

Klasifikasi Hashtag Buzzer/Bot Menggunakan Algoritma Random Forest dengan Atribut Komunitas untuk Mengurangi Disinformasi Pada Twitter

1st Fadhiil Nugroho Yudianto
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia

fadhiilnugroho@students.telkomuniversity.ac.id

2nd Yuliant Sibaroni
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia

yuliant@telkomuniversity.ac.id

Abstrak

Saat ini persebaran informasi terjadi secara cepat pada sosial media. Namun, tidak semua informasi yang beredar pada sosial media belum tentu merupakan informasi yang benar. Twitter sebagai salah satu media sosial terbesar didunia juga mengalami masalah yang sama. Banyak sekali informasi dan juga hashtag yang tersebar pada twitter merupakan informasi dan hashtag yang disebarluaskan oleh buzzer. Dalam tugas akhir ini, penulis membangun sebuah sistem klasifikasi yang dapat membedakan antara sebuah hashtag natural dan hashtag yang dibuat oleh buzzer pada sosial media twitter. Penulis memakai algoritma random forest dengan memanfaatkan atribut komunitas untuk melakukan klasifikasi hashtag pada sistem yang dibuat. Penulis juga akan mencari tahu apa saja atribut yang paling berpengaruh dalam proses klasifikasi tersebut. Pada penelitian ini, didapatkan bahwa sistem mampu mendeteksi hashtag buzzer dengan tingkat akurasi sebesar 98% dan tingkat f-measure sebesar 98% berdasarkan dataset yang telah penulis dapatkan. Didapatkan juga bahwa atribut atribut yang berpengaruh dalam proses klasifikasi antara lain adalah atribut seperti `unique_user_pertweet` dan `q1_rasio`

Kata kunci : buzzer, random forest, hashtag, trending topic, twitter

Abstract

Right now, the circulation of information happened so fast in the social media space. However, not all information on the social media is true and valid. Twitter as one of the biggest social media also had the same problem as stated before. A lot of hashtags on twitter are coming and spreading from buzzer. In this paper, we are going to make a classification system that can differentiate between a hashtag that come from buzzer and a hashtag that come naturally on twitter. We will use random forest algorithm with community feature from previous study as the foundation. We also going to see what is the most influential attribute that affecting the classification process. In this study, the system that we created can distinguished the hashtag that appear from buzzer with 98% accuracy and 98% f-measure based on the dataset that we gather. We also get the most influential

attribute in this classification are `unique_user_pertweet` and `q1_rasio`

Keywords: buzzer, random forest, hashtag, trending topic, twitter

I. PENDAHULUAN

Pada tahun 2017, 143 juta dari 262 juta rakyat Indonesia telah terkoneksi dengan internet. Diantara 143 juta tersebut, 87.13% telah dapat mengakses sosial media dan menjadikannya sebuah gaya hidup[1]. Sosial media sendiri adalah salah satu medium yang efektif dalam menyebarkan informasi secara luas. Twitter sebagai salah satu sosial media terbesar di Indonesia, memiliki jumlah pengguna indonesia sebanyak 29.4 juta pada tahun 2012[2]. Sektor *trending topic* pada Twitter merupakan salah satu akses pengguna Twitter dalam mencari tahu informasi tentang apa saja kejadian kejadian yang sedang terjadi di sekitar kita[3]. Namun belakangan ini, sektor *trending topic* pada Twitter dipenuhi oleh hashtag hashtag yang tidak relevan dan juga sering kali hashtag tersebut berisi informasi yang tidak benar. Hal tersebut dikarenakan hashtag hashtag tersebut berasal dari buzzer.

Di Indonesia sendiri, efek dari buzzer telah dirasakan pada event pemilihan gubernur Jakarta di tahun 2012[1]. Sejak saat itu, fenomena buzzer di Indonesia terus mengalami peningkatan tiap tahunnya. Puncaknya bisa dibilang terjadi pada pemilihan presiden pada tahun 2019[4]. Buzzer dipakai oleh politisi untuk menyerang lawan politiknya dengan cara menyebarkan informasi palsu dan menyebarkan informasi pribadi lawan politiknya [5][1]. Efek dari buzzer di Indonesia telah mencapai tahap yang sangat buruk hingga dapat memecah belah masyarakat [6]. Penelitian yang dilakukan oleh [7] menunjukkan bahwa peran buzzer di media sosial dalam kegiatan Pemilu dan Pilkada telah mencapai tahap yang parah. Kemunculan buzzer dalam dunia

perpolitikan juga menunjukkan tren yang selalu naik tiap tahunnya[8].

Penelitian yang dilakukan oleh [9] menunjukkan bahwa kita dapat membedakan sebuah hashtag menjadi sebuah event dan meme menggunakan algoritma random forest dengan *accuracy* sistem yang mencapai 90%. Namun penelitian tersebut tidak membahas sama sekali tentang faktor bot dan buzzer dalam penyebaran informasi hashtag. Berdasarkan penelitian yang dilakukan oleh [10] menunjukkan bahwa random forest dapat mendeteksi akun penyebar disinformasi pada pemilihan presiden Prancis pada tahun 2017. Penelitian yang dilakukan oleh [11] menunjukkan bahwa kita dapat mendeteksi bot pada twitter dengan melihat bagaimana similaritas tweet yang dihasilkan oleh sebuah akun. Sedangkan penelitian yang dilakukan oleh [1] menunjukkan bahwa kita dapat menjangkau akun buzzer pada sosial media berdasarkan beberapa atribut seperti aktivitas akun dan hashtag postingan akun tersebut.

