

**PERBANDINGAN ALGORITMA CBOW DAN SKIP-GRAM  
PADA METODE WORD EMBEDDING FASTTEXT  
UNTUK SENTIMEN ANALISIS**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**TRI SUSILO**

**18.11.1823**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2022**

**PERBANDINGAN ALGORITMA CBOW DAN SKIP-GRAM PADA  
METODE WORD EMBEDDING FASTTEXT  
UNTUK SENTIMEN ANALISIS**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**Tri Susilo**

**18.11.1823**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2022**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA CBOW DAN SKIP-GRAM PADA  
METODE WORD EMBEDDING FASTTEXT  
UNTUK SENTIMEN ANALISIS**

yang disusun dan diajukan oleh

**Tri Susilo**

**18.11.1823**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 17 November 2022

**Dosen Pembimbing,**

**Dr. Arief Setyanto, S.Si., M.T**

**NIK. 190302036**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA CBOW DAN SKIP-GRAM PADA  
METODE WORD EMBEDDING FASTTEXT  
UNTUK SENTIMEN ANALISIS**

yang disusun dan diajukan oleh

**Tri Susilo**

**18.11.1823**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 17 November 2022

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Erni Seniwati, S.Kom, M.Cs**  
**NIK. 190302231**

**Mardhiva Hayaty, S.T., M.Kom.**  
**NIK. 190302108**

**Dr. Arief Setyanto, S.Si, M.T**  
**NIK. 190302xxx**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 17 November 2022

**DEKAN FAKULTAS ILMU KOMPUTER**

**Hanif Al Fatta, S.Kom., M.Kom.**  
**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

**Nama mahasiswa : Tri Susilo**  
**NIM : 18.11.1823**

Menyatakan bahwa Skripsi dengan judul berikut:

### **Perbandingan Algoritma CBOV dan Skip-gram Pada Metode Word Embedding Fasttext Untuk Sentimen Analisis**

Dosen Pembimbing : Dr. Arief Setyanto, S.Si., MT

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 17 November 2022

Yang Menyatakan,



Tri Susilo

## HALAMAN PERSEMBAHAN

Puji syukur kepada Allah SWT atas berkat dan rahmat yang telah diberikan kepada saya serta telah mengabulkan do'a saya sehingga saya dapat mencapai apa yang saya raih hingga saat ini. Oleh karena itu dengan rasa syukur dan bangga saya ucapkan terimakasih kepada :

1. Diri saya sendiri, yang telah mampu bertahan dan berjuang hingga dapat menyelesaikan studi dengan baik, walaupun dalam perjalanannya terdapat berbagai rintangan akhirnya diri ini mampu untuk melewati itu semua hingga pada titik ini.
2. Kedua Orang tua saya, Bapak, Ibu yang selalu mendidik, pendukung, dan mendoakan saya sehingga saya menjadi pribadi yang dewasa dan dapat menyelesaikan pendidikan hingga saat ini.
3. Bapak Ibu Dosen pembimbing, penguji dan pengajar yang selalu mencurahkan waktunya untuk selalu mengamalkan ilmu, memberikan pengalaman dan mengarahkan saya hingga saya dapat mencerahkan hidup saya. Semoga ilmu yang telah diajarkan menjadi ladang amal dan semoga menjadi ilmu yang berkah untuk kami.

## KATA PENGANTAR

Assalamualaikum Wr. Wb

Puji syukur selalu tercurahkan kepada Ilahi Rabbi Allah Swt. Yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat melaksanakan dan menyusun skripsi ini dengan baik. Shalawat dan salam semoga senantiasa tercurahkan kepada kepada junjungan kita Nabi Muhammad SAW beserta keluarga, para sahabat dan para pengikutnya seluruh umat Islam.

