

BAB 2

TINJAUAN PUSTAKA

2.1 Kerangka Teori

Tahap ini bertujuan untuk memberikan informasi dan teori – teori yang berkaitan dengan topik yang diangkat pada tugas akhir ini.

2.1.1 Media Sosial Twitter

Menurut kamus Cambridge sosial media adalah situs web dan program komputer yang memungkinkan orang berkomunikasi dan berbagi informasi di internet menggunakan komputer atau ponsel. Sedangkan menurut Philip Kotler dan Kevin Lane Keller Media sosial adalah media yang digunakan oleh konsumen untuk berbagi teks, gambar, suara, dan video informasi baik dengan orang lain maupun perusahaan dan vice versa [9].

Twitter adalah jejaring sosial atau *microblog* yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan (*tweet*). *Twitter* didirikan pada bulan maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan juli 2006. *Twitter* telah menjadi alat yang menarik bagi berbagai kalangan untuk mengikuti keinginan para pengguna terhadap setiap kondisi secara langsung. Hal ini menjadi sumber data yang potensial untuk digunakan oleh jutaan orang. *Twitter* membuat segala tersedia dalam sebuah *data stream*, yang dapat dimanfaatkan dengan menggunakan teknik *stream mining*. Secara prinsip hal ini membuat kita mengetahui opini publik secara umum [10]. *Micro-blogging* adalah sumber informasi baru yang menarik untuk pengolahan data [11].

2.1.2 Corona Virus Desease 2019 (COVID-19)

COVID-19 merupakan penyakit yang diidentifikasi menyerang saluran pernapasan. Penyakit ini pertama kali dideteksi kemunculannya di Wuhan,

Tiongkok. Sebagaimana diketahui bahwa COVID-19 bukanlah jenis virus baru.

Akan tetapi dalam penjelasan ilmiah suatu virus mampu bermutasi membentuk susunan genetik yang baru. DNA dari COVID-19 ini memiliki kemiripan dengan DNA pada hewan kelelawar. Diyakini pula bahwa virus ini muncul dari pasar basah (*wet market*) di Wuhan, dimana dijual banyak hewan langka Asia dari berbagai jenis bahkan untuk menjaga kesehatannya ada yang dipotong langsung di pasar agar dibeli dalam keadaan segar. Kemudian pasar ini dianggap sebagai tempat berkembang biaknya virus akibat dekatnya interaksi hewan dan manusia. WHO mengumumkan COVID-19 sebagai pandemi pada tanggal 11 Maret 2020 [12].

2.1.1 Visualisasi Data

Data yang banyak, kompleks, serta tidak teratur akan sulit untuk dimengerti oleh manusia. Membaca data akan menjadi masalah ketika data secara inkremental bertambah atau berubah. Tujuan visualisasi data adalah menunjukkan mana hal-hal yang penting secara jelas, melihat pola, dan menunjukkan informasi dari berbagai dimensi. Visualisasi yang efektif harus dapat menjelaskan data yang disajikan dengan baik dan dapat membangkitkan rasa ingin tahu dari seorang pembacanya [13].

Visualisasi data tidak semata-merta mengubah data mentah menjadi grafik visual, visualisasi data memerlukan perencanaan. Menurut Jack Dougherty dan Ilya Ilyankoi dalam buku "*Hands-On Data Visualization Interactive Storytelling from Spreadsheets to Code*" Visualisasi data secara luas didefinisikan sebagai metode pengkodean informasi kuantitatif, relasional, atau spasial ke dalam gambar [14]. Sedangkan Menurut buku "*Fundamentals of Data Visualization*" data visual adalah bagian dari *art* dan *science*, yaitu bagaimana kita mendapatkan seni yang tepat tanpa membuat si bagian *science* salah, ataupun sebaliknya. Utamanya data visual harus menyampaikan data dengan akurat dan pada saat yang sama harus estetik [15]. Sejalan dengan buku "*Fundamentals of Data Visualization*" menurut buku "*Storytelling With Data : a data visualization guide for business*

professionals” visualisasi data bertempat duduk di persimpangan antara *science* dan *art* [16].

