



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

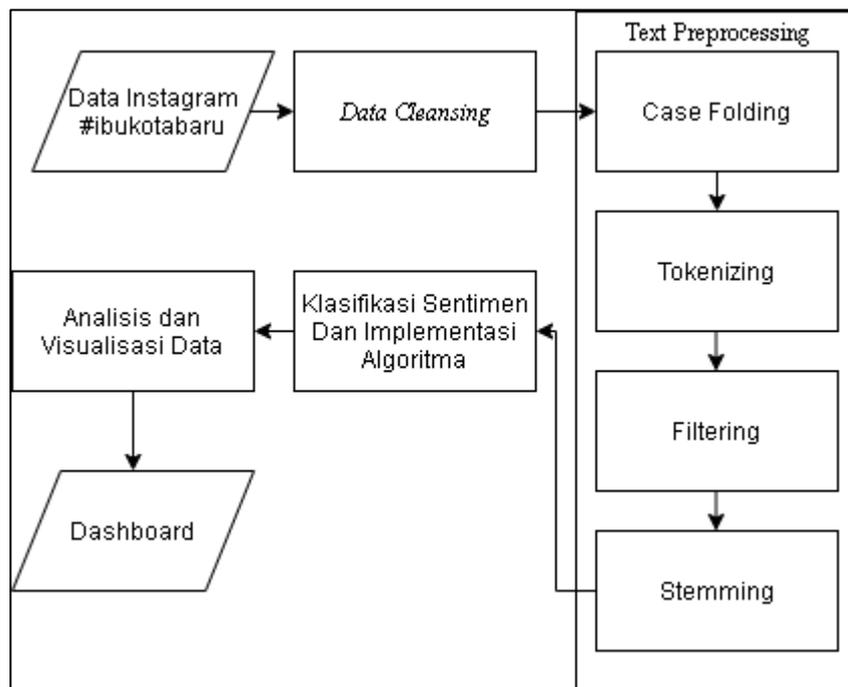
Objek penelitian pada penelitian ini adalah *hashtag* di aplikasi *Instagram* dengan kata kunci #ibukotabaru. Setelah itu akan diperoleh hasil dengan pembuatan sebuah *dashboard* menggunakan aplikasi *Power BI* dimana *dashboard* tersebut berfungsi untuk memberikan informasi sentimen para warga Indonesia mengenai pemindahan ibukota di Indonesia.

3.2 Metode Penelitian

3.2.1 Alur Penelitian

Pada penelitian ini menggunakan alur penelitian yang digunakan pada penelitian yang dilakukan oleh Retno Sari dengan judul penelitian “Analisis Sentimen *Review* Restoran menggunakan Algoritma *Naïve Bayes* berbasis *Particle Swarm Optimization*” (Sari, 2019) dan alur penelitian tersebut diubah sesuai dengan tujuan penelitian ini. Tahap pertama adalah pengumpulan data dari *Instagram* dengan memanfaatkan *instagram-scraper* berbahasa *python*. Kemudian dilakukan pembersihan data menggunakan aplikasi R. Selanjutnya adalah tahap *text preprocessing* untuk memperoleh data yang siap dipakai untuk menentukan sentimen dari kata tersebut. Kemudian hasil dari *text preprocessing* akan di klasifikasi menjadi 3 sentimen, yaitu positif, negatif, dan netral dan diimplementasikan algoritma *Naïve Bayes*, *Support Vector Machine* dan *Decision Tree C4.5*. Setelah itu hasil klasifikasi sentimen akan

dibuatkan sebuah *dashboard* yang berguna untuk memberikan informasi kepada pembaca. Untuk lebih jelasnya, alur penelitian dalam penelitian ini dapat dilihat pada gambar 3.1 berikut.



Gambar 3.1 Alur Penelitian

3.2.1.1 Pengumpulan Data

Dalam tahap ini, akan dilakukan pengumpulan data dari aplikasi *Instagram* dengan kata kunci #ibukotabaru. Pengumpulan data yang dilakukan dibantu dengan sebuah *instagram-scraper* berbahasa *python* kemudian diimplementasikan pengumpulan datanya menggunakan *software* yang bernama R.

