekstraksi fitur numerik seperti typo_ratio.
2. **Tokenisasi**: Menggunakan spaCy untuk tokenisasi dan ekstraksi fitur linguistik lanjutan termasuk distribusi part-of-speech (POS) dan identifikasi entitas bernama.
3. **Stemming dan Lemmatisasi**: Penerapan berbagai metode stemming (Porter, Lancaster, Snowball) dan lemmatizer spaCy untuk menyederhanakan kata ke bentuk dasar.
4. **Penghapusan Stopwords**: Fokus pada penghapusan stopword dengan tetap mempertahankan konteks penting seperti kata negasi dan istilah relevan dalam dunia kripto. Penelitian [18] mengusulkan metode Term Based Random Sampling untuk membuat daftar stopword kustom yang mengungguli pendekatan tradisional.
5. **Ekstraksi Fitur Lanjutan**: Menambah fitur kompleksitas linguistik seperti distribusi POS tag [19], metrik kompleksitas teks, dan variasi panjang kalimat.
6. **Seleksi Fitur**: Menggunakan metode statistik ANOVA (f_classif) dan mutual information [20] untuk mempertahankan sekitar 15 fitur terbaik dengan daya pembeda tinggi.

Pemodelan

Proses pemodelan menggunakan algoritma klasik seperti: Logistic Regression (regularisasi L2), Random Forest (n_estimators=100), SVM, XGBoost (penyesuaian bobot kelas otomatis), dan deeplearning CNN-LSTM. Pipeline pelatihan memproses data teks melalui TF-IDF dan fitur numerik melalui standardisasi sebelum masuk ke classifier.

Arsitektur CNN-LSTM menggabungkan lapisan konvolusi untuk mengekstraksi pola leksikal dengan LSTM untuk menganalisis urutan temporal kata [21]. Model ini dilengkapi dengan dynamic thresholding berbasis cost-sensitive learning untuk menangani ketidakseimbangan data [22]. Arsitektur CNN-LSTM yang dikembangkan terdiri dari layer embedding dengan dimensi 128, diikuti oleh dua layer konvolusi dengan 64 filter dan kernel size 3, layer LSTM dengan 64 unit untuk menangkap dependensi temporal, dan layer dense untuk klasifikasi akhir. Gambar 2 menunjukkan arsitektur lengkap model CNN-LSTM yang digunakan dalam penelitian ini.



Gambar 2. Arsitektur CNN-LSTM

Konfigurasi hyperparameter model CNN-LSTM dioptimasi melalui grid search dengan fokus pada performa F1-score. Tabel 1 menyajikan detail hyperparameter yang digunakan, termasuk parameter pra-pemrosesan, arsitektur jaringan, strategi jaringan.

Tabel 1. HyperParameter cnn-lstm

Kategori Parameter	Parameter	Nilai
Pra-pemrosesan Teks	max_features	10,000
	max_len	100
	embedding_dim	128
	oov_token	'<OOV>'
Layer CNN (Conv1D)	filters	64
	kernel_size	3
	activation	'relu'
	padding	'same'
Layer LSTM	units	64
	return_sequences	True
	dropout	0.2
	recurrent_dropout	0.2
Layer Dense	units	64
	activation	'relu'
Layer Output Dense	units	2
	activation	'softmax'
Dropout Rate (Setelah Conv1D + MaxPooling)	Rate	0.2
Dropout Rate (Setelah Dense + BatchNorm)	Rate	0.4
Parameter Pelatihan	initial_lr	0.0015

	epochs	50
	batch_size	32
	optimizer	Adam
	loss	'categorical_crossentropy'
Callback (EarlyStopping)	monitor	'val_loss'
	patience	7
	restore_best_weights	True
Callback (ReduceLROnPlateau)	monitor	'val_loss'
	factor	0.5
	patience	3
	min_lr	0.00001
Cost-Sensitive Learning	cost_matrix	[[0, 2], [1, 0]]
	class_weight	'balanced'

Validasi

Untuk memastikan generalisasi model, digunakan metrik utama F1-score (weighted makro). Evaluasi akhir dilakukan menggunakan 20% data uji yang telah distratifikasi, menghasilkan classification report, dan visualisasi berupa perbandingan antar model. Untuk mengatasi ketidakseimbangan kelas, diterapkan cost-sensitive learning dengan matriks biaya [[0,2], [1,0]] dan dynamic thresholding.

