

Analisis Pengguna Berpengaruh di Media Sosial Twitter Menggunakan Metode TwitterRank

Elsha Nursyifa Anindya¹, Warih Maharani², Jati Hiliamsyah Husen³

^{1,2,3} Fakultas Informatika, Universitas Telkom, Bandung

¹elshansyfa@students.telkomuniversity.ac.id, ²wmaharani@telkomuniversity.ac.id,

³jatihusen@telkomuniversity.ac.id

Abstrak

Twitter adalah jejaring sosial yang berkembang pesat saat ini, karena pengguna dapat berinteraksi dengan pengguna lainnya (disebut user) dimanapun dan kapanpun. Twitter memberi akses pada para pengguna untuk mengirimkan pesan singkat yang terdiri dari maksimal 140 karakter (disebut tweet). Melalui tweet ini, pengguna dapat berinteraksi dan menyebarkan informasi kepada pengguna lainnya. Penelitian ini menggunakan metode TwitterRank yang bertujuan untuk melakukan klasifikasi dan mengevaluasi pengguna yang paling berpengaruh pada penyebaran informasi berdasarkan hashtag #KartiniDay. Metode ini dihitung berdasarkan tweet yang dibuat pengguna dengan tiga indikator, yaitu Followers Influence (FI), Retweet Influence (RI), dan Favorite Influence (FVI). Kemudian, pengguna diklasifikasikan kedalam beberapa kelompok pengguna berdasarkan nilai indikator yang didapat, yaitu kelompok pengguna highly influential user, influential user, potential influential user, dan not influential user. Hasil pengukuran pengaruh pengguna dengan menggunakan perhitungan akurasi yaitu 0.614 atau 61.4%, perhitungan presisi yaitu 0.266 atau 26.6%, perhitungan recall yaitu 0.0878 atau 8.78%, dan perhitungan f-measure yaitu 0.071 atau 7.1%.

Kata kunci : TwitterRank, hashtag, peringkat, klasifikasi.

Abstract

Twitter is a social network that is growing rapidly nowadays, because users can interact with other users (called users) wherever and whenever. Twitter gives users access to send short messages consisting of up to 140 characters (called tweets). Through this tweet, users can interact and spread information to other users. This research uses TwitterRank method which aims to identify the ranking of the most influential users on information spreading based on the hashtag #KartiniDay. This method is calculated based on tweets made by the user with three indicators, namely Followers Influence (FI), Retweet Influence (RI), and Favorite Influence (FVI). Then, users are classified into several groups of users based on the obtained indicator values, namely the group of highly influential users, influential users, potential influential users, and not influential users. The results of measuring the influence of the user by using an accuracy calculation are 0.614 or 61.4%, the precision calculation is 0.266 or 26.6%, the recall calculation is 0.0878 or 8.78%, and for the f-measure calculation is 0.071 or 7.1%.

Keywords: TwitterRank, hashtag, rank, classification.

1. Pendahuluan

Latar Belakang

Twitter adalah jejaring sosial yang berkembang pesat saat ini, karena pengguna dapat berinteraksi dengan pengguna lainnya (disebut user) dimanapun dan kapanpun. Selain itu, pengguna juga dapat menyebarkan dan memperoleh informasi dengan cepat melalui sebuah tweet, yang dapat berupa tulisan, gambar, video, dan yang lainnya. Penyebaran informasi pada jejaring sosial twitter dapat dimanfaatkan dalam berbagai bidang. Contohnya, twitter dapat digunakan sebagai wadah untuk melakukan kampanye politik. Selain itu, banyak juga pengguna twitter yang memiliki pengaruh besar dalam penyebaran informasi mengenai suatu produk, suatu kejadian, atau informasi lainnya. Maka dari itu, perlu diketahui pengguna mana saja yang memiliki pengaruh besar dalam penyebaran suatu informasi, agar informasi yang disebarkan lebih tepat sasaran, dalam arti informasi yang disebarkan sampai ke orang yang tepat.

Pada penelitian terdahulu, pengguna yang paling berpengaruh dalam penyebaran informasi kepada pengguna lain adalah pengguna yang memiliki pengikut (disebut followers) dalam jumlah besar [11]. Namun, dengan hanya melihat jumlah followers bukan merupakan indikator yang baik, karena diduga sebagian besar followers mengikuti

pengguna tertentu hanya untuk berbasa-basi dan tidak mempengaruhi persebaran informasi apapun [13]. Informasi pengguna yang terdapat pada jejaring sosial twitter juga tidak dapat diandalkan sepenuhnya, biodata twitter atau foto palsu misalnya [4]. Selain itu, tingkat kepopuleran pengguna juga berbeda-beda. Pengguna tertentu dapat menjadi sangat populer dalam waktu yang sangat singkat, bahkan dapat tetap populer dalam jangka waktu yang lama, ataupun sebaliknya.

Maka, penelitian ini menggunakan metode TwitterRank yang bertujuan untuk melakukan klasifikasi dan mengevaluasi pengguna yang paling berpengaruh berdasarkan suatu hashtag. Hari Kartini merupakan hari bersejarah yang harus tetap diingat dan disebarluaskan di berbagai kalangan pengguna twitter, maka hashtag #KartiniDay digunakan pada penelitian ini untuk mengetahui pengguna mana saja yang paling berpengaruh untuk menyebarkan informasi yang berkaitan dengan Hari Kartini. Penelitian ini menggunakan metode TwitterRank karena metode ini dianggap sebagai metode yang paling masuk akal dibandingkan dengan metode lainnya, karena perhitungannya dilakukan menggunakan beberapa indikator penting [8]. Metode ini dihitung berdasarkan tweet yang dibuat pengguna dengan tiga indikator, yaitu followers, retweet, dan favorite tweet. Kemudian, pengguna diklasifikasikan ke dalam beberapa kelompok pengguna berdasarkan nilai indikator yang didapat.

