 ISSN:2302-8149
e-ISSN:2540-9719

Pembobotan TF-IDF untuk Mendeteksi Akun Spammer di Twitter berdasarkan Tweet dan Representasi Retweet dari Tweet

TF-IDF Weighting to Detect Spammer Accounts on Twitter based on Tweets and Retweet Representation of Tweets

[1]Arif Mudi Priyatno*, [2] Lidya Ningsih
[1]Bisnis Digital, Fakultas Ekonomi dan Bisnis, Universitas Pahlawan Tuanku Tambusai
Jl. Tuanku Tambusai No.23, Kabupaten Kampar, Riau, Indonesia 28412
[2]Pascasarjana Ilmu Komputer, FMIPA, IPB University
Jl. Raya Dramaga, Babakan, Kec. Dramaga, Kota Bogor, Jawa Barat, Indonesia 16680
*e-mail: arifmudi11@gmail.com

(***received*:** 25 Maret2022, ***revised*:** 9 April 2022, ***accepted*:** 2 Juli 2022)

Abstrak

Twitter merupakan salah satu layanan media sosial yang sering digunakan (popular) sebagai sarana komunikasi antar pengguna. Kepopuleran twitter tersebut membuat spammer melakukan spam demi tujuan dan keuntungan pribadi. Bot spammer merupakan penyalahgunaan user pada media sosial Twitter. Spammer menyebarkan spam secara bertubi-tubi pada pengguna lain. Spam ini dilakukan bertujuan untuk mencapai trending topik. Aktivitas spam dilakukan dengan meniru pola perilaku pengguna asli agar tidak terdeteksi sebagai tindakan penyalahgunaan Twitter. Penelitian ini mengusulkan pembobotan TF-IDF untuk mendeteksi akun spammer di Twitter berdasarkan tweet dan representasi retweet dari tweet. Tujuan dari penelitian ini adalah untuk mendeteksi Bot Spammer atau Human menggunakan teknik klasifikasi meggunakan algoritma naive bayes. Hasil percobaan terbaik pada pembagian 70% data latih dan 30% data uji mendapatkan akurasi 92% dengan precision dan recall sebesar 100% dan 87.5%. Hal ini menunjukan berhasil mendeteksi akun bot spammer di Twitter.

Kata kunci: *Twitter, TF-IDF, Spam, Tweet, Retweet Representation*

Abstract

Twitter is a social media service that is often used (popular) as a means of communication between users. Twitter's popularity makes spammers spam for personal purposes and gains. Bot spammers are user abuse on Twitter social media. Spammers spread spam repeatedly to other users. This spam is done with the aim of achieving trending topics. Spam activity is carried out by imitating the behavior patterns of real users so that they are not detected as acts of Twitter abuse. in this paper proposed a TF-IDF weighting to detect spammer accounts on Twitter based on tweets and retweet representation of tweets. The purpose of this study is to detect Bot Spammers or Humans using a classification technique using the Naive Bayes algorithm. The best experimental results in the division of 70% training data and 30% test data obtained 92% accuracy with precision and recall of 100% and 87.5%, respectively. This shows that it has successfully detected spammer accounts on Twitter.

Keywords: *Twitter, TF-IDF, Spam, Tweet, Retweet Representation*

1 Pendahuluan

Informasi merupakan hal yang sangat dibutuhkan pada saat ini, terutama pada media digital atau dalam jejaring internet. Menurut hasil survey yang telah dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) tahun 2016, media sosial merupakan konten yang paling populer diakses dalam jaringan internet yakni 97,4% dari seluruh internet [1]. Saat ini, online social network (OSN) telah menjadi rutinitas sehari-hari masyarakat umum. setiap pengguna OSN dapat menghabiskan

banyak waktu mereka di OSN yang mereka minati untuk menyimpan dan menyebarkan informasi pribadi. Twitter merupakan salah satu OSN yang paling populer digunakan [2]. Twitter memungkinkan penggunanya untuk mengirimkan dan membaca pesan. Perbedaan Twitter yang paling menonjol dari OSN lain yakni Twitter membatasi posting atau tweets (pesan berbasis teks) hanya 280 karakter [3].

