IMPLEMENTASI NAÏVE BAYES UNTUK ANALISIS SENTIMEN PUBLIK TERHADAP BENTUKAN KABINET BARU INDONESIA 2019 PADA MEDIA ONLINE

Yella Guspita[1], Leon Abdillah[2]
Fakultas Teknik Ilmu Komputer, Universitas Bina Darma
Email: yellaguspita04@gmail.com[1], leon.abdillah@yahoo.com [2]

ABSTRACT

The increasingly widespread use of social media, such as making social media a source of information that is in great demand by the Indonesian people. One of the uses of this data is to find out the opinions or sentiments of social media users on a topic. One of the important topics to be discussed is the formation of the new 2019 Indonesian cabinet which will be used as the domain in this research. This research describes a sentiment analysis result which is divided into three classes, namely positive, negative and neutral, with data collection from social media information which will then be classified using the Naïve Bayes algorithm. The results of this study will later display information on the analysis of public sentiment towards the formation of the new Indonesian cabinet 2019.

Keywords: *Sentiment Analysis, Online Media, Nave Bayes.*

ABSTRAK

Makin maraknya penggunaan media sosial sepertimenjadikan media sosial tersebut sebagai sumber informasi yangsangat diminati masyarakat indonesia. Salah satu pemanfaatan data ini adalah untukmengetahui opini atau sentimen pengguna media sosialterhadap suatu topik. Salah satu topik yang penting untuk dibahas adalah pembentukan kabinet baru Indonesia 2019 yang akan dijadikan sebagai domain dalam penelitian ini. Peneltian ini memaparkan sebuah hasil analisis sentimen yang terbagi menjadi tiga kelas yaitu positif,negatif dan netral, dengan pengumpulan data dari informasi media sosialyang kemudian selanjutnya akan dilakukan klasifikasi menggunakan algoritma Naïve Bayes. Hasil dari penelitian ini nantinya akan menampilkan informasi analisis sentimen masyarakat tehadap pembentukan kabinet baru indonensia 2019.

Kata kunci: Analisis Sentimen, Media Online, Naïve Bayes.

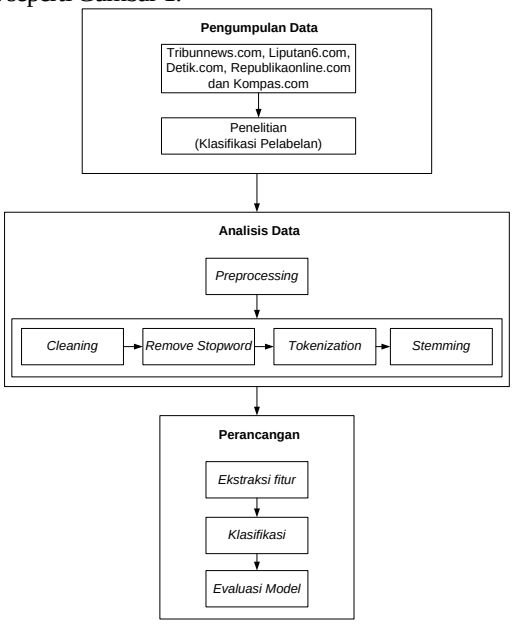
1. PENDAHULUAN

Banyaknya opini masyarakat pada media sosial yang membahas mengenai pembentukan kabinet baru 2019. Opini tersebut bersifat subyektif. Opini tersebut merupakan textual data yang dapat dianalisa dan dimanfaatkan untuk mendapatkan informasi guna memantau sentimen masyarakat terhadap pemerintahan baru Joko Widodo 2019 di media sosial menggunakan teknik text mining yaitu sentimen analisis. Sentimen tersebut dikategorikan menjadi sentimen positif, netral dan negatif. Penelitian ini memodelkan proses klasifikasi sentimen pengguna media sosial terhadap pembentukan kabinet baru 2019 pemerintahan Joko Widodo untuk mengetahui kondisi sentiment masyarakat indonesia menggunakan metode Naïve Bayes Classifier (NBC).[1]