Topik dan Batasannya

Berdasarkan latar belakang yang telah diuraikan di atas, perumusan masalah dari penelitian ini adalah apa saja faktor yang mempengaruhi peringkat pengguna twitter dalam penggunaan hashtag #KartiniDay menggunakan metode TwitterRank, dan bagaimana pengaruh hasil perhitungan indikator followers, retweet, dan favorite tweet terhadap klasifikasi pengguna menggunakan metode TwitterRank. Alasan penggunaan hashtag #KartiniDay dalam penelitian ini adalah karena pengambilan data tweet yang bertepatan dengan peringatan Hari Kartini yang menjadi salah satu trending topic di media sosial Twitter.

Adapun batasan masalah dari penelitian ini adalah menggunakan data tweet para pengguna twitter yang menggunakan hashtag #KartiniDay dalam tweet mereka dengan twitterscraper tanpa harus menggunakan Twitter API. Pengambilan data tweet dilakukan real-time tiga hari setelah hari kartini, yaitu pada tanggal 24 April 2020 yang menghasilkan sejumlah 121 tweet sejak tanggal 21 April 2020 hingga 24 April 2020 beserta metadatanya, yaitu username, jumlah tweet, jumlah retweet, jumlah followers, jumlah favorite tweet, dan waktu tweet tersebut dibagikan.

Tujuan

Tujuan dari penelitian ini adalah melakukan klasifikasi dan mengevaluasi pengguna yang paling berpengaruh mengenai suatu hashtag pada interval waktu tertentu. Pengaruh penyebaran informasi para pengguna twitter ini dihitung menggunakan metode TwitterRank, dengan perhitungan indikator-indikator yang telah dijelaskan sebelumnya. Hasil dari perhitungan indikator tersebut akan dijadikan acuan untuk mengklasifikasi kelompok pengguna dari yang sangat berpengaruh (highly influential user), berpengaruh (influential user), berpotensi mempengaruhi (potentially influential user), dan tidak berpengaruh (not influential user).

Organisasi Tulisan

Pada bagian selanjutnya, akan dijelaskan mengenai penelitian atau studi yang terkait dengan penelitian ini, beserta penulis, judul, dan hasil penelitian tersebut. Pada bagian tiga, dijelaskan sistem yang digunakan dan dibangun pada penelitian ini beserta teori-teori terkait. Bagian empat akan menjelaskan hasil evaluasi dari sistem yang digunakan. Sistem evaluasi yang digunakan pada penelitian ini adalah dengan perhitungan akurasi, presisi, recall, dan f-measure. Lalu, bagian terakhir adalah kesimpulan dan saran dari hasil penelitian ini.

2. Landasan Teori

2.1 Penelitian Terkait

Penelitian ini dilakukan dengan melihat beberapa penelitian terkait yang pernah dilakukan untuk dijadikan sebagai bahan perbandingan. Berdasarkan penelitian/studi terkait yang tercantum pada tabel, pada tahun 2010, Meeyoung Cha et al. [3] melakukan penelitian yang bertujuan untuk mengetahui pengaruh pengguna twitter terhadap suatu topik dengan mengukur peringkat pengguna menggunakan tiga indikator, yaitu perhitungan in-degree, perhitungan jumlah retweet, dan perhitungan jumlah mention dengan rumusnya masing-masing.

Selanjutnya, pada tahun 2011, Isabel Anger dan Christian Kittl [1] melakukan penelitian serupa yang bertujuan untuk mengetahui pengaruh pengguna twitter terhadap suatu topik dengan mengukur peringkat pengguna, dengan indikator yang berbeda dengan penelitian yang dilakukan oleh Meeyoung Cha et al. [3]. Indikator yang digunakan pada penelitian ini adalah menghitung rasio following/followers, rasio retweet, rasio mention, dan Klout score dari tiap pengguna.

Lalu, pada tahun 2017, Xingjun Ma et al. [6] melakukan penelitian yang juga bertujuan untuk mengetahui pengaruh pengguna twitter terhadap suatu topik dengan mengukur peringkat pengguna dengan menggunakan beberapa fitur yang telah dipilih dari 10 fitur yang ada. Fitur yang digunakan yaitu user attributes (retweet, favorite, dll), temporal activity (jumlah tweet dalam satu jam, aktifitas pengguna, dll), dan topical similarity (kesamaan topik antara dua pengguna).

Penelitian-penelitian terkait menyebutkan bahwa metode pada penelitian mereka cukup baik dibanding dengan metode yang digunakan pada penelitian lain. Namun, pada penelitian terkait belum ada perhitungan evaluasi yang dapat membuktikan bahwa metode dan perhitungan yang digunakan memang merupakan metode yang baik.

2.2 Twitter

Twitter adalah salah satu media sosial yang memberikan akses kepada para penggunanya untuk menyebarkan dan membagikan informasi melalui sebuah tulisan (tweet) mengenai pikiran dan pengalaman mereka [2]. Tweet yang dibagikan berupa sebuah pesan singkat yang hanya dapat terdiri dari maksimal 140 karakter. Tweet tersebut juga dapat berupa sebuah gambar atau foto, video, dan yang lainnya.

Pada jejaring sosial twitter, terdapat beberapa istilah yang digunakan. Istilah tersebut dapat dilihat dan dipahami pada tabel sebagai berikut [7].

