
Analisis Sentimen *Netizen Twitter* terhadap Pemberitaan PPN Sembako dan Jasa Pendidikan dengan Pendekatan *Social Network Analysis* dan *Naive Bayes Classifier*Joko Ade Nursiyono⁽¹⁾, Chusnul Chotimah⁽²⁾

Badan Pusat Statistik Provinsi Jawa Timur

Jl. Raya Kendangsari Industri No 43-44 Telp. (031) 8439343 Surabaya 60292

e-mail: joko.ade@bps.go.id dan chus_chot@bps.go.id

ABSTRAK

Pandemi covid-19 yang terjadi memberikan dampak di berbagai bidang kehidupan. Salah satu dampaknya penerimaan negara semakin tertekan hebat. Padahal di sisi lain negara dalam proses pemulihan ekonomi nasional (PEN) yang membutuhkan dana sangat besar. Sehingga pemerintah ingin menggenjot pendapatan negara dari pajak pertambahan nilai (PPN). Jika pemungutan PPN dapat dilakukan dengan seoptimal mungkin, maka akan meningkatkan penerimaan negara. Rencana tersebut mengakibatkan maraknya pemberitaan mengenai pengenaan PPN sembako dan jasa pendidikan di Indonesia. Pemberitaan tersebut secara otomatis memicu opini di masyarakat. Salah satu cara untuk melihat opini masyarakat adalah melalui media sosial *Twitter*. Penelitian ini bertujuan untuk mengkaji *network* dan sentimen *netizen Twitter* tentang PPN Sembako dan jasa pendidikan menggunakan metode *betweenness centrality* dan *Naive Bayes Classifier*. *Social Network Analisis* (SNA) menghasilkan 5 kluster dengan *record* ke-90 merupakan *bottleneck node*, yaitu aktor utama penyebaran informasi antar kluster. Model *Naive Bayes Classifier* memberikan hasil *Accuracy Classified* sebesar 74.865 persen, sementara persentase *Incorrectly Classified Instance* sebesar 25.135 persen. Hasil klasifikasi berdasarkan emosi terbentuk 5 ekspresi *fear*, *sadness*, *surprise*, *joy*, dan *anger* dengan emosi kata terbanyak adalah emosi *anger* (amarah), artinya mayoritas respon masyarakat terhadap kebijakan pengenaan PPN sembako dan jasa pendidikan diidentifikasi oleh *R Studio* sebagai wujud keamarahan.

Kata kunci : *PPN, Sembako, Sentimen, Emosi Kata, Social Network Analisis, Naive Bayes Classifier.*

ABSTRACT

The COVID-19 pandemic has had an impact on many areas of life. One of the effects of state revenues is getting severely depressed. Whereas on the other hand, the country is in the process of national economic recovery (PEN) which requires very large funds. So the government wants to boost state revenue from value-added tax (VAT). If VAT collection can be carried out optimally, it will increase state revenue. The plan resulted in widespread news regarding the imposition of VAT on necessities and education services in Indonesia. The news automatically triggers public opinion. One way to see public opinion is through social media Twitter. This study aims to examine more deeply the network and sentiments of Twitter netizens about VAT for necessities and education services using betweenness centrality and Naive Bayes Classifier methods. The results of Social Network Analysis (SNA) resulted in 5 clusters with the 90th record being a bottleneck node, which is the main actor in disseminating information between clusters. The Naive Bayes Classifier model gives for Accuracy Classified it is 74,865 percent, while the percentage for Incorrectly Classified Instance is 25,135 percent. The results of the classification based on emotions formed 5 expressions of fear, sadness, surprise, joy, and anger and the emotion with the most words was anger, meaning that the majority of public responses to the policy of imposing VAT on necessities and education services were identified by R Studio as an anger emotion.

