

BAB I

PENDAHULUAN

1.1. Latar Belakang

Ratusan ribu entri data dan *posting*-an terjadi setiap harinya di media sosial, hal ini dikarenakan orang-orang banyak menggunakan media sosial sebagai sarana mereka untuk mengekspresikan pikiran, minat, dan opini mereka atas banyak hal yang terjadi di lingkungannya [1]. Survei menyebutkan bahwa Indonesia memiliki sekitar 150 juta pengguna aktif media sosial hingga tahun 2019 [2]. Hal itu mendorong adanya banyak penelitian untuk menggunakan media sosial sebagai sumber data untuk dianalisis dalam bentuk sebuah analisis sentimen [3], yaitu sebuah proses pemahaman, ekstraksi, dan memproses data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung pada suatu kalimat pendapat [4].

Media sosial seperti Twitter merupakan sebuah sarana *microblogging* yang memungkinkan orang untuk berkomunikasi dengan pesan singkat yang kira-kira berisi suatu pemikiran atau ide secara langsung dan dapat dengan mudah diinterpretasikan pesannya, terima kasih kepada *limit* yang diberikan oleh Twitter pada satu *tweet* yang hanya boleh berisikan kurang lebih 140-280 karakter [5].

Meskipun analisis sentimen banyak populer di kalangan media sosial Twitter, namun juga tidak sedikit juga peneliti yang tertarik untuk melakukan analisis sentimen dari media sosial lain seperti halnya Instagram [6]–[8]. Penelitian yang dilakukan pada Instagram biasanya berfokus pada data visual, meskipun tidak menutup kemungkinan data tekstual pun juga dapat menjadi perhatian [9].

Seperti yang sudah diketahui sebelumnya, media sosial merupakan tempat produksi data yang besar dan sebagian besar data yang ada merupakan data yang tidak terstruktur [10]. Secara sederhana data yang tidak terstruktur merujuk pada fakta atau data yang tidak dapat diletakkan pada tabel (baris-kolom) basis data tradisional dan biasanya berbentuk dokumen, video, ataupun elemen multimedia yang lain [11]. Peneliti lain menyebutkan bahwa data tidak terstruktur dianggap sebagai data yang tidak memiliki skema yang didefinisikan terlebih dahulu atau bahkan tidak memiliki susunan di dalam strukturnya [12]. Untuk menggali dan mengolah data yang tidak terstruktur tersebut lalu mengubahnya menjadi informasi yang bermanfaat untuk bisnis, penggunaan *text mining* mulai populer khususnya di bidang *big data analytics* [13]. *Text mining* terdapat pada lapisan kedua bersamaan dengan *web mining* dan *bio mining* pada *big data mining*, yang dimana memiliki karakteristik 4V, (1) *Variety*, (2) *Velocity*, (3) *Variability*, dan (4) *Veracity* [14]. Melihat karakteristik tersebut, sebuah basis data relasional biasa dan teknik sederhana analisis data menggunakan SQL saja dinilai kurang cukup untuk mendukung pengolahan data khususnya data yang tidak terstruktur seperti ulasan pada media sosial yang terbilang sangat besar dan beragam [15].

Oleh karena itu, penelitian yang berkaitan dengan perancangan basis data yang dapat menangani data tidak terstruktur mulai dilakukan tidak hanya untuk bisa mengelola data tersebut namun juga memastikan konsistensi data seperti halnya yang ada pada basis data relasional [16]. Berdasarkan hal tersebut, penelitian ini akan menggunakan data pada media sosial seperti opini dalam *tweet* dan komen Instagram sebagai sampel evaluasi atas rancangan basis data tidak terstruktur yang nantinya dibangun. Perancangan dimulai dari model konseptual dan model logikal menggunakan pendekatan UML hingga nantinya model diimplementasikan pada salah basis data fisik NoSQL yang telah terkenal dikalangan pengembang aplikasi, MongoDB [17]. Kemudian, model dari hasil penelitian ini akan diuji ke dalam sebuah aplikasi analisis sentimen berbasis web menggunakan pendekatan *black-box testing*. Model analisis sentimen yang digunakan adalah dengan pendekatan leksikon. Pendekatan ini telah diuji dengan hasil yang baik dalam melakukan analisis sentimen pada penelitian sebelumnya [18]. Diharapkan dengan adanya penelitian ini dapat membuka peluang penelitian lainnya ke depan dalam melakukan perancangan, pengembangan, dan evaluasi kinerja basis data dalam menangani data yang tidak terstruktur khususnya media sosial menggunakan metode implementasi lainnya.

1.2. Rumusan Masalah

Masalah-masalah yang ditemukan dan kiranya akan dibahas dalam penelitian kali ini antara lain:

1. Bagaimana hasil rancangan basis data tidak terstruktur untuk menangani data *tweet* dan komen?
2. Bagaimana hasil implementasi basis data tidak terstruktur ke dalam sebuah aplikasi *sentiment analysis*?
3. Bagaimana hasil kinerja basis data tidak terstruktur ketika dievaluasi menggunakan simulasi *black-box testing*?

1.3. Batasan Masalah

Hal-hal yang menjadi batasan pada penelitian kali ini ialah:

1. Penelitian hanya dilakukan untuk merancang dan mengimplementasikan model basis data tidak terstruktur pada aplikasi *sentiment analysis* sederhana dan hanya berfokus pada *Database Development*.
2. Rancangan basis data akan diimplementasikan ke dalam sebuah aplikasi berbasis web.
3. *Text mining* untuk *sampling* dilakukan sebelum implementasi basis data dan aplikasi.
4. Implementasi model *sentiment analysis* yang digunakan terbatas menggunakan pendekatan *lexicon-based* saja.

5. Sistem basis data tidak terstruktur yang digunakan untuk penelitian ini hanyalah menggunakan MongoDB.

1.4. Tujuan Penelitian

Tujuan dari penelitian kali ini adalah untuk merancang, membangun dan mengimplementasi sistem basis data ke dalam sebuah aplikasi berbasis web untuk *sentiment analysis* media sosial Twitter dan Instagram agar dapat digunakan oleh banyak objek dan kasus penelitian ataupun berbagai bidang sesuai kebutuhan perusahaan/organisasi. Pengguna dari aplikasi yang dibangun nanti adalah entitas dari sebuah organisasi/perusahaan yang ingin melakukan analisis terhadap opini pelanggannya atas produk/jasa yang telah diberikan.

1.5. Manfaat Penelitian

Adapun manfaat dari penelitian kali ini antara lain:

1. Menemukan solusi sistem basis data yang tepat untuk sebuah aplikasi *sentiment analysis*.
2. Membantu perusahaan/organisasi dalam mengetahui sikap serta reaksi *customer* terhadap suatu topik tertentu dari komen pada Instagram ataupun *tweet* yang di-*post* dalam Twitter.
3. Mendukung pengambilan keputusan perusahaan dalam menganalisis teks dan opini *customer*.