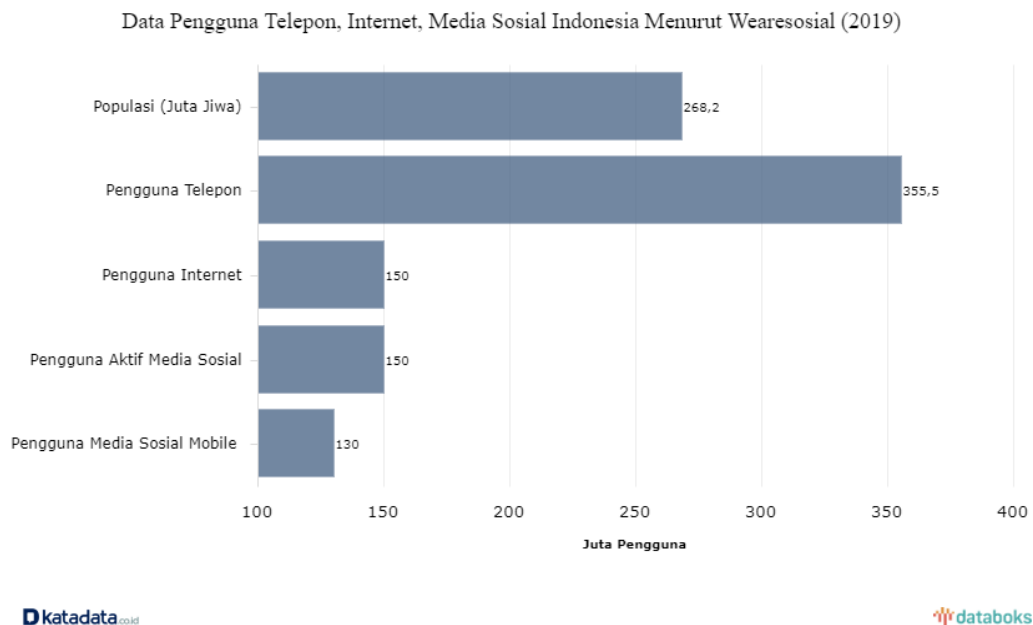


Bab I PENDAHULUAN

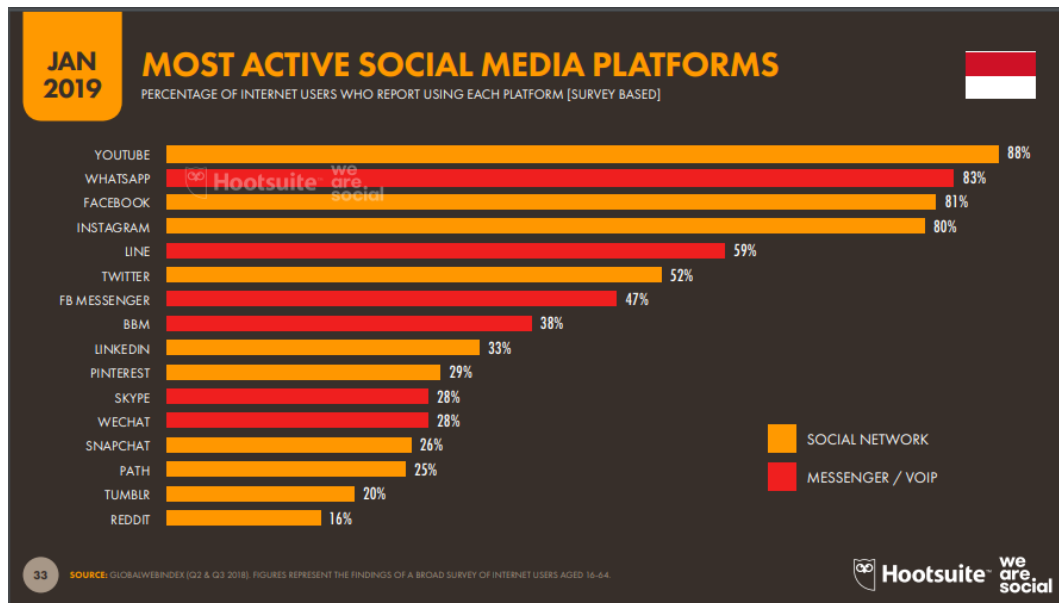
Bab pendahuluan menjelaskan mengenai dasar yang perlu diketahui dalam melakukan penelitian ini. Selain itu, pemaparan latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah dari dilakukannya penelitian dijelaskan dalam bab pendahuluan.