Tabel 2. Hyperlane data testing

Kategori	**Parameter**	**Nilai**
Threshold Optimization	cost_matrix	[[0, 2], [1, 0]]
	optimal_threshold	Dinamis (dihitung)
	default_threshold	0.5
Evaluation Metrics	average	'weighted'
	output_dict	True
Data Split	test_size	Dari *preprocessing*
	validation_split	Implisit

HASIL

Dataset berhasil dikumpulkan dengan total 20.038 tweet yang setelah deduplikasi menjadi 19.798 tweet dari 10 akun (5 human expert dan 5 AI agents). Data dibagi menjadi 80% training dan 20% testing dengan stratifikasi untuk mempertahankan proporsi kelas.

Tabel 3. Distribusi akun

Jumlah tweet					
Type account	**Username**	**Crawling**	**Deduplikasi**	**train**	**test**
human	Human_01	2013	1968	80%	20%
	Human_02	2001	2001		
	Human_03	2001	1999		
	Human_04	2002	1931		
	Human_05	2002	2000		
bot	Bot_01	2009	2003	80%	20%
	Bot_02	2010	1998		
	Bot_03	2000	1998		
	Bot_04	2000	1908		
	Bot_05	2000	1992		
Total		**20038**	**19798**		

Word cloud menunjukkan perbedaan karakteristik linguistik yang jelas antara kedua model. Full model dengan TF-IDF didominasi oleh terminologi kripto spesifik seperti "BITCOIN", "whale", "price", dan "volume", menunjukkan fokus bot pada konten finansial

Gambar 3. Wordcloud tf idf

teknis. Linguistic model menampilkan distribusi kata yang lebih beragam dengan penekanan pada kata-kata umum seperti "is", "the", dan "with", mengindikasikan gaya komunikasi yang lebih natural pada akun human. Perbedaan ukuran dan frekuensi kata

ini menjadi indikator kuat untuk membedakan pola komunikasi bot dan human dalam domain cryptocurrency.

Tabel 4. Perbandingan Fitur Kombinasi(Fulmodel) Antara Akun Manusia dan Bot

	min		***max***		***mean***		***std dev***			
	h	*b*	*h*	*b*	*h*	*b*	*h*	*b*	***Dif(%)***	***n***
capital_token_ratio	0.00	0.00	1.00	1.00	0.33	0.09	0.29	0.11	73.95	9899
like_count	0.00	0.00	149725	2748	1643.44	29.01	3892.36	82.41	98.23	9899
caps_ratio	0.00	0.00	1.00	0.90	0.21	0.04	0.25	0.05	79.52	9899
capital_tokens	0.00	0.00	52.00	47.00	5.70	2.50	6.27	3.72	56.05	9899
punctuation_ratio	0.00	0.00	0.30	0.20	0.03	0.02	0.02	0.01	44.51	9899
retweet_count	0.00	0.00	32942	1177	219.91	3.10	711.58	21.52	98.59	9899
mention_count	0.00	0.00	16.00	35.00	0.63	2.08	0.98	3.20	69.86	9899
reply_count	0.00	0.00	44085	4344	241.56	12.44	835.37	57.25	94.85	9899
pos_noun_ratio	0.00	0.00	1.00	1.00	0.18	0.23	0.12	0.10	22.92	9899
text_length	4.00	4.00	436.00	692	117.93	170.36	78.37	92.71	30.78	9899
ari_score	-12.89	-9.57	85.05	176.89	2.18	4.07	4.22	5.26	46.56	9899
token_diversity	0.34	0.12	1.00	1.00	0.93	0.88	0.09	0.13	5.35	9899
lancaster_reduction_ratio	0.00	0.00	0.58	0.55	0.11	0.14	0.07	0.06	18.57	9899
fk_grade	-3.01	-3.40	43.80	36.71	5.19	6.53	4.19	4.40	20.44	9899

