

# EKSTRAKSI FITUR MENGGUNAKAN MODEL *WORD2VEC* PADA *SENTIMENT ANALYSIS* KOLOM KOMENTAR KUISIONER EVALUASI DOSEN OLEH MAHASISWA

Muhammad Rusli<sup>1</sup>, Mohammad Reza Faisal<sup>2</sup>, Irwan Budiman<sup>3</sup>,  
Radityo Adi Nugroho<sup>4</sup>, Andi Farmadi<sup>5</sup>  
Prodi Ilmu Komputer FMIPA ULM  
Jl. A. Yani Km 36 Banjarbaru, Kalimantan Selatan  
Email: [rusli150397@gmail.com](mailto:rusli150397@gmail.com)

## *Abstract*

*This research is about Sentiment Analysis using the Word2vec model. this research was conducted by Fauzi (2019). But in his research the use of the Word2vec model produces an accuracy of 70%, because the data used is small. In little data Word2vec cannot grasp the similarity of meaning well. So that related research was conducted which used lecturer evaluation comment data and also Wikipedia article data in Indonesian language as Word2vec model. In this study a comparison of average extraction features of Word2vec and Bag of Centroid base Word2vec was done and a combination of the two was then performed using the Support Vector Machine method. The application of Word2vec Average base feature extraction in the lecturer evaluation commentary data resulted in an accuracy of 84,8%. Then using the Bag of Centroid base feature extraction using Word2vec Hierarchy Clustering produces the best accuracy of 81,6% with a total of 75 features. The result of merging the two feature extractions produces an accuracy of 85,3%.*

**Keywords:** *Sentiment Analysis, Word2vec, Feature extraction*

## *Abstrak*

*Penelitian ini mengenai Sentiment Analysis menggunakan model Word2vec. penelitian ini pernah dilakukan oleh Fauzi (2019). Namun pada penelitiannya penggunaan model Word2vec menghasilkan akurasi 70%, karena data yang digunakan sedikit. Dalam data yang sedikit Word2vec tidak dapat menangkap kemiripan makna dengan baik. Sehingga dilakukan penelitian terkait yang mana menggunakan data komentar evaluasi dosen dan juga data artikel Wikipedia berbahasa Indonesia sebagai model Word2vec. Pada penelitian ini dilakukan perbandingan ekstraksi fitur Average base Word2vec dan Bag of Centroid base Word2vec dan juga dilakukan penggabungan keduanya kemudian dilakukan klasifikasi menggunakan metode Support Vector Machine. Penerapan ekstraksi fitur Average base Word2vec pada data komentar evaluasi dosen menghasilkan akurasi sebesar 84,8%. Kemudian menggunakan ekstraksi fitur Bag of Centroid base Word2vec menggunakan Hirarki Clustering menghasilkan akurasi terbaik sebesar 81,6% dengan jumlah 75 fitur. Hasil penggabungan kedua ekstraksi fitur menghasilkan akurasi sebesar 85,3%.*

**Kata kunci:** *Sentiment Analysis, Word2vec, Ekstraksi fitur.*

## 1. PENDAHULUAN

*Sentiment Analysis* merupakan cabang penelitian dari *Text Mining* yang berguna untuk mengklasifikasi dokumen teks berupa opini berdasarkan sentimen. Tugas dasar *Sentiment Analysis* adalah mengklasifikasikan beberapa teks dari dokumen, kalimat atau fitur, kalimat dari fitur tersebut bisa bersifat positif, negatif dan netral. Pada tahapan *Sentiment Analysis* terdapat ekstraksi fitur yang berguna mengubah teks menjadi data yang dapat diklasifikasi.

Ada banyak cara dalam ekstraksi fitur seperti TF, TF-IDF, *Lexical Description*, Unigram, Bigram dan sebagainya. Pada beberapa penelitian *Word2vec* digunakan dalam melakukan ekstraksi fitur. *Word2vec* merupakan algoritma *word embedding*, yaitu pemetaan dari kata menjadi *vector*. *Vector* ini nantinya digunakan untuk berbagai macam tugas *Natural Language Processing*. Pada *Word2vec* dalam pembobotan kata digunakan nilai rata-rata *vector* yang mewakili kata tersebut. Pada penelitian Fauzi [1] mengenai *Sentiment Analysis* menggunakan model *Word2vec* menghasilkan akurasi 70%, hal ini karena data yang digunakan dalam data pelatihan *Word2vec* sedikit. Dalam data yang kecil, *Word2vec* tidak dapat menangkap kata-kata yang mirip dengan baik.

