Vol. 03 No. 02 Desember 2021 P-ISSN: 2656-7377 || E-ISSN: 2714-8467

KLASIFIKASI AKUN BUZZER PEMILU PADA MEDIA SOSIAL TWITTER BERDASARKAN DATA TWEET MENGGUNAKAN ALGORITMA C4.5

Agus Maulana¹, Sri Kuswayati²
Program Studi Teknik Informatika^{1,2}
Sekolah Tinggi Teknologi Bandung^{1,2}
sugaanaluam@gmail.com¹, srikuswayati@sttbandung.ac.id²

Abstrak

Kemajuan teknologi membawa manusia kepada perkembangan aktifitas sosial, dari yang semula beraktifitas secara tatap muka berubah menjadi secara daring. Perkembangan ini juga membawa perubahan kepada aktifitas politik seperti pemilihan umum, baik dalam tahap pemilihan kepada daerah, pemilihan gubernur, pemilihan legislative, hingga pemilihan presiden. Aktifitas politik saat ini mengiring opini masyarakat dan membuka kesempatan masyarakat untuk berkomentar dan mengemukakan pendapat melalui Media Sosial baik untuk mendukung pilihannya ataupun menjatuhkan lawan. Aktifitas inilah yang kemudian mengahdirkan buzzer-buzzer media sosial, lebih jauh lagi aktifitas buzzer ini dijadikan sebuah profesi yang dapat menghasilkan uang. Aktifitas buzzer ini kemudian menyimpan pola perilaku dari penggunaan Twitter yang kemudian dapat dicontoh dan diikuti oleh pengguna-pengguna Twitter lainnya yang tertarik untuk menjadi akun buzzer Twitter. Aktor politik juga dapat mencari dan menggali akun yang memang berpotensi menjadi buzzer yang bisa ia gunakan jasanya dalam melakukan kampanye politik. Penelitian ini ditujukan untuk merancang sebuah model dan membangun sebuah sistem yang dapat memenuhi kebutuhan yang sudah dijelaskan diatas. Data mining dan Algoritma Klasifikasi C4.5 adalah metode yang tepat untuk menyelesaikan permasalahan ini.

Kata kunci: Algoritma C4.5, Data Mining, Twitter, Buzzer, Pemilu

Abstract

Advanced technology have brought people to the growth of social activities, from what were originally face-to-face activities to online activities. This growth also brought changes to political activities such as general elections, both in the regional elections, legislative elections, and presidential elections. Current political activities accompany public opinion and open up opportunities for the people to comment and express opinions through social media either to support their choice or to bring down their opponents. This activity will creates social media buzzers, and this buzzer activity is turned into a profession that can make money. This buzzer activity then stores behavioral patterns from the use of Twitter which can then be imitated and followed by other Twitter users who are interested in becoming a Twitter buzzer account. Political actors can also search for and dig up accounts that have the potential to become buzzers that they can use to carry out political campaigns. This research is intended to design a model and build a system that can meet the needs described above. Data mining and C4.5 Classification Algorithm is the right method to solve this problem. Keywords: C4.5 Algorithm, Data mining, Twitter, Buzzer, Elections.

I. PENDAHULUAN

Menurut Bambang Arianto, fenomena tingginya penggunaan media sosial melahirkan para aktor komunikator yang memiliki pengaruh hingga bisa membangun citra positif seperti peran *Buzzer* dan *Influencer* [1]. *Buzzer* pada awalnya digunakan untuk mempromosikan suatu produk tertentu dengan atau tanpa imbalan. Namun, sejak tahun 2014 ketika pemilihan umum (pemilu), jasa *buzzer* mulai dilirik oleh aktor-aktor politik, tujuannya untuk memenangkan pilkada, pileg, hingga pilpres [2]. Peran *buzzer* pemilu pada media sosial Twitter adalah untuk memperluas suatu informasi melalui aktifitas *retweet* terkait narasi dan *hashtag* harian hingga dapat dilihat oleh masyarakat dalam bentuk *trending topic*.