Keywords : *VAT, Basic Food, Sentiment, Word Emotion, Social Network Analysis, Naive Bayes Classifier.*

1. PENDAHULUAN

Pandemi covid-19 yang terjadi memberikan dampak di berbagai bidang kehidupan. Menurut Neilmardin (2021), salah satu dampaknya membuat penerimaan kas negara semakin tertekan hebat. Padahal di sisi lain negara dalam proses pemulihan

ekonomi nasional (PEN) yang membutuhkan dana sangat besar. Sehingga pemerintah ingin menggenjot pendapatan negara dari pajak pertambahan nilai (PPN). Tarif PPN Indonesia yang dipatok sebesar 10 persen terlalu rendah jika dibandingkan negara lain di dunia yang menerapkan

Pajak Pertambahan Nilai atau *Value Added Tax (VAT)* dan *Goods and Services Tax (GST)*. Struktur penerimaan negara dari PPN selama ini memiliki kontribusi sebesar 42 persen terhadap penerimaan negara. Jika pemungutan PPN dapat dilakukan dengan seoptimal mungkin, maka akan meningkatkan penerimaan negara sebagai bekal berbagai program pembangunan. Rencana tersebut mengakibatkan maraknya pemberitaan mengenai pengenaan PPN sembako dan jasa pendidikan di Indonesia. Pemberitaan tersebut secara otomatis memicu opini di masyarakat.

Opini yang terbentuk tentu menimbulkan sentimen sekaligus pro dan kontra di media sosial. Tingkat dan kecepatan penyebaran informasi ditentukan oleh seberapa besar pengaruh sebuah akun seseorang dalam jejaring media sosial, terutama *Twitter*. Melalui pendekatan *Social Network Analysis (SNA)*, penelitian ini bertujuan untuk menentukan aktor utama sekaligus pengaruhnya dalam model penyebaran informasi (Setatama, Mahdi, S. dan Tricahyono, Dodie, 2017) atau opini antar klaster. Selain itu, untuk mengetahui sentimen yang terbentuk dari wacana pengenaan PPN sembako maupun jasa pendidikan, penelitian ini menggunakan pengklasifikasian emosi kata model *Naive Bayes Classifier*. Menurut Nurzahputra, Aldi dan Muslim, Much A.(2016), dalam *text mining*, pendekatan *Naive Bayes Classifier* bisa digunakan untuk mengklasifikasikan opini ke dalam opini positif dan negatif.

2. METODE PENELITIAN

Set Data

Set data yang digunakan dalam penelitian ini bersumber dari *Twitter* yang diperoleh dengan teknik *Crawling* menggunakan *R Studio* versi 4.0.3. Dari sebanyak 5.000 target *records* yang akan digunakan sebagai sampel, didapatkan sebanyak 4.090 *records* yang terdiri atas 90 variabel. Adapun periode *records* yang tercatat sebagai sampel penelitian ini mulai 10 Juni 2021 - 21 Juni 2021. Setelah itu dilakukan pembersihan data teks (*text cleaning*) dengan tahapan seperti pada gambar 1.

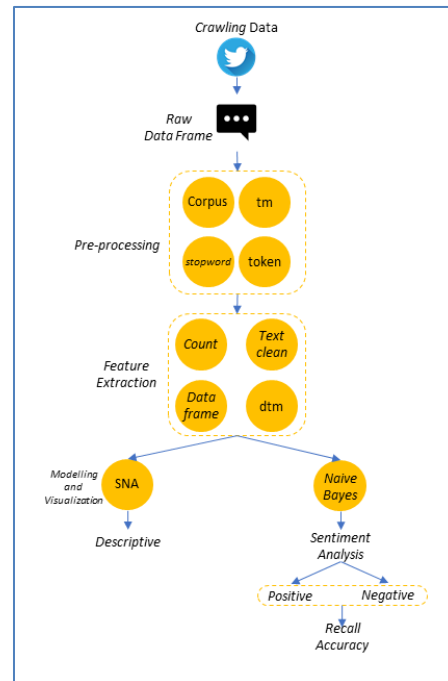
Import data

Import data dilakukan dengan menggunakan library *readxl* kemudian menyeleksi kalimat *tweet* pada data untuk kemudian dijadikan sebagai data *corpus* melalui library *corpus*.

Text Cleaning

Pada data *corpus*, dilakukan pembersihan kalimat dari tautan atau URL, *mention*, *hashtags* atau #,

menghapus kelebihan spasi, dan nomor menggunakan library *tm*.