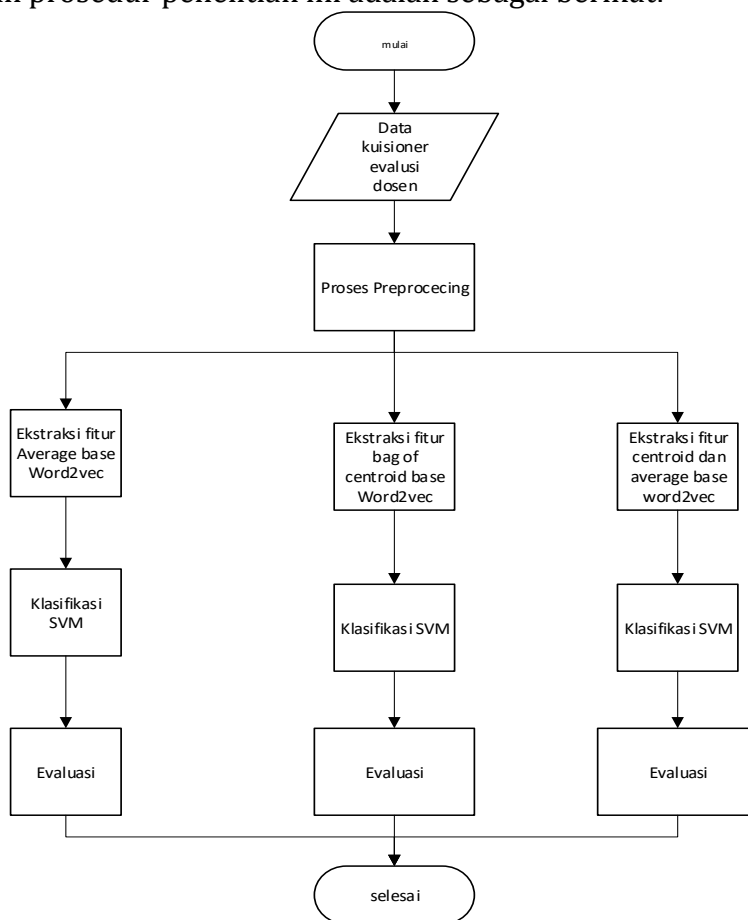
Rossiello [2], pada penelitiannya mengusulkan metode berbasis *centroid* untuk peringkasan teks dalam eksploitasi kemampuan *Word Embedding*. Pada penelitian tersebut menggunakan metode *Word2vec* dalam mempelajari hubungan semantik antar kata, kemudian dikluster menjadi beberapa *centroid* yang mana akan terkumpul berdasarkan kemiripan makna yang dianggapnya sama. Pada penelitian tersebut menghasilkan kinerja yang lebih baik dibandingkan dengan model pembelajaran yang lebih kompleks.

Berdasarkan penelitian Rossiello [2], perlu dilakukan penelitian terkait menggunakan *centroid base Word2vec* dalam ekstraksi fitur yang akan digunakan dalam klasifikasi *Sentiment Analysis* yang mana akan kita bandingkan tingkat akurasi dengan ekstraksi fitur menggunakan *Average base Word2vec*. Pada penelitian mengenai ekstraksi fitur kadang dilakukan penggabungan beberapa proses ekstraksi fitur untuk mendapatkan informasi yang lebih lengkap, sehingga perlu dilakukan penelitian juga bagaimana tingkat akurasi ekstraksi fitur dalam *Sentiment Analysis* jika *Bag of Centroid base Word2vec* dan *Average base Word2vec* digabung dalam melakukan ekstraksi fitur.

