

SKRIPSI

**DETEKSI KOMUNITAS DAN POLA PENYEBARAN TWEETS
DARI DATA TWITTER**



Fariz Armesta

NPM: 2017730011

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2022**

UNDERGRADUATE THESIS

**COMMUNITY DETECTION AND TWEETS SPREAD
PATTERNS FROM TWITTER DATA**



Fariz Armesta

NPM: 2017730011

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2022**

LEMBAR PENGESAHAN

DETEKSI KOMUNITAS DAN POLA PENYEBARAN TWEETS DARI DATA TWITTER

Fariz Armesta

NPM: 2017730011

Bandung, 18 Januari 2022

Menyetujui,

Pembimbing

Digitally signed
by Veronica Sri
Moertini

Dr. Veronica Sri Moertini

Ketua Tim Penguji
Digitally signed
by Cecilia Esti
Nugraheni

Dr.rer.nat. Cecilia Esti Nugraheni

Anggota Tim Penguji
Digitally signed
by Keenan
Adiwijaya Leman

Keenan Adiwijaya Leman, M.T.

Mengetahui,

Ketua Program Studi

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

DETEKSI KOMUNITAS DAN POLA PENYEBARAN TWEETS DARI DATA TWITTER

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 18 Januari 2022



Fariz Armesta
NPM: 2017730011

ABSTRAK

Twitter merupakan salah satu jejaring sosial yang populer digunakan didunia pada saat ini. Dikarenakan hal tersebut, Twitter terus mengakumulasi jumlah data yang begitu banyak setiap harinya. Hal tersebut memunculkan suatu permasalahan yaitu bagaimana mendapatkan suatu *insights* dari data yang telah terkumpul tersebut.

Suatu solusi yang berpotensi untuk menemukan *insights* dari data Twitter adalah dengan melakukan pencarian dan analisis komunitas pada data Twitter. Komunitas dapat terbentuk dari interaksi yang dilakukan antara pengguna-pengguna Twitter. Dari komunitas yang terbentuk tersebut dapat dimanfaatkan berdasarkan sifat dan karakteristik dari komunitas tersebut.

Dalam melakukan hal tersebut, dikarenakan jumlah data Twitter yang dapat diproses berjumlah besar, pemilihan teknologi yang tepat sangatlah penting dalam melakukan analisis terhadap data. Teknologi yang dapat digunakan dalam melakukan analisis terhadap data yang berjumlah sangat besar tersebut adalah seperti Spark dan Hadoop MapReduce. Hal tersebut dikarenakan kedua teknologi tersebut dapat memproses data dengan secara terdistribusi, yang membuat pemrosesan data menjadi lebih efisien.

Dengan menggunakan teknologi-teknologi tersebut berhasil didapatkan berbagai jenis komunitas yang terbentuk dari data Twitter tersebut. 4 data Twitter yang digunakan terbentuk berdasarkan 3 buah topik perbincangan, yaitu topik pemanasan global, olahraga, *video game*, dan gabungan antara olahraga, *video game*, dan musik. Dari data-data yang digunakan, ditemukannya persamaan hasil yang didapatkan. Komunitas-komunitas tersebut terbentuk berdasarkan dari 3 tipe interaksi antara pengguna Twitter menggunakan *tweet*, yaitu interaksi *reply*, *retweet*, dan *quote*. Dimana setiap jenis komunitas tersebut memiliki keunggulan dan kekurangannya masing-masing.

Kata-kata kunci: Komunitas, Twitter, Spark, data besar, graf

ABSTRACT

Twitter is one of the most popular social networks used in the world today. Due to this, Twitter continues to accumulate huge amounts of data every day. This raises a problem, namely how to get an insight from the data that has been collected.

A potential solution for finding insights from Twitter data is to perform community searches and analyzes on Twitter data. Communities can be formed from interactions between Twitter users. From the formed community, it can be utilized based on the nature and characteristics of the community.

In doing so, due to the large amount of Twitter data that can be processed, choosing the right technology is very important in analyzing the data. Technologies that can be used to analyze such large amounts of data are Spark and Hadoop MapReduce. This is because both technologies can process data in a distributed manner, which makes data processing more efficient.