2.1.1 Proses Visualisasi Data

Proses visualisasi data pada penelitian ini mengikuti tahapan yang dimuat dalam buku “*Visualizing Data*” oleh Ben Fry, adapun tahapan dalam visualisasi data menurut buku tersebut sebagai berikut [8]:

1. *Aquire*

Tahap ini adalah tahap dimana data dikumpulkan dari berbagai sumber. Tahap ini menjelaskan bagaimana suatu data diperoleh, apakah data diambil dari sebuah berkas atau data berasal dari internet.

2. *Parse*

Tahap ini adalah tahap dimana data dikemas dalam struktur tertentu dan dikelompokkan kedalam suatu kageteri.

3. *Filter*

Tahap ini adalah proses dimana data yang tidak berhubungan dengan informasi yang dibutuhkan akan disampaikan untuk dihilangkan.

4. *Mine*

Tahap ini adalah tahap dimana menggunakan metode statistika atau data mining untuk jalan mencari pola atau dijabarkan kedalam konteks matematis.

5. *Represent*

Tahap ini adalah tahap dimana data diubah menjadi data yang berbentuk visual dasar, seperti tabel, bar, list atau tree.

6. *Refine*

Pada tahap ini adalah tahap dimana model visual dibuat lebih jelas lagi dan menarik.

7. *Interact*

Tahap ini adalah dimana penggunaan metode untuk memanipulasi data ditambahkan, atau pengguna dapat menampilkan data yang diinginkan.

2.1.2 Konsep Data

Menurut kamus Cambridge data adalah kumpulan dari fakta-fakta atau angka yang digunakan untuk membantu pengambilan keputusan. Data adalah

representasi fakta dunia nyata yang mewakili suatu objek seperti manusia (pegawai, siswa, pembeli, pelanggan), barang, hewan, peristiwa, konsep, keadaan, dan sebagainya, yang diwujudkan dalam sebuah bentuk angka, huruf, simbol, teks, gambar, bunyi, atau kombinasinya [13].

2.1.1 Basis Data

Basis Data adalah sekumpulan data yang saling berhubungan satu sama lain, yang diorganisir sedemikian rupa yang nantinya akan digunakan kembali dengan cepat. Tujuan dalam adanya basis data adalah untuk mempercepat pengambilan data, efisiensi ruang penyimpanan, akurat, ketersediaan, kelengkapan, keamanan dan bersifat multiuser. Adapun sistem yang akan mengelola data tersebut biasa disebut DBMS (*Database Management System*) [17].

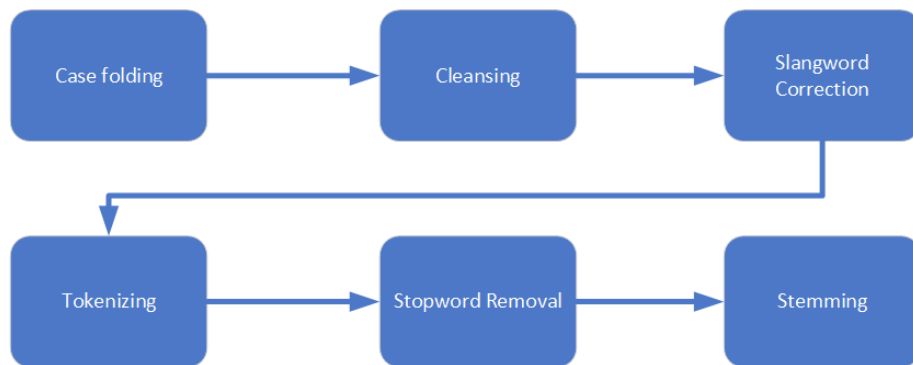
2.1.2 Statistika Deskriptif

Statistika deskriptif adalah salah satu ilmu yang menjelaskan tahapan dalam pengumpulan, mengolah, penyajian dan menganalisis data kuantitatif secara deskriptif, sehingga menghasilkan informasi yang lebih jelas. Statistika deskriptif yaitu terdiri dari [18]:

- a. Menampilkan data dalam bentuk tabel, gambar, diagram, atau grafik.
- b. Mengukur mean, median, dan modus.
- c. Ukuran dispersi atau deviasi.