Penggunaan Twitter sebagai media informasi dan publikasi memiliki keterbatasan, seperti dalam segmentasi pengguna seperti apa yang dapat digunakan akunnya sebagai jasa *buzzer*. Keterbatasan pada Twitter tentu akan mempengaruhi terhadap klasifikasi akun *buzzer* yang dapat menjadi acuan bagi aktor-aktor politik dalam menggunakan jasa akun *buzzer* selama periode pemilihan umum. Dengan demikian, diperlukan metode yang dapat menterjemahkan karakteristik dari pengguna Twitter berdasarkan *tweet* yang dimiliki. Sehingga dapat ditemukan pola data untuk mengklasifikasikan suatu akun berpotensi menjadi akun *buzzer* atau tidak.

Dalam melakukan segmentasi pada media sosial Twitter di Indonesia akan memakan waktu yang cukup lama karena banyaknya pengguna twitter di Indonesia. Untuk itu, dalam menyelesaikan permasalahan efisiensi waktu ini dapat diselesaikan dengan menggunakan cara *Data mining*. *Data mining* sebagai proses menemukan struktur yang menarik dalam data. Struktur yang dapat mengambil banyak bentuk, termasuk seperangkat aturan, sebuah grafik atau jaringan, sebuah *tree*, satu atau beberapa persamaan, dan banyak lagi. Struktur dapat menjadi bagian dari dasbor visual yang kompleks atau sederhana [3]. Salah satu algoritma yang dapat diaplikasikan pada proses klasifikasi menggunakan *data mining* ini adalah *Decision Tree* C4.5. *Decision Tree* C4.5 adalah algoritma yang digunakan untuk mengklasifikasikan data dan digambarkan dengan diagram pohon keputusan [4].

Vol. 03 No. 02 Desember 2021

P-ISSN: 2656-7377 || E-ISSN: 2714-8467

II. TINJAUAN PUSTAKA

Abdul Bari Azed dalam jurnalnya menyebutkan, pemilihan umum adalah salah satu lembaga yang berfungsi sebagai sarana penyampaian hak-hak demokrasi rakyat [8]. Dalam konferensi *International Commission of Jurist* tahun 1965 disebutkan bahwa salah satu dari enam syarat bagi Negara demokrasi adalah menyelenggarakan pemilihan umum, yang mana Negara demokrasi adalah suatu bentuk pemerintahan dimana warga Negara melaksanakan hak yang sama melalui wakil-wakil yang dipilih dan bertanggung jawab melalui proses pemilihan yang bebas. Sehingga, eksistensi lembaga pemilihan umum sudah diakui oleh negara-negara yang bersendikan asas kedaulatan rakyat.

1. Buzzer

Kehadiran media baru seperti media sosial di masayarakat saat ini sejatinya merupakan angin segar dalam bidang periklanan. Kreatifitas para praktisi periklanan diuji dalam menyajikan satu tampilan iklan yang berbeda melalui media sosial. Dewasa ini, tantangan gaya periklanan melalui media sosial ini sepertinya telah terjawab oleh beberapa pihak, dengan munculnya profesi-profesi terkait periklanan media sosial. Salah satu profesi baru yang mulai dikenal dan dimanfaatkan dalam periklanan adalah profesi sebagai *Buzzer*. Mengutip dalam jurnal yang dibuat oleh Erika Dwi, *Buzzer* merupakan orang yang dibayar oleh pemilik produk/jasa yang akan diiklankan atau oleh agensi iklan untuk mendengungkan atau mendesas-desuskan produk/jasa yang dimaksud, sehingga produk atau jasa tersebut bias dikenal orang, dibicarakan orang, dan masuk dalam benak sasaran [9].

Peran *buzzer* sebagai profesi periklanan kemudian dimanfaatkan dalam kancah politik, seiring dengan pesatnya perkembanagan media sosial dan politik. Akmaliah menegaskan definisi *Buzzer* mengalami pergeseran makna dan tujuan sejak tahun 2014. Pada tahun tersebut partai politik banyak menggunakan jasa mereka untuk kampanye Pilpres 2014. Penggunaan *Buzzer* untuk kepentingan politik ini mengubah definisi dan makna tersebut. Sebelumnya, ia memiliki konotasi positif, kemudian berubah menjadi cenderung negatif. Memang pada tahun 2012, khususnya Pilkada DKI, *Buzzer* digunakan dalam politik, tapi maknanya masih cenderung positif seiring dengan belum maraknya kampanye hitam [10].

