

ANALISIS SENTIMEN PADA *TWITTER* TERHADAP KINERJA KOMISI PEMBERANTASAN KORUPSI (KPK) DI INDONESIA DENGAN METODE *NAIVE BAYES*

Rahmat Taufiq H.R.¹, Ari Eko Wardoyo², Mudafiq Riyan Pratama³
Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember
idondidong@gmail.com, Arieiko@unmuhjember.ac.id, mudafiq.riyan@unmuhjember.ac.id

ABSTRAK

The Corruption Eradication Commission (KPK) in its performance in handling corruption gets little attention from the public. One of them on social media twitter, in timelines of twitter there is a positive sentiment and negative sentiment towards the KPK. Such societal sentiments can not be known for certain by the KPK to measure public confidence in its performance. Twitter sentiment analysis on the performance of the Corruption Eradication Commission (KPK) in Indonesia using naive bayes method is proposed to accommodate the problem. Research begins by collecting tweet data, validating sentiment directives, analyzing and compiling positive and negative dictionaries and classifying them using naive bayes. The results show that as many as 50.96% of twitter accounts have positive sentiments and 49.03% of twitter accounts have negative sentiment towards KPK. While the results of analysis on the test data showed 65.51% accuracy, precision 51.35% and 90.47% recall.

Keywords: Sentiment, Twitter, KPK, Naive bayes.

Komisi Pemberantasan Korupsi (KPK) dalam kinerjanya menangani tidak pidana korupsi banyak mendapat sorotan dari masyarakat. Salah satunya pada media sosial *twitter*, di linimasa ada yang memberikan sentimen positif dan ada yang memberikan sentimen negatif terhadap KPK. Sentimen masyarakat seperti ini yang belum bisa diketahui pasti oleh KPK untuk mengukur kepercayaan publik terhadap kinerjanya. Analisis sentimen pada *twitter* terhadap kinerja komisi pemberantasan korupsi (KPK) di Indonesia dengan metode *naive bayes* diajukan untuk mengakomodasi permasalahan tersebut. Penelitian dimulai dari *crawling* data *tweet*, validasi arah sentimen, analisis dan menyusun kamus positif dan negatif serta menganalisisnya menggunakan *naive bayes*. Hasil menunjukkan bahwa sebanyak 50,96% akun *twitter* memiliki sentimen positif dan 49,03 akun *twitter* memiliki sentimen negatif. Sedangkan hasil analisis pada data uji menunjukkan tingkat akurasi 65,51%, tingkat presisi 51,35% dan *recall* 90,47%.

Kata kunci : Sentimen, *Twitter*, KPK, *Naive bayes*.

1. PENDAHULUAN

Opini masyarakat terhadap KPK pada media sosial yang masif ini belum terakomodasi atau digambarkan secara sederhana sehingga KPK tidak mengetahui persentase sentimen masyarakat terhadap kinerjanya, oleh karena itu belum bisa menjadi bahan pertimbangan dalam perbaikan kinerja yang diperoleh melalui opini atau kritik dan saran masyarakat pada media sosial khususnya *twitter*.

Dengan ketenaran *Twitter* tersebut maka tidak heran banyak orang yang memberikan pendapatnya tentang berbagai hal termasuk pendapat kinerja KPK dalam memberantas korupsi di Indonesia dan juga *Twitter* berfokus pada pendekatan analisis sentimen yang dirancang khusus untuk *tweet*. Analisis sentimen adalah proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini baik yang bersifat positif dan negatif.

2. TINJAUAN PUSTAKA

2.1. Tindak Pidana Korupsi

Pengertian tindak pidana korupsi menurut Pasal 2 ayat (1) dan Pasal 3 Undang-undang No. 31 Tahun 1999 tentang Pemberantasan Tindak Pidana Korupsi adalah sebagai berikut :

1. Pasal 2 ayat (1) : “Setiap orang yang secara melawan hukum melakukan perbuatan memperkaya diri sendiri atau orang lain atau suatu korporasi yang dapat merugikan keuangan negara atau perekonomian negara, dipidana dengan penjara seumur hidup atau pidana penjara paling singkat 4 (empat) tahun dan paling lama 20 (dua puluh) tahun dan denda paling sedikit Rp. 200.000.000,00 (dua ratus juta rupiah) dan denda paling banyak Rp. 1.000.000.000,00 (satu milyar rupiah).”