By using these technologies, various types of communities were formed from the Twitter data. The 4 Twitter data used are formed based on 3 topics of conversation, namely the topic of global warming, sports, video games, and a combination of sports, video games, and music. From these data, similarity between the results are obtained. These communities are formed based on 3 types of interactions between Twitter users using *tweet*, namely *reply*, *retweet*, and *quote* interactions. Where each type of community has its own advantages and disadvantages.

Keywords: Community, Twitter, Spark, big data, graph

KATA PENGANTAR

Dengan memanjatkan puji dan dan syukur kehadirat Allah SWT yang telah melimpahkan rahmat, dan hidayah-Nya sehingga skripsi ini dapat diselesaikan. Penyelesaian skripsi ini tidak lepas dari bantuan dan bimbingan berbagai pihak yang telah membantu dan mendukung selama menjalani perkuliahan di UNPAR dan pengerjaan skripsi ini. Pada kesempatan ini Penulis ingin berterima kasih kepada semua pihak yang telah membantu dalam pengerjaan skripsi ini, ucapan terima kasih sebesar-besarnya kepada yang terhormat :

1. Kepada Ibu dan Ayah yang telah memberi dukungan dalam bentuk material maupun emosional selama ini.
2. Kepada Ibu Dr. Veronica Sri Moertini selaku dosen pembimbing skripsi ini. Terima kasih atas ilmu, pengalaman, bimbingan, dan kesabaran yang telah diberikan selama ini.
3. Kepada Ibu Dr.rer.nat. Cecilia Esti Nugraheni dan Bapak Keenan Adiwijaya Leman, M.T. selaku penguji. Terima kasih atas masukan, kritik, dan saran yang telah diberikan sehingga skripsi ini dapat menjadi lebih baik.
4. Kepada seluruh dosen dan staf UNPAR yang telah mengajar dan membantu Penulis selama menuntut ilmu di UNPAR.
5. Kepada teman-teman dan sahabat-sahabat di UNPAR, terima kasih atas pertemanan dan bantuan-bantuannya selama, semoga pertemanan ini dapat terus berjalan dan terus saling membantu satu sama lain.
6. Kepada sahabat-sahabat *online* Fatih, Florian, Archie, Taylor, Payton, dan Barry terima kasih telah mendukung dan menemani Penulis selama ini, semoga pertemanan ini dapat terus berlanjut dan dapat terus saling membantu satu sama lain seperti suatu keluarga.
7. Kepada seseorang yang tidak bisa saya cantumkan namanya, terima kasih atas bantuanmu selama ini, selama kurang lebih 9 tahun ini, semoga kita dapat mendapatkan kebahagiaan yang selalu kita impikan.
8. Terakhir kepada seluruh pihak yang tidak tercantumkan namanya pada bagian ini, Terima kasih atas segala bantuan dan dukungannya.

Sebagai penutup, ucapan terima kasih Penulis berikan kepada pembaca dokumen skripsi ini. Semoga skripsi ini dapat bermanfaat bagi pembaca dokumen skripsi ini baik dalam kehidupan pribadinya, bisnis, maupun sebagai referensi penelitian di masa depan. Mohon maaf jika terdapat kesalahan penulisan, sistematika pengerjaan, atau metode-metode yang dipaparkan pada skripsi ini.

Bandung, Januari 2022

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Batasan Masalah	2
1.5 Metodologi	2
1.6 Sistematika Pembahasan	3
2 LANDASAN TEORI	5
2.1 Twitter	5
2.1.1 Definisi	5
2.1.2 Kegunaan Twitter	5
2.1.3 Bentuk-Bentuk <i>Tweet</i>	6
2.2 Komunitas	9
2.3 Graf	11
2.3.1 Tipe dan Struktur Graf	11
2.3.2 Karakteristik Graf	12
2.4 Algoritma Analisis Graf	15
2.4.1 Connected Components	15
2.4.2 PageRank	16
2.5 Big Data	19
2.5.1 Pengaruh Big Data	20
2.5.2 Penggerak Big Data	20
2.5.3 Teknologi Pada Big Data	20
2.6 Hadoop dan MapReduce	21
2.6.1 Hadoop	21
2.6.2 MapReduce	21
2.7 Scala	22
2.7.1 Tipe Data	22
2.7.2 Variabel	23
2.7.3 Operator	23
2.8 Spark	24
2.8.1 DataFrame	24
2.8.2 Partisi	25
2.8.3 Lazy Evaluation	25
2.8.4 Arsitektur Spark	26