2.1.3 Text Processing

Tahap *text processing* diperlukan untuk membersihkan sumber data dari data yang tidak diperlukan [19]. Proses ini bertujuan agar data yang digunakan nantinya bersih dari sesuatu yang tidak berpengaruh atau *noise*. *Text processing* juga bertujuan agar data yang digunakan memiliki dimensi lebih kecil sehingga bisa diolah lebih lanjut. tahapan *text processing* yang dilakukan yaitu *case folding*, *cleansing*, *slangword correction*, *tokenizing*, *stopword removal*, dan *stemming*. Tahapan *text processing* dapat dilihat pada gambar 2.1.



Gambar 1.1 Tahapan *text processing*

Penjelasan dari masing-masing tahapan *text processing* yaitu :

1. *Case Folding*

Case folding merupakan proses *text processing* yang dilakukan untuk menyeragamkan *case* karakter pada dokumen. pada proses ini, semua huruf besar (*uppercase*) dijadikan huruf kecil (*lowercase*).

2. *Cleansing*

Cleansing merupakan proses *text processing* yang dilakukan untuk menghapus komponen yang tidak penting atau tidak berpengaruh dalam proses klasifikasi. Komponen yang dihapus yaitu *username*, *URL*, *special character*, dan *hashtag*.

3. *Slangword Correction*

Slangword correction merupakan proses *text processing* yang dilakukan untuk mengganti kata-kata slang menjadi kata yang lebih baku.

4. *Tokenizing*

Tokenizing merupakan proses *text processing* yang dilakukan untuk memisahkan atau memotong dokumen menjadi bagian-bagian yang disebut dengan token.

5. *Stopword Removal*

Stopword removal merupakan proses *text processing* yang dilakukan untuk menghilangkan kata yang dianggap tidak berpengaruh dalam menentukan kategori dalam suatu dokumen.

1. *Stemming*

Stemming merupakan proses *text processing* yang dilakukan untuk mengubah kata menjadi kata dasar.

2.1.1 Algoritma Nazief dan Adriani

Algoritma Nazief & Adriani dikembangkan pertama kali oleh Bobby Nazief dan Mirna Adriani. Algoritma ini berdasarkan pada aturan morfologi Bahasa Indonesia yang luas, yang dikumpulkan menjadi satu grup dan di-*enkapsulasi* pada imbuhan/*affixes* yang diperbolehkan (*allowed affixes*) dan imbuhan/*affixes* yang tidak diperbolehkan (*disallowed affixes*). Algoritma ini menggunakan kamus kata dasar dan mendukung recoding, yakni penyusunan kembali kata-kata yang mengalami proses *stemming* berlebih [20].

Langkah-langkah algoritma Nazief & Adriani adalah:

1. Kata yang belum di-*stemming* dicari pada kamus, jika ditemukan, kata tersebut dianggap sebagai kata dasar yang benar dan algoritma dihentikan.
2. Hilangkan *Inflectional suffixes*, yaitu dengan menghilangkan *particle* (“-lah”, ”-kah”, “-tah” atau “-pun”), kemudian hilangkan *inflectional possessive pronoun suffixes* (“-ku”, “-mu” atau ”-nya”). Cek kata di dalam kamus kata dasar, jika ditemukan, algoritma dihentikan, jika tidak lanjut ke langkah 3.
3. Hapus *Derivational Suffix* (“-i” atau ”-an”,”). Jika kata ditemukan dalam kamus kata dasar, maka algoritma berhenti. Jika tidak, maka lanjut ke langkah 3a:
 - a. Jika akhiran “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
 - b. Akhiran yang dihapus (“-i”, “- an” atau “-kan”) dikembalikan, lanjut ke langkah 4.

