

Perbandingan Ekstraksi Fitur dengan Pembobotan Supervised dan Unsupervised pada Algoritma Random Forest untuk Pemantauan Laporan Penderita COVID-19 di Twitter

¹Sulastris Norindah Sari, ^{2,*}Mohammad Reza Faisal, ³Dwi Kartini, ⁴Irwan Budiman, ⁵Triando Hamonangan Saragih & ⁶Muliadi

^{1,2,3,4,5,6} Program Studi Ilmu Komputer, Universitas Lambung Mangkurat, Banjarbaru, Indonesia

Abstrak — Dimasa sekarang masyarakat sudah berani melaporkan dirinya terpapar COVID-19 melalui unggahan di media sosial seperti Twitter. Hal ini dapat dimanfaatkan oleh masyarakat sekitar atau lembaga kesehatan untuk memberikan bantuan terhadap pelapor. Pemantauan laporan penderita COVID-19 di Twitter dapat dilakukan secara otomatis dengan algoritma machine learning untuk klasifikasi teks. Pada kasus klasifikasi teks, algoritma machine learning menerima input berupa data terstruktur hasil ekstraksi fitur dengan teknik unigram dengan pembobotan. Metode pembobotan kata unsupervised merupakan pembobotan yang tidak memperhatikan letak term di kelas positif atau negatif. Kemudian metode pembobotan ini dikembangkan menjadi pembobotan supervised, karena dalam proses pembobotannya metode ini membobotkan term dengan memperhatikan letak term di kelas positif atau negatif. Pada riset ini dilakukan perbandingan kedua jenis pembobotan pada klasifikasi data tweet gejala covid dengan algoritma machine learning yaitu Random Forest. Dari hasil penelitian didapat hasil kinerja klasifikasi dengan pembobotan supervised Delta TF-IDF terbukti lebih bagus dengan akurasi sebesar 88,5% sedangkan dengan pembobotan unsupervised TF-IDF diperoleh hasil akurasi 87,9%.

Kata Kunci: klasifikasi teks; laporan mandiri gejala covid-19; Delta TF-IDF; TF-IDF; random forest

Abstract — Nowadays people have the courage to tell themselves that they have been exposed to COVID-19 through uploads on social media and many people tell stories on Twitter. The unsupervised word weighting method is a weighting method that does not pay attention to the position of the term in the positive or negative class. Then this weighting method was developed into supervised weighting, because in the process of weighting this method weights the terms by paying attention to the position of the term in the positive or negative class. Therefore, a study will be carried out by comparing the two types of weighting on the classification of tweet data with Covid symptoms using the Random Forest method. From the results of the study, it can be seen that the classification performance with the Delta weighting TF-IDF proved better with an accuracy of 88.5%, while the TF-IDF obtained an accuracy of 87.9%.

Keywords: text classification; self-reported covid-19 symptoms; Delta TF-IDF; TF-IDF; random forest.

* Corresponding author :
Mohammad Reza Faisal
Universitas Lambung Mangkurat, Banjarbaru, Indonesia
reza.faisal@ulm.ac.id

1. PENDAHULUAN

Pada masa sekarang masyarakat sudah mulai berani memberitahukan dirinya positif Covid-19 dengan mempublikasikannya di sosial media, dalam hal tersebut maka dapat dimanfaatkan data-data hasil kiriman cerita pengguna baik itu di Facebook, Twitter atau media sosial yang lain. Dengan adanya informasi orang-orang terdampak covid maka dapat bermanfaat nantinya untuk kemungkinan dipantau untuk dibantu oleh pemerintah atau keluarga kerabat terdekat. Menurut [1] dari twitter bisa didapatkan tweet berupa tanggapan ataupun opini akan suatu informasi yang penting, sehingga setiap tanggapan atau sentimen pengguna twitter dapat dikategorikan untuk dimanfaatkan. Dengan adanya analisis sentimen dapat mempermudah pemegang kebijakan atau pemerintah dalam meningkatkan kualitas

©2023 Ilmu Komputer Unila Publishing Network all rights reserved layanan kepada masyarakat [2]. Tujuan dasar analisis sentimen dapat mengklasifikasikan sejumlah teks dari dokumen dengan berupa fitur yang digunakan untuk diprediksi kedalam kelas positif atau negatif [3].

