



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

3.1 Sifat dan Jenis Penelitian

Dari sudut pandang metodologi, paradigma pemrosesan informasi mengacu pada proses komputasi yang dianggap sebagai pendekatan yang bersifat instrumental-fudamental untuk memahami kompleksitas sosial (Cioffi-Revilla, 2018, p. 3). Hal tersebut membuat penelitian ini sesuai apabila disebut komputasional karena meneliti masalah sosial melalui alogaritma komputasi. Cioffi-Revilla mengemukakan bahwa simulasi sosial mampu merepresentasikan sistem sosial dan sosio-teknologi-natural dengan cara yang mana tidak dapat dicapai oleh pendekatan metodologi yang lain(2018, p. 233). Penelitian ini adalah kuantitatif dengan Paradigma konstruktivisme yang menurut Suyanto (2016, p. 65-66). Adalah:

memandang ilmu sosial sebagai analisis sistematis terhadap aksi yang bermakna sosial melalui pengamatan langsung terhadap pelaku sosial dalam setting yang alamiah untuk menafsirkan bagaimana para pelaku sosial itu menciptakan dan memelihara dunia sosial mereka . Pendekatan kuantitatif merupakan salah satu jalur yang dapat merealisasikan cara pandang konstruktivisme yang merupakan hasil bentukan masyarakat seperti sentimen dalam kolom komentar yang dibuat oleh masyarakat.

3.2 Metode Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan metode algoritma komputasi. Mengapa digunakan simulasi komputer sebagai metodologi untuk investigasi saintifik untuk kompleksitas sosial? Teori formal sosial yang kompleks terkadang lebih mudah diteruskan melalui pemodelan komputasi dibandingkan melalui solusi tertutup (Cioffi-Revilla, 2018, p. 223). Cioffi-Revilla menyatakan bahwa Kode kode yang diaplikasikan dalam bahasa pemrograman terpilih atau sistem simulasi menyediakan formalisme yang kuat untuk membuat teori, bereksperimen, dan memahami kompleksitas sosial (2018, p. 233).

Penelitian kuantitatif meneliti suatu kasus atau fenomena yang hadir dalam masyarakat secara mendalam dengan penggunaan angka, tabel, grafik, dan diagram untuk menampilkan hasil data/ informasi yang diperoleh. Menurut Lincoln dan Guba (1985) dalam *Metode Penelitian Sosial : Berbagai Alternatif Pendekatan*, dalam penelitian kuantitatif, masalah penelitian kerap dipahami sebagai suatu kondisi yang berasal dari hubungan antara dua faktor atau lebih yang menyebabkan situasi yang membingungkan (Suyanto, 2005, p. 17). Suyanto (2005, p.136) juga menyebutkan:

pada umumnya, para peneliti lebih sering menyebut penelitian kuantitatif sebagai penelitian yang hanya menggunakan instrumen dan/atau bermain main dengan angka angka saja.

Penelitian ini dilakukan dengan bantuan dari rekan peneliti yang kapabel untuk membuat program Naïve bayes. mereka adalah mahasiswa jurusan Human Computer Interaction (HCI) dari universitas Surya, yaitu Juan Sabanari dalam pembuatan dan penanganan tampilan dan fungsi software, dan Jonathan Fernando

dalam pembobotan data dan kalkulasi algoritma komputasi. Dengan skema penghitungan Naïve Bayes Classifier peneliti akan mengukur jenis sentimen yang muncul pada kolom komentar CNN Indonesia, yang pada penelitian ini berada dalam suasana politik menjelang pemilu

3.3 Populasi

Mengenai populasi, Gulo (2002, p. 76-77) menyatakan bahwa,

Populasi terdiri dari sekumpulan obyek yang menjadi pusat perhatian, yang mana mengandung informasi yang ingin diketahui dan Keseluruhan Satuan analisi yang merupakan sasaran penelitian disebut sebagai populasi.

Penghitungan yang dilakukan berasal dari kolom komentar portal berita online CNN Indonesia dimana terdapat komentar yang bermuatan sentimen didalamnya sebagai populasi.

