

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Virus Corona di temukan di kota wuhan pada desember 2019, yang mana virus ini menyerang sistem pernapasan manusia [1]. Pada maret 2021 virus ini masuk ke Indonesia dan kasus pasien positif bertambah setiap harinya [2]. Segala upaya pemerintah dalam menangani penyebaran virus dalam masyarakat diterapkan salah satunya Pembatasan Sosial Berskala Besar (PSBB) yang dianggap mampu menanggulangi dan mencegah penyebaran virus di Indonesia [3].

Dalam penanganan Covid-19, opini publik dalam menilai kebijakan yang diterapkan pemerintah sangatlah penting. Salah satu portal berita menginformasikan bahwa opini publik memiliki pengaruh besar dalam mengevaluasi kebijakan yang diterapkan [4]. Dari pemberitaan tersebut menyatakan bahwa masyarakat tidak puas dengan kebijakan yang diterapkan dalam menanggulangi covid-19. Kebijakan yang telah diterapkan oleh pemerintah pun akhirnya berganti dimana sebelumnya penanganan virus yang disebut PSBB menjadi PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat) karena evaluasi yang telah dilakukan oleh pemerintah [5]. Selama hampir satu tahun setengah dalam menangani Virus Covid-19 pemerintah sudah beberapa kali mengganti kebijakan dalam penanganan Virus Covid-19 yang mana tujuannya diharapkan dengan adanya pembaharuan kebijakan dapat menangani penyebaran Virus[6]. Dapat disimpulkan bahwa opini publik terhadap penanganan virus Covid-19 yang telah dilaksanakan mempengaruhi kebijakan yang diambil oleh pemerintahan selanjutnya.

Penyebaran informasi mengenai penanganan virus corona di masyarakat menggunakan berbagai social media, salah satunya yaitu Twitter. Dimana Twiter digunakan oleh sebagian besar orang untuk menyatakan pendapat, kritikan serta memberikan informasi suatu berita yang disebut Tweet, yang nantinya dimanfaatkan untuk menilai suatu perusahaan atau pemerintah terhadap suatu

objek. Dikarenakan banyaknya data yang ada maka digunakan metode analisis sentiment guna menganalisis, memahami, dan mengklasifikasi pendapat, evaluasi, penilaian, sikap, dan emosi setiap opini dari penggunaan Twitter terhadap kebijakan yang diterapkan.

Namun dalam analisis sentimen level kalimat dengan metode pembelajaran mesin menunjukkan performa tidak cukup akurat. Pada analisis sentimen level kalimat hanya difokuskan untuk menentukan sentimen positif dan negatif tetapi tidak pada aspek yang terdapat dalam kalimat tersebut, contohnya jika suatu opini mengandung beberapa aspek dan kelas sentiment di dalamnya maka memerlukan pendekatan sentimen analisis berbasis aspek [6]. Maka nantinya dalam teknik analisis sentimen, akan dilakukan analisis data berdasarkan aspek dan mengidentifikasi kelas sentimen yang dikaitkan ke masing-masing aspeknya.

Terdapat beberapa penelitian terdahulu yang melakukan analisis sentiment berbasis aspek antara lain C. H. Yutika dkk. yang mana menganalisis sentiment Review Female Daily berbasis aspek dengan metode TF-IDF dan Naïve Bayes diperoleh nilai F1-Score sebesar 62,81% [8]. Mubarok dkk, mengklasifikasikan sentimen berdasarkan aspek pada studi kasus ulasan restoran pada *subtask SemEval* menggunakan metode *part-of-speech (POS) tagging* pada *data preprocessing* dan menggunakan metode *Chi Square* untuk seleksi fitur dan algoritma *Naïve Bayes* yang menghasilkan nilai analisis sentimen berbasis aspek dengan F1-Measure sebesar 78,12%, F1-Measure untuk klasifikasi aspek 88,13%, dan F1-Measure untuk klasifikasi sentimen 75%[9]. Penelitian selanjutnya mengenai analisis sentiment berbasis aspek yaitu P. H. Prastyo dkk. analisis sentimen terhadap respon pemerintah dalam penanganan covid-19 menggunakan SVM dengan Normalized Poly Kernel memberikan akurasi tertinggi berdasarkan sentiment pada aspek dengan rata-rata akurasi 82,00%, presisi 82,24%, recall 82,01%, dan f-measure sebesar 81,84% [10]. Ni Made A. J. Astari dkk, dilakukan analisis sentimen twitter tentang dampak virus corona menggunakan metode Naive Bayes Classifier menunjukkan hasil akurasi 67% dan error rate sebesar