Berdasarkan riset-riset sebelumnya, dapat dilihat bahwa algoritma random forest dan penggunaan atribut komunitas banyak digunakan pada penelitian-penelitian sebelumnya. Metode klasifikasi random forest telah digunakan pada penelitian [9][10][1] sedangkan atribut komunitas telah digunakan pada penelitian [9][10]. Atribut komunitas sendiri adalah atribut yang dapat menggambarkan sebuah hashtag dari segi komunitas seperti jumlah user yang menyebarkan hashtag dan rata-rata jumlah follower[9]. Oleh karena itu, pada Tugas akhir ini, penulis mencoba untuk mengklasifikasikan hashtag pada platform twitter dengan algoritma *random forest* dengan memanfaatkan atribut komunitas. Penulis menggabungkan berbagai temuan yang ada pada penelitian sebelumnya dan menggabungkan beberapa langkah yang ada didalam penelitian tersebut dengan harapan dapat meningkatkan tingkat keberhasilan penelitian ini.

Tujuan penulis melakukan penelitian ini adalah

- a. mengetahui komposisi data train yang optimal dalam pembuatan model klasifikasi
- b. mengetahui apakah atribut komunitas berpengaruh dalam proses klasifikasi

Batasan dari penelitian ini adalah data yang dipakai terbatas pada hashtag yang muncul pada *trending topic* yang muncul pada daerah Indonesia. Dataset trending topic yang dipakai hanya terbatas pada hashtag dan tidak termasuk kata kata biasa.

II. KAJIAN TEORI

Twitter adalah sebuah microblogging service dimana penggunaannya dapat menggunakannya secara gratis[2]. Twitter merupakan penyedia jasa microblogging terbesar di dunia mengalahkan jasa lain seperti Jaiku dan Qaiku[12]. Pada tahun 2008 Twitter mengalami pertumbuhan yang sangat cepat

dengan memiliki 500.000 pengguna aktif, naik 1.382% dibanding tahun sebelumnya. Pada tahun 2014 sendiri, Twitter telah memiliki 284 juta pengguna aktif tiap bulannya[2]. Di Indonesia sendiri, jumlah pengguna Twitter telah mencapai angka 50 juta pengguna[13]. Namun sebanding dengan jumlah user yang ada, jumlah disinformasi yang ada pada twitter juga ikut bertambah tiap tahunnya.

Upaya untuk mengurangi disinformasi pada sosial media twitter telah dimulai sejak tahun 2010[14] dengan adanya deteksi spam pada twitter dengan menggunakan algoritma SVM. Selain itu penelitian [11] juga dapat mendeteksi akun bot dengan analisa *entropy component*, *Account Properties Component*, dan *Spam Detection Component*. Saat ini deteksi akun pada twitter telah dapat mendeteksi akun buzzer pada twitter[15][16]. Hal ini dikarenakan perilaku akun buzzer sendiri memiliki kemiripan dengan akun bot[5]. Penelitian tersebut tentu sangat berguna dikarenakan buzzer pada sosial media twitter sering kali menyebarkan fitnah[7] dan politik negatif [17]. Efek dari buzzer di Indonesia sendiri telah mencapai tahap yang sangat buruk hingga dapat memecah belah masyarakat [6].

Istilah buzzer berasal dari ranah pemasaran yang awalnya berupa istilah buzz marketing atau teknik pemasaran barang atau jasa untuk menghasilkan bisnis dengan pergerakan informasi dari mulut ke mulut[17]. Namun belakangan ini, istilah buzzer sendiri telah mengalami pergeseran arti di sosial media. Buzzer sekarang memiliki arti menjadi akun media sosial baik yang dikelola individu maupun perusahaan dimana akun tersebut memiliki *follower* dalam jumlah banyak dan turut melakukan kampanye politik dengan menyebar berbagai berita hoax serta ujaran kebencian[6]. Selain itu menurut penelitian yang dilakukan oleh Bradshaw[5], buzzer di Indonesia sendiri telah dikategorikan sebagai *low cyber troop army*. Hal ini berarti bahwa buzzer di Indonesia sering melakukan disinformasi dan menyebarkan disinformasi tersebut secara luas.

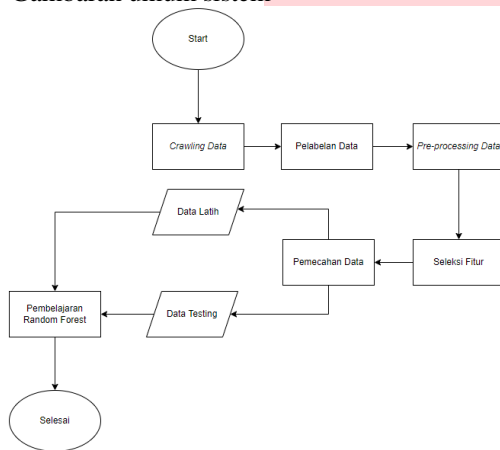
Klasifikasi hashtag sendiri sudah mulai dicoba pada tahun 2014. Penelitian pertama [18] mencoba untuk mengklasifikasi hashtag menggunakan metode *naïve bayes* dengan pembobotan TF-IDF. Penelitian tersebut mampu mengklasifikasi hashtag, namun hasil *accuracy* dan *precision* yang didapatkan masih kurang baik. Penelitian kedua [9] justru mendapatkan hasil yang lebih baik. Penelitian tersebut mampu mendapatkan nilai *precision*, *accuracy*, dan *recall* yang cukup tinggi untuk klasifikasi hashtag menggunakan metode random forest dibandingkan algoritma algoritma lain seperti SVM, *naïve bayes*, dan k-NN. Penelitian [9] memanfaatkan atribut komunitas dalam membedakan hashtag meme dan hashtag event dalam penelitian yang dilakukannya. Atribut komunitas sendiri adalah atribut yang dapat menggambarkan sebuah hashtag dari segi komunitas

seperti jumlah user yang menyebarkan hashtag, rata rata jumlah follower, dll.

Berdasarkan beberapa penelitian sebelumnya, penulis mencoba mengembangkan sebuah sistem yang dapat mengklasifikasi sebuah hashtag secara natural atau sebuah hashtag yang diciptakan oleh buzzer pada platform twitter. Sistem ini menggunakan penelitian [9] sebagai acuan utama dalam pemilihan metode klasifikasi hashtag. Selain itu, penulis juga memanfaatkan fitur atribut komunitas yang telah digunakan pada dua penelitian sebelumnya oleh [9] dan [10]. Penulis juga mengambil beberapa atribut khusus yang ada pada penelitian lainnya seperti atribut *account reputation* pada penelitian [11] untuk meningkatkan tingkat *accuracy* sistem.

