



---

**ANALISIS SENTIMEN DI TWITTER TERKAIT TIM NASIONAL SEPAK BOLA  
INDONESIA MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE***

***SENTIMENT ANALYSIS ON TWITTER RELATED TO THE INDONESIAN  
NATIONAL FOOTBALL TEAM USING THE SUPPORT VECTOR MACHINE METHOD***

Galang Sumantri, Prodi Matematika FMIPA UNY

Bambang Sumarno Hadi Marwoto\*, Prodi Matematika FMIPA UNY

\*e-mail: bambang@uny.ac.id

**Abstrak**

Timnas sepak bola Indonesia sering gagal bersaing di berbagai turnamen besar internasional. Sentimen masyarakat terhadap prestasi Timnas yang diekspresikan melalui twitter dapat digunakan sebagai salah satu cara untuk menilai perkembangan sepak bola di Indonesia. Penelitian ini bertujuan untuk mengetahui performa metode *Support Vector Machine* (SVM) dengan ekstraksi fitur word embeddings Word2vec dan FastText dalam analisis sentimen terkait Timnas sepak bola Indonesia. Data dalam penelitian ini menggunakan data teks berupa tweet terkait keikutsertaan Timnas di ajang piala AFF tahun 2018, 2020, dan 2022 yang dikumpulkan dengan metode *crawling*. Metode SVM diawali dengan tahap *preprocessing* dan ekstraksi fitur. Metode ekstraksi fitur dalam penelitian ini menggunakan word embeddings Word2Vec dan FastText. Hasil penelitian, metode SVM menghasilkan akurasi terbaik hingga mencapai 84%, presisi 82%, *recall* 81%, dan *F1 score* sebesar 81%. FastText memiliki performa yang sedikit lebih baik daripada Word2Vec untuk fitur ekstraksi pada analisis sentimen menggunakan SVM, perbedaannya adalah FastText dapat mengenali kata-kata yang tidak ada dalam korpus sedangkan Word2Vec tidak. Model terbaik dihasilkan dengan menggunakan *word embeddings* FastText dengan model Skip-gram.

**Kata kunci:** analisis sentimen, *Support Vector Machine* (SVM), Twitter, FastText, Word2Vec.

**Abstract**

*The Indonesian national football team often fails to compete in major international tournaments. Public sentiment on the national team's achievements, which is expressed through Twitter can be used as a way to assess the development of football in Indonesia. This study aims to determine the performance of the Support Vector Machine (SVM) method with word embeddings Word2vec and FastText feature extraction in sentiment analysis related to the Indonesian national football team. The data in this study uses text data in the form of tweets related to the participation of the National Team in the 2018, 2020 and 2022 AFF Cup which was collected using the crawling method. The SVM method begins with the preprocessing and feature extraction stages. The feature extraction method in this study uses word embeddings Word2Vec and FastText. The results of the study, the SVM method produces the best accuracy up to 84%, 82% precision, 81% recall, and 81% F1 score. FastText has slightly better performance than Word2Vec for feature extraction on sentiment analysis using the support vector machine, the difference is that FastText can recognize words that are not in the corpus while Word2Vec cannot. The best model is generated by using FastText word embeddings with the Skip-gram model.*

**Keywords:** sentiment analysis, *Support Vector Machine* (SVM), Twitter, FastText, Word2Vec.

## PENDAHULUAN

Prestasi bidang olahraga pada berbagai tingkatan, terutama tingkat internasional menjadi sumber kebanggaan bagi suatu negara. Prestasi ini menjadi salah satu bukti nyata dari upaya sebuah negara dalam mempersiapkan atlet-atlet terbaik dan program pelatihan yang efektif. Prestasi olahraga juga memiliki dampak luas, seperti memperkuat citra positif suatu negara di mata dunia. Prestasi olahraga yang diraih dengan gemilang di tingkat internasional merupakan salah satu aspek penting dalam membangun reputasi bangsa di kancah dunia.