Tabel 1. Istilah yang digunakan pada media sosial Twitter

No.	Istilah	Keterangan
1	User atau pengguna	Para pemilik akun dan pengguna twitter.
2	Tweet atau cuitan	Pesan singkat berupa tulisan, gambar, atau video yang dibagikan oleh para pengguna twitter.
3	Timeline atau linimasa	Halaman utama twitter setelah pengguna masuk ke akun twitter mereka masing-masing. Berisi tweet dari followers pengguna tersebut.
4	Follow atau mengikuti	Mengikuti pengguna lain, yang berarti tweet para pengguna tersebut akan muncul di timeline.
5	Unfollow atau berhenti mengikuti	Berhenti melihat tweet pengguna lain pada timeline.
6	Retweet	Membagikan tweet pengguna lain agar muncul pada timeline.
7	Reply	Merespon atau membalas tweet pengguna lain.
8	Favorite	Pengguna twitter menjadikan tweet pengguna yang ia follow sebagai tweet favorit dan tweet tersebut akan muncul pada daftar favorite di profil mereka masing-masing.
9	@	Digunakan jika kita ingin menyebutkan username pengguna lain pada tweet, agar tweet tersebut muncul pada tabulasi 'interaction' pada profil mereka.
10	#	Disebut hashtag atau tagar, digunakan untuk mengelompokkan tweet. Topik yang populer diperbincangkan di twitter biasanya disertai tagar atau hashtag.

2.3 Metode TwitterRank

TwitterRank merupakan metode yang bertujuan untuk melakukan klasifikasi dan mengevaluasi pengguna yang paling berpengaruh berdasarkan suatu topik pada media sosial twitter. Metode ini dihitung berdasarkan tweet yang dibuat pengguna dengan tiga indikator. Indikator-indikator yang dimaksud adalah Followers Influence (FI), Retweet Influence (RI), dan Favorite Influence (FVI) [8].

Followers Influence (FI) adalah indikator perhitungan jumlah followers yang dimiliki pengguna. Pada umumnya, diasumsikan bahwa semakin banyak jumlah followers yang dimiliki pengguna, maka tweet yang dibuat akan semakin berpengaruh kepada pengguna lainnya. Namun, saat ini jumlah followers bukan merupakan indikator yang baik untuk mengukur pengaruh pengguna berdasarkan suatu topik. Selain itu, telah dilakukan observasi bahwa semakin banyak jumlah pengguna yang diikuti (disebut following), maka semakin kecil kemungkinan bahwa pengguna tersebut membaca seluruh tweet yang dibuat oleh para following.

u_i, \dots, u_n adalah followers dari pengguna U dan $aF w_i$ adalah jumlah following aktif u_i dalam interval waktu i . Asumsikan bahwa peluang pengguna u_i membaca tweet yang dibuat pengguna U adalah kebalikan dari $aF w_i$.

aF_{w_i} selalu bernilai ≥ 1 untuk setiap $i = 1, \dots, n$, karena pengguna u_i pasti memiliki pengguna U sebagai following pengguna u_i tersebut.

$$FI_U = \sum_{i=1}^n \frac{1}{aF_{w_i}} \quad (1)$$

Semakin tinggi nilai FI_U , maka semakin tinggi pengaruh pengguna tersebut. Nilai FI_U yang tinggi dapat dicapai oleh pengguna yang memiliki jumlah followers dan jumlah following aktif yang besar.

Retweet Influence (RI) adalah indikator perhitungan jumlah pengguna yang mengirim ulang sebuah tweet. Jika tweet yang dibuat oleh seorang pengguna memiliki jumlah retweet melebihi jumlah followers yang dimiliki pengguna tersebut, hal ini merupakan indikasi bahwa tweet telah sampai keluar lingkungan pengguna, yang berarti bahwa tweet yang dibuat pengguna tersebut dianggap penting oleh pengguna lain [8]. Namun, menurut hasil observasi yang dilakukan sebelumnya, para pengguna seringkali melakukan retweet hanya untuk alasan pertemanan, bukan melakukan retweet berdasarkan ketertarikan terhadap konten yang dibuat [13].

T adalah tweet yang ditulis oleh pengguna U dalam interval waktu i . Pengaruh tweet diperkirakan berdasarkan jumlah retweet (Rt_T) dari tweet T dengan rumus sebagai berikut.

$$RI_T = \frac{RT_T}{aF_U} \quad (2)$$

aF_U adalah jumlah followers aktif pengguna U dalam interval waktu i . Jika $aF_U = 0$, maka $RI_T = 0$. Jika $0 \leq RI_T < 1$, tweet memiliki sedikit jumlah retweet dari followers aktif pengguna. Jika $RI_T > 1$, tweet telah di-retweet oleh pengguna yang bukan merupakan followers dari pengguna U , dan dapat diasumsikan bahwa banyak pengguna lain yang tertarik dengan tweet tersebut.

Nilai retweet influence pengguna dihitung sebagai rata-rata dari nilai RI_T . T_1, \dots, T_p adalah jumlah p tweet dari pengguna U dalam interval waktu i . Rumus dari retweet influence-nya adalah sebagai berikut.

$$RI_U = \frac{\sum_{i=1}^p RI_{T_i}}{p} \quad (3)$$

Jika $p = 1$, maka $RI_U = RI_T$. Semakin tinggi nilai RI_U , maka semakin tinggi pengaruh pengguna yang diperoleh dengan membagikan tweet T_1, \dots, T_p .

Favorite Influence (FVI) adalah indikator perhitungan jumlah pengguna yang memfavoritkan sebuah tweet. Twitter memberikan fitur kepada para penggunanya agar dapat menandai sebuah tweet sebagai tweet yang disukai (favorit). Ketika seorang pengguna menandai sebuah tweet sebagai favorit, maka tweet tersebut akan tersimpan secara otomatis kedalam daftar favorit milik pengguna tersebut. Tindakan tersebut merupakan tindakan yang disengaja dan disadari. Artinya, tweet yang ditandai sebagai favorit telah dibaca dan dianggap menarik [10].

u_1, u_2, \dots, u_m adalah daftar m pengguna yang menandai tweet T (yang dibuat oleh pengguna U) sebagai favorit, jumlah favorit dari tweet T disebut F_{v_T} , dan F_{s_i} adalah jumlah followers milik pengguna u_i . Pengaruh favorite influence dihitung berdasarkan jumlah favorit dari tweet T yang dibuat pengguna U dan berdasarkan jumlah followers milik pengguna u_i , dengan rumus sebagai berikut.

$$FVI_T = \sum_{i=1, F_{s_i}=0}^m \log(F_{s_i}) \quad (4)$$

Nilai favorite influence pengguna dihitung sebagai rata-rata dari nilai FVI_T . T_1, \dots, T_r adalah jumlah r tweet dari pengguna U dalam interval waktu i . Rumus dari favorite influence-nya adalah sebagai berikut.

$$FVI_U = \frac{\sum_{i=1}^r FVI_{T_i}}{r} \quad (5)$$

Jika $r = 1$, maka $FVI_U = FVI_{T_1}$. Semakin tinggi nilai FVI_U , maka semakin tinggi pengaruh pengguna tersebut.