I.1 Latar Belakang

Berdasarkan hasil riset Wearesosial dan Hootsuite yang dirilis januari 2019, pengguna media sosial di Indonesia mencapai 150 juta atau sebesar 56% dari total populasi. Hasil survey GlobalWebIndex pada pengguna internet di Indonesia dalam rentang usia 16-64 tahun, menunjukkan bahwa ada beberapa platform media sosial yang aktif digunakan oleh masyarakat Indonesia. Salah satunya adalah instagram. Menurut hasil survei WeAreSocial.net dan Hootsuite, instagram merupakan platform media sosial dengan jumlah pengguna terbanyak ke empat di Indonesia. (Riyanto, 2019). Selain sebagai jejaring sosial untuk berbagi foto, instagram digunakan untuk pemasaran produk bisnis.



Gambar I.1 Grafik data pengguna telepon, internet, media sosial Indonesia



Gambar I.2 Grafik pengguna aktif platform media sosial yang paling banyak digunakan

Pemasaran melalui media sosial adalah kegiatan yang berurusan dengan informasi seperti mengevaluasi data konsumen, menilai tingkat konversi, dan merencanakan konten media sosial yang paling efektif berdasarkan temuan dari data yang dianalisis. Menurut Walter Moore seorang konsultan manajemen terkemuka dan pakar pemasaran digital mengatakan pengguna instagram hampir 95 juta melakukan unggahan baru setiap harinya. Pada instagram terdapat alat seperti *sosial insight* yang memberi akses ke analisis terperinci untuk memahami strategi pemasaran instagram yang tidak terdapat pada *platform* media sosial lainnya. Pada instagram memungkinkan untuk mempelajari aspek-aspek unik seperti demografis dan mengoptimalkan upaya pemasaran untuk *audience* tertentu. Hal tersebut dapat dilakukan dengan cara melakukan analisis data. Namun banyaknya data yang diambil dari berbagai sumber membuat data yang dihasilkan memiliki berbagai format seperti data terstruktur, data tidak terstruktur atau data semi-terstruktur (Jagdale, Sonawane, & Khan, 2014). Sebagian besar data dengan format itu berasal dari web yang baru dibuat (Mahanama, et al., 2018).

Data dengan format tidak terstruktur mulai tumbuh lebih cepat seiring dengan perubahan zaman di era teknologi (Bavota, 2016). Alasan besar jumlah data yang muncul dalam berbagai macam format terutama disebabkan oleh evolusi situs jejaring sosial seperti Facebook, Instagram, Twitter, dan lain-lain. Evolusi ini

memungkinkan pengguna untuk membuat konten secara bebas sehingga menyebabkan ekspansi dalam volume data melalui internet (Jagdale, Sonawane, & Khan, 2014). Perluasan data yang muncul dari berbagai format data yang tidak terstruktur dan semi-terstruktur membuat sulit untuk melakukan analisis data. Data yang memiliki format tidak terstruktur atau semi-terstruktur akan sulit untuk dianalisis, sehingga sulit untuk menghasilkan informasi yang berharga.

Data merupakan komponen penting berisi informasi yang dapat mempengaruhi suatu organisasi dalam pengambilan keputusan (Fávero & Belfiore, 2019). Pada dasarnya sebuah data dapat menjelaskan fakta dan angka-angka yang di olah oleh suatu perusahaan setiap harinya. Data menjadi fakta dan statistik sederhana yang dikumpulkan. Data dapat digunakan untuk mengukur atau mencatat berbagai kegiatan baik internal maupun eksternal. Data menjadi dasar untuk semua pelaporan dikarenakan data menjadi suatu hal yang sangat penting. Tanpa data tidak dapat membuat perencanaan dan implementasi kegiatan dengan tepat sasaran. Semakin berkembangnya informasi akan menimbulkan semakin banyaknya format dan bentuk data yang dihasilkan, untuk mengolah sebuah data agar dapat menciptakan sebuah pengetahuan atau *knowledge* yang berguna dalam meningkatkan nilai bisnis suatu perusahaan perlu adanya pengolahan terhadap suatu data. Proses pengolahan data agar dapat menghasilkan informasi berharga berada pada serangkaian prosedur untuk melakukan *data mining* (Jagdale, Sonawane, & Khan, 2014).

Data mining adalah serangkaian proses dalam studi teknik berbasis data untuk menemukan pola pada data mentah dengan volume yang besar. *Data mining* adalah proses menemukan informasi yang bermanfaat dan mendapatkan pola dengan menggunakan algoritma *data mining* tertentu. *Data mining* berbicara tentang perspektif yang berbeda dari apa yang dianalisis untuk mendapatkan informasi yang bermanfaat. *Data mining* menggunakan proses *Knowledge Discovery in Database* (KDD) yang melibatkan *data cleaning*, *data integration*, *data selection and data transformation* dalam proses pengolahan data. Langkah-langkah tersebut berada pada tahap *data preprocessing* sebelum *data mining* dilakukan (Jagdale, Sonawane, & Khan, 2014).

Data preprocessing adalah solusi yang dapat digunakan untuk mengatasi data yang umumnya terjadi pada saat setelah *data collecting* dilakukan terhadap sebagian besar situs jejaring sosial. *Data preprocessing* perlu dilakukan untuk menghasilkan informasi yang berharga (Paul, Vishwakarma, & Verma, 2015). *Data preprocessing* dilakukan setelah *file log* beralih ke *file database*. Tujuan dari dilakukannya tahap *data preprocessing* ini yaitu untuk menghapus hal-hal yang tidak dibutuhkan atau *noise* dari catatan log. Proses ini terutama meliputi tahap penting pada *data cleaning* (Han & Xia, 2014). *Data preprocessing* menjadi langkah penting dalam pengolahan data sebelum terjadi proses *mining* (Guan, Ji, Qian, Ma, & Hong, 2017). Karena kunci agar sebuah data dapat memiliki makna yang bernilai bergantung pada bagaimana data tersebut diolah, dan hal tersebut terjadi pada proses *data preprocessing* (Wei & Tian, 2012). *Data preprocessing* memproses data mentah untuk dijadikan *dataset* yang bernilai bagi pemilik kepentingan. Proses menemukan pola dalam data dimungkinkan setelah *dataset* besar telah divalidasi (Paul, Vishwakarma, & Verma, 2015).

Dataset yang sudah tervalidasi memungkinkan untuk ditampilkan secara visual dalam bentuk grafik. Dari sini dapat diketahui *insight* apa yang akan dihasilkan dari *dataset* yang digunakan. Dari *insight* tersebut dapat membantu perusahaan atau organisasi dalam pengambilan keputusan yang valid sesuai data yang ada.

Penelitian ini membahas perancangan *data preprocessing* berbasis *website* menggunakan teknik *web scraping* untuk menghasilkan visualisasi *dataset* pada media sosial instagram. Objek pada penelitian ini adalah *website* media sosial instagram dengan akses publik. Diharapkan dengan adanya *website* sebagai *tools* untuk melakukan *data preprocessing* dengan teknik *web scraping* mampu menghasilkan *insight* dari visualisasi data menggunakan *tools* tableau berdasarkan *dataset* yang dihasilkan agar dapat memberikan informasi berharga untuk pengambilan keputusan atau untuk penelitian selanjutnya pada data mining.

I.2 Perumusan Masalah

Berdasarkan uraian masalah yang telah dijelaskan pada latar belakang. Permasalahan yang akan dikaji dalam penelitian adalah sebagai berikut:

1. Bagaimana melakukan pengumpulan data pada media sosial instagram dengan teknik web *scraping* untuk menghasilkan *dataset*?
2. Bagaimana melakukan *data preprocessing* dari hasil web *scraping* media sosial instagram hingga menjadi sebuah *dataset*?
3. Bagaimana melakukan visualisasi data menggunakan *dataset* dari media sosial instagram untuk menghasilkan sebuah *insight* yang bernilai bagi pengambilan keputusan.

I.3 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut:

1. Melakukan pengumpulan data pada media sosial instagram dengan teknik web *scraping* untuk menghasilkan *dataset*.
2. Melakukan *data preprocessing* dari hasil web *scraping* media sosial instagram hingga menjadi sebuah *dataset*.
3. Melakukan visualisasi data menggunakan *dataset* dari media sosial instagram untuk menghasilkan sebuah *insight* yang bernilai bagi pengambilan keputusan.

I.4 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat membawa beberapa manfaat untuk industri/objek TA yang dikaitkan dengan kepentingan bisnis serta mampu mengkaitkan dengan manfaat dari sisi ilmu pengetahuan pada domain Sistem Informasi diantaranya adalah sebagai berikut:

1. Industri/objek

Secara manfaat, hasil penelitian ini diharapkan dapat menjadi referensi atau saran untuk pengembangan teknologi data preprocessing dengan *tools* untuk melakukan pengelolaan data pada industri/objek. Penelitian ini juga dapat digunakan sebagai referensi atau pengembangan lebih lanjut.

2. Ilmu pengetahuan

Secara manfaat, hasil penelitian ini diharapkan dapat menjadi referensi atau saran pembelajaran untuk pengembangan pada bidang ilmu teknologi khususnya pada domain Sistem Informasi.

I.5 Batasan Masalah

Dalam melakukan penelitian diperlukan batas atau ruang lingkup dalam pengerjaannya. Batasan dalam melakukan penelitian ini adalah sebagai berikut:

1. Tugas akhir ini melakukan pengambilan data pada media sosial instagram dengan akun publik.
2. Tugas akhir ini melakukan *scraping data* pada media sosial instagram dengan menggunakan *tools* perantara berbasis website untuk melakukan *data collection* dan *data preprocessing*.
3. Tugas akhir ini melakukan pengolahan data *JavaScript Object Notation* (JSON) hingga menjadi sebuah *dataset* ternormalisasi dengan format *Comma Separated Values* (CSV).
4. Tugas akhir ini melakukan visualisasi data untuk menghasilkan *insight* sebagai hasil akhir dari penelitian.

I.6 Sistematika Penulisan

Penulisan ini akan dijabarkan dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Bab ini memberikan uraian latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian batasan implementasi, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan kajian literatur pendukung yang relevan dengan permasalahan yang ada dan menjelaskan teori-teori terkait dengan penelitian. Bab kedua juga mengulas kembali beberapa *related work* yang pernah dilakukan oleh peneliti sebelumnya.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai metodologi penelitian. Metodologi penelitian membahas mengenai konseptual dan sistematika dari penelitian yang dilakukan. Pemilihan dan penjelasan dari metode yang digunakan akan dijabarkan pada bab ini.

BAB IV ANALISIS SISTEM

Bab ini berisi tentang model dari sistem atau penelitian yang akan dilakukan. Bab ini menjelaskan alur yang akan dilakukan dalam proses penelitian ini.

BAB V PENGUJIAN SISTEM DAN ANALISIS

Bab ini menjelaskan setiap hasil yang didapat dari penelitian yang sudah dikerjakan beserta dengan analisisnya. Pada bab ini juga menjelaskan implementasi dari sistem dilakukan. Selain itu detail informasi pada sistem juga dijelaskan pada bab ini. Hasil dari pembuatan logika, pengujian sistem, analisa, dan evaluasi termasuk dalam bab ini.

BAB VI KESIMPULAN DAN SARAN

Bab ini menyimpulkan hasil dari dilakukannya penelitian. Saran untuk peneliti selanjutnya dapat dituliskan pada bab ini.