Ruang lingkup media sosial sangatlah luas di dalamnya kita dapat mengakses berbagai kebutuhan informasi, karena pada dasarnya media yang satu ini dapat memberikan informasi yang cepat walaupun belum tentu akurat. Beberapa media sosial yang sedang aktif dalam dunia perpolitikan saat ini menurut penelitian penulis antara lain Tribunnews.com, Liputan6.com, Detik.com, Republikaonline.com dan Kompas.com. Media ini yang akan peneliti jadikan dataset analissi sentiment mengenai pembentukan kabinet baru Indonesia 2019. Facebook merupakan media sosial yang paling banyak digunakan di Indonesia untuk kampanye presiden. Pada pemilihan presiden periode 2014-2-19, Prabowo Subianto berhasil mendapatkan jumlah penggemar terbanyak di Facebook. Sedangkan Joko Widodo berhasil mendapatkan jumlah pengikut terbanyak di Twitter. Pengguna facebook didominasi oleh pengguna pada rentang usia 18-24 tahun, dan kota terbanyak yang menggunakan Facebook adalah Jakarta.[2]

2. METODOLOGI PENELITIAN

Metode penelitian ini memiliki suatu rancangan bagaimana alur sistem penelitian ini akan berjalan. Alur penelitian akan terlihat seperti Gambar 1.



Gambar 1. Alur Peneltian

2.1 Pengumpulan Data

Pada tahap ini dilakukan pengelolaan data yang sudah diambil, data yang akan diolah didapat dari media sosial seperti Tribunnews.com, Liputan6.com, Detik.com, Republikaonline.com dan Kompas.com . Pada penelitian ini peneliti menggunakan bahasa pemrogamam Python untuk melakukan scraping data dalam pengumpulan komentar masyarakat mengenai pembentukan kabinet 2019 Joko Widodo. Hasil pengumpulan data didapatkan ±1000 komentar yang sudah di filtering untuk dilakukan analisis data lebih lanjut.**[3]**

2.1.1 Penelitian Data

Hasil dari proses crawling data dalam file excel pada Gambar 3.5 di atas yang nantinya akan dilakukan proses labelling data untuk menentukan klasifikasi pendapat atau pandangan dari hasil tweet yang telah di crawling tadi. Pada proses labelling ini dibedakan menjadi 2 kelas. Yaitu class positif, class negative dan class netral.[4]

2.2 Analisis Data

Pada tahap ini dilakukan analisis data yang sudah diambil, melalui beberapa tahapan. Berikut tahapan yang akan dilalui.

2.2.1 Pengumpulan Data

Preprocessing adalah tahapan proses untuk membersihkan data dari kata-kata atau tweet yang tidak di perlukan serta kata-kata yang tidak memiliki makna. Proses ini dilakukan sesuai dengan isi data dari proses pengambilan data atau crawling data Twitter. Adapun proses beberapa langkah dari proses preprocessing memilki urutan sebagai berikut :

a. Cleaning

Cleaning adalah proses penghapusan simbol, tanda baca, huruf kapital dan bilangan angka yang sering muncul pada tweet pengguna Twitter sehingga data tersebut menjadi data yang tidak efektif dan tidak memiliki arti seperti : (#\\S+", "", x, !,()).[5]

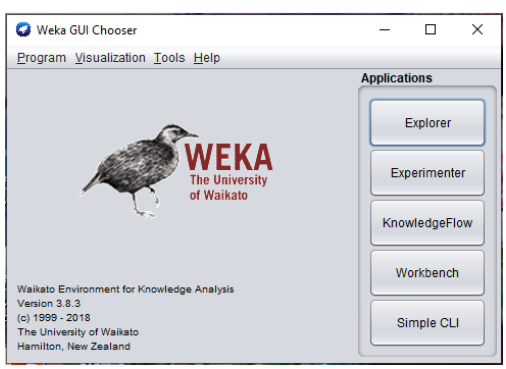
2.2.2 *Algoritma Naïve Bayes Classifier*

Data yang telah melalui tahap pemrosesan teks kemudian dapat dilanjutkan ke tahap selanjutnya yaitu klasifikasi dengan algoritma Naïve Bayes Classifier. Data dalam bentuk teks akan muncul dua hasil klasifikasi teks yang mengandung positif, netral, dan negatif. Berikut ini adalah perhitungan algoritma Naïve Bayes Classifier.

1. Tahap awal dalam proses Naïve Bayes Classifier adalah menghitung probabilitas masing-masing kelas dari keseluruhan data training.
2. Proses testing. Proses ini intuk mengetahui keakuratan model yang dibangun pada proses training, umumnya digunakan data yang disebut test set untuk memprediksi label. Metode Naïve Bayes Classifier terdiri dari dua tahap dalam proses klasifikasi teks, tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sample dokumen berupa pemilihan vocabulary yaitu kata yang dimungkinkan muncul dalam koleksi dokumen sample yang menjadi representasi dokumen. [6]

3. HASIL DAN PEMBAHASAN

Tools yang peneliti gunakan yaitu Weka 3.8, Tampilan dari Weka 3.8 seperti Gambar 2 di bawah ini.