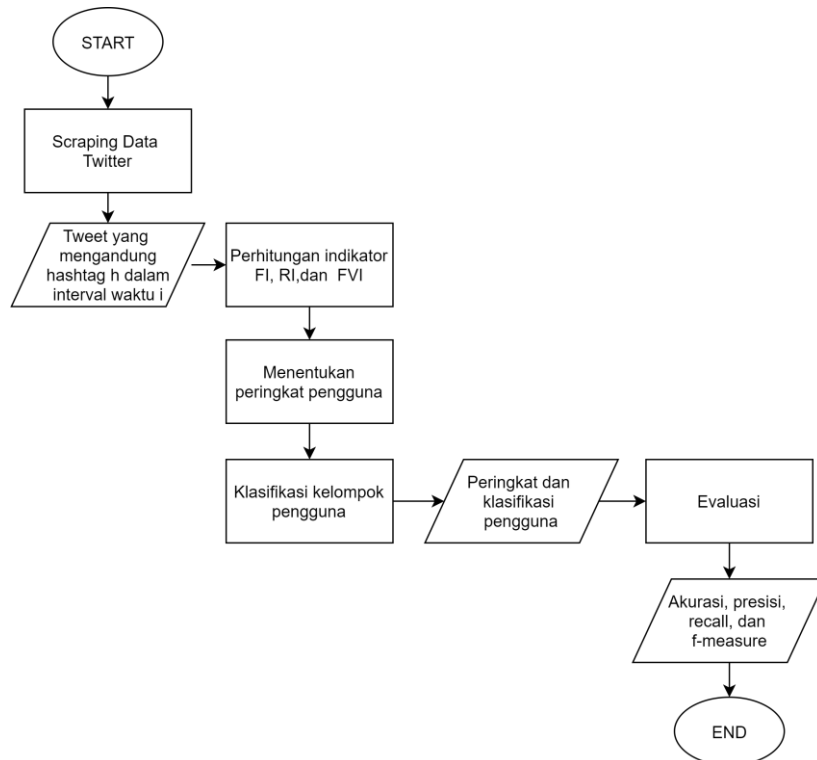
Untuk memilih dan menghasilkan peringkat pengguna, metode TwitterRank bekerja dengan cara sebagai berikut: U_1, \dots, U_1 adalah daftar pengguna yang memiliki tweet dengan hashtag h dalam interval waktu i . Untuk setiap pengguna, telah dihitung tiga indikator yaitu FI_{U_i} , RI_{U_i} , FVI_{U_i} yang telah dijelaskan sebelumnya, dengan menandai beberapa pengguna dengan nilai tertinggi untuk tiap indikatornya. Kemudian, beberapa pengguna de-

ngan nilai paling tinggi diklasifikasikan kedalam beberapa kategori. Kategori tersebut adalah highly influential user, yaitu pengguna yang memiliki nilai tinggi di ketiga indikator, influential user, yaitu pengguna yang memiliki dua nilai tinggi diantara ketiga indikator, potential influential user, yaitu pengguna yang hanya memiliki satu nilai tinggi diantara ketiga indikator, dan not influential user untuk pengguna yang tidak memiliki nilai tinggi di semua indikator.

3. Sistem yang Dibangun

3.1 Gambaran Umum Sistem

Gambaran umum sistem adalah proses sistem secara garis besar dengan menggunakan flowchart atau diagram. Gambaran umum sistem pada penelitian ini adalah sebagai berikut.



Gambar 1. Flowchart gambaran umum sistem

3.2 Tahapan Sistem

Sistem yang dibangun dalam penelitian ini memiliki beberapa tahapan pengerjaan. Tahapan yang dilakukan beserta penjelasannya adalah sebagai berikut.

3.2.1 Scraping Data Twitter

Dalam penelitian ini, data yang digunakan diambil menggunakan twitterscraper, tanpa menggunakan twitter API. Pengambilan data pada penelitian ini dipilih dengan menggunakan twitterscraper karena pengambilan data menggunakan twitterscraper tidak memerlukan twitter API, meskipun kekurangan dari penggunaan twitterscraper sendiri adalah adanya limitasi pengambilan data tweet yang bersangkutan.

Data yang diambil adalah data tweet para pengguna twitter yang menggunakan hashtag #KartiniDay yang di scrap secara real-time pada tanggal 24 April 2020. Tweet yang terkumpul adalah tweet sejak tanggal 22 April 2020 hingga tanggal 24 April 2020 sebanyak 121 tweets beserta metadatanya. Metadata yang dibutuhkan adalah total tweet, angka followers, retweet, dan favorite tweet dari masing-masing pengguna twitter beserta username, tanggal dan waktu tweet tersebut di bagikan. Berikut adalah sampel data hasil scraping menggunakan twitterscraper.

