

BAB I PENDAHULUAN

1.1. Latar Belakang

Situs *microblogging* seperti Twitter memiliki peran penting dalam penyebaran informasi mengenai suatu kejadian, baik alami maupun yang berhubungan dengan aktivitas manusia. Volume dan kecepatan *tweet* pada saat suatu kejadian berlangsung sangat tinggi sehingga masyarakat yang terdampak dan petugas profesional mengalami kesulitan dalam pemrosesan informasi (Imran *et al.* 2013).

Berkembangnya media sosial mendorong munculnya situs *microblogging* seperti Twitter yang dapat berperan sebagai sistem sensor dengan menjadikan penggunaanya sebagai sensor dan *tweet* yang memiliki informasi tentang suatu lokasi. Penelitian yang telah dilakukan oleh Crooks *et al.* (2012) bahwa media sosial Twitter bisa menjadi sensor sistem yang dapat memberikan informasi kepada masyarakat tentang terjadinya gempa bumi. Hal tersebut terjadi karena Twitter memiliki *hashtag* tertentu, *longitude*, dan *latitude* yang bisa memberikan informasi lokasi *tweet* tersebut di-*posting* pada Gempa bumi dapat dideteksi dengan cara melakukan pemantauan *tweet*. Untuk Gempa bumi dengan skala richter 3 atau lebih Sakaki *et al.* (2010). Probabilitas yang dihasilkan dari *Japan Meteorologi Agency* (JMA) tinggi yaitu 96%.

Penyakit menular akut adalah penyakit yang dapat ditularkan melalui berbagai media. Penyakit jenis ini merupakan masalah kesehatan yang besar di hampir semua negara berkembang karena angka kesakitan dan kematiannya yang relatif tinggi dalam kurun waktu yang relatif singkat. Penyakit menular umumnya bersifat akut (mendadak) dan menyerang semua lapisan masyarakat. Penyakit jenis ini diprioritaskan mengingat sifat menularnya yang bisa menyebabkan wabah dan menimbulkan kerugian yang besar. Penyakit menular merupakan hasil perpaduan berbagai faktor yang saling mempengaruhi (Widoyono, 2011: 3).

Infeksi Saluran Pernapasan Akut adalah penyakit menular akibat infeksi pada saluran pernapasan yang seringkali disebabkan oleh virus, namun bisa juga disebabkan oleh bakteri. Influenza adalah salah satu penyebab ISPA menyerang sistem pernapasan, sehingga gejala yang timbul berupa hidung tersumbat, bersin, batuk, dan sakit tenggorokan. Anak-anak juga akan mengalami demam tinggi selama beberapa hari, lemas, sakit kepala, dan nyeri otot. Pada kasus yang lebih berat, kondisi ini dapat berkembang menjadi pneumonia. (Alodokter, 2018),

Berdasarkan Profil Kesehatan Indonesia yang diterbitkan oleh Kementerian Kesehatan Republik Indonesia (2018), terdapat 5 penyakit yang termasuk penyakit menular, yaitu tuberkulosis, diare, hepatitis, ISPA, dan pneumonia. Pada tahun 2017 ditemukan jumlah kasus ISPA sebanyak 1.846.180 kasus, meningkat bila dibandingkan semua kasus ISPA yang ditemukan pada tahun 2016 sebanyak 1,801.968 kasus. (Kemenkes, 2018)

Kebakaran hutan di Indonesia secara umum disebabkan oleh dua faktor, yaitu faktor kelalaian dan faktor kesengajaan. Kedua faktor ini sama-sama diakibatkan oleh aktivitas manusia. Faktor kelalaian terjadi sebagai akibat kurang hati-hatinya dalam beraktivitas di dalam hutan, sedangkan faktor kesengajaan terjadi sebagai akibat dari aktivitas pembukaan lahan baru dengan cara membakar atau peremajaan tanaman industri di wilayah industri. Bila dibandingkan, kebakaran hutan karena faktor kelalaian manusia jauh lebih kecil dibanding dengan faktor kesengajaan membakar hutan (Bahri, 2002).

Kejadian kabut asap akibat kebakaran hutan dan lahan di Kalimantan Barat telah dikenal sejak tahun 1997 dan merupakan bencana nasional yang terjadi setiap tahun hingga kini (Adam, 2012). Wilayah hutan dan lahan yang berpotensi terbakar di Indonesia antara lain Riau, Jambi, Sumatra Utara, Sumatera Selatan, Kalimantan Barat, Kalimantan Timur, dan Kalimantan Selatan. Hal ini biasanya terjadi pada musim kemarau atau pada masa peralihan (Bahri, 2002).

Polusi adalah masalah yang berbahaya bagi aktivitas kehidupan manusia baik di dalam maupun di luar lingkungan. Polusi udara memiliki implikasi negatif bagi kesehatan manusia secara umum. Polusi udara dapat memicu penyakit seperti infeksi saluran pernafasan akut yang lebih besar. Akibat polusi udara di Indonesia diperkirakan sekitar 89.526.421 orang berisiko terkena polusi udara, sehingga meningkatkan risiko ISPA (Hidayat, 2012).

Beberapa penelitian yang menggunakan data dari media sosial Twitter telah dilakukan sebelumnya. Denatari (2015) melakukan penelitian tentang klasterisasi data teks Twitter untuk kasus pertanian di Indonesia. Data teks Twitter terbagi menjadi 2 jenis, yaitu data *tweet* sejumlah 51 data dan data konten *Uniform Resource Locator* (URL) sejumlah 51 data. Kedua jenis data tersebut dibandingkan dan dikelompokkan dengan algoritma *hierarchical clustering* untuk mendapatkan klaster terbaik.

Pada penelitian Susanto *et al.* (2014) dilakukan visualisasi data teks Twitter berbasis Bahasa Indonesia menggunakan teknik klasterisasi. Topik sentimen yang diteliti pada Twitter adalah Pemilu 2014. Teknik yang digunakan yaitu klasterisasi dengan algoritma K-Means, *cascade* K-Means, dan *Self Organizing Map* (SOM) *Kohonen* kemudian membuat visualisasinya dan dilakukan perbandingan ketiga metode tersebut.

Zwitch (2013) telah menganalisis kata-kata yang digunakan oleh pengunjung dalam mengunjungi suatu situs (kata kunci pencarian) dan kata-kata apa saja yang digunakan saat berada di situs tersebut (pencarian situs) dengan algoritma KMeans. Tujuan dari penelitiannya adalah untuk menentukan label pada setiap klaster. Penentuan nilai *k* dilakukan untuk memperkirakan topik setiap klaster sebagai asumsi awal pencarian kata-kata yang dominan sebelum dilakukan penentuan label

Terpstra *et al.* (2012) melakukan penelitian mengenai kemungkinan analisis secara *real time* dan otomatis dari pesan Twitter selama terjadinya situasi darurat. Analisis dilakukan dengan menggunakan *tool* ekstraksi informasi yang berhasil mendapatkan 97.000 *tweet* yang dikirim sebelum, saat, dan setelah badai terjadi pada festival Pukkelpop 2011 di Belgia. *Tool* ekstraksi dapat menganalisis *tweet* melalui

tampilan geografis, jenis isi pesan (kerusakan, korban), dan jenis *tweet* (seperti *retweet*).

Pada penelitian ini dilakukan klasterisasi *tweet* dengan algoritma K-Means untuk kasus penyebaran penyakit menular akut (studi kasus ISPA, kebakaran hutan dan polusi). Algoritma K-Means digunakan karena dapat mengelompokkan objek secara menyeluruh sehingga klaster yang terbentuk tidak bersifat hirarki. Sebelum proses klasterisasi, dilakukan tahap praproses seperti *case folding*, penghapusan simbol khusus *tweet*, penghapusan angka dan tanda baca, penghapusan *white space* dan *single character*, *tokenizing*, penghapusan *stopwords*, normalisasi, dan *stemming*. Sehingga mengubah data *tweet* menjadi *Term Document Matix* (TDM).

Pengujian dilakukan dengan mengevaluasi hasil analisis klasterisasi menggunakan metode *elbow* yang berdasarkan nilai *Sum Squared Error* (SSE) dan analisis *silhouette* yang berdasarkan nilai koefisien *silhouette*.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana menerapkan algoritma klasterisasi K-Means untuk mengolah data *tweet*?
2. Bagaimana menerapkan analisis hasil klasterisasi data *tweet* untuk kasus penyebaran menular langsung (studi kasus ISPA, kebakaran hutan dan polusi) dengan membandingkan pola kasus yang ada.

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Menerapkan algoritma klasterisasi K-Means untuk mengolah data *tweet*.
2. Menerapkan analisis hasil klasterisasi data *tweet* untuk kasus penyebaran penyakit menular akut (studi kasus ISPA, kebakaran hutan dan polusi) dari hasil membandingkan pola kasus yang ada.

1.4 Manfaat Penelitian

Hasil klasterisasi dan analisis *tweet* dengan algoritma K-Means untuk kasus penyebaran penyakit menular akut diharapkan dapat bermanfaat bagi:

1. Masyarakat yang terdampak bisa mengetahui *tweet* kejadian berlangsung dan sebagai sistem peringatan dini tentang penyebaran penyakit menular akut (studi kasus ISPA, kebakaran hutan dan polusi) di Indonesia.
2. Tenaga profesional bermanfaat untuk memudahkan proses pengambilan informasi tentang *tweet* dan sebagai sistem peringatan dini tentang penyebaran penyakit menular akut (studi kasus ISPA, kebakaran hutan, dan polusi) di Indonesia.

1.5 Batasan Masalah

Batasan masalah penelitian ini adalah:

1. Penelitian ini menggunakan data teks dari media sosial Twitter.
2. Data teks yang digunakan yaitu *tweet* berbahasa Indonesia.
3. Data teks yang digunakan adalah data *tweet* yang menggunakan bahasa indonesia yang diambil selama Januari 2019 – Desember 2019.