

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Mikroblog adalah salah satu bentuk blog yang memungkinkan penggunanya untuk berbagi konten digital. Perbedaan mikroblog dengan blog biasa terletak pada ukuran posting. Mikroblog memiliki batasan karakter yaitu antara 140-200 karakter, sedangkan pada blog biasa jumlah karakter tidak dibatasi. Mikroblog sangat populer dikalangan jejaring sosial yang digunakan untuk membuat atau memperbarui konten dan menciptakan rasa komunitas *online*.

Twitter merupakan mikroblog paling populer saat ini yang terus berkembang sejak awal kemunculannya di tahun 2006. Twitter memungkinkan penggunanya untuk berbagi informasi dengan tulisan sepanjang maksimum 140 karakter yang disebut *tweet*. Berdasarkan data bulan Juli 2014, Twitter telah memiliki 645 juta pengguna aktif, menghasilkan 9.100 *tweets* setiap detik dan mengatasi 2,1 milyar *search engine queries* setiap harinya (Statisticbrain, 2014). Indonesia merupakan negara peringkat ketiga di Asia Pasifik dengan pengguna twitter aktif yaitu sebesar 22,1 juta (Statista, 2015). Maka informasi dari twitter dapat digunakan sebagai sumber untuk mencari pengetahuan yang berharga.

Twitter menyediakan sumber informasi yang besar dan mudah didapat. Sifat informasinya mengalir secara *stream* dan *up-to-date*. Namun, karena topik di twitter sangat beragam, maka dibutuhkan pemrosesan *tweet* untuk memperoleh informasi yang berharga. Selain topiknya beragam, cara penulisan juga beragam. Banyak *tweet* yang mengandung *noise* seperti *emoticon*, singkatan, bahasa gaul, *link* dan kata-kata yang tidak bermakna (Doan, dkk., 2012).

Pengetahuan yang diperoleh dari twitter dapat menunjang berbagai bidang, salah satunya bidang kesehatan. Topik yang dipilih pada penelitian ini yaitu mengenai penyebaran *Influenza Like Illness* (ILI) di Indonesia. ILI bukan sebuah penyakit, melainkan diagnosa medis atau gejala terhadap suatu penyakit yang dapat menyebabkan wabah. Pada banyak negara yang memiliki empat musim, pemantauan terhadap ILI dilakukan secara musiman karena influenza dapat

menyebabkan kematian. Untuk negara tropis seperti Indonesia, influenza dianggap sebagai hal biasa, namun pemantauan terhadap ILI penting dilakukan karena menjadi titik awal dalam deteksi penyebaran penyakit yang mewabah seperti flu burung (Sampurno, dkk., 2011). Kasus flu burung di Indonesia mulai menjangkit manusia pada tahun 2005. Sejak saat itu tercatat sebanyak 197 kasus flu burung menjangkit manusia dengan jumlah korban meninggal sebanyak 165 orang dan merupakan kasus kematian paling tinggi di dunia (Depkes RI, 2014). Kejadian tersebut membuktikan perlu adanya tindakan pencegahan untuk memperbaiki perencanaan kesehatan sebelum sebuah wabah menular.

Pemantauan terhadap penyebaran ILI di Indonesia telah dilakukan oleh badan terkait melalui *sentinel surveillance* (WHO, 2012). Namun demikian, pemantauan masih menggunakan sistem manual yaitu dengan teknik wawancara. Selain itu pemantauan baru dilakukan pada daerah ujicoba dan belum mencakup seluruh wilayah Indonesia (Kusriastuti, 2012).

Pemantauan ILI secara otomatis telah dilakukan sebelumnya dengan menggunakan data dari twitter. Cullota menggunakan teknik klasifikasi untuk membuang data yang merupakan *false alarm*. Pertama dengan membuat dataset yang diklasifikasikan ke dalam dua kelas yaitu *positive* dan *negative* secara manual. Hasil klasifikasi dengan *10-fold cross validation* menghasilkan akurasi sebesar 84,29%. Hasil klasifikasi untuk *tweet* positif influenza kemudian dikorelasikan dengan data CDC dan menghasilkan korelasi sebesar 78% (Cullota, 2010). Penelitian yang dilakukan oleh Aramaki, dkk. difokuskan pada pembagian *tweet* influenza negatif dan positif. Kalimat yang dinyatakan positif jika kalimat tersebut *present tense* atau *recent past*. Dilakukan klasifikasi dengan beberapa classifier, dihasilkan *f-measure* tertinggi sebesar 75,6% menggunakan SVM. Data hasil klasifikasi dengan kelas positif kemudian dibandingkan dengan data CDC dan memperoleh korelasi sebesar 89% (Aramaki, dkk). Namun demikian, pemantauan ILI di Indonesia dengan menggunakan twitter belum pernah dilakukan sebelumnya. Oleh karena itu, dengan memanfaatkan twitter akan dibuat sistem yang dapat mengklasifikasikan ILI dan visualisasi berupa sistem pemantau.