Gambar 2. Weka 3.8

3.1 Pengolahan Data

Proses pertama adalah mengambil hasil pengolahan data pembentukan kabinet 2019 Joko Widodo yang sudah melalui tahapan pelabelan apakah termasuk Positif dan Negatif. Hasil proses pengolahan data akan tampak seperti Gambar 2.

	A	B	C	D	E	F	G	H	I	J
1	id_pembu	id_status	nama_per	waktu_sta	id_komen	id_user_k	komentar	waktu_ko	label	
2	2.34E+10	233830611	Bang Nafi	########	101520762	7.95E+14	Oke pak boss!!	########	positif	
3	2.34E+10	233830611	Sumy Kak	########	101520762	3.55E+14	Jangan dusta	########	negatif	
4	2.34E+10	233830611	Roberto R	########	101520762	2.98E+14	PROBOWO is the	########	positif	
5	2.34E+10	233830611	Sugiarmar	########	101520762	2.99E+14	ilmu padi	########	netral	
6	2.34E+10	233830611	Rio Corps	########	101520762	1.52E+15	Hidup Pak Prabov	########	positif	
7	2.34E+10	233830611	Permas At	########	101520762	4.73E+12	Ganteng bgt Pak :	########	positif	
8	2.34E+10	233830611	Taufik	########	101520762	1.51E+15	enak, di lhat tapa	########	netral	
9	2.34E+10	233830611	Anca Ardi	########	101520762	7.25E+14	ayo pak rujuk lagi	########	positif	
10	2.34E+10	233830611	Rizaldi Pil	########	101520762	8.85E+14	Prabowo insya Al	########	positif	
11	2.34E+10	233830611	Ichy Da ?o	########	101520762	5.46E+14	HIDUP JOKOWI.	########	negatif	
12	2.34E+10	233830611	Andref Se	########	101520762	3.12E+14	Semoga Pak Prab	########	positif	
13	2.34E+10	233830611	Ade Surya	########	101520762	3.12E+14	Prabowo Subiant	########	positif	
14	2.34E+10	233830611	Yazid Bust	########	101520762	9.33E+14	100% dukung Pral	########	positif	
15	2.34E+10	233830611	Kencrets T	########	101520762	4.65E+14	Jokowiku mana jo	########	negatif	
16	2.34E+10	233830611	Marco Reu	########	101520762	6.46E+14	hidup jokowi	########	negatif	
17	2.34E+10	233830611	Danny Cha	########	101520762	1.02E+16	Duh gagal pertam	########	netral	
18	2.34E+10	233830611	Ocu Ardy I	########	101520762	8.61E+14	Br 1 menit 1000 o	########	netral	
19	2.34E+10	233830611	Indra Detr	########	101520762	6.14E+14	semoga terpilih ja	########	positif	
20	2.34E+10	233830611	Marshal M	########	101520762	7.06E+14	mantap pak...	########	positif	
21	2.34E+10	233830611	Hugh Butl	########	101520762	1.48E+15	Saya punya	########	positif	
22	2.34E+10	233830611	MahaKing	########	101520762	5.51E+14	Prabowo RI 1 # ga	########	positif	
23	2.34E+10	233830611	Susan Rah	########	101520762	7.14E+14	Senyum nyaaa...	########	positif	
24	2.34E+10	233830611	Salahudin	########	101520762	7.76E+14	mntapp,,, yakinn	########	positif	
25	2.34E+10	233830611	Hariri Thai	########	101520762	1.56E+15	onde mande gant	########	positif	

Gambar 3. Excel Hasil Proses Pengolahan Data Komentar

Dari proses data komentar yang sudah dikumpulkan dilakukan proses transformasi seperti pada penejelasan bab sebelumnya sehingga data yanag akan diolah menjadi seperti Gambar 3.3 di bawah