1. Hapus *Derivational Prefix* (“be-”, “di-”, “ke-”, “me-”, “pe-”, “se-” dan “te-“).
Jika kata yang didapat ditemukan didalam database kata dasar, maka proses dihentikan, jika tidak, maka lakukan *recoding*.
Tahapan ini dihentikan jika memenuhi beberapa kondisi berikut:
 - a. Terdapat kombinasi awalan dan akhiran yang tidak diijinkan
 - b. Awalan yang dideteksi sama dengan awalan yang dihilangkan sebelumnya.
 - c. Tiga awalan telah dihilangkan
2. Jika semua langkah telah dilakukan tetapi kata dasar tersebut tidak ditemukan pada kamus, maka algoritma ini mengembalikan kata yang asli sebelum dilakukan *stemming*.

2.1.1 NoSQL

NoSQL merupakan istilah yang digunakan untuk menyimpan data yang tidak memakai konsep relasional [21]. NoSQL tidak dibangun di atas tabel, sehingga NoSQL tidak bisa memakai SQL (Structured Query Language) untuk manipulasi datanya [21].

NoSQL memiliki beberapa tipe penyimpanan. Tipe-tipe penyimpanan dari basis data NoSQL, yaitu:

1. *Coloumn-Oriented*

Basis data NoSQL *Coloumn-Oriented* menyimpan data sebagai kolom yang terbalik dengan baris pada basis data relasional

2. *Document Store*

Basis data NoSQL *document store* memperbolehkan manipulasi dari data yang semi-structured. format yang digunakan biasanya yaitu XML, JSON, BSON atau YAML.

3. *Key Value Store*

Basis data NoSQL *key value store* sangat mirip dengan *document store*. *key value store* memungkinkan penyimpanan sebuah nilai terhadap sebuah kunci. perbedaan *key value store* dan *document store* yaitu *key value store*

mengharuskan penentuan kunci dan kunci digunakan untuk menerima data/nilai.

1. *Graph*

Basis data NoSQL *graph* menginterpretasikan hubungan data di dalam grafik. sebuah simpul di dalam grafik memungkinkan memiliki lebih dari satu tautan yang menggambarkan hubungan data.

2.1.1 Analisis dan Desain Berorientasi Objek

Analisis dan Desain Berorientasi Objek (Object Oriented Analysis and Design) adalah cara baru dalam memecahkan suatu masalah dengan model yang dibuat menurut konsep. Dasar pembuatannya sendiri adalah objek yang merupakan kombinasi antara struktur data dan perilaku dalam satu entitas. Adapun dalam pemrograman berorientasi objek berbagai konsep seperti: *class, object, abstract, encapsulation, polymorphism, inheritance* dan UML (*Unified Modeling Language*). UML sendiri merupakan salah satu standar Bahasa yang banyak digunakan di dunia industri untuk mendefinisikan kebutuhan (requirement), membuat rancangan analisis & desain, serta menggambarkan arsitektur dalam pemrograman berbasis objek. Terdapat beberapa diagram yang digunakan proses pembuatan perangkat lunak berorientasi objek diantaranya, *use case diagram, activity diagram, class diagram* dan *sequence diagram* [22].

2.1.2 Usability Testing

Usability testing adalah metode evaluatif yang memungkinkan tim untuk mengamati pengalaman individu dengan aplikasi digital. Metode ini dirancang untuk membantu identifikasi bagian antarmuka yang paling membingungkan orang sehingga bisa diperbaiki dan diuji ulang sebelum diluncurkan [23].