33% dan menunjukkan metode yang digunakan menghasilkan performa yang stabil. [11]

Maka dalam penelitian ini dilakukan kombinasi antara metode klasifikasi Naïve Bayes Classifier yang menghasilkan performa yang stabil serta penggunaan aspek untuk analisis teks yang akan mengkategorikan data berdasarkan aspek. Atas dasar hal tersebut, maka akan dibangun sebuah sistem klasifikasi sentimen berdasarkan aspek terhadap sentimen publik pada objek tertentu yang disampaikan di Twitter dalam bahasa Indonesia atas opini publik terhadap penanganan Virus Corona menggunakan metode Naïve Bayes Classifier.

1.2 Perumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah yang ada adalah seberapa baik Naive Bayes Classifier dalam analisis sentimen berbasis aspek pada kasus penanganan virus Covid 19.

1.3 Maksud dan Tujuan

Berdasarkan Permasalahan yang ada, maksud dari penelitian ini adalah untuk mengimplementasikan metode Naive Bayes Classifier untuk analisis sentimen berbasis aspek pada pengguna twitter terhadap penanganan virus Corona. Adapun tujuan dari penelitian ini adalah Untuk mengetahui performansi metode Naive Bayes Classifier dan penggunaan aspek dalam menganalisis sentimen terhadap penanganan Virus Covid 19.

1.4 Batasan Masalah

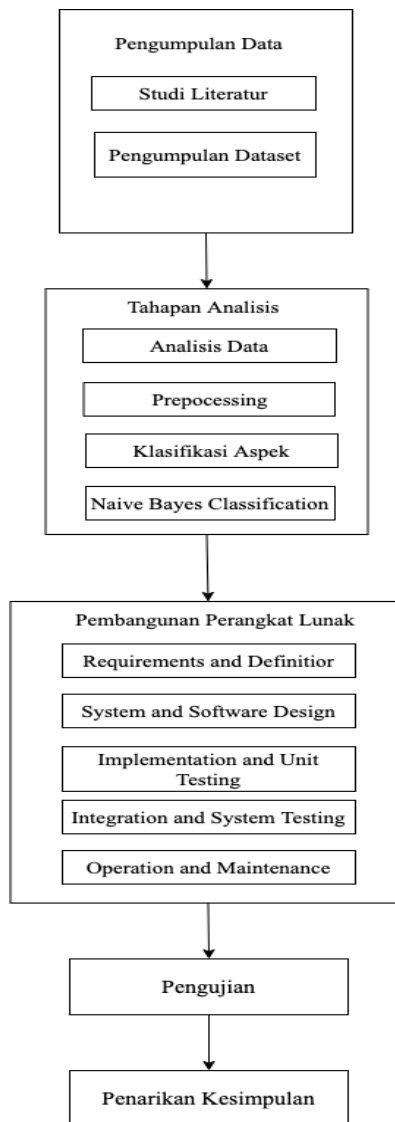
Agar penelitian yang dilakukan semakin terarah sesuai dengan tujuan yang diinginkan tercapai, maka batasan masalah dalam penelitian ini yaitu sebagai berikut.

1. Data yang digunakan diambil dari twitter dengan memanfaatkan twitter API.
2. Periode tweet yang digunakan adalah pada bulan Maret 2020 sampai Maret 2021
3. Penelitian ini hanya melakukan analisis terhadap tweets berbahasa Indonesia.

4. Akun Twitter yang dianalisa adalah @KawalCOVID19 dan @BNPB_Indonesia sebagai akun Informasi Penanganan Virus Covid 19 di Indonesia.
5. Data tweet diambil dari Twitter yang mengandung keyword “PSBB”, “Corona”, “Covid19”, dan “Penanganan”
6. Terdapat dua kelas sentiment yaitu sentimen positif dan sentiment negative.
7. Berdasarkan persebaran aspek pada dataset yang digunakan, maka pada penelitian ini setiap tweet akan diklasifikasikan dalam empat kelas aspek yaitu Kesehatan, Ekonomi, Transportasi, dan Pendidikan.
8. Format file yang digunakan pada data latih dan uji adalah .csv.