3	EKSPLORASI TEKNOLOGI DAN DATA TWITTER UNTUK MENGIDENTIFIKASI KOMUNITAS	27
3.1	Cytoscape	28
3.2	<i>Dataset</i>	30
3.2.1	<i>Dataset</i> Pemanasan Global	30
3.2.2	<i>Dataset</i> <i>Video Game</i>	31
3.2.3	<i>Dataset</i> Olahraga	31
3.3	Eksplorasi dan Penyiapan Data	31
3.3.1	Eksplorasi Data	34
3.3.2	Penyiapan Data Graf	34
3.4	Eksperimen untuk Mengidentifikasi Komunitas	41
3.4.1	Eksperimen Menggunakan Data Sintetik	42
3.4.2	Eksperimen pada <i>Dataset</i> Pemasanan Global	43
3.4.3	Eksperimen pada <i>Dataset</i> <i>Video Game</i>	47
3.4.4	Eksperimen pada <i>Dataset</i> Olahraga	49
3.5	Ringkasan hasil Eksperimen-Eksperimen	52
4	ANALISIS DATA TWITTER UNTUK PENCARIAN KOMUNITAS	53
4.1	Pengumpulan Data	53
4.2	Eksplorasi Data Hasil Pengumpulan	54
4.3	Penyiapan Data	57
4.4	Analisis Hasil Eksperimen pada Interaksi <i>Retweet</i>	58
4.4.1	Analisis Graf <i>component</i> "223338299654" pada Interaksi <i>Retweet</i>	59
4.4.2	Analisis Graf <i>component</i> "309237646155" pada Interaksi <i>Retweet</i>	60
4.4.3	Analisis Graf <i>component</i> "24051817168" pada Interaksi <i>Retweet</i>	60
4.4.4	Analisis Graf <i>component</i> "180388627386" pada Interaksi <i>Retweet</i>	61
4.4.5	Analisis Graf <i>component</i> "747324311311" pada Interaksi <i>Retweet</i>	62
4.4.6	Analisis Graf <i>component</i> "197568496561" pada Interaksi <i>Retweet</i>	63
4.5	Analisis Hasil Eksperimen pada Interaksi <i>Reply</i>	64
4.5.1	Analisis Graf <i>component</i> "1005022349438" pada Interaksi <i>Reply</i>	65
4.5.2	Analisis Graf <i>component</i> "60129542595" pada Interaksi <i>Reply</i>	65
4.5.3	Analisis Graf <i>component</i> "609885357515" pada Interaksi <i>Reply</i>	66
4.5.4	Analisis Graf <i>component</i> "618475291697" pada Interaksi <i>Reply</i>	67
4.6	Analisis Hasil Eksperimen pada Interaksi <i>Quote</i>	68
4.7	Insights yang Didapatkan dan Manfaatnya	69
5	KESIMPULAN	73
5.1	Kesimpulan	73
5.2	Saran	74
	DAFTAR REFERENSI	75
	A KODE PROGRAM	77