Sepak bola merupakan salah satu cabang olahraga yang mendapatkan perhatian banyak pihak. Sepak bola adalah olahraga yang bertipe tim atau beregu sehingga dapat dimainkan oleh banyak orang. Sepak bola memungkinkan partisipasi dari berbagai usia dan latar belakang sosial. Aturan dalam sepak bola juga relatif sederhana sehingga mudah untuk dimainkan. Sepak bola di masyarakat Indonesia cukup populer dan memiliki banyak penggemar. Berdasarkan survei yang dilakukan oleh Ipsos (2022), Indonesia menempati urutan pertama sebagai negara dengan proporsi penggemar sepak bola terbesar di dunia dengan proporsi sebesar 69%.

Timnas senior sepak bola Indonesia sering gagal berprestasi di berbagai turnamen besar internasional. Timnas sepak bola Indonesia terdiri dari Timnas senior dan Timnas dengan kelompok usia seperti Timnas U-19 dan U-23. Timnas senior sepak bola Indonesia adalah tim dengan pemain terpilih dari berbagai usia yang menjadi wakil resmi negara Indonesia dalam kompetisi sepak bola internasional. Meskipun penggemar sepak bola di Indonesia tergolong besar, Timnas senior sepak bola Indonesia lebih banyak gagal meraih gelar juara di setiap turnamen yang diikutinya. Sejak kemerdekaannya, Timnas senior sepak bola Indonesia belum pernah lolos kualifikasi turnamen sepak bola pada tingkat dunia yaitu Piala Dunia (*World Cup*). Pada turnamen tingkat benua yaitu Piala AFC (*Asian Football Confederation*), Timnas senior sepak bola Indonesia juga jarang berhasil lolos tahap kualifikasi. Peluang Timnas senior sepak bola Indonesia untuk juara yang paling besar berada di turnamen tingkat ASEAN (*Association of Southeast Asian Nations*). Turnamen tingkat ASEAN tersebut adalah Piala AFF (*ASEAN Football Federation*). Namun, kenyataannya Timnas senior sepak bola Indonesia kesulitan memenangkannya.

Pada era digital sekarang media sosial menjadi salah satu wadah yang paling mudah untuk berekspresi atau menyampaikan pendapat. Hampir setiap penduduk di Indonesia, saat ini punya media sosial. Berdasarkan data dari Social (2022) pada bulan februari 2022, di Indonesia terdapat 191,4 juta pengguna media sosial atau 88,5% dari populasi penduduk yang berumur lebih dari 13 tahun. Salah satu media sosial yang aktif sering digunakan adalah twitter. Pada data Social (2022) juga menyatakan bahwa twitter termasuk dalam lima platform teratas yang paling banyak digunakan dengan persentase sebesar 58% pada rentang usia 16-64 tahun. Jumlah pengguna aktif twitter yang banyak membuat twitter mempunyai sentimen yang banyak dan beragam.

Pendapat atau opini yang dituangkan ke dalam twitter dikenal sebagai tweet. Kumpulan tweet yang banyak tersebut sering dimanfaatkan untuk melakukan penilaian terhadap suatu topik atau tujuan-tujuan tertentu. Proses penilaian ini sering disebut analisis sentimen. Analisis sentimen ini biasanya mengkategorikan sebuah sentimen kedalam sentimen dengan nada “positif” dan sentimen dengan nada “negatif”. Banyaknya sentimen yang ada di media sosial tentu akan memakan banyak waktu dan tenaga apabila dilakukan secara manual. Salah satu cara untuk mengatasi hal tersebut adalah melakukan analisis sentimen dengan waktu yang singkat adalah dengan analisis sentimen menggunakan pendekatan machine learning (Ravi & Ravi, 2015).

SVM sering dianggap sebagai pengklasifikasi yang mempunyai akurasi terbesar dalam masalah klasifikasi teks, termasuk analisis sentimen (Basari et al., 2013). Dalam tulisannya, Joachims (2005) memberikan bukti secara teoritis dan secara eksperimen bahwa SVM cocok untuk data teks. Salah satunya adalah pada pengolahan data teks didapati jumlah fitur yang

sangat besar (lebih dari 10000) dan SVM cenderung tidak tergantung pada besarnya dimensi data. Selain itu, sebagai metode yang berbasis statistik, SVM memiliki landasan teori yang dapat dianalisa dengan jelas (Nugroho et al., 2003). Oleh karena itu, penelitian ini menggunakan metode SVM sebagai pengklasifikasinya.