III. METODE

A. Gambaran umum sistem



GAMBAR 1 flowchart sistem

Sistem yang dibangun adalah sebuah sistem klasifikasi hashtag yang dapat membedakan antara hashtag yang muncul secara natural dan hashtag yang muncul dengan bantuan buzzer. Sistem tersebut memiliki diagram flow seperti gambar diatas.

B. Crawling Data

Data yang dipakai dalam penelitian ini adalah data tweet yang berasal dari Twitter. Proses pengumpulan data atau *crawling* data dilakukan dengan menggunakan aplikasi python dengan memanfaatkan library *Tweepy*. *Tweepy* sendiri memanfaatkan Twitter Public API untuk mengambil data pada Twitter. Data yang diambil adalah data-data tweet yang berkaitan dengan hashtag yang sedang trending pada suatu hari. Berikut adalah beberapa contoh dari hashtag yang diambil

TABEL 1 contoh hashtag yang diambil

Hashtag	
#17an	#Oposisiplangaplongo
#Adzanbukanmainan	#UCLfinal
#Hajat4nAsetku	#Whatif
#KRINanggala402	#Whatsapp
#ReshufflePresiden	#Ygletlisadoherwork

Dapat dilihat pada TABEL 1 bahwa data yang diambil cukup beragam. Mulai dari topik olahraga (#UCLfinal), bencana(#KRINanggala402), K-pop(#Ygletlisadoherwork), politik(#ReshufflePresiden), hingga event internasional(#Whatsapp). Dataset yang beragam ini penting agar sistem yang dibuat mampu mengenali berbagai macam buzzer yang ada pada sosial media.

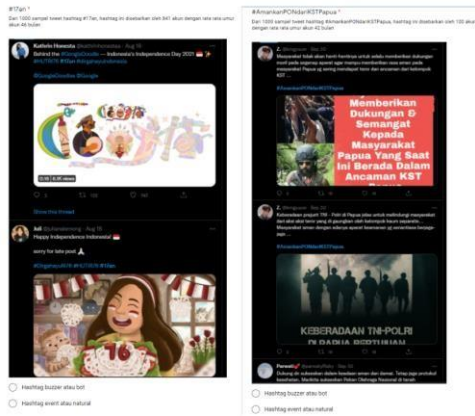
TABEL 2 atribut yang diambil ketika crawling

Atribut	
Created	Following
Id_Tweet	Reply_Tweet_Id
User_Id	Reply_User_Name
Username	Reply_User_Id
User_Screen_Name	Entities
Status_Count	Url
Verified	User_Created_At
Follower	Jum_Mention

Pada TABEL 2 dapat dilihat bahwa tidak semua atribut yang kita perlukan dapat diambil langsung dari twitter. Atribut atribut seperti follower, following, dan username memang dapat kita ambil langsung dari twitter. Namun untuk mengaplikasikan penelitian [11] pada penelitian ini, kita perlu membuat beberapa atribut baru seperti atribut *Account_Reputation*, dan atribut *Umur_Akun*. Penulis melakukan crawling data sebanyak 202 hashtag dimana tiap hashtag 1000 tweet.

Data yang dipakai dalam penelitian ini diambil pada periode bulan April 2021 sampai dengan bulan September 2021. Data diambil dengan cara mengamati apa saja hashtag yang sedang menjadi trending topik di suatu hari pada akun pribadi penulis dan menurut website www.trends24.in/indonesia/. Selanjutnya penulis mengambil 1000 tweet dari satu hashtag tersebut. Kemudian data satu hashtag tersebut disimpan kedalam satu file dimana artinya penulis mendapatkan total 202 file data untuk tiap hashtag yang berbeda.

C. Pelabelan Data



GAMBAR 2 contoh pertanyaan survey

Pelabelan data dilakukan menggunakan survey secara anonim yang dilaksanakan menggunakan google form. Dalam survey tersebut, responden diberikan sebuah screenshot yang berisi tweet dari hashtag tersebut. Lalu, responden mengkategorikan hashtag tersebut menjadi hashtag event atau hashtag buzzer. Hasil survey tersebut kemudian dilihat modus atau hasil terbanyak untuk setiap hashtag dalam melakukan pelabelan data. Dalam proses pelabelan ini, didapatkan bahwa 63% hashtag dilabelkan oleh ketiga responden secara sama dan 27% hashtag dilabelkan secara majority vote Link survey yang digunakan penulis dapat dilihat pada <https://forms.gle/RfFXuiDSaKEaycMA6>.

D. Preprocessing Data

Data yang telah diambil selama proses pengumpulan data pertama tama perlu digabungkan terlebih dahulu. Hal ini dikarenakan data dari hasil pengumpulan tersebut tersebar kedalam 202 file yang berbeda dimana satu file tersebut berisi data untuk satu data hashtag yang berisi 1000 data tweet. Cara penggabungan datanya juga tidak perlu mengambil seluruh nilai data tersebut. Penulis melakukan kompresi data dengan mengambil beberapa nilai statistik dasar dari beberapa atribut terutama dari atribut komunitas. Atribut komunitas sendiri adalah atribut yang melambangkan seberapa pentingnya sebuah akun jika dilihat dari segi komunitas jaringan[9]. Beberapa contoh atribut komunitas yang diolah pada tahap ini adalah atribut follower, following, dan status_count Oleh karena itu, penulis banyak mengambil nilai statistik dari atribut follower, following, status, dan beberapa atribut turunan lainnya. Nilai statistik yang diambil sendiri adalah jumlah rata rata dari 1000 tweet, nilai kuartil tiga dari 1000 tweet, nilai kuartil satu dari 1000 tweet, dan lain lain. berikut adalah contoh ilustrasi dari satu data hashtag sebelum dilakukan kompresi data dan sesudah dilakukan kompresi data