Popularitas Twitter tersebut menarik pengguna yang tidak bertanggung jawab untuk melakukan kegiatan spam [4]. Spam merupakan pesan yang dikirimkan oleh pengguna (*Spammer*) ke pengguna lain secara bertubi-tubi. Pesan yang dikirim tersebut tidak diinginkan oleh penerima pesan. Spammer merupakan pelaku kegiatan spam yang memiliki tujuan dan maksud tertentu. Spammer melakukan kegiatan dengan cara membuat banyak akun palsu. Spammer melakukan tweet konten yang sama serta mengandung mention dalam jumlah besar kepada pengguna lain secara acak. Spammer melakukan follow pengguna yang tidak terkait. Spammer mengirim pesan yang tidak diinginkan dan menyamarkan komponen yang berbahaya (misalnya spam menggunakan shortener URL untuk mengganti URL berbahaya) [5].

Spammer menggunakan tagar (hashtag) yang sedang popular agar dapat dilihat oleh pengguna lainnya [6]. Hal ini dilakukan agar pengguna lain mengikuti tweets tersebut tetapi dengan URL yang tidak diminta mengarah ke situs web yang tidak terkait. Jumlah pesan spam yang terus meningkat dapat memperburuk pengalaman pengguna twitter, karena spam dapat mencemari informasi pada twitter dan membuang sumber daya dari pengguna [7].

Bot merupakan suatu program otomatis yang tidak memerlukan operator manusia untuk melaksanakan tugasnya. Spam pada twitter Sebagian besar dihasilkan oleh bot. Bot dapat melakukan spam secara otomatis, sehingga sangat membantu spammer untuk menghasilkan banyak pesan spam di Twitter [8].

Akun twitter dapat dikatakan sebagai akun bot dapat dicirikan berdasarkan waktu pembuatan setiap tweet (umumnya terjadwal), hal ini dikarenakan akun bot adalah script program yang berjalan secara otomatis sesuai dengan perintah yang telah diberikan [9]. Ciri berdasarkan tweet yang dilakukan, tweet dilakukan secara berulang, sehingga memiliki kesamaan yang tinggi. Ciri berdasarkan hashtag, tweet dilakukan mengikuti trending topik (hashtag) yang sedang berlangsung. Hal ini dilakukan agar tweet terlihat oleh pengguna lainnya.

Penelitian ini mengusulkan pembobotan TF-IDF untuk mendeteksi akun spammer di Twitter berdasarkan tweet dan representasi retweet dari tweet. Tweet menjadi ciri yang digunakan untuk menunjukan kesamaan tweet yang telah dilakukan. Ciri tweet digabungan dengan ciri hashtag yang digunakan berdasarkan representasi retweet tersebut. Ciri-ciri tersebut diklasifikan menggunakan metode naïve bayes. Strategi ini diharapkan mampu untuk malakukan deteksi terhadap bot spammer pada twitter, sehingga dapat meningkatkan pengalaman pengguna serta mengurangi user bot spammer yang telah berkeliaran di twitter.

2 Tinjauan Literatur

Penelitian terdahulu yang telah dilakukan tentang bot spammer diantaranya dapat dilihat pada Tabel 1. Penelitian terdahulu memberikan penulis pengetahuan yang lebih dalam, sehingga penelitian terdahulu menjadi refrensi dalam melakukan penelitian.

Tabel 1. Penelitian Terkait

Nama Peneliti	Hasil Penelitian dan Perbedaan dengan Usulan
S. D. Priyani, E. Ripmiatin, dan S. Arifin	Paper [10] menggunakan fitur tweet dengan menghapus semua hal kecuali tweet dari user itu sendiri (baik itu url, mention, retweet, hashtag, dll). Fitur lainnya yang digunakan yaitu *time interval entropy* serta melakukan deteksi spammer menggunakan *Unigram Matching*. Perbedaan: penelitian usulan pada fitur tweet tidak menghapus retweet dan hashtag pada tweet tersebut, serta melakukan deteksi spammer dengan menggunakan *naïve bayes*.
I. Syafii, A. Setyanto, dan S. Raharjo	Paper [11] menggunakan fitur tweet dengan menghapus semua hal kecuali tweet dari user itu sendiri. Fitur tweet dilihat kesamaannya dengan menggunakan metode *Similarity Smith*