Hasil analisis menunjukkan perbedaan yang signifikan antara akun human dan bot. Fitur engagement seperti like_count (98.23% perbedaan), retweet_count (98.59%), dan reply_count (94.85%) menunjukkan bahwa akun human memiliki interaksi sosial yang jauh lebih tinggi. Sebaliknya, fitur linguistik seperti capital_token_ratio (73.95% perbedaan) dan caps_ratio (79.52%) mengindikasikan bahwa akun human lebih ekspresif dalam penggunaan huruf kapital. Bot cenderung memiliki mention_count dan text_length yang lebih tinggi, menunjukkan karakteristik posting yang lebih terstruktur.

Tabel 5. Perbandingan Fitur Linguistik Saja Antara Akun Manusia dan Bot

	min		*max*		*mean*		*std dev*			
	h	*b*	*h*	*b*	*h*	*b*	*h*	*b*	*Dif(%)*	***n***
text_length	4.00	4.00	436.00	692.00	117.93	170.36	78.37	92.71	30.78	9899
sentence_count	0.00	0.00	12.00	14.00	2.00	2.92	1.26	2.52	31.48	9899
punctuation_ratio	0.00	0.00	0.30	0.20	0.03	0.02	0.02	0.01	44.51	9899
token_diversity	0.34	0.12	1.00	1.00	0.93	0.88	0.09	0.13	5.35	9899
avg_word_length	2.33	1.75	34.00	22.33	5.69	5.54	1.59	1.39	2.79	9899
word_length_variance	0.00	0.00	1099.13	400.00	5.12	6.58	11.87	14.63	22.14	9899
flesch_score	133.6	470.9	121.22	121.22	68.21	68.41	30.03	26.81	0.29	9899
flesch_kincaid	-3.40	-3.40	32.39	79.20	4.57	5.35	3.90	4.03	14.71	9899
ari_score	- 12.89	-9.57	85.05	176.89	2.18	4.07	4.22	5.26	46.56	9899
avg_sentence_length	0.00	0.00	42.00	37.00	5.93	6.94	4.15	4.99	14.67	9899
noun_ratio	0.00	0.00	1.00	1.00	0.29	0.31	0.17	0.13	4.01	9899
verb_ratio	0.00	0.00	1.00	1.00	0.23	0.20	0.14	0.10	12.62	9899
adj_ratio	0.00	0.00	1.00	1.00	0.08	0.10	0.11	0.08	17.54	9899

