

ANALISIS SENTIMEN TERHADAP PENGGEMAR K-POP DI MEDIA SOSIAL TWITTER MENGGUNAKAN NAIVE BAYES (STUDI KASUS PENGGEMAR GRUP BTS)

Puji Nurmawati¹, Endang Supriyati², Tri Listyorini³
^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknik,
Universitas Muria Kudus
pujinurmawati07@gmail.com, endang.supriyati@umk.ac.id,
trilistyorini@umk.ac.id

Abstrak— Perkembangan *Twitter* sebagai platform yang digandrungi masyarakat terbukti dengan data Statista [1] yang menunjukkan bahwa Indonesia peringkat tujuh di dunia dengan banyak pengguna mencapai 13.2 juta. Banyak pengguna mengutarakan pendapat di dalam *Twitter*. Termasuk juga dengan ujaran kebencian hingga perundungan. Menjadikan penggemar BTS atau yang biasa disebut *Army* sebagai objek analisis karena grup BTS yang telah mendunia ini memiliki pengikut yang cukup banyak di *Twitter* yakni 31 juta pengikut. Karena hal ini maka analisis sentimen dilakukan Kesulitan dalam melakukan analisis data yang begitu banyak dari *Twitter* jika dilakukan dengan cara manual, maka dari itu komputerisasi dibutuhkan. Menggunakan sebuah program yang dibuat dengan bahasa pemrograman *Python* analisis dapat dilakukan lebih efektif dan efisien. Dari analisis yang dilakukan menggunakan klasifikasi algoritma *Naive Bayes* terdapat polaritas sentimen negatif sebanyak 34,2%, netral 58,5%, dan positif 7,3%. Dari 1000 data yang diambil sesuai hasil polaritas *tweet* yang bernilai negatif sesuai hasil polaritas sebanyak 342. Dengan tingkat akurasi sebesar 75%. Dari penelitian ini diharapkan dapat membantu dalam proses analisis sentimen dan sesuai dalam mengatasi masalah yang ada.

Kata Kunci— Analisis Sentimen, *Army*, *Twitter*, *Python*, *Naive Bayes*.

I. PENDAHULUAN

Perkembangan *Twitter* sebagai platform yang digandrungi masyarakat terbukti dengan data [1] yang menunjukkan bahwa Indonesia peringkat tujuh di dunia dengan banyak pengguna mencapai 13.2 juta. Banyak pengguna mengutarakan pendapat di dalam *Twitter*. Termasuk juga dengan ujaran kebencian hingga perundungan. Menjadikan penggemar BTS atau yang biasa disebut *Army* sebagai objek analisis karena grup BTS yang telah mendunia ini memiliki pengikut yang cukup banyak di *Twitter* yakni 31 juta pengikut. Karena hal ini maka analisis sentimen dilakukan Kesulitan dalam melakukan analisis data yang begitu banyak dari *Twitter* jika dilakukan dengan cara manual, maka dari itu komputerisasi dibutuhkan. Menggunakan sebuah program yang dibuat dengan bahasa pemrograman *Python* analisis dapat dilakukan lebih efektif dan efisien. Analisis dalam penelitian ini dilakukan menggunakan klasifikasi algoritma *Naive Bayes*. Langkah-langkah pada analisis ini ada 3 proses utama, yaitu *scraping*, *text pre-*

processing, dan klasifikasi. Untuk selanjutnya hasil klasifikasi berupa positif, negatif, dan netral.

II. METODOLOGI PENELITIAN

Pada penelitian ini data yang diambil merupakan *tweet* dari *Twitter* yang diambil dengan cara *scraping*. *Scraping* sendiri adalah sebuah kegiatan mengekstraksi data yang didapat dari sebuah situs web [2]. Dari data inilah akan dilakukan analisis dengan mengacu pada nilai sentimen data yang ada. Ini disebut dengan analisis sentimen. Dalam analisis yang dilakukan, perhitungannya menggunakan algoritma klasifikasi *Naive Bayes*.

