

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Sistem klasifikasi makna cuitan (selanjutnya disingkat menjadi SKMC) yang dimaksud dalam tulisan ini adalah program komputer yang dibangun menggunakan bahasa pemrograman python dan algoritma SVM (*support vector machine*), bertujuan mengklasifikasikan makna cuitan: bermakna positif, negatif, atau netral. Berdasarkan pada penelitian-penelitian terdahulu, selama ini penelitian terkait pembangunan SKMC masih menggunakan prosedur standar. Prosedur standar yang dimaksud adalah mengumpulkan data cuitan, membagi cuitan sebagai data *training* dan data *testing*, lalu melabeli data dengan kategori positif, negatif, netral dan selanjutnya sistem akan mempelajari pelabelan tersebut. Berdasarkan hasil mempelajari pelabelan data, sistem dapat memprediksi makna cuitan: apakah suatu cuitan masuk dalam klasifikasi bermakna positif, negatif, atau netral.

Jika pada saat pelabelan data terjadi kekeliruan pemaknaan tentu membuat akurasi prediksi SKMC kurang baik. Oleh karena itu, pada proses pelabelan data diperlukan pemaknaan cuitan dengan tepat. Untuk memaknai cuitan dengan tepat, tentunya SKMC memerlukan *database* kosakata bermakna positif, negatif, atau netral.

Terkait dengan pemaknaan cuitan pada SKMC yang telah dijelaskan di atas, dalam penelitian ini dilakukan penelusuran kata-kata yang masuk dalam medan makna menyenangkan dan kata-kata yang berasosiasi menyenangkan dari cuitan-

cuitan menyenangkan. Dari penelusuran ini terbangun *database* kosakata dalam medan makna menyenangkan dan kosakata berasosiasi menyenangkan. *Database* ini dimanfaatkan pada SKMC untuk membantu meningkatkan akurasi prediksi SKMC dalam mengklasifikasikan makna cuitan (bermakna menyenangkan atau tidak). Untuk mempermudah pembahasan dalam tulisan ini, SKMC yang telah dimodifikasi dengan *database* kosakata dalam medan makna menyenangkan dan kosakata berasosiasi menyenangkan tersebut, dalam tulisan ini selanjutnya disingkat menjadi Sidikta (sistem prediksi cuitan).

Pada penelitian ini, penelusuran kosakata dibatasi hanya pada kosakata dalam medan makna menyenangkan dan berasosiasi menyenangkan karena beberapa alasan berikut ini.

- 1) Banyaknya penelitian yang menganalisis cuitan untuk mengetahui respons warganet terhadap suatu produk (suka, senang, atau tidak terhadap suatu produk) sehingga suatu hari *database* medan makna menyenangkan dan kosakata berasosiasi menyenangkan ini akan membantu penelitian semacam itu.
- 2) Terkait dengan makna asosiasi, maka pembicaraan berhubungan dengan medan makna dari suatu kata. Misalnya, kita berbicara tentang *libur*, maka bisa saja muncul asosiasi terhadap *jalan-jalan*, *gunung*, dan *santai*, sehingga kata-kata tersebut bisa dimasukkan dalam satu medan makna libur. Keterkaitan tersebut, menjadi alasan pentingnya menelusuri medan makna menyenangkan dan kosakata berasosiasi menyenangkan dalam penyusunan database-nya.

- 3) Sejauh ini belum ada *database* kosakata dalam medan makna menyenangkan dan berasosiasi menyenangkan.

Penelitian-penelitian terdahulu yang mendukung batasan kajian ini, dan menjadi latar belakang penelitian ini dipaparkan di bawah ini.

1.1.1 Latar Belakang Empiris

Istilah cuitan/twit/kicau didefinisikan sebagai pesan yang dikirim atau diunggah pada layanan jejaring sosial, Twitter. Pesan ini biasanya berupa pendapat atau informasi yang pada umumnya membahas topik atau isu terkini.

Twitter telah dimanfaatkan untuk berbagai keperluan, misalnya sebagai sarana protes. Twitter juga telah diadopsi sebagai alat komunikasi dan pembelajaran di lingkungan pendidikan, seperti di perguruan tinggi (Grosbeck & Holotescu, 2008).

Karena membahas isu terkini dan manfaatnya juga banyak, maka hingga kini Twitter sebagai media sosial masih populer serta masih digemari warganet. Hal ini sesuai dengan tulisan Basri dan Syafrizal (2017) bahwa Twitter merupakan jejaring sosial kedua yang banyak diminati anak muda setelah jejaring sosial Facebook.