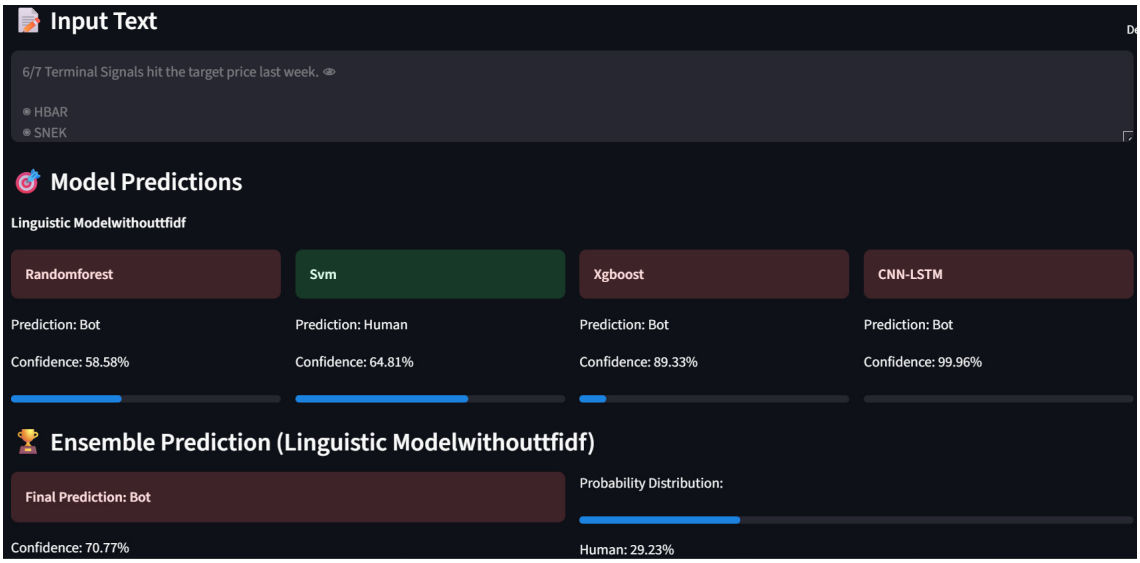
Fokus pada fitur linguistik murni mengungkap pola yang menarik. Bot memiliki text_length rata-rata lebih panjang (170.36 vs 117.93 karakter) dan sentence_count yang lebih tinggi (2.92 vs 2.00), menunjukkan gaya komunikasi yang lebih verbose. Skor keterbacaan seperti ari_score menunjukkan perbedaan 46.56%, di mana bot cenderung menggunakan struktur kalimat yang lebih kompleks. Fitur token_diversity yang relatif similar (5.35% perbedaan) mengindikasikan bahwa bot sudah cukup canggih dalam variasi kosakata.

Tabel 6. Hasil Training

Model Name	Algoritma	Accuracy	ScoreF1	Precision	Recall	AUC
fulmodel + tfidf	RandomForest	0.9707	0.9707	0.9707	0.9707	0.9947
	LogisticRegression	0.8740	0.8738	0.8764	0.8740	0.9524

	SVM	0.8856	0.8856	0.8859	0.8856	0.9567
	XGBoost	0.9722	0.9722	0.9723	0.9722	0.9961
fulmodel No tfidf	CNN-LSTM (Opt)	0.9109	0.9108	0.9122	0.9109	0.9725
	RandomForest	0.9712	0.9712	0.9712	0.9712	0.9940
	LogisticRegression	0.8525	0.8517	0.8603	0.8525	0.9346
	SVM	0.9045	0.9045	0.9045	0.9045	0.9673
	XGBoost	0.9699	0.9699	0.9700	0.9699	0.9951
linguistic + tfidf	RandomForest	0.8359	0.8359	0.8359	0.8359	0.9260
	LogisticRegression	0.7798	0.7798	0.7799	0.7798	0.8709
	SVM	0.8263	0.8262	0.8267	0.8263	0.9019
	XGBoost	0.8503	0.8502	0.8505	0.8503	0.9277
linguistic no tfidf	CNN-LSTM (Opt)	0.9159	0.9159	0.9161	0.9159	0.9738
	RandomForest	0.8164	0.8164	0.8165	0.8164	0.9011
	LogisticRegression	0.7220	0.7205	0.7269	0.7220	0.7903
	SVM	0.8093	0.8093	0.8094	0.8093	0.8838
	XGBoost	0.8126	0.8126	0.8127	0.8126	0.8981

Sebagai bagian dari validasi akhir sistem klasifikasi akun bot dan human, dilakukan uji coba langsung melalui aplikasi demo Streamlit yang telah dibangun. Pada percobaan ini, digunakan sebuah teks tweet bertema crypto dengan label asli bot. Teks tersebut kemudian dimasukkan ke dalam aplikasi, dan hasil prediksi dari lima model yaitu Random Forest, SVM, XGBoost, CNN-LSTM, serta ensemble ditampilkan secara bersamaan. Mayoritas model berhasil memprediksi akun sebagai bot, dengan tingkat keyakinan tertinggi diberikan oleh CNN-LSTM (99.96%) dan XGBoost (89.33%). Hanya SVM yang memprediksi sebagai human dengan probabilitas 64.81%. Hasil ini menunjukkan bahwa pendekatan ensemble dan model deep learning memiliki performa yang lebih stabil untuk menangani pola bahasa yang khas dari akun bot. Visualisasi hasil ditampilkan pada Gambar 4 berupa tampilan demo Streamlit, yang memperlihatkan



Gambar 4. Demo Streamlit

input teks, label asli, dan prediksi dari masing-masing model.