 ISSN:2302-8149
e-ISSN:2540-9719

	Waterman dan proses deteksi spammer menggunakan *k-Nearest Neighbour*. Perbedaan: penelitian usulan pada fitur tweet tetap mempertahankan hashtag dan retweet yang dilakukan. Pembobotan fitur tweet dan representasi retweet menggunakan TF-IDF, serta proses deteksi spammer menggunakan *naïve bayes classifier*.
H. Shukla, N. Jagtap, dan B. Patil	Paper [12] menggunakan metadata profile user seperti jumlah follower, jumlah follow dan lain sebagainya. Fitur metadata dilakukan proses seleksi fitur dengan menggunakan metode *Principal Component Analysis* (PCA). Proses deteksi spammer menggunakan 5 *machine learning* yaitu random forest, KNN, AdaBoost, logistic regression, dan naive Bayes. Perbedaan: Penelitian ususlan menggunakan fitur tweet dan representasi retweet. Proses klasifikasi menggunakan metode *naïve bayes classifier*.
H. Kurniawan	Penelitian [13] melakukan deteksi spammer dengan menggunakan metode *decision Tree*. Data yang digunakan yaitu verified-2019, botwiki-2019, cresci-rtbust-2019, dan botometer-feedback- 2019. Perbedaan: Penelitian usulan mengambil data secara langsung dengan menggunakan library python. Proses deteksi spammer dengan menggunakan *naïve bayes classifier*.

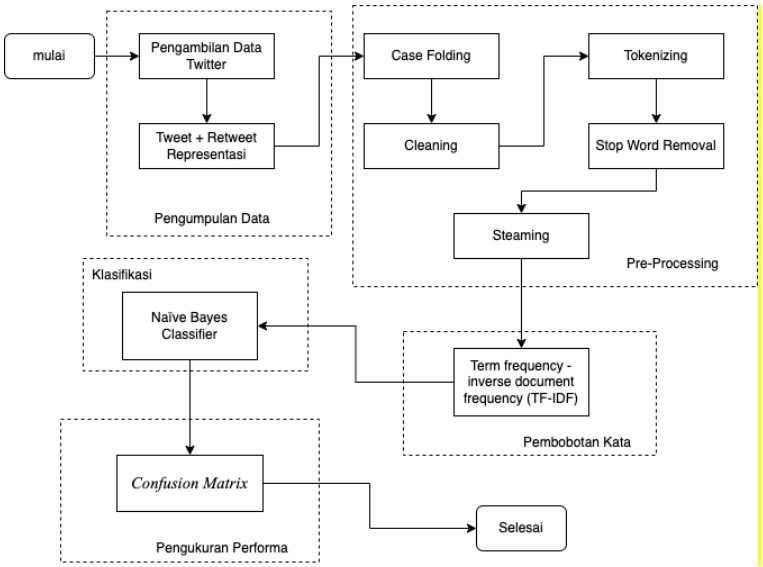
Berdasarkan penelitian diatas terfokus kepada fitur tweet, profile dan waktu interval melakukan tweet. Selain fitur tersebut, fitur hashtag dan fitur retweet dapat dimanfaatkan untuk melakukan deteksi akun spammer. sehingga penelitian ini menambahkan fitur representasi retweet dengan ciri hashtag. Fitur representasi retweet akan digunakan bersamaan dengan fitur tweet utama dari akun user twitter tersebut.

Penelitian ini mengusulkan pembobotan TF-IDF untuk mendeteksi akun spammer di Twitter berdasarkan tweet dan representasi retweet dari tweet. Tweet menjadi ciri yang digunakan untuk menunjukan kesamaan tweet yang telah dilakukan. Ciri tweet digabungan dengan ciri hashtag yang digunakan berdasarkan representasi retweet tersebut. Ciri-ciri tersebut diklasifikan menggunakan metode naïve bayes.

3 Metode Penelitian

Proses penelitian ini diantara yaitu pengambilan data, *pre-processing*, pembobotan menggunakan TF-IDF, klasifikasi menggunakan *Naïve Bayes Clasifier*, dan pengukuran performa menggunakan *Confusion Matrix*. Berikut ini bagan proses dari penelitian ini:

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 1. Tahapan Penelitian

3.1 Pengumpulan Data

Pengumpulan data dilakukan untuk mendapatkan data pada sosial media twitter. Proses pengambilan data menggunakan twint pada library python. Data yang didapatkan berupa data Bot Spammer dan Human. Data pada penelitian ini diambil dari 32 akun twitter. Data terdiri dari 18 akun bot spammer dan 14 akun human. Jumlah data tweet yang digunakan sebanyak 21.000 tweet.