Tabel 2. Sampel data yang diambil menggunakan twitterscraper

Total Tweet	Username	Fav	RT	Flwrs	Flwg	Tweet	Timestamp
95	@HilihKintil	3	0	31	23	Semua perempuan dilahirkan untuk menjadi cantik. Sebagai seorang lelaki jantan kalian berhak memilih dan memiliki.... https://t.co/aUjW96QQoh	Apr 24 2020 09:06
1.023	@mubaadalah news	1	1	188	33	Kartini itu bukan hanya tokoh perempuan yang memperjuangkan emansipasi, tp jg tokoh intelektual yang membela keadilan..... https://t.co/JNUa2AOSPI	Apr 22 2020 07:12
86	@Panwaslih LSM	1	0	60	45	Selamat Hari Kartini untuk seluruh wanita hebat di Indonesia. Tetaplah menjadi wanita yang kuat dan pejuang yang hebat..... https://t.co/xPciHDPfJZ	Apr 23 2020 09:20
47	@cakyyus	1	0	0	23	Selamat hari dimana setiap hari wanita kuat (hatinya) #kartiniday..... https://t.co/UBIWYIUxcX	Apr 23 2020 06:23
184	@NaufalNFM	4	0	30	41	Selamat hari Kartini! Maaf ya selama beradâ.... #KartiniMasaKini #kartini..... https://t.co/0Oac5dAQOx	Apr 22 2020 15:05
9	@AndreFeb87	2	1	7	52	Selamat hari ibu Kartini.... 21 April 2020.. @andrefebrian 87 #jeparahits #jeparahariini #jeparaoceanpark..... https://t.co/v6mtBu8IyJ	Apr 22 2020 07:40
596	@panjipurnamaaa	2	0	87	62	Selamat Hari Ibu Kartini gais! #harikartini #harikartini2020 #kartiniday #kartinidays #KartiniMasaKini..... https://t.co/Sh70qVY16l	Apr 22 2020 02:20
1	@28Fhasya	0	0	1	25	” Wanita cantik ialah wanita yang membangun kekuatan dari masalah,tumbuh kuat dengan doa dan tersenyum selama beradâ.... https://t.co/kL2grmhWNR	Apr 22 2020 03:24
2.913	@Daydenn.17	1	0	149	504	”Agama memang menjauhkan kita dari dosa, tapi berapa banyak dosa yang kita lakukan atas nama agama. R.A Kartini..... https://t.co/gKyhoUGS4X	Apr 22 2020 03:20
95	@JINGGAPRMEDIA	0	0	7	9	Selamat Hari Kartini untuk semua wanita di Indonesia #jinggaprmedia #kartiniday https://t.co/LAw1G75gES	Apr 22 2020 02:19

3.2.2 Perhitungan Indikator FI, RI, dan FVI

Setelah mendapatkan dataset dengan metadata yang dibutuhkan, dilakukan perhitungan indikator Followers Influence (FI), Retweet Influence (RI), dan Favorite Influence (FVI). Perhitungan tiap indikator dilakukan sesuai dengan rumus-rumus yang telah dijelaskan pada bagian sebelumnya.

Untuk perhitungan Followers Influence, dilakukan perhitungan yang mengacu pada rumus (1) dari bagian sebelumnya. Hasil perhitungan untuk indikator Followers Influence dari sampel data yang ada dan telah diurutkan dari angka FI terbesar ke angka FI terkecil adalah sebagai berikut.

Tabel 3. Hasil perhitungan indikator FI

Username	FI
@mubaadalah news	2,30848840905997
@HilihKintil	1,31331798914354
@JINGGAPRMEDIA	1,03574431286811
@Panwaslih LSM	0,986371215511134
@Daydenn 17	0,894120361041916
@NaufalNFM	0,613386026780744
@panjipurnamaaa	0,613250014642813
@AndreFeb87	0,0645847110828795
@28Fhasya	0,000477099236641221
@cakyyus	0

Selanjutnya, untuk indikator Retweet Influence, dilakukan perhitungan yang mengacu pada rumus (2) dan (3) dari bagian sebelumnya. Hasil perhitungan indikator Retweet Influence dari sampel data yang ada dan telah diurutkan dari angka RI terbesar ke RI terkecil adalah sebagai berikut.

Tabel 4. Hasil perhitungan indikator RI

Username	RI
@mubaadalah news	5,19955907739023
@AndreFeb87	0,015873015873015872
@HilihKintil	0
@JINGGAPRMEDIA	0
@Panwaslih LSM	0
@Daydenn 17	0
@NaufalNFM	0
@panjipurnamaaa	0
@28Fhasya	0
@cakyyus	0

Indikator terakhir yang akan dihitung adalah Favorite Influence. Perhitungan indikator Favorite Influence ini mengacu pada rumus (4) dan (5) dari bagian sebelumnya. Hasil perhitungan indikator Favorite Influence dari sampel data yang ada dan telah diurutkan dari angka FVI terbesar ke FVI terkecil adalah sebagai berikut.

Tabel 5. Hasil perhitungan indikator FVI

Username	FVI
@AndreFeb87	0,4720550268943733
@HilihKintil	0,10655096756244113
@NaufalNFM	0,08686870495459517
@Panwaslih LSM	0,0161197018734871
@panjipurnamaaa	0,01267944726847802
@mubaadalah= news	0,0022508163176872395
@Daydenn 17	0,0017177982512686094
@JINGGAPRMEDIA	0
@28Fhasya	0
@cakyyus	0

Setelah itu, dipilih enam nilai tertinggi dari masing-masing pengguna pada masing-masing indikator yang ditandai dengan angka yang dicetak tebal pada tabel berikut. Angka nol berarti pengguna tidak memiliki pengaruh sama sekali pada indikator tersebut.