Gambar 1. Bagan pembersihan data teks (*text cleaning*)

Stopwords Bahasa Indonesia

Pada struktur data yang sama, untuk menghapus kata hubung (konjungsi) atau kata-kata lain yang tidak bermakna, termasuk kata singkatan dilakukan *tokenize* kemudian mengeliminasi dengan fungsi *custome*, yaitu *stopwords-Indonesia*.

Export Clean Text

Setelah ketiga tahapan di atas dilakukan, pada tahap akhir adalah mengubah kembali *tweets* dalam bentuk *data frame* dan menyimpannya dalam *file excel* dengan nama: *databersih*.

Analisis Jaring Sosial (Social Network Analysis)

Social Network Analysis (SNA) merupakan salah satu analisis untuk melihat tingkat keeratan dan pengklasifikasian topik yang dibahas dalam jejaring sosial. Menurut Alamsyah, Andri dan Bratawisnu, Made K. (2018), analisis yang dapat dilakukan terkait jejaring sosial yang terbentuk adalah properti jaringan sosial. Properti sosial yang terbentuk berupa *nodes*, *edges*, dan *degree* untuk selanjutnya dianalisis *Centrality*-nya.

Tabel 1. Pengertian properti jaringan

Properti Jaringan	Definisi
<i>Nodes</i>	<i>Nodes</i> adalah aktor individu dalam jaringan yang dihubungkan oleh <i>ties</i> (Oktora, Rio dan Alamsyah, Andry, 2014)
<i>Edges</i>	<i>Edges</i> merepresentasikan interaksi dua aktor (<i>nodes</i>) atau lebih (Anam, Khairul M, dkk., 2021)
<i>Degree</i>	<i>Degree</i> merupakan ukuran keeratan grup di dalam jaring (<i>Network</i>)
<i>Centrality</i>	<i>Centrality</i> adalah proses kalkulasi terhadap relasi setiap individu dari sebuah jejaring sosial (Susanto, Budi. dkk., 2012)

Centrality

Centrality pada dasarnya menentukan *node* utama penyebaran sebuah informasi. Proses kalkulasi *centrality* dapat dihitung dengan pendekatan jumlah interaksi atau *edge* antar *node*, jarak terpendek yang ditempuh oleh sebuah *node* ke seluruh *node*, berdasarkan skor *eigen vector* jaring antar *node*, dan ukuran peran sebuah *node* terhadap seluruh *node*. Untuk menentukan *centrality*, dalam penelitian ini menggunakan pendekatan *betweenness centrality* yang merupakan modifikasi dari *degree centrality*. Menurut Tabassum et al. (2018), *degree centrality* adalah jumlah interaksi atau *edge* yang terjadi pada suatu *node* tertentu dengan menggunakan formula berikut:

$$C_D(i) = \sum_{j=1}^n a_{ij} \tag{1}$$

di mana:

C_D : *Degree centrality node ke-i*

a_{ij} : nilai baris ke-*i* dan kolom ke-*j* matriks *adjacency*

Degree centrality ditentukan oleh matriks *adjacency*. Menurut Santoso, Budi. dkk. (2012), nilai yang ada di setiap *cell* matriks *adjacency* menunjukkan informasi atas hubungan atau relasi antar aktor atau individu. Menurut Mamas, Christoporos (2018), struktur matriks *adjacency* untuk menentukan *degree centrality Social Network Analysis* (SNA) adalah sebagai berikut:

	Node 1	Node 2	Node 3	Node 4	Node 5
Node 1	—	1	1	0	0
Node 2	1	—	0	0	0
Node 3	0	0	—	1	1
Node 4	0	1	0	—	1
Node 5	0	1	1	1	—

