

## Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan *Random Forest Classifier*

<sup>1</sup>Debby Alita, <sup>2</sup>Auliya Rahman

<sup>1,2</sup> Informatika, Fakultas Teknik dan Ilmu Komputer Universitas Teknokrat Indonesia  
Jln. H. Zainal Abidin Pagaram 9-11 Labuhanratu  
Telp (0721)707059  
e-mail : [debbyalita@teknokrat.ac.id](mailto:debbyalita@teknokrat.ac.id), Rahman.isnain@gmail.com

---

**Abstract** — Kalimat sindiran atau sarkasme masih sering digunakan oleh kalangan publik untuk mengungkapkan maksud isi hati dan pikiran baik itu yang disampaikan secara langsung maupun tidak langsung. Sarkasme dilakukan untuk menyindir dan menyakiti hati seseorang dengan menggunakan bahasa atau kata yang didalamnya mengandung kata positif tetapi maknanya negatif sehingga sering sekali terjadi opini salah diklasifikasikan. Penelitian ini melakukan kombinasi antara proses sentimen analisis dengan deteksi sarkasme untuk pengklasifikasian opini yang terdapat pada *Twitter*. Proses analisis sentimen dilakukan dengan tahapan *preprocessing* dan ekstraksi fitur dan diklasifikasikan dengan menggunakan metode *Support Vector Machine* dilanjutkan dengan proses pendeteksian sarkasme yang dilakukan tahapan ekstraksi fitur dengan 4 set fitur yaitu *sentiment related*, *punctuation-relate*, *lexical and syntactic*, dan *pattern-relate* dan diklasifikasikan dengan menggunakan metode *Random Forest Classifier*. Hasil penelitian ini didapatkan peningkatan nilai rata-rata akurasi sebesar 16,61 %, nilai presisi sebesar 5,45 %, nilai recall sebesar 9,64% dan kenaikan nilai F1score sebesar 11,27% dengan jumlah data sebanyak 2.027 dengan rincian data dengan label positif berjumlah 1023, data dengan label negatif berjumlah 587 dan data dengan label netral berjumlah 462. Data sarkasme didapatkan dari tweet dengan label positif yang kemudian diberikan label sarkasme atau tidak sarkasme dan didapat hasil label dengan jumlah keseluruhan berlabel sarkasme berjumlah 354 dan tidak sarkasme berjumlah 669.

**Keywords:** analisis sentimen, deteksi sarkasme, *Support Vector Machine*, *Random Forest Classifier*

---

### 1. PENDAHULUAN

Gaya bahasa memegang peran penting dalam mengartikan maksud dan tujuan dari penyampaian seseorang melalui suatu opini baik itu yang disampaikan secara langsung maupun tidak langsung. Menganalisis opini seseorang bukanlah hal yang mudah karena banyak sekali kalimat sindiran yang disampaikan secara langsung tanpa basa-basi, sedangkan yang lain menggunakan metode tidak langsung dalam menyinggung poin yang hendak disampaikan untuk berbagai tujuan [1]. Opini yang mengandung sindiran yang pahit dan celaan yang getir disebut juga dengan opini sarkasme. Sarkasme merupakan gaya bahasa penyindiran dengan menggunakan kata-kata yang akan menyakiti hati seseorang [2]. Salah satu jenis sarkasme yang digunakan adalah jenis sarkasme yang berlainan makna yaitu kalimat sarkasme berbentuk kalimat positif yang memiliki makna negatif. Opini sarkasme ini sering sekali disampaikan seseorang secara tidak langsung melalui media sosial.

Penggunaan media sosial terutama twitter hingga saat ini terus mengalami peningkatan pengguna yang signifikan dan setiap orang dapat mengakses informasi secara cepat melalui twitter dengan mencari hashtag yang sedang menjadi trending topic [3]. Analisis data menggunakan media sosial Twitter sudah menjadi hal yang sangat diminati oleh para peneliti *text mining* karena twitter memberikan kemudahan kepada para programmer untuk secara bebas dalam mendapatkan data dan ada beberapa alasan lain mengapa twitter unggul, yaitu mudah diadaptasi dan digunakan, piranti penelitian yang handal dan mesin pencari yang realtime serta analisis data yang cepat. Salah satu bentuk analisis data yang paling sering dilakukan adalah analisis sentimen [4].

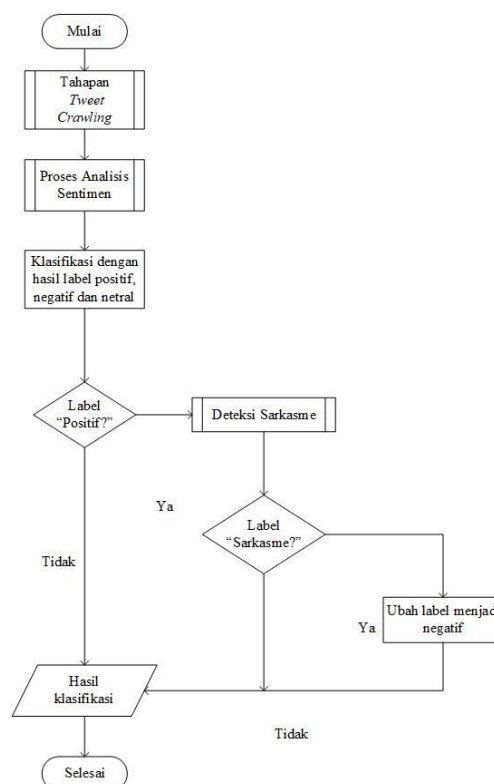
Analisis sentimen adalah proses memahami dan mengolah data tekstual secara otomatis untuk mendapatkan informasi [5]. Analisis sentimen dilakukan untuk mendeteksi opini terhadap suatu subjek dan objek (misalnya individu, organisasi ataupun produk) dalam sebuah kumpulan data [6]. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian ataupun aplikasi mengenai analisis sentimen berkembang pesat. Negara Amerika kurang lebih 20-30 perusahaan memfokuskan layanannya pada analisis sentimen [7]. Analisis sentimen merupakan proses klasifikasi dan faktanya tidak semudah proses klasifikasi biasa karena penggunaan Bahasa yang mengandung kata ambigu, tidak adanya intonasi dalam sebuah teks, dan perkembangan dari bahasa itu sendiri.

Pada penelitian ini akan dilakukan klasifikasi data teks berbahasa Indonesia di twitter dengan melakukan pendeteksian sarkasme pada opini publik di twitter. Setelah dilakukan pengklasifikasian terhadap tweet, tweet yang bernilai positif akan dilakukan proses pendeteksian sarkasme dengan 4 fitur yaitu *sentiment related*, *punctuation-relate*, *lexical and syntactic*, dan *pattern-relate* dengan menggunakan metode *Random Forest Classifier*.

## 2. METODOLOGI PENELITIAN

### 2.1 Alur Penelitian

Alur penelitian merupakan gambaran dari tahapan pekerjaan dari awal penelitian itu dimulai hingga berakhir. Alur penelitian ini digambarkan dengan suatu bagan yang saling berurutan yang disebut juga dengan bagan alir dokumen. Proses penelitian dimulai dari awal pengambilan data kemudian tahapan preprocessing data pada proses analisis sentiment, ekstraksi fitur dan klasifikasi, kemudian dilanjutkan dengan proses pendeteksian sarkasme dengan dilakukan tahapan preprocessing data berlabel positif dengan 4 fitur dan terakhir proses klasifikasi sarkasme dan pengujian. Secara garis besar langkah-langkah yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian Pendeteksian Sarkasme

## 2.2. Analisis Data

Data didapatkan dari proses *tweet crawling* dengan memanfaatkan API yang disediakan oleh *twitter*. Data yang digunakan pada proses penelitian ini berjumlah 2027 data tentang pelayanan publik yaitu mengenai jaringan telekomunikasi seluler, layanan BPJS dan layanan PLN. Proses *scraping* data dilakukan dengan memasukkan data nama grup, kata kunci, tandai, tanda pagar, tanggal dan tahun.

## 2.3 Analisis Sentimen

Analisis sentimen atau bisa disebut juga *opinion mining* merupakan sebuah cabang penelitian di domain text mining yang mulai banyak dilakukan pada tahun 2013. Analisis sentimen adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual. Jika diberikan satu set dokumen teks yang berisi opini (atau sentimen) mengenai suatu objek, maka *opinion mining* bertujuan untuk mengekstrak atribut dan komponen dari objek yang telah dikomentari pada setiap dokumen dan untuk menentukan apakah komentar tersebut positif, negatif atau netral [8]. [9] juga menyebutkan bahwa *opinion mining* adalah bagian pekerjaan yang melakukan *review* dan berkaitan dengan perlakuan opini, sentimen, dan subjektivitas dari sebuah teks.

Proses mempersiapkan teks dokumen atau dataset mentah disebut juga dengan proses *text preprocessing*. *Text preprocessing* berfungsi untuk mengubah data teks yang tidak terstruktur atau sembarang menjadi data yang terstruktur [10]. Pada penelitian ini proses yang dilakukan dalam tahapan *preprocessing* adalah sebagai berikut :

### 1. Case Folding

*Case Folding* adalah proses penyamaan case dalam sebuah dokumen. Hal ini dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (dalam hal ini huruf kecil atau *lowercase*).

### 2. Tokenizing

*Tokenizing* adalah proses pemotongan sebuah dokumen menjadi bagian-bagian, yang disebut dengan token. Pada saat bersamaan *tokenizing* juga berfungsi untuk membuang beberapa karakter tertentu yang dianggap sebagai tanda baca.

### 3. Stopword Removal

*Stopword removal* adalah proses penghilangan kata-kata yang tidak berkontribusi banyak pada isi dokumen (Yates dan Neto, 1999). Kata-kata yang termasuk ke dalam *stopword* dihilangkan karena memberikan pengaruh yang tidak baik dalam proses text mining seperti kata-kata "and", "I", "you", "with", "she", "he", dan lain-lain.

### 4. Convert Emoticon

*Convert Emoticon* adalah proses konversi tweet atau kalimat yang mengandung *emoticon* ke dalam string yang bersesuaian. Jenis *emoticon* yang akan diproses adalah *emoticon* western style, dikarenakan jenis tersebut banyak digunakan atau menjadi standar.

### 5. Slangword

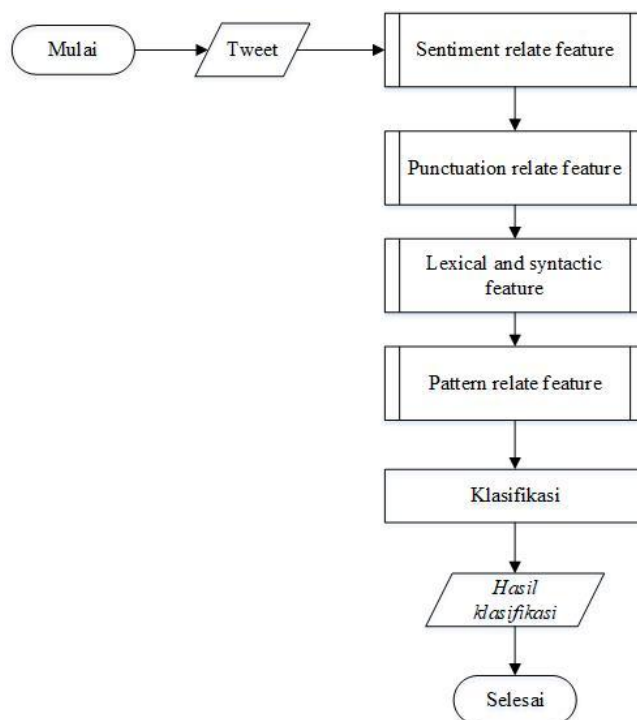
*Slangword* merupakan perubahan kata berbahasa Indonesia yang tidak baku di tweet dan tidak ada dalam kamus, kata tersebut akan ditranslasikan ke kata terdekat dengan menggunakan kamus yang dibuat dengan melihat pola kemunculan kata-kata baku.

Selanjutnya dilakukan proses *Feature extraction* atau ekstraksi fitur yang merupakan proses merubah sebuah data di proyeksikan ke dalam fitur baru dengan ruang dimensi yang lebih rendah [11]. Terdapat empat fitur yang akan digunakan dalam penelitian ini adalah Unigram, POS *Tagging*, TF dan TF-IDF. Terakhir adalah melakukan tahapan klasifikasi pada proses analisis sentimen dengan menggunakan metode *Random Forest Classifier*

## 2.4 Deteksi Sarkasme

Pendeteksian sarkasme dilakukan pada *tweet* yang diklasifikasikan positif pada analisis sentiment [12]. Fitur yang akan diekstraksi pada pendeteksian sarkasme terdiri dari fitur *sentiment-relate*, *punctuation-relate*, *lexical* dan *syntactic*, dan *pattern-relate*. *Tweet* yang akan diekstraksi fitur pendeteksian sarkasmenya

merupakan *tweet* awal yang sama seperti *tweet* awal pada analisis sentimen. Tahapan pendeteksian sarkasme dapat dilihat pada gambar 2.



Gambar 2. Tahapan Pendeteksian Sarkasme

#### 2.4.1. Sentiment-Relate Feature

Pada set fitur *sentiment-relate*, terdapat 10 fitur yang akan diekstraksi.

- 1) Sentimen Kontras  
Menentukan jumlah kata berupa kata opini yang memiliki kecenderungan bersifat positif atau negatif yang didefinisikan oleh [13]
- 2) Menentukan jumlah *emoticon* positif pada satu kalimat *tweet*.
- 3) Menentukan jumlah *emoticon* negatif pada satu kalimat *tweet*.
- 4) Menentukan jumlah *emoticon* sarkasme pada satu kalimat *tweet*.
- 5) Menentukan jumlah hashtag positif pada *tweet*
- 6) Menentukan jumlah hashtag negatif pada *tweet*  
Keterangan untuk fitur kelima dan keenam yaitu jumlah hashtag positif dan negatif dihitung dan dijadikan fitur. Hashtag positif didapat dari data *tweet* dengan label positif dan hashtag negatif didapat dari data *tweet* dengan label negatif.
- 7) Menentukan nilai kontras antar komponen kata positif dengan kata negatif dalam satu *tweet*
- 8) Menentukan nilai kontras antar komponen hashtag positif dengan hashtag negatif dalam satu *tweet*
- 9) Menentukan nilai kontras antar komponen kata positif dengan hashtag negatif atau sebaliknya dalam satu *tweet*
- 10) Menentukan nilai kontras antar komponen kata positif dengan *emoticon* negatif atau sebaliknya dalam satu *tweet*

#### 2.4.2. Punctuation-Relate Feature

Pada set fitur *punctuation-relate*, terdapat 6 fitur yang akan diekstraksi.

- 1) Menentukan jumlah tanda seru dengan ketentuan jika hanya ada satu tanda seru maka tidak dihitung.
- 2) Menentukan jumlah tanda tanya.
- 3) Menentukan jumlah titik.
- 4) Menentukan jumlah kata yang semua hurufnya capital.
- 5) Menentukan jumlah kutipan kata dengan ciri satu kutip atau dua kutip.

- 6) Menentukan jumlah pengulangan huruf yang sama lebih dari 2 kali pada suatu kata.

#### 2.4.3. Lexical and Syntactic Feature

Pada set fitur *lexical and syntactic*, terdapat 4 fitur yang akan diekstraksi, yaitu:

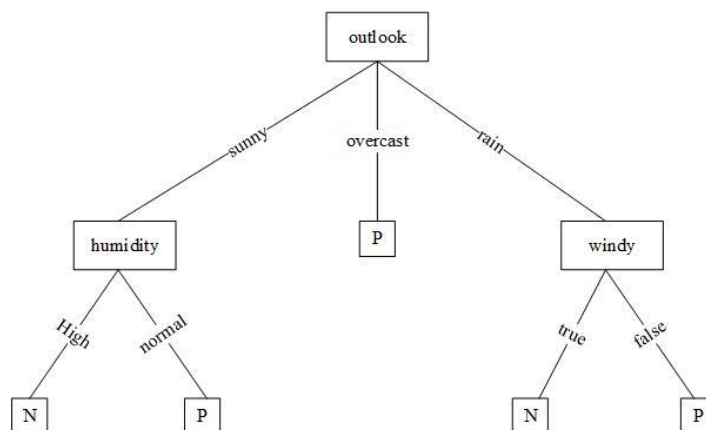
- 1) Menentukan jumlah tawa
- 2) Menentukan jumlah kata seru
- 3) Keberadaan kata yang jarang digunakan
- 4) Keberadaan pola umum kalimat sarkasme

#### 2.4.4. Pattern-Relate Feature

Cara untuk mengekstrak set fitur *pattern-relate* sama seperti pada pola umum kalimat sarkasme dan selanjutnya membuat daftar pola sarkasme serta mengelompokkan jumlah pola sarkasme berdasarkan panjang pola sarkasme.

## 2.5. Klasifikasi

*Random forest classifier* merupakan metode klasifikasi yang terdiri dari kumpulan pohon keputusan yang nantinya akan dijadikan vote untuk mendapatkan hasil terakhir dari pendeteksian sarkasme dengan pendukung berupa data latih dan fitur acak yang independen dengan fitur yang berbeda-beda [14]. Pohon keputusan dibuat dengan menentukan node akar dan berakhir dengan beberapa node daun untuk mendapatkan hasil akhir. Berikut merupakan ilustrasi dari pohon keputusan yang dapat dilihat pada Gambar 3.



Gambar 3. Contoh Pohon Keputusan

Selanjutnya setelah proses klasifikasi dilakukan proses pengujian untuk mengukur performa dari metode klasifikasi yang digunakan menggunakan k-fold cross validation dengan mengukur nilai *accuracy*, *precision*, *recall*, dan *F1-Score* [15]. Pada Gambar 4 menggambarkan contoh untuk prediksi dua kelas dengan menggunakan *confusion matrix*.

		<i>Predicted Class</i>	
		<i>Class 1</i>	<i>Class 2</i>
<i>Actual Class</i>	<i>Class 1</i>	<i>True Positive</i>	<i>False Negative</i>
	<i>Class 2</i>	<i>False Positive</i>	<i>True Negative</i>

Gambar 4. *Confusion Matrix*

### 3. HASIL DAN PEMBAHASAN

Pengujian dilakukan dengan mengukur akurasi dari hasil sentimen mengenai opini publik yang dilakukan oleh sistem. Berdasarkan hasil pengujian yang dilakukan maka dapat diketahui parameter mana yang menghasilkan akurasi terbaik.

Total data yang dihasilkan melalui proses scraping adalah sebanyak 2.281 data yang kemudian melalui proses preprocessing jumlah data menjadi 2.072 dengan rincian data dengan label positif berjumlah 1023, data dengan label negatif berjumlah 587 dan data dengan label netral berjumlah 462. Data tersebut terdiri dari data telekomunikasi data seluler, data PLN, BPJS dan data politik berupa data tentang Presiden Jokowi dan Ahok. Distribusi data untuk sentimenya dapat dilihat pada Tabel 1.

Tabel 1. Distribusi data analisis sentimen

Keyword	Jumlah Data		
	Positif	Negatif	Netral
Indosat	428	212	62
Telkomsel	39	55	74
XL	58	32	70
PLN	93	41	92
BPJS	183	39	40
Jokowi	122	80	50
Ahok	100	128	74

Data dengan polaritas positif pada data asli kemudian akan dilakukan pelabelan untuk data sarkasme dan tidak sarkasme. Distribusi untuk data sarkasme dapat dilihat pada Tabel 2

Tabel 2. Distribusi data sarkasme

Keyword	Jumlah Data	
	Sarkasme	Tidak Sarkasme
Indosat	50	378
Telkomsel	13	26
XL	31	27
PLN	57	36
BPJS	52	131
Jokowi	69	53
Ahok	82	18

Data sarkasme didapatkan dari tweet dengan label positif yang kemudian diberikan label sarkasme atau tidak sarkasme dan didapat hasil label dengan jumlah keseluruhan berlabel sarkasme berjumlah 354 dan tidak sarkasme berjumlah 669.

#### 3.1. Hasil Pengujian Analisis Sentimen

Hasil pengujian analisis sentimen dengan menggunakan *support vector machine* memberikan nilai akurasi tertinggi pada pengujian 2-Fold cross validation dan nilai presisi tertinggi pada pengujian 10-Fold cross validation, sedangkan untuk nilai recall dan F1score terbaik dilakukan melalui pengujian 9-Fold cross validation dengan nilai akurasi sebesar 62,95 %, nilai presisi sebesar 67,84%, nilai recall 52,89% dan nilai F1Score 50,84%.

Nilai rata-rata yang didapatkan dari pengujian 2-Fold cross validation hingga 10-Fold cross validation adalah nilai akurasi sebesar 62,25 %, nilai presisi sebesar 54,94 %, nilai recall sebesar 49,26 % dan nilai F1score sebesar 47,44 %. Hasil pengujian dapat dilihat pada Tabel 3

Tabel 3. Hasil Pengujian Analisis Sentimen

<b>Fold Cross Validation</b>	<b>Akurasi</b>	<b>Presisi</b>	<b>Recall</b>	<b>F1score</b>
2	62,95 %	42,35 %	37,26 %	36,33 %
3	61,80 %	51,37 %	47,36 %	45,53 %
4	61,92 %	52,56 %	48,05 %	46,32 %
5	62,40 %	55,66 %	50,85 %	49,14 %
6	61,92 %	54,83 %	50,18 %	48,32 %
7	62,09 %	55,25 %	51,79 %	49,39 %
8	62,28 %	57,27 %	52,32 %	50,37 %
9	62,34 %	57,38 %	52,89 %	50,84 %
10	62,59 %	67,84 %	52,65 %	50,70 %
<b>Rata-rata</b>	62,25 %	54,94 %	49,26 %	47,44 %

### 3.2 Hasil Pengujian Pendeteksian Sarkasme

Pada pengujian deteksi sarkasme menggunakan metode *random forest classifier* dan diekstraksi dengan 4 fitur yaitu sentiment relate feature, punctuation relate feature, lexical and syntactic feature, dan pattern relate feature diperoleh nilai ukuran kualitas dari pengujian deteksi sarkasme yang dapat dilihat pada Tabel 4.

Tabel 4. Hasil pengujian deteksi sarkasme

<b>Fold cross validation</b>	<b>Akurasi</b>	<b>Presisi</b>	<b>Recall</b>	<b>F1score</b>
2	59,50 %	56,95 %	58,22 %	55,95 %
3	62,55 %	59,39 %	61,09 %	58,80 %
4	61,19 %	58,72 %	60,59 %	57,90 %
5	60,10 %	58,54 %	60,36 %	57,24 %
6	60,10 %	58,50 %	60,28 %	57,12 %
7	61,30 %	59,53 %	61,56 %	58,40 %
8	60,21 %	58,80 %	60,71 %	57,25 %
9	59,44 %	58,09 %	59,92 %	56,56 %
10	61,14 %	58,69 %	60,60 %	57,75 %
<b>Rata-rata</b>	60,61 %	58,58 %	60,37 %	57,44 %

Dari Tabel 4 dapat dilihat bahwa hasil pengujian deteksi sarkasme dengan metode random forest classifier memberikan nilai akurasi tertinggi dengan pengujian 3-Fold cross validation dengan nilai sebesar 62,55%. Nilai presisi dan recall tertinggi didapat dan melalui pengujian 7-Fold cross validation dengan nilai presisi sebesar 59,53% dan nilai recall sebesar 61,56% dan nilai F1score tertinggi melalui pengujian 3fold cross validation dengan nilai 58,80%.

Nilai rata-rata yang didapatkan dari pengujian 2-Fold cross validation hingga 10-Fold cross validation adalah nilai akurasi sebesar 60,61%, nilai presisi sebesar 58,58%, nilai recall sebesar 60,37% dan nilai F1score sebesar 57, 44%.



### 3.3 Hasil Pengujian Analisis Sentimen dengan Pendeteksian Sarkasme

Pada pengujian analisis sentimen dengan deteksi sarkasme digunakan tweet yang telah dilabeli positif, negatif, netral, kemudian untuk kalimat yang berlabel positif akan dilakukan proses label sarkasme dan tidak sarkasme yang kemudian dilakukan proses klasifikasi dan diperoleh nilai ukuran kualitas dari pengujian analisis sentimen dengan pendeteksian sarkasme yang dapat dilihat pada Tabel 5

Tabel 5 Hasil pengujian analisis sentimen dengan pendeteksian sarkasme

<b>Fold Cross Validation</b>	<b>Akurasi</b>	<b>Presisi</b>	<b>Recall</b>	<b>F1score</b>
2	75,82 %	52,44 %	50,65 %	50,52 %
3	77,79 %	58,84 %	56,90 %	56,90 %
4	77,49 %	58,52 %	56,89 %	56,69 %
5	77,01 %	59,68 %	58,80 %	58,59 %
6	77,61 %	62,10 %	60,20 %	60,17 %
7	77,26 %	62,23 %	60,96 %	60,71 %
8	77,20 %	63,19 %	61,83 %	61,45 %
9	77,19 %	62,56 %	61,42 %	61,09 %
10	77,67 %	64,01 %	62,45 %	62,32 %
<b>Rata-rata</b>	77,22 %	60,39 %	58,90 %	58,71 %

Pada tabel 5 hasil pengujian analisis sentimen dikombinasikan dengan pendeteksian sarkasme memberikan nilai akurasi tertinggi pada pengujian 3-Fold cross validation dan nilai presisi, recall dan F1score terbaik dilakukan melalui 10-Fold cross validation dengan nilai akurasi sebesar 77,79 %, nilai presisi sebesar 64,01%, nilai recall 62,45% dan nilai F1Score 62,32%. Nilai rata-rata yang didapatkan dari pengujian 2-Fold cross validation hingga 10-Fold cross validation adalah nilai akurasi sebesar 77,22 %, nilai presisi sebesar 60,39%, nilai recall sebesar 58,90 % dan nilai F1score sebesar 58,71 %.

## 4. KESIMPULAN

Proses analisis sentimen dengan pendeteksian sarkasme berhasil dikombinasikan pada data Twitter tentang jaringan telekomunikasi seluler, layanan perusahaan dan politik. Dengan menggunakan pendeteksian sarkasme pada analisis sentimen dapat meningkatkan nilai rata-rata dari akurasi, presisi, recall dan f1score walaupun ada satu nilai yang terjadi penurunan tetapi tidak mempengaruhi nilai rata-rata. Kenaikan nilai rata-rata akurasi sebesar 16,61 %, kenaikan nilai presisi sebesar 5,45 %, kenaikan nilai recall sebesar 9,64% dan kenaikan nilai F1score sebesar 11,27%.

## UCAPAN TERIMA KASIH

Tim Peneliti mengucapkan terimakasih yang sebesar-besarnya pada Universitas Teknokrat Indonesia yang telah memberikan dukungan dan pembiayaan kepada kami sehingga terselenggaranya penelitian ini.

## DAFTAR PUSTAKA

- [1] F. Prawira, "THE EFFECT OF SARCASM DETECTION ON SENTIMENT ANALYSIS QUALITY MEASUREMENT IN TWITTER THE EFFECT OF SARCASM DETECTION ON SENTIMENT," 2017.



- [2] E. Inderasari, F. Achسانی, and B. Lestari, "BAHASA SARKASME NETIZEN DALAM KOMENTAR AKUN INSTAGRAM ' LAMBE TURAH ,'" vol. 8, no. 1, 2019, doi: 10.22460/semantik.vXiX.XXX.
- [3] D. Alita, S. Priyanta, and N. Rokhman, "Analysis of Emoticon and Sarcasm Effect on Sentiment Analysis of Indonesian Language on Twitter," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 5, no. 2, p. 100, 2019, doi: 10.20473/jisebi.5.2.100-109.
- [4] S. Styawati and K. Mustofa, "A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 3, p. 219, 2019, doi: 10.22146/ijccs.41302.
- [5] R. R. SURYONO and I. BUDI, "P2P Lending Sentiment Analysis in Indonesian Online News," vol. 172, no. Siconian 2019, 2020, doi: 10.2991/aisr.k.200424.006.
- [6] A. R. Isnain, A. Sihabuddin, and Y. Suyanto, "Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 14, no. 2, p. 169, 2020, doi: 10.22146/ijccs.51743.
- [7] K. Liu, W. Li, and M. Guo, "Emoticon Smoothed Language Models for Twitter Sentiment Analysis," *Aaai*, pp. 1678–1684, 2012, [Online]. Available: <http://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/viewFile/5083/5319>.
- [8] Y. Yunitasari, A. Musdholifah, and A. K. Sari, "Sarcasm Detection For Sentiment Analysis in Indonesian Tweets," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 1, p. 53, 2019, doi: 10.22146/ijccs.41136.
- [9] N. W. S. Saraswati, "Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis," *Semin. Nas. Sist. Inf. Indones.*, pp. 587–591, 2013.
- [10] N. D. Putranti and E. Winarko, "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine," *Ijcss*, vol. 8, no. 1, pp. 91–100, 2014, doi: 10.22146/ijccs.3499.
- [11] A. F. Hidayatullah and A. Sn, "Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter," *Semin. Nas. Inform. 2014*, vol. 2014, no. August 2013, pp. 0–8, 2014.
- [12] M. Bouazizi and T. Otsuki, "A Pattern-Based Approach for Sarcasm Detection on Twitter," *IEEE Access*, vol. 4, pp. 5477–5488, 2016, doi: 10.1109/ACCESS.2016.2594194.
- [13] M. Bouazizi and T. Ohtsuki, "Opinion Mining in Twitter How to Make Use of Sarcasm to Enhance Sentiment Analysis," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 - ASONAM '15*, 2015, pp. 1594–1597, doi: 10.1145/2808797.2809350.
- [14] I. Klyueva, "Improving Quality of the Multiclass SVM Classification Based on the Feature Engineering," *Proc. - 2019 1st Int. Conf. Control Syst. Math. Model. Autom. Energy Effic. SUMMA 2019*, pp. 491–494, 2019, doi: 10.1109/SUMMA48161.2019.8947599.
- [15] H. Sulistiani, K. Muludi, and A. Syarif, "Implementation of Dynamic Mutual Information and Support Vector Machine for Customer Loyalty Classification," *J. Phys. Conf. Ser.*, vol. 1338, no. 1, 2019, doi: 10.1088/1742-6596/1338/1/012050.