2. Data Mining

Menurut Kantardzic, dalam buku ketiganya disebutkan bahwa *data mining* adalah suatu proses berulang dengan penemuan-penemuan baik secara manual ataupun otomatis [5]. Disebutkan juga bahwa data mining paling berguna dalam ekslorasi analisis skenario yang belum ditentukan sebelumnya sehingga menemukan hasil yang menarik. *Data mining* adalah penemuan baru, berharga dan bukan hal sepele untuk data dalam jumlah banyak, melainkan upaya manusia dalam menemukan hasil terbaik melalui komputerisasi. Dalam prakteknya *data mining* memiliki dua tujuan, pertama untuk *prediction* dan yang kedua untuk *description*.

3. CRISP-DM (CRoss Industry Standard Process for Data Mining)

Data Mining adalah proses kompleks yang membutuhkan berbagai alat dan orang yang berbeda. Keberhasilan proyek data mining bergantung pada perpaduan yang tepat antara alat yang baik dan analis yang terampil. Selain itu, diperlukan metodologi yang kuat dan manajemen proyek yang efektif. Model proses dapat membantu untuk memahami dan mengelola interaksi di sepanjang proses yang kompleks ini.

Model referensi CRISP-DM untuk *data mining* memberikan gambaran umum siklus hidup dari proyek *data mining*. Ini berisi fase proyek, tugas masing-masing, dan *output*nya. Siklus hidup proyek data mining dipecah dalam enam fase. Urutan fase tidak ketat. Tanda panah hanya menunjukkan ketergantungan yang paling penting dan sering di antara fase, tetapi dalam proyek tertentu, itu tergantung pada hasil dari setiap fase di mana fase tersebut, atau tugas tertentu dari fase tersebut, harus dilakukan selanjutnya [7].

4. Classification

Salah satu tugas utama *data mining* adalah Klasifikasi/ *Classification*. Menurut Kantardzic, klasifikasi adalah fungsi prediktif untuk mengklasifikasikan sebuat data menjadi salah satu dari beberapa kelas yang sudah ditentukan [5]. Setiap klasifikasi yang diberikan kepada suatu sampel didasarkan kepada nilai atribut yang dimiliki sampel tersebut. Dengan kata lain, klasifikasi adalah proses memberikan label kepada data atau *record* yang tidak berlabel. Misalnya, klasifikasi sederhana seperti mengelompokkan catatan penagihan pelanggan menjadi dua kelas khusus: mereka yang membayar tagihan mereka di dalamnya 30 hari dan mereka yang membutuhkan waktu lebih dari 30 hari untuk membayar.

III. ANALISIS DAN PERANCANGAN

Tahapan awal dalam proses data mining adalah dengan analisa proses CRISP-DM sebagai metode standar yang digunakan dalam proses data mining. Dalam penelitian ini tahapan tersebut dijelaskan lebih lengkap, sebagai berikut:

1. Bussiness Understanding

Aktifitas yang dilakukan pada tahapan ini dimulai dari pemahaman terkait regulasi yang diberlakukan oleh pihak Twitter. Twitter memberikan peluang bagi developer aplikasi untuk mengolah data yang dimiliki dengan

P-ISSN: 2656-7377 || E-ISSN: 2714-8467

menggunakan API yang sudah disediakan. Dalam API tersebut diperlukan kode autentikasi sehingga setiap developer yang menggunakan data tweet dapat diketahui oleh pihak Twitter sehingga mengurangi resiko salah guna data.

nsumer API keys			Regenerate	
API key:	101-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1			
API secret key:	ALC: 7103 C. C. ANY THE PROPERTY OF THE PROPER			
ccess token & access token secret		Revoke	Regenerate	
	ken and secret when you first generate it in order to make your acc which will invalidate your existing tokens.	ount more secure.	You can revoke or	
Access token:	(00000000000000000000000000000000000000			
Access token secret:	100000000000000000000000000000000000000		Last generated: Oct 26, 2020	

Gambar 1. Kode Autentikasi Twitter API Untuk Developer

Kode autentikasi Twitter API dapat diperoleh melalui halaman https://developer.twitter.com/ developer dapat masuk menggunakan akun twitter yang akan digunakan sebagai developer user. Kemudian, kode autentikasi tersebut disimpan pada kode program.