Perjalanan Panjang telah dilalui penulis dalam rangka menyelesaikan skripsi ini dengan judul “Perbandingan Algoritma CBOW dan Skip-gram Pada Metode Word Embedding Fasttext Untuk Sentimen Analisis”. Banyak hambatan yang dialami dalam penyusunannya, namun berkat kehendak-Nya penulis berhasil menyelesaikan skripsi ini. Oleh karena itu, dengan penuh kerendahan hati, pada kesempatan ini pattutlah kiranya penulis mengucapkan terima kasih kepada :

1. Kedua orang tua saya serta keluarga besar yang telah mendoakan, memberikan semangat dan memberikan memotivasi kepada saya.
2. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta
3. Hanif Al Fatta, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Ibu Windha Mega Pradnya D, M.Kom. selaku Ketua Program Studi S1- Informatika Universitas Amikom Yogyakarta.
5. Bapak Dr. Arief Setyanto, S.Si., M.T\_ selaku dosen pembimbing yang selalu bijaksana memberikan bimbingan serta waktunya selama penulisan skripsi ini.
6. Bapak dan Ibu Dosen Universitas Amikom Yogyakarta serta Staff dan Karyawan Universitas Amikom Yogyakarta.
7. Teman-teman kost Laundry Spot yang selalu memberikan dukungan selama pengerjaan skripsi
8. Karyawan serta Student Staff UPT yang menemani dan memberikan dukungan selama pengerjaan skripsi.

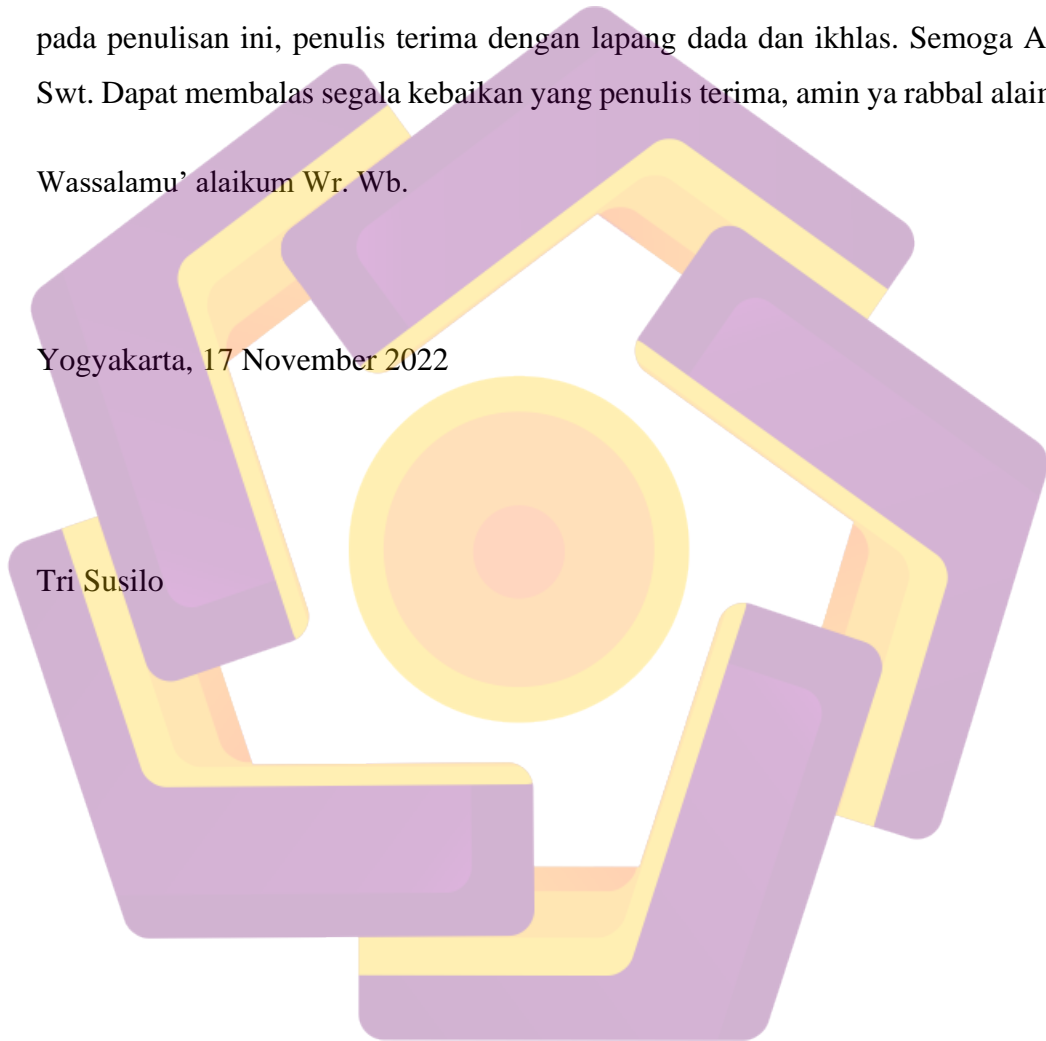
9. Serta teman-teman kelas 18 informatika 01 yang telah memberikan pengalaman yang mengesankan selama masa kuliah.