Pada penelitian ini digunakan metode SVM dalam klasifikasi *Sentiment Analysis* karena pada beberapa penelitian sebelumnya menghasilkan tingkat akurasi yang sangat baik. Untuk data dalam pembelajaran *Word2vec* digunakan corpus Wikipedia dan isi komentar pada kuisisioner evaluasi dosen yang ada pada portal mahasiswa FMIPA ULM tahun ajaran 2017/2018 semester genap.

## 2. METODOLOGI PENELITIAN

Adapun prosedur penelitian ini adalah sebagai berikut:



Gambar 1. Alur Penelitian

### 2.1 Pengumpulan Data

Data yang digunakan pada penelitian ini yaitu data kolom komentar evaluasi dosen oleh mahasiswa semester genap tahun ajaran 2017/2018 FMIPA ULM dengan jumlah 9154 data.

### 2.2 Preprocessing

Dari semua data, teks adalah bentuk yang paling tidak terstruktur sehingga kita harus melakukan banyak pembersihan. Langkah-langkah *preprocessing* ini membantu mengubah noise dari fitur dimensi tinggi ke ruang dimensi rendah untuk memperoleh informasi seakurat mungkin dari teks. Yang mana langkah *preprocessing* sebagai berikut :

#### a. Data Cleaning

Proses mendeteksi dan memperbaiki atau menghapus data yang rusak atau tidak akurat agar dapat menambah akurasi proses klasifikasi.

#### b. Removing

Menghapus karakter-karakter yang tidak berkontribusi pada *sentiment analysis* sehingga hanya menyisakan karakter alfabet.

c. *Case folding*

Mengubah setiap kata menjadi kedalam bentuk yang standar, huruf kapital (*upper case*) atau huruf kecil (*lower case*).

d. Ekstraksi *Fitur*

Proses mengubah data teks yang tidak terstruktur menjadi data yang lebih terstruktur, sehingga data tersebut dapat diklasifikasi.

### 2.3 *Data Mining*

Data mining dilakukan untuk menemukan suatu pola atau informasi dari data yang tersimpan dalam jumlah yang besar dengan menggunakan teknik atau metode tertentu. Pada penelitian ini dilakukan klasifikasi menggunakan *Support Vector Machine* dan ekstraksi fitur menggunakan model *Word2vec* untuk *Sentiment Analysis*.

### 2.5 *Evaluation*

*Evaluation* merupakan tahapan penilaian hasil dari teknik *data mining* yang telah dilakukan sebelumnya.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil

#### 3.1.1 Pengumpulan Data

Data yang digunakan pada penelitian ini yaitu data kolom komentar evaluasi dosen oleh mahasiswa semester genap tahun ajaran 2017/2018 FMIPA ULM dengan jumlah 9154 data. Selain data komentar juga digunakan data artikel Wikipedia yang berjumlah 36000 data. Data komentar dan data artikel digunakan sebagai korpus pada model *Word2vec*.

#### 3.1.2 *Preprocessing*

*Preprocessing* pada penelitian ini yaitu penghapusan simbol atau *removing*, *Case Folding*, *data cleaning*. Penghapusan simbol (*Removing*) merupakan proses menghapus karakter karakter yang tidak perlu. Penghapusan simbol-simbol berfungsi untuk menghapus karakter khusus dalam komentar seperti tanda baca (seperti: koma (,), titik(.), tanda tanya (?), tanda seru (!) dan sebagainya), angka numerik (0 - 9), dan karakter lainnya(seperti: \$, %, \*, dan sebagainya). Setelah melalui tahapan *removing*, tahapan *preprocessing* selanjutnya yaitu *case folding*. *Case folding* adalah tahapan mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk yang standar. Pada penelitian ini *case folding* digunakan untuk mengubah semua kata menjadi huruf kecil (*lowercase*). Kemudian tahapan *preprocessing* selanjutnya yaitu *data cleaning*. *Data cleaning* pada penelitian ini yaitu menghapus data yang tidak berkontribusi pada saat klasifikasi agar meningkatkan hasil akurasi. Adapun *data cleaning* disini menghapus komentar yang sama, menghapus komentar yang hanya karakter dan tidak menunjukkan sentimen. Setelah data melalui tahapan *preprocessing* diperoleh data sebanyak 2823 data komentar. Dari 2823 data komentar, terdapat 1242 data komentar positif dan juga 1581 data komentar negatif.