PEMBAHASAN

Hasil penelitian menunjukkan bahwa akun bot dan human memiliki karakteristik yang berbeda secara signifikan, baik dari aspek linguistik maupun metrik interaksi. Bot cenderung menggunakan struktur bahasa yang lebih kaku dengan variasi kata yang lebih rendah dan skor keterbacaan yang lebih tinggi, menunjukkan gaya penulisan yang lebih teknis dan tidak alami.

Temuan ini sejalan dengan hipotesis bahwa meskipun bot mencoba meniru manusia, mereka masih meninggalkan anomali yang dapat dikenali [15]. Fitur seperti capital_token_ratio dan punctuation_ratio menunjukkan bahwa akun human lebih

ekspresif dalam gaya penulisan, sementara bot lebih terstruktur dan konsisten.

Penggunaan kombinasi fitur leksikal, metrik interaksi, dan TF-IDF terbukti sangat efektif. Kata-kata seperti "bitcoin" atau "whale" dari TF-IDF menjadi indikator kuat akun bot [23]. Model yang menggabungkan berbagai jenis fitur mencapai akurasi hingga 97%, jauh lebih tinggi dibandingkan model yang hanya mengandalkan metadata konvensional.

Arsitektur CNN-LSTM dengan dynamic thresholding berhasil memberikan hasil kompetitif dalam menangani data tidak seimbang [24]. Model ini mampu menangkap pola sekuensial dan lokal dari teks mentah, dengan penyesuaian ambang batas yang mengurangi kesalahan klasifikasi kelas minoritas. Implementasi pada platform Streamlit menunjukkan bahwa model dapat digunakan secara praktis untuk deteksi. Pendekatan ensemble yang menggabungkan prediksi dari multiple algorithms memberikan stabilitas keputusan yang lebih baik. Keterbatasan penelitian ini meliputi fokus pada data bahasa Inggris, limitasi API, dan evolusi bot yang terus berkembang. Penelitian masa depan dapat mengeksplorasi deteksi multilingual dan adaptasi terhadap perkembangan teknologi bot yang semakin canggih.

KESIMPULAN

Penelitian ini berhasil mengembangkan sistem klasifikasi akun bot dan human di Twitter dalam domain cryptocurrency dengan tingkat akurasi tinggi. Model XGBoost dengan kombinasi fitur lengkap dan TF-IDF mencapai performa terbaik dengan akurasi 97.22% dan F1-score 0.9722. Arsitektur hybrid CNN-LSTM juga menunjukkan hasil yang kompetitif dengan F1-score 91.59% pada konfigurasi linguistik.

Temuan utama menunjukkan bahwa akun bot dan human memiliki perbedaan karakteristik yang dapat diidentifikasi melalui fitur linguistic, metrik interaksi, dan leksikal. Fitur seperti capital_token_ratio, punctuation_ratio, dan engagement metrics terbukti menjadi indikator kuat pembeda kedua jenis akun. Pendekatan fusi fitur yang menggabungkan aspek leksikal, metrik interaksi, dan TF-IDF jauh lebih efektif dibandingkan pendekatan berbasis metadata konvensional.

Kontribusi praktis penelitian ini adalah penyediaan sistem deteksi bot yang dapat diimplementasikan oleh platform media sosial untuk memitigasi manipulasi pasar, serta membantu investor dalam menginterpretasi dinamika komunitas cryptocurrency secara lebih cerdas. Secara teoritis, penelitian ini memperkuat pentingnya fitur linguistik dan metrik interaksi sebagai indikator kunci dalam deteksi bot di ranah kripto.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada pihak yang telah berkontribusi dalam penyediaan akses data dan dukungan teknis dalam pengembangan sistem klasifikasi ini.