Tabel 6. Hasil perhitungan ketiga indikator (FI, RI, FVI) dengan enam nilai tertinggi masing-masing pengguna pada masing-masing indikator

Username	FI	RI	FVI
@NaufalNFM	0,613386026780744	0	0,08686870495459517
@HilihKintil	1,31331798914354	0	0,10655096756244113
@panjipurnamaaa	0,613250014642813	0	0,01267944726847802
@Daydenn 17	0,894120361041916	0	0,0017177982512686094
@AndreFeb87	0,0645847110828795	0,015873015873015872	0,4720550268943733
@mubaadalah news	2,30848840905997	5,19955907739023	0,0022508163176872395
@Panwaslih LSM	0,986371215511134	0	0.0161197018734871
@JINGGAPRMEDIA	1,03574431286811	0	0
@28Fhasya	0,000477099236641221	0	0
@cakyyus	0	0	0

Untuk pengguna yang memiliki angka tertinggi pada ketiga indikator, maka pengguna tersebut diklasifikasi sebagai highly influential user. Pada sampel data penelitian ini, dapat dilihat bahwa pengguna yang diklasifikasikan sebagai highly influential user adalah @mubaadalah_news dan @AndreFeb87.

Selanjutnya, pengguna yang memiliki angka tertinggi hanya pada kedua indikator, maka pengguna tersebut diklasifikasi sebagai influential user. Pada sampel data penelitian ini, pengguna yang diklasifikasikan sebagai influential user adalah @HilihKintil, dan @Panwaslih LSM.

Lalu, pengguna yang memiliki angka tertinggi pada satu indikator diklasifikasikan sebagai potential influential user. Pengguna yang diklasifikasikan sebagai potential influential user pada sampel data penelitian ini adalah @NaufalNFM, @panjipurnamaaa, @Daydenn_17, dan @JINGGAPRMEDIA.

Terakhir, pengguna lainnya yang tidak memiliki angka tertinggi pada ketiga indikator diklasifikasikan sebagai not influential user.

3.2.3 Evaluasi Performansi

Setelah pengguna twitter diklasifikasi, akan dilakukan evaluasi dari performa sistem atau metode yang digunakan. Dalam penelitian ini, evaluasi performansi diukur menggunakan accuracy (akurasi), precision (presisi), recall, dan f-measure dengan penjelasan mengacu pada tabel confusion matrix sebagai berikut. [12]

Tabel 7. Confussion Matrix

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

True Positive (TP) adalah jumlah data positif yang terklasifikasi benar oleh sistem. True Negative (TN) adalah jumlah data negatif yang terklasifikasi benar oleh sistem. False Positive (FP) adalah jumlah data positif namun terklasifikasi salah oleh sistem. Terakhir, False Negative (FN) adalah jumlah data negatif namun terklasifikasi salah oleh sistem.

1. Accuracy (Akurasi)

Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual. Rumus akurasi adalah sebagai berikut.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

2. Precision (Presisi)

Presisi adalah tingkat ketepatan antara informasi yang diminta dan jawaban yang diberikan oleh sistem. Rumus presisi adalah sebagai berikut.

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (7)$$

3. Recall

Recall adalah tingkat keberhasilan sistem dalam menemukan kembali suatu informasi. Rumus recall adalah sebagai berikut.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

4. F1-Measure (F1-Score)

F1-Measure (F1 Score) adalah penyetaraan nilai precision dan recall karena kedua nilai tersebut biasanya memiliki perbedaan yang cukup tinggi. Rumusnya adalah sebagai berikut.

$$\text{F1-Score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (9)$$

4. Evaluasi

Setelah dilakukan evaluasi dengan akurasi, presisi, recall, dan f-measure (f-score), hasil dan analisis dari evaluasi penelitian ini adalah sebagai berikut.

4.1 Hasil Pengujian

Pengujian dari penelitian ini dilakukan dengan menggunakan data pembandingan peringkat dan klasifikasi yang ditentukan oleh orang-orang yang mengisi sebuah kuisisioner. Kuisisioner tersebut berisi daftar pengguna twitter yang telah diambil data tweet-nya sesuai batasan-batasan yang telah dijelaskan sebelumnya, yang mengandung hashtag #KartiniDay didalamnya.

Orang yang mengisi kuisisioner tersebut mengurutkan pengguna yang ada pada daftar, dari yang paling berpengaruh hingga yang berpotensi memiliki pengaruh menurut angka-angka following, followers, retweet, dan favorite tweet menurut opini mereka. Menurut buku yang ditulis oleh Notoatmodjo [9], jumlah responden paling sedikit untuk uji coba adalah 20 orang. Maka, pada penelitian ini, kuisisioner diisi oleh 23 orang dan hasil kuisisioner tersebut akan digunakan sebagai data pembandingan.

Dari 121 pengguna twitter yang ada pada dataset yang digunakan, diambil peringkat 30 pengguna teratas dan menetapkan 5 pengguna teratas dengan klasifikasi highly influential user, 10 pengguna selanjutnya dengan klasifikasi influential user, 10 pengguna selanjutnya dengan klasifikasi potential influential user dan 5 pengguna terakhir dengan klasifikasi not influential user. Evaluasi dihitung dengan mengukur rata-rata akurasi, presisi, recall, dan f-measure dari seluruh responden.

Angka rata-rata hasil pengujian dari seluruh responden dengan pengukuran akurasi, presisi, recall, dan f-measure dapat dilihat pada tabel sebagai berikut.

Tabel 8. Angka rata-rata hasil pengujian dan evaluasi sistem dari seluruh responden

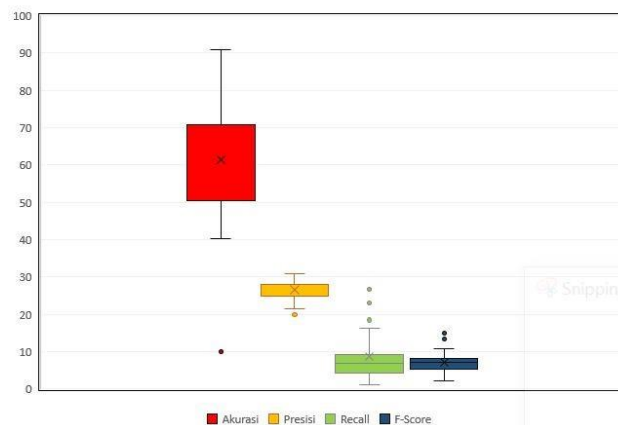
	Akurasi	Presisi	Recall	F-Measure
Minimum	10%	20%	1,12%	2,12%
Q1	55,5%	25,1%	5,35%	5,23%
Median	70,7%	27,9%	6,8%	7,07%
Q3	70,7%	27,9%	9,3%	7,78%
Maksimum	90,91%	30,8%	26,8%	14,9%
Mean	61,4%	26,6%	8,78%	7,1%