Penulis menyadari betul bahwa penulisan skripsi ini masih jauh dari kesempurnaan, walau bagaimana pun penulis berusaha memberikan yang terbaik dari ketidaksempurnaan yang ada. Demikian segala saran dan kritik yang tertuju pada penulisan ini, penulis terima dengan lapang dada dan ikhlas. Semoga Allah Swt. Dapat membalas segala kebaikan yang penulis terima, amin ya rabbal alamin.

Wassalamu' alaikum Wr. Wb.

Yogyakarta, 17 November 2022

Tri Susilo





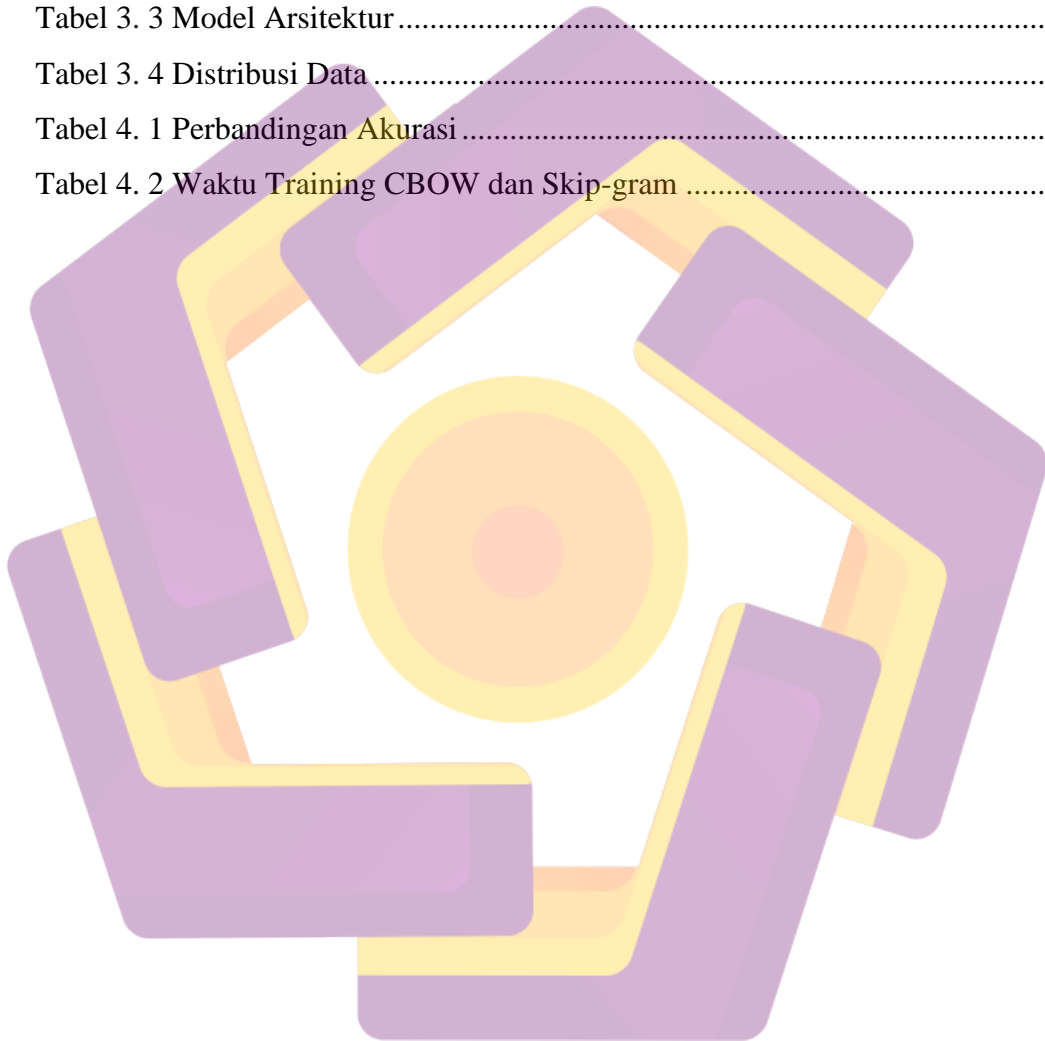
## DAFTAR ISI

|   |      |
|---|------|
| HALAMAN JUDUL.....                        | 1    |
| HALAMAN PERSETUJUAN.....                  | ii   |
| HALAMAN PENGESAHAN.....                   | iii  |
| HALAMAN PERNYATAAN KEASLIAN SKRIPSI ..... | iv   |
| HALAMAN PERSEMBAHAN .....                 | v    |
| KATA PENGANTAR .....                      | vi   |
| DAFTAR ISI.....                           | viii |
| DAFTAR TABEL.....                         | x    |
| DAFTAR GAMBAR .....                       | xi   |
| DAFTAR ISTILAH .....                      | xii  |
| INTISARI.....                             | xiii |
| ABSTRACT.....                             | xiv  |
| BAB I PENDAHULUAN.....                    | 1    |
| 1.1 Latar Belakang .....                  | 1    |
| 1.2 Rumusan Masalah .....                 | 2    |
| 1.3 Batasan Masalah.....                  | 2    |
| 1.4 Tujuan Penelitian.....                | 2    |
| 1.5 Manfaat Penelitian.....               | 3    |
| 1.6 Metode Penelitian.....                | 3    |
| 1.7 Sistematika Penulisan.....            | 3    |
| BAB II LANDASAN TEORI .....               | 5    |
| 2.1 Tinjauan Pustaka .....                | 5    |

|  |  |           |
|--|--|-----------|
| 2.2                                      | Landasan Teori .....                           | 7         |
| 2.2.1                                    | <i>Natural Language Processing (NLP)</i> ..... | 7         |
| 2.2.2                                    | Sentimen Analisis .....                        | 7         |