3.2.1.2 Data Cleansing

Tahap ini akan dilakukan pembersihan data untuk atribut yang tidak mempengaruhi sentimen. Contoh atribut yang tidak penting adalah (#), dan

karakter simbol (~,!,@,\$,%,^,&*,?,-, dll). Pembersihan juga dilakukan terhadap data-data yang duplikat.

3.2.1.3 *Text Preprocessing*

Pada tahap *text preprocessing*, akan dibagi menjadi 4 tahap, yaitu:

1. *Case Folding*

Pada tahap *case folding* akan dilakukan pengkonversian teks ke dalam bentuk standar yaitu mengubah semua huruf dokumen menjadi huruf kecil. Sebagai contoh pada penelitian ini, pengguna yang ingin mendapatkan informasi “IBUKOTABARU” dan mengetik “IbukotaBaru”, “IbuKotaBaru”, “ibukotabaru”, tetap diberikan hasil yang sama yaitu “ibukotabaru”.

2. *Stemming*

Pada tahap *stemming* akan dilakukan pengelompokan kata-kata yang lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau *form* yang berbeda karena mendapatkan imbuhan yang berbeda.

3. *Tokenizing*

Pada tahap *tokenizing* akan dilakukan pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Sebagai contoh pada penelitian ini adalah ada kalimat dengan kata-kata “Setuju dengan rencana pemindahan ibukota baru”. Setelah dilakukan *tokenizing*, akan menjadi “setuju” “dengan” “rencana” “pemindahan” “ibukota” “baru”.

4. *Filtering*

Pada tahap *filtering* akan dilakukan penghapusan kata-kata umum seperti “yang”, “dan”, “di”, “dari”. Sebagai contoh pada penelitian ini adalah ada kalimat dengan kata-kata “Semoga ibukota baru yang dipindahkan akan membawa dampak yang baik bagi negara”. Setelah dilakukan *filtering* akan menjadi “semoga” “ibukota” “baru” “dipindahkan” “membawa” “dampak” “baik” “bagi” “negara”

3.2.1.4 Klasifikasi Sentimen dan Implementasi Algoritma

Dalam tahap ini, akan dilakukan klasifikasi sentimen dan implementasi algoritma dalam melakukan klasifikasi sentimen masyarakat Indonesia terhadap pemindahan ibukota baru. Algoritma yang akan dipakai pada penelitian ini ada 3, yaitu *Naïve Bayes*, *Support Vector Machine* dan *Decision Tree C4.5*. Klasifikasi sentimen tersebut dibagi menjadi 3 sentimen, yaitu positif, negatif dan netral.

3.2.1.5 Analisis dan Visualisasi Data

Dalam tahap ini, akan dilakukan analisis dan visualisasi data berdasarkan klasifikasi yang telah dilakukan pada tahap sebelumnya. Visualisasi data tersebut berupa sebuah *dashboard* menggunakan aplikasi *Power BI* yang berguna untuk memberikan informasi mengenai sentimen masyarakat Indonesia terhadap pemindahan ibukota baru di Indonesia.

3.2.2 Problem Solving

Dalam melakukan penyelesaian masalah dalam penelitian ini, digunakan algoritma *Naïve Bayes*, *Support Vector Machine*, dan *Decision Tree C4.5*. Ketiga algoritma pada penelitian ini digunakan untuk mencari tingkat akurasi

yang tertinggi dari ketiga algoritma tersebut dikarenakan ketiga algoritma ini sangat populer dalam bidang klasifikasi teks sentimen analisis. Berikut adalah perbandingan algoritma klasifikasi antara algoritma *Naïve Bayes*, *Support Vector Machine*, dan *Decision Tree C4.5* (Imran, 2015; Varghese, 2018):