Text mining merupakan salah satu metode machine learning yang berfungsi untuk mengekstrak data teks yang tidak terstruktur sehingga didapatkan informasi atau pengetahuan. Pada penelitian [4] yang berjudul A Text Classification Approach for the Automatic Detection of Twitter Posts Containing Self-reported COVID-19 Symptoms melakukan penelitian pada data tweet tentang gejala Covid-19 dan membaginya menjadi dua kelas yaitu positif dan negatif.

Menurut [5] skema pembobotan dapat dibagi menjadi dua kategori berdasarkan penggunaan informasi kelas dalam dokumen. Yang pertama adalah pembobotan istilah tanpa pengawasan atau *Unsupervised Term Weighting* (UTW). UTW tidak menggunakan informasi kelas atau label untuk menghasilkan bobot. Metode UTW yang paling populer adalah TF-IDF (*Term Frequency-Inverse Document Frequency*). TF-IDF ini telah digunakan secara efektif dalam studi pencarian informasi namun, tidak terlalu cocok untuk teks klasifikasi. Kategori kedua dari skema pembobotan adalah pembobotan istilah yang diawasi atau *Supervised Term Weighting* (STW). Metode STW menggunakan informasi kelas dari dataset untuk menghitung pembobotan kata, contoh metode STW ini adalah Delta TF-IDF, TF-IDF-ICF dan TF-RF [6][7], [8]. Kekurangan metode TF-IDF adalah pada saat pembobotan tidak dapat memberikan nilai bobot yang berbeda pada term positif maupun negatif.

Penelitian [9] menjelaskan bahwa menggunakan pembobotan TF-IDF-ICF juga harus ada metode *cosine similarity* sebagai perhitungan jarak. Penelitian ini menggunakan metode klasifikasi KNN untuk mengklasifikasikan data ide kreatif perusahaan menjadi dua kelas yaitu Teknik dan Non-teknik. Hasil akurasi diperoleh nilai sebesar 93% yang berarti membuktikan kinerja pembobotan TF-IDF-ICF cukup baik digunakan dalam penelitian ini. Metode pembobotan *supervised* tidak hanya dapat digunakan dalam permasalahan pengklasifikasian sentimen namun juga dapat digunakan dalam pencarian dokumen teks seperti pada permasalahan yang ada dalam penelitian [10] yaitu pencarian teks pada terjemahan ayat Al-Qur'an dengan menggunakan TF-RF dan Bray-Curtis Distance memberikan peningkatan pada kinerja dibandingkan sebelum menggunakannya terbukti dengan diperoleh nilai MAP yang meningkat pada k-rank 5 sebesar 61,80%.

Pembobotan TF-IDF banyak digunakan dalam berbagai kasus *text mining* namun metode pembobotan *unsupervised* ini kurang bagus dalam kinerjanya dibuktikan dalam penelitian [11] menyebutkan bahwa metode pembobotan *supervised* TF-RF mengungguli kinerjanya dibandingkan dengan metode TF-IDF dengan diperoleh hasil akurasi cukup baik sebesar 67% menggunakan metode klasifikasi KNN dan ekstraksi fitur unigram.

Delta TF-IDF merupakan metode keterbaruan dari metode pembobotan TF-IDF yang dapat memberikan nilai bobot yang berbeda pada kata yang berlabel positif dan negatif. Dibuktikan dalam penelitian [6] melakukan penelitian mengenai analisis sentimen pada komentar video official trailer dengan menggunakan metode pembobotan Delta TF-IDF serta menerapkan teknik stemming dengan metode klasifikasi SVM diperoleh akurasi sebesar 82,88%, model romance sebesar 82,91%, model komedi sebesar 91,4%, model horor sebesar 80,03% dan model gabungan sebesar 85,2%.

Random Forest merupakan salah satu metode klasifikasi yang dapat digunakan dalam proses *text mining*. Metode ini adalah kumpulan dari Decision Tree. Random Forest mengkombinasikan hasil dari setiap pohon keputusan untuk memprediksi klasifikasi. Pada penelitian analisis sentiment [12] digunakan pembobotan TF-IDF dan menggunakan beberapa metode klasifikasi, didapatkan hasil klasifikasi terbaik dengan menggunakan metode Random Forest dengan hasil akurasi sebesar 75,37%.