	A	B	C
1	komentar	label	
2	Oke-pak-boss!!	positif	
3	Jangan-dusta	negatif	
4	PROBOWO-is-the-WINNER-of-PEM	positif	
5	ilmu-padi	netral	
6	Hidup-Pak-Prabowo-	netral	
7	Ganteng-bgt-Pak---	positif	
8	enak,-di-lhat-tapa-beban	netral	
9	ayo-pak-rujuk-lagi-sama-bu-titiek	netral	
10	Prabowo-insya-Allah-jadi-presiden	positif	
11	HIDUP-JOKOWI	netral	
12	Semoga-Pak-Prabowo-Jadi-Preside	netral	
13	Prabowo-Subianto---engkau-inspir	positif	
14	Jokowiku-mana-jokowiku	negatif	
15	hidup-jokowi-	negatif	
16	Duh-gagal-pertamax	netral	
17	Br---menit------org-udah-like	netral	
18	semoga-terpilih-jadi---	positif	
19	mantap-pak	netral	
20	Prabowo-RI-----gak-ada-yg-Laen	positif	
21	onde-mande-ganteng-bana-presid	positif	
22	Kmi-ada-di-luar-negri-siap-mmena	netral	
23	Bapak-org-yg-sangat-presistent-me	netral	
24	Saya-dan-keluarga-saya-di-tangera	positif	
25	benar-pak,senyum-itu-membawa-b	positif	

Gambar 4. Excel Hasil Data yang akan Dilakukan Uji

3.2 Preprocessing dengan Weka

3.2.1 Memasukkan data uji ke dalam Weka

Pada tahap ini peneliti melakukan input data ke dalam Weka, untuk menginput data ke dalam weka, berikut input data ke dalam weka.

1.Karena software yang digunakan untuk data mining merupakan WEKA maka dataset di atas terlebih dahulu di-export kedalam format (.xlsx) kemudian dari format (.xlsx) akan dirubah lagi menjadi format “csv” (Comma Delimited).

2.Kemudian pada weka klik open file, yang bisa diinput adalah format CSV (Comma Delimited).

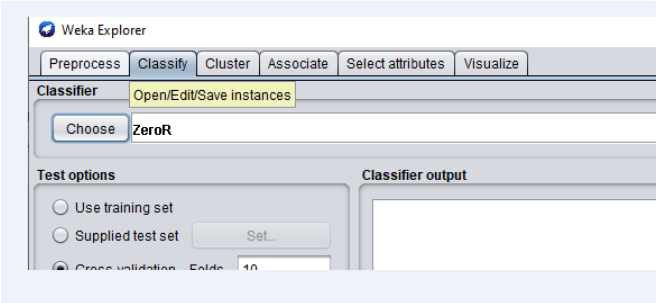
3.3 Pembahasan

Sekarang peneliti akan meakukan pembahasan klasifikasi menggunakan metode Naïve Bayes pada pemrograman Phyton.

3.3.1 Implementasi Klasifikasi Naïve Bayes Pada Weka

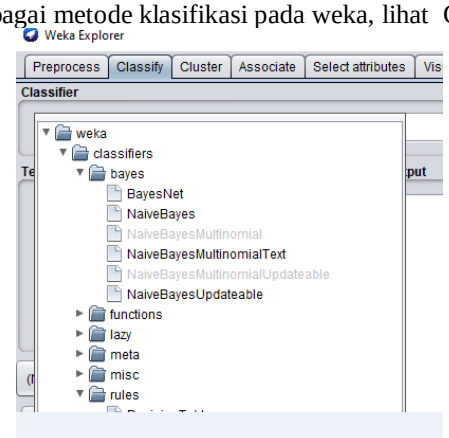
Pada tahapan ini peneliti akan melakukan klasifikasi data setelah semua proses Preprocessing selesai, metode yang akan digunakan yaitu Naïve Bayes pada weka.

1. Halaman Classify pada weka, lihat Gambar 4.13 di bawah ini.



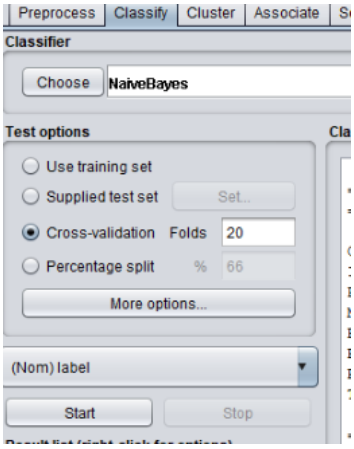
Gambar 5. Menggunakan Metode Naïve Bayes