2. Pasal 3 : “Setiap orang yang dengan tujuan menguntungkan diri sendiri atau orang lain atau suatu korporasi, menyalahgunakan kewenangan, kesempatan, atau sarana yang ada padanya karena jabatan atau kedudukan yang dapat merugikan keuangan negara atau perekonomian negara, dipidana dengan pidana penjara seumur hidup atau

pidana penjara paling singkat 1 (satu) tahun dan paling lama 20 (dua puluh) tahun dan atau denda paling sedikit Rp. 50.000.000,00 (lima puluh juta rupiah) dan paling banyak Rp. 1.000.000.000,00 (satu milyar rupiah).”

2.2. Komisi Pemberantasan Korupsi (KPK)

Berdasarkan Pasal 3 Undang-undang no. 30 Tahun 2002 tentang Komisi Pemberantasan Tindak Pidana Korupsi, yang dimaksud dengan Komisi Pemberantasan Korupsi (KPK) adalah lembaga negara yang dalam melaksanakan tugas dan wewenangnya bersifat independen dan bebas dari pengaruh kekuasaan manapun.

2.3. Sentimen Analisis

Opinion mining atau analisis sentimen merupakan salah satu bidang dari ilmu komputer yang mempelajari komputasi linguistik, pengolahan bahasa alami, dan *text mining* yang bertujuan untuk menganalisa emosi, penilaian, sikap, pendapat, sentimen, evaluasi seseorang terhadap seorang pembicara atau penulis berkenaan dengan suatu produk, layanan, organisasi, individu, tokoh publik, topik, acara, ataupun kegiatan tertentu (Liu, 2012).

2.4. Twitter

Twitter adalah sebuah situs web yang dimiliki dan dioperasikan oleh *Twitter Inc.*, yang menawarkan jaringan sosial berupa *microblog* sehingga memungkinkan penggunaannya untuk mengirim dan membaca pesan *Tweets* (*Twitter*, 2013). *Mikroblog* adalah satu jenis alat komunikasi *online* dimana pengguna dapat memperbarui status tentang mereka yang sedang memikirkan dan melakukan sesuatu, apa pendapat mereka tentang suatu objek atau fenomena tertentu.

2.5. Naive bayes

Algoritma *Naive bayes* adalah salah satu algoritma yang terdapat pada teknik data mining klasifikasi. *Naive bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris yaitu Thomas bayes, *Naive bayes* memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya, sehingga dikenal dengan Teorema Bayes. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya.

2.6. Naive bayes classifier

Bayes Classifier (NBC) merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing variabel.

$$V_{map} = \frac{\text{argmax}_{V_j \in V} P(V_j) \prod P(W_k | V_j)}{P(V_j) \frac{|docs_j|}{|contoh|}}$$

3. METODOLOGI PENELITIAN

3.1. Tahapan Penelitian

- a. Pengumpulan Data
tahap selanjutnya setelah identifikasi masalah adalah pengumpulan data. Data yang dikumpulkan didapat dari hasil *crawling* opini masyarakat terdapat KPK dengan kata kunci KPK dan mention @KPK_RI pada API *Twitter*.
- b. Pembuatan Kamus data
Tahap ini menyusun kamus data positif, negatif dan negasi dari analisis data *tweet* yang sudah divalidasi oleh kuesioner.
- c. Preprocessing
 - 1) Case Folding ialah proses merubah huruf kapital (uppercase) menjadi huruf kecil (lowercase).
 - 2) Tokenizing, Pada proses tokenizing setiap kata pada *tweet* dipisahkan, pada proses ini tahap yang dilakukan adalah memisahkan setiap kata yang dipisahkan oleh spasi.
 - 3) Cleansing, *Tweet* yang ditujukan terhadap akun @KPK_RI memiliki berbagai komponen atau karakteristik *tweet* yang khas seperti “@”, link, “#” dan RT. Komponen-komponen tersebut tidak memiliki pengaruh apapun terhadap sentimen, maka akan dibuang.
 - 4) *Stopword removal*, Pada tahap ini *tweet* masih mengandung kata yang dianggap tidak dapat memerikan pengaruh dalam menentukan suatu kategori sentimen.
 - 5) *Convert negation*, merupakan proses konversi kata-kata negasi yang terdapat pada suatu *tweet*, karena kata negasi mempunyai pengaruh dalam merubah nilai sentimen pada suatu *tweet*.
 - 6) *Stemming*, adalah proses mengubah kata berimbuhan ke bentuk asalnya (kata dasar). Algoritma yang digunakan untuk proses *stemming* berbahasa Indonesia adalah Nazief dan Andriani.
- d. Implementasi *naive bayes*
Tahap selanjutnya adalah tahap implementasi metode, dari hasil preprocessing akan diklasifikasikan kedalam dua class; positif dan negatif menggunakan metode *Naive bayes classifier*.
- e. *Confusion matrix*
Hasil dari klasifikasi menggunakan *Naive bayes classifier* selanjutnya akan diukur tingkat akurasi, presisi dan *recall*, hasil klasifikasi ini dibandingkan dengan hasil klasifikasi manual dari kuesioner.

4. IMPLEMENTASI DAN PENGUJIAN

4.1. Pengumpulan Data

Pengumpulan (*crawling*) data *tweet* diambil melalui aplikasi Rstudio pada API *Twitter* dengan rentang waktu dari tanggal 31 Januari 2018 sampai 31 Mei 2018 dengan jumlah keseluruhan 806 *tweet*, serta pembagian data yaitu 748 data latih dan 58 data uji.

4.2. Pembuatan Kamus data

Dari 748 *tweet* data latih selanjutnya dianalisis setiap kata didalamnya lalu dimasukkan ke tiap-tiap kamus data diantaranya kamus positif, negatif dan *stopword*.

Tabel 4.5 Potongan hasil analisis dan pengumpulan kamus sentimen positif

Tangkap	Sikat	Sanggup
Bongkar	Sikatt	Bawa
Tuntas	percaya	Gizi
Periksa	tegak	kritik
Bener	bukti	beriman
Suka	ditindaklanjuti	lirik
mantap	selamat	harap
lanjutken	ringkussssss	dioooooooooorrrr
mantaaaap	ringkus	dor
berantas	hajarr	semoga
lanjutkannn	Hajar	smoga
lanjutkan	Usut	selidiki
kuat	lancar	endus
mohon	Melek	habisi
tolong	Bangun	pantau

Tabel 4.6 Hasil analisis dan pengumpulan kamus sentimen negatif

jatah	piye	pemborosan
pencemaran	mirisss	tembakmati
blunder	popularitas	beranienggak
propaganda	tuduhan	beraninggak
cengeng	lolos	jawabdong
kebal	tebangpilih	tidakpunya akal
mangkrak	nggakmampu	gakditerusin
bosen	manaberani	tidakadatinda kan
bocah	gakberes	kokdiam
sicebong	gberani	tidakmemerik sa
kafir	tidaklayak	tdkmemeriks a
bobroknya	sotoy	gakberani
No	apakabar	mencarikesal ahan

pesimis	apakbr	kokbisa
bingung	menjerit	hukumanmati
terkutuklah	macamapa	berteletele
direzim	tidakberani	dapetjatah
takut	kokminta	pilihtebang
kpan	bubarin	tidakbernyali
lupa	diganti	ndakbernyali
plongaplongo	taktau	gakpercaya
kurangpuas	nggakditangka p	soksuci
gakmanggil	dibiarkankabur	gadaygbsperca aya
tidakditindakla njuti	kagakbernyali	gakadaygbisa percaya
tdkditindaklanj uti	segerasadar	masihpantask ah
tidakmampu	mainopini	gktaumalu
emangberani	pilihkasus	gaktaumalu
manaberani	manamampu	diulur
kurangvitamin	takberguna	gkpercaya
gaditangkap	dudukmanis	gakpercaya
gakditangkap	dirusak	

Tabel 4.7 Potongan hasil analisis dan pengumpulan kamus *stopword*

Toh	Br	intern	km
Turut	bwh	ekster n	dil
Tutur	gini	sumbe r	dg
Tuturnya	mas	ed	bu
Ucap	mba	bd	dr
Ucapnya	nya	utk	aj
Ujar	Tsk	bngettt	emang
Ujarnya	Krn	banget	bgt
umum	Yg	aja	amp
umumnya	Ke	klo	cc
ungkap	wih	ko	org
ungkapny a	trus	iki	bi
untuk	Teb	tuhh	saj
usah	sdh	seh	de
usai	dong	dah	t
waduh	Bs	gua	k
wah	sdh	ganj	jreeengg
wahai	Sk	ahok	bingitss
waktu	Mu	ganjar	gitu
waktunya	Di	hahaha	gin

walau	kan	sy	hahahahahaha
walaupun	Sih	min	jga
wong	Sll	kl	tan
yaitu	sllu	yth	lha
yakin	Tuh	milu	msh
yakni	Jrg	deh	nih
yang	Jd	lo	hai
adl	yaa	pade	s
adlh	smpai	dgn	ahhhh
adpn	pucuk	tp	iya
agr	blbi	si	n

Tabel 4.8 Potongan hasil analisis dan pengumpulan kamus negasi positif

gak benar
gak bener
jangan lembek
jangan lelah
jgn lelah
jangan dibubarkan
jgn dibubarkan
jgn lemah

Tabel 4.9 Potongan hasil analisis dan pengumpulan kamus negasi negatif

nggak mampu	gak manggil	gak benar
mana berani	tidak ditindak lanjuti	gak bener
gak beres	tdk ditindaklanjuti	kaga tuntas
g berani	tidak mampu	gk percaya
tidak layak	ga ditangkap	gak percaya
tidak berani	tdk benar	gada yg bs percaya
tidak punya akal	gada yg percaya	gak ada yg bisa percaya
gak diterusin	kaga tuntas	gk tau malu

4.3. Preprocessing

a. Case folding

Tabel 4.10 Potongan hasil case folding pada tweet

<i>Tweet</i>	@nilamerah1 @theo_putraaa @mawisevendry @kumparan @Dahnilanzar @KPK_RI penjara karena korupsi? kurang info apa
--------------	--

	kurang vitamin mas/mbanya?
<i>Case folding</i>	@nilamerah1 @theo_putraaa @mawisevendry @kumparan @dahnilanzar @kpk_ri penjara karena korupsi? kurang info apa kurang vitamin mas/mbanya?

b. Cleansing

Tabel 4.11 Potongan hasil Cleansing pada tweet

<i>Tweet</i>	kapan yaa @kpk_ri bongkar tuntas kasus besar periksa smpai ke pucuk??... blbi, pelindo,temanbok, sumberwaras,dll..... https://t.co/priec8vii
<i>Cleansing</i>	kapan yaa bongkar tuntas kasus besar periksa smpai ke pucuk blbi pelindo temanbok sumberwaras dll

c. Convert negation

Tabel 4.12 Potongan hasil analisa dan pengumpulan kamus negasi positif

<i>Tweet</i>	@kpk_ri tangkapin aja koruptor2 tidak bermoral itu..rakyat dibelakangmu
Negasi positif	tangkapin aja koruptor tidakbermoral itu..rakyat dibelakangmu

Tabel 4.12 Potongan hasil analisa dan pengumpulan kamus negasi negatif

<i>Tweet</i>	Kl sampe jutaan begini yg rusak(katanya), pasti ada kerugian Negara!! Gak beres berarti! @bpkri
--------------	--

	@KPK_RI sudah harus... https://t.co/GUbqL9MvuN
Negasi negatif	kl sampe jutaan begini yg rusak katanya pasti ada kerugian negara gakberes berarti sudah harus

d. *Stemming*

Tabel 4.13 Potongan hasil *Stemming*

<i>Tweet</i>	"Usut tuntas kasus ini wahai @KPK_RI @bpkri @BPKPgoid Jika kalian bernyali. Apakah terdapat indikasi merugikan uan... https://t.co/gWYamCeznw "
<i>Stemming</i>	usuttuntas kasus ini wahai jika kalian nyali . apakah dapat indikasi rugi uan

e. *Stopword removal*

Tabel 4.14 Potongan hasil penghapusan *stopword*

<i>Tweet</i>	@tohang_bang @mohmahfudmd Berarti ntar tgl muda gajian dong.. ha..ha..ha.. tolong di cek @KPK_RI dan zakat/sedekah... https://t.co/cjMHHq13MQ
--------------	---

<i>Stopword removal</i>	berarti tgl muda gajian tolong cek zakat sedekah
-------------------------	---

f. *Tokenizing*

Tabel 4.15 Potongan hasil tokenizing pada *tweet*

<i>Tweet</i>	berarti tgl muda gajian tolong cek zakat sedekah							
<i>Tokenizing</i>	berarti	tgl	muda	gajian	tolong	cek	zakat	sedekah

4.4. Implementasi *naive bayes*

Tabel 4.16 Potongan pembobotan tiap kata menggunakan term frekuensi

<i>Tweet</i>	"@KPK_RI Alhamdulillah, Lembaga beriman..Ttd. @Allahpencipta"		
<i>Tokenizing</i>	Alhamdulillah	Lembaga	Beriman
TF	1	1	1

Tabel 4.17 Hasil hitung prior probability

	jumlah positif	Jumlah negatif	Jumlah	probabilitas positif	probabilitas negatif
Kata keseluruhan	3146	1536	2000	0.488239034	0.635727908

4.5. *Confusion matrix*

Tabel 4.19 Hasil hitung persentase sentimen positif dan negatif

Sentimen	Persentase	Jumlah data
Positif	50.969529085873 %	184
Negatif	49.030470914127 %	177

5. PENUTUP

5.1. Kesimpulan

- Algoritma *Naive bayes* dapat digunakan dalam mengklasifikasikan data *tweet* untuk analisis sentimen pada *twitter* terhadap kinerja komisi pemberantasan korupsi (KPK) di Indonesia. Pada data latih dengan jumlah 748 *tweet*, hasil menunjukkan 50.96% akun *twitter* bersentimen positif dan 49.03% akun *twitter* bersentimen negatif terhadap KPK.
- Pengujian pada data uji dengan jumlah 58 *tweet* menggunakan *confusion matrix* menunjukkan tingkat akurasi 65.51%, presisi 51.35% dan *recall* 90.47%.
- Pada tahap pembentukan kamus negatif, positif dan negasi perlu analisis per kata, ini diakibatkan oleh ragam bahasa yang digunakan cukup banyak serta sentimen yang dimaksud terbentuk secara kontekstual.
- Pembentukan kamus negasi mendapatkan tambahan dalam susunan ungkapan negasi. Yang sebelumnya untuk mengungkapkan ingkaran hanya membutuhkan kata negasi dan dilanjutkan kata bermakna positif atau negatif dalam kasus ini besar kemungkinan sebuah negasi tersusun atas tiga kata.
- Tingkat akurasi dan presisi yang tergolong rendah juga diakibatkan oleh terdapat temuan kamus sentimen positif dan negatif baru yang mengakibatkan sentimen tidak terdeteksi atau tidak bisa terklasifikasi. Semakin banyak data latih yang dianalisis, semakin banyak pula pola penulisan, gaya bahasa serta sentimen masyarakat yang dapat diklasifikasi.

5.2. Saran

- Crawling tweet* dapat menggunakan program yang sama dalam hal ini menggunakan PHP agar dari tahap preprocessing atribut ke preprocessing data tidak pindah aplikasi.
- Data latih yang digunakan dan sudah tervalidasi lebih banyak sehingga pengklasifikasian lebih

akurat dikarenakan kondisi penggunaan bahasa dalam media sosial beragam sekali.

- Penyimpanan data dapat menggunakan database supaya hasil hitung data tersimpan dan pada setiap mengecek ulang hasil hitung tidak melakukan penghitungan dari awal.
- Proses *stemming* dapat menggunakan algoritma lain atau mengembangkan *stemming* yang sudah ada agar beberapa kata yang belum bisa diklasifikasi selanjutnya bisa.

Daftar Pustaka

- Ankur Goel, Jyoti Gautam, Sitiesh Kumar. (2016). *Real Time Sentiment Analysis of Tweets Using Naive Bayes*. JSS Academy of Technical Education India.
- Chowdhury, A., Frieder, O., Grossman, D., & McCabe, M. C. (2002). *Collection statistics for fast duplicate document detection*. ACM Transactions on Information Systems (TOIS), 20(2), 171-191.
- Danil, E. (2011). *Korupsi: Konsep, tindak pidana, dan pemberantasannya*. RajaGrafindo Persada.
- Evi Hartanti. (2007). *Tindak Pidana Korupsi*, Jakarta edisi kedua: Sinar graфика.
- Harlili., Yudi Wibisono. (2013). *Sistem Analisis Opini Microblogging Berbahasa Indonesia*. UPI Bandung.
- Huma Parveen, Prof. Shikha Pandey. (2016). *Sentiment Analysis on Twitter Data-set using Naive Bayes Algorithm*. Rungta College of Engineering and Technology India.
- Kohavi, R., & Provost, F. (1998). *Confusion matrix*. *Machine learning*, 30(2-3), 271-274.
- KPK. (2008). *Laporan Tahunan 2007*. Jakarta.
- Liu, B. (2012). *Sentiment analysis and opinion mining*. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- Manalu, B. U. (2014). *Analisis Sentimen Pada Twitter Menggunakan Text Mining*. Analisis Sentimen pada Twitter Menggunakan Text Mining.
- Mulyana, I., Ramadona, S., & Herfina, H. (2013). *Penerapan Terms Frequency-Inverse Document Frequency Pada Sistem Peringkasan Teks Otomatis Dokumen Tunggal Berbahasa Indonesia*. Proceedings of KNASTIK.
- Nabila, S. (2017). *Analisis Sentimen pada Twitter*

- tentang Jasa Transportasi Online di Indonesia dengan Metode Support Vector Machine*. Skripsi, Fakultas Ilmu Komputer.
- Nomor 31, U. U. R. I. (1999). *Tentang pemberantasan tindak pidana korupsi*. Undang-Undang Nomor, 20.
- Nomor 30, U. U. R. I. (2002). *Tentang Komisi Pemberantasan Tindak Pidana Korupsi*. Lembaran Negara Republik Indonesia Nomor, 4250.
- Saraswati, N. W. S. (2011). *Text mining dengan metode naïve bayes classifier dan support vector machines untuk sentiment analysis*. Universitas Udayana Denpasar.
- Suprpto, F. (2015). *Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode Naïve Bayes (Studi Kasus Sentiment Analysis Dari Media Twitter)*. SESINDO 2015, 2015.
- Surachim dan Suhandi Cahaya. (2011). *Strategi dan Teknik Korupsi*, cetakan pertama. Jakarta: Sinar grafika.
- Syamsuddin, A. (2011). *Tindak Pidana Khusus*. Sinar Grafika.
- Wang, A. H. (2010, July). *Don't follow me: Spam detection in twitter*. In *Security and cryptography (SECRYPT), proceedings of the 2010 international conference on* (pp. 1-10). IEEE.
- Wibisono, Y. (2005). *Aplikasi Berita Berbahas Indonesia Memakai Naive Bayes Classifier*. UPI Bandung.
- Young Gyo Jung, Kyung Tae Kim, Byungjun Lee, Hee Yong Youn. (2016). *Enhanced Naive Bayes Classifier for Real-time Sentiment Analysis with SparkR*. Sungkyunkwan University Korea.