Tabel 3.1 Perbandingan Algoritma

| <i>Naïve Bayes</i> | <i>Support Vector Machine</i> | <i>Decision Tree C4.5</i> |
|--|--|--|
| <i>Generative model.</i> | <i>Discriminative model.</i> | <i>Discriminative model.</i> |
| Tidak mengabaikan nilai kunci dalam <i>data training</i> dan membutuhkan sedikit data untuk <i>data training</i> . | Semakin banyak <i>data training</i> yang digunakan maka akan menyebabkan berkurangnya akurasi. | Pemangkasan model keputusan dapat mengabaikan beberapa nilai kunci dalam data training, yang dapat menyebabkan berkurangnya akurasi. |
| Hasil yang baik diperoleh di sebagian besar kasus dalam klasifikasi teks. | Hasil yang baik diperoleh di sebagian besar kasus dalam klasifikasi teks. | Hasil yang baik diperoleh sebagian besar kasus dalam prediksi rekomendasi. |
| Biasanya digunakan dalam klasifikasi teks, memfilter spam, rekomendasi sistem. | Biasanya digunakan dalam klasifikasi dan regresi. | Biasanya digunakan dalam prediksi dan klasifikasi tapi hanya 2 faktor. |
| <i>Supervised learning algorithm.</i> | <i>Supervised learning algorithm.</i> | <i>Supervised dan unsupervised learning algorithm.</i> |

Algoritma *Naïve Bayes* dikenal dengan pembuatan model yang sederhana namun berkinerja baik, terutama di bidang klasifikasi dokumen. Algoritma *Support Vector Machine* sendiri dikenal dengan algoritma yang banyak digunakan dalam polaritas teks. Sedangkan metode *Decision Tree* dengan algoritma *C4.5* sendiri dikenal sebagai *machine learning* yang biasanya digunakan dalam sebuah prediksi *data mining*. Dari perbandingan dapat disimpulkan bahwa algoritma

Naïve Bayes cocok untuk melakukan klasifikasi teks seperti yang akan dilakukan di dalam penelitian ini.

3.3 Variabel Penelitian

Variabel penelitian pada penelitian ini dibagi menjadi 2 variabel, yaitu *dependent* dan *independent*. Variabel *dependent* merupakan variabel yang berkaitan dengan apa yang menjadi ukuran atau faktor dari hasil daripada penelitian ini, yaitu sentimen masyarakat Indonesia terhadap pemindahan ibukota baru yang didapat dari #ibukotabaru di aplikasi *Instagram*. Sedangkan variabel *independent* hasil daripada penelitian ini yaitu sebuah *dashboard* dengan aplikasi *Power BI* dimana *dashboard* tersebut berfungsi untuk memberikan informasi sentimen masyarakat Indonesia terkait pemindahan ibukota baru

3.4 Teknik Pengumpulan Data

Data yang diperoleh merupakan sumber data yang diambil secara langsung dari aplikasi *Instagram* dengan kata kunci pencarian #ibukotabaru dengan memanfaatkan *python* yang kemudian diimplementasikan ke dalam *software* bernama R.

3.5 Teknik Pengambilan Sampel

Teknik pengambilan sampel yang digunakan pada penelitian ini adalah teknik *random sampling*. Teknik ini digunakan karena data yang diambil bersifat homogen, sehingga elemen manapun yang terpilih menjadi sampel dapat mewakili populasi. Data yang diambil untuk penelitian ini adalah kata-

kata sentimen dari aplikasi *Instagram* dengan kata kunci pencarian #ibukotabaru.

3.6 Teknik Analisis Data

Teknik analisis data yang digunakan pada penelitian ini menggunakan *software* yang bernama R. Selain R, ada juga *software* lain dalam melakukan analisis data yaitu *Python*. Berikut adalah perbandingan antara R dengan *Python* (Zeolearn, 2019):

Tabel 3.2 Perbandingan R dengan Python

| R | <i>Python</i> |
|--|---|
| Lebih diunggulkan dalam analisis teks. | Lebih diunggulkan dalam analisis yang berulang. |
| Data visualisasi sangat unggul dikarenakan R mempunyai <i>library package</i> yang mendukung visualisasi data. | <i>Python</i> juga memiliki <i>library package</i> tetapi masih dibawah R dalam konteks visualisasi data. |
| R merupakan juara dalam <i>data science</i> | <i>Python</i> lebih unggul dalam prediksi akurasi dan <i>machine learning</i> . |
| Analisis data dan perhitungan statistic | Manipulasi data dan <i>data mining</i> |
| Lebih banyak digunakan dalam dunia pendidikan | Lebih banyak digunakan dalam perusahaan maupun organisasi |

Dapat dilihat dari tabel 3.2 diatas, R sangat bagus dalam melakukan analisa teks dan juga visualisasi data sehingga dipilihlah aplikasi R untuk membantu melakukan klasifikasi sentimen dalam penelitian ini.