A. Metode Pengumpulan Data

Metode pengumpulan data yang dilakukan dalam penelitian ini ialah menggunakan studi pustaka. Pedoman dalam penelitian ini berdasarkan penelitian-penelitian sebelumnya baik berupa artikel maupun jurnal. Dari berbagai metode pengumpulan data ini nanti penulis akan dapat mengetahui bagaimana sistem ini akan dibuat untuk melakukan sebuah analisis sentimen terhadap *Army*.

B. Perancangan Program Analisis

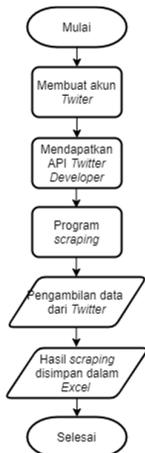
Alur secara umum untuk membuat program analisis dapat dilihat pada Gbr diagram alir berikut:



Gbr 1. Diagram Alir Program Secara Umum

Gbr (1) menjelaskan tentang alur secara umum yang akan dilakukan dalam program. Dimulai dengan melakukan *scraping* dalam mencari data. Data yang didapat harus dipersiapkan terlebih dahulu agar dapat diolah pada tahap selanjutnya, ini disebut *text pre-processing*. Setelah data siap diolah maka dapat diklasifikasikan sesuai kelasnya apakah nanti hasilnya positif, negatif, atau netral.

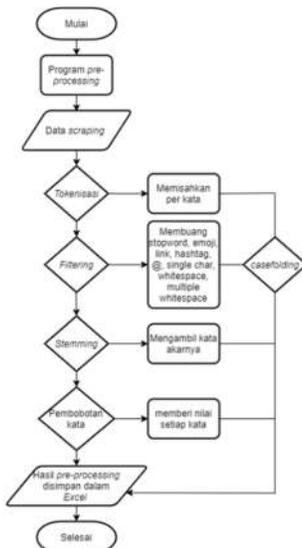
1) *Scraping*: Pada proses *scraping*, alur yang akan dijalankan dapat dilihat seperti Gbr dibawah:



Gbr 2. Diagram Alir Proses Scraping

Pada Gbr (2) menunjukkan proses pengambilan data dari *Twitter* atau disebut *scraping*. Sebelum melakukan *scraping* data kita membutuhkan program *scraping* dan API dari akun *Twitter* kita. Data yang diambil disimpan kedalam file *csv*.

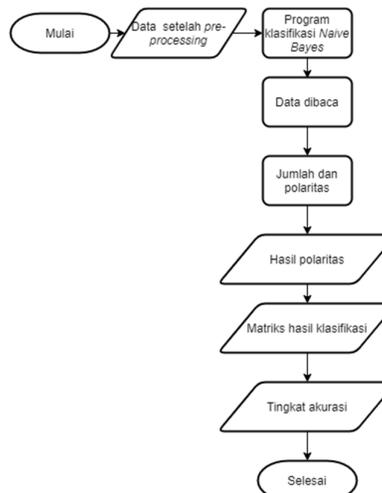
2) *Text Pre-processing*: Data dari *scraping* selanjutnya akan di *Pre-processing* sebelum dapat diklasifikasikan. Alurnya adalah sebagai berikut :



Gbr 3. Diagram Alir Proses Text Pre-processing

Gbr (3) menunjukkan alur dalam proses *text pre-processing*. Sebuah proses sebelum klasifikasi dimana data yang didapat akan disiapkan sesuai langkah-langkah yang ada agar pengklasifikasian data dapat dilakukan dengan akurat. Tahapannya ada 4: tokenisasi, filterisasi, stemming, dan pembobotan kata.

3) *Klasifikasi*: Dalam pengklasifikasian data yang ada akan melalui alur sebagai berikut:



Gbr 4. Diagram Alir Proses Klasifikasi

Pada diagram alur diatas, nantinya akan menghasilkan nilai polaritas dari datanya dan akurasi pengklasifikasian, dan matriks hasil sentimen. Algoritma yang digunakan dalam penelitian ini adalah algoritma klasifikasi *Naïve Bayes*. klasifikasi *Naïve Bayes* sendiri memiliki persamaan [3] umum sebagai berikut:

$$P(H|X) = \frac{P(x | H)P(H)}{P(X)} \quad (1)$$

- X : Data dengan class yang belum diketahui.
- H : Hipotesis data merupakan suatu *class* spesifik.
- P(H|X) : Probabilitas hipotesis berdasar kondisi (posteriori probability).
- P(H) : Probabilitas hipotesis (prior probability).
- P(X|H) : Probabilitas berdasarkan kondisi pada hipotesis.
- P(X) : Probabilitas H.