Twitter memberikan akses kepada siapapun untuk mengambil cuitan sebagai data penelitian. Twitter bahkan menyediakan sarana berupa API (*application programming interface*) yang bersifat *open source* untuk *crawling* data. Dengan sarana tersebut, pengumpulan data dapat dilakukan dengan waktu yang singkat. Oleh karena itu, cuitan telah banyak digunakan sebagai data dalam

penelitian, misalnya survei tentang elektabilitas calon presiden atau survei tentang sentimen konsumen terhadap suatu produk. Survei-survei tersebut menggunakan cuitan sebagai data karena ribuan cuitan dapat diperoleh dengan mudah dan cepat.

Zain dkk. (2020) dalam penelitiannya mengidentifikasi cuitan-cuitan hoaks yang berhubungan dengan pemilihan presiden 2019 agar masyarakat terhindar dari berita hoaks. Salah satu alasan Twitter dipilih karena banyak pihak yang memanfaatkan media sosial (Twitter) sebagai media kampanye. Hal itu sering dimanfaatkan oleh pihak-pihak yang tidak bertanggung jawab untuk menyebarkan berita hoaks. Oleh karena itu, perlu adanya pendeteksi cuitan-cuitan hoaks.

Selain penelitian di atas, kajian lain yang memanfaatkan cuitan Twitter adalah riset Herdiani dkk. (2016) yang memaparkan pengukuran indeks kebahagiaan masyarakat Kota Bandung. Penelitiannya menggunakan parameter tertentu yang disesuaikan dengan hal-hal yang membuat masyarakat Bandung bahagia. Dalam penelitiannya diusulkan adanya sistem untuk mengukur indeks kebahagiaan menggunakan data dari cuitan dengan perhitungan sesuai dengan parameter-parameter yang ada.

Curini dkk. (2015) memanfaatkan cuitan warganet untuk mengetahui kondisi masyarakat, mereka membangun cara pengukuran tingkat kebahagiaan dengan menggunakan teknik statistik yang inovatif dengan data dari cuitan masyarakat Italia. Penelitian lain terkait penggunaan data cuitan juga telah dilakukan oleh Mitchell dkk. (2013). Hasil penelitiannya menyatakan bahwa media sosial dapat berpotensi untuk memperkirakan tingkat *real-time* dan perubahan dalam mengukur skala populasi. Penelitian yang lain adalah penelitian Ponilan dkk.

(2016) yang menyusun ontologi dari cuitan di Twitter. Dalam penelitiannya dijelaskan penggunaan ontologi untuk menghitung indeks kebahagiaan masyarakat Bandung. Hal itu serupa dengan penelitian Setyawan dkk. (2016), ontologi yang disusun oleh keduanya, baik Ponilan maupun Setyawan, dibatasi pada kata-kata yang termasuk dalam 10 kriteria indeks kebahagiaan, yaitu 1) hubungan sosial, 2) keadaan lingkungan, 3) keharmonisan keluarga, 4) kesehatan, 5) ketersediaan waktu, 6) kondisi keamanan, 7) kondisi rumah dan aset, 8) pekerjaan, 9) pendapatan rumah tangga, dan 10) pendidikan.

Sejalan dengan penelitian di atas, penelitian yang menggunakan cuitan Twitter sebagai data, adalah tulisan Ramadan dan Setiawan (2019) yang menganalisis sentimen warganet terhadap acara di stasiun televisi, SCTV. Penelitian Adiati dkk. (2019) juga menggunakan cuitan Twitter sebagai data untuk menganalisis opini masyarakat terhadap partai politik peserta pemilu 2019. Demikian juga penelitian Bastomy dkk. (2019), penelitiannya menggunakan cuitan warganet sebagai data untuk identifikasi karakter presiden melalui analisis sentimen pada Twitter.

Asriadie dkk. (2017) menggunakan cuitan Twitter sebagai data untuk menganalisis klasifikasi emosi. Dalam penelitiannya dijelaskan bahwa pemilihan cuitan sebagai data karena Twitter merupakan salah satu jejaring sosial yang masih populer di masyarakat hingga saat ini. Twitter dapat digunakan untuk berbagai hal, dari mengobrol, berbagi informasi, hingga berkeluh kesah. Menurut penelitiannya, cuitan dapat memberikan informasi tentang emosi dari penulisnya. Oleh karena itu,

penelitiannya menjadikan Twitter sebagai sumber data dalam penelitiannya terkait klasifikasi emosi.

Selain penelitian di atas, penelitian berikut ini juga memanfaatkan cuitan Twitter sebagai data. Prameswari dan Setiawan (2019) menggunakan cuitan untuk memprediksi klasifikasi kepribadian. Analisis dilakukan melalui kata-kata pada cuitan yang telah diunggah warganet di Twitter.