Tabel 3 contoh data sebelum di preprocessing

Text	User_id	Follower	Following	Status_count	Rasio	Reputat ion	Umur_aku n
Iya be git u #u p	12 34 56 7	98	12	1239	7, 61 5	0,89	0.9
Ha ha ha #u p	12 34 56 2	31 21 2	20	3123	1. 48 6, 3	0,99	9
Be git u ... #u p	71 29 12 8	68 12	420 1	1912 1	1, 62 1	0,61 85	1.1
Ck ck ck #u p	01 23 81 2	59 1	344	2301 2	2, 41 65 9	0,63 2	2.1
Up ga n #u p	20 23 14 1	62	23	8890	2, 62 5	0,72 9	0.2

Pada Tabel 3, data awal yang ada masih berisi data individual tweet yang berasal dari beragam user. Namun dikarenakan satu data hashtag terdapat 1000 baris data, kita perlu mencari cara untuk mengkompresi data tersebut tanpa menghilangkan informasi berharga yang ada didalamnya.

TABEL 4 contoh data setelah dilakukan preprocessing

Hasht ag	Q1_ follo wer	Q2_ follo wer	Q3_ follo wer	Me an_ fo llo we r	Q1_ ra sio	Q2_ ra sio	. . .
#u p	80	591	3	19.012	2,1 23	2,6 25

Pada TABEL 4, penulis mengambil nilai rata rata, median, kuartil satu, kuartil tiga, dan jangkauan antar kuartil dari atribut komunitas penting dan beberapa rangkuman data lainnya untuk tetap menjaga informasi dari atribut tersebut.

TABEL 5 atribut komunitas dataset setelah dilakukan preprocessing

Atribut	
Nama	Mean_Reputation
Unique_User_Pertweet	JAK_Reputation
Url_Pertweet	Q1_Rasio
Reply_Pertweet	Q2_Rasio
Reply_User_Pertweet	Q3_Rasio
Rt_Pertweet	Mean_Rasio
Unique_Rt_Pertweet	JAK_Rasio
Verified_User_Pertweet	Q1_Umur_Akun
Duration	Q2_Umur_Akun
Q1_Reputation	Q3_Umur_Akun
Q2_Reputation	Mean_Umur_Akun
Q3_Reputation	JAK_Umur_Akun

Pada TABEL 5, penulis mengambil atribut yang berisi informasi komunitas seperti jumlah follower, jumlah following dan rasio tweet per user untuk diolah pada langkah berikutnya. Daftar atribut komunitas yang lengkap dapat dilihat pada Lampiran 1.

Langkah selanjutnya yang ada dalam preprocessing ini adalah scalling. Data scalling adalah proses merubah rentang data menjadi suatu rentang tertentu tanpa merubah isi data itu sendiri. Scalling dilakukan agar rentang data menjadi seragam antara suatu atribut dengan atribut yang lainnya[21]. Hal ini dapat membantu mengurangi waktu eksekusi dalam melatih model dikarenakan angka yang diproses menjadi lebih kecil dibandingkan sebelum dilakukan scalling. Berikut adalah rumus dari scalling

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

D.1

- x_{scaled} = hasil scalling
- x = nilai atribut x
- x_{max} = nilai maksimal atribut x
- x_{min} = nilai minimal atribut x

E. Seleksi Fitur

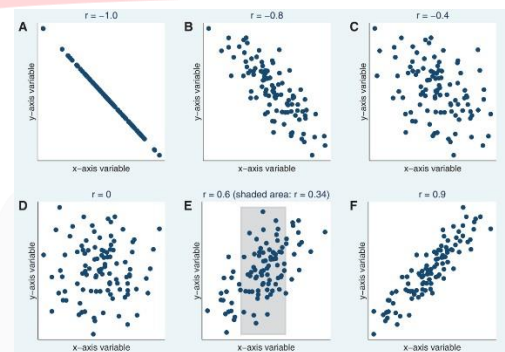
Seleksi fitur menggunakan metode pearson correlation untuk mencari fitur atau atribut mana yang memiliki korelasi yang tinggi terhadap label data. Penggunaan pearson correlation pada langkah ini telah dilakukan sebelumnya pada[22]. Seleksi fitur dilakukan agar model dapat dilatih dengan lebih cepat dan mengurangi kompleksitas dari model itu sendiri[23]. Penulis membandingkan hasil klasifikasi yang didapat dengan data yang memiliki jumlah atribut berbeda berdasarkan nilai korelasi tertinggi dari atributnya. Berikut adalah rumus dari pearson correlation

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

E.1

- r = nilai pearson correlation
- x = nilai atribut x
- y = nilai atribut y
- n = jumlah data berpasangan

Pada penelitian ini, nilai x dan y yang ada pada akan berisi nilai dari atribut atribut yang ada pada TABEL 5 dan Lampiran 1. Nilai y akan berisi nilai dari atribut kelas dan nilai x akan berisi atribut atribut lainnya. Jika nilai atribut y ikut naik dengan naiknya nilai atribut x, maka nilai pearson correlation antara kedua atribut tersebut adalah 1. Namun, jika nilai atribut y turun dengan naiknya nilai atribut x, maka nilai pearson correlation antara kedua atribut tersebut akan menjadi -1. Jika sebaran data kedua atribut tersebut acak, maka nilai pearson correlation antara kedua atribut tersebut mendekati 0. Penulis akan memakai atribut atribut yang nilai pearson correlationnya mendekati 1 dan -1 dalam langkah seleksi fitur ini.