3.2 Pre-processing

Setelah dilakukan pengumpulan data, Langkah selanjutnya adalah preprocessing data. Tahap ini dilakukan untuk mempersiapkan teks menjadi data yang dapat diolah pada tahap-tahap berikutnya. Hasil *pre-processing* mendapatkan teks yang bersih dari noise. Tahap preprocessing ini dilakukan karena data tweets yang tidak baku serta masih terdapat noise pada data teks tersebut. proses *pre-processing* juga dilakukan untuk mendapatkan parameter-parameter yang diinginkan pada penelitian. Gambar 1 menunjukkan proses dari klasifikasi bot spammer mulai dari proses pengumpulan data, preprocessing data, pembagian data latih dan data uji, pembobotan TF-IDF dan klasifikasi.

Berikut merupakan tahap pre-processing yang dilakukan pada penelitian klasifikasi bot spammer pada twitter [14].

1. Case Folding: tahap pertama yang dilakukan untuk menyeragamkan huruf dengan cara mengubah huruf besar (uppercase) pada tweet menjadi huruf kecil (lowercase).
2. Cleaning: tahap yang dilakukan untuk menghapus beberapa variable yang tidak digunakan pada proses penelitian seperti URL, tanda baca, emoticon, angka, dan lainnya yang dianggap tidak digunakan.
3. Tokenizing: Proses yang dilakukan untuk memisahkan kalimat menjadi token (kata per kata).
4. Stop Word Removal: merupakan tahap yang dilakukan untuk menghilangkan kata-kata umum atau yang tidak dibutuhkan dalam penelitian. Proses stop word removal dilakukan melalui pengecekan hasil parsing deskripsi. Jika kata-kata tersebut termasuk tidak penting (stoplist) maka akan di remove Contohnya dan, atau dll.
5. Steaming: merupakan proses yang digunakan untuk menggabungkan kata token kembali kedalam bentuk kata baku.

3.3 Pembobotan Kata

Term frequency-inverse document frequency (TF-IDF) merupakan teknik pembobotan berbasis statistik dengan menggabungkan dua konsep dalam perhitungannya, yaitu frekuensi kemunculan kata dan inverse. Pembobotan TF-IDF sering diterapkan pada permasalahan penggalian informasi. Ide dasar TF-IDF adalah memberikan bobot pada setiap kalimat, selanjutnya kalimat tersebut diurutkan berdasarkan bobot teratas dengan bobot paling besar akan dipilih sebagai hasil. Bobot kalimat diperoleh dari penjumlahan bobot term pada sebuah kalimat [15].

 ISSN:2302-8149
e-ISSN:2540-9719

Metode TF-IDF merupakan suatu cara untuk memberikan hubungan suatu kata (term) terhadap dokumen dengan menggabungkan dua konsep untuk perhitungan bobot yaitu frekuensi kemunculan kata tertentu dan invers frekuensi dokumen yang mengandung kata.

Rumus umum dari *term frequency* (TF) – *inverse document frequency* (IDF) dapat dilihat pada persamaan 1, 2, dan 3.

$$TF_{dt} = \frac{n_{d,t}}{\sum_k n_{d,t}} \quad (1)$$

$$IDF_t = \log (N/df_t) \quad (2)$$

$$W_{dt} = TF_{dt} * IDF_t \quad (3)$$

Dimana d merupakan dokumen ke-d, t merupakan kata ke-t dari kata kunci, W meruapkan bobot dokumen ke-d terhadap kata ke-t, TF merupakan banyaknya kata yang dicari pada sebuah dokumen, dan IDF merupakan Inverse Document Frequency. t merupakan total dokumen atau banyak dokumen yang mengandung kata yang dicari.

3.4 Klasifikasi

Naïve Bayes atau *Naïve Bayes Classifier* ditemukan oleh Thomas Bayes pada tahun 1770. Naïve Bayes merupakan salah satu teknik klasifikasi dengan metode probabilitas dan statistic untuk memprediksi peluang pada masa depan berdasarkan pengalaman sebelumnya, sehingga dikenal dengan istilah teorema Bayes. Teorema tersebut dikombinasikan dengan Naïve yang diasumsikan sebagai kondisi antar atribut saling bebas [16].