Seperti penelitian Prameswari dkk, serta Asriadie, dkk. penelitian Rizkia dkk. (2019) juga memanfaatkan cuitan Twitter sebagai data. Dalam hal ini data untuk mengukur kepuasan pelanggan Indihome. Data berupa opini pelanggan terhadap layanan operator tersebut. Analisis sentimen dilakukan untuk mengklasifikasikan opini pelanggan terhadap layanan tersebut ke dalam kelas positif, negatif dan netral, sehingga diperoleh tingkat kepuasan pelanggan.

Alasan pemilihan cuitan warganet dipilih sebagai data penelitian, karena adanya kemudahan dalam akses pengumpulan data dari Twitter. Selain akses yang mudah, data yang dapat diambil juga banyak (bisa mencapai ribuan bahkan jutaan cuitan). Demikian pula dengan penelitian pada tulisan ini, berdasarkan kelebihan Twitter seperti yang telah dijelaskan di atas, cuitan Twitter dipilih untuk menjadi data yang ditelaah pada penelitian ini.

1.1.2 Latar Belakang Teoretis

Salah satu penelitian terkait medan makna telah dilakukan oleh Varlamova dkk. (2017). Dari penelitiannya dapat disimpulkan bahwa medan makna dapat berguna untuk 1) pengayaan kosakata kuantitatif dan kualitatif; 2) membantu dalam

pemilihan kata yang benar, 3) menyusun leksikal, 4) menggunakan lexis berbagai medan makna dalam konteks yang berbeda tanpa kesalahan, 5) menjamin keterampilan bahasa secara sadar dan otomatis dari para siswa, dan 6) meningkatkan kosakata aktif. Secara singkat dari hasil penelitiannya, medan makna membantu proses pembelajaran bahasa kedua.

Selain penelitian di atas, Gitonga dkk. (2018) juga melakukan penelitian terkait medan makna. Kajiannya menemukan bahwa medan makna secara terperinci digunakan dalam khotbah-khotbah gereja Gikūyū. Berdasarkan analisis medan makna, penelitiannya menghasilkan beberapa implikasi berikut: 1) ada manfaat dari tinjauan teoretis yang luas dari medan makna khotbah gereja (yang sebelumnya hanya tampak sepintas), 2) ada konsekuensi metodologis untuk mempelajari medan makna yang membahas khotbah gereja, 3) kemampuan untuk menggunakan medan makna dengan benar dan tepat adalah bagian penting dari kompetensi linguistik, dan 4) ada kemudahan memahami item kosakata yang dimiliki medan makna yang sama karena mampu membentuk pola kata-kata yang saling terkait dalam pikiran seseorang.

Penelitian-penelitian pendahulu tersebut menghasilkan manfaat yang beragam terkait medan makna, sedangkan dalam penelitian ini, *database* kosakata medan makna yang terbentuk dimanfaatkan untuk meningkatkan akurasi Sidikta dalam memprediksi klasifikasi cuitan menyenangkan.

Dickins (2014) dalam penelitiannya menjelaskan bahwa ada tiga jenis fenomena yang biasanya menjadi fokus dari analisis makna asosiatif: (1) berbasis ekstralinguistik, (2) berbasis linguistik, dan (3) berbasis implikatur skalar. Analisis

berbasis ekstralingistik, misalnya pada kata *nurse* dalam bahasa Inggris yang berasosiasi dengan *female nurse* ‘perawat wanita’. Analisis berbasis linguistik, misalnya kata *vernietigen* yang bersinonim dengan *vernieten* ‘menghancurkan’ namun *vernieten* ada kecenderungan merujuk ke ‘berbagai jenis penghancuran’. Analisis berdasarkan implikatur skalar, misalnya kata *some* ‘beberapa’ yang memiliki kecenderungan ditafsirkan ‘tidak semua’. Merujuk pada penelitian Dickins dalam menganalisis makna asosiasi tersebut, kajian makna asosiasi pada penelitian ini pun berbasis linguistik dengan menelusuri makna kata secara leksikal lalu menganalisis asosisasinya berdasarkan pada cuitan yang menyenangkan.

Ufimtseva (2014), mengkaji makna asosiasi skala besar dengan partisipasi penutur asli bahasa Rusia. Penelitiannya membawa manfaat untuk memperkenalkan objek baru dari analisis linguistik, yang memungkinkan adanya pandangan baru tentang bahasa dan budaya serta peran bahasa dalam pembentukan mentalitas etnis. Terkait dengan pemanfaatan penelitian makna asosiasi, penelitian dalam tulisan ini juga memanfaatkan kosakata makna asosiasi menyenangkan dari cuitan warganet Bandung untuk meningkatkan akurasi sistem prediksi makna cuitan menyenangkan. Dengan demikian, diharapkan manfaat dari penelitian ini menambah kontribusi Linguistik dalam bidang ilmu lain, khususnya Informatika.