### 3.1.3 Pembuatan Model *Word2vec*

Pembuatan model *word2vec* ini berguna untuk menjadikan setiap kata pada data yang kita miliki menjadi *vector*. *Vector* pada model *word2vec* dapat ditentukan jumlah fiturnya. Pada penelitian ini dibuat model *word2vec* dengan jumlah fitur 100, karena default dari *word2vec* dengan jumlah fitur 100.

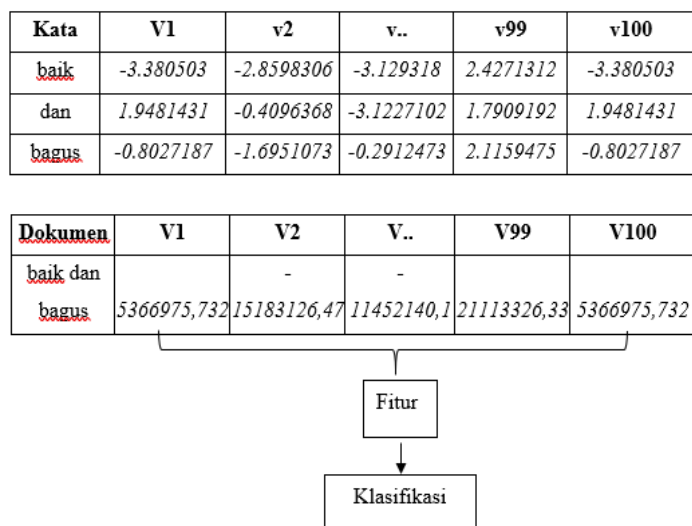
Data artikel Wikipedia dalam bentuk data *.txt*, Kemudian dengan bantuan package *Gensim* pada bahasa pemrograman *Python* terbantu model *word2vec* dengan jumlah fitur 100. Adapun contoh model yang dihasilkan pada tabel 1.

Tabel 1. *Vector* model *Word2vec*

Kata	v1	v..	v100
baik	-3.380503	-2.8598306	2.4271312
dan	1.9481431	-0.4096368	1.7909192
bagus	-0.8027187	-1.6951073	2.1159475
semoga	-2.2682748	-0.8437226	-1.4933019
lebih	2.2986956	-1.6426915	-2.0371377
lagi	0.8926442	0.8561693	0.39170513

### 3.1.4 Ekstraksi Fitur *Average base Word2vec*

Ekstraksi fitur menggunakan *Average base Word2vec* setiap komentar terdiri dari kata-kata, setiap kata memiliki nilai *vector*, nilai *vector* dari masing-masing kata dari setiap komentar dilakukan rata-rata yang mana sehingga setiap komentar memiliki 100 fitur yang sama. Hasil rata-rata *vector* tersebut akan mewakili fitur komentarnya saat proses klasifikasi Adapun contoh proses ekstraksi fitur *Average base Word2vec* pada gambar 1.