Penelitian sebelumnya terkait analisis sentimen mengenai tim nasional sepak bola Indonesia adalah penelitian yang dilakukan oleh Astiningrum et al. (2020). Penelitian tersebut menggunakan algoritma Naïve Bayes untuk klasifikasi dan diterapkan pada data twitter ketika kualifikasi piala dunia, pertandingan antara Indonesia vs Malaysia. Hasil dari penelitian tersebut didapatkan akurasi tertinggi sebesar 87%. Penelitian lainnya dilakukan oleh Pratama et al. (2022) yang membahas mengenai Timnas sepak bola Indonesia pada Piala AFF 2020 menggunakan algoritma KNN. Hasil dari penelitian tersebut menghasilkan model klasifikasi dengan akurasi sebesar 67,49%, *precision* 78,99% dan *recall* 47,69%. Penelitian serupa dilakukan oleh Hadiyan Sidik et al. (2022) yang juga membahas mengenai Timnas sepak bola Indonesia pada Piala AFF 2020, namun menggunakan algoritma Naïve Bayes. Pada penelitian tersebut menghasilkan akurasi sebesar 64,41%. Bedanya dengan penelitian serupa sebelumnya, penelitian ini menggunakan algoritma SVM sebagai pengklasifikasinya dan diterapkan pada Piala AFF 2022.

Metode ekstraksi fitur yang digunakan dalam proses mengubah teks menjadi vektor agar dapat diproses menggunakan SVM juga cukup beragam. Perkembangan fitur ekstraksi terbaru adalah *word embeddings* yang menggunakan pembelajaran jaringan syaraf tiruan untuk mempelajari representasi vektor kata yang mencerminkan hubungan semantik dan sintaktik antara kata-kata dalam dokumen. Metode *word embeddings* yang cukup populer adalah Word2Vec dan FastText. Word2Vec merupakan model *word embeddings* yang algoritmanya diciptakan oleh Mikolov et al. (2013) dan FastText dikembangkan berdasarkan Word2Vec oleh Facebook AI Research. Penelitian ini akan menggunakan kedua model *word embeddings* ini sebagai fitur ekstraksi dalam metode SVM untuk memperoleh kinerja yang terbaik.