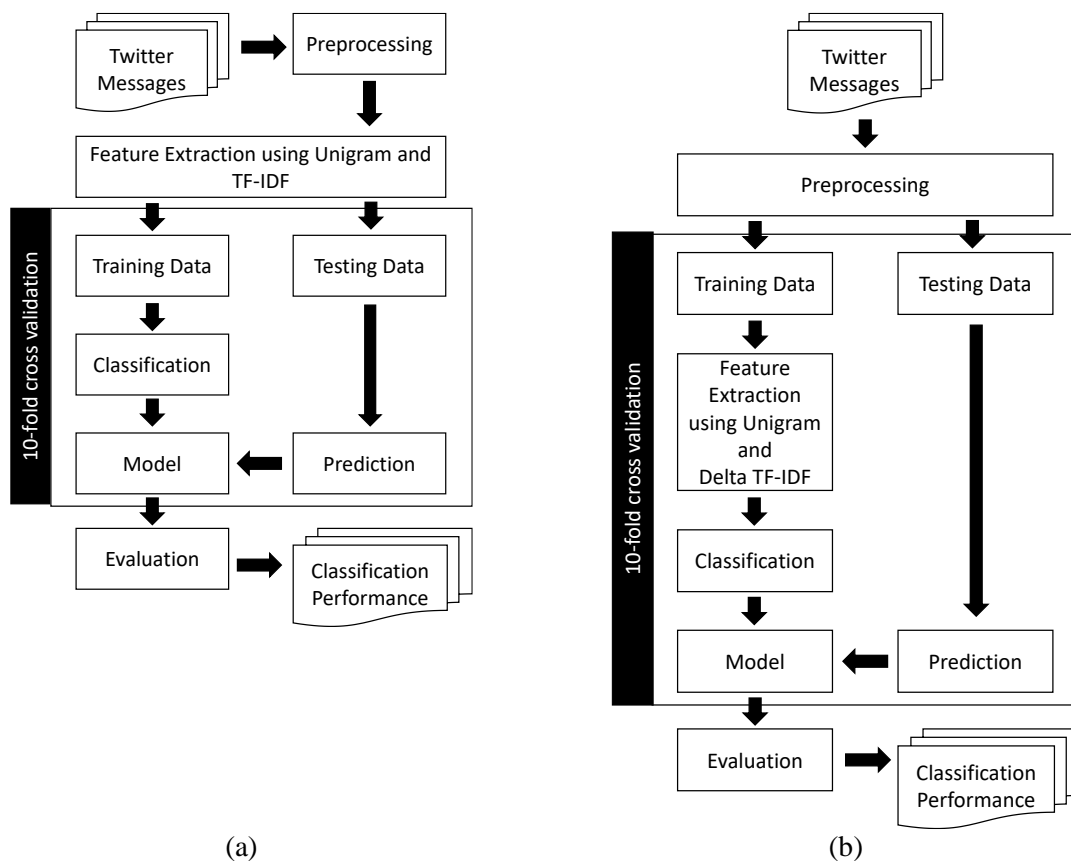
Berdasarkan paparan diatas telah dilakukan penelitian tentang klasifikasi untuk penentuan data tweet termasuk positif atau negatif bergejala COVID dan beberapa penelitian menyebutkan bahwa kinerja pembobotan *supervised* lebih baik dibandingkan dengan pembobotan *unsupervised* dalam kasus

©2023 Ilmu Komputer Unila Publishing Network all rights reserved

pengklasifikasian. Dari permasalahan tersebut, maka akan dilakukan penelitian terkait yaitu : (1) berapa tingkat akurasi klasifikasi data tweet gejala COVID menggunakan metode Random Forest dengan pembobotan TF-IDF dan (2) berapa tingkat akurasi klasifikasi data tweet gejala COVID menggunakan metode Random Forest dengan pembobotan Delta TF-IDF.