DAFTAR GAMBAR

2.1	Gambar contoh <i>tweet</i> yang dikirim oleh pengguna dengan nama Joko Widodo	7
2.2	Gambar contoh <i>retweet</i> yang dilakukan oleh pengguna dengan nama Kang Santuy	7
2.3	Gambar contoh <i>quote</i> yang dilakukan oleh pengguna dengan nama Kang Santuy	8
2.4	Gambar contoh <i>reply</i> yang dilakukan oleh pengguna dengan nama Kang Santuy	9
2.5	Gambar 3 tipe dasar graf [1]	11
2.6	Gambar 3 tipe jaringan graf [1]	12
2.7	Gambar graf <i>connected</i> dan <i>disconnected</i> [1]	12
2.8	Gambar graf <i>unweighted</i> dan <i>weighted</i> [1]	13
2.9	Gambar graf <i>directed</i> dan <i>undirected</i> [1]	13
2.10	Contoh gambar graf <i>acyclic</i> dan <i>cyclic</i> [1]	14
2.11	Contoh 3 gambar graf beserta perhitungan <i>density</i> setiap graf tersebut [1]	14
2.12	Gambar <i>cluster</i> yang ditemukan menggunakan Connected Component [1]	16
2.13	Gambar perhitungan PageRank pada suatu graf [1]	19
2.14	Gambar representasi arsitektur Spark[2]	26
3.1	Gambar visualisasi pada Cytoscape [3]	28
3.2	Gambar visualisasi galFiltered.sif pada Cytoscape [3]	29
3.4	Gambar 5 baris pertama data pemanasan global bulan Mei	30
3.3	Gambar visualisasi VizMapper pada Cytoscape [3]	30
3.5	Gambar contoh baris data <i>tweet</i>	31
3.6	Gambar kata kunci <i>dataset video game</i>	32
3.7	Gambar kata kunci <i>dataset olahraga</i>	32
3.8	Gambar diagram <i>flowchart</i> pergerakan data	32
3.9	Gambar diagram <i>flowchart</i> penyiapan data	35
3.10	Gambar diagram <i>flowchart</i> pembuatan <i>nodes</i>	37
3.11	Gambar diagram <i>flowchart</i> pembuatan <i>edges</i>	39
3.12	Gambar diagram <i>flowchart</i> proses eksperimen pada data	41
3.13	Gambar visualisasi graf pada data sintetik	44
3.14	Gambar <i>bar chart</i> jumlah Connected Component untuk setiap interaksi pada data pemanasan global	44
3.15	Gambar visualisasi graf tidak berarah interaksi <i>retweet</i> data pemanasan global	45
3.16	Gambar visualisasi graf berarah interaksi <i>retweet</i> data pemanasan global	46
3.17	Gambar visualisasi graf berarah interaksi <i>reply</i> data pemanasan global	46
3.18	Gambar <i>bar chart</i> jumlah Connected Component untuk setiap interaksi pada data <i>video game</i>	47
3.19	Gambar visualisasi graf tidak berarah interaksi <i>retweet</i> data <i>video game</i>	48
3.20	Gambar visualisasi graf tidak berarah interaksi <i>reply</i> data <i>video game</i>	48
3.21	Gambar visualisasi graf berarah interaksi <i>reply</i> data <i>video game</i>	49
3.22	Gambar <i>bar chart</i> jumlah Connected Component untuk setiap interaksi pada data olahraga	50
3.23	Gambar visualisasi graf tidak berarah interaksi <i>retweet</i> data olahraga	50
3.24	Gambar visualisasi graf tidak berarah interaksi <i>reply</i> data olahraga	51

3.25	Gambar visualisasi graf berarah interaksi <i>quote</i> data olahraga	52
4.1	Gambar daftar kata kunci mengenai aliran musik	54
4.2	Gambar contoh baris <i>tweet</i>	54
4.3	Gambar contoh baris <i>tweet</i> bertipe <i>retweet</i>	55
4.4	Gambar jumlah tiap interaksi pada data besar	56
4.5	Gambar contoh baris <i>tweet</i> bertipe <i>retweet</i>	57
4.6	Gambar visualisasi graf tidak berarah interaksi <i>retweet</i> data besar	58
4.7	Gambar profil Twitter akun BlackettPromo dan 4DaPeopleMusic	59
4.8	Gambar contoh <i>tweet</i> JosephDeMauro	60
4.9	Gambar contoh <i>tweet</i> mytenace	61
4.10	Gambar profil Twitter akun LivEchoLFC	62
4.11	Gambar profil Twitter akun Wapbaze_Com	63
4.12	Gambar profil Twitter akun Kidzcoolit	64
4.13	Gambar visualisasi graf tidak berarah interaksi <i>reply</i> data besar	64
4.14	Gambar profil Twitter akun jhooniq	65
4.15	Gambar profil Twitter akun DOBBYK1M	66
4.16	Gambar profil Twitter akun SB19TeamBooster	67
4.17	Gambar profil Twitter akun MAbillboard	68
4.18	Gambar visualisasi graf tidak berarah interaksi <i>quote</i> data besar	68
4.19	Gambar profil Twitter akun FremontCyril	69

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Twitter adalah salah satu jejaring sosial yang sangat populer digunakan pada saat ini. Dengan pada tahun 2012, 140 juta pengguna aktif dan 340 juta kicauan (*tweet*) dikirim setiap harinya pada jejaring sosial Twitter¹. Hal tersebut membuat jumlah data yang diproduksi oleh Twitter berjumlah sangat besar, dari data-data tersebut berpotensi untuk ditemukannya berbagai informasi-informasi yang berguna dan dapat dimanfaatkan.