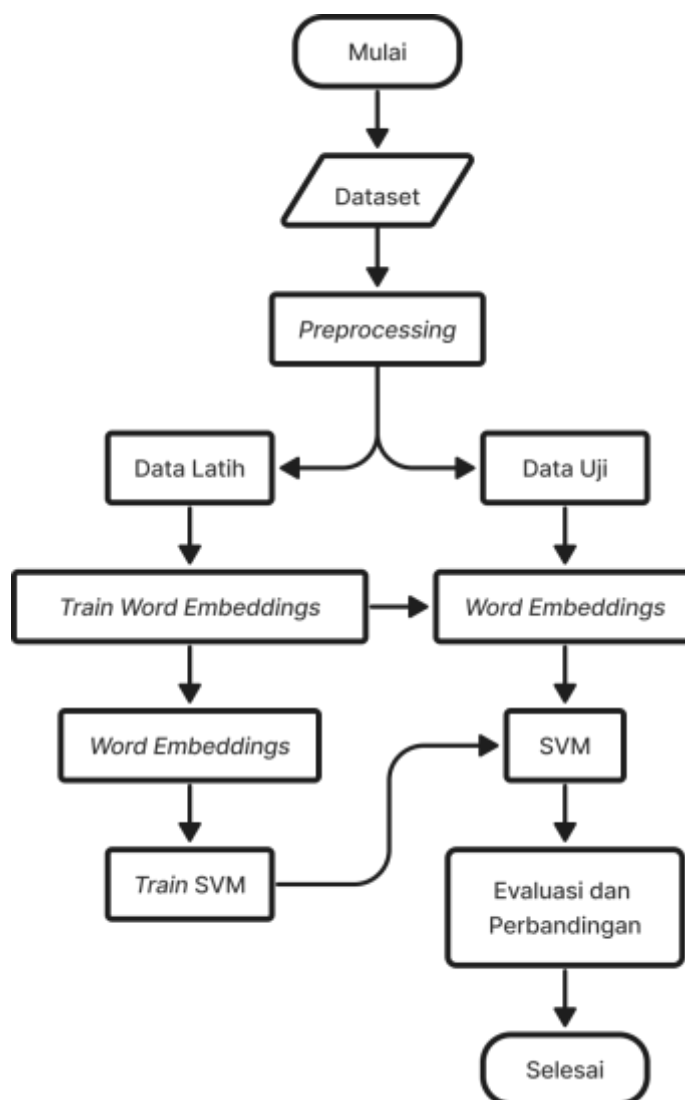
## **METODE**

### **Deskripsi Data**

Data yang digunakan dalam penelitian ini berasal dari twitter. Data yang digunakan adalah data teks berupa tweet yang berisi sentimen pengguna twitter. Sentimen yang twitter yang digunakan adalah sentimen yang membahas mengenai Timnas sepak bola Indonesia. Teknik pengumpulan data menggunakan metode *crawling* pada saat pelaksanaan Piala AFF 2018, 2020, dan 2022. Keyword yang digunakan untuk mengumpulkan data adalah “Timnas Indonesia AFF”. Label data yang digunakan dalam penelitian ini adalah positif dan negatif. Label data diperoleh dengan *majority voting* yang diperoleh dari pelabelan yang dilakukan secara manual oleh 3 orang.

### **Tahapan Penelitian**

Pada Gambar 1 terdapat beberapa proses yang dilakukan dalam penelitian ini setelah dilakukan pengumpulan dan pelabelan data, yaitu *preprocessing* data, *word embeddings*, klasifikasi SVM, serta pengujian metode menggunakan *confusion matrix* dan dibandingkan hasil akurasi, *recall*, presisi, dan *F1 score* setiap modelnya.



**Gambar 1.** Flowchart tahapan penelitian

1. *Preprocessing* Data

Pada tahap ini dilakukan *preprocessing* data untuk mengoptimalkan pengolahan data. Terdapat 3 tahapan *preprocessing* data pada penelitian ini antara lain *case folding*, *filtering*, dan *convert slangword*.

2. Pembagian Data

Pengujian dilakukan dengan membagi komposisi data latih sebesar 80% dan data uji sebesar 20% dari dataset. Data latih digunakan untuk membuat model klasifikasi sedangkan data uji digunakan untuk menguji performa klasifikasi setiap modelnya.

3. *Word Embeddings*

*Word embeddings* yang digunakan dalam penelitian ini adalah Word2Vec dan FastText. Word2Vec adalah sebuah model *machine learning* yang digunakan untuk mempelajari representasi vektor dari kata-kata dalam teks yang algoritmanya diciptakan oleh Mikolov et al. (2013). Model Word2Vec bekerja dengan cara mempelajari hubungan antara kata-kata dalam sebuah corpus teks. FastText (Bojanowski et al., 2017) merupakan pengembangan ide dari model Word2Vec dengan menambahkan fitur-fitur baru, seperti kemampuan untuk menghasilkan representasi vektor kata yang lebih akurat dan efisien,

serta kemampuan untuk memproses kata-kata yang tidak ditemukan dalam kamus atau yang jarang muncul dalam teks. Model Word2Vec dan FastText dapat dilatih menggunakan dua jenis arsitektur *neural network*, yaitu *Continuous Bag-of-Words* (CBOW) dan Skip-gram.

#### 4. Support Vector Machine

Metode klasifikasi sentimen dalam penelitian ini menggunakan metode SVM. Konsep SVM apabila dijelaskan secara sederhana adalah usaha untuk mencari *hyperplane* terbaik sebagai pemisah dua buah kelas pada *input space*. SVM mempunyai dasar prinsip klasifikasi linear, di mana SVM dapat memisahkan kasus klasifikasi secara linear. Namun, SVM telah dikembangkan agar dapat digunakan pada kasus nonlinear dengan memasukkan konsep kernel pada ruang kerja yang memiliki dimensi yang lebih tinggi (Octaviani et al., 2014). Kernel yang digunakan dalam penelitian ini adalah kernel linear. Fungsi dari kernel linear ditunjukkan dalam persamaan 1 (Suyanto, 2018).

$$K(x, x_k) = x_k^T x \quad (1)$$

Fungsi kernel pada persamaan 1 akan menghasilkan matriks kernel. Setiap elemen matriks kernel tersebut digunakan untuk menggantikan *dot-product*  $x_i \cdot x_j$  dalam persamaan dualitas Langrange:

$$\max L_D(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (2)$$

Dengan kendala

$$\alpha_i \geq 0 (i = 1, 2, 3, \dots, l), \sum_{i=1}^l \alpha_i y_i = 0 \quad (3)$$

Kemudian dengan persamaan tersebut didapatkan vektor  $w$  dan skalar  $b$  yang berarti telah ditemukan persamaan *hyperplane* yang paling optimum.

#### 5. Evaluasi dan Perbandingan Model

Hasil klasifikasi dievaluasi dengan menggunakan *confusion matrix*. Kemudian dilakukan penghitungan persentase performa klasifikasi berupa persentase akurasi, presisi, *recall*, dan *F1 score*. Hasil peforma kedua *word embeddings* dibandingkan sehingga dapat disimpulkan *word embeddings* yang memiliki peforma lebih baik.

## HASIL DAN PEMBAHASAN

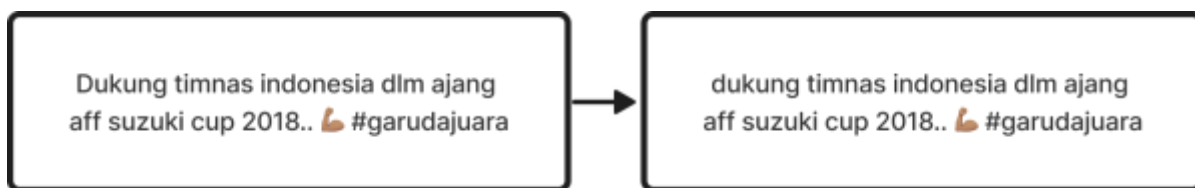
### Hasil

Algoritma klasifikasi yang digunakan sebagai pendekatan analisis sentimen pada penelitian ini yaitu SVM dengan menerapkan teknik *word embeddings* Word2Vec dan FastText menggunakan dataset tweet terkait Piala AFF. Baik Word2Vec maupun FastText terdapat dua arsitektur yang dapat digunakan untuk membangun model *word embeddings*. Maka, terdapat empat model klasifikasi yang dibandingkan untuk memperoleh fitur ekstraksi yang tepat dalam mengklasifikasikan sentimen Timnas Indonesia. Proses pembentukan model tersebut melalui proses-proses sebagai berikut.

#### 1. Preprocessing Data

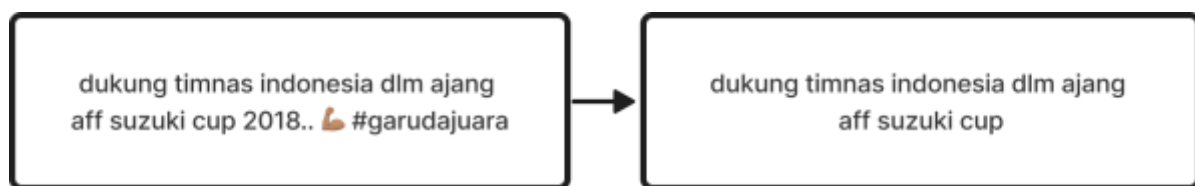
Tahap pertama yang dilakukan dalam *preprocessing* data adalah *case folding*. *Case folding* bertujuan untuk membuat semua kata menjadi seragam hurufnya yaitu dengan mengubah menjadi *lowercase* atau huruf kecil. Gambar 2 merupakan contoh dari proses *case*

*folding.*



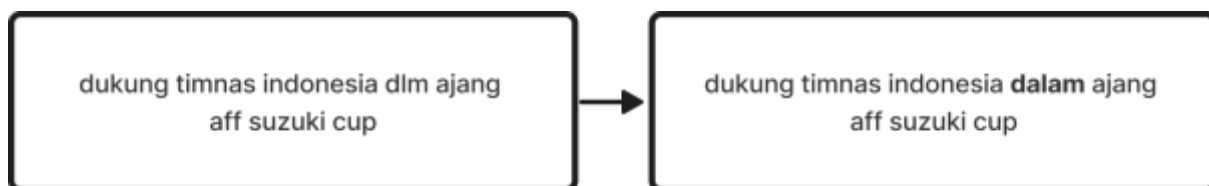
**Gambar 2.** *Case folding*

Tahap yang kedua adalah *filtering*. *Filtering* merupakan tahap penghapusan *hashtag* (#), kata berawalan @ yang biasanya berisi *username*, tanda baca dan angka, *emoticon*, dan string dengan satu karakter. Gambar 3 merupakan contoh dari proses *filtering*.