Berdasarkan pada asumsi penyederhanaan, nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, jika atribut diberikan nilai output, maka probabilitas mengamati secara bersama. Hal ini adalah produk dari probabilitas individu. Keuntungan penggunaan metode ini yaitu hanya membutuhkan jumlah data pelatihan yang kecil dalam proses pengklasifikasian. Metode ini bekerja jauh lebih baik dalam kebanyakan situasi pada data yang kompleks.

Teorema Bayes merupakan teorema yang mengacu pada konsep probabilitas bersyarat dengan notasi pada persamaan 4.

$$P\ (H|X) = \frac{\mathrm{P(X|H)}\ x\ P(H)}{\mathrm{P(X)}} \quad (4)$$

X meruapkan data dengan class yang belum diketahui. H merupakan Hipotesis data atau suatu classs pesifik. P(H|X) meruapakan probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas). P(H) merupakan probabilitas hipotesis H (prior probabilitas). P(X|H) merupakan probabilitas X berdasarkan kondisi pada hipotesis H. P(X) merupakan Probabilitas X.

3.5 Pengukuran Performa

Confusion Matrix merupakan metode untuk melihat keberhasilan algoritma yang digunakan. Confusion matrix dilakukan dengan mencari nilai precision, recall, accuracy dan F-Measure. Perhitungan confusion matrix dilakukan dengan menghitung true positive (TP), false positive (FP), true negative (TN), dan false negative (FN), hal ini sesuai dengan Table 2.

Tabel 2. *Confusion Matrix*

Kategori X	Predicted	
Aktual	TP	FP
	FN	TN

Sensitivity (*recall*) digunakan untuk membandingkan jumlah TP terhadap jumlah record yang positif, precision adalah perbandingan jumlah TN terhadap jumlah record yang negative. Persamaan menghitung akurasi, presisi, dan *recall* dapat dilihat pada persamaan 5, 6, dan 7 [17].

$$Akuasi = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}}\ x\ 100 \quad (5)$$

 ISSN:2302-8149
e-ISSN:2540-9719

$$Presisi = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+TN} \quad (7)$$

Dimana TP (*True Positif*) merupakan jumlah data positif yang terdeteksi benar. TN (*True Negatif*) merupakan jumlah data negative yang terdeteksi benar. FP (*False Positif*) meruapkan jumlah data positif yang terdeteksi salah. FN (*False Negatif*) merupakan jumlah data negative yang terdeteksi salah.

4 Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini merupakan data dari sosial media twitter. Data diambil menggunakan twint pada python. Data yang digunakan adalah data tweets dan representasi retweet dengan jumlah data 21000 dari 32 akun, yakni 18 akun bot spammer dan 14 akun human.

Data akun yang berjumlah 32 akun dilakukan pelabelan. Proses pelabelan akun bot spammer atau human dilakukan secara manual. Pelabelan dilakukan dengan memberi label kategori 1 sebagai bot spammer dan -1 sebagai human. Data tweets dan representasi retweet yang diambil pada penelitian menggunakan python kemudian disimpan kedalam file berformat CSV. Pembagian data latih dan data uji pada penelitian ini terbagi atas 4 macam yakni 50%, 60%, 70%, dan 80%. Tabel 3 merupakan tabel jumlah tweet hasil pembagian data latih dan data uji.

Tabel 3. Data Latih dan Data Uji

No	Persentase	Data Latih	Data Uji	Total
1	50 %	10500	10500	21000
2	60 %	12600	8400	21000
3	70 %	14700	6300	21000
4	80 %	16800	4200	21000

Data tweet dilakukan proses pre-processing dengan menghapus yang tidak dibutuhkan dan tetap mempertahankan tweet, retweet, dah hashtag. Tabel 4 merupakan hasil dari proses preprocessing.