3.4 Teknik Pengumpulan Data

teknik yang peneliti gunakan adalah teknik observasi di portal berita online CNN Indonesia yang memuat konten politik pilpres 2019. Pengamatan (observasi) merupakan metode pengumpulan data dengan mencatat informasi yang mereka saksikan dalam penelitian. Peristiwa yang disaksikan itu bisa dengan dilihat, didengarkan, dirasakan, yang lalu kemudian dicatat dengan objektif (Gulo, 2002, p. 116). peranan pengamat menurut Gulo (2002, p.116) dapat dibedakan berdasarkan hubungan partisipatifnya dengan kelompok yang diamatinya, yaitu :

- a. partisipan penuh
menyamakan diri dengan orang yang diteliti. Dengan demikian pengamat dapat merasakan dan menghayati apa yang diamati oleh responden. Tidak jarang seorang pengamat tinggal bersama dengan kelompok masyarakat yang diamatinya dalam waktu yang cukup lama sehingga ia dianggap sebagai bagian dari masyarakat yang bersangkutan.
- b. Partisipan sebagai pengamat

Masing masing pihak, baik pengamat maupun yang diamati, menyadari peranannya. Peneliti sebagai pengamat membatasi diri dalam berpartisipasi sebagai pengamat, dan responden menyadari bahwa dirinya adalah obyek pengamatan. Oleh karena itu, pengamat membatasi aktivitasnya dalam kelompok responden.

- c. pengamat sebagai partisipan
peneliti hanya berpartisipasi sepanjang yang dibutuhkan dalam penelitiannya.
- d. Pengamat sempurna (complete observer)
peneliti hanya menjadi pengamat tanpa partisipasi dengan yang diamati. Ia mempunyai jarak dengan responden yang diamatinya.

Data yang dikumpulkan berupa komentar dari kolom komentar berita seputar kampanye pilpres 2019 CNN Indonesia dalam rentang waktu 23 September 2018 - 13 April 2019. Semua komentar di salin kedalam tabel yang sudah diberi kategori agar nantinya kumpulan komentar tersebut tinggal dites melalui software naïve Bayes.

3.5 Teknik Analisa Data

Kolom komentar berita politik seputar pilpres 2019 terlebih dahulu diberi patokan pribadi untuk menentukan apakah sebuah kata mengandung sentimen negatif atau positif. Filtering yang digunakan berasal dari kolom komentar Youtube CNN Indonesia dengan tema yang sama. Skor (1) diberikan kepada komentar bermuatan sentimen negatif, sedangkan komentar bermuatan sentimen positif diberikan skor (0) Setelah memiliki skor, tiap kata pada kolom komentar diproses menggunakan program Naïve Bayes Classifier. Tahap ini bisa juga disebut preprocessing (pelabelan, tokenisasi dan pembobotan).

Algoritma yang sesuai dalam hal ini adalah multinomial naive-bayes yang merupakan salah satu algoritma yang paling populer untuk digunakan dalam

klasifikasi teks. Setiap fitur dalam multinomial naïve-bayes adalah kata-kata dari kumpulan data.

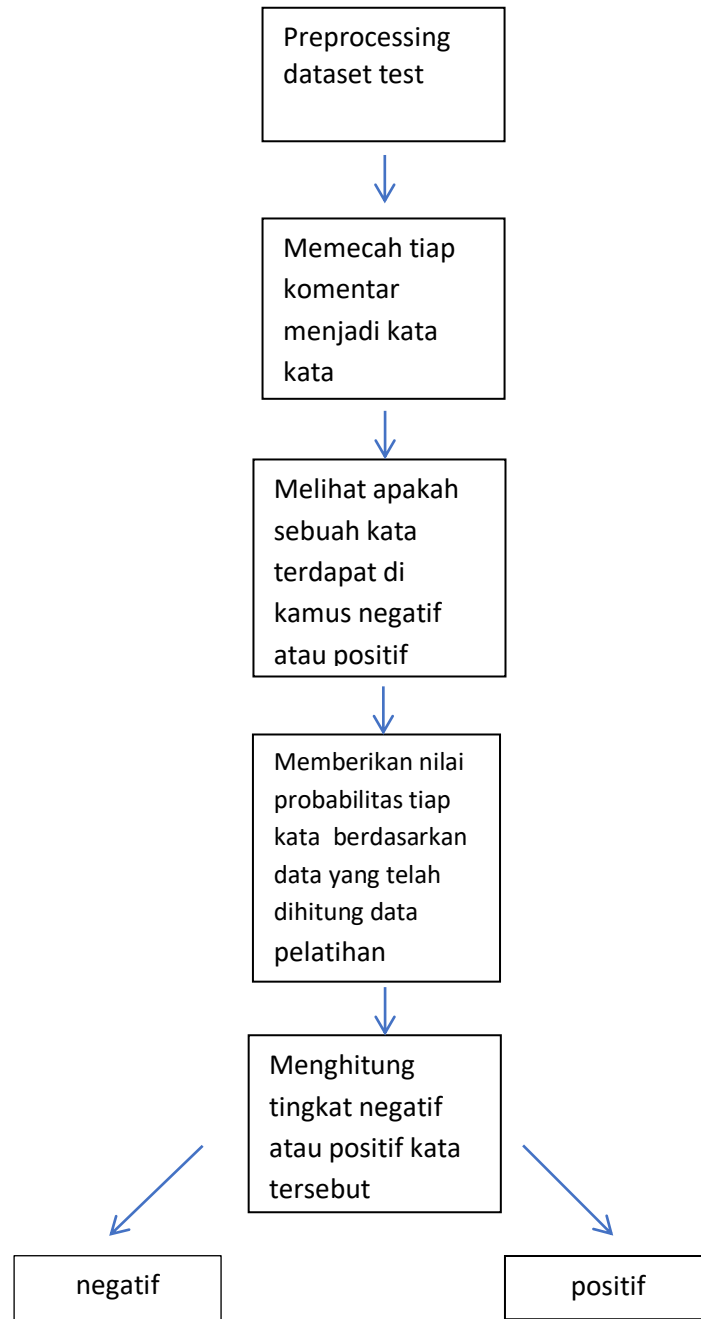
$$P(w_i|c) = \frac{\text{count}(w_i,c)}{\sum_{w \in V} \text{count}(w,c)} \quad (1)$$