Gambar 2. Struktur matriks *adjacency*

Betweenness Centrality

Betweenness centrality pada dasarnya merupakan ukuran untuk melihat peranan sebuah *node* dalam jaring atau *network*. Menurut Pratama, Faustinus Y. (2018), *betweenness centrality* dapat digunakan untuk melihat *node* yang berperan sebagai *bottleneck*. Sebuah *node* dikatakan mempunyai peran penting jika menjadi *communication bottleneck*. Ukuran ini juga dapat digunakan untuk mengidentifikasi *boundary spanners*, yaitu aktor atau *node* yang berperan sebagai penghubung (*bridge*) antara komunitas. Ukuran ini dihitung dengan menjumlahkan semua garis atau *path* yang mengandung *bottleneck node*. Rumus di bawah ini digunakan dalam menghitung *betweenness centrality*:

$$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{j \in G, G \neq 1} \sum_{k \in G, k \neq i, k \neq j} \frac{n_{jk}(i)}{n_{jk}} \tag{2}$$

di mana:

C_i^B : bobot nilai *betweenness centrality*

N : jumlah *node* dalam suatu jaringan

n_{jk} : jumlah jalur (*path*) terpendek dari *node ke-j* ke *node ke-k*

$n_{jk}(i)$: jumlah jalur (*path*) terpendek dari *node ke-j* ke *node ke-k* yang melewati simpul ke-*i*

Analisis Sentimen

Analisis sentimen merupakan *tool* analisis untuk data yang bertipe teks dan kontekstual. Menurut Rozi, Imam F, dkk. (2012), analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini.

Naive Bayes Classifier

Naive Bayes Classifier (NBC) merupakan algoritma pengklasifikasian yang sering digunakan dalam permasalahan teks (Rozi, Imam F, dkk., 2012). Adapun atribut yang digunakan dalam NBC adalah data *training*, *prior probability*, serta *posterior probability*. Untuk menjelaskan teori *Naive Bayes*, proses pengklasifikasian memerlukan sejumlah

petunjuk untuk menentukan kelas yang cocok bagi sampel yang dianalisis. Secara umum, peluang teorema *Bayes* yang disesuaikan dirumuskan berikut:

$$P(C|F_1, F_2, F_3, \dots, F_n) = \frac{P(C) \cdot P(F_1, F_2, F_3, \dots, F_n|C)}{P(F_1, F_2, F_3, \dots, F_n)} \quad (3)$$

di mana C merepresentasikan kelas dan $F_1, F_2, F_3, \dots, F_n$ merupakan karakteristik yang diperlukan sebagai dasar pengklasifikasian. $P(C|F_1, F_2, F_3, \dots, F_n)$ pada rumus tersebut merupakan *posterior probability*, yaitu peluang masuknya sampel karakteristik tertentu dalam kelas C . Sedangkan $P(C)$ merupakan *prior probability*, yaitu besarnya peluang munculnya kelas C sebelum sampel tersebut masuk. $P(F_1, F_2, F_3, \dots, F_n|C)$ merupakan peluang munculnya karakter-karakter sampel kelas C . Dan $P(F_1, F_2, F_3, \dots, F_n)$ disebut sebagai peluang munculnya karakter sampel secara global atau disebut *evidence*. Setelah sentimen untuk masing-masing *record* terbentuk, kebaikan model *Naive Bayes Classifier* dalam penelitian ini diukur menggunakan *Recall Accuracy*. Menurut Rozi, Imam F, dkk. (2012), *Recall* merupakan rasio jumlah dokumen yang relevan yang ditemukan terhadap total jumlah dokumen relevan yang dikumpulkan. Nilai *recall* biasanya ditunjukkan dalam satuan persen (%). Ukuran ini merupakan *proxy* performa model *Naive Bayes Classifier* dalam mengevaluasi kecocokan pengklasifikasian sentimen seluruh sampel. Berikut ini adalah rumus penghitungan *recall accuracy* (Handayani, Fitri dan Pribadi, Feddy S., 2015) :

$$Recall Accuracy = \frac{TP}{TP + FP} \quad (4)$$

di mana:

TP : klasifikasi sentimen yang benar sesuai model *Naive Bayes*

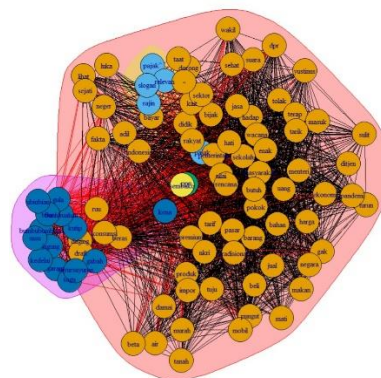
FP : klasifikasi sentimen yang salah sesuai model *Naive Bayes*

Berdasarkan nilai *recall accuracy*, semakin besar nilai persentasenya, maka menunjukkan hasil pengklasifikasian model *Naive Bayes Classifier* semakin baik.