Tabel 4. Hasil Preprocessing data

No	Tweet	Hasil Pre-processing
1	b'Pemprov DKI Jakarta mendukung penerapan Tilang Elektronik atau Electronic Traffic Law Enforcement (ETLE), yg terdiri dari Integrated Vehicle Registration and Identification System (IVRIS) dan SMS Info 8893.\n\nMari wujudkan budaya tertib berlalu lintas.\n\nhttps://t.co/Olqo2JzJax https://t.co/YHWSyZ86j8'	pemprov dki jakarta dukung terap tilang elektronik electronic traffic law enforcement etle yg diri integrated vehicle registration and identification system ivris sms info mari wujud budaya tertib lalu lintas
2	b'RT @JSCLounge: Menjelang musim hujan dan rawan banjir, website https://t.co/WIw2homOaQ membuat fitur baru CCTV lokasi rawan banjir di beber\xe2\x80\xa6'	jsclounge jelang musim hujan rawan banjir website buat fitur baru cctv lokasi rawan banjir beber
3	b'RT @DKIJakarta: Musim hujan telah datang. Selain waspada, ada beberapa hal yang perlu dipersiapkan untuk antisipasi sebelum terjadi banjir.\xe2\x80\xa6'	dkijakarta musim hujan datang waspada beberapa yang perlu siap antisipasi jadi banjir
4	b'rt @kemnakerri: #siaranperskemnaker \n\ndki jakarta raih 4 penghargaan ketenagakerjaan dalam integra	kemnakerri siaranperskemnakerd ki jakarta raih harga ketenagakerjaan integra

 ISSN:2302-8149
e-ISSN:2540-9719

	2018\nhttps://t.co/i4e5rkiver'

Data yang telah dilakukan preprocessing, selanjutnya data dilakukan pembobotan dengan menggunakan metode TF-IDF. Tabel 5 menunjukan hasil dari proses pembobotan menggunakan mrtode TF-IDF.

Tabel 5. hasil Proses Pembobotan menggunakan TF-IDF

No	Hasil Preprocessing	Hasil TF-IDF yang ada nilai	
1	pahlawan tanda jasa	(0, 9140)	0.6195557572919819
		(0, 20876)	0.5916699461547401
		(0, 15455)	0.5158268492659195
2	lingkung dekat benteng lawan keras perempuan	(0, 16138)	0.41080641284392816
		(0, 10402)	0.3516807263780032
		(0, 11607)	0.38179084815684905
		(0, 2310)	0.5008417908843724
		(0, 4378)	0.37930333911134895
		(0, 11889)	0.40875540799717086
3	foto karya pak ahmad muqowwam wakil ketua dpd ri enak kalo gak diupload	(0, 4952)	0.41044768875943016
		(0, 6403)	0.2371005881337948
		(0, 9793)	0.24271212147841392
		(0, 5602)	0.2750532853676139
		(0, 17952)	0.17379319199896467
		(0, 5124)	0.3362956097695575
		(0, 10546)	0.2278147245585022
		(0, 22711)	0.23067480515581876
		(0, 14028)	0.41044768875943016
		(0, 360)	0.2879949921064729
		(0, 15473)	0.1925944065050121
		(0, 9998)	0.2371005881337948
		(0, 6261)	0.21823286187693333
4	selamat cak alimasykurmusa sukses selalu semiga malam bisa hadir ikut	(0, 8145)	0.2286832556268507
		(0, 7143)	0.23968811898291956
		(0, 2724)	0.24213326792531772
		(0, 12406)	0.24563493280938045
		(0, 19140)	0.48727471027406133
		(0, 647)	0.48727471027406133
		(0, 3372)	0.3431563310033689
		(0, 19014)	0.26074130342064533
		(0, 20361)	0.26507268169161646
		(0, 19023)	0.20102096064877192

Data Hasil Proses TF-IDF seperti pada Table 5, maka selanjutnya klasifikasi antara bot spammer dan human. Tabel 6 merupakan hasil klasifikasi menggunakan algoritma *Naïve Bayes*.

Tabel 6. hasil Klasifikasi menggunakan *Naïve Bayes*

Kemungkinan Prediksi		Data	Prediksi
Bot	**Human**		
30.4	69.6	Human	Human
44.1	55.9	Human	Human
29.33	70.69	Human	Human
67.83	32.16	Human	Bot
52.2	47.8	Human	Bot

Hasil klasifikasi bot spammer ditentukan dengan melihat hasil probabilitas dapat dilihat pada Table 6, di mana hasil prediksi merupakan nilai probabilitas yang lebih tinggi. Table 6 pada data ke 5 merupakan kelas Human, namun karena nilai probabilitas bot spammer lebih tinggi maka hasil prediksinya adalah bot.

