

BAB I

PENDAHULUAN

1.1. Latar Belakang

Indonesia merupakan negara yang mendukung dengan adanya resolusi Majelis Umum PBB. Resolusi tersebut menjadi salah satu bentuk dukungan terbanyak yang pernah diajukan kepada Majelis Umum PBB. Tetapi, sikap pemerintah ternyata cukup bertolak belakang dengan kebanyakan reaksi warganet Indonesia di media sosial. Dalam penglihatan pengamat masyarakat Indonesia terpecah dalam menanggapi fenomena invasi Rusia. Radityo beranggapan publik cenderung mendukung adanya invasi Rusia karena beberapa variabel. Salah satunya yang menjadi sorotan warganet yaitu sentimen agama yang mendasari dukungan publik Indonesia terhadap Rusia. Seperti yang ramai di media sosial, media dan publikasi Rusia sering menampilkan citra yang bersahabat dengan Islam akhir-akhir ini. Contohnya, Presiden Rusia yaitu Putin mengangkat tokoh muslim Ramzan Kadyrov sebagai Presiden Chechnya, di negara bagian di Rusia. Namanya begitu dikenal di kalangan umat Muslim, sering masuk ke dalam sorotan media, termasuk media Indonesia. Hal itu yang membuat dampak positif bagi Rusia yang menjadi awal mula citra persahabatan dengan Islam. Radit juga menyebutkan di dalam youtube-nya, Rusia juga di kaitkan sebagai bangsa Rum yang nantinya akan beraliansi dengan Umat Muslim di akhir zaman [1].

Dengan adanya berbagai segmentasi yang beredar dimasyarakat seperti itu maka akan bermunculan sebuah opini-opini masyarakat yang sangat beragam mulai dari opini yang pro mengenai adanya invasi yang dilakukan oleh Rusia ke Ukraina hingga opini yang kontra mengenai hal tersebut. Masyarakat Indonesia pun banyak yang melakukan *tweet* mengenai fenomena tersebut mulai dari yang mengecam perbuatan Rusia hingga yang mendukung perbuatan pergerakan invasi Rusia.

Bertambahnya data yang muncul setiap detiknya dapat dimanfaatkan dalam memperoleh sebuah informasi. Proses pengolahan data atau data mining

bisa dilakukan dengan berbagai macam cara, salah satunya dengan klusterisasi [2]. *Data mining* merupakan sebuah cabang dari *machine learning*, *data mining* adalah suatu proses untuk olah data sehingga mampu menjadi sebuah informasi. Sederhananya, *data mining* merupakan sebuah langkah untuk bisa mengekstraksi data dan dijadikan sebagai informasi namun masih bersifat implisit atau belum diketahui [3].

Dalam *data mining* memiliki beragam teknik untuk dapat mengolah sebuah data menjadi informasi, salah satunya dengan teknik *clustering*. *Clustering* merupakan teknik untuk menganalisis dan mengkaji data untuk menyelesaikan permasalahan dengan mengelompokkan data membagi dari suatu dataset ke dalam subset. *Clustering* memiliki berbagai macam algoritma seperti k-means, mini Batch K-means, k-means++, k-medoids, Fuzzy C-means dan lainnya [4].

Algoritma mini Batch K-means adalah alternatif algoritma K-means untuk pengelompokan masif kumpulan data. Keuntungan dari algoritma ini adalah untuk mengurangi biaya komputasi dengan tidak menggunakan semua dataset setiap iterasi tetapi subsampel dengan ukuran tetap. Strategi ini mengurangi jumlah perhitungan jarak periterasi [5].

Algoritma mini batch K-means sering diterapkan kedalam berbagai macam permasalahan penelitian terkait tentang *clustering*, beberapa penelitiannya adalah penelitian Klasifikasi Kategori Kemiskinan Di Malaysia Menggunakan Model Pengelompokan. Penelitian tersebut bertujuan untuk Pembangunan dari sistem tersebut adalah sistem yang dapat memprediksi klasifikasi kategori kemiskinan berdasarkan dari hasil data input yang dimasukkan oleh *user* [6]. Selain itu, terdapat penelitian lain yang berjudul Implementasi metode algoritma k-means *clustering* pada analisis sentimen keluhan pengguna Indosat. Penelitian ini bertujuan dapat menjadi acuan terhadap *provider* dalam melihat berbagai macam keluhan yang ada dari konsumen mereka menyebabkan pihak *provider* bisa meningkatkan pelayanan [7].