2. Data Understanding

Aktifitas yang dilakukan pada tahapan ini dimulai dari pemahaman terkait data tweet yang akan dijadikan data awal, karena data *tweet* yang diambil perlu untuk dilakukan filterasi sehingga data yang diambil adalah data yang sesuai dengan kebutuhan bisnis yang sedang diteliti. Selain itu, data *tweet* juga tidak menampilkan informasi lengkap sehingga perlu dilakukan proses untuk menemukan *hidden information* dalam *tweet*, seperti jenis kelamin dan domisili.

3. Data Preparation

Aktifitas yang dilakukan pada tahapan ini adalah melakukan data cleaning atau proses membersihkan data dari anomali dan atribut yang tidak digunakan untuk proses selanjutnya. Kemudian data yang sudah dibersihkan disimpan kedalam database sehingga menjadi sebuah data training yang dapat diolah. Proses data preparation lebih lengkap akan dijelaskan sebagai berikut:

a. Transformasi perubahan nilai, transformasi yang dilakukan untuk menghitung nilai dari data agar dapat digunakan dalam pemodelan klasifikasi.

1. Usia akun

Merupakan atribut usia yang diambil berdasarkan atribut create_at_user. Pada tahapan ini usia akun diperoleh berdasarkan kapan waktu data diambil dikurang dengan kapan waktu akun Twitter pertama kali dibuat. Dengan menggunakan rumus persamaan sebagai berikut:

Usia Akun = Waktu data diambil - Waktu akun dibuat

2. Rentang waktu rata-rata tweet (menit)

Merupakan rata-rata rentang waktu dari total tweet pengguna. Dengan mengambil total tweet, lalu menghitung interval waktu dari tweet ke-1 hingga tweet ke-n, dijabarkan dalam rumus persamaan sebagai berikut:

$$Avg\ Tweet = Waktu\ Tweet\ (ke-n) - Waktu\ Tweet\ (ke1)$$

3. Jumlah URL

Merupakan rata-rata jumlah URL dari totsl tweet pengguna. Rumus persamaan untuk mendapatkan nilai jumlah URL adalah sebagai berikut:

nilai jumlah URL adalah sebagai berikut:
$$Avg\ URL = \frac{\Delta URL1 + \Delta URL2 + \cdots + \Delta URLn}{total\ tweet}$$

4. Jumlah re-tweet

Merupakan rata-rata jumlah re-tweet dari total tweet pengguna. Rumus persamaan untuk mendapatkan nilai jumlah re-tweet adalah sebagai berikut:

$$Avg Re - Tweet = \frac{\Delta RT1 + \Delta RT2 + \cdots + \Delta RTn}{total tweet}$$

b. Transformasi perubahan bentuk, ransformasi yang digunakan untuk mengubah data input berupa tipe numerik kedalam bentuk diskrit, dengan mengelompokkan sesuai kriteria dan interval dari nilai atribut yang ada.

Vol. 03 No. 02 Desember 2021

P-ISSN: 2656-7377 || E-ISSN: 2714-8467

1. Followers

Pada penelitian ini jumlah Follower di kategorikan menjadi 2 kategori, high dan low. Untuk menentukan nilai diskrit pada atribut follower ditentukan berdasarkan jumlah follower dengan persamaan sebagai berikut

$$Follower = (LOW < 500 \ge HIGH)$$

Following

Pada penelitian ini jumlah Following di kategorikan menjadi 2 kategori, high dan low. Untuk menentukan nilai diskrit pada atribut following ditentukan berdasarkan jumlah following dengan persamaan sebagai berikut