 ISSN:2302-8149
e-ISSN:2540-9719

Tabel 7. Hasil klasifikasi data 50:50

Label	Human	Bot	Presisi	Recall	F1 measure
Human	5	3	1	0.62	0.77
Bot	0	10	0.77	1	0.87

Hasil klasifikasi 50% data latih dan 50% data uji pada Table 7 didapatkan *recall* dan *precision* adalah 100% dan 77%. Pengujian algoritma *naïve bayes* pada pembagian data 50:50 ini menghasilkan akurasi sebesar 83%, hal ini dikarekan terdapat kesalahan 3 akun human yang terprediksi sebagai bot.

Tabel 8. Hasil klasifikasi data 40:60

Label	Human	Bot	Presisi	Recall	F1 measure
Human	4	3	1	57	0.73
Bot	0	8	0.73	1	0.84

Hasil klasifikasi pembagian 60% data latih dan 40% data uji pada Table 8 menghasilkan *precision* dan *recall* sebesar 73% dan 100%. Akurasi didapatkan sebesar 83% dengan kesalahan prediksi 3 akun human yang terprediksi sebagai bot.

Tabel 9. Hasil klasifikasi data 30:70

Label	Human	Bot	Presisi	Recall	F1 measure
Human	4	1	1	0.8	0.89
Bot	0	7	0.88	1	0.93

Hasil presisi dan recall klasifikasi pada pembagian data latih dan data uji 70%:30% pada Table 9 adalah 87.5% dan 100%. Akurasi didapatkan 92%, hal ini dikarenakan kesalahan prediksi terdapat pada 1 akun human yang terprediksi sebagai bot.

Tabel 10. Hasil klasifikasi data 20:80

Label	Human	Bot	Presisi	Recall	F1 measure
Human	2	2	1	0.5	0.67
Bot	0	5	0.71	1	0.83

Table 10 pembagian data latih 80% dan data uji 20% mendapatkan hasil precision dan recall sebesar 72% dan 100%. Hasil klasifikasi yang didapatkan berdasarkan akurasi adalah 78% dengan kesalahan prediksi pada 2 akun human yang terprediksi sebagai akun bot.

Pengujian yang dilakukan pada 4 jenis pembagian data, algoritma naïve bayes masih terdapat kesalahan dalam mengklasifikasikan, kesalahan tersebut terjadi pada akun human yang diprediksi sebagai akun bot. Hasil klasifikasi yang terbaik didapatkan pada pembagian data 70% data latih dan 30% data uji dengan nilai akurasi, recall, dan presisi sebesar 92%, 87.5%, dan 100%.

5 Kesimpulan

Penelitian ini menggunakan algoritma naïve bayes untuk melakukan klasifikasi akun bot dan human dengan pembobotan TF-IDF. Data yang digunakan adalah data tweet dan representasi retweet yang berjumlah 2100, terdiri dari 32 akun dengan 18 akun bot dan 14 akun human. Pembobotan TF-IDF dan klasifikasi menggunakan Algoritma naïve bayes berhasil dalam klasifikasi bot spammer pada twitter. Hal ini dikarena memiliki hasil akurasi, recall, dan presisi sebesar 92%, 87.5%, dan 100%. Penggunaan fitur tweet dan representasi retweet pada penelitian ini mampu mendapatkan hasil yang baik. Berdasarkan hasil penelitian yang didapat maka dapat disimpulkan pembobotan TF-IDF dan klasifikasi algoritma naive bayes mampu mendeteksi akun spammer pada twitter.

Referensi

[1] A. M. Priyatno, M. M. Muttaqi, F. Syuhada, and A. Z. Arifin, “Deteksi Bot Spammer Twitter Berbasis Time Interval Entropy dan Global Vectors for Word Representations Tweet’s

Hashtag," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 5, no. 1, p. 37, Jan. 2019, doi: 10.26594/register.v5i1.1382.

[2] R. Gilmary, A. Venkatesan, and G. Vaiyapuri, "Detection of Automated Behavior on Twitter Through Approximate Entropy and Sample Entropy," *Pers. Ubiquitous Comput.*, Sep. 2021, doi: 10.1007/s00779-021-01647-9.