Persamaan (1) di atas menjelaskan tentang algoritma yang akan digunakan dalam penelitian ini. Algoritma ini dipilih karena tingkat akurasi yang cukup tinggi [4].

III. HASIL DAN PEMBAHASAN

1) *Scraping*: menggunakan data yang didapat dengan *scraping* dari *Twitter* analisis dalam penelitian ini dilakukan. Kata kunci untuk analisis tentu saja objek kita yakni *Army BTS*.

Dengan *scraping* didapat 1000 data. Hasil data *scraping* dapat dilihat seperti dibawah:

TABEL I
 DATA SCRAPING

Keyword	Tanggal	Username	Tweet
Army bts	12/13/2020 4:28	LovitaCendana	Pertama kali aku ngestan kepop ya sama bts, terus orang-orang pada bilang: “kok mau sih jadi Army? Kan Army toxic. Ya sesantainya aja gw jawab: “Gw ngestan BTS bukan fansnya.

2) *Text Pre-processing*: data yang didapat kemudian disiapkan sebelum pengklasifikasian. Ada 4 langkah yang akan dilakukan, yang pertama tokenisasi. Tokenisasi adalah proses memisahkan suatu teks menjadi per kata yang dipisahkan spasi dan tanda baca. Kedua yakni *Filtering*. *Filtering* atau penyaringan dari data yang telah di tokenisasi meliputi *Stopword*, *emoji*, *link*, *hashtag*, *@*, *single character*, *whitespace*, *multiple whitespace*, dan *casefolding*. *Stopword* sendiri merupakan kata-kata tak bermakna [5]. Sedangkan *casefolding* adalah sebuah proses menjadikan seluruh huruf menjadi huruf kecil (*lowercase*) [6]. Ketiga yakni *stemming*. *Stemming* adalah proses memetakan kata yang ada kedalam kata dasarnya (*root*) [5]. Terakhir yakni pembobotan kata yang bertujuan untuk memberikan bobot pada setiap kata yang terdapat pada dokumen teks yang akan diproses [7]. Hasil dari *text pre-processing* sebagai berikut:

TABEL 2
 DATA TEXT PRE-PROCESSING

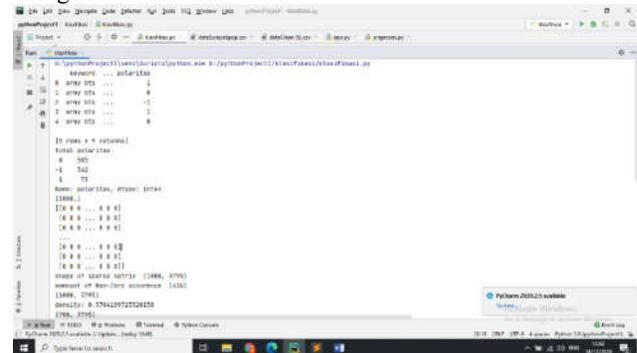
Dataset Awal	Pertama kali aku ngestan kepop ya sama bts, terus orang-orang pada bilang: “kok mau sih jadi Army? Kan Army toxic. Ya sesantainya aja gw jawab: “Gw ngestan BTS bukan fansnya.
Hasil Text Pre-processing	ngestan pop bts orang orang Army Army toxic santai gw gw ngestan bts fansnya

3) *Klasifikasi*: klasifikasi yang dilakukan dalam analisis menggunakan perbandingan 70:30 antara data *training* dengan data *testing*. Dengan melakukan *training* pada *Dataset* yang ada akan menghasilkan sebuah pola atau model pada data *testing*. Data yang telah selesai dilakukan pengklasifikasian akan menghasilkan polaritas dan analisis sentimen sebagai berikut:

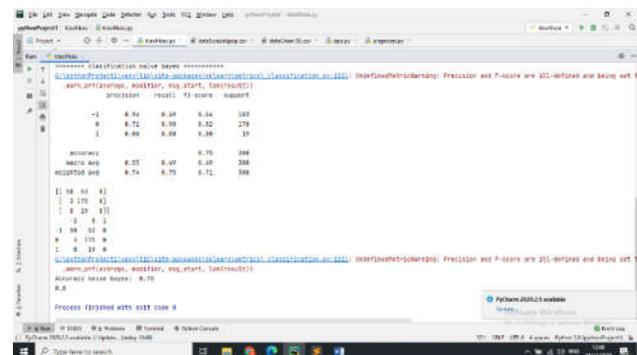
TABEL 3
 DATA TEXT PRE-PROCESSING

Sentimen	Polaritas
Netral	585
Negatif	342
Positif	73

Hasil klasifikasi juga bisa dilihat dari jalannya program sebagai berikut:



Gbr 5. Hasil Polaritas



Gbr 6. Hasil Klasifikasi dan Confusion Matrix

Hasil klasifikasi yang ditunjukkan pada tabel (3) yang merupakan polaritas dari setiap *class* yang ada yakni netral, negatif, dan positif. Ditunjukkan bahwa polaritas pada *class* netral terdapat 585 *Dataset*, negatif 342 *Dataset*, dan positif 73 *Dataset*. Untuk Gbr (4) menunjukkan polaritas yang ada di dalam program. Sedangkan Gbr (5) berisi hasil klasifikasi yang memiliki akurasi 75% dan *confusion matrix* yang menunjukkan data matriks dari setiap *class* setelah dilakukan *testing* menggunakan model.

IV. KESIMPULAN

Kesimpulan yang dapat diambil dari analisis sentimen menggunakan klasifikasi algoritma *Naïve Bayes* dengan data yang diambil dari *Twitter* dan *Army* BTS sebagai objek penelitian adalah sebagai berikut:

- Dari 1000 data yang diambil *tweet* yang bernilai negatif berdasarkan nilai polaritas sebanyak 342.
- Analisis sentimen dari anggapan pengguna *Twitter* tentang *Army* BTS memiliki tingkat akurasi sebesar 75%.
- terdapat polaritas sentimen negatif sebanyak 34,2%, netral 58,5%, dan positif 7,3%.

UCAPAN TERIMA KASIH

Ucapan terimakasih kepada Tuhan, kedua orang tua, dosen pembimbing, teman-teman, adik, seta idola saya yang telah memberi dukungan dan inspirasi.

REFERENSI

- [1] J. Clement, "statistics/number-of-active-Twitter-users-in-selected-countries," 29 Oktober 2020. [Online]. Available: <https://www.statista.com>.
- [2] A. Priadana och A. W. Murdiyanto, "Pemantauan Tren Hashtag pada Instagram Menggunakan Web Scraping," *Jurnal Pekommas*, 2020.
- [3] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di *Twitter*," *Integer Journal*, pp. 32-41, 2017.
- [4] F. M. Matulatuwa, "Text Mining Dengan Metode Lexicon Based Untuksentiment Analysis Pelayanan Pt. Pos Indonesia Melalui Media Sosial *Twitter*," *Jurnal Masyarakat Informatika Indonesia*, 2017.
- [5] E. Putra, "Klasifikasi Sentimen Masyarakat Terhadap Transgender Berdasarkan Komentar Di Instagram Menggunakan Metode Naive Bayes Classifier," *Universitas Islam Negeri Sultan Syarif Kasim Riau*, 2019.
- [6] N. Y. Faradillah, R. P. Kusumawardani och I. Hafidz, "EKSPERIMEN SISTEM KLASIFIKASI ANALISA SENTIMEN *TWITTER* PADA AKUN RESMI PEMERINTAH KOTA SURABAYA BERBASIS PEMBELAJARAN MESIN," *Seminar Nasional Sistem Informasi Indonesia*, 2016.
- [7] W. Nurjanah, R. S. Perdana och M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial *Twitter* menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah *Retweet*," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, pp. 1750-1757, 2017.