Salah satu bentuk informasi yang berpotensi didapatkan dari data Twitter tersebut adalah informasi mengenai komunitas. Terdapat dua buah arti berbeda dari penggunaan arti komunitas [4]. Yang pertama adalah komunitas berdasarkan teritorial dan geografis seperti daerah, atau kota [4]. Yang kedua adalah "relasional", berkaitan dengan "kualitas karakter hubungan manusia, tanpa referensi pada lokasi" [4]. Informasi-informasi, seperti sub komunitas atau komunitas yang mungkin berada di dalam suatu komunitas dapat ditemukan, yang lalu informasi tersebut dapat digunakan dalam suatu kegiatan bisnis, seperti pemasaran suatu produk. Dikarenakan potensi informasi yang dapat ditemukan dari data Twitter tersebut maka diperlukannya suatu cara yang efisien untuk mendeteksi suatu komunitas dari data Twitter.

Untuk mendapatkan informasi-informasi tersebut data Twitter perlu diolah dan dianalisis dengan menggunakan peralatan yang tepat. Data mengenai teks Twitter dapat didapatkan melalui berbagai teknik, salah satu teknik tersebut adalah Tweepy. Tweepy merupakan sebuah *library* yang berguna dalam mengakses Twitter API². Dikarenakan sifat dari data Twitter yang besar, perlu dipilihnya peralatan yang tepat sehingga proses pengolahan dan analisis dapat dijalankan dengan seefisien mungkin. Peralatan yang akan digunakan adalah Hadoop MapReduce dan Spark, dengan digunakannya kedua perangkat lunak tersebut, data dapat diproses secara paralel dengan efisien.

Hadoop menyediakan suatu sistem *file* terdistribusi dan kerangka kerja untuk melakukan analisis dan transformasi terhadap kumpulan data yang sangat besar menggunakan paradigma MapReduce [5]. Transformasi ini dilakukan dikarenakan kita ingin data yang dimiliki menjadi berbentuk graf, bentuk graf dipilih dikarenakan interaksi-interaksi yang dapat terjadi antar pengguna-pengguna Twitter dapat terlihat dengan lebih jelas menggunakan bentuk data tersebut. Selain hal tersebut, dengan graf pun dapat ditemukannya pengguna-pengguna mana saja yang paling berpengaruh pada komunitas tersebut. Analisis terhadap data graf tersebut dapat dilakukan dengan menggunakan *library* yang telah tersedia pada perangkat lunak Spark, yaitu *library* GraphFrames. Dengan menggunakan teknologi-teknologi tersebut, data berpotensi untuk diolah dalam mendapatkan komunitas-komunitas yang mungkin berada di dalam data. Hasil dari skripsi ini adalah sebuah kesimpulan mengenai berbagai macam bentuk komunitas yang dapat terbentuk pada data Twitter beserta manfaat yang dapat digunakannya.

¹https://blog.twitter.com/official/en_us/a/2012/twitter-turns-six.html

²<https://www.tweepy.org/>

1.2 Rumusan Masalah

Adapun rumusan masalah dari deskripsi yang telah dipaparkan adalah:

1. Bagaimana cara mengumpulkan data Twitter?
2. Bagaimana cara mengeksplorasi dan menyiapkan data Twitter?
3. Bagaimana cara mendeteksi komunitas dari data graf yang dibentuk dari data Twitter?
4. Bagaimana cara mengevaluasi komunitas yang didapatkan dari hasil analisis?
5. Bagaimana cara menginterpretasikan hasil dari analisis yang telah didapat?
6. Bagaimana memanfaatkan hasil analisis data Twitter?

1.3 Tujuan

Adapun tujuan penelitian dari rumusan masalah yang telah dipaparkan adalah:

1. Melakukan pengumpulan data Twitter.
2. Melakukan eksplorasi dan penyiapan data pada data Twitter.
3. Mendeteksi komunitas yang terdapat pada data graf.
4. Mengevaluasi komunitas yang didapatkan dari hasil analisis.
5. Menginterpretasikan hasil dari analisis yang telah didapatkan.
6. Merumuskan pemanfaatan hasil analisis dari data Twitter.