[3] T. Ruan, Q. Kong, S. K. McBride, A. Sethjiwala, and Q. Lv, "Cross-Platform Analysis of Public Responses to the 2019 Ridgecrest Earthquake Sequence on Twitter and Reddit," *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, 2022, doi: 10.1038/s41598-022-05359-9.

[4] S. Bazzaz Abkenar, E. Mahdipour, S. M. Jameii, and M. Haghi Kashani, "A Hybrid Classification Method for Twitter Spam Detection based on Differential Evolution and Random Forest," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 21, pp. 1–20, 2021, doi: 10.1002/cpe.6381.

[5] M. Heidari, J. H. J. Jones, and O. Uzuner, "Online User Profiling to Detect Social Bots on Twitter," *arXiv*, Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.05966

[6] A. S. Alhassun and M. A. Rassam, "A Combined Text-Based and Metadata-Based Deep-Learning Framework for the Detection of Spam Accounts on the Social Media Platform Twitter," *Processes*, vol. 10, no. 3, p. 439, 2022, doi: 10.3390/pr10030439.

[7] A. M. Priyatno, "Spammer Detection based on Account, Tweet, and Community Activity on Twitter," *J. Ilmu Komput. dan Inf.*, vol. 13, no. 2, pp. 97–107, Jul. 2020, doi: 10.21609/jiki.v13i2.871.

[8] L. D. Samper-Escalante, O. Loyola-González, R. Monroy, and M. A. Medina-Pérez, "Bot Datasets on Twitter: Analysis and Challenges," *Appl. Sci.*, vol. 11, no. 9, pp. 1–25, 2021, doi: 10.3390/app11094105.

[9] Y. Wu, Y. Fang, S. Shang, J. Jin, L. Wei, and H. Wang, "A Novel Framework for Detecting Social Bots with Deep Neural Networks and Active Learning," *Knowledge-Based Syst.*, vol. 211, p. 106525, 2021, doi: 10.1016/j.knosys.2020.106525.

[10] S. D. Priyani, E. Ripmiatin, and S. Arifin, "Implementation of Cosine Similarity and Time Interval Entropy Method to Identify Bot Spammer Account on Twitter," *ITSMART J. Teknol. dan Inf.*, vol. 6, no. 2, pp. 51–57, 2017, doi: 10.20961/itsmart.v6i2.14320.

[11] I. Syafii, A. Setyanto, and S. Raharjo, "Deteksi Bot Spammer pada Twitter menggunakan Smith Waterman Similarity dan Time Interval Entropy," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 3, pp. 633–638, 2018, doi: 10.29207/resti.v2i3.549.

[12] H. Shukla, N. Jagtap, and B. Patil, "Enhanced Twitter Bot Detection using Ensemble Machine Learning," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Jan. 2021, pp. 930–936. doi: 10.1109/ICICT50816.2021.9358734.

[13] H. Kurniawan, "Deteksi Twitter Bot menggunakan Klasifikasi Decision Tree," *J. Sustain. J. Has. Penelit. dan Ind. Terap.*, vol. 9, no. 1, pp. 31–37, May 2020, doi: 10.31629/sustainable.v9i1.2347.

[14] Oryza Habibie Rahman, Gunawan Abdillah, and Agus Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter menggunakan Support Vector Machine," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 17–23, 2021, doi: 10.29207/resti.v5i1.2700.

[15] N. S. Mohd Nafis and S. Awang, "An Enhanced Hybrid Feature Selection Technique using Term Frequency-Inverse Document Frequency and Support Vector Machine-Recursive Feature Elimination for Sentiment Classification," *IEEE Access*, vol. 9, no. Ml, pp. 52177–52192, 2021, doi: 10.1109/ACCESS.2021.3069001.

[16] E. Sutoyo, A. P. Rifai, A. Risnumawan, and M. Saputra, "A Comparison of Text Weighting Schemes on Sentiment Analysis of Government Policies: A Case Study Of Replacement Of National Examinations," *Multimed. Tools Appl.*, vol. 81, no. 5, pp. 6413–6431, 2022, doi: 10.1007/s11042-022-11900-9.

[17] I. M. De Diego, A. R. Redondo, R. R. Fernández, J. Navarro, and J. M. Moguerza, "General Performance Score for classification Problems," *Appl. Intell.*, no. January 2021, pp. 1–15, Jan. 2022, doi: 10.1007/s10489-021-03041-7.