GAMBAR 3 nilai pearson correlation berdasarkan sebaran data atribut x dan y

TABEL 6 interpretasi nilai pearson correlation

Nilai korelasi	Interpretasi
0.00 – 0.09	Tidak ada korelasi
0.10 – 0.39	Korelasi lemah
0.40 – 0.69	Korelasi sedang
0.70 – 0.89	Korelasi kuat
0.90 – 1.00	Korelasi sangat kuat

Pada TABEL 6, didapatkan beberapa interpretasi dari nilai pearson correlation [24]. Didapatkan bahwa korelasi dengan nilai dibawah 0.4 merupakan nilai korelasi yang lemah. Untuk korelasi dengan nilai diantara 0.4 sampai dengan 0.7 merupakan korelasi sedang. Dan terakhir, edangkan nilai korelasi diatas 0.7 merupakan nilai korelasi yang kuat. Selain itu, nilai pada table tersebut juga berlaku untuk angka yang negatif. Artinya jika suatu atribut memiliki nilai korelasi sebesar -0.7, maka atribut tersebut memiliki nilai korelasi yang kuat [24]

F. Pembelajaran Random Forest

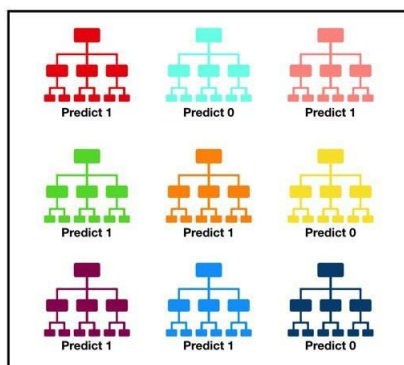
Random forest merupakan salah satu algoritma machine learning yang masuk ke dalam kategori tree-based ensemble learning[25]. Ensemble learning

adalah suatu sistem dimana sistem membuat beberapa pohon keputusan dan melihat hasil mayoritas untuk menentukan hasil akhir dari sistem[25][26]. Random forest pertama kali diperkenalkan oleh Leo Breiman yang terinspirasi oleh penelitian yang dilakukan oleh Amit dan Geman[26]. Walaupun begitu, Random forest sendiri merupakan perkembangan lanjut dari metode bagging dan dibuat sebagai kompetisi dari metode boosting[27].

TABEL 7 keputusan yang dibuat berdasarkan ensemble learning

	Pohon 1	Pohon 2	Pohon 3	Pohon 4	Pohon 5	Pohon 6	Hasil akhir
Kasus 1	1	1	0	0	1	1	1
Kasus 2	0	0	1	0	0	1	0
Kasus 3	1	1	0	1	1	1	1
Kasus 4	0	0	1	0	0	1	0

Dapat dilihat pada TABEL 7, sistem yang dibuat berdasarkan ensemble learning akan melihat keenam pohon keputusan yang ada di dalam sistem dalam mengambil keputusan akhir. Jika ada lebih dari empat pohon keputusan yang menganggap ya, maka keputusan akhir adalah ya. Sebaliknya, jika ada lebih dari empat pohon keputusan yang menganggap tidak, maka keputusan akhir adalah tidak.



Tally: Six 1s and Three 0s
Prediction: 1

GAMBAR 4 ilustrasi random forest [28]

Sebagai bagian dari ensemble learning, random forest juga bekerja dengan membuat berbagai macam pohon keputusan dan melihat hasil mayoritas vote untuk melakukan klasifikasi[29]. Jumlah dari pohon keputusan yang ada pada random forest juga beragam berdasarkan desain atau oleh pembuat sistem[29].

Tiap pohon memakai subset atribut yang dipilih secara random dan berbeda dengan pohon yang lainnya. Pemilihan atribut tersebut akan memakai metode random *subspace* yang telah dikembangkan oleh Ho[30] Selain itu, tiap pohon juga menggunakan subset data training yang berbeda juga antara satu pohon dengan pohon lainnya. Pohon yang dibuat pada algoritma random forest mengikuti aturan *binary recursive partitioning*. Hal ini berarti cabang dari tiap *node* yang ada pada pohon di random forest hanya memiliki dua cabang dan dipisah berdasarkan *impurity* dari sebuah *node*[26]. *Impurity* sendiri memiliki arti seberapa besar variasi yang ada pada sebuah atribut yang menjadi rule pada *node* tersebut. Pada penelitian ini sendiri penulis membuat sebuah random forest yang terdiri dari 250 pohon keputusan dan memiliki kedalaman pohon maksimal sebesar 7 keputusan.

G. Evaluasi Sistem

Dalam pembuatan suatu sistem, tentu dibutuhkan sebuah metode evaluasi untuk mengukur seberapa bagus performa dari sebuah sistem. Dalam masalah sistem klasifikasi, pengukuran *accuracy*, *precision*, dan *recall*, *f-measure* merupakan salah satu kriteria dalam mengukur performa sistem[32]. Setelah sistem berhasil melakukan klasifikasi dataset menggunakan data training, selanjutnya kita perlu mengukur tingkat performa dari sistem tersebut untuk tahap evaluasi. Pada tahap ini, kita membuat sebuah confusion matrix dengan mencari nilai true positive (TP), true negative (TN), false positive (FP) dan false negative (FN) menggunakan data testing.

TABEL 8 confusion matrix

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	TP (True Positive)	FN (False Negative)
Negatif	FP (False Positive)	TN (True Negative)

Dapat dilihat pada TABEL 8, TP berisi nilai berapa banyak hashtag dengan label buzzer yang berhasil diklasifikasikan oleh sistem menjadi hashtag buzzer. TN berisi nilai berapa banyak hashtag natural yang berhasil diklasifikasikan oleh sistem menjadi hashtag buzzer. FP berisi nilai seberapa banyak hashtag natural yang diklasifikasikan menjadi hashtag buzzer oleh sistem. FN berisi nilai seberapa banyak hashtag buzzer yang diklasifikasikan oleh sistem menjadi hashtag natural.

