

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Penelitian ini dilaksanakan untuk melakukan *social network analysis* terhadap aplikasi Twitter untuk mencari pengguna Twitter yang paling berpengaruh terhadap topik pembicaraan mengenai resesi. Penelitian ini menggunakan aplikasi Twitter dikarenakan Twitter merupakan salah satu media sosial paling populer dan sarana berita *online* yang memungkinkan pengguna untuk mengekspresikan pandangan mereka dan berpartisipasi dalam berbagai masalah berbeda di dunia serta berperan penting dalam acara dan proses sosial pada saat ini [45]. Data yang digunakan dalam penelitian ini adalah data hasil *scraping* Twitter dengan *keyword* resesi menggunakan aplikasi NodeXL. Data yang diambil berjumlah 306028 data yang diambil dari periode Juli 2022 sampai dengan Desember 2022. Terdapat 2 *sheet* data yang akan digunakan pada penelitian ini yaitu *sheet edges* dan *sheet vertices*. Pada *sheet edges* berisi data-data untuk membentuk sebuah garis pada jaringan yang dinamakan *edge*. Data tersebut diantaranya *vertex 1*, *vertex 2*, *relationship*, *relationship date*, *tweet*, *URLs in tweet*, dan beberapa data pendukung lainnya. Beberapa deskripsi dari masing-masing variabel data di *sheet edges* dapat dilihat pada tabel 3.1 seperti berikut:

Tabel 3. 1 *Data Variable in Sheet Edges*

No	Variable	Description
1	<i>Vertex 1</i>	Akun yang memberikan aksi relasi
2	<i>Vertex 2</i>	Akun yang menerima aksi relasi dari <i>vertex 1</i>
3	<i>Relationship</i>	Tipe relasi yang dilakukan seperti <i>mention</i> , <i>retweet</i> , <i>tweet</i> , <i>replies to</i> , dan <i>mentions in retweet</i>
4	<i>Relationship Date</i>	Tanggal dari aksi relasi yang dilakukan

No	Variable	Description
5	<i>Tweet</i>	Isi dari pesan yang diunggah
6	<i>URLs in Tweet</i>	<i>Link website</i> yang dicantumkan dalam <i>tweet</i>

Pada *sheet vertices* berisi data-data untuk membentuk sebuah representasi dari sebuah individu dalam jaringan yang dinamakan *vertice* atau *node*. Data tersebut diantaranya adalah *vertex*, *name*, *user ID*, *followed*, *followers*, *tweets*, *favorites*, *location*, dan beberapa data pendukung lainnya. Beberapa deskripsi dari masing-masing variabel data pada *sheet vertices* dapat dilihat pada tabel 3.2 seperti berikut:

Tabel 3. 2 *Data Variable in Sheet Vertices*

No	Variable	Description
1	<i>Vertex</i>	<i>Username</i> dari akun Twitter
2	<i>Name</i>	Nama akun Twitter yang ditampilkan pada halaman akun
3	<i>User ID</i>	<i>Unique ID</i> untuk merepresentasikan suatu akun
4	<i>Followed</i>	Jumlah akun yang diikuti oleh akun Twitter tersebut
5	<i>Followers</i>	Jumlah akun yang mengikuti akun Twitter tersebut
6	<i>Tweets</i>	Jumlah <i>tweet</i> yang diunggah oleh akun Twitter
7	<i>Favorites</i>	Jumlah <i>likes</i> yang didapatkan akun Twitter
8	<i>Location</i>	Lokasi tempat tinggal akun Twitter tersebut

Variabel yang digunakan untuk membuat jalur interaksi antar dua *node* yaitu variabel *vertex 1* dan *vertex 2* pada *sheet edges*. Variabel *vertex 1* mengacu pada akun yang melakukan aksi relasi seperti *reply tweet* ke akun twitter yang didefinisikan sebagai *vertex 2*. *Vertex 2* berfungsi sebagai penerima informasi yang dituju dari *vertex 1*. Variabel yang digunakan untuk membuat *node* yaitu variabel *vertex* pada *sheet vertices*. Variabel yang lainnya

tidak digunakan pada penelitian ini karena variabel tersebut berfungsi sebagai data pendukung dalam *dataset*. Data tersebut mencakup informasi lebih lengkap mengenai informasi akun Twitter dan isi dari *tweet* yang diunggah yang dapat digunakan untuk penggunaan algoritma lainnya seperti *topic modelling*.

