



## Studi Komparatif Metode Ekstraksi Fitur pada Analisis Sentimen Maskapai Penerbangan Menggunakan *Support Vector Machine* dan *Maximum Entropy*

Mona Cindo<sup>1</sup>, Dian Palupi Rini<sup>2</sup>, Ermatita<sup>3</sup><sup>1,2,3</sup> Magister Teknik Informatika, Ilmu Komputer, Universitas Sriwijaya<sup>1</sup>monacindo@gmail.com, <sup>2</sup>dprini@unsri.ac.id, <sup>3</sup>ermatitaz@yahoo.com**Abstract**

Almost all companies use social media to improve their product services and provide after-sales services that allow their customers to review the quality of their products. By using Twitter social media to be an important source for tracking sentiment analysis. Sentiment analysis is one of the most popular studies today, using sentiment analysis companies can analyze customer satisfaction to improve their services. This study aims to analyze airline sentiments with five different features such as pragmatic, lexical n-gram, POS, sentiment, and LDA using the Support Vector Machine and Maximum Entropy methods. The best results can be obtained using the Maximum Entropy method using all feature extraction with an accuracy of 92.7% and in the Support Vector Machine method, the accuracy obtained is 89.2%.

Keywords: sentiment analysis, twitter, airline, feature extraction, support vector machine, maximum entropy

**Abstrak**

Hampir semua perusahaan menggunakan media sosial untuk meningkatkan layanan produknya, dan memberikan layanan *after sales* yang memungkinkan pelanggannya dapat me-review kualitas produknya. Dengan menggunakan media sosial twitter untuk menjadi sumber penting untuk melacak analisis sentimen. Analisis sentimen adalah salah satu penelitian paling populer saat ini, dengan menggunakan analisis sentimen perusahaan dapat menganalisis kepuasan pelanggan untuk meningkatkan layanannya. Penelitian ini bertujuan untuk menganalisis sentimen maskapai penerbangan dengan lima fitur berbeda seperti topik pragmatic, lexical n-gram, POS, sentimen, dan LDA menggunakan metode *Support Vector Machine* dan *Maximum Entropy*. Hasil terbaik di dapat pada metode *Maximum Entropy* menggunakan semua ekstraksi fitur dengan akurasi 92,7% dan pada metode *Support Vector Machine* akurasi yang diperoleh sebesar 89,2%.

Kata kunci: analisis sentimen, twitter, maskapai penerbangan, ekstraksi fitur, *support vector machine*, *maximum entropy*

© 2019 Jurnal RESTI

**1. Pendahuluan**

Dalam beberapa tahun terakhir, media sosial telah berkembang luas, dan penggunaannya meningkat pesat. Terlihat dari orang-orang yang sangat bergantung pada media sosial di mana mereka dapat dengan mudah mengekspresikan pendapat dan perasaan mereka [1]. Media sosial adalah bagian penting dari rutinitas sehari-hari bagi orang-orang saat ini. Media sosial dan internet dapat digunakan untuk berbagai hal, seperti iklan, menyebarkan pendapat politik, gaya hidup, dan keuangan, mendapatkan komentar pengguna tentang ulasan produk, menyebarkan berita [2]. Saat ini, jalur udara merupakan salah satu jalur yang paling sering digunakan untuk transportasi. Kebanyakan perusahaan

maskapai penerbangan bersaing untuk mendapatkan keunggulan dengan meningkatkan kualitas pelayanan mereka. Beberapa maskapai penerbangan Amerika fokus pada kualitas pelayanan dan pengalaman pengguna layanan mereka [1]. Dengan menggunakan media sosial sebagai sumber penting mendapatkan tingkat kepuasan pelanggan, terutama media sosial seperti Twitter untuk memahami apakah pelanggan menyukai atau tidak layanan yang telah disediakan oleh maskapai tersebut.