| 2.2.3                                    | Word2vec .....                                 | 8         |
| 2.2.4                                    | Fasttext .....                                 | 8         |
| 2.2.5                                    | Long Short Term Memory (LSTM).....             | 14        |
| <b>BAB III METODE PENELITIAN.....</b>    |  | <b>17</b> |
| 3.1                                      | Objek Penelitian .....                         | 17        |
| 3.2                                      | Alur Penelitian.....                           | 17        |
| 3.2.1                                    | <i>Preprocessing</i> .....                     | 18        |
| 3.2.2                                    | <i>Convert to Numeric</i> .....                | 20        |
| 3.2.3                                    | <i>Training CBOW and Skip-gram</i> .....       | 20        |
| 3.2.4                                    | Dataset Splitting .....                        | 24        |
| 3.2.5                                    | Model LSTM.....                                | 25        |
| 3.2.6                                    | <i>Testing</i> .....                           | 25        |
| 3.3                                      | Data dan Alat.....                             | 26        |
| 3.3.1                                    | Data Penelitian .....                          | 26        |
| 3.3.2                                    | Alat/instrumen.....                            | 26        |
| <b>BAB IV HASIL DAN PEMBAHASAN .....</b> |  | <b>27</b> |
| 4.1                                      | Hasil Pengujian dan Pembahasan .....           | 27        |
| <b>BAB V KESIMPULAN .....</b>            |  | <b>35</b> |
| 5.1                                      | Kesimpulan.....                                | 35        |
| 5.2                                      | Saran.....                                     | 35        |
| Daftar Pustaka .....                     |  | 36        |