3.2 Metode Penelitian

3.2.1 Alur Penelitian

Alur penelitian berisi mengenai tahapan yang dilakukan selama proses penelitian dari awal hingga akhir. Tahapan alur penelitian mengadopsi kerangka kerja dari CRISP-DM. Berikut merupakan tahapan proses pengerjaan penelitian yang dilakukan:

1. *Business Understanding*

Tahapan pertama yang dilakukan pada penelitian ini adalah melakukan identifikasi masalah yang terjadi. Pada penelitian ini, masalah yang diangkat adalah akibat dari pandemi COVID-19 yang memungkinkan terjadinya isu resesi. Dengan banyak perbincangan tentang isu resesi yang akan terjadi, banyak masyarakat takut dengan terjadinya resesi karena berdampak pada kehidupan ekonomi mereka seperti terjadinya PHK massal. Selain itu, juga dengan banyak informasi yang beredar maka makin besar juga berita hoax dapat tersebar secara luas. Penelitian menggunakan metode *social network analysis* dengan menggunakan *centrality measure* dan *community detection* bertujuan agar masyarakat dapat mengetahui akun *news platform* mana yang menampilkan berita yang akurat dan terpercaya.

2. *Data Understanding*

Ketika sudah mengetahui permasalahan yang terjadi, maka dapat dilakukan pengumpulan data berdasarkan permasalahan pada identifikasi masalah. Data yang diambil adalah data *user* dari Twitter dengan kata kunci resesi. Data diambil dengan

menggunakan *tools* NodeXL dan didapatkan sebanyak 306.026 data.

3. *Data Preparation*

Pada tahapan ini, dilakukan preparasi data seperti melakukan pembersihan data yang tidak diperlukan dan juga dilakukan *filter data* yaitu hanya menggunakan data akun yang berelasi langsung dengan *news platform* lokal.

4. *Modeling*

Setelah data berhasil dikumpulkan, tahapan selanjutnya adalah melakukan analisis data menggunakan algoritma *social network analysis*. Algoritma *social network analysis* ini dilakukan dengan menghitung nilai *centrality* dan menentukan komunitas yang terbentuk. Perhitungan *centrality* dan *community detection* dilakukan dengan menggunakan *tools* Python. Untuk memperindah hasil visualisasi dari hasil perhitungan *centrality* dan *community detection* dapat dilakukan dengan menggunakan *tools* Gephi.

5. *Evaluation*

Setelah melakukan *modeling* data, tahapan selanjutnya adalah melakukan evaluasi hasil dan pembahasan terhadap model yang telah dibuat. Evaluasi yang didapatkan bersifat subjektif karena algoritma *social network analysis* merupakan tipe *unsupervised learning* sehingga tidak ada perhitungan hasil dari evaluasi. Evaluasi dilakukan dengan melihat hasil dari grafik yang terbentuk dan dari nilai perhitungan *centrality* yang dapat disimpulkan siapa yang memiliki pengaruh terhadap penyebaran informasi resesi.

6. *Deployment*

Pada tahap *deployment*, hasil analisis data yang telah terbentuk direpresentasikan dalam bentuk sistem atau visualisasi sehingga dapat dimanfaatkan oleh pengguna sebagai media informasi. Pada tahap ini dapat dilakukan

pembuatan *dashboard* mengenai akun Twitter yang paling berpengaruh terhadap penyebaran informasi resesi sehingga dari kalangan pembaca atau pemerintah dapat melihat dan memilih akun tersebut sebagai sumber informasi terpercaya.

3.2.2 Metode *Data Mining*

Pada penelitian ini, metode *data mining* yang digunakan adalah CRISP-DM. Selain itu, terdapat beberapa metode *data mining* yang bisa digunakan seperti KDD dan SEMMA. Untuk perbandingan metode *data mining* antara CRISP-DM, KDD, dan SEMMA dapat dilihat pada tabel 3.3 [46].