Dengan digunakannya metode Mini Batch K-means dalam mensegmentasi opini masyarakat Indonesia terkait invasi Rusia. Pengelompokan dibentuk dengan cara digabungkannya teknik klasterisasi pada Mini Batch K-means diperuntukan untuk pengelompokkan data [8].

Mengumpulan data dari media sosial dinilai bisa memberikan efisiensi pada segala hal apabila dibandingkan dengan harus dilakukan sebuah survey secara tradisional. Dengan adanya hal itu dapat biaya yang harus dikeluarkan untuk perolehan data menjadi lebih kecil, lalu perolehan data dapat dilakukan secara *real time*, dan juga hasil data yang dimiliki memiliki informasi yang lebih baik untuk menggambarkan opini-opini warga banyak yang sebenarnya terjadi [9].

Penelitian segmentasi opini masyarakat ini menggunakan algoritma Mini Batch K-means. Algoritma ini merupakan salah satu algoritma klasterisasi yang biasa digunakan pada teks. Algoritma ini dipilih karena algoritma Mini Batch K-means dapat mengatasi jumlah data yang besar. Selain itu algoritma Mini Batch K-means juga dapat menghasilkan kualitas *cluster* yang lebih optimal dibandingkan dengan algoritma pengklasteran lainnya. Dan juga merupakan salah satu metode yang cukup efisien dalam hal kompleksitasnya sehingga memiliki ketepatan yang relatif tinggi terhadap ukuran objek, lalu algoritma ini menjadi lebih terukur dan juga akan efisien dalam olahan suatu objek data jumlah besar. Selain itu algoritma Mini Batch K-means ini tidak terpengaruh akan urutan objek [10].

Oleh sebab itu, dibutuhkan suatu metode pengolahan topik secara otomatis agar dapat mempermudah masyarakat mengetahui sebuah informasi. Salah satu metode otomatis dalam pendeteksian topik adalah dengan menggunakan metode K-means. Namun, data yang cukup besar menjadi permasalahan dalam mendeteksi sebuah topik secara otomatis, sehingga dibutuhkan suatu solusi lain dalam proses pengolahan data tersebut. Pada penelitian ini proses yang digunakan yaitu mini batch K-means yang hanya bisa menggunakan sekelompok kecil data dalam proses *clustering*. Dalam beberapa

penelitian akurasi dari metode ini menunjukkan hasil yang cukup baik dalam waktu komputasi yang jauh lebih singkat [11].

Dalam penelitian sebelumnya mengenai *clustering* penelitian ini dilakukan pengelompokan data obat di Puskesmas Karangsambung yang dimanfaatkan untuk bentuk referensi perencanaan obat yang datang ke puskesmas. Dalam pengelompokannya data dibagi kedalam tiga macam yaitu lambat, sedang dan cepat. Hasil yang didapatkan dalam proses kedua algoritma ini membuktikan bahwa algoritma K-Means ini dapat memiliki hasil Silhouette Coefficient yang jauh lebih tinggi sebesar 0,367 sedangkan untuk hasil K-Medoids sebesar 0,325 yang membuktikan bahwa hasil *clustering* dengan metode K-Means akan jauh lebih berkualitas dibandingkan K-Medoids [12].

Pada penelitian ini dalam pemberian nilai vector akan menggunakan *skipgram* dari *fasttext embedding*, lalu menambahkan metode *elbow* dalam menentukan jumlah kluster optimal dan metode untuk meningkatkan tingkat kualitas yaitu dengan cara melakukan evaluasi *cluster* melalui *Silhouette Coefficient* yaitu metode yang digunakan untuk melihat kualitas dan kekuatan kluster, seberapa baik suatu objek diposisikan dalam suatu kluster [13].