## DAFTAR TABEL

|   |    |
|---|----|
| Tabel 2. 1 Tinjauan Literatur .....                   | 7  |
| Tabel 3. 1 Tunning Parameter CBOW dan Skip-gram ..... | 20 |
| Tabel 3. 2 <i>Tunning Parameter</i> .....             | 25 |
| Tabel 3. 3 Model Arsitektur .....                     | 25 |
| Tabel 3. 4 Distribusi Data .....                      | 26 |
| Tabel 4. 1 Perbandingan Akurasi .....                 | 27 |
| Tabel 4. 2 Waktu Training CBOW dan Skip-gram .....    | 29 |



## DAFTAR GAMBAR

|  |    |
|--|----|
| Gambar 2. 1 Arsitektur Skip-gram.....                            | 9  |
| Gambar 2. 2 ilustrasi Windowing Skip-gram .....                  | 9  |
| Gambar 2. 3 one-hot encoded vector .....                         | 10 |
| Gambar 2. 4 Feedfoward Neural Network Skip-gram.....             | 11 |
| Gambar 2. 5 Arsitektur CBOW.....                                 | 12 |
| Gambar 2. 6 Ilustrasi Windowing CBOW .....                       | 12 |
| Gambar 2. 7 Feedfoward Neural Network CBOW.....                  | 13 |
| Gambar 2. 8 Struktur LSTM.....                                   | 14 |
| Gambar 3. 1 Alur Penelitian.....                                 | 18 |
| Gambar 3. 2 Ilustrasi Skip-gram .....                            | 21 |
| Gambar 3. 3 one-hot encoding .....                               | 21 |
| Gambar 3. 4 matriks $W$ dan $W'$ .....                           | 22 |
| Gambar 3. 5 Ilustrasi CBOW .....                                 | 23 |
| Gambar 3. 6 Nilai $W$ CBOW .....                                 | 24 |
| Gambar 3. 7 Nilai $W'$ CBOW .....                                | 24 |
| Gambar 4. 1 Grafik Akurasi CBOW dan Skip-gram.....               | 27 |
| Gambar 4. 2 Grafik Perbandingan Waktu.....                       | 29 |
| Gambar 4. 3 Grafik Training CBOW dan Skip-gram Dimensi 50 .....  | 30 |
| Gambar 4. 4 Grafik Training CBOW dan Skip-gram Dimensi 100 ..... | 31 |
| Gambar 4. 5 Grafik Training CBOW dan Skip-gram Dimensi 150 ..... | 32 |
| Gambar 4. 6 Grafik Training CBOW dan Skip-gram Dimensi 200 ..... | 33 |
| Gambar 4. 7 Grafik Training CBOW dan Skip-gram Dimensi 300 ..... | 34 |