**Gambar 3.** *Filtering*

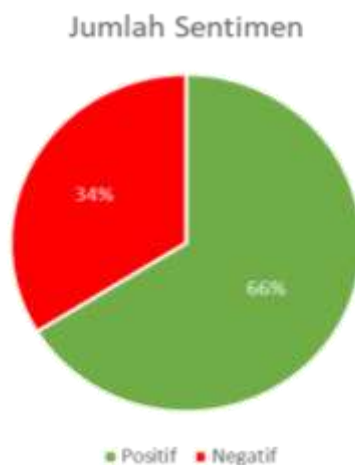
Tahap yang ketiga adalah *convert slangword*. *Convert slangword* merupakan proses merubah berbagai macam penulisan kata *slang* yang memiliki makna yang sama agar memiliki penulisan yang sama pula. Kamus *slangword* dalam penelitian ini menggunakan kamus yang didapatkan dari *colloquial Indonesian lexicon* dalam penelitian yang dilakukan oleh Salsabila et al. (2018). Gambar 4 merupakan contoh dari proses *convert slangword*.



**Gambar 4.** *Convert slangword*

## 2. Pembagian Data

Dataset pada penelitian ini sebanyak 1621 data dengan rincian 1073 data bernada positif dan 548 data bernada negatif. Visualisasi perbandingan label positif dan negatif dapat dilihat pada Gambar 5.



**Gambar 5.** Visualisasi perbandingan jumlah sentimen

Pembagian data latih dan data uji dalam penelitian ini dilakukan secara acak. Pada tahap pembagian data ini menghasilkan data *train* sebanyak 1.296 untuk data latih dengan rincian 855 positif dan 441 negatif serta data tes sebanyak 325 dengan rincian 218 positif dan 107 negatif.

### 3. Word Embeddings

Pembentukan *word embeddings* awal dilatih menggunakan 100.000 data wikipedia bahasa Indonesia. Proses pembentukan semuanya dilakukan menggunakan *library* gensim dengan *vector size* 300. Pada *word embeddings* FastText dilatih dengan *min-n* (panjang minimum *n*-gram yang digunakan untuk melatih representasi kata) sebanyak 3 dan *max-n* (panjang maksimum *n*-gram yang digunakan untuk melatih representasi kata) sebanyak 6. Model *word embeddings* yang sudah dilatih menggunakan dataset wikipedia kemudian dilatih juga menggunakan dataset spesifik terkait Timnas Indonesia pada data latih. Model yang dihasilkan dalam tahap *word embeddings* ini sebanyak 4 model yaitu model Word2Vec dengan arsitektur CBOW, Word2Vec dengan arsitektur Skip-gram, FastText dengan arsitektur CBOW dan FastText dengan arsitektur Skip-gram. *Output* yang dihasilkan dalam proses *word embeddings* ini berupa *vector* dengan *size* 300 yang akan digunakan sebagai *input* model klasifikasi.

### 4. Support Vector Machine

Metode SVM yang digunakan dalam penelitian ini menggunakan kernel linear. Kernel linear dipilih untuk mempercepat waktu komputasi dan dimensi data *input* sudah tergolong tinggi. Pemilihan parameter C yang optimal dipilih menggunakan parameter terbaik pada *5-Fold Cross Validation* pada data latih. Nilai C yang diuji dalam penelitian ini sejumlah 10 pada interval tertutup [0.1, 10]. Parameter dari hasil *5-Fold Cross Validation* terbaik digunakan sebagai parameter untuk membangun model dengan data latih secara keseluruhan. Masing-masing model SVM dari input *word embeddings* yang berbeda digunakan untuk memprediksi data uji yang belum pernah terlihat sama sekali saat pembentukan model klasifikasi.

### 5. Evaluasi dan Perbandingan Model

Setiap prediksi yang dihasilkan oleh setiap model dievaluasi menggunakan *confusion matrix*. Hasil dari *confusion matrix* digunakan untuk mendapatkan nilai akurasi, presisi, *recall*, dan *F1 score* yang merepresentasikan performa model. Perbandingan performa setiap model ditunjukkan pada Tabel 1.

**Tabel 1.** Perbandingan performa setiap model

	Arsitektur	Akurasi	Presisi	Recall	F1 Score
Word2Vec	CBOW	80,3%	79%	75%	76%
	Skip-gram	81,8%	80%	78%	79%
FastText	CBOW	80,9%	79%	76%	77%
	Skip-gram	<b>84%</b>	<b>82%</b>	<b>81%</b>	<b>81%</b>

### Pembahasan

Pada perbandingan performa setiap model Tabel 1, terlihat bahwa model FastText dengan arsitektur Skip-gram memiliki akurasi sebesar 84%, presisi 82%, *recall* 81% dan *F1 score* sebesar 81%. Model FastText dengan arsitektur Skip-gram ini merupakan model dengan performa terbaik. Kemudian pada model Word2Vec dengan arsitektur Skip-gram memiliki akurasi sebesar 81,8%, presisi 80%, *recall* 78%, dan *F1 score* sebesar 79%. Hal ini menunjukkan bahwa FastText memiliki performa yang lebih bagus pada arsitektur Skip-gram daripada Word2Vec.

Pada *word embeddings* FastText dengan arsitektur CBOW, model yang dihasilkan mendapatkan akurasi sebesar 80,9%, presisi 79%, *recall* 76%, dan *F1 score* 77%. Kemudian pada *word embeddings* Word2Vec dengan arsitektur CBOW memiliki akurasi sebesar 80,3%, presisi 79%, *recall* 75%, dan *F1 score* 76%. Hal ini menunjukkan bahwa FastText juga memiliki performa yang lebih bagus pada arsitektur CBOW daripada Word2Vec meskipun tidak terlalu jauh.

Arsitektur Skip-gram sedikit lebih baik daripada arsitektur CBOW apabila dilihat dari performa model pada data uji, khususnya pada dataset dalam penelitian ini. Skip-gram dan CBOW memiliki tujuan berbeda. Skip-gram bertujuan untuk memprediksi konteks (*output*) di sekitar *current word* (*input*), sedangkan CBOW memprediksi kata (*output*) ketika diberikan konteks di sekitar kata tersebut (*input*). Skip-gram memiliki keunggulan yaitu dapat bekerja dengan baik pada data latih yang sedikit, selain itu Skip-gram dapat merepresentasikan kata yang dianggap langka. Sedangkan CBOW memiliki keunggulan yaitu memiliki akurasi yang sedikit lebih baik untuk *frequent words* dan memiliki waktu *training* yang lebih cepat dibandingkan dengan Skip-gram (Nurdin et al., 2020).

Arsitektur Skip-gram memiliki performa yang lebih baik karena dalam dataset pada penelitian ini hanya menggunakan 1621 data yang kemudian dibagi menjadi 1296 data latih dan 325 sebagai data uji. Meskipun *word embeddings* sudah terlebih dahulu dilatih menggunakan data korpus Wikipedia Bahasa Indonesia, namun pada dataset ini juga terdiri dari kata-kata *slang* atau kata-kata gaul yang jarang muncul. *Preprocessing* pada tahap *convert slangword* memang sudah dilakukan sebelumnya, tetapi kamus yang digunakan juga masih cukup terbatas sehingga meskipun sudah dilakukan tahap *convert slangword* sebagian kata *slang* masih ada.

Pada Tabel 1 juga menunjukkan bahwa FastText sedikit lebih unggul dari Word2Vec baik pada arsitektur CBOW maupun Skip-gram meskipun tidak terlalu jauh. Performa model yang dibangun menggunakan model FastText lebih bagus karena pada FastText dapat menangani kata *Out of Vocabulary* (OOV) dimana FastText membangun vektor dari *n*-gram karakter dari kata-kata yang ada dalam kamus. Sedangkan Word2Vec belum mampu menangani hal tersebut dan apabila dipaksakan akan menghasilkan eror.

## SIMPULAN

Berdasarkan penelitian yang dilakukan, diperoleh kesimpulan sebagai berikut.

1. Berdasarkan pengujian yang dilakukan, performa akurasi sistem analisis sentimen terkait tim nasional sepak bola indonesia dengan pendekatan *machine learning* dengan metode Support Vector Machine didapatkan akurasi mencapai 84%, presisi 82%, *recall* 81% dan *F1 score* sebesar 81%.
2. Metode *word embeddings* FastText memiliki peforma yang sedikit lebih baik daripada Word2Vec apabila digunakan untuk fitur ekstraksi pada analisis sentimen menggunakan Support Vector Machine, perbedaannya adalah FastText dapat mengenali kata-kata yang tidak ada dalam korpus sedangkan Word2Vec tidak dapat menangani hal tersebut. Kata-kata yang tidak di dalam korpus FastText akan membangun vektor dari subword hasil dekonstruksi kata yang lain. Pada analisis sentimen dalam penelitian ini, model terbaik dihasilkan dengan menggunakan *word embeddings* FastText dengan model Skip-gram.

## UCAPAN TERIMA KASIH

Terima kasih kepada koordinator Prodi Matematika dan seluruh Dosen Prodi Matematika yang telah memberikan ilmu dan bimbingan hingga terselesainya artikel ini.



## DAFTAR PUSTAKA

- Astiningrum, M., Haniah, M., & Pradana, Y. rahmat yoga. (2020). Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter. *Seminar Informatika Aplikatif Polinema (Siap)*, 35—39.
- Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. (2013). Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. *Procedia Engineering*, 53, 453–462. <https://doi.org/10.1016/j.proeng.2013.02.059>
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5, 135–146. [https://doi.org/10.1162/tacl\\_a\\_00051](https://doi.org/10.1162/tacl_a_00051)
- Hadiyan Sidik, M., Widiyanesti, S., & Puteri Ramadhani, D. (2022). *Analisis Sentimen dan Topic Modelling Terhadap Tim Nasional Indonesia di Kejuaraan AFF Suzuki Cup 2020 Berdasarkan Opini Pengguna Twitter Analysis of Sentiment and Topic Modeling of the Indonesian National Team in the 2020 AFF Suzuki Cup Championship Base* (Vol. 9).
- Ipsos. (2022). More than half of adults across 34 countries plan to watch the 2022 FIFA World Cup. *Ipsos.Com*. Diambil dari <https://www.ipsos.com/en-id/more-half-adults-across-34-countries-plan-watch-2022-fifa-world-cup>
- Joachims, T. (2005). Text categorization with support vector machines: Learning with many relevant features. In *Machine Learning: ECML-98: 10th European Conference on Machine Learning Chemnitz, Germany, April 21–23, 1998 Proceedings* (hal. 137–142). Springer.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, 1–12.
- Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support vector machine. *Proceeding Indones. Sci. Meeting Cent. Japan*.
- Nuridin, A., Aji, B. A. S., Bustamin, A., & Abidin, Z. (2020). Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks. *Jurnal Tekno Kompak*, 14(2), 74–79.
- Octaviani, P. A., Yuciana Wilandari, & Ispriyanti, D. (2014). Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang. *Jurnal Gaussian*, 3(8), 811–820. Diambil dari [http://download.portalgaruda.org/article.php?article=286497&val=4706&title=PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE \(SVM\) PADA DATA AKREDITASI SEKOLAH DASAR \(SD\) DI KABUPATEN MAGELANG](http://download.portalgaruda.org/article.php?article=286497&val=4706&title=PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG)
- Pratama, A. E., Ariesta, A., & Gata, G. (2022). Analisis Sentimen Masyarakat terhadap Tim Nasional Indonesia pada Piala AFF 2020 Menggunakan Algoritma K-Nearest Neighbors Aditya. *Jurnal TICOM: Technology of Information and Communication*, 10(3), 187–196.
- Ravi, K., & Ravi, V. (2015). *A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems* (Vol. 89). Elsevier B.V. <https://doi.org/10.1016/j.knosys.2015.06.015>
- Salsabila, N. A., Winatmoko, Y. A., Septiandri, A. A., & Jamal, A. (2018). Colloquial Indonesian Lexicon. In *2018 International Conference on Asian Language Processing (IALP)* (hal. 226–229). <https://doi.org/10.1109/IALP.2018.8629151>
- Social, W. A. (2022). Hootsuite (We are Social): Indonesian Digital Report.
- Suyanto, M. L. (2018). Tingkat Dasar dan Lanjut. *Informatika Bandung*.