2.1 Kerangka Pikir

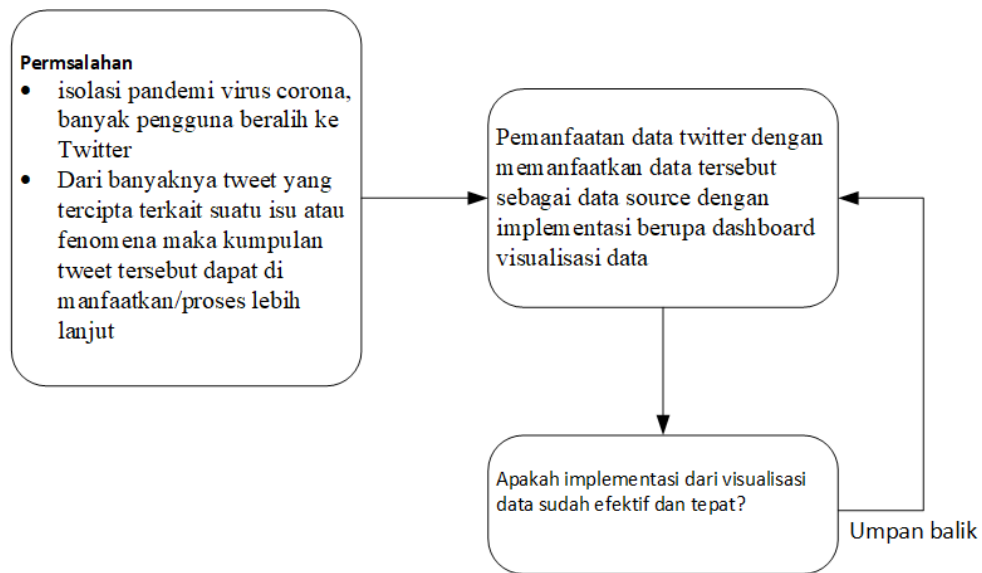
Perkembangan media sosial semakin pesat. Jumlah pengguna media sosial di Indonesia ini meningkat sebesar 12 juta atau 8,1% antara April 2019 dan Januari

2020. Dari beberapa media sosial yang ada, Twitter merupakan salah satu media sosial yang populer digunakan dan pada Januari 2020, Indonesia tercatat masuk

dalam peringkat kedelapan pengguna twitter terbanyak. Di Indonesia sendiri twitter digunakan sebagai platform penyebaran berita, ini dapat dilihat dari tagar-tagar trending topic Indonesia yang selalu terkait dengan isu atau fenomena yang sedang terjadi saat itu.

Di saat isolasi pandemi virus corona, banyak pengguna media sosial lain kembali beralih ke Twitter. Twitter sebagai sumber berita dan hiburan telah membantu meningkatkan keterlibatan selama pandemi karena konsumen yang tinggal di rumah menggunakan platform untuk berita dan informasi secara real-time.

Dari banyaknya tweet yang tercipta terkait suatu isu atau fenomena, maka kumpulan tweet tersebut dapat di proses lebih lanjut dengan *knowledge discovery*, sehingga dapat menjadi sebuah pengetahuan baru yang nantinya akan ditampilkan atau disajikan melalui visualisasi data yang tepat. Dari apa yang dapat dilihat pada media sosial twitter saat ini, publik banyak membicarakan isu terkait COVID-19, baik isu COVID-19 yang kaitannya respon terhadap pemerintah atau sekedar opini terkait keadaan akibat pandemi ini. Karena banyaknya data terkait isu ini dan bertujuan untuk memanfaatkan data ini, peneliti membuat visualisasi data yang dataset atau sumber data nya diambil dari twitter. Berdasarkan tinjauan kepustakaan dan penelitian terdahulu yang telah diuraikan oleh peneliti, kerangka pemikiran melalui paradigma penelitian digambarkan sebagai berikut :



Gambar 2.2 Paradigma Penelitian

2.1 Hipotesis

Hipotesis adalah dugaan awal atau jawaban sementara terhadap masalah penelitian yang kebenarannya harus diuji secara empiris dan dibuktikan melalui penelitian.

Berdasarkan pada kerangka pemikiran teoritis diatas, maka hipotesis penelitian yang diajukan dalam penelitian ini

1. Implementasi visualisasi data berupa dashboard telah tepat dalam pemanfaatan data twitter.
2. Visualisasi data yang dibentuk sudah tepat untuk segmen pengguna yang dituju.