4.2 Analisis Hasil Pengujian

Berdasarkan tabel 8, pengujian sistem dengan pengukuran akurasi mendapatkan hasil rata-rata dari seluruh responden sebesar 0.614 atau 61.4%, lalu pengujian sistem dengan pengukuran presisi mendapatkan hasil rata-rata dari seluruh responden sebesar 0.266 atau 26.6%. Selanjutnya, pengujian sistem dengan pengukuran recall mendapatkan hasil rata-rata dari seluruh responden sebesar 0.0878 atau 8.78% dan terakhir pengujian sistem dengan pengukuran f-measure mendapatkan hasil rata-rata dari seluruh responden sebesar 0.071 atau 7.1%.

Untuk hasil akurasi sebesar 61.4%, dapat digambarkan bahwa tingkat kedekatan antara nilai prediksi dengan nilai aktual adalah hanya sebanyak 61 dari 100 data yang ada. Hasil akurasi ini dapat dikatakan cukup rendah karena hanya setengah dari jumlah total data yang memiliki nilai aktual dan nilai prediksi yang tidak terlalu jauh. Selanjutnya, hasil presisi pada penelitian ini adalah 0.266 atau 26.6%, yang berarti ketepatan antara nilai aktual dengan nilai yang diprediksi sistem hanya tepat 26 hingga 27 data dari 100 data yang ada. Lalu untuk nilai recall pada penelitian ini adalah 0,0878 atau 8.78%. Maka, dapat dikatakan bahwa sistem hanya dapat menemukan kembali 8 hingga 9 data dari 100 data yang digunakan. Pengukuran terakhir untuk hasil pengujian dan evaluasi dari sistem pada penelitian ini adalah dengan f-measure, yaitu penyetaraan nilai antara presisi dan recall yaitu sebesar 0.071 atau 7.1%. Sehingga, untuk nilai presisi, recall, dan f-measure pada penelitian ini dapat dikatakan sangat rendah karena data dengan nilai prediksi yang tepat tidak mencapai setengah dari jumlah total data yang ada.

Analisis hasil pengujian pada penelitian ini diinterpretasikan melalui sebuah boxplot. Boxplot adalah salah satu cara dalam statistik deskriptif untuk menggambarkan data numerik secara grafik dengan lima ukuran, yaitu nilai hasil observasi terkecil, kuartil terendah (Q1), median (nilai tengah), kuartil tertinggi (Q3), dan nilai hasil observasi terbesar [5]. Satu ukuran terakhir yaitu nilai ekstrem atau disebut juga sebagai outlier dari data hasil pengukuran, namun tidak semua boxplot menampilkan ukuran tersebut. Jarak antara bagian-bagian dari box menunjukkan derajat penyebaran dan kecondongan dalam data. Dalam penggambarannya, boxplot dapat digambarkan secara horizontal maupun vertikal. Hasil akurasi, presisi, recall, dan f-measure dari seluruh responden dapat dilihat pada box plot sebagai berikut.



Gambar 2. Box plot hasil evaluasi

Dari hasil pengujian yang didapat, box plot pengukuran akurasi sistem lebih tinggi dan lebih lebar dari metode evaluasi lainnya, yang berarti pada penelitian ini memiliki nilai akurasi yang lebih tinggi dibanding hasil pengukuran yang lainnya. Sedangkan, box plot hasil pengukuran presisi, recall, dan f-measure lebih sempit dan lebih kecil, karena nilai pengukuran yang lebih rendah dibandingkan nilai akurasi sistem pada penelitian ini.

Hasil klasifikasi pengguna twitter pada sampel data yang digunakan pada tabel 2 dapat dilihat pada tabel sebagai berikut.

Tabel 9. Hasil klasifikasi dengan metode TwitterRank

Total Tweet	Username	Fav	RT	Followers	Following	Klasifikasi
1.023	@mubaadalah news	0	0	188	33	Highly Influential
9	@AndreFeb87	2	1	7	52	Highly Influential
95	@HilihKintil	3	0	31	23	Influential
86	@Panwaslih LSM	1	0	61	45	Influential
184	@NaufalNFM	4	0	30	41	Potential Influential
596	@panjipurnamaaa	2	0	87	62	Potential Influential
95	@JINGGAPR MEDIA	0	0	7	9	Potential Influential
2.913	@Daydenn 17	1	0	149	504	Potential Influential
1	@28Fhasya	0	0	1	25	Not Influential
47	@cakyyus	1	0	0	23	Not Influential

Berdasarkan tabel 9, hasil klasifikasi sistem dengan menggunakan metode TwitterRank menghasilkan empat kelompok pengguna berdasarkan sampel data pada tabel 2. Pengguna @mubaadalah.news yang merupakan pengguna yang paling berpengaruh memiliki jumlah followers yang paling banyak diantara pengguna lainnya. Ada juga beberapa pengguna yang memiliki jumlah favorite, retweet, followers, dan following terkecil dengan angka kurang dari 10. Namun, tidak semua dari pengguna-pengguna tersebut adalah pengguna yang tidak berpengaruh. Sebagai contoh, pengguna @AndreFeb87 adalah pengguna yang dengan jumlah favorite, retweet, followers, dan following terkecil dengan angka kurang dari 10, namun pengguna @AndreFeb87 merupakan pengguna dengan kelompok kasifikasi yang sangat berpengaruh (highly influential user). Sehingga, dapat dikatakan bahwa besar atau kecilnya jumlah favorite, retweet, followers, dan following, tidak menentukan seberapa berpengaruhnya seorang pengguna terhadap penyebaran suatu topik atau hashtag.