3. HASIL DAN PEMBAHASAN

Hasil pengolahan dengan menggunakan *Social Network Analysis* (SNA) terbentuk 5 klaster. Klaster 1 memiliki 73 anggota, klaster 2 memiliki 5 anggota, klaster 3 memiliki 1 anggota, dan klaster 4 memiliki 1 anggota, serta klaster 5 memiliki 14 anggota.

Hasil dari *centrality summary* didapatkan nilai minimum sebesar -0.028, kemudian kuantil pertama sebesar 0.010, dan nilai maksimumnya sebesar 0.073. Berdasarkan *summary* tersebut, nilai maksimum *centrality* terdapat pada *record* ke-90 dan sekaligus merupakan *bottleneck node*, yaitu aktor utama penyebaran informasi antar klaster dengan kalimat *tweet*, “Pemerintah berencana mengenakan Pajak Pertambahan Nilai (PPN) terhadap sejumlah kebutuhan masyarakat banyak, termasuk di antaranya sembako dan jasa pendidikan (sekolah). Apa muara dari permasalahan ini? Simak analisis selengkapnya.” Dari 5 klaster yang terbentuk, *record* ke-90 ini masuk dalam klaster sendiri, yaitu klaster 4 (terlihat *node* warna kuning pada gambar 3).



Gambar 3. Hasil olah *social network analisis* (SNA)

Analisis Sentimen

Twitter merupakan salah satu media sosial yang biasa digunakan untuk melihat opini masyarakat. Berdasarkan hasil olah data dari *twitter*, terdapat beberapa kata yang paling banyak muncul dalam *tweet* terkait isu PPN sembako dan jasa pendidikan. Kata yang sering digunakan yaitu “pajak”, “ppn”, “pemerintah”, “kena”, “rakyat”, dan tentunya masih banyak kata-kata juga digunakan dalam *tweet*. Artinya, secara umum masyarakat menggunakan satu kata dari kelima kata tersebut dalam unggahan *tweet*-nya terkait kebijakan pengenaan PPN sembako dan jasa pendidikan (lihat gambar 4).

Classifier dalam penelitian ini diukur menggunakan *Recall Accuracy*. Hasil penghitungan *Recall*

Accuracy menunjukkan bahwa untuk *Accuracy Classified* sebesar 74.865 persen sementara persentase untuk *Incorrectly Classified Instance* sebesar 25.135 persen. Pada gambar 7, terlihat bahwa dari 4090 data *twitter* terkait topik ppn sembako dan jasa pendidikan, ada sebanyak 3062 data yang berhasil diklasifikasikan sentimennya dengan benar dan sebanyak 1028 data *twitter* lainnya tidak berhasil diklasifikasikan dengan benar.

Pada tabel 2 dapat dilihat bahwa dari 413 *tweet* yang diklasifikasikan bersentimen negatif ternyata semuanya masuk sentimen netral, dari 615 *tweet* bersentimen positif, ternyata semuanya masuk klasifikasi sentimen netral, hanya sebanyak 3062 *tweet* saja yang diklasifikasikan menurut sentimen yang benar (yaitu netral).