1.5 Metodologi Penelitian

Metodologi penelitian dalam penyusunan tugas akhir ini dibagi menjadi lima tahap yaitu, tahap pengumpulan data, tahap analisis metode, tahap pembangunan perangkat lunak, tahap pengujian dan tahap. Berikut tahapan penelitian yang akan dilakukan dengan penarikan kesimpulan alur seperti pada Gambar 1.1.



Gambar 1.1 Alur Penelitian

1.5.1 Metode Pengumpulan Data

Metode pengumpulan data yang dilakukan dalam penelitian ini yaitu sebagai berikut.

1. Studi literatur

Metode pengumpulan data yang dilakukan penelitian ini yaitu studi literatur. Studi ini dilakukan dengan cara mempelajari, meneliti dan menelaah berbagai literatur-literatur dari perpustakaan yang bersumber dari buku-buku, teks, jurnal ilmiah, situs-situs internet, dan bacaan-bacaan yang berkaitan dengan penelitian.

2. Pengumpulan Dataset.

Pengumpulan Dataset adalah proses pengumpulan data yang akan digunakan sebagai data masukan yang nantinya akan diproses oleh aplikasi yang akan dibangun. Dataset yang digunakan berupa data primer yang diambil secara langsung dari akun-akun twitter yang telah di tentukan, serta tweet yang di-mention terkait virus covid19 dengan memanfaatkan stream API (*Application Programming Interface*) yang disediakan oleh twitter.

1.5.2 Metode Tahapan Analisis

Adapun metode-metode yang akan digunakan sebagai berikut:

1. Analisis Data

Analisis Data, yaitu menganalisis pola kalimat dari data yang didapat, data yang dianalisis adalah kalimat data yang diambil dari kumpulan *tweets* bahasa Indonesia yang diambil dari *Twitter*. Data *tweets* ini diperoleh dengan membuat program crawling yang menggunakan *Twitter API*

2. Preprocessing

Preprocessing merupakan tahapan awal yang dapat dilakukan dalam tahap pengolahan data. Tujuan preprocessing adalah mentransformasikan data ke dalam suatu format data yang lebih mudah untuk diproses sesuai dengan kebutuhan penelitian.

3. Klasifikasi Aspek

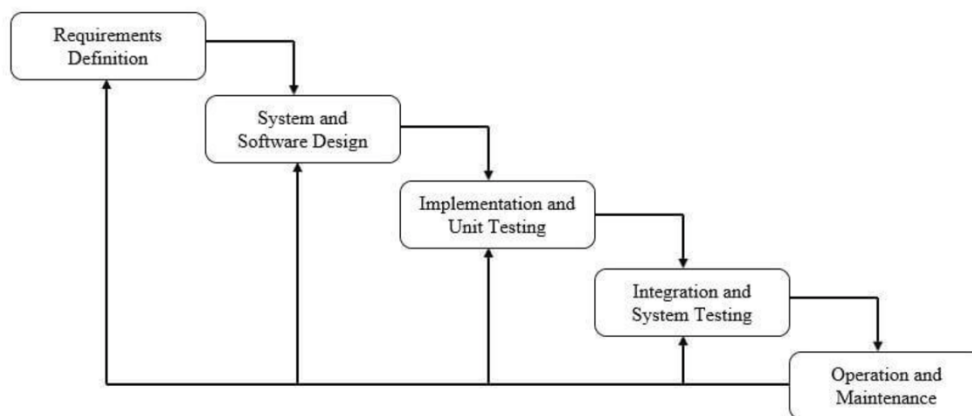
Klasifikasi Aspek merupakan tahapan pelabelan data yang dilakukan terhadap data teks uraian opini publik untuk setiap aspek. Klasifikasi aspek digunakan untuk menentukan jenis-jenis aspek yang terdapat pada teks komentar, dimana proses analisis sentimen akan lebih terperinci dan mendalam, karena teks telah dikelompokkan menjadi beberapa aspek yang berbeda - beda. Jika terdapat beberapa aspek-aspek yang berbeda tersebut, maka dibutuhkan klasifikasi multilabel menggunakan metode klasifikasi binari yang bertujuan melakukan pengidentifikasian dan menentukan kata yang memiliki kategori yang sama dan sehingga dapat menentukan aspek yang relevan [12].

4. Naive Bayes Classification

Langkah selanjutnya adalah proses klasifikasi yang menggunakan metode Naïve Bayes Classifier untuk menentukan termasuk pendapat positif dan yang termasuk pendapat negatif.