Tabel 3. 3 Perbandingan CRISP-DM, KDD, dan SEMMA

Indicator	CRISP-DM	KDD	SEMMA
<i>Phase</i>	<ol style="list-style-type: none"> 1. <i>Business Understanding</i> 2. <i>Data Understanding</i> 3. <i>Data Preparation</i> 4. <i>Modeling</i> 5. <i>Evaluation</i> 6. <i>Deployment</i> 	<ol style="list-style-type: none"> 1. <i>Selection</i> 2. <i>Preprocessing</i> 3. <i>Transformation</i> 4. <i>Data Mining</i> 5. <i>Interpretation / Evaluation</i> 6. <i>Knowledge</i> 	<ol style="list-style-type: none"> 1. <i>Sample</i> 2. <i>Explore</i> 3. <i>Modify</i> 4. <i>Model</i> 5. <i>Assess</i>
<i>Strengths</i>	<ol style="list-style-type: none"> 1. Proses tahapan yang jelas 2. Cocok untuk <i>project</i> atau <i>dataset</i> yang besar 3. Proses iteratif 4. Memiliki dokumentasi dengan studi kasus serupa yang akan membantu prosesnya 	<ol style="list-style-type: none"> 1. Proses iteratif 2. Mendukung berbagai teknik <i>data mining</i>, seperti <i>Neural Network</i> 	<ol style="list-style-type: none"> 1. Dukungan pengguna yang baik 2. Mendukung berbagai teknik <i>data mining</i> 3. Proses iteratif

Indicator	CRISP-DM	KDD	SEMMA
<i>Limitations</i>	1. Proses yang panjang 2. Tahapan <i>data preparation</i> dan <i>modeling</i> berbeda dari <i>data mining</i> tradisional sehingga tidak ada dalam dokumentasi CRISP-DM	1. Membutuhkan latar belakang pengetahuan di <i>data mining</i>	1. Tidak memperhitungkan pemahaman bisnis dari suatu masalah.

Berdasarkan tabel perbandingan pada tabel 3.1, penelitian ini menggunakan metode CRISP-DM dikarenakan tahapan dalam proses CRISP-DM yang jelas dan terarah sehingga dapat membantu proses penelitian ini. Selain itu, penelitian ini menggunakan *dataset* yang besar dan CRISP-DM memiliki panduan untuk melakukan *data mining* pada *dataset* yang besar sehingga sangat cocok untuk digunakan dalam penelitian ini [24].

3.3 Teknik Pengumpulan Data

3.3.1 Populasi dan Sampel

3.3.1.1 Populasi

Populasi adalah keseluruhan objek penelitian atau objek yang akan diteliti [47]. Populasi yang digunakan pada penelitian ini adalah seluruh data pengguna aktif Twitter yang melakukan interaksi terkait dengan topik resesi.

3.3.1.2 Sampel

Sampel merupakan sebuah teknik atau prosedur yang digunakan untuk memilih sejumlah item atau individu yang lebih kecil dari jumlah populasi yang telah ditentukan untuk dijadikan sumber data dalam melakukan penelitian [48]. Teknik pengambilan sampel yang dilakukan adalah *purposive sampling*.

Purposive sampling merupakan sampel yang karakteristiknya

ditentukan untuk tujuan yang relevan dengan penelitian [49]. Sampel yang digunakan pada penelitian ini adalah data twitter dimana data yang digunakan adalah informasi akun-akun yang melakukan interaksi atau yang berhubungan dengan portal berita dengan topik seputar resesi.

3.3.2 Periode Pengambilan Data

Pada penelitian ini, data yang diambil merupakan data Twitter dengan periode waktu dimulai dari bulan Juli 2022 – Januari 2023.

3.4 Teknik Analisis Data

Pada tahapan ini, terdapat beberapa *tools* yang dapat digunakan. Berikut merupakan kegunaan dari *tools* yang digunakan dalam melakukan analisis data, yaitu:

Tabel 3. 4 Tools yang Digunakan dalam Penelitian

Indikator	NodeXL	Python	Gephi
<i>Usage</i>	Digunakan untuk mengambil data dalam jumlah yang sangat banyak dan melakukan visualisasi dalam bentuk jaringan.	Digunakan dalam perhitungan statistika jaringan seperti menghitung berbagai jenis <i>centrality</i> dan menghitung jumlah <i>community</i>	Melakukan visualisasi data berbentuk jaringan

Tahapan pertama dalam melakukan analisis data adalah tahap *scraping data*. Pada tahap ini dilakukan pengumpulan data yang dibutuhkan dalam penelitian. Pengumpulan data dilakukan dengan cara menarik data dari media sosial Twitter. Berdasarkan kegunaan aplikasi pada tabel 3.4, *tools* yang digunakan adalah NodeXL. Aplikasi NodeXL digunakan pada penelitian ini

karena dapat menarik data Twitter dalam jumlah besar sehingga banyak data yang akan didapatkan.

Tahapan setelah melakukan *scraping data* adalah menentukan nilai *centrality* dari masing-masing akun Twitter untuk mengetahui *influence* dari akun tersebut. Mengacu pada tabel 3.4, *tools* yang digunakan untuk perhitungan *centrality* adalah Python. Tahap ini menggunakan aplikasi Python dibandingkan R Studio adalah karena Python adalah bahasa yang mudah untuk dimengerti. Selain itu terdapat *library* NetworkX yang mempermudah dan mendukung dalam penghitungan *centrality* dari jaringan.