DAFTAR PUSTAKA

[1] J. Zhang dan C. Zhang, "Do cryptocurrency markets react to issuer sentiments? Evidence from Twitter," *Res. Int. Bus. Finance*, vol. 61, hlm. 101656, Okt 2022, doi: 10.1016/j.ribaf.2022.101656.

[2] L. Nizzoli, S. Tardelli, M. Avvenuti, S. Cresci, M. Tesconi, dan E. Ferrara, "Charting the Landscape of Online Cryptocurrency Manipulation," *IEEE Access*, vol. 8, hlm. 113230–113245, 2020, doi: 10.1109/ACCESS.2020.3003370.

[3] M. Mirtaheri, S. Abu-El-Haija, F. Morstatter, G. V. Steeg, dan A. Galstyan, "Identifying and Analyzing Cryptocurrency Manipulations in Social Media," *IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 3, hlm. 607–617, Jun 2021, doi: 10.1109/TCSS.2021.3059286.

[4] I. Inuwa-Dutse, B. Shehu Bello, I. Korkontzelos, Dan R. Heckel, "The Effect Of Engagement Intensity And Lexical Richness In Identifying Bot Accounts On Twitter," *IADIS Int. J. WWWINTERNET*, vol. 16, no. 2, hlm. 50–65, Des 2018, doi: 10.33965/ijwi_2018161204.

[5] D. A. Belokurov, E. S. Shamakova, dan V. S. Kolomoitcev, "Using Machine Learning Techniques to Identify Bot Accounts on a Social Network," dalam *2021 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, Mei 2021, hlm. 1–5. doi: 10.1109/WECONF51603.2021.9470605.

[6] A. Khalisa, C. K. Karismasari, H. H. Ikhsan, dan N. Saraswati, "Pengaruh Behavioral Factors Terhadap Pengambilan Keputusan Investasi Finansial Individu," *Indones. Bus. Rev.*, vol. 3, no. 1, Art. no. 1, Jul 2020, doi: 10.21632/ibr.3.1.15-35.

[7] P. G. Pratama dan N. A. Rakhmawati, "Social Bot Detection on 2019 Indonesia President Candidate's Supporter's Tweets," *Procedia Comput. Sci.*, vol. 161, hlm. 813–820, 2019, doi: 10.1016/j.procs.2019.11.187.

[8] S. Castillo *dkk.*, "Detection of Bots and Cyborgs in Twitter: A Study on the Chilean Presidential Election in 2017," G. Meiselwitz, Ed., dalam Lecture Notes in Computer Science, vol. 11578. Cham: Springer International Publishing, 2019, hlm. 311–323. doi: 10.1007/978-3-030-21902-4_22.

[9] F. Wei dan U. T. Nguyen, "Twitter Bot Detection Using Neural Networks and Linguistic Embeddings," *IEEE Open J. Comput. Soc.*, vol. 4, hlm. 218–230, 2023, doi: 10.1109/OJCS.2023.3302286.

[10] K.-C. Yang, O. Varol, P.-M. Hui, dan F. Menczer, "Scalable and Generalizable Social Bot Detection through Data Selection," dalam *Proceedings of the AAAI Conference on Artificial Intelligence*, Apr 2020, hlm. 1096–1103. doi: 10.1609/aaai.v34i01.5460.

[11] Institute for Computer Science, University of Göttingen, Germany dan J. Knauth, "Language-Agnostic Twitter Bot Detection," dalam *Proceedings - Natural Language Processing in a Deep Learning World*, Incoma Ltd., Shoumen, Bulgaria, Okt 2019, hlm. 550–558. doi: 10.26615/978-954-452-056-4_065.

[12] S. Feng, H. Wan, N. Wang, dan M. Luo, "BotRGCN: Twitter Bot Detection with Relational Graph Convolutional Networks," dalam *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Nov 2021, hlm. 236–239. doi: 10.1145/3487351.3488336.