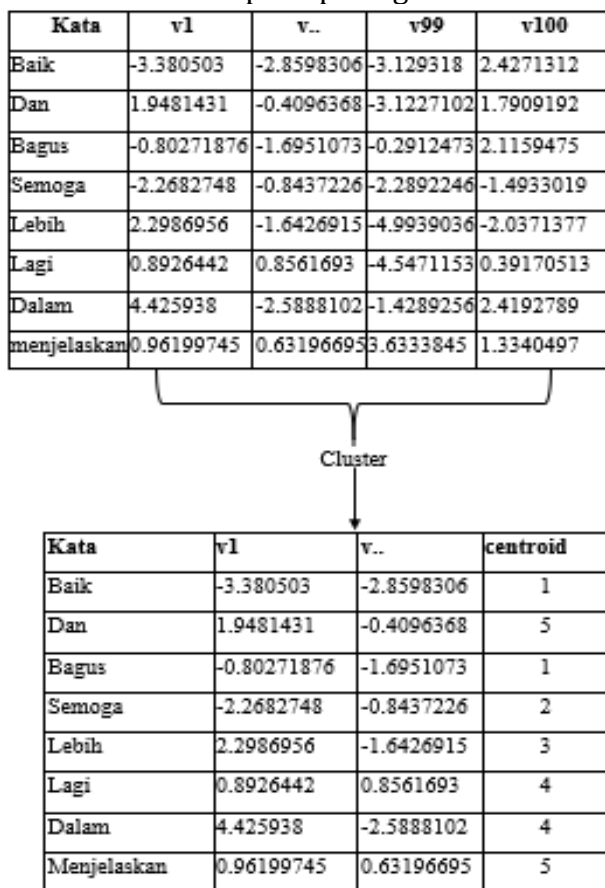


Gambar 1. Proses *average* berbasis model *Word2vec*

Ekstraksi fitur menggunakan *Average base Word2vec* pada analisis sentimen komentar ini memiliki tingkat akurasi sebesar 84,8%.

### 3.1.4 Ekstraksi fitur menggunakan *Bag of Centroid base Word2vec*.

Pada ekstraksi fitur menggunakan *Bag of Centroid base Word2vec* setiap kata kita *cluster* menggunakan algoritma *clustering*. Pada penelitian ini digunakan dua algoritma clustering yang mana *K-Means* dan *Hierarchical Clustering*. Dari setiap kata yang pada data komentar, dilakukan *cluster* sehingga mendapatkan label *cluster*. Tahapan cluster vector Word2vec seperti pada gambar 2.



Gambar 2. Pengelompokan kata berdasarkan cluster

Dari kata-kata yang ada pada komentar kita *cluster* sehingga mendapatkan label berdasarkan *centroid* yang terkumpul. Adapun hasil kata dengan label centroid terdapat pada tabel 2.

Tabel 2. Contoh kata dengan label *centroid*

Kata	<i>Centroid</i>
Baik	1
Dan	5
Bagus	1
Agar	3
Jangan	2
Terlalu	3
Cepat	1
Dalam	2
Menjelaskan	3
Soal	4

Kemudian dari data komentar setiap katanya dikelompokkan berdasarkan *centroid* yang telah ada. Jika jumlah *cluster* 5 maka jumlah fitur juga 5, adapun hasil ekstraksi fitur *Bag of Centroid base Word2vec* disajikan pada tabel 3.

Tabel 3. Contoh hasil ekstraksi fitur menggunakan *Bag of Centroid base Word2vec*

Komentar	C1	C2	C3	C4	C5	Label
Baik dan bagus	2	0	0	0	1	Positif
Agar jangan terlalu cepat dalam menjelaskan soal soal	1	2	3	2	0	Negatif

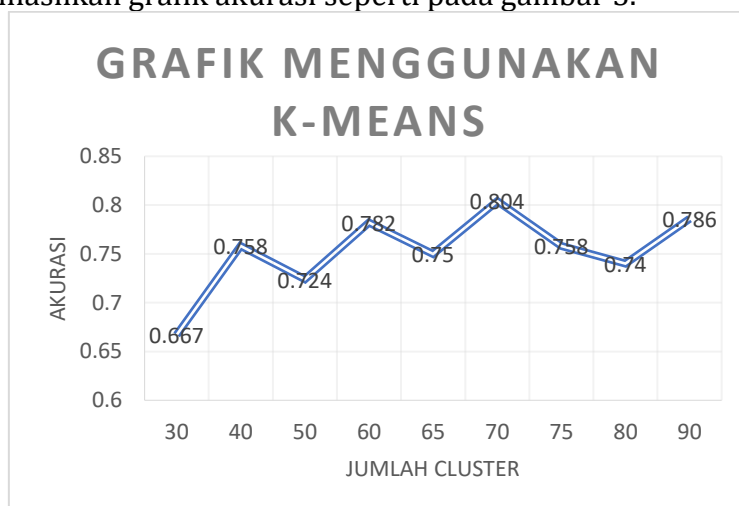
Kata “baik” berada pada *centroid* 1 maka C1 yang awalnya 0 ditambah 1, kata “dan” berada pada *centroid* 5 nilai C1 ditambah 1, kemudian kata bagus berada pada C1 sehingga nilai C1 yang awalnya 1 bertambah 1 sehingga nilai C1 =2. Begitu seterusnya hingga 2823 komentar menjadi fitur yang terstruktur.