Dengan menggunakan pola rumus Nazrul (2018), Persamaan (1) adalah persamaan kemungkinan dalam multinomial naïve-bayes. Di mana $\text{count}(w_i,c)$ adalah banyaknya kemunculan kata i di kelas c . Penyebut adalah jumlah kata yang dikenal dengan kosakata di kelas c . Multinomial naïve-bayes mengimplementasikan bayes theorem yang menghitung probabilitas kata yang diberikan kelas sebagai kemungkinannya.

Tetapi jika sistem menerima kata-kata baru, sistem akan menciptakan probabilitas 0 dan mempengaruhi perhitungan. Laplacian Smoothing membantu algoritma agar tidak terpengaruh oleh probabilitas dengan nilai 0. Laplacian Smoothing dalam multinomial naïve-bayes menambahkan 1 sebagai konstanta pada nominator. Maka, algoritma akan memberikan nilai probabilitas dalam kemungkinan dan tidak merusak nilai hasil untuk prediksi.

Secara teknis, software Naïve Bayes yang dibuat oleh Juan Sabanari sudah memuat algoritma yang telah disusun oleh Jonathan Fernando yang mana menjadi filter positif – negatif untuk teks yang dimasukkan kedalam aplikasi. Naïve bayes selanjutnya memprediksi tiap kata yang telah diberi nilai, apakah kata tersebut akan menjadi negatif atau positif tergantung oleh kombinasi probabilitas kata yang masuk. Semakin banyak kata yang masuk, semakin banyak pula Naïve Bayes mempelajari kata dan kombinasi baru untuk prediksi selanjutnya.

Bagan 2. Alur kerja prediksi Multinomial Naïve Bayes



3.6 Uji Keabsahan Data

Solokova, Japkowicz dan Szpakowicz (2006) mengajukan metode *Confusion matrix* untuk menghitung keakuratan, presisi, Recall, dan F-measure dengan menghitung perbedaan label sebelum prediksi dan setelah prediksi (p. 1). Data teks yang telah dikategorisasikan selanjutnya berikan pelatihan dengan algoritma multinomial naive-bayes yaitu 265 data pelatihan dan 65 data pengujian. Hasilnya adalah tabel evaluasi performa yang ditampilkan pada tabel 2. Hasil rata rata diatas 80% artinya akurasi prediksi adalah 4 dari 5.

Tabel 2. Akurasi

Evaluasi Performa	Hasil
Accuracy	85.2459016393%
Precision	80.487804878%
<i>Recall</i>	97.0588235294%
F-Measure	88.0%

Sumber : Jonathan Fernando

Hasil evaluasi performa menjadi indikator keakuratan Naïve bayes dalam Memprediksi. Nilai recall yang cukup tinggi yaitu 97% yang berarti bahwa algoritma ini ketat dalam memprediksi data negatif. Kelemahannya adalah terkadang prediksi data salah mengklasifikasi data sebagai data negatif sedangkan data asli bukanlah negatif, begitu pula sebaliknya dengan data positif. F-Measure menunjukkan nilai keseimbangan antara presisi dan recall, yang dapat menunjukkan ketidakseimbangan prediksi. Hasil F Measure adalah 88% yang seimbang (Fernando,2019, p. 302).