## DAFTAR ISTILAH

|                           |  |
|---------------------------|--|
| <i>Case Folding</i>       | Mengubah huruf seluruh data dalam bentuk huruf kecil.  |
| <i>Data Cleanning</i>     | Proses untuk memastikan kebenaran,konsistensi, dan kegunaan suatu data yang ada dalam <i>dataset</i> .   |
| <i>Tokenization</i>       | adalah proses membagi teks yang berupa kalimat menjadi kepingan-kepingan kata.   |
| <i>Data Normalization</i> | Merubah kata-kata yang tidak baku atau slang menjadi kata yang baku.   |
| <i>Stemming</i>           | Proses mengubah infleksi kata kebentuk kata dasarnya.  |
| <i>Stopword</i>           | Kumpulan kata yang memiliki frekuensi kemunculan tinggi namun kurang memiliki makna.   |
| <i>Vector Size</i>        | Ukuran dimensi vektor yang mewakili setiap token atau kata.  |
| <i>Window</i>             | Jarak maksimum antara kata target dengan kata disekitarnya. Jika nilai window yang digunakan besar, maka semakin Banyak kata yang terkait. Dan apabila semakin kecil nilai Window,maka akan semakin spesifik terhadap suatu konteks kata . |
| <i>Minimal Count</i>      | Frekuensi minimal kemunculan kata yang akan dirubah kedalam bentuk vektor.   |
| <i>Worker</i>             | Jumlah <i>logical core</i> yang digunakan saat melatih <i>word embedding</i> .   |
| <i>Epoch</i>              | adalah jika semua training data kita sudah selesai di train. Satu epoch tercapai ketika semua batch selesai melewati jaringan saraf .  |
| <i>Batch_Size</i>         | Jumalh training sample yang digunakan yang digunakan Dalam satu batch untuk sekali iterasi   |
| $\pm$                     | Nilai standar deviasi  |

## INTISARI

Sentimen analisis merupakan salah satu bidang dari NLP yang digunakan untuk klasifikasi data tekstual. Pada proses analisis sentimen terdapat suatu proses untuk mengubah teks menjadi vektor, metode yang paling sering digunakan untuk mengubah teks menjadi vektor adalah *Bag of Words*. Metode BoW ini sendiri memiliki kelemahan seperti hilangnya konteks dari kalimat, urutan kata, dan hubungan semantik antar kata dan membutuhkan ukuran penyimpanan yang besar. *Fasttext* hadir untuk mengatasi masalah tersebut dengan konsep *word to vector*, dan mempunyai kemampuan untuk menangani kata-kata yang belum pernah muncul sebelumnya.

Penelitian ini menggunakan metode *CBOW* dan *Skip-gram* pada *fasttext* untuk membentuk *word vector* di corpus. Beberapa ukuran dimensi dari word vector yaitu dimensi 50,100,150,200, dan 300 digunakan untuk mengetahui pengaruhnya terhadap akurasi yang dihasilkan.

Akurasi terbaik diperoleh pada ukuran dimensi *word vector* 300 dengan akurasi 82,5% dan akurasi terendah sebesar 74,8% pada ukuran dimensi word vektor 50. Untuk ukuran dimensi *word vector* 50,150,200 dan 300 metode *Skip-gram* lebih unggul dibandingkan metode *CBOW*. Sedangkan metode *CBOW* unggul di ukuran dimensi *word Vector* 100 dengan hasil akurasi 77,5% dan 74,9% untuk metode *Skip-gram*.

**Kata Kunci:** Word Embedding, Fasttext, LSTM, CBOW, Skip-gram, Sentimen Analisis, Teks Indonesia.

## ABSTRACT

*Sentiment analysis is a field of NLP that is used for textual data classification. In the process of sentiment analysis there is a process for converting text into vectors, the most frequently used method for converting text into vectors is the Bag of Words. The BoW method itself has weaknesses such as loss of context from sentences, word order, and semantic relationships between words and requires a large storage size. Fasttext is here to solve this problem with the word to vector concept, and has the ability to handle words that have never appeared before.*

*This study uses the CBOW and Skip-gram methods on fasttext to form word vectors in the corpus. Several dimensions of the word vector, namely dimensions of 50,100,150,200, and 300 are used to determine the effect on the resulting accuracy.*

*The best accuracy was obtained on a word vector dimension of 300 with an accuracy of 82.5% and the lowest accuracy was 74.8% on a word vector dimension of 50. For a word vector dimension of 50,150,200 and 300 the Skip-gram method is superior to the CBOW method. Whereas the CBOW method excels in the Vector 100 word dimension with an accuracy of 77.5% and 74.9% for the Skip-gram method.*

**Keywords:** Word Embedding, Fasttext, LSTM, CBOW, Skip-gram, Sentiment Analysis, Indonesian Text.