

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Berita merupakan salah satu sumber informasi mengenai kejadian terkini yang ada pada media massa seperti surat kabar, televisi, dan media *online*. Seiring berkembangnya teknologi internet menyebabkan kemudahan untuk mendapatkan suatu informasi berita salah satunya melalui sosial media [1]. Salah satu sosial media yang paling berpengaruh di dunia adalah *twitter*. Di Indonesia terdapat sebanyak 15,7 juta pengguna *twitter* dan Indonesia menjadi negara peringkat ke 6 sebagai negara dengan pengguna *twitter* terbanyak di bawah Amerika Serikat, Brazil, dan negara-negara lainnya [2]. *Twitter* menjadi salah satu sosial media yang paling digemari dan sering dijadikan sebagai tempat untuk berbagi informasi, baik itu informasi berupa komentar ataupun pesan [3]. *Twitter* merupakan aplikasi sosial media yang didirikan oleh Jack Dorsey pada Maret 2006 yang memungkinkan penggunaanya berbagai informasi atau berita secara realtime dengan kicauan (*tweets*) yang terdiri dari 140 karakter dengan rata-rata pengguna *twitter* memposting sebanyak 500 juta *tweets* per hari dan 80% postingan menggunakan perangkat *mobile* [Fauzo Zakiy Mhd. 2019]. Berita-berita yang diposting di media *twitter* tidak terdapat kategori penggolongan ataupun pengelompokkan jenis berita sehingga memerlukan waktu yang tidak sedikit untuk melakukan pencarian sesuai kategori yang diinginkan.

Maka salah satu solusi yang dapat diterapkan untuk dapat mengatasi masalah tersebut adalah melalui *clustering* terhadap informasi berita yang ada di *twitter*. *Clustering* merupakan proses pengelompokan benda serupa ke dalam kelompok yang berbeda, atau lebih tepatnya partisi dari sebuah data set kedalam *subset*, sehingga data dalam setiap subset memiliki arti yang bermanfaat [4]. Banyaknya berita yang tersedia di *twitter* perlu dikelompokkan agar dapat memudahkan pengguna dalam melakukan pencarian berita. *K-Means Clustering* merupakan salah satu metode *clustering* yang digunakan dalam mengelompokkan data, karena *K-Means Clustering* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya. *Clustering text* ini

melibatkan data teks yang tidak terstruktur, teknik dalam *text minning* dapat dijadikan sebagai solusi untuk menemukan kata atau pola yang diinginkan untuk dijadikan kunci dalam proses *clustering*. *text minning* itu sendiri adalah proses ekstraksi pola berupa informasi dan pengetahuan yang berguna dari sejumlah besar sumber data teks [4]. Proses pada *text mining* tersebut diantaranya adalah *case folding*, *tokenization*, *filtering* dan *stemming* memiliki tujuan untuk mereduksi atau mengekstrak data serta mengurangi noise pada data [5]. Metode ini dipilih pada penelitian ini karena tergolong cukup akurat serta memiliki waktu proses komputasi yang relatif singkat. Metode *K-Means Clustering* akan digunakan pada penelitian ini untuk melakukan pengelompokan berita pada sosial media *Twitter* [5]. Perbedaan penelitian ini dari penelitian sebelumnya yaitu pada penelitian ini sistem pengelompokkan berita kesehatan pada penelitian ini dibangun dengan menggunakan bahasa pemrograman *Python* versi 3.9.6 dengan *text editor* yang digunakan adalah *Python IDLE (Integrated DeveLopment Environment)*. *Output* hasil *clustering* akan ditampilkan menggunakan *Python Shell*, selain itu pada penelitian ini dilakukan pengelompokkan terhadap berita kesehatan *twitter* berdasarkan klaster kata positif dan negatif [7]

Penelitian sebelumnya dengan metode *K-Means Clustering* sudah pernah diterbitkan dengan judul “Klasifikasi Berita *Twitter* Menggunakan Metode *Improved Naïve Bayes*”. Pada penelitian ini digunakan delapan kategori berita berbahasa Indonesia [6]. Penelitian berikutnya menerapkan metode *K-Means Clustering* dalam pengelompokkan berita kesehatan dengan judul “*Analysis of Health-Tweets using K-Means Clustering*”.. Namun penelitian sebelumnya belum melakukan penelitian untuk menganalisis klaster terbaik dalam melakukan pengelompokkan berita kesehatan dari sosial media *twitter*.

Berdasarkan uraian diatas maka akan dianalisis dan dibangun sebuah sistem dengan judul **“Pengelompokkan Berita Kesehatan Pada Sosial Media *Twitter* Dengan Metode *K-Means Clustering*”**.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang diuraikan di atas, maka rumusan masalah pada penelitian ini yaitu:

1. Bagaimana cara menganalisis dan menentukan klaster terbaik terhadap pengelompokan berita kesehatan dari sosial media *twitter*?
2. Bagaimana cara melakukan pengelompokan berita kesehatan menggunakan metode *K-Means Clustering*?

1.3. Batasan Masalah

Berikut ini akan diuraikan batasan masalah pada penelitian ini agar penelitian berjalan sesuai dengan perencanaan yaitu:

1. Metode yang digunakan adalah metode *K-Means Clustering*.
2. Dataset yang digunakan diambil dari *UCI Machine Learning Repository* yaitu *Health News in Twitter Data Set* [8].
3. Analisis dilakukan dengan menggunakan *Python* sebagai bahasa pemrograman untuk melakukan perhitungan *K-Means Clustering*.

1.4. Tujuan Penelitian

Tujuan yang diharapkan dari penelitian ini yaitu:

1. Untuk mengetahui klaster terbaik dalam melakukan pengelompokan berita kesehatan dari sosial media *twitter*.
2. Untuk merancang sistem pengelompokan berita kesehatan menggunakan metode *K-Means Clustering*.

1.5. Manfaat Penelitian

Manfaat penelitian bagi penulis yaitu:

1. Menjadi portofolio bagi penulis untuk mendapatkan pekerjaan.
2. Menambah pengetahuan penulis dalam metode *data mining*, khususnya pada metode *K-Means Clustering*.

Manfaat penelitian bagi masyarakat yaitu:

1. Membantu *user* dalam menemukan berita kesehatan *Twitter* yang diinginkan.
2. Memudahkan pengguna *Twitter* dalam melakukan pencarian berita kesehatan berdasarkan klasterisasi yang terbentuk.