Tabel 2. Klasifikasi prediksi sentimen model *naive bayes*

Prediksi	Negatif	Netral	Positif
Negatif	0	0	0
Netral	413	3062	615
Positif	0	0	0

Sumber: hasil olah *R Studio* versi 4.0.3

4. KESIMPULAN DAN SARAN

1. Pengolahan data tweet ppn sembako menggunakan Social Network Analisis (SNA) menghasilkan 5 kluster dengan record ke-90 merupakan bottleneck node yaitu aktor utama penyebaran informasi antar kluster.
2. Hasil analisis sentimen secara umum masyarakat menggunakan satu kata dari kelima kata yang sering digunakan dalam tweet terkait pemberitaan ppn sembako dan jasa pendidikan yaitu pemerintah, pajak, ppn, didik, rakyat dan tentunya masih banyak kata-kata juga digunakan dalam tweet.
3. Hasil penghitungan Recall Accuracy menunjukkan bahwa untuk Accuracy Classified sebesar 74.865 persen sementara persentase untuk Incorrectly Classified Instance sebesar 25.135 persen.
4. Hasil klasifikasi berdasarkan emosi terbentuk 5 ekspresi fear, sadness, surprise, joy, dan anger dan emosi kata yang paling banyak adalah emosi anger (amarah), artinya, mayoritas respon masyarakat terhadap kebijakan pengenaan PPN sembako dan jasa pendidikan diidentifikasi oleh R Studio sebagai wujud keamarahan.

5. Kendati masih wacana, ternyata di media sosial Twitter sudah terbentuk sentimen. Dengan demikian, sebaiknya pengumuman kebijakan PPN sembako dan jasa pendidikan diumumkan secara detil agar tidak menimbulkan salah persepsi di masyarakat.

DAFTAR PUSTAKA

Anam, Khairul M., dkk. 2021. Analisis Kesiapan Masyarakat Pada Penerapan Smart City di Sosial Media Menggunakan SNA, Jurnal Rekayasa Sistem dan Teknologi Informasi, 5(1), 69-81.

Alamsyah, Andry. dan Bratawisnu, Made. K. 2018. Social Network Analysis Untuk Analisa Interaksi User Media Sosial Mengenai Bisnis E-Commerce (Studi Kasus: Lazada, Tokopedia, dan Elevenia). Jurnal Manajemen dan Bisnis, 2(2), 107-115.

DJP Buka 3 Alasan Pemerintah Pungut PPN Sembako dan Sekolah. 2021. <https://www.cnnindonesia.com/ekonomi/20210614101335-532-653968/djp-buka-3-alasan-pemerintah-pungut-ppn-sembako-dan-sekolah>, diakses tanggal 21 Juni 2021.

Handayani, Fitri dan Pribadi, Feddy, S. 2015. Implementasi Algoritma Naïve Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110, Jurnal Teknik Elektro, 7(1), 19-24.

Mamas, Christoforos, 2018. The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation, Chapter Title: "Matrices (in Social Network Analysis)", SAGE Publication, Inc, Thousand Oaks.

Nurzahputra, Aldi dan Muslim, Much A. 2016. Analisis Sentimen pada Opini Mahasiswa Menggunakan Natural Language Processing, Seminar Nasional Ilmu Komputer (SNIK 2016)-Semarang, 10 Oktober 2016, 114-118.

Oktora, Rio dan Alamsyah, Andy. 2014. Pola Interaksi dan Aktor yang Paling Berperan pada Event JGTC 2013 melalui Media Social Twitter (Studi menggunakan Metode Sosial Networking Analysis), Jurnal Manajemen Indonesia, 14(3), 201-209.

Pratama, Faustinus Y. 2018. Simulasi Jejaring Jalan Kota Pontianak Dengan Betweenness Centrality dan Degree Centrality, Jurnal TIN Universitas Tanjungpura, 2(1), 1-6.

Rozi, Imam F, dkk. 2012. Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi, Jurnal EECCIS, 6(1), 37-43.

Setatama, Mahdi, S. dan Tricahyono, Dodie, 2017. Implementasi Social Network Analysis dalam Penyebaran Country Branding “Wonderful Indonesia”, 2(1), 91-104.

Susanto, Budi, dkk. 2012. Penerapan Network Analysis dan Penentuan Centrality Studi Kasus Social Network Twitter, Jurnal Informatika, 8(1), 1-13.

Tabassum, Shazia, dkk. 2018. Social Network Analysis: an Overview, Original Article Jurnal Sectoin.