Berdasarkan uraian latar belakang diatas. Untuk mengetahui apakah algoritma Mini Batch K-means dapat melakukan klusterisasi pada teks dengan optimal, sehingga perlu adanya penelitian. Dengan permasalahan tersebut maka dirumuskan sebuah penelitian berjudul **“SEGMENTASI OPINI MASYARAKAT INDONESIA TERHADAP INVASI RUSIA PADA MEDIA SOSIAL MENGGUNAKAN ALGORITMA MINI BATCH K-MEANS”**.

1.2 Rumusan Masalah

Rumusan masalah didasarkan kepada latar belakang sebagai berikut:

1. Bagaimana mengimplementasikan algoritma *Mini Batch* K-means untuk menghasilkan kluster optimal dengan *elbow* dan *silhouette coefficient* terhadap opini masyarakat Indonesia mengenai invasi Rusia?
2. Bagaimana interpretasi hasil kluster optimal opini masyarakat Indonesia terhadap invasi Rusia yang dihasilkan algoritma *Mini Batch* K-means?

1.3 Tujuan

Tujuan dari penelitian ini sebagai berikut:

1. Mengetahui bagaimana proses pengimplementasian algoritma *Mini Batch* K-means dalam menghasilkan kluster optimal dengan metode *elbow* dan *silhouette coefficient* terhadap opini masyarakat mengenai invasi Rusia.
2. Mengetahui bagaimana interpretasi hasil kluster optimal opini masyarakat Indonesia terhadap invasi Rusia dengan algoritma *Mini Batch* K-means.

1.4 Batasan Masalah

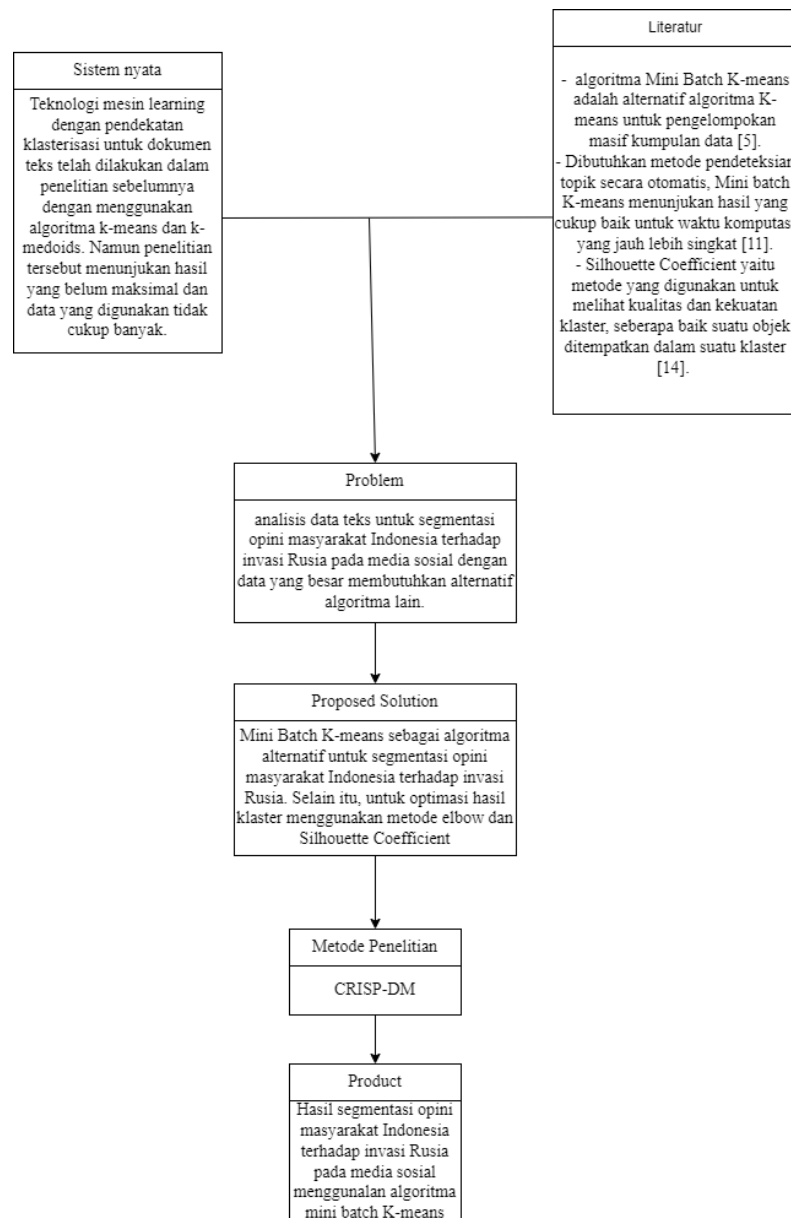
Batasan masalah dalam penelitian ini bertujuan untuk penelitian ini bisa menjadi lebih terarah sesuai dengan tujuan yang di harapkan. Berikut adalah beberapa batasan masalah pada penelitian mengenai Segmentasi Opini Masyarakat Indonesia Terhadap Invasi Rusia Dengan Algoritma *Mini Batch* K-means adalah:

1. Penelitian ini menggunakan data yang diambil dari media sosial Youtube dan Twitter.
2. Komentar atau *tweet* yang digunakan untuk dianalisis hanyalah yang berbahasa Indonesia
3. Algoritma yang digunakan pada penelitian ini adalah *Mini Batch* K-means

4. *Keyword* yang digunakan dalam analisis yaitu “invasi rusia dan perang rusia” yang diambil pada bulan Februari hingga Maret 2022.
5. Metode CRISP-DM yang digunakan hanya sampai tahapan *Evaluation*.

1.5 Kerangka Pemikiran

Kerangka pemikiran dari penelitian tugas akhir ini bisa dilihat pada gambar 1.1 berikut:



Gambar 1.1 Kerangka Pemikiran

Pada Gambar 1.1 menggambarkan mengenai kerangka pemikiran yang menjadi dasar dilakukannya penelitian. Pada sistem nyata menyatakan terdapat pengklasteran yang dilakukan pada penelitian sebelumnya tetapi hasil belum baik, hal tersebut menyebabkan perlu adanya pembaharuan terhadap penelitian sebelumnya namun menggunakan algoritma yang berbeda. Algoritma yang digunakan adalah Mini Batch K-means dan juga menambahkan metode *elbow* dalam pemilihan jumlah kluster optimal lalu metode Silhouette Coefficient untuk meningkatkan kualitas hasil *clustering*. Dengan metode pengembangan CRISP-DM produk yang diinginkan adalah hasil pengujian *clustering* menggunakan algoritma Mini Batch K-means terkait invasi Rusia.

1.6 Sistematika Penulisan

Sistematika penulisan yang disusun pada penelitian yang berjudul Segmentasi Opini Masyarakat Indonesia Terhadap Invasi Rusia pada Media Sosial Menggunakan Algoritma Mini Batch K-means ini terdiri dari 5 bab, pada setiap bab yang dirancang memiliki tujuan tersendiri yang ingin dicapai. Berikut adalah sistematika penulisan yang dirancang:

BAB I PENDAHULUAN

Pada bab pertama yang berisikan latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan masalah dari penelitian, kerangka pemikiran dan sistematika penulisan.

BAB II KAJIAN LITERATUR

Pada bab ini berisikan pembahasan penelitian terdahulu serta konsep-konsep dan teori-teori pendukung dalam perancangan sistem yang akan dibuat serta proses analisis kebutuhan yang nantinya akan digunakan dalam pemodelan perancangan.

BAB III METODOLOGI PENELITIAN

Pada bab menjelaskan mengenai metode yang digunakan dalam penyusunan tugas akhir serta analisis dari sistem yang akan dibuat kemudian akan melakukan proses perancangan berdasarkan analisis

kebutuhan yang diketahui. Metodologi penelitian ditampilkan berdasarkan analisis kebutuhan dengan menggunakan metode CRISP-DM.

BAB IV HASIL DAN PEMBAHASAN

Bab keempat merupakan pembahasan dan juga pemaparan mengenai hasil dari analisis menggunakan algoritma yang dipilih serta evaluasi dari hasil analisis tersebut.

BAB V PENUTUP

Setelah dilakukan penelitian, maka dalam penelitian akan disampaikan kesimpulan dan juga saran yang diajukan untuk peningkatan dalam penelitian ini sebagai bagian dari bab terakhir yaitu penutup.