1.5.3 Metode Pembangunan Perangkat Lunak

Metode yang akan digunakan pada penelitian ini model klasik yang bersifat sistematis, berurutan dalam membangun software. Waterfall Model terdiri dari 5 tahapan untuk pengembangan. Berikut adalah penjelasan dari tahap – tahap yang dilakukan di dalam model ini menurut Sommerville: [13]



Gambar 1.2 Alur Model *Waterfall* Sommerville

Adapun tahapan dalam melaksanakan metode waterfall adalah.

1. *Requirements and definition*

Tahap ini merupakan tahap mengumpulkan kebutuhan secara lengkap kemudian menganalisis dan mengidentifikasi kebutuhan yang harus dipenuhi oleh program yang akan dibangun.

2. *System and software design*

Tahap ini merupakan tahap desain yang dilakukan setelah persyaratan lengkap telah terkumpul.

3. *Implementation and unit testing*

Tahap ini merupakan tahap menerjemahkan rancangan program ke dalam kode-kode dengan menggunakan bahasa pemrograman yang telah ditentukan. Program yang dibangun langsung diuji secara unit.

4. *Integration and system testing*

Tahap ini merupakan penyatuan unit-unit program kemudian diuji secara keseluruhan (pengujian sistem).

5. *Operation and maintenance*

Tahap ini merupakan tahap mengoperasikan program di lingkungannya dan melakukan pemeliharaan, seperti penyesuaian atau perubahan akibat adaptasi dengan keadaan yang sebenarnya.

1.5.4 Metode Pengujian

Dalam tahapan ini akan dilakukan pengujian akurasi dari metode algoritma Naive Bayes Classification yang akan diimplementasikan ke sistem analisis sentiment berdasarkan aspek terhadap sentiment masyarakat dalam penanganan virus Covid 19 dengan menampilkan akurasinya.

1.5.5 Penarikan Kesimpulan

Pada tahap ini akan dilakukan penarikan kesimpulan yang berdasarkan hasil dari tahap pengujian dalam penerapan metode Naive Bayes Classification pada kasus analisis sentimen berdasarkan aspek.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum mengenai penelitian yang dikerjakan. Sistematika penulisan penelitian sebagai berikut:

BAB 1 PENDAHULUAN

Pada bab ini berisi penjelasan mengenai latar belakang permasalahan, rumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, serta sistematika penulisan dalam penelitian tentang analisis sentimen berbasis aspek pada Twitter.

BAB 2 LANDASAN TEORI

Pada bab ini pembahasan umum rangkaian penjelasan yang berhubungan teori-teori atau tinjauan dari literatur pendukung yang akan digunakan untuk optimasi dalam menganalisis sentimen pada Twitter berbasis aspek. Adapun teori-teori yang akan ditulis pada Bab ini yaitu teori analisis sentimen, preprocessing (cleansing, case folding, tokenizing, convert negation, remove stopword dan stemming), klasifikasi aspek, metode Naive Bayes Classifier.

BAB 3 ANALISIS DAN PERANCANGAN

Pada bab ini dijelaskan secara rinci mulai dari analisis data, gambaran umum sistem, analisis proses, analisis pengujian, spesifikasi kebutuhan perangkat

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Pada bab ini ditampilkan implementasi dari hasil analisis dan perancangan yang sudah dilakukan sebelumnya menjadi sebuah aplikasi. Aplikasi digunakan untuk melakukan pengujian akurasi analisis sentimen terhadap tweet yang diatur dalam sebuah skenario pengujian. Informasi yang ditampilkan sebagai hasil dari pengujian yaitu persentase akurasi dari masing-masing pengujian pada nilai yang digunakan serta informasi tentang kelas aspek yaitu Kesehatan, Ekonomi, Transportasi, dan Pendidikan dan informasi kelas sentimen positif dan negative.

BAB 5 KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dan saran tentang penelitian yang dilakukan. Pada bagian kesimpulan, berisi tentang apakah penerapan metode berpengaruh terhadap tingkat persentase akurasi dalam klasifikasi untuk analisis sentimen tweet berbasis aspek pada Twitter atau tidak. Sedangkan pada bagian saran, berisi tentang hal-hal yang belum bisa terselesaikan pada penelitian yang telah dilakukan serta hal-hal yang bisa dilakukan untuk pengembangan selanjutnya.