[13] J. Gong, J. Peng, J. Qu, S. Du, dan K. Wang, "Enhancing Twitter Bot Detection via Multimodal Invariant Representations," 2024, doi: 10.48550/ARXIV.2408.03096.

[14] Ni Made Tara Okta Adriana, I Made Agus Dwi Suarjaya, dan Dwi Putra Githa, "Analisis Sentimen Publik Terhadap Aksi Demonstrasi di Indonesia Menggunakan Support Vector Machine Dan Random Forest," *Decode J. Pendidik. Teknol. Inf.*, vol. 3, no. 2, hlm. 257–267, Jun 2023, doi: 10.51454/decode.v3i2.187.

[15] B. Wu, L. Liu, Y. Yang, K. Zheng, dan X. Wang, "Using Improved Conditional Generative Adversarial Networks to Detect Social Bots on Twitter," *IEEE Access*, vol. 8, hlm. 36664–36680, 2020, doi: 10.1109/ACCESS.2020.2975630.

[16] J. Pfeffer, A. Mooseder, J. Lasser, L. Hammer, O. Stritzel, dan D. Garcia, "This Sample seems to be good enough! Assessing Coverage and Temporal Reliability of Twitter's Academic API," arXiv, 2022. doi: 10.48550/ARXIV.2204.02290.

[17] C. Barrie dan J. Ho, "academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint," *J. Open Source Softw.*, vol. 6, no. 62, hlm. 3272, Jun 2021, doi: 10.21105/joss.03272.

[18] R. M. Iqbal, Y. A. Sari, dan E. Santoso, "Pembentukan Daftar Stopword Goffman Transition Point dengan Pembobotan Emoji pada Analisis Sentimen di Twitter," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 9, no. 5, hlm. 1101–1108, Okt 2022, doi: 10.25126/jtiik.2022954706.

[19] W. N. Ibrahim, M. S. Anuar, A. Selamat, dan O. Krejcar, "BOTNET DETECTION USING INDEPENDENT COMPONENT ANALYSIS," *IIUM Eng. J.*, vol. 23, no. 1, hlm. 95–115, Jan 2022, doi: 10.31436/iiumej.v23i1.1789.

[20] M. Liandana, I. M. D. Susila, dan Y. P. Atmojo, "Pengenalan Aktivitas Manusia dengan Seleksi Fitur Analysis of Variance (ANOVA) dan Mutual Information (MI) pada Data Sensor Accelerometer Berbasis Machine Learning," *J. Sist. Dan Inform. JSI*, vol. 18, no. 2, Art. no. 2, Mei 2024, doi: 10.30864/jsi.v18i2.606.

[21] E. W. Pamungkas dan V. Patti, "Cross-domain and Cross-lingual Abusive Language Detection: A Hybrid Approach with Deep Learning and a Multilingual Lexicon," dalam *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, F. Alva-Manchego, E. Choi, dan D. Khashabi, Ed., Florence, Italy: Association for Computational Linguistics, Jul 2019, hlm. 363–370. doi: 10.18653/v1/P19-2051.

[22] Z. L. Thakker dan S. H. Buch, "Effect of Feature Scaling Pre-processing Techniques on Machine Learning Algorithms to Predict Particulate Matter Concentration for Gandhinagar, Gujarat, India," *Int. J. Sci. Res. Sci. Technol.*, hlm. 410–419, Feb 2024, doi: 10.32628/IJSRST52411150.

[23] A. Aguilera, P. Quinteros, I. Dongo, dan Y. Cardinale, "CrediBot: Applying Bot Detection for Credibility Analysis on Twitter," *IEEE Access*, vol. 11, hlm. 108365–108385, 2023, doi: 10.1109/ACCESS.2023.3320687.

[24] H. Zhou, "Research of Text Classification Based on TF-IDF and CNN-LSTM," *J. Phys. Conf. Ser.*, vol. 2171, no. 1, hlm. 012021, Jan 2022, doi: 10.1088/1742-6596/2171/1/012021.