Following =
$$(LOW < 500 \ge HIGH)$$

3. Usia akun

Nilai diskrit pada atribut usia akun adalah "baru" jika nilai usia akun ≤ 730 hari, dan "lama" jika nilai usia akun > 730 hari. Rumus persamaannya sebegai berikut:

$$Usia\ akun = (baru \le 730hari > lama)$$

4. Rentang waktu tweet

Nilai diskrit pada atribut rentang waktu tweet adalah "aktif" dan "pasif". Perubahan nilai ditentukan melalui persamaan berikut:

$$Avg\ Tweet = (pasif \leq 1200menit > aktif)$$

5. Jumlah URL

Nilai diskrit pada atribut Jumlah URL adalah "spam" jika nilai jumlah URL > 0.5 dan "nonspam" jika nilai jumlah URL ≤ 0.5 . Untuk persamaan yang digunakan sebagai berikut:

$$Avg\ URL = (nonspam \le 0.5 > spam)$$

6. Jumlah re-tweet

Nilai diskrit pada atribut Jumlah URL adalah "sosial" jika nilai jumlah Re-tweet > 0.5 dan "non-sosial" jika nilai jumlah Re-tweet ≤ 0.5 . Untuk persamaan yang digunakan sebagai berikut:

$$Avg\ Re-Tweet=(nonsosial \le 0.5 > sosial)$$

c. Pelabelan data, adalah tahapan untuk memberikan kelas pada data. kelas yang dimaksud adalah kelas "buzzer" dan kelas "non-buzzer". Pelabelan data dilakukan secara manual dengan merujuk kepada atribut buzzer sebagaimana tercantum dalam tabel 3.1. Atribut yang digunakan merupakan hasil penelitian yang dilakukan oleh Ibrahim, dkk. (2015), dengan tingkat akurasi 86%.

TABEL I ATRIBUT UNTUK MENENTUKAN AKUN *BUZZER*

No	Atribut
1	Frekuensi tweet harian pengguna
2	Usia akun pengguna (sejak pembuatan)
3	Banyaknya URL yang disematkan dalam tweet
4	Bayanknya retweet yang dilakukan dalam tweet
5	Banyaknya Follower (pengikut)
6	Banyaknya Following (mengikuti)

Selaras dengan atribut yang disampaikan diatas, Arianto (2020) menyebutkan bahwa tipologi dari buzzer politik adalah memiliki jumlah follower lebih besar dari 500 follower. Frekuensi tweet juga menjadi salah satu tipologi yang disebutkan, rata-rata frekuensi tweet dari akun yang dikategorikan buzzer adalah kurang dari 1200 menit, dengan usia akun yang berbilang baru.

4. Modelling

Dalam tahap pemodelan, aktifitas yang dilakukan adalah pemilihan data yang sudah tersimpan kedalam database dan menyesuaikan parameter yang ada untuk mendapatkan nilai yang optimal, dan menjadi data yang sesuai dengan format yang akan digunakan. Dalam hal ini, proses yang dilakukan adalah mencari entropi dan information gain, yaitu dengan cara:

P-ISSN: 2656-7377 || E-ISSN: 2714-8467

Menghitung nilai total entropi, rumusnya yaitu:

Menghitung nilai total entropi, rumusnya yaitu:
$$I total \ entropy = \sum_{i=1}^{jumlah \ atribut} \frac{nilai \ atribut \ ke-i}{nilai \ atribut \ Label} * I \ atribut \ [5]$$

Keterangan:

I total entropy: jumlah total nilai entropy dari seluruh atribut

I atribut : jumlah nilai informasi dari tiap atribut

Kemudian untuk menhitung nilai I atribut adalah dengan rumus berikut:

$$I \ atribut = \sum_{i=1}^{atribut} (-pi * log2(pi))$$
 [6]

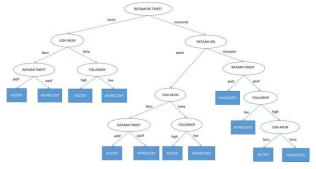
Menghitung nilai information gain dengan rumus sebagai berikut:

$$Gain(R, A) = I a tribut Label - I total entropy [7]$$

Keterangan: R: Label A: atribut

I atribut Label: jumlah nilai informasi dari atribut Label Akun I total entropy: jumlah nilai entropy dari seluruh atribut

Dari hasil perhitungan yang dilakukan pada penjelasan diatas, maka dapat dibentuk satu pohon keputusan seperti pada gambar 2 berikut.