Pada penelitian ini, dapat dikatakan bahwa hasil pengukuran pengaruh pengguna dengan menggunakan metode pada penelitian ini kurang baik. Hal ini dikarenakan angka setiap indikator pada dataset yang digunakan kurang lengkap. Misalnya, pada indikator retweet atau favorite tweet, masih banyak angka 0 pada sebagian besar data tweet yang didapat, sehingga pengukuran pada indikator tersebut tidak teruji secara maksimal. Selain itu, data pembanding yang digunakan adalah peringkat pengguna pada dataset menurut opini para responden dengan menggunakan google form. Berdasarkan hasil wawancara dengan beberapa responden, opini mereka mengenai peringkat pengguna twitter hanya dilihat berdasarkan pengguna mana yang memiliki angka paling besar pada satu indikator saja tanpa memperhitungkan indikator lainnya. Sebagai contoh, responden hanya mengurutkan pengguna twitter pada dataset dari pengguna dengan angka followers terbesar hingga terkecil. Dalam mengurutkan pengguna yang berpengaruh, responden belum tentu benar-benar memperhitungkan dan memperhatikan satu atau lebih indikator yang ada, karena bisa saja peringkat yang dibuat responden yang digunakan sebagai data pembanding adalah peringkat yang diurutkan secara sembarangan.

Selain itu, adanya limitasi saat pengambilan data juga merupakan salah satu kendala pada penelitian ini. Sehingga pengambilan dataset tidak dapat dilakukan secara maksimal.

5. Kesimpulan

Penelitian ini dilakukan dengan tujuan melakukan klasifikasi dan mengevaluasi pengguna twitter terhadap suatu hashtag pada waktu tertentu. Percobaan ini menggunakan dataset yang mengandung hashtag #KartiniDay sebanyak 121 tweets. Sesuai hasil klasifikasi pengguna pada penelitian ini, dapat dikatakan bahwa besar atau kecilnya jumlah favorite, retweet, followers, dan following, tidak menentukan seberapa berpengaruhnya seorang pengguna terhadap penyebaran suatu topik atau hashtag.

Hasil pengujian peringkat pengaruh pengguna twitter dan klasifikasi pengguna twitter dengan menggunakan metode TwitterRank menghasilkan akurasi sebagai hasil pengujian sistem paling besar yaitu sebesar 0.614 atau 61.4%, lalu pengujian sistem dengan pengukuran presisi mendapatkan hasil sebesar 0,266 atau 26.6%. Selanjutnya, pengujian sistem dengan pengukuran recall mendapatkan hasil sebesar 0.0878 atau 8.78% dan terakhir pengujian sistem dengan pengukuran f-measure mendapatkan hasil sebesar 0,071 atau 7.1%.

Maka, dapat dikatakan bahwa metode pada penelitian ini kurang baik untuk mengukur dan mengklasifikasi pengguna yang berpengaruh berdasarkan suatu topik. Hal ini dikarenakan kurang akuratnya data pembanding yang digunakan. Karena data pembanding yang digunakan merupakan sebuah opini yang tidak dapat diukur kebenarannya. Selain itu, limitasi saat pengambilan data tweets juga merupakan salah satu kendala pada penelitian ini.

Saran untuk penelitian selanjutnya yaitu sebaiknya data diambil dalam jumlah yang lebih besar. Selain itu, untuk data pembanding yang digunakan, sebaiknya dilakukan survei secara langsung untuk pengurutan pengguna twitter oleh responden agar proses pengurutan pengguna lebih tertata dan tidak dilakukan secara sembarangan.

Daftar Pustaka

- [1] I. Anger and C. Kittl. Measuring influence on twitter. Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies, pages 1–4, 2011.
- [2] T. J. Bristol. Twitter: Consider the possibilities for continuing nursing education. The Journal of Continuing Education in Nursing, pages 199–200, 2010.
- [3] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring user influence in twitter: The million follower fallacy. Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, pages 10–17, 2010.

- [4] A. Gupta, H. Lamba, P. Kumaguru, and A. Joshi. Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. Proceedings of the 22nd International Conference on World Wide Web Companion, pages 729–736, 2013.
- [5] Junaidi. Deskripsi data melalui box-plot. 2015.
- [6] X. Ma, C. Li, J. Bailey, and S. Wijewickrema. Finding influentials in twitter: A temporal influence ranking model. ArXiv, 2017.
- [7] F. Maclean, D. Jones, G. Carin-Levy, and H. Hunter. Understanding twitter. British Journal of Occupational Therapy, pages 295–298, 2013.
- [8] M. Montangero and M. Furini. Trank: ranking twitter users according to specific topics. 12th Annual IEEE Consumer Communication and Networking Conference (CNCC), pages 767–772, 2015.
- [9] S. Notoatmodjo. Metodologi Penelitian Kesehatan. Rineka Cipta, 2012.
- [10] K. Rosman. On twitter, more 'favoriting': Why more people are pushing the favorite button. The Wallstreet Journal, 2013.
- [11] G. Stringhini, G. Wang, M. Egele, C. Kruegel, G. Vigna, H. Zheng, and B. Y. Zhao. Follow the green: Growth and dynamics in twitter follower marketers. Proceeding of the 2013 Conference on Internet Measurement Conference, pages 163–176, 2013.
- [12] B. J. P. Utomo, M. A. Bijaksana, and A. Romadhony. Answer classification using multiple features extraction. e-Proceeding of Engineering, 3:3700–3710, 2016.
- [13] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twiterrank: Finding topic-sensitive influential twitter. Proceeding of The Third ACM International Conference on Web Search and Data Mining, pages 261–270, 2010.