Untuk mengukur sebuah performa dari suatu sistem klasifikasi, penulis perlu mencari nilai dari TP,

TN, FN, dan TN. Setelah mendapatkan nilai tersebut, penulis dapat mengukur nilai *accuracy*, *precision*, dan *recall*, *f-measure* berdasarkan beberapa rumus berikut.

a. *Accuracy*

Accuracy menggambarkan berapa banyak jumlah item yang diklasifikasi secara benar

$$Accuracy = \frac{TP}{TP+TN+FP+FN} \tag{G.1}$$

b. *Precision*

Precision menggambarkan berapa banyak jumlah item yang diklasifikasi secara benar dibagi oleh total data yang diklasifikasi positif.

$$Precision = \frac{TP}{TP+FP} \tag{G.2}$$

c. *Recall*

Recall menggambarkan berapa banyak jumlah item yang diklasifikasi secara benar dibagi oleh total data asli yang positif.

$$Recall = \frac{TP}{TP+FN} \tag{G.3}$$

d. *F-Measure* atau *F1-score*

F-Measure atau *F1-score* menggambarkan perbandingan *precision* dan *recall* yang telah dibobotkan.

$$F - Measure = \frac{2 \times precision \times recall}{precision + recall} \tag{G.4}$$

IV. HASIL DAN PEMBAHASAN

A. *Experimen*

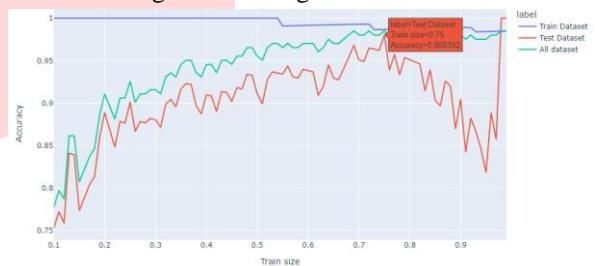
Pada penelitian ini, penulis melakukan beberapa eksperimen untuk mencari tahu bagaimana caranya mendapatkan komposisi data train yang optimal untuk model dan apakah atribut komunitas berpengaruh dalam proses klasifikasi. Beberapa eksperimen yang dilakukan untuk mengetahui kedua hal tersebut tersebut adalah

- a. Membandingkan model yang dilatih dengan jumlah data training yang berbeda.
- b. Membandingkan model yang dilatih dengan dataset penuh dan dataset yang telah dilakukan seleksi fitur.
- c. Membandingkan model yang dilatih dengan dataset yang telah dilakukan scaling dan yang tidak.
- d. Melihat fitur yang berpengaruh terhadap proses klasifikasi.

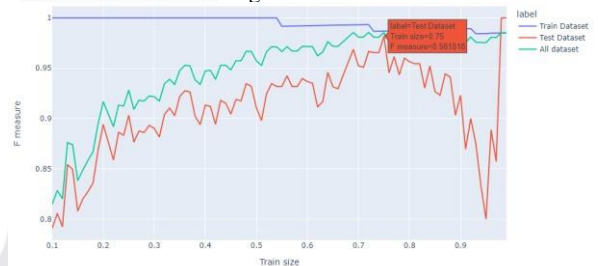
B. Hasil pengujian

- a. Hasil pengujian model yang dilatih dengan jumlah data training yang berbeda

Didapatkan bahwa model yang dilatih dengan persentase data training lebih besar mendapatkan hasil *accuracy* dan *f1-score* yang lebih besar juga. Sebagai contoh, pada model yang dilatih dengan persentase data training sebesar 40%, *accuracy* yang didapatkan sebesar 90%. Bandingkan dengan model yang dilatih hanya dengan persentase data training sebesar 20% yang mendapatkan nilai *accuracy* sebesar 85%. Namun, kenaikan hasil *accuracy* dan *f1-score* mulai berhenti pada tingkat 75% dan mulai menurun setelahnya. Hal ini disebabkan oleh overfitting yang terjadi pada model yang dilatih dengan data training diatas 75%. Oleh karena itu penulis akan memakai jumlah data training sebesar 75% untuk mencegah overfitting.



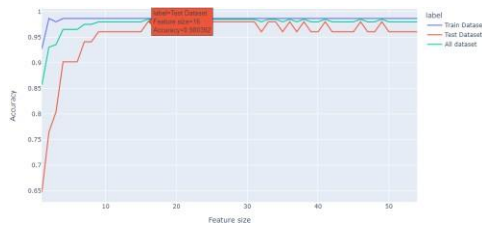
GAMBAR 5 *accuracy* model dengan berbagai ukuran data training



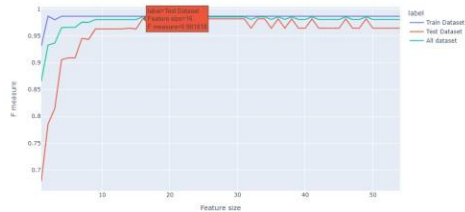
GAMBAR 6 *f measure* model dengan berbagai ukuran data training

- b. Hasil pengujian model yang dilatih dengan dataset penuh dan dataset yang telah dilakukan seleksi fitur

Didapatkan bahwa hasil *accuracy* dan *f measure* yang didapatkan oleh model yang dibuat dengan seleksi fitur akan terus naik dan mulai mengalami stagnasi setelah nilai fitur yang dipakai diatas 16. Hal ini membuktikan bahwa langkah seleksi fitur mampu mempertahankan dan meningkatkan nilai *accuracy* dari sebuah model dengan mengurangi fitur atau atribut yang dipakai dalam melatih model[23].



GAMBAR 7 accuracy model dengan berbagai fitur



GAMBAR 8 f measure model dengan berbagai fitur

Tabel 9 fitur yang dipilih dari proses seleksi fitur

Fitur	
Q1_Rasio	Q3_Follower_Rate
Unique_User_Pertweet	Q2_Rasio
JAK_Status	Unique_Rt_Pertweet
Q3_Status	Jak_Following_Rate
Url_Pertweet	Q3_Umur_Akun
Q2_Follower	Q3_Following_Rate
JAK_Reputation	JAK_Umur_Akun
JAK_Follower_Rate	Q1_Reputation

Dapat dilihat pada Tabel 9 bahwa fitur-fitur yang diambil berdasarkan proses seleksi fitur banyak diisi oleh atribut komunitas. Fitur-fitur seperti Q3_status, JAK status, dan Q2_Follower adalah contoh dari atribut komunitas tersebut.

c. Hasil pengujian model yang dilatih dengan dataset yang telah dilakukan scaling