1.4 Batasan Masalah

Batasan yang digunakan dalam penelitian ini adalah:

1. Pemrosesan data tidak bersifat *stream processing*, yaitu suatu pemrosesan kepada suatu data baru secara berkelanjutan untuk mendapatkan suatu hasil [2].
2. Proses *profiling* akun-akun komunitas tidak dianalisis menggunakan *Natural language processing* (NLP) pada *content* (isi) dari *tweet* atau deskripsi akun-akun tersebut, tetapi dilakukan secara manual.

1.5 Metodologi

Metodologi untuk menyelesaikan masalah dalam penelitian ini adalah:

1. Studi Literatur
Pencarian terhadap referensi-referensi yang dapat digunakan sebagai acuan dalam mempelajari permasalahan, teori, dan peralatan-peralatan yang akan digunakan.
2. Mengumpulkan Data untuk Eksperimen
Mengumpulkan beberapa data untuk eksplorasi dan eksperimen tahap awal pada data Twitter. Data sintetik akan dibuat untuk memastikan hasil dari teknologi yang digunakan adalah benar.
3. Menyiapkan Data untuk Eksperimen
Mengubah dan menyajikan data agar data dapat digunakan dengan baik oleh teknologi-teknologi yang akan digunakan.
4. Percobaan Menggunakan Data Sintetik
Percobaan menggunakan data sintetik dilakukan untuk memastikan hasil dari teknologi yang digunakan adalah benar.
5. Merancang Metode untuk Menganalisis Graf
Membuat rancangan bagaimana langkah-langkah memproses data yang dimiliki hingga mendapatkan hasil analisis yang akan diinterpretasi untuk mendapatkan informasi.
6. Eksplorasi dan Percobaan untuk Mengidentifikasi Komunitas
Melakukan eksplorasi dan percobaan terhadap teknologi yang dimiliki dengan menggunakan beberapa *dataset* untuk mengetahui lebih dalam bagaimana cara teknologi tersebut bekerja.

Dan melakukan percobaan terhadap beberapa *dataset* tersebut untuk mencoba mendapatkan komunitas yang mungkin berada didalamnya.

7. Menyiapkan Data untuk Besar

Mengubah dan menyajikan data besar dengan metode yang telah dilakukan pada tahapan eksperimen supaya data dapat digunakan dengan baik oleh teknologi-teknologi yang akan digunakan.

8. Analisis Pada Data Twitter untuk Mencari Komunitas

Penerapan metode yang untuk menganalisis graf pada data besar untuk mendapatkan hasil.

9. Evaluasi Hasil Analisis

Melakukan evaluasi pada hasil yang didapatkan untuk mengidentifikasi komunitas-komunitas yang terbentuk.

10. Menginterpretasikan Hasil Analisis

Hasil dari analisis tersebut akan diinterpretasi untuk mencoba mendapatkan jawaban dari beberapa permasalahan yang sudah dirumuskan sebelumnya.

11. Penulisan Dokumen

Pembuatan dokumen penelitian.

1.6 Sistematika Pembahasan

Sistematika pembahasan untuk menyelesaikan masalah pada penelitian ini adalah:

1. Pada Bab 2 akan diberikan penjelasan dan contoh dari teori-teori dan teknologi-teknologi yang digunakan pada penelitian ini. Hal-hal yang akan diberikan penjelasan antara lain mengenai Twitter, komunitas, graf, algoritma analisis graf, Big Data, Hadoop dan MapReduce, Scala, dan Spark.
 - (a) Pada Sub-bab 2.1 mengenai Twitter akan dijelaskan mengenai beberapa hal berikut:
 - i. Definisi dari Twitter.
 - ii. Kegunaan dari Twitter.
 - iii. Bentuk-bentuk dari *Tweet*.
 - (b) Pada Sub-bab 2.2 mengenai komunitas akan diberikan penjelasan mengenai hal-hal berikut:
 - i. Berbagai definisi yang mendefinisikan suatu komunitas.
 - ii. Apa yang dimaksud dengan *Sense of Community* (Rasa Kebersamaan).
 - (c) Sub-bab 2.3 mengenai graf akan menjelaskan hal-hal berikut ini:
 - i. Tipe dan struktur dari graf.
 - ii. Berbagai karakteristik dari graf.
 - (d) Pada Sub-bab 2.4 mengenai algoritma analisis graf akan dijelaskan dua buah algoritma utama, yaitu:
 - i. Connected Component.
 - ii. PageRank.
 - (e) Sub-bab 2.5 akan menjelaskan berbagai hal mengenai Big Data, seperti berikut:
 - i. Definisi dan Karakteristik dari Big Data.
 - ii. Pengaruh Big Data.
 - iii. Penggerak Big Data.
 - iv. Teknologi pada Big Data.
 - (f) Sesuai dengan namanya, Sub-bab 2.6 akan menjelaskan mengenai dua buah utama yaitu:
 - i. Hadoop.
 - ii. MapReduce.
 - (g) Sub-bab 2.7 mengenai bahasa pemrograman Scala, akan menjelaskan hal-hal berikut:
 - i. Tipe Data pada Scala.
 - ii. Variabel pada Scala.
 - iii. Operator pada Scala.

- (h) Pada Sub-bab 2.8 akan dijelaskan hal-hal berikut:
 - i. DataFrame pada Spark.
 - ii. Partisi pada Spark.
 - iii. *Lazy Evaluation*.
 - iv. Arsitektur Spark.
 - v. GraphFrames.
- 2. Bab 3 akan dilakukan berbagai eksperimen pada beberapa data. Juga akan diberikannya penjelasan mengenai salah satu teknologi yang digunakan untuk memvisualisasikan hasil yang didapat, yaitu Cytoscape. Pada Bab 3 ini akan terdapat 5 buah sub-bab sebagai berikut ini:
 - (a) Pada Sub-bab 3.1 akan diberikan penjelasan mengenai berbagai hal tentang teknologi ini. Mulai dari cara memuat suatu data untuk divisualisasikan pada Cytoscape, hingga melakukan pengaturan terhadap tampilan.
 - (b) pada Sub-bab 3.2 akan diberikan deskripsi dan penjelasan mengenai tiga buah dataset, yaitu dataset berikut:
 - i. *Dataset Pemanasan Global*.
 - ii. *Dataset Video Game*.
 - iii. *Dataset Olahraga*.
 - (c) Sub-bab 3.3 diberikan penjelasan mengenai dua buah hal utama, yaitu:
 - i. Eksplorasi data.
 - ii. Penyiapan data graf.
 - (d) Pada Sub-bab 3.4 akan dilakukannya 4 buah eksperimen, yang dilakukan pada 4 buah data yang berbeda.
 - i. *Dataset Sintetik*.
 - ii. *Dataset Pemanasan Global*.
 - iii. *Dataset Video Game*.
 - iv. *Dataset Olahraga*.
 - (e) Sub-bab 3.5 akan diberikan ringkasan dari hasil eksperimen-eksperimen yang telah dilakukan.
- 3. Bab 4 akan dilakukannya analisis pada data Twitter untuk melakukan pencarian komunitas. Pada bagian ini akan diberikan penjelasan dari pengumpulan data yang digunakan hingga berbagai *insights* yang didapatkan dari hasil analisis.
 - (a) Pengumpulan data.
 - (b) Penyiapan data.
 - (c) Analisis hasil eksperimen pada interaksi *retweet*. Pada bagian ini, akan dilakukan analisis terhadap setiap *component* (komunitas) yang didapatkan secara visual dan menggunakan algoritma Page Rank.
 - (d) Analisis hasil eksperimen pada interaksi *reply*. Pada bagian ini, akan dilakukan analisis terhadap setiap *component* yang didapatkan secara visual dan menggunakan algoritma Page Rank.
 - (e) Pada Sub-bab 4.6 analisis hasil eksperimen pada interaksi *quote*, akan diberikan penjelasan terhadap hasil analisis pada komunitas yang terbentuk secara visual dan menggunakan algoritma Page Rank.
 - (f) Sub-bab 4.7 akan dipaparkan berbagai *insights* dan manfaat yang dapat digunakan beserta contoh-contoh kasus pemanfaatan dari *insights* tersebut.