Gambar 2. Hasil Pembentukan Pohon Keputusan

Evaluation

Kemudian setelah model dibuat, tahapan selanjutnya adalah melakukan evaluasi dari hasil model-model yang telah terbentuk. Model-model ini kemudian diterjemahkan menjadi informasi yang penting atau bernilai tinggi. Proses evaluasi dan validasi ini juga berguna sebelum diputuskan bahwa model yang ditemukan layak untuk dilakukan deployment.

Deployment

Tahapan ini dapat dilakukan ketika model data mining sudah divalidasi. Dari model-model yang dihasilkan, perlu diuji dengan data yang berbeda dan dilakukan evaluasi kembali mengenai keakuratan dari hasil model tersebut.

IV. KESIMPULAN

Kesimpulan yang bisa didapatkan dari penelitian ini mengacu kepada tujuan penelitian yang terlah dirancang. Diantaranya:

- Sistem perangkat lunak yang dibangun sesuai dengan hasil analisa dan perancangan sistem. Sistem dapat menghasilkan informasi akun twitter berdasarkan data tweet pengguna. Hasil pohon keputusan dapat diterapkan menjadi rule (aturan) pada sistem yang berjalan, sehingga seluruh komponen sistem dapat berfungsi sebagai mana mestinya.
- Sistem dapat mengidentifikasi setiap informasi yang didapatkan yang kemudian dilakukan penyesuaian data dan klasifikasi yang diharapkan. Sehingga model klasifikasi dengan menggunakan algoritma C4.5 dapat diterapkan untuk mengklasifikasikan akun buzzer berdasarkan nilai atribut yang terdapat pada setiap akun pengguna Twitter dengan cara menghitung nilai entropy dan gain.
- Dengan model yang diterapkan, aplikasi dapat melakukan fungsi sesuai yang diharapkan. Hasil uji model dapat dilihat melalui hasil pengujian berdasarkan Confusion Matrix secara bertahap dengan menggunakan 10.540 data tweet didapatkan nilai akurasi secara berurutan 89%, 87%, 88%, dan 88%. Sehingga rata-rata nilai akurasi dari algoritma C4.5 untuk klasifikasi akun *buzzer* pada pengguna Twitter sebesar 88.5%.

NARATIF(Jurnal Ilmiah Nasional Riset Aplikasi dan Teknik Informatika)

Vol. 03 No. 02 Desember 2021 P-ISSN: 2656-7377 || E-ISSN: 2714-8467

REFERENSI

- [1] Bambang Arianto (2020), Peran Buzzer Media Sosial Dalam Memperkuat Ekosistem Pemasaran Digital, Jurnal Optimum, Vol 10, No.1.
- [2] Felicia, Riris Loisa (2018), Peran Buzzer Politik dalam Aktivitas Kampanye di Media Sosial Twitter, Koneksi, Vol.2, No.2.
- [3] Richard J. Roiger (2017), *Data Mining: A Tutorial Based Primer* 2nd Edition, CRC Press, Taylor & Francis Group, New York.
- [4] Yi yang & Wenguang Chen (2016), Taiga: Performance Optimization of the C4.5 Decision Tree Construction Algorithm. Tsinghua Science and Technology Vol.21 No.4.
- [5] Mehmed Kantardzic (2020), Data Mining: Concepts, Models, Methods, and Algorithms, 3rd Edition, John Wiley & sons, New Jersey.
- [6] Mochamad Ibrahim, Omar Abdillah, Alfan F. Wicaksono, Mirna Adriani (2015), Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Result in A Twitter Nation, IEEE 15th International Conference on Data Mining Workshops.
- [7] Rudiger Wirth & Jochen Hipp, CRISP-DM: Towards a Standard Process Model for Data Mining.
- [8] Abdul Bari Azed (2006), Sistem Pemilihan Umum di Indonesia. Vol.17. No.2.
- [9] Erika Dwi Setya Watie (Januari 2012), Periklanan Dalam Media Baru(Advertising In The New Media). THE MESSENGER, Volume IV, No.1.
- [10] Wahyudi Akmaliah (Juni 2018), Bukan Sekedar Penggaung (Buzzer) Media Sosial dan Transformasi Arena Politik. MAARIF Vol.13, No 1.