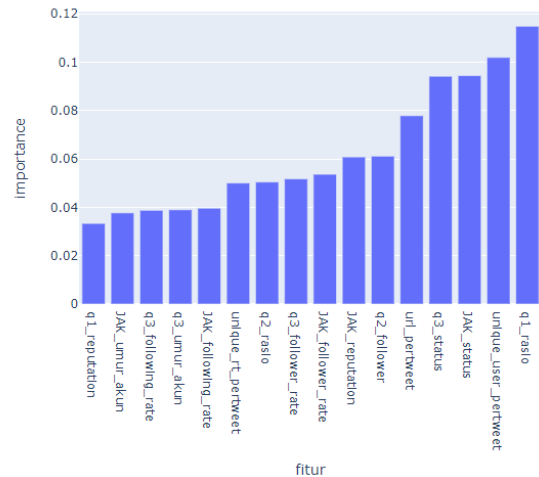
TABEL 10 perbandingan performa scaling dan tanpa scaling

	Scaling	Tanpa Scaling
Waktu Eksekusi (detik)	0.47	0.47
Accuracy	98%	98%
Recall	100%	100%
Precision	96.4%	96.4%
F measure	98.1%	98.1%

Dapat dilihat pada TABEL 10 bahwa tidak perbedaan yang signifikan antara model yang dilatih dengan dataset yang telah dilakukan scaling dan yang tidak. Kedua model mendapatkan hasil accuracy, precision, recall, dan f measure yang identik serta waktu eksekusi yang mirip. Kedua model tersebut dilatih dengan parameter yang sama.

d. Hasil pengujian fitur yang berpengaruh teradap proses klasifikasi

Didapatkan bahwa sistem dapat menunjukkan apa saja atribut yang mempengaruhi hasil klasifikasi menggunakan metode feature importance. Feature importance sendiri adalah metode untuk mengetahui atribut atribut apa saja yang mempengaruhi hasil klasifikasi berdasarkan nilai gini impurity dari tiap atribut tersebut.



GAMBAR 9 atribut paling berpengaruh

C. Evaluasi dan analisis

Berdasarkan hasil eksperimen yang dilakukan; didapatkan bahwa sistem dapat membedakan antara hashtag natural dengan hashtag buzzer dengan baik. Hasil evaluasi dilakukan dengan membagi data menjadi data latih dan data testing dengan rincian jumlah data latih sebanyak 75% (151 baris) dan data testing sebanyak 25% (51 baris) berdasarkan hasil eksperimen pertama. Random forest classifier juga telah di set dengan nilai maksimal depth tree sebesar 7 dan pembobotan antara data hashtag natural dengan data hashtag buzzer sebesar 1:2. Hal ini dilakukan agar sistem lebih memprioritaskan pengklasifikasian data hashtag buzzer dibanding data hashtag natural.

TABEL 11 confusion matrix model terbaik

Actual Class	Predicted class	
	Natural	Buzzer
Natural	23	1
Buzzer	0	27

Berdasarkan Tabel 11, dari 24 data hashtag natural yang ada pada dataset testing, 23 diantaranya berhasil diklasifikasikan secara benar oleh sistem dan 1 gagal diklasifikasikan secara benar. Dan dari 27 data hashtag buzzer yang ada pada dataset testing, semuanya berhasil diklasifikasikan secara benar oleh sistem. Berdasarkan rumus accuracy yang ada pada persamaan (3.3), maka didapatkan nilai accuracy sebesar 98%. Hasil yang didapatkan lebih baik dibanding dengan hasil penelitian [9] yang berhasil

mendapatkan nilai *accuracy* sebesar 89% pada sistem klasifikasi hashtag .

V. KESIMPULAN

Dari hasil penelitian dan pembahasan yang telah dijabarkan pada bab-bab sebelumnya, maka dapat ditarik kesimpulan bahwa sistem yang dibuat mampu membedakan antara hashtag buzzer dan hashtag non-buzzer dengan nilai *accuracy* yang cukup tinggi. Selain itu, sistem juga dapat mencari atribut yang paling berpengaruh dalam klasifikasi hashtag buzzer dan non-buzzer ini. Atribut atribut tersebut antara lain adalah Unique_User_Pertweet dan url_pertweet dari sebuah hashtag. Didapatkan juga bahwa tidak ada perbedaan antara model yang dilatih dengan data scaling dan model yang dilatih tanpa scaling. Hal ini memungkinkan penulis untuk mencoba untuk melakukan klasifikasi menggunakan model tersebut terhadap data baru yang belum pernah penulis lihat sebelumnya.

REFERENSI

- [1] A. J. Panatra, F. B. Chandra, W. Darmawan, H. L. H. S. Warnars, W. H. Utomo, and T. Matsuo, "Buzzer Detection to Maintain Information Neutrality in 2019 Indonesia Presidential Election," *Proc. - 2019 8th Int. Congr. Adv. Appl. Informatics, IIAI-AAI 2019*, pp. 873–876, 2019, doi: 10.1109/IIAI-AAI.2019.00177.
- [2] K. M. Carley, M. Malik, M. Kowalchuck, J. Pfeffer, and P. Landwehr, "Twitter Usage in Indonesia," *SSRN Electron. J.*, 2018, doi: 10.2139/ssrn.2720332.
- [3] M. Cheong, "What are you Tweeting about?": A survey of Trending Topics within Twitter," *Search*, pp. 1–12, 2009, [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.158.4679>.
- [4] A. Suciati, A. Wibisono, and P. Mursanto, "Twitter Buzzer Detection for Indonesian Presidential Election," *ICICOS 2019 - 3rd Int. Conf. Informatics Comput. Sci. Accel. Informatics Comput. Res. Smarter Soc. Era Ind. 4.0, Proc.*, 2019, doi: 10.1109/ICICoS48119.2019.8982529.
- [5] S. Bradshaw and P. N. Howard, "The Global Disinformation Order 2019 Global Inventory of Organised Social Media Manipulation," *Univ. Oxford*, p. 25, 2019, [Online]. Available: <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf>.
- [6] S. Sugiono, "Fenomena Industri Buzzer Di Indonesia: Sebuah Kajian Ekonomi Politik Media," *Commun. J. Ilmu Komun.*, vol. 4, no. 1, pp. 47–66, 2020, doi: 10.15575/cjik.v4i1.7250.
- [7] C. Juditha, "Buzzer di Media Sosial Pada Pilkada dan Pemilu Indonesia Buzzer in Social Media in Local Elections and Indonesian Elections," pp. 199–212, 2019.
- [8] B. Arianto, "Salah Kaprah Ihtwal Buzzer: Analisis Percakapan Warganet di Media Sosial," *J. Ilm. Ilmu Pemerintah.*, vol. 5, no. 1, pp. 1–20, 2020, doi: 10.14710/jiip.v5i1.7287.
- [9] D. Kotsakos, P. Sakkos, I. Katakis, and D. Gunopulos, "#tag: Meme or event?," *ASONAM 2014 - Proc. 2014 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min.*, no. Asonam, pp. 391–394, 2014, doi: 10.1109/ASONAM.2014.6921615.
- [10] E. Ferrara, "Disinformation and social bot operations in the run up to the 2017 French presidential election," *First Monday*, vol. 22, no. 8, 2017, doi: 10.5210/fm.v22i8.8005.
- [11] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?," *IEEE Trans. Dependable Secur. Comput.*, vol. 9, no. 6, pp. 811–824, 2012, doi: 10.1109/TDSC.2012.75.
- [12] T. A. Small, "WHAT THE HASHTAG? Tamara A. Small WHAT THE HASHTAG? A content analysis of Canadian politics on Twitter," vol. 4462, no. May, 2011, doi: 10.1080/1369118X.2011.554572.
- [13] C. Juditha, B. Besar, P. Komunikasi, and B. Makassar, "FENOMENA TRENDING TOPIC DI TWITTER: ANALISIS WACANA TWIT # SAVEHAJILULUNG TRENDING TOPIC PHENOMENON ON TWITTER: DISCOURSE ANALYSIS OF TWEET # SAVEHAJILULUNG," vol. 16, no. 2, pp. 138–154, 2015.
- [14] G. Magno and T. Rodrigues, "Benevenuto2010," *Ceas*, no. Seventh annual Collaboration, Electronic messaging, AntiAbuse and Spam Conference, p. 10, 2010.
- [15] M. Ibrahim, O. Abdillah, A. F. Wicaksono, and M. Adriani, "Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in a Twitter Nation," *Proc. - 15th IEEE Int. Conf. Data Min. Work. ICDMW 2015*, pp. 1348–1353, 2016, doi: 10.1109/ICDMW.2015.113.
- [16] M. T. Juzar and S. Akbar, "Buzzer Detection on Twitter Using Modified Eigenvector Centrality," *Proc. 2018 5th Int. Conf. Data Softw. Eng. ICODSE 2018*, pp. 1–5, 2018, doi: 10.1109/ICODSE.2018.8705788.
- [17] R. Mustika, "Pergeseran Peran Buzzer Ke Dunia Politik Di Media Sosial," *Diakom J. Media dan Komun.*, vol. 2, no. 2, pp. 144–151, 2019, doi: 10.17933/diakom.v2i2.60.

- [18] M. Jeon, S. Jun, and E. Hwang, "Hashtag Recommendation Based on User Tweet and Hashtag Classification on Twitter," pp. 325–336, 2014, doi: 10.1007/978-3-319-11538-2.
- [19] G. Valkanas, A. Saravanou, and D. Gunopulos, "A faceted crawler for the Twitter service," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8787, pp. 178–188, 2014, doi: 10.1007/978-3-319-11746-1_13.
- [20] J. E. Sembodo, E. B. Setiawan, and A. Baizal, "Data Crawling Otomatis pada Twitter," no. October 2018, pp. 10–16, 2016, doi: 10.21108/INDOSC.2016.111.
- [21] A. Ambarwari, Q. Jafar Adrian, and Y. Herdiyeni, "Analysis of the Effect of Data Scaling on the Performance of the Machine Learning Algorithm for Plant Identification," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 117–122, 2020, doi: 10.29207/resti.v4i1.1517.
- [22] I. M. Nasir *et al.*, "Pearson correlation-based feature selection for document classification using balanced training," *Sensors (Switzerland)*, vol. 20, no. 23, pp. 1–18, 2020, doi: 10.3390/s20236793.
- [23] L. Yu and H. Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy," vol. 5, pp. 1205–1224, 2004.
- [24] P. Schober and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesth. Analg.*, vol. 126, no. 5, pp. 1763–1768, May 2018, doi: 10.1213/ANE.0000000000002864.
- [25] A. Cutler, D. R. Cutler, and J. R. Stevens, "Ensemble Machine Learning," *Ensemble Mach. Learn.*, no. February 2014, 2012, doi: 10.1007/978-1-4419-9326-7.
- [26] L. E. O. Breiman, "Random Forests," pp. 5–32, 2001.
- [27] L. E. O. Breiman, "Bagging Predictors," vol. 140, pp. 123–140, 1996.
- [28] T. Yui, "Understanding Random Forest. How the Algorithm Works and Why it Is So Effective," *Towards Data Science*, 2019. <https://towardsdatascience.com/understanding-random-forest-58381e0602d2> (accessed Jan. 31, 2022).
- [29] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest?," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7376 LNAI, pp. 154–168, 2012, doi: 10.1007/978-3-642-31537-4_13.
- [30] T. K. Ho, "The Random Subspace Method for Constructing Decision Forests," vol. 20, no. 8, pp. 832–844, 1998.
- [31] L. Rokach and O. Maimon, "Chapter 9," no. August 2015, 2005, doi: 10.1007/0-387-25465-X.
- [32] S. Visa, B. Ramsay, A. Ralescu, and E. Van Der Knaap, "Confusion matrix-based feature selection," in *CEUR Workshop Proceedings*, 2011, vol. 710, pp. 120–127.