1.2 Rumusan Masalah Penelitian

Rumusan masalah penelitian ini berfokus pada upaya menginvestigasi medan makna dan makna asosiasi dari cuitan. Cuitan menyenangkan menjadi sampel data dalam proses penelitian ini. Rumusan penelitian meliputi:

- a. bagaimana medan makna kata menyenangkan dipetakan berdasarkan penelusuran dari cuitan warganet;
- b. bagaimana makna asosiasi menyenangkan dipetakan berdasarkan penelusuran dari cuitan warganet;
- c. apa kontribusi hasil penelusuran medan makna dan makna asosiasi terhadap Sidikta?

1.3 Tujuan Penelitian

Terkait dengan rumusan masalah di atas, berikut ini tujuan yang ingin dicapai melalui penelitian ini.

- a. Memaparkan proses pemetaan medan makna kata menyenangkan berdasarkan penelusuran dari cuitan warganet.
- b. Memaparkan proses pemetaan makna asosiasi menyenangkan berdasarkan penelusuran dari cuitan warganet.
- c. Memaparkan kontribusi hasil penelusuran medan makna dan makna asosiasi menyenangkan terhadap Sidikta.

1.4 Manfaat Penelitian

Seperti yang telah dijabarkan dalam latar belakang, cuitan banyak dimanfaatkan sebagai data penelitian untuk mengetahui respons warganet, misalnya respons positif/negatif terhadap suatu produk atau respons terhadap pasangan calon pemimpin. Salah satu cara untuk mengetahui makna respons-respons dalam cuitan adalah menggunakan SKMC. Oleh SKMC cuitan diklasifikasikan dalam makna

positif, negatif, atau netral. Terkait dengan SKMC, manfaat penelitian ini dipaparkan berikut ini.

- a. Penelusuran cuitan menyenangkan bermanfaat untuk membangun *database* kosakata dalam medan makna menyenangkan dan berasosiasi menyenangkan.
- b. *Database* kosakata dalam medan makna menyenangkan dan berasosiasi menyenangkan, dimanfaatkan untuk meningkatkan persentase akurasi SKMC.
- c. Klasifikasi makna menyenangkan atau tidak menyenangkan dari ribuan cuitan warganet dapat diketahui secara cepat.
- d. Pemerintah/lembaga dapat mengetahui respons masyarakat/konsumen terhadap suatu kebijakan/produk tanpa membaca respons secara satu per satu.
- e. Ribuan respons warganet dapat diketahui maknanya secara umum dengan cepat dan *real time*.
- f. Penelitian ini dapat membantu pengembangan penelitian lain, misalnya untuk pembangunan *database* kosakata berasosiasi rekreasi atau kemacetan.

1.5 Struktur Organisasi Disertasi

Penelitian ini akan dipaparkan dalam lima bab secara berurutan, yaitu Bab Pendahuluan, Bab Tinjauan Pustaka, Bab Metode Penelitian, Bab Analisis Data, dan Bab Penutup.

Dalam Bab Pendahuluan dipaparkan studi terkait analisis cuitan di Twitter yang pernah dilakukan oleh para peneliti sebelumnya. Tren penelitian sebelumnya memanfaatkan data cuitan di Twitter untuk tujuan tertentu dengan menggunakan teknik statistik atau menggunakan algoritma tertentu. Sedangkan penelitian ini mencoba berkontribusi dalam penelusuran medan makna dan makna sosiasi pada identifikasi makna menyenangkan dari cuitan.

Bab Tinjauan Pustaka memaparkan pustaka-pustaka yang menjadi landasan berpikir pada penelitian ini, meliputi teori semantik, konsep medan makna, dan makna asosiasi. Setelah konsep-konsep dipaparkan, selanjutnya disusunlah prosedur penelitian (dari pengambilan hingga pengolahan data) yang dipaparkan dalam Bab 3 Metodologi.

Berdasarkan pada metode yang telah dipaparkan sebelumnya, data cuitan akan dipaparkan dan dianalisis. Proses tersebut dipaparkan pada Bab 4 Temuan dan Pembahasan. Dari hasil pembahasan, akan ditarik kesimpulan yang diambil berdasarkan pada pembahasan untuk menjawab masalah: 1) penelusuran medan makna menyenangkan dan makna asosiasi menyenangkan dalam identifikasi cuitan bermakna menyenangkan dan 2) pengaruh *database* kosakata (dalam medan makna menyenangkan dan kata berasosiasi menyenangkan) dalam membantu sistem memprediksi cuitan bermakna menyenangkan.