2. METODOLOGI PENELITIAN

Penelitian ini terdiri dari dua percobaan, pada percobaan pertama data akan di ekstraksi fitur TF-IDF terlebih dahulu selanjutnya dilakukan tahapan splitting data. Alur percobaan 1 ini dapat dilihat pada bagian (a) pada Gambar 1. Namun berbeda pada tahapan klasifikasi dengan pembobotan Delta TF-IDF, di percobaan ini data akan diberikan nilai bobot pada saat proses training sehingga bobot nilai testing didapatkan dari tahapan training yang dilanjutkan klasifikasi data menggunakan Random Forest. Alur percobaan 2 ini dapat dilihat pada bagian (b) pada Gambar 1. Perbedaan terjadi karena Delta TF-IDF adalah teknik berbasis supervised sehingga proses harus dilakukan pada data training saja.



Gambar 1. Alur penelitian.

Adapun tahapan pada metodologi ini yaitu pengumpulan data, preprosesing, ekstraksi fitur, klasifikasi, dan evaluasi yang penjelasannya dapat dilihat di bawah ini.

2.1. Pengumpulan Data

Data yang digunakan adalah data komentar mengenai gejala COVID-19. Data dikumpulkan melalui proses *crawling* dengan API Twitter menggunakan package RTweet [13] pada bahasa pemrograman R dan kemudian dilakukan pelabelan manual sehingga data dibagi menjadi dua kelas yaitu kelas positif dan negatif. Pelabelan merupakan proses membuat data hasil *crawling* ke dalam kelas atau label yang sudah ditentukan yaitu kelas positif dan negatif. Sampel data tersebut dapat dilihat pada Tabel 1.

Tabel 1. Contoh data

Text	Label
GUE POSITIF OMICRON AAAAAA	Positif
Gejala covid pertama yg gue bener2 drop bgtttt tp gue ga isoman smpe akhirnya sembuh sendiri dan ga nular ke bnyk org, skrng covid lagi dan gejalanya cuma flu radng demam mual muntah gue isoman dan nularin ke beberapa org wkwkkwwkkwwkk	Positif
aku udh negatif covid 🙏👍	Negatif
Hah apaan orang gw kemaren pilek + batuk swab bisa negatif kok	Negatif

Data yang berlabel positif dan negatif masing-masing mempunyai karakteristik untuk di kategorikan berdasarkan bergejala atau tidak COVID.

2.2. Preprocessing

Preprocessing merupakan proses pembersihan data sebelum data akan diproses ketahapan selanjutnya agar data yang digunakan lebih terstruktur dan prosesnya menjadi lebih mudah. Pada umumnya untuk tahapan preprocessing terdapat *cleaning*, *case folding*, *stopword removal* dan *tokenizing* [13]. Untuk tahapan preprocessing dilakukan beberapa proses.

Proses pertama adalah *cleaning* merupakan proses membuat mengeliminasi komponen yang tidak sesuai pada data, seperti : tag link html, hastag, mention, simbol, angka, tanda baca yang tidak sesuai dan komponen yang tidak relevan lainnya. Proses kedua adalah *case folding* yaitu melakukan perubahan huruf yang awalnya kapital menjadi huruf kecil semua. Preprocessing selanjutnya adalah *tokenizing* yang merupakan proses membuat unit kata per kata dari membagi setiap dokumen yang ada untuk membangun vektor kata tanpa mempertimbangkan adanya duplikasi kata. Proses yang terakhir adalah *stopword removal* atau disebut juga sebagai filtering yang bertujuan untuk mengeliminasi kata-kata yang tidak penting dari hasil tokenisasi. *Stopword* pada umumnya menggunakan pendekatan *Bag-of-Words* untuk membuang kata yang tidak merupakan deskriptif.

2.3. Ekstraksi Fitur

Ekstraksi fitur merupakan proses mengubah data yang berupa teks yang tidak terstruktur menjadi data yang terstruktur untuk dapat diproses selanjutnya ke tahapan klasifikasi. Penelitian ini melakukan perbandingan ekstraksi fitur jenis pembobotan term (kata). Pembobotan kata dibagi menjadi dua jenis yaitu : pembobotan supervised dan pembobotan unsupervised. Dengan membandingkan dua jenis pembobotan term maka dilakukan perbandingan antara pembobotan TF-IDF dan Delta TF-ID.