2. Memilih Naïve Bayes sebagai metode klasifikasi pada weka, lihat Gambar 4.14 di bawah ini.



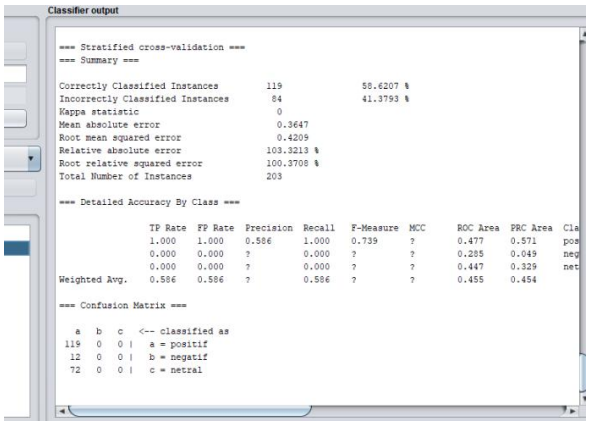
Gambar 6. Menggunakan Metode Naïve Bayes

3. Selanjutnya menjalankan fungsi Naive Bayes, dengan presentase split sebesar 20%.



Gambar 7. Fungsi Naïve Bayes

4. Hasil keberhasilannya 58% dengan error 41%.



```
Classifier output

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         119               58.6207 %
Incorrectly Classified Instances        84               41.3793 %
Kappa statistic                          0
Mean absolute error                      0.3647
Root mean squared error                  0.4209
Relative absolute error                103.3213 %
Root relative squared error            100.3708 %
Total Number of Instances              203

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Cla
                 1.000    1.000    0.586      1.000    0.739      ?        0.477     0.571     pos
                 0.000    0.000    ?          0.000    ?          ?        0.285     0.049     neg
                 0.000    0.000    ?          0.000    ?          ?        0.447     0.329     net
Weighted Avg.    0.586    0.586    ?          0.586    ?          ?        0.455     0.454

=== Confusion Matrix ===

   a   b   c   <-- classified as
 119   0   0 |   a = positif
  12   0   0 |   b = negatif
  72   0   0 |   c = netral
```

Gambar 8. Hasil Klasifikasi Naive Bayes

Dari hasil proses pengklasifikasian data ini, peneliti menggunakan data dokumen data komentar, jumlah data yang dijadikan sebagai bahan pengujian diambil dari data training sebesar 20% sama dengan 0,20 yang diacak dari 33% atau 0,33 dari data training. Proses pengklasifikasian data ini dilakukan dengan menggunakan perhitungan probabilitas dari setiap kelas, maka peneliti baru bisa mendapatkan hasil jelas dari prediksi data yang di-input. Tahapan akhir setelah melakukan semua proses pengklasifikasian, maka barulah bisa menghitung dari performa dari algoritme yang dipergunakan.

4. KESIMPULAN

a. Metode Naive Bayes dapat mengklasifikasikan data berupa teks, terutama teks yang berasal dari data komentar di media sosial seperti Tribunnews.com, Liputan6.com, Detik.com, Republikaonline.com dan Kompas.com.
b. Jumlah kata dalam tiap kelas training sangat mempengaruhi hasil klasifikasi pada data testing, oleh karena itu keseimbangan data perlu dijaga.
c. Kosakata tidak baku dapat mempengaruhi hasil klasifikasi terhadap suatu kelas testing apabila suatu kelas training memiliki data jumlah kata tidak baku lebih banyak dibanding kelas training lain.

DAFTAR PUSTAKA

[1] M. Kamil and W. Cholil, “Analisis Perbandingan Algoritma C4.5 dan Naive Bayes pada Lulusan Tepat Waktu Mahasiswa di Universitas Islam Negeri Raden Fatah Palembang,” *JI. Jurnal. Informatika*, vol. 7, no. 2, pp. 97–106, Sep. 2020, doi: 10.31294/ji.v7i2.7723.
[2] F. Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter,” *ISI*, vol. 3, no. 1, p. 50, Jun. 2018, doi: 10.35314/isi.v3i1.335.
[3] D. Wahyudi and A. H. Mirza, “IMPLEMENTASI DATA MINING DENGAN NAIVE BAYES CLASSIFIER UNTUK MENDUKUNG STRATEGI PROMOSI (Studi kasus Universitas Bina Darma Palembang),” p. 10.
[4] N. C. Siregar, R. R. A. Siregar, and M. Y. D. Sudirman, “Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ),” vol. 3, no. 1, p. 9, 2020.
[5] T. Kaihatu, “Analisa Kesenjangan Kualitas Pelayanan dan Kepuasan Konsumen Pengunjung Plaza Tunjungan Surabaya.” 2008.
[6] M. Basuki, H. Mz, S. Aprilyanti, and M. Junaidi, “PERANCANGAN SISTEM KESEIMBANGAN LINTASAN PRODUKSI DENGAN PENDEKATAN METODE HEURISTIK,” *Jurnal Teknologi*, vol. 11, no. 2, p. 9, 2019.