Tahapan selanjutnya adalah melakukan visualisasi dalam bentuk jaringan. Pada tahapan ini, visualisasi yang dihasilkan bertujuan untuk mempermudah dalam membaca hasil akhir dari penelitian. Untuk melakukan visualisasi dan mengacu pada tabel 3.4, *tools* yang digunakan dalam melakukan penelitian adalah Gephi. Keunggulan dari menggunakan *tools* Gephi dikarenakan aplikasi ini dapat memproses data ke dalam jaringan dengan jumlah yang sangat besar. Selain itu, pada aplikasi Gephi, pengguna dapat memilih berbagai jenis tipe jaringan untuk memperindah visualisasi dan juga cara penggunaannya yang mudah namun dapat menghasilkan kesimpulan terhadap pengaruh akun Twitter dengan sangat baik melalui visualisasi tersebut.

Selain dari penggunaan *tools* untuk menganalisis data, terdapat juga beberapa teknik dalam menganalisis data untuk memberikan hasil yang optimal. Beberapa teknik diantaranya adalah menghitung nilai *centrality* dan mengelompokkan data dengan teknik *community detection*. Pada tahapan *centrality* digunakan untuk menghitung seberapa pentingnya setiap *node* yang diberikan dalam jaringan. *Centrality* juga dapat memotong data yang *noisy*, mengungkap bagian jaringan yang membutuhkan perhatian tetapi semuanya bekerja secara berbeda. Setiap ukuran *centrality* memiliki definisi dan kepentingan tersendiri. Pada bagian *community detection*, algoritma ini berfungsi untuk mendeteksi serta mengidentifikasi adanya komunitas pada

suatu jaringan sosial. Adapun beberapa jenis dari *community detection* dapat dilihat pada tabel 3.5 [22].

Tabel 3. 5 Jenis *Community Detection*

Indikator	Greedy Modularity	Walktrap	Label Propagation	Louvain Communities
<i>Method</i>	Menggunakan modularitas untuk menentukan pembagian optimal dalam jaringan. setiap <i>node</i> dipindahkan ke komunitas yang memaksimalkan fungsi modularitas. Komunitas-komunitas ini kemudian digabungkan sampai fungsi modularitas tidak dapat lagi ditingkatkan.	Menghitung matriks transisi dan melakukan perjalanan acak selama jumlah langkah tertentu. Setiap <i>node</i> dianggap komunitasnya sendiri lalu bergabung dengan komunitas yang berdekatan dengan cara yang meminimalkan jumlah jarak antar komunitas.	Memberi setiap <i>node label</i> yang unik. Setiap <i>node</i> mengadopsi label yang sama dengan mayoritas tetangganya dengan hasil seri diputus secara acak. Proses ini dilakukan secara iteratif sampai setiap simpul memiliki label yang sama dengan mayoritas tetangganya.	Menukar <i>node</i> antar komunitas dan mengevaluasi perubahan modularitas sampai tidak ada perbaikan lagi. Lalu menggabungkan komunitas menjadi simpul laten dan mengidentifikasi bobot <i>edge</i> dengan simpul laten.
<i>Type of Graph</i>	<i>Directed and Undirected Graph</i>	<i>Directed and Undirected Graph</i>	<i>Undirected Graph</i>	<i>Undirected Graph</i>

Berdasarkan tabel 3.5 mengenai jenis-jenis dari *community detection*, penelitian ini menggunakan jenis *greedy modularity* dan *walktrap*. Alasan dari digunakan jenis *community detection* tersebut adalah dikarenakan kedua jenis tersebut dapat diimplementasikan pada tipe *directed graph* maupun *undirected graph*. Tipe *graph* yang digunakan pada penelitian ini adalah

directed graph dikarenakan objek penelitian yang digunakan lebih berfokus kepada relasi antara aktor yang satu dengan yang lainnya sehingga menggunakan *directed graph* agar bisa melihat kemana arah relasi suatu aktor terhadap aktor yang lain. Selain itu, algoritma *label propagation* dan *louvain* bersifat tidak deterministik yang dimana hasil yang diberikan pasti berbeda setiap kali dijalankan. Hal ini kurang cocok untuk diterapkan kedalam penelitian karena penelitian ini ingin mengidentifikasi komunitas yang relatif stabil terhadap pembahasan topik resesi.



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA