

DETEKSI EMOSI MEDIA SOSIAL MENGGUNAKAN TERM FREQUENCY- INVERSE DOCUMENT FREQUENCY

*SOCIAL MEDIA EMOTION DETECTION USING TERM FREQUENCY-INVERSE DOCUMENT
FREQUENCY*

Arif Nur Rohman^{*1}, Riska Dwi Handayani², Ryan Dwi Y.P³, Kusri⁴

^{1,2,3,4} Magister Teknik Informatika, Universitas AMIKOM Yogyakarta

e-mail: ^{*1}arifrahman2592@gmail.com, ²riska.1116@students.amikom.ac.id,

³ryan.prakoso@students.amikom.ac.id, ⁴kusri@amikom.ac.id

Abstrak

Pada saat ini, manusia cenderung mengekspresikan pendapat, dan emosi melalui media sosial. Keterbukaan ekspresi pada media sosial membuat batasan pribadi seseorang menjadi lebur. Orang tidak lagi sungkan menulis kehidupan pribadinya melalui postingan status pembaharuan untuk dilihat oleh orang lain. Penulis mencoba menggunakan data dari media sosial agar dapat dilakukan analisis untuk mendapatkan informasi kepribadian termasuk emosi. Sebelum dianalisis, data dilakukan pra pemrosesan membuang symbol dan icon, normalisasi teks, stemming dan membuang stopword terlebih dahulu untuk memperbaiki data dari media sosial yang menggunakan bahasa tidak baku. Pra pemrosesan memperoleh hasil akhir 365 dataset yang dipisah menjadi 265 data latih dan 100 data uji. Tiap data sudah memiliki label emosi yang diberikan secara manual oleh pengguna facebook selaku penulis dari postingan. Perhitungan dilakukan dengan TF IDF untuk mendapatkan bobot tiap kata atau token, lalu label emosi data latih dengan nilai tertinggi diberikan pada data yang diuji. Dari 100 data uji diperoleh akurasi 59%, selebihnya 41% tidak akurat. Emosi dianggap hal penting yang berpengaruh pada seseorang dalam pengambilan keputusan, sehingga deteksi emosi seseorang dapat dimanfaatkan sebagai input sistem pendukung keputusan penerimaan pegawai dan konseling kepribadian.

Kata kunci—Deteksi Emosi, Media Sosial, TF-IDF

Abstract

At this time, humans tend to express their opinions, and emotions through sosial media. Disclosure of expression to sosial media makes the limitation of one's personal boundaries become melting. People no longer hesitate to write their personal lives through post-renewal status for others to see. Authors try to use data from sosial media in order to do analysis to get personality information including emotions. Before analysing, data is pre-processing to remove symbols and icons, text normalization, stemming and discard recognized in advance to improve data from sosial media using unraw language. Pre-processing gets the final result of 365 datasets that are split into 265 training data and 100 test data. Each data already has an emotional label that is manually assigned by Facebook users as the author of the Post. Calculations are done with TF IDF to get the weight of each word or token, then the Trainer Data Emotion label with the highest value given to the data tested. From 100 test data acquired accuracy of 59%, the rest 41% inaccurate. Emotions are considered an important thing that affects someone in decision making, so that one's emotional detection can be utilized as an input sistem of employee acceptance decision support and personality counseling.

Keywords—Emotion Detection, Sosial Media, TF-IDF

1. PENDAHULUAN

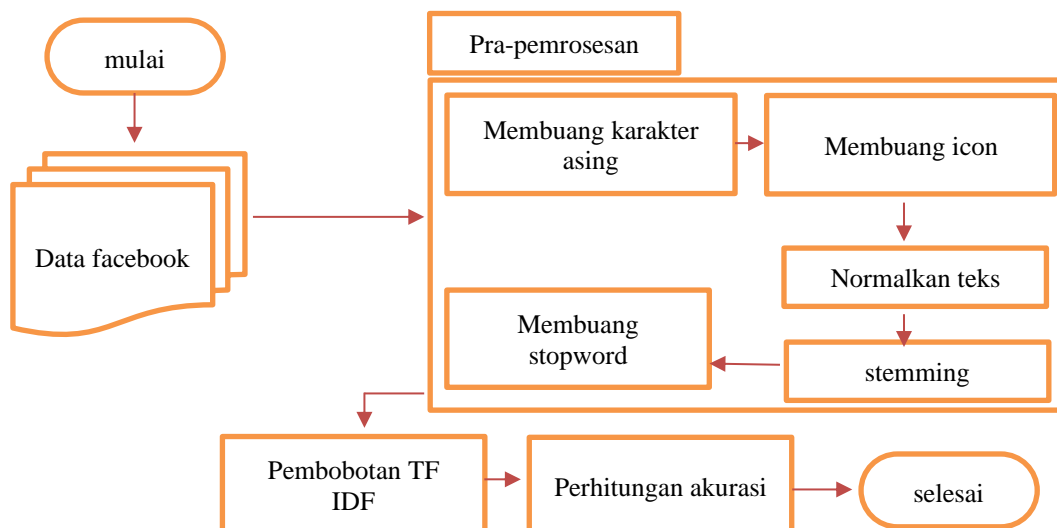
Emosi adalah suatu perasaan atau gejala jiwa yang muncul di dalam diri seseorang sebagai akibat dari adanya rangsangan, baik dari dalam diri sendiri maupun dari luar. Emosi dapat bervariasi berdasarkan kepribadian, jenis kelamin, lokasi, etnis, budaya, situasi, di samping banyak parameter psikologis, sosial dan individu lainnya. Deteksi emosi merupakan bagian dari prediksi kepribadian yang bermanfaat untuk rekrutmen karyawan, konseling dan profil psikologis [1]. Deteksi emosi dalam teks merupakan bidang penelitian yang banyak diminati dalam kajian linguistik dan komputer sains. Deteksi emosi yang ada saat ini menggunakan wajah, suara dan teks [2].

Para peneliti memanfaatkan media sosial seperti twitter dan facebook untuk menjadi sumber data dalam deteksi emosi. Media sosial memiliki data dalam bentuk video, foto dan umumnya teks yang mencerminkan ekspresi pengguna. Teks merupakan bentuk data yang paling banyak digunakan karena mudah diolah. Selain itu, media sosial juga memiliki data dalam jumlah yang besar. Ada 2 cara untuk analisis teks berdasarkan deteksi emosi, yaitu model kategori dari dimensi. Setiap model memiliki kelebihan dan kekurangan serta perbedaan konsep dalam prediksi emosi [3]. Dalam model kategori, model Ekman merupakan yang paling banyak digunakan. Model Ekman membagi emosi dalam enam label dasar, yaitu bahagia, marah, takut, jijik, sedih, dan terkejut [4]. Label-label ini bersifat universal pada budaya yang berbeda. Pendekatan yang digunakan dalam deteksi emosi pada teks media sosial antara lain pendekatan keyword, pendekatan kamus, pendekatan *machine learning* dan pendekatan *hybrid* [5]. Pendekatan hybrid adalah gabungan dari pendekatan kamus dan *machine learning*.

Saat ini penelitian yang ada memanfaatkan twitter sebagai sumber data deteksi emosi, sedangkan Facebook merupakan media sosial dengan pengguna aktif mencapai 2,4 miliar akses tiap bulan dan memiliki batasan karakter teks yang lebih panjang dari twitter. Sehingga perlu dilakukan penelitian deteksi emosi media sosial menggunakan data dari facebook. Penulis akan mencoba menggunakan teks status pembaharuan pengguna facebook sebagai dataset. Pengguna facebook tidak terbatas pada wilayah dan budaya tertentu sehingga model emosi yang penulis pilih adalah model Ekman. Penelitian sebelumnya pernah dilakukan menggunakan dataset yang bersumber dari Facebook dengan pendekatan kamus menggunakan EmoLex [6]. Maka pada penelitian ini penulis menggunakan Term Frequency-Inverse Document Frequency (TF-IDF).

2. METODE PENELITIAN

Gambaran umum deteksi emosi menggunakan data facebook disajikan pada gambar 1.



Gambar 1. Metode penelitian

Penulis meminta izin pengguna facebook yang datanya akan digunakan dalam penelitian ini. Pengguna yang dipilih berusia 20 sampai 30 tahun tanpa memperhatikan jenis kelamin. Data diambil dari facebook menggunakan tool scrapping arifScrapper yang dibuat mandiri oleh peneliti dan

dipublikasikan di <https://github.com/arifrahman2592>. Kemudian data diberikan kepada pengguna selaku author untuk diberikan label emosi. Setelah itu data dipisah menjadi data latih dan data uji.

Setelah data peroleh, selanjutnya dilakukan pra-pemrosesan sebagai berikut:

- a) Membuang symbol dan kode ASCII
Teks dari facebook mengandung karakter yang tidak diinginkan seperti #!%^:). karakter seperti ini dibuang dan huruf dikonversi menjadi *lowercase*.
- b) Membuang icon
Penelitian ini berfokus pada teks, sehingga icon dibuang. Daftar icon disajikan dalam tabel 1.

Tabel 1. Daftar icon yang dibuang

:)	:[=P	:-D	;))	-_-	:v	B-
:]	=(:-p	:o	;-)	>:o	:3	>:(
=)	:-(-	:-P	:O	8-)	>:O	8	>:-(-
:-)	:p	:D	:-o	B-)	>:-O	B	:/
:(:P	=D	:-O	^_^	>:-o	8-	:-/
O:)	O:-)	:*	:-*	<3	(y)	(Y)	:poop:

- c) Normalkan teks
Pengguna Facebook sering menggunakan bahasa yang tidak baku, singkatan dan typo dalam menulis status pembaharuan, sehingga teks perlu dinormalkan. Untuk menormalkan teks penulis menggunakan tool dari [prosa.ai](https://api.prosa.ai/v1/normals), sebuah layanan berbasis API untuk memperbaiki teks bahasa Indonesia. Layanan ini dapat diakses melalui <https://api.prosa.ai/v1/normals>. Contoh penggunaan normalisasi pada kalimat “Aq suka mkn nasi goreng di dkt sklh” menjadi “Aku suka makan nasi goreng di dekat sekolah”. Meskipun dengan inputan kalimat, namun tool ini bekerja dengan menormalkan tiap kata, kata yang gagal dinormalkan akan dikembalikan apa adanya.
- d) Stemming
Stemming merupakan salah satu tahapan dalam proses text pre-processing. Stemming bertujuan untuk mentransformasikan kata menjadi kata dasarnya [7] (root) dengan menghilangkan semua imbuhan kata (affixes) meliputi awalan kata (prefixes), sisipin kata (infixes), akhiran kata (suffixes) dan atau menghilangkan awalan dan akhiran kata (confixes) pada kata turunan. Teks yang sudah dinormalkan, selanjutnya dilakukan stemming menggunakan tool stemmer dari Sastrawi, sebuah *library* yang ditulis menggunakan Bahasa python. Library ini menerapkan algoritma Nazief dan Adriani. Cara kerja algoritma ini menghilangkan Inflection Suffixes (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”). Jika kata berupa *particles* (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus Possesive Pronouns (“-ku”, “-mu”, atau “-nya”). Selanjutnya juga menghilangkan derivational *suffix* (imbuhan turunan) dengan cara menghilangkan imbuhan -i, -kan, -an. Menghilangkan derivational *prefix* (awalan turunan) dengan cara menghilangkan awalan be-, di-, ke-, me-, pe-, se- dan te-. Namun bila semua proses tersebut gagal, maka algoritma mengembalikan teks apa adanya tanpa proses stemming.
- e) Membuang stopwords
Dalam NLP (Natural Language Processing) stopwords merupakan kata yang diabaikan dalam pemrosesan, kata-kata ini biasanya disimpan ke dalam stop lists. Karakteristik utama dalam pemilihan stop word biasanya adalah kata yang mempunyai frekuensi kemunculan yang tinggi misalnya kata penghubung seperti “dan”, “atau”, “tapi”, “akan” dan lainnya. Tidak ada aturan pasti dalam menentukan stop word yang akan digunakan, penentuan stop word bisa disesuaikan dengan kasus yang sedang diselesaikan. Kata dasar yang tersisa dilakukan tahap pembuangan stopwords. Total stopwords yang digunakan ada 758 kata yang bersumber

dari basis data Sastrawi yang dipublikasikan melalui <https://github.com/sastrawi/sastrawi>. Contoh stopword Sastrawi disajikan dalam tabel 2.

Tabel 2. Contoh stopword Sastrawi

ada	agaknya	akhiri	amatlah
adalah	agar	akhirnya	anda
adanya	akan	aku	andalah
adapun	akankah	akulah	antar
agak	akhir	amat	antara

Pada tabel 2, kata penghubung, kata pengganti orang dan kata yang sering muncul dianggap stopwords dan harus dihilangkan sehingga perhitungan emosi dilakukan dengan kata kata yang tepat [8].

Pada dokumen yang besar, skema yang paling sukses dan secara luas digunakan untuk pemberian bobot term adalah skema pembobotan atau Term Weighting TF-IDF. Hal yang perlu diperhatikan dalam pencarian informasi dari koleksi dokumen yang heterogen adalah pembobotan term. Term dapat berupa kata, frase atau unit hasil indexing lainnya dalam suatu dokumen yang dapat digunakan untuk mengetahui konteks dari dokumen tersebut, maka untuk setiap kata tersebut diberikan indikator, yaitu term weight. Sebagai ukuran statistik, TF-IDF memberikan bobot skor untuk mengevaluasi seberapa penting suatu frasa dalam satu set dokumen yang terdeteksi. Penulis menggunakan bobot TF-IDF karena skema pembobotan ini telah terbukti mencapai kinerja yang lebih baik daripada algoritma lain seperti TextRank, SingleRank dan ExpandRank [9]. Perhitungan akurasi dilakukan dengan membandingkan hasil pelabelan emosi menggunakan TF-IDF dengan manual oleh pengguna.

3. HASIL DAN PEMBAHASAN

3.1. Dataset dan Pra-pemrosesan

Teks yang diambil dari facebook bersumber dari fitur status pembaharuan. Teks yang diambil hanya yang berbahasa Indonesia sebagai teks asli lalu dibuang simbol dan icon yang tidak diinginkan. Penulis menggunakan akun Facebook penulis pribadi sebagai masukan *tool* arifScraping untuk mendapatkan status pembaharuan dari pengguna Facebook yang ditargetkan. Data yang terambil disimpan pada basis data agar mudah dilakukan pra pemrosesan tahap berikutnya. Hal ini juga bermanfaat apabila terjadi masalah saat pengambilan data, penulis tidak perlu mengulangi dari awal. Tiap pengguna diperoleh antara 20 sampai 30 status pembaharuan dengan total diperoleh 400 dataset. Data teks yang diambil tidak hanya murni status pembaharuan yang berbentuk teks, namun juga status pembaharuan yang mengandung gambar walaupun yang diambil hanya teksnya saja. Selanjutnya penulis melakukan seleksi secara manual untuk menghapus dataset yang menggunakan selain bahasa Indonesia, umumnya bahasa inggris dan daerah. Dari seleksi manual tersisa 365 dataset. Teks tiap tahapan pra pemrosesan tepat disimpan agar dapat diketahui perubahan teks pada tiap tahapannya mulai dari teks asli hingga teks hasil stopwords. Bentuk dari teks asli dan yang sudah dibuang simbolnya seperti pada tabel 1.

Tabel 1. Teks asli dan yang sudah dibuang simbolnya

teks_asli	Buang simbol dan ASCII	Teks buang icon
Cinta kepada Allah dengan KETAATAN. Cinta kepada Rasul dengan keperibadian MULIA. Cinta kepada Jihad dengan PERJUANGAN :) ♡.	cinta kepada allah dengan ketaatan cinta kepada rasul dengan keperibadian mulia cinta kepada jihad dengan perjuangan ♡	cinta kepada allah dengan ketaatan cinta kepada rasul dengan keperibadian mulia cinta kepada jihad dengan perjuangan

Jangan terlalu sibuk mencari KESEMPURNAAN, jika KESEDERHANAAN mampu membuatmu bahagia	jangan terlalu sibuk mencari kesempurnaan jika kesederhanaan mampu membuatmu bahagia	jangan terlalu sibuk mencari kesempurnaan jika kesederhanaan mampu membuatmu bahagia
Ya ALLAH Sekiranya aku jatuh cinta. Maka jatuhkanlah cintaku pada seseorang yang melabuhkan cintanya pada-Mu. Agar bertambah kekuatanku untuk mencintai-Mu #doa	ya allah sekiranya aku jatuh cinta maka jatuhkanlah cintaku pada seseorang yang melabuhkan cintanya pada-mu agar bertambah kekuatanku untuk mencintai-mu	ya allah sekiranya aku jatuh cinta maka jatuhkanlah cintaku pada seseorang yang melabuhkan cintanya padamu agar bertambah kekuatanku untuk mencintaimu

Penulis menggunakan program yang dibuat secara mandiri dengan Bahasa PHP untuk menghilangkan simbol, ASCII dan icon pada semua dataset.

Selanjutnya teks dinormalkan untuk memperbaiki kosa kata yang tidak baku, *stemming* untuk mendapatkann kata dasar, dan buang *stopword* untuk menghilangkan kata yang dianggap kurang memiliki makna seperti pada tabel 2.

Tabel 2. Teks normal, *stemming* dan buang *stopword*.

Teks normal	Teks stemming	Teks stopwords	keyword	label
cinta kepada allah dengan ketaatan cinta kepada rasul dengan keperibadian mulia cinta kepada jihad dengan perjuangan	cinta kepada allah dengan taat cinta kepada rasul dengan keperibadian mulia cinta kepada jihad dengan juang	cinta allah taat cinta rasul keperibadian mulia cinta jihad juang	Cinta rosul	bahagia
jangan terlalu sibuk mencari kesempurnaan jika kesederhanaan mampu membuat mu bahagia	jangan terlalu sibuk cari sempurna jika sederhana mampu buat mu bahagiadan	sibuk cari sempurna sederhana mu bahagia	kesempurnaan	sedih
ya allah sekiranya aku jatuh cinta maka jatuhkanlah cinta ku pada seseorang yang melabuhkan cinta nya pada mu agar bertambah kekuatan ku untuk mencintai mu	ya allah sekira aku jatuh cinta maka jatuh cinta ku pada orang yang labuh cinta nya pada mu agar tambah kuat ku untuk cinta mu	ya allah sekira jatuh cinta jatuh cinta ku orang labuh cinta nya mu kuat ku cinta mu	Jatuh cinta	bahagia

Kolom keyword harus diisi oleh pemilik status pembaharuan saat memberikan label emosi. Kolom keyword berisi kosa kata yang berasal dari status pembaharuan yang terkait dengan label emosinya. Kolom keyword digunakan penulis sebagai validasi bahwa pengguna Facebook mengerti dengan label emosi yang diberikan pada teks status pembaharuan. Teks hasil pra pemrosesan yang digunakan dalam deteksi emosi adalah teks hasil stopwords. Teks hasil stopwords merupakan tapat terakhir pada pra pemrosesan. Total dataset yang digunakan sebanyak 365 status pembaharuan, 265 digunakan untuk data latih dan 100 untuk data uji.

3.2. Model Emosi Ekman

Model emosi Ekman ditemukan oleh Paul Ekman, psikolog amerika serikat yang mendeteksi emosi dari perubahan fisiologis pada wajah manusia. Ekman membagi emosi manusia dalam 6 emosi dasar, yaitu bahagia, takut, marah, sedih, jijik dan terkejut [9]. Selain 6 emosi dasar, terdapat banyak sekali emosi sekunder, yaitu gabungan dari 2 atau lebih emosi primer. Misalnya kecewa merupakan gabungan dari emosi marah dan sedih. Terharu merupakan gabungan dari emosi senang dan sedih. Jumlah dataset yang diperoleh penulis pada tiap label emosi disajikan pada tabel 3.

Tabel 3. Jumlah dataset pada tiap label emosi

Label emosi	Jumlah dataset
Bahagia	161
Jijik	25
Kaget	8
Marah	25
Sedih	122
Takut	24
Total dataset	365

Jumlah dataset pada tiap label tidak sama, terdapat label dengan jumlah dataset sangat banyak seperti bahagia dan sedih. Ada pula yang terlalu sedikit seperti kaget, marah dan takut. Hal ini dapat berpengaruh pada hasil akurasi deteksi emosi. Terdapat banyak dokumen dan token yang menunjukkan emosi bahagia dan sedih, Namun sangat sedikit token yang menunjukkan emosi kaget, marah dan takut.

3.3. Perhitungan TF-IDF

Dalam pencarian informasi, TF-IDF adalah statistik numerik yang dimaksudkan untuk mencerminkan betapa pentingnya sebuah kata atau term dalam koleksi atau *corpus*. Dataset dipisah menjadi 2, yaitu data latih dan data uji. Tiap 1 data uji akan diujikan pada semua data latih. Masing-masing data latih dan uji disebut dokumen. Penulis membuat program secara mandiri untuk melakukan perhitungan TF-IDF pada semua dataset yang ada. Penulis memberikan contoh perhitungan TF-IDF 1 data uji pada 3 data latih. Penulis pilih dokumen dengan karakter yang pendek seperti pada table 4.

Tabel 4. data uji dan data latih

Dokumen uji	monyet cinta remaja	Ini yang dicari label emosinya
Dokumen latih 1	cinta allah taat cinta rasul	Bahagia
Dokumen latih 2	sibuk cari sempurna sederhana mu bahagia	Sedih
Dokumen latih 3	ya allah sekira jatuh cinta bahagia	Bahagia

Setiap kata atau term dalam dokumen latih (dokumen 1,2 dan 3) dipisah tiap kata atau disebut token. Token yang muncul lebih dari 1 kali tetap ditulis 1 kali untuk menghindari hasil ganda pada 1 token yang sama. Setiap token dicari TF-IDF dengan rumus sebagai berikut:

$$TFIDF = TF \times IDF = TF \times \log_e \frac{|D|}{DF}$$

Gambar 2. Term Frequency – Inverse Document Frequency

Dimana,

- TF = term frequency, banyaknya sebuah kata/istilah/token muncul
- DF = document frequency, jumlah dokumen di mana kata/istilah/token tersebut muncul. Minimal 1 dokumen

Konsekuensinya TF tinggi maka TFI-DF akan naik, maka istilah/kata/token tersebut penting. DF tinggi artinya TFI-DF menjadi rendah [10]. Selanjutnya hasil TF-IDF digunakan untuk menacari bobot masing-masing token dengan cara mengkalikan IDF dengan jumlah kemunculan token pada suatu dokumen. Hasil perhitungan bobot tiap token disajikan dalam tabel 5.

Tabel 5. Hasil perhitungan bobot token

token	tf			df	D/df	IDF	W (bobot)		
	D1	D2	D3				D1	D2	D3
cinta	2		1	2	1,5	0,176091	0,352183	0	0,176091
allah	1		1	2	1,5	0,176091	0,176091	0	0,176091
taat	1			1	3	0,477121	0,477121	0	0
rosul	1			1	3	0,477121	0,477121	0	0
sibuk		1		1	3	0,477121	0	0,477121	0
cari		1		1	3	0,477121	0	0,477121	0
sempurna		1		1	3	0,477121	0	0,477121	0
sederhana		1		1	3	0,477121	0	0,477121	0
mu		1		1	3	0,477121	0	0,477121	0
bahagia		1		1	3	0,477121	0	0,477121	0
ya			1	1	3	0,477121	0	0	0,477121
sekira			1	1	3	0,477121	0	0	0,477121
jatuh			1	1	3	0,477121	0	0	0,477121
bahagia			1	1	3	0,477121	0	0	0,477121
total bobot tiap dokumen							1,482516	2,862728	2,260667

Pada tabel 5, tiap token yang ada di semua dokumen dicari keberadaannya di masing-masing dokumen (TF) dan hitung jumlah yang muncul di setiap dokumen. Setelah diketahui jumlah yang muncul di tiap dokumen, jumlahkan berapa dokumen suatu token muncul (DF), token yang muncul lebih dari 1 kali pada 1 dokumen yang sama, dianggap 1. Selanjutnya jumlah dokumen (D) dibagi dengan kemunculan dokumen (DF) lalu di $\log(10)$ untuk mendapatkan nilai IDF. Nilai IDF tiap token dikalikan dengan jumlah kemunculan di suatu dokumen untuk mendapatkan bobot (W) token pada dokumen itu.

Pada tabel 5 dari 3 dokumen yang dicontohkan sudah diketahui bobot masing-masing kata atau token. Selanjutnya 1 dokumen uji akan dihitung bobotnya berdasarkan bobot token pada tabel 5. Token yang terdapat pada dokumen uji adalah monyet, cinta, dan remaja, maka token token tersebut dicarikan bobotnya berdasarkan pembobotan token pada tabel hasil perhitungan bobot token. Perhitungan bobot dokumen uji yang mengandung 3 token tersebut disajikan pada tabel 6.

Tabel 6. Perhitungan bobot dokumen uji

token	dokumen		
	D1	D2	D3
monyet	0	0	0
cinta	0,3521825	0	0,176091
remaja	0	0	0
total	0,3521825	0	0,176091

Untuk mencari dokumen latih mana dengan total bobot tertinggi pada dokumen uji. Dokumen uji dipecah menjadi token. Token pada dokumen uji diberikan bobot pada tiap dokumen latih, lalu tiap dokumen dijumlahkan semua bobot tokennya, dokumen dengan jumlah tertinggi dipilih sebagai hasil.

Pada tabel 6 dokumen 1 (D1) memiliki total tertinggi yaitu 0,3521825 sehingga dokumen latih diberikan label emosi yang sama dengan dokumen 1 yaitu bahagia.

3.4. Perhitungan Akurasi

100 dokumen uji yang dihitung satu per satu menggunakan pembobotan token 265 dokumen latih untuk mengetahui label emosi dokumen uji berdasarkan bobot tertinggi tiap dokumen latih yang mengandung token dokumen uji. Hasil pelabelan emosi menggunakan TF-IDF dibandingkan dengan pelabelan manual pemilik status Facebook yang diujikan untuk mengetahui akurasi. Pelabelan manual sudah disiapkan saat pembuatan dataset. Label yang sama antara manual dan TF-IDF dinyatakan sebagai akurat selain itu dinyatakan tidak. Hasil pencocokan label disajikan pada tabel 7.

Tabel 7. Perhitungan dokumen latih

Kata	Label Manual	Label TF IDF	Akurasi
orang hebat ribu hebat	marah	marah	akurat
ganteng ganteng iman	bahagia	bahagia	akurat
hubung sama	bahagia	bahagia	akurat
moga kelak aroma kopi hangat ku rumah	bahagia	bahagia	akurat
ya malu	marah	sedih	tidak
bahagia ummat	bahagia	bahagia	akurat
mustahil tuju	sedih	sedih	akurat
..
..
libur asyik wifi id corner	bahagia	bahagia	akurat
alhamdulillah sukses kejordatapaksucizamzam	bahagia	bahagia	akurat
ketemu nabillah jfestpwt2016 kualitas	bahagia	bahagia	akurat
cari cari beasiswa nipon senang	bahagia	sedih	tidak
impi peluh bunga mekar perlahan usaha keras khianat harap terkabulshonichi	takut	bahagia	tidak
congratulation menang semangat joyful kawan try to be the best	bahagia	bahagia	akurat
pimpin nya aja contoh negara maju	jijik	takut	tidak
moga ramai fap akrab kasih rekomendasi anime populer giveawayfap	bahagia	bahagia	akurat
bukti isis cipta amerika	marah	marah	akurat

Hasil pencocokan pada tabel 7 dari 100 dokumen uji diperoleh 59 (59%) dokumen dengan label akurat dan 41 (41%) dokumen tidak akurat. Pada penelitian sebelumnya menggunakan dataset dari Facebook dengan kamus EmoLex [6] mendapatkan nilai akurasi 61,53%. Hasil akurasi yang diperoleh pada penelitian ini lebih rendah dikarenakan jumlah dataset tiap label emosi tidak sama. Emosi bahagia memiliki dataset dengan label terbanyak, sehingga peneliti melakukan skenario pengujian khusus emosi bahagia saja. Dari 47 dataset dengan label bahagia diperoleh hasil 41 atau 87,23% akurat. Hal ini mengindikasikan bahwa jumlah dataset berpengaruh besar dalam deteksi emosi menggunakan TF-IDF.

4. KESIMPULAN

Update status pengguna facebook dapat dijadikan sebagai data untuk mendeteksi emosi seseorang. Pengambilan data menggunakan *scrapping* dapat menyingkat waktu, namun harus dilakukan pra-pemrosesan terlebih dahulu untuk membuang symbol dan icon yang tidak diperlukan. Selain itu teks dari facebook banyak menggunakan Bahasa yang tidak baku, sehingga perlu melakukan normalisasi teks, kata yang dianggap penting harus dibuang menggunakan stopword. TF IDF dapat digunakan untuk menghitung bobot tiap kata/token untuk menentukan label emosi

dokumen uji berdasarkan dokumen latih. Hasil akurasi pelabelan emosi dengan TF IDF dibandingkan dengan manual sebesar 59%. Hasil pengujian pada dataset khusus emosi bahagia memperoleh hasil 87,23% akurat. Hal ini menunjukkan bahwa akurasi deteksi emosi menggunakan TF-IDF bergantung pada banyaknya dataset yang digunakan. Hasil label emosi pada penelitian ini dapat digunakan sebagai input sistem pendukung keputusan penerimaan pegawai, mengingat salah satu manfaat deteksi kepribadian termasuk emosi adalah untuk perekrutan pegawai.

5. SARAN

Berdasarkan hasil akurasi yang diperoleh pada penelitian ini, diperlukan normalisasi teks yang lebih baik dari yang sudah digunakan, mengingat teks dari media sosial tidak baku. Selain itu perlu menggunakan jumlah dataset yang sama pada tiap label emosi agar setiap data latih memiliki kekuatan yang sama dalam menghitung data uji. Pembuatan dataset dapat menggunakan metode lain yang lebih valid dalam hal pelabelan emosi agar dapat meningkatkan nilai akurasi deteksi emosi menggunakan TF-IDF.

DAFTAR PUSTAKA

- [1] P. S. Dandannavar, S. R. Mangalwede, and P. M. Kulkarni, "Sosial Media Text - A Source for Personality Prediction," *2018 Int. Conf. Comput. Tech. Electron. Mech. Syst.*, pp. 62–65, 2018.
- [2] T. Rabeya, S. Ferdous, H. S. Ali, and N. R. Chakraborty, "A survey on emotion detection: A lexicon based backtracking approach for detecting emotion from Bengali text," in *20th International Conference of Komputer and Information Technology, ICCIT 2017*, 2018, vol. 2018-Janua, pp. 1–7.
- [3] F. H. Rachman, R. Sarno, and C. Fatichah, "CBE: Corpus-based of emotion for emotion detection in text document," in *Proceedings - 2016 3rd International Conference on Information Technology, Komputer, and Electrical Engineering, ICITACEE 2016*, 2017, pp. 331–335.
- [4] M. F. Alhamid, S. Alsahli, M. Rawashdeh, and M. Alrashoud, "Detection and Visualization of Arabic Emotions on Sosial Emotion Map," in *Proceedings - 2017 IEEE International Symposium on Multimedia, ISM 2017*, 2017, vol. 2017-Janua, pp. 378–381.
- [5] S. Al-Saaqa, H. Abdel-Nabi, and A. Awajan, "A Survey of Textual Emotion Detection," in *2018 8th International Conference on Komputer Science and Information Technology, CSIT 2018*, 2018, pp. 136–142.
- [6] A. N. Rohman, E. Utami, and S. Raharjo, "Deteksi Emosi Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing," pp. 70–76, 2019.
- [7] L. Povoda, R. Burget, and M. K. Dutta, "Sentiment analysis based on Support Vector Machine and Big Data," in *2016 39th International Conference on Telecommunications and Signal Processing, TSP 2016*, 2016, pp. 543–545.
- [8] S. X. Mashal and K. Asnani, "Emotion intensity detection for sosial media data," *Proc. Int. Conf. Comput. Methodol. Commun. ICCMC 2017*, vol. 2018-Janua, no. Iccmc, pp. 155–158, 2018.
- [9] C. S. Montero, H. Haddad, and M. Mozgovoy, *Detecting the Likely Causes Behind the Emotion Spikes of Influential Twitter Users*, vol. 1. Springer International Publishing.
- a) Buku : O. W. Purbo, 2019, *Text Mining Analisis MedSos Kekuatan Brand dan Intelegen di Internet*, Andi, Yogyakarta.