Twitter memungkinkan pengguna untuk mengirim dan membaca pesan yang disebut dengan Tweet. Dengan maksimal 140 karakter teks yang ditampilkan di halaman pengguna Twitter memudahkan pengguna

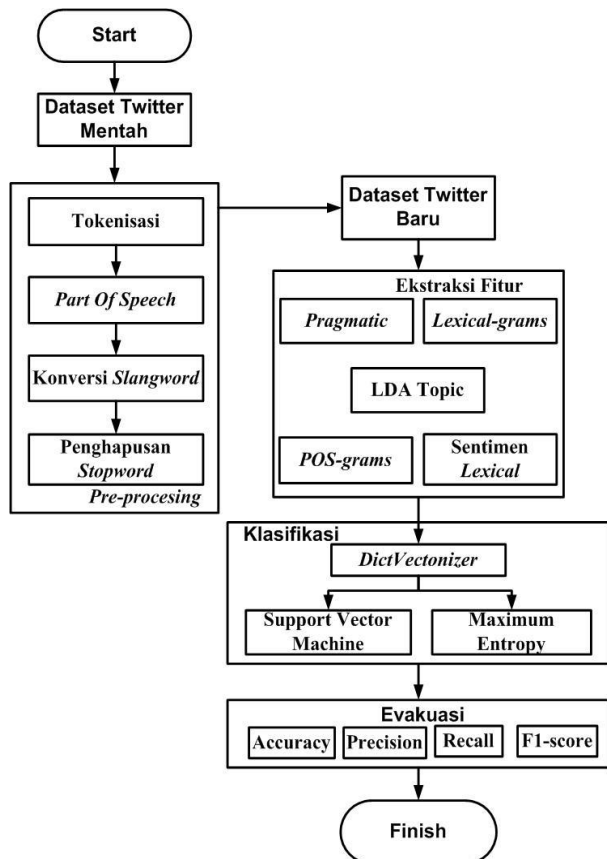
untuk berbagi pemikiran dan pendapat dengan yang lain secara cepat dan mudah [3]. Untuk mendeteksi kalimat opini, terutama opini tentang produk layanan perusahaan maskapai penerbangan dalam konten media online dapat dilakukan dengan menganalisis sentimen pengguna yang disampaikan melalui pesan yang ditulis di jejaring sosial Twitter.

Analisis sentimen adalah proses memahami, mengekstraksi, dan secara otomatis memproses data tekstual untuk mendapatkan informasi yang dibahas [4]. Pada penelitian sebelumnya, emotikon digunakan pengguna sebagai cara mudah untuk mengekspresikan emosi secara singkat. Diskusi tentang analisis sentimen ini sebagian besar menggunakan *Naïve Bayes* (NB), *Maximum Entropy* dan *Support Vector Machine* [5]. Penelitian sebelumnya membahas analisis sentimen di facebook menggunakan fitur pendeteksian emotikon, fitur berbasis leksikon (n-gram), dan fitur penandaan POS. dengan menggunakan *Support Vector Machine* [6]. Penelitian [7], menemukan *Maximum Entropy* memiliki tingkat efektivitas yang lebih baik daripada *Support Vector Machine* dan metode klasifikasi lainnya. Penelitian sebelumnya membahas analisis sentimen opini maskapai penerbangan menggunakan metode *Support Vector Machine* sebesar 40% [8].

Seperti yang disebutkan di atas beberapa penelitian sebelumnya telah menggunakan beberapa fitur seperti *emoticon*, n-grams. Fitur tersebut dianggap kurang lengkap karena dibutuhkanannya perhitungan kata-kata yang lebih mendalam seperti memperhatikan tanda baca, *interjection* (kalimat seruan), kalimat negasi, dan topik yang dibahas sehingga dapat mengetahui emosi sebenarnya yang disampaikan oleh penulis. Oleh karena itu, penelitian ini menambahkan fitur pragmatic yang terdiri dari perhitungan emotikon, tanda baca, kalimat seruan, dan kalimat negasi. Selain itu, penelitian ini juga menerapkan LDA topik sebagai salah satu fitur untuk mendeteksi suatu topik pembicaraan pada suatu tweets sehingga nantinya diharapkan penelitian ini memiliki tingkat akurasi yang lebih tinggi dari penelitian sebelumnya. Penelitian ini menggunakan dataset opini maskapai penerbangan Amerika. Untuk mendapatkan hasil yang terbaik, penelitian ini menggunakan dua metode pembandingan yaitu *Support Vector Machine* dan *Maximum Entropy*. Selain itu, penelitian ini menggunakan 5 ekstraksi fitur yaitu fitur *pragmatic*, *lexical n-grams*, *lexical sentiment*, *POS-grams*, dan LDA topik yang diharapkan dapat meningkatkan akurasi dari analisis sentimen.

## 2. Metode Penelitian

Untuk melakukan analisis sentimen pada Twitter, dibutuhkan beberapa pendekatan untuk mendapatkan hasil yang baik. Gambar 1 menunjukkan tahapan analisis sentimen pada penelitian ini.



Gambar 1. Tahapan Analisis Sentimen

Pada tahapan Gambar 1 menggambarkan tahapan pertama yaitu mengumpulkan data dan diikuti oleh *preprocessing*, ekstraksi fitur, klasifikasi, dan tahapan terakhir adalah evaluasi.

### 2.1. Pengumpulan Data

Pengumpulan data merupakan salah satu bagian yang penting untuk proses analisis sentimen. Dalam penelitian ini, peneliti menggunakan data twitter tentang opini maskapai penerbangan yang tersedia di internet. Distribusi dataset dapat ditunjukkan pada Tabel 1.

Tabel 1. Distribusi Dataset Twitter

AIRLANES OPINION DATASET				
Source : <a href="https://www.figure-eight.com/data-for-everyone/">https://www.figure-eight.com/data-for-everyone/</a>				
Positif	Negatif	Total	Training	Testing
1510	1923	3433	3090	343

### 2.2. Data Pre-Processing

Tujuan dari *preprocessing* pada penelitian ini adalah untuk menyederhanakan pengambilan fitur dengan menghapus, dan mengubah data yang tidak dibutuhkan untuk mempermudah memproses data. *Preprocessing* ini terdiri dari *Tokenization*, *Part-Of-Speech*, *Slang Words*, dan *Stopword*.

*Tokenization* berfungsi untuk memecah komentar menjadi satuan kata. Proses tokenisasi dilakukan dengan melihat setiap spasi yang ada dalam komentar, maka berdasarkan spasi tersebut kata-kata dapat dipecah [9].

*Part-of-Speech* merupakan tahapan untuk merubah kalimat dalam dataset menjadi *Part-of-Speech Tagging* [10].

*Slangword Conversion* berfungsi untuk mengubah kata tidak baku ke kata baku.

*Stopword Removal* berfungsi untuk menghilangkan kata-kata yang tidak penting dalam proses klasifikasi dan penentuan alasan atau kumpulan kata-kata yang sering muncul tetapi jika dihapus tidak mengubah makna dari tweet.

### 2.3. Ekstraksi Fitur

Ekstraksi fitur bertujuan untuk mengenali karakteristik kalimat Twitter dan menjadikannya fitur. Penelitian ini menggunakan lima ekstraksi fitur, yaitu *Pragmatic*, *Lexical-Gram*, *POS-gram*, *Lexical Sentiment* dan *Latent Dirichlet Allocation* (LDA Topic).

Pengguna Twitter dapat dengan mudah membagikan pendapat mereka dalam bentuk *tweet*, hanya menggunakan 140 karakter di setiap tweet [11]. Hal ini dapat menyebabkan pengguna twitter menjadi terbatas dalam mengungkapkan perasaan atau pendapat mereka. Pengguna Twitter sering menggunakan *hashtag*, emotikon, atau tanda baca untuk mengekspresikan perasaan dan pendapat mereka [12].

Fitur *Pragmatic* berfungsi menghitung tanda baca seperti penggunaan huruf besar, tanda tanya (?), Tanda seru (!), Kutipan (") dan ("), tagar (#), dan emotikon [13]. selain kalimat negasi seperti "no", "none", "never" dan lain-lain.

*Lexical-gram* (n-gram) merupakan salah satu proses fitur untuk mengekstraksi sentimen dari pesan yang ditulis oleh pengguna [6]. satu set kata yang muncul bersama dalam sebuah teks [14]. Fitur *Lexical-gram* dapat digunakan dengan menggunakan *tokenization* di *python library* untuk dapat menyelesaikan kalimat.

*Part-of-Speech* merupakan proses pemberian kelas pada sebuah kata dengan membagi kalimat atau paragraf menjadi kata-kata [15]. Sebuah kalimat dipecah menjadi beberapa bagian seperti kata sifat (D), kata benda (N), dan kata kerja (V). setelah dibagi menjadi beberapa bagian, maka kata-kata yang dipisahkan dibuat menjadi fitur *Part-Of-Speech* seperti fitur *lexical-gram*.

Sentimen *Lexical* Fitur menggunakan kamus *SentiWordNet* leksikal yang memiliki nilai sentimen untuk setiap kata [16]. Teks ini dapat secara otomatis memberikan skor untuk setiap kalimat. Dalam proses ini, pertama, menggunakan *tokenization* pada *python*

*library*. Proses *tokenization* merupakan proses pemotongan kata. Kemudian dengan menggunakan teks *post\_tag* (*Part-Of-Speech*) kalimat tersebut dibagi berdasarkan ADJ (*Adjective*), ADV (*Adverb*), NN (*Noun*). Setelah kalimat tersebut diuraikan maka kata tersebut hanya dapat dinilai berdasarkan kamus *SentiWordNet*.

*Latent Dirichlet Allocation* (LDA Topic) adalah model probabilistik generatif dari sebuah *corpus*. Model generatif yang memungkinkan pembaruan dibuat Didiskusikan oleh kelompok yang tidak menjelaskan beberapa bagian data. Ide dasar sehingga dokumen direpresentasikan sebagai campuran acak untuk topik laten, di mana setiap topik ditandai oleh distribusi kata. LDA memiliki besar dalam bidang pembelajaran statistik dan peningkatan bahasa dengan cepat menjadi salah satu teknik pemodelan teks probabilistik yang paling populer [17] [18].

### 2.4. Klasifikasi

Dalam penelitian ini menggunakan dua metode klasifikasi yaitu, *Support Vector Machine* dan *Maximum Entropy* untuk mendapatkan hasil klasifikasi terbaik.

Saat menganalisis data, *Support Vector Machine* menentukan batas keputusan dan menggunakan *kernel* untuk melakukan perhitungan *input*. Tujuan dari metode *Support Vector Machine* adalah untuk menemukan pemisah linier di ruang pencarian yang dapat memisahkan berbagai kelas [19].

Metode *Maximum Entropy* bertujuan untuk memaksimalkan entropi dalam sistem dengan memprediksi distribusi kondisi dari label di setiap kelas. Klasifikasi ini menangani fitur yang tumpang tindih seperti regresi logistik [20].

Distribusi ini kemudian didefinisikan sebagai MaxEnt yang tidak membuat asumsi tentang fitur-fiturnya. MaxEnt dapat didefinisikan sebagai:

$$P_{ME}(c | d, \lambda) = \frac{\exp[\sum_i \lambda_i f_i(c, d)]}{\sum_c \exp[\sum_i \lambda_i f_i(c, d)]} \quad (1)$$

c adalah kelas, b adalah twitter dan  $\lambda$  adalah bobot vektor.

### 2.5. Evaluasi

Pada langkah ini, peneliti menggunakan beberapa matriks untuk menganalisis ekstraksi dan klasifikasi fitur. peneliti menggunakan *confusion matrix* untuk menghitung *precision*, *recall*, *f1-score*, dan *accuracy* klasifikasi untuk menemukan kinerja klasifikasi.

## 3. Hasil dan Pembahasan

Penelitian ini menggunakan dataset *open source* yang didapat di internet. Data ini terdiri dari 1510 *tweet* positif dan 1923 *tweet* negatif. Total dari seluruh data adalah 3433 *tweet*. Pada proses evaluasi, data dibagi

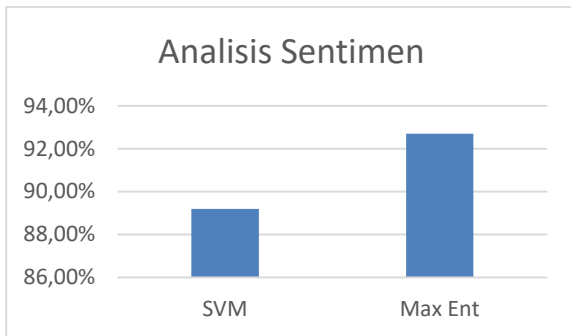
menjadi 3090 dataset latih dan 343 dataset uji. Berikut contoh dari dataset yang telah dikumpul dapat dilihat pada Tabel 2.

Tabel 2. Contoh Dataset Maskapai Penerbangan

Label	Tweet
Negatif	@united smiling ? NOT !!
Negatif	@united I knew I should've flew @VirginAmerica lol
Positif	@united is my favorite airline .
Positif	@united love the new 1st class breakfast !

### 3.1. Hasil Analisis Sentimen

Pada penelitian ini, peneliti menggunakan dua metode klasifikasi sebagai pembanding. Untuk menyatukan semua fitur, paper ini menggunakan algoritma yang disebut *One Hot Encoding*. Algoritma ini tersedia di *sklearn kit* (library dari *Python*) yang disebut dengan *DictVectorizer*. Berikut hasil akurasi analisis sentimen pada metode *Maximum Entropy* dan *Support Vector Machine* ditunjukkan pada Gambar 2.



Gambar 2. Hasil Akurasi Analisis Sentimen

Gambar 2 menjelaskan hasil analisis sentimen dengan menggunakan *matrix* akurasi. Berdasarkan gambar tersebut, *Maximum Entropy* memberikan performa yang lebih baik dengan akurasi 92,70% dibanding dengan *Support Vector Machine* yang hanya 89,20%. Untuk mendapat hasil lebih spesifik penelitian ini juga menggunakan *confusion matrix*. Hasil dari *confusion matrix* dapat dilihat di Tabel 3.

Tabel 3. Hasil Confusion Matrix Analisis Sentimen

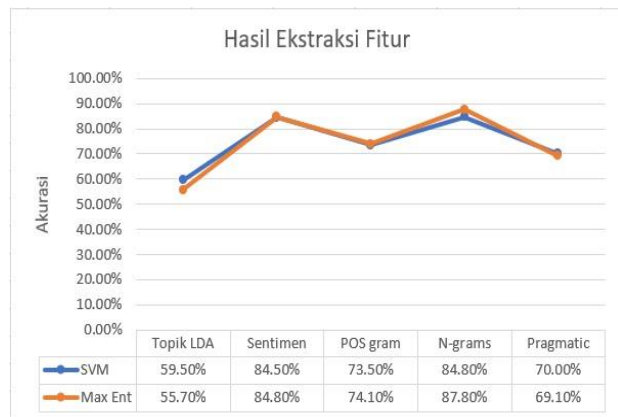
	TP	TN	FP	FN
<b>SVM</b>	156	150	15	22
<b>Max Ent</b>	162	156	9	16

Dari total 171 data opini positif dan 172 data opini negatif, sistem ini berhasil memprediksi 162 data opini positif data dan 156 data opini negatif pada metode *Maximum Entropy*. Sedangkan pada metode *Support Vector Machine* data yang berhasil diprediksi sebagai opini positif adalah 156 dan negatif 150. Berdasarkan *matrix* tersebut dapat disimpulkan pendekatan ini berhasil dengan baik dengan menentukan analisis sentimen. Dari penjelasan tersebut, data opini yang

paling mudah diprediksi oleh metode tersebut adalah opini positif.

### 3.2. Hasil Ekstraksi Fitur

Untuk melihat fitur yang paling berguna, penelitian ini juga melakukan evaluasi pada tiap-tiap fitur. Hasil dari masing-masing fitur dapat dilihat pada Gambar 3.



Gambar 3. Hasil Fitur Individu

Hasil dari gambar tersebut menjelaskan bahwa fitur *n-grams* merupakan fitur yang paling penting di penelitian ini. Fitur *n-grams* menghasilkan akurasi sebesar 87,80% dengan menggunakan metode *Maximum Entropy*. Tabel 4 merupakan contoh hasil dari fitur *n-grams*.

Tabel 4. Contoh Hasil Fitur N-grams

Dataset	Hasil N-grams
@united Great ! I'm ready to go home	@united 1:0, Great 1:0, ! 1:0, I 1:0, am 1:0, ready 1.0, to 1.0, go 1.0, home 1.0
@united you suck .	@united 1:0, you 2:0, suck 1:0,
@SouthwestAir you're the best	@SouthwestAir 1:0 are 1:0, best 1:0

Berdasarkan hasil dari *n-grams* tersebut kalimat yang paling sering terdeteksi sebagai kalimat opini adalah “Great”, “Appreciate”, “Worst” dan “lack”.

Hasil terbaik kedua terdapat pada fitur sentimen. Pada fitur sentimen hasil akurasi yang didapat sebesar 84,80% pada metode *Maximum Entropy*. Fitur ini memanfaatkan kamus leksikal bernama *SentiWordNet* yang berisi tentang kumpulan kalimat opini yang telah diberi skor positif dan negatifnya. Tabel 5 menunjukkan contoh hasil dari fitur sentimen.

Tabel 5. Contoh Hasil Fitur Sentimen

	@united you suck . @SouthwestAir you're the best .
Suck	Positif sentimen : 0.44791 Negatif sentimen : 0.8125 Sentimen : - 0.36458
Best	Positif sentimen : 0.0625 Negatif sentiment : 0.03125 Sentimen : 0.59375 Sentimen = 0.22917

Berdasarkan tabel tersebut, kata “*suck*” memiliki probabilitas – 0.36458 yang berarti kata tersebut pasti negatif. Sedangkan pada kalimat *best* probabilitas kata tersebut 0.59375 yang berarti kata tersebut lebih mengarah ke positif. Pada fitur ini, kamus leksikal hanya dapat menilai kata *Noun*, *Adjective*, dan *Adverb*.

Hasil terbaik ketiga terdapat pada fitur *POS-gram* dengan nilai akurasi 74,10% pada metode *Maximum Entropy*. Fitur ini mempunyai fungsi yang sama dengan n-grams fitur, akan tetapi pada fitur ini yang dihitung hanya hasil dari *Part-of-Speech* dari tiap Tweet. Tabel 6 menggambarkan contoh dari hasil fitur *POS-gram*.

Tabel 6. Contoh Hasil Fitur POS-grams

@united	Gate	Agent	Alavera	is	Amazing
@	^	^	^	V	A

Pada tabel 6 hasil *tag* dikonversikan sebagai “@”, nama sebagai “^”, *is* sebagai “*Verb*”, dan *amazing* sebagai “*Adverb*”.

Hasil terbaik keempat terdapat pada fitur *Pragmatic* dengan hasil akurasi 69,10% pada metode *maximum entropy*. Berbeda dengan hasil fitur sebelumnya, fitur *pragmatic* memiliki hasil terbaik pada metode *Support Vector Machine*. Fitur ini mendeteksi tanda baca seperti “!”, “?” sekaligus *emoticon* sebagai suatu bentuk yang merepresntasikan emosi seseorang.

Terakhir hasil yang paling kecil terdapat pada fitur LDA topik. Sama seperti fitur *pragmatic*, fitur ini memiliki hasil yang lebih baik pada metode *Support Vector Machine*. Fitur ini menghubungkan suatu kalimat pada suatu topik seperti pada kalimat “@united I'm not booked on a 1:30 return flight ..” yang dihubungkan sebagai tema “waktu”.

### 3.1. Hasil Kombinasi Fitur

Untuk mendapatkan hasil terbaik penelitian ini juga melakukan evaluasi dari beberapa fitur kombinasi. Adapun fitur yang dikombinasikan dapat dilihat pada Table 7.

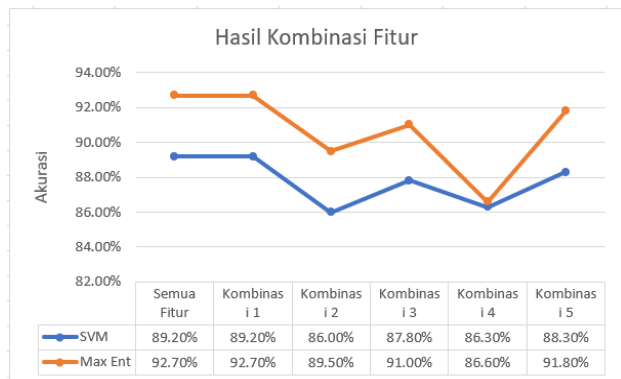
Tabel 7. Kombinasi Fitur

Nama Fitur	Kombinasi Fitur
All feature	Semua Fitur dikombinasikan
1 <sup>st</sup> combination	Pragmatic, N-grams, POS grams, Sentiment
2 <sup>nd</sup> combination	Pragmatic, N-grams, POS grams, LDA topic
3 <sup>rd</sup> combination	Pragmatic, N-grams, Sentiment, LDA topic
4 <sup>th</sup> combination	Pragmatic, POS grams, Sentiment, LDA topic
5 <sup>th</sup> combination	N-grams, POS grams, Sentiment, LDA topic

Pada tabel di atas, penelitian ini melakukan evaluasi dari kombinasi fitur dengan menggunakan dua metode klasifikasi. Hasil evaluasi dapat di lihat di Gambar 4.

Dari Gambar 4, hasil terbaik ada pada seluruh kombinasi fitur dan kombinasi pertama dengan akurasi 92,70% pada metode *Maximum Entropy*. Hasil dari seluruh fitur kombinasi dan kombinasi fitur pertama

sangat identik. Hal ini membuktikan bahwa fitur LDA topik hanya memberikan kontribusi yang sedikit di penelitian ini.



Gambar 4. Hasil Fitur Kombinasi

## 4. Kesimpulan

Dalam penelitian ini, hasil terbaik didapatkan pada semua fitur dengan akurasi 92,7% pada metode *Maximum Entropy* sedangkan pada metode *Support Vector Machine* akurasi yang didapat sebesar 89,20%. Hasil dari semua fitur yang diperoleh sama dengan hasil kombinasi fitur 1. Ini membuktikan bahwa fitur topik LDA tidak memberikan kinerja yang signifikan untuk dataset ini. Untuk hasil ekstraksi fitur individual, hasil terbaik diberikan oleh fitur n-gram. Fitur ini dapat menunjukkan kata-kata yang menunjukkan emosi kata seperti “*great*”, “*appreciate*”, dan “*amazing*” untuk sentimen positif dan “*worst*”, “*issue*”, dan “*lack*” untuk sentimen negatif. Untuk penelitian yang akan datang, penelitian ini membutuhkan metode lain seperti *deep learning* untuk meningkatkan kinerja analisis sentimen.

## Daftar Rujukan

- [1] C. Paper and N. Inspired, “Online Social Media-based Sentiment Analysis for US Airline companies Online Social Media-based Sentiment Analysis for US Airline companies,” no. April, 2017.
- [2] A. Alarifi, M. Alsaleh, and A. M. Al-Salman, “Twitter turing test: Identifying social machines,” *Inf. Sci. (Ny)*, vol. 372, pp. 332–346, 2016.
- [3] M. Badri, “Komunikasi Pemasaran UMKM Di Era Media Sosial. Corporate and Marketing Communication,” no. January, p. Jakarta: Pusat Studi Komunikasi dan Bisnis Progra, 2011.
- [4] D. Bandorski *et al.*, “Contraindications for video capsule endoscopy,” *World J. Gastroenterol.*, vol. 22, no. 45, pp. 9898–9908, 2016.
- [5] E. Kouloumpis, T. Wilson, and J. Moore, “Twitter sentiment analysis: The good the bad and the omg!,” *Proc. Fifth Int. AAAI Conf. Weblogs Soc. Media (ICWSM 11)*, pp. 538–541, 2011.
- [6] A. Ortigosa, J. M. Martín, and R. M. Carro, “Sentiment analysis in Facebook and its application to e-learning,” *Comput. Human Behav.*, vol. 31, no. 1, pp. 527–541, 2014.
- [7] B. Gupta, I. M. Negi, K. Vishwakarma, G. Rawat, P. Badhani, and B. Tech, “Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python,” *Int. J. Comput. Appl.*, vol. 165, no. 9, pp. 975–8887, 2017.
- [8] A. M. Pravina, I. Cholissodin, and P. P. Adikara, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada

- Dokumen Twitter Menggunakan Algoritme Support Vector Machine ( SVM ),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 3, pp. 2789–2797, 2019.
- [9] A. Rachmat and Y. Lukito, “Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes,” *J. Inform. dan Sist. Inf. Univ. Ciputra*, vol. 02, no. 02, pp. 26–34, 2016.
- [10] P. Baid, “Sentiment Analysis of Movie Reviews using Machine Learning Techniques,” no. December 2017, 2018.
- [11] M. Desai and M. A. Mehta, “Techniques for sentiment analysis of Twitter data: A comprehensive survey,” *Proceeding - IEEE Int. Conf. Comput. Commun. Autom. ICCCA 2016*, no. April 2016, pp. 149–154, 2017.
- [12] T. Jain, N. Agrawal, G. Goyal, and N. Aggrawal, “Sarcasm detection of tweets: A comparative study,” *2017 10th Int. Conf. Contemp. Comput. IC3 2017*, vol. 2018-Janua, no. August, pp. 1–6, 2018.
- [13] M. S. M. Suhaimin, M. H. A. Hijazi, R. Alfred, and F. Coenen, “Natural language processing based features for sarcasm detection: An investigation using bilingual social media texts,” *ICIT 2017 - 8th Int. Conf. Inf. Technol. Proc.*, pp. 703–709, 2017.
- [14] P. B. Awachate and V. P. Kshirsagar, “Improved Twitter Sentiment Analysis Using N Gram Feature Selection and Combinations,” *Int. J. Adv. Res. Comput. Commun. Eng. ISO*, vol. 3297, no. 9, pp. 154–157, 2007.
- [15] A. G. Prasad, S. Sanjana, S. M. Bhat, and B. S. Harish, “Sentiment analysis for sarcasm detection on streaming short text data,” *2017 2nd Int. Conf. Knowl. Eng. Appl. ICEA 2017*, vol. 2017-Janua, no. 2009, pp. 1–5, 2017.
- [16] C. Musto, G. Semeraro, and M. Polignano, “A comparison of lexicon-based approaches for sentiment analysis of microblog,” *CEUR Workshop Proc.*, vol. 1314, pp. 59–68, 2014.
- [17] Z. Tong and H. Zhang, “A Text Mining Research Based on LDA Topic Modelling,” pp. 201–210, 2016.
- [18] D. M. Blei, B. B. Edu, A. Y. Ng, A. S. Edu, M. I. Jordan, and J. B. Edu, “technique...Latent Dirichlet Allocation,” *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
- [19] J. Jayalekshmi and T. Mathew, “Facial expression recognition and emotion classification system for sentiment analysis,” *2017 Int. Conf. Networks Adv. Comput. Technol. NetACT 2017*, no. July, pp. 1–8, 2017.
- [20] V. A. Kharde and S. S. Sonawane, “Sentiment Analysis of Twitter Data: A Survey of Techniques,” *Int. J. Comput. Appl.*, vol. 139, no. 11, pp. 975–8887, 2016.