Klasifikasi teks berfungsi untuk mengelompokkan suatu teks ke dalam suatu kategori atau kelas tertentu. Visualisasi informasi merupakan penggunaan interaksi visual dari komputer untuk merepresentasikan data yang abstrak agar lebih dipahami. Tujuan visualisasi yaitu untuk membantu orang dalam penerimaan informasi sehingga memperoleh keputusan yang lebih baik (Few, dkk., 2010).

Metode yang digunakan pada penelitian ini yaitu menggunakan teknik klasifikasi dengan algoritma *improved kNN* (*K-Neareast Neighbors*). Eksperimen dilakukan dengan mengadaptasi proses klasifikasi yang dilakukan Cullota (2010) dan Aramaki, dkk. (2011). Proses pertama yaitu mengklasifikasikan tweet dalam satu tahap yaitu memisahkan tweet flu dan tweet bukan flu. Selain itu dilakukan klasifikasi dalam dua tahap. Pertama untuk memisahkan antara *tweet* sakit (menyatakan sakit) dan *tweet* bukan sakit (membicarakan sakit secara umum). Kemudian tweet dengan kategori sakit diklasifikasikan kembali menjadi *tweet* flu (menyatakan flu) dan *tweet* bukan flu (menyatakan sakit, namun bukan flu). Nantinya akan dibandingkan hasil klasifikasi dalam satu tahap dan hasil klasifikasi dalam dua tahap.

kNN merupakan salah satu algoritma klasifikasi teks yang banyak digunakan karena efektivitasnya dan kemudahan penggunaannya. Proses klasifikasi pada algoritma *kNN* yaitu dengan mencari nilai *kemiripan* antara data uji dengan data latih. Kekurangan pada algoritma kNN yaitu membutuhkan alokasi memori yang besar dan lamanya waktu komputasi karena setiap satu data masukan akan dibandingkan dengan seluruh data latih. Untuk mengatasi masalah tersebut, maka pada *improved kNN* ditambahkan *similarity function* untuk mengukur kedekatan ketetanggan. Pemilihan *improved kNN* ini dikarenakan akurasi menjadi lebih tinggi pada jenis *short text* dibandingkan *classifier* lain seperti Naïve Bayes dan SVM (Khamar, 2013).

Dari penjelasan di atas, penulis akan melakukan penelitian tentang penyebaran ILI di Indonesia dengan memanfaatkan twitter menggunakan teknik klasifikasi dengan algoritma *improved kNN*.

## 1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas pada penelitian ini adalah :

1. Bagaimana cara mengklasifikasikan *tweet* dengan menggunakan *improved kNN*?
2. Apakah visualisasi informasi yang disajikan dapat mempermudah orang untuk memantau penyebaran ILI?

## 1.3 Batasan Masalah

Adapun batasan masalah terhadap penelitian ini, diantaranya:

1. Data diambil dari pengguna twitter berbahasa Indonesia.
2. Data lokasi tweet diambil dari profil pengguna.
3. Pengambilan data terbatas pada *tweet* yang mengandung kata kunci yang telah ditentukan yaitu ‘flu’, ‘batuk’, ‘sakit tenggorokan’, ‘sakit kepala’, ‘pilek’, ‘meler’ dan ‘bersin’.

## 1.4 Tujuan Penelitian

Tujuan diadakannya penelitian ini yaitu:

1. Dapat mengklasifikasikan *tweet* menggunakan *improved kNN*.
2. Dapat memvisualisasikan informasi sehingga memudahkan orang untuk memantau penyebaran influenza.

## 1.5 Manfaat Penelitian

Dengan adanya penelitian ini, diharapkan dapat memberikan manfaat sebagai berikut.

1. Memberikan alat yang memudahkan dalam pemantauan *Influenza Like Illness* di Indonesia.
2. Dapat menjadi rujukan untuk penelitian selanjutnya yang memiliki keterkaitan dengan penelitian ini, baik langsung atau tidak langsung.

## 1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut.

## BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematikan penulisan.

## BAB II TINJAUAN PUSTAKA

Bab ini berisi tentang teori dan konsep terkait penelitian.

## BAB III METODOLOGI PENELITIAN

Bab ini berisi langkah-langkah yang akan dilakukan dalam penelitian.

## BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Bab ini berisi uraian dan pembahasan terhadap penelitian yang dilakukan.

## BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan terkait penelitian dan saran dari penulis untuk kegiatan penelitian selanjutnya terkait dengan topik yang sedang dibahas.