TF-IDF merupakan proses pembobotan kata pada *text mining* dengan mencari nilai TF (*Term-Frequency*) yaitu menghitung jumlah kemunculan kata pada dokumen. Sedangkan IDF (*Invers Document Frequency*) yaitu pengurangan jumlah dominasi kata yang muncul pada dokumen sehingga perhitungan yang dilakukan adalah kebalikan dari frekuensi dokumen yang sering muncul. Rumus TF-IDF dapat dilihat pada formula (1).

$$W_{d,t} = TF_{d,t} \times IDF_t = TF \times \log(n/dft) \quad (1)$$

Dimana W adalah bobot term t pada dokumen d; TF adalah jumlah frekuensi term t pada dokumen t; IDF adalah nilai IDF dari term t; Dft adalah jumlah dokumen yang mengandung term t; n adalah jumlah dokumen.

Metode pembobotan Delta TF-IDF merupakan keterbaruan dari metode TF-IDF. Delta TF-IDF meningkatkan pentingnya kata-kata yang tidak merata antara kelas positif dan negatif. Pada metode ini di pembobotan memberikan nilai yang berbeda pada dataset kelas positif dan negated sehingga hasil klasifikasi menjadi lebih relevan. Setelah dilakukan pembobotan maka akan terlihat perbedaan nilai antara term yang positif maupun negatif. Rumus Delta TF-IDF dapat dilihat pada formula (2).

$$\delta tf - idf(t_i) = tf(t_i, d_j) \times \log_2 \left(\frac{N_p \times C + 0.5}{A \times N_n + 0.5} \right) \quad (2)$$

Dimana N_p adalah jumlah dokumen dengan class positif; N_n adalah jumlah dokumen dengan class negatif; A adalah jumlah dokumen dengan class label adalah positif dimana term t_i ditemukan paling tidak sekali; C adalah jumlah dokumen dengan class label adalah negatif dimana term t_i ditemukan paling tidak sekali.

2.4. Klasifikasi

Random Forest merupakan algoritma klasifikasi dari kumpulan pohon keputusan atau biasa disebut Decision Tree. Random Forest dikembangkan dari pohon keputusan untuk peningkatan karena cenderung *overfitting* [14]. Random Forest melakukan prediksi klasifikasi analisis sentimen dengan cara mengkombinasikan hasil dari pohon keputusan. Pada Random Forest, untuk mengetahui kelas dari prediksi didapatkan dari suara mayoritas keseluruhan pohon [15]. Kelebihan algoritma Random Forest ini salah satunya dapat memberikan klasifikasi yang lebih pada data yang berjumlah besar. Parameter yang dapat digunakan untuk Random Forest adalah ntree dan mtry dengan pembentukan model menggunakan package randomForest pada Rstudio [16]. Ntree adalah banyaknya pohon keputusan yang digunakan pada data sample untuk klasifikasi data. Sedangkan mtry apabila sebanyak n berarti bahwa pada proses splitting pembentukan pohon klasifikasinya, hanya sebanyak n variabel secara acak yang dicek atau dibandingkan untuk menentukan pembagian data terbaik.

2.5. Evaluasi

Untuk mengetahui hasil perbandingan dari masing-masing pembobotan maka digunakan confusion matrix berupa akurasi, *sensitivity* dan *specificity*. Confusion matrix adalah sebuah tabel yang digunakan untuk mengukur kinerja model klasifikasi Random Forest untuk data testing yang nilai sebenarnya diketahui. Pada penelitian digunakan package caret untuk menggunakan *confusion matrix* pada Rstudio [14]. Dengan nilai seperti true positive, true negative, false positive dan false negative akan ditampilkan kedalam tabel matrix untuk mengetahui seberapa banyak proporsi yang diprediksi model benar dan salah, sehingga dapat dihitung akurasi, *sensitivity* dan *specificity* setelahnya. Confusion matrix dapat dilihat pada Tabel 2.

Tabel 2. Confusion matrix

Actual Class	Predicted Class	
	Class	
Positive	True Poitive (TP)	False Negative (FN)
Negative	False Negative (FP)	True Negatif (TN)

Dimana:

True Positive (TP) : Kelas yang digolongkan sebagai positif dan faktanya benar adalah positif;

True Negative (TN) : Kelas yang digolongkan sebagai negatif dan faktanya benar adalah negatif;

False Positive (FP) : Kelas yang digolongkan sebagai positif dan faktanya sebagai negatif;

False Negative (FN) : Kelas yang digolongkan sebagai negatif dan faktanya sebagai positif.

Setelah nilai *confusion matrix* didapatkan dari hasil model klasifikasi maka dapat dilakukan perhitungan evaluasi untuk mengukur kinerja hasilnya. Adapun Rumus dari akurasi, *sensitivity*, dan *specificity* menggunakan *confusion matrix* adalah sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (3)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% \quad (5)$$

Metode berisi uraian tentang cara menjalankan penelitian. Pada bagian ini ditulis cara memperoleh data, algoritma atau rumusan yang digunakan dalam penelitian atau cara mengolah data, dan cara mengevaluasi atau menilai hasil penelitian. Metode yang sudah umum tidak perlu dituliskan secara detail, tetapi cukup merujuk ke buku acuan. Prosedur penelitian harus dituliskan dalam bentuk kalimat berita, bukan kalimat perintah.

3. HASIL DAN PEMBAHASAN

Setelah dilakukan pengumpulan data, maka akan dibersihkan terlebih dahulu pada tahapan preprocessing kemudian akan di ekstraksi fitur menggunakan pembobotan kata TF-IDF dan Delta TF-IDF, selanjutnya akan diklasifikasikan menggunakan metode klasifikasi Random Forest dan juga evaluasi hasil.

3.1. Hasil

Penelitian pertama dilakukan dengan membuat model klasifikasi dengan menggunakan algoritma Random Forest dan ekstraksi fitur dengan pembobotan TF-IDF. Pada model ini, dilakukan proses ekstraksi fitur pada dataset menggunakan metode TF-IDF. Untuk Rumus perhitungan TF-IDF yang digunakan yaitu Rumus (1) sebagai pembobotan setiap fitur kata dengan total 2262 fitur pada penelitian ini. Dari hasil pembobotan dapat dilihat pada Tabel 3 dengan bobot numerik di setiap fiturnya.

Tabel 3. Hasil ekstraksi fitur dengan pembobotan TF-IDF

No	aaa	aaaaa	...	zezehahaha	zombie	Label
1	0	0	...	0	0	Positif
2	0	0	...	0	0	Positif
...
999	0.8380821	0	...	0	0	Negatif
1000	0	0	...	0	0	Negatif

Dari hasil pembobotan TF-IDF pada didapatkan 2262 fitur kata dengan bobot numerik yang selanjutnya akan dilanjutkan ke tahapan klasifikasi. Hasil tahap klasifikasi dapat dilihat pada *confusion matrix* pada Tabel 4.

Tabel 4. Confusion matrix dari model klasifikasi penelitian 1

Confusion Matrix		Aktual	
		Negatif	Positif
Predeksi	Negatif	410	31
	Positif	90	469

Penelitian kedua adalah pembuatan model klasifikasi dengan Random Forest dengan menggunakan data dari hasil ekstraksi fitur pada dataset menggunakan metode Delta TF-IDF. Setelah data dilakukan pembobotan maka akan menghasilkan 2262 fitur dengan bobot numerik untuk diproses selanjutnya ke tahapan klasifikasi. Rumus yang digunakan dalam pembobotan ini yaitu rumus (2). Adapun untuk contoh data yang sudah dibobotkan dapat dilihat pada Tabel 5.

©2023 Ilmu Komputer Unila Publishing Network all rights reserved
 Tabel 5. Hasil ekstraksi fitur dengan pembobotan Delta TF-IDF

No	aaa	aaaaa	...	zezehahaha	zombie	Label
1	0	0	...	0	0	Positif
2	0	0	...	0	0	Positif
...
999	11.55	0	...	0	0	Negatif
1000	0	0	...	0	0	Negatif

Berdasarkan Tabel 5 didapatkan bobot masing-masing fitur kata melalui tahapan pembobotan Delta TF-IDF untuk dapat dilanjutkan pada tahapan klasifikasi. Hasil kinerja klasifikasi model ini dapat dilihat pada Tabel 6.

Tabel 6. Confusion matrix dari model klasifikasi penelitian 2

Confusion Matrix		Aktual	
		Negatif	Positif
Predeksi	Negatif	425	40
	Positif	75	460

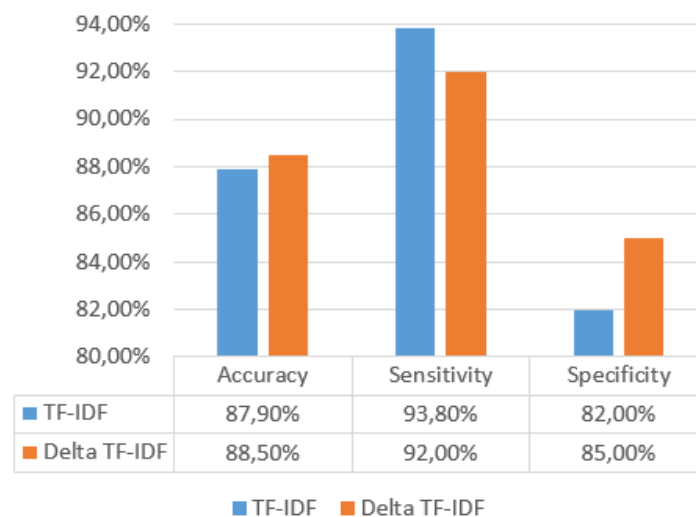
Dengan menghitung confusion matrix dari Tabel 4 dan Tabel 6 didapat kinerja klasifikasi yang ditunjukkan pada Tabel 7.

Tabel 7. Kinerja klasifikasi

Kinerja Klasifikasi	TF-IDF	Delta TF-IDF
Akurasi	87,900%	88,500%
Sensitivity	93,800%	92,000%
Specificity	82,000%	85,000%

3.2. Diskusi

Perbandingan kinerja klasifikasi dari kedua penelitian dapat dilihat pada Gambar 2. Dari perbandingan akurasi dapat dilihat bahwa data terstruktur hasil ekstraksi fitur unigram dengan pembobotan Delta TF-IDF lebih baik dibandingkan teknik pembobotan unsupervised TF-IDF. Pada penelitian ini juga dapat dilihat kinerja klasifikasi lain seperti sensitivity dan specificity dengan pembobotan Delta TF-IDF lebih tinggi dibandingkan TF-IDF.



Gambar 2. Perbandingan kinerja klasifikasi

©2023 Ilmu Komputer Unila Publishing Network all rights reserved

Berdasarkan hasil sensitivity atau berdasarkan informasi dari perbandingan *confusion matrix* dapat dilihat bahwa pembobotan Delta TF-IDF dapat meningkatkan ketepatan klasifikasi kelas negatif namun menurunkan kinerja klasifikasi kelas positif. Peningkatan klasifikasi kelas negative dapat dilihat pada **Error! Reference source not found.** dimana nilai True Negative adalah 410 dan terjadi kenaikan nilai True Negative menjadi 425 seperti yang ditunjukkan di **Error! Reference source not found.** Penurunan klasifikasi kelas positive dapat dilihat pada nilai True Positive pada **Error! Reference source not found.** adalah 469 sedangkan nilai True Positive pada **Error! Reference source not found.** adalah 460.

Walaupun kinerja klasifikasi kelas positif menurun saat menggunakan teknik ekstraksi fitur unigram dengan pembobotan supervised, namun secara umum kedua teknik pembobotan ini bekerja lebih baik mengklasifikasikan kelas positif dibandingkan kelas negatif.

4. KESIMPULAN

Dari hasil penelitian yang dilakukan didapatkan akurasi klasifikasi Random Forest dengan TF-IDF didapatkan nilai sebesar 87,9% dan akurasi klasifikasi Random Forest dengan Delta TF-IDF sebesar 88,5%. Dari perbandingan hasil akurasi menunjukkan bahwa pengaruh ekstraksi fitur Delta TF-IDF lebih baik dibandingkan ekstraksi fitur TF-IDF pada model Random Forest untuk mengklasifikasikan data tweet gejala COVID-19. Berdasarkan hasil percobaan dengan pembobotan Delta TF-IDF prediksi pada data negatif menjadi naik sedangkan prediksi pada data positif sedikit menurunkan.

Pada penelitian selanjutnya akan dilakukan percobaan untuk menguji apakah metode supervised lainnya dapat meningkatkan prediksi pada data positif. Selain itu akan dilakukan klasifikasi menggunakan deep learning untuk meningkatkan kinerja deteksi laporan penderita COVID-19 dari Twitter.

UCAPAN TERIMA KASIH

Penelitian ini adalah hasil dari tugas akhir di Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Penelitian ini dibiayai dari Program Dosen Wajib Meneliti (PDWM) dari hibah PNPB Universitas Lambung Mangkurat.

DAFTAR PUSTAKA

- [1] N. U. L. Hafidhoh and A. N. Sissandhy, "Klasifikasi Fasilitas Umum di Jawa Tengah pada Twitter dengan Algoritma Agglomerative Hierarchical Clustering dan Naive Bayes Classifier," vol. 4, no. 2, pp. 60–68, 2021.
- [2] Nuryani and D. Mahayana, "Analisis Sentimen Berbasis Aspek dengan Deep Learning Ditinjau dari Sudut Pandang Filsafat Ilmu," *Jumanji*, vol. 4, no. 2, pp. 70–85, 2021.
- [3] M. Rusli, M. R. Faisal, and I. Budiman, "Ekstraksi Fitur Menggunakan Model Word2vec untuk Analisis Sentimen pada Komentar Facebook," *Soliter*, vol. 2, pp. 104–109, 2019.
- [4] M. A. Al-garadi, Y. Yang, S. Lakamana, and A. Sarker, "A Text Classification Approach for the Automatic Detection of Twitter Posts Containing Self-reported COVID-19 Symptoms," *Open Rev.*, vol. 003, pp. 1–5, 2020.
- [5] F. Carvalho and G. P. Guedes, "TF-IDFC-RF: A Novel Supervised Term Weighting Scheme," pp. 1–28, 2020.
- [6] M. Alkaff, A. Rizky Baskara, and Y. Hendro Wicaksono, "Sentiment Analysis of Indonesian Movie Trailer on YouTube Using Delta TF-IDF and SVM," *2020 5th Int. Conf. Informatics Comput. ICIC 2020*, 2020, doi: 10.1109/ICIC50835.2020.9288579.

©2023 Ilmu Komputer Unila Publishing Network all rights reserved

- [7] C. E. Joergensen Munthe, N. Astuti Hasibuan, and H. Hutabarat, “RESOLUSI: Rekayasa Teknik Informatika dan Informasi Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace,” *Media Online*, vol. 2, no. 3, pp. 110–115, 2022.
- [8] M. Alfian Rosid, E. Pramana, U. Muhammadiyah Sidoarjo, and S. Tinggi Teknik Surabaya, “Centroid Based Classifier Dengan Fitur Tf-Idf-Icf Untuk Klasifikasi Keluhan Mahasiswa Pada Aplikasi E-Complaint Di Universitas Muhammadiyah Sidoarjo,” vol. 1, no. 1, 2015.
- [9] R. R. M. Putri, R. Y. Herlambang, and R. C. Wihandika, “Implementasi Metode K-Nearest Neighbour Dengan Pembobotan TF.IDF.ICF Untuk Kategorisasi Ide Kreatif Pada Perusahaan,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 2, p. 97, 2017, doi: 10.25126/jtiik.201742296.
- [10] I. Istiqlaliyyah, Y. A. Sari, and M. A. Fauzi, “Pencarian Teks pada Terjemahan Ayat Al- Qur ’ an dengan Menggunakan TF-RF dan Bray-Curtis Distance,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 5, pp. 4357–4363, 2019.
- [11] R. Ramadhan, Y. A. Sari, and P. P. Adikara, “Perbandingan Pembobotan Term Frequency-Inverse Document Frequency dan Term Frequency-Relevance Frequency terhadap Fitur N-Gram pada Analisis Sentimen,” vol. 5, no. 11, pp. 5075–5079, 2021.
- [12] M. M. Alvanof and R. Triandi, “Analisa Dan Deteksi Konten Hoax Pada Media Berita Alvanof, Mulia Mahendra, and Rahmat Triandi. 2020. ‘Analisa Dan Deteksi Konten Hoax Pada Media Berita.’ Jurnal Teknologi Terapan & Sains 4.0 Universitas Malikussaleh 1:2.,” *J. Teknol. Terap. Sains 4.0 Univ. Malikussaleh*, vol. 1, p. 2, 2020.
- [13] M. W. Kearney, A. Heiss, and F. Briatte, “Package ‘rtweet,’” URL <https://cran.r-project.org/web/packages/rtweet/rtweet.pdf> [accessed 2019-03-19], 2020.
- [14] M. Kuhn, “caret: Classification and Regression Training.” 2021.