Hasil akurasi ekstraksi fitur *Bag of Centroid base Word2vec* menggunakan algoritma *K-Means* dengan jumlah fitur yang bervariasi disajikan pada tabel 4.

Tabel 4. Hasil akurasi dari ekstraksi fitur *Bag of Centroid base Word2vec* menggunakan *K-Means*

Jumlah fitur	Akurasi
30	0,667
40	0,758
50	0,724
60	0,782
65	0,75
70	0,804
75	0,758
80	0,74
90	0,768

Dari tabel 4 dihasilkan grafik akurasi seperti pada gambar 3.



Gambar 3. Grafik akurasi menggunakan *K-Means*

Dalam menggunakan *K-Means* diperoleh akurasi terbaik saat jumlah *cluster* 70 dengan hasil akurasi 0,804. Dari hasil grafik pada gambar 3, jumlah *cluster* atau *centroid* tidak berpengaruh banyak dalam hasil kinerja klasifikasi, hal ini

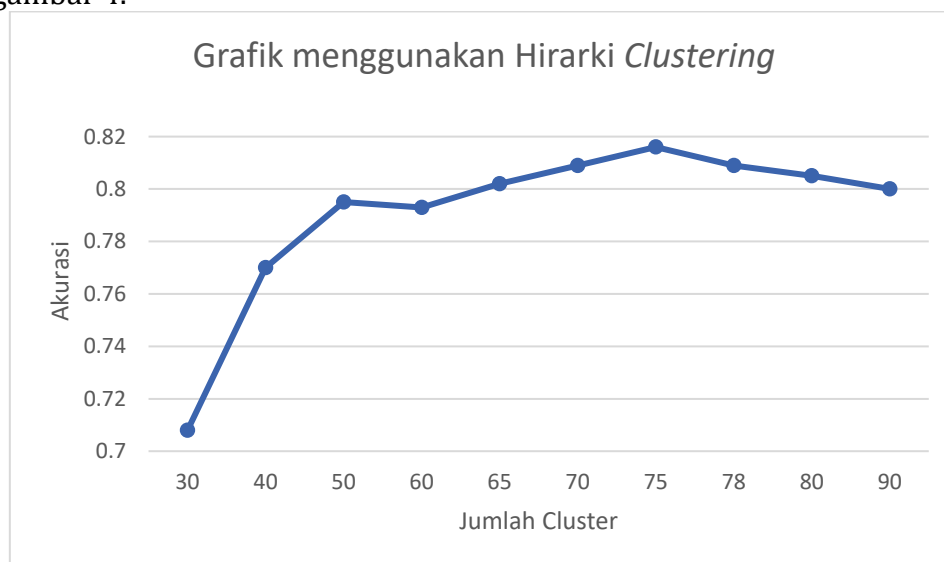
kemungkinan karena pada algoritma *K-Means* dalam penentuan nilai *K* awal acak sehingga nilai *cluster* pada setiap prosesnya berbeda-beda. Nilai *cluster* pada proses ini sangat berpengaruh besar dalam menentukan akurasi klasifikasi.

Karena menggunakan algoritma *K-Means* hasil akurasinya terbilang kurang terstruktur, dalam penelitian ini dilakukan percobaan menggunakan algoritma *clustering* lain yaitu Hirarki *Clustering* dengan metode ward. Hasil menggunakan algoritma Hirarki *Clustering* dengan fitur yang bervariasi terdapat pada tabel 5.

Tabel 5. Hasil akurasi dari ekstraksi fitur *Bag of Centroid base Word2vec* menggunakan Hirarki *Clustering*

Jumlah Cluster	Akurasi
30	0,708
40	0.77
50	0.795
60	0.793
65	0.802
70	0.809
75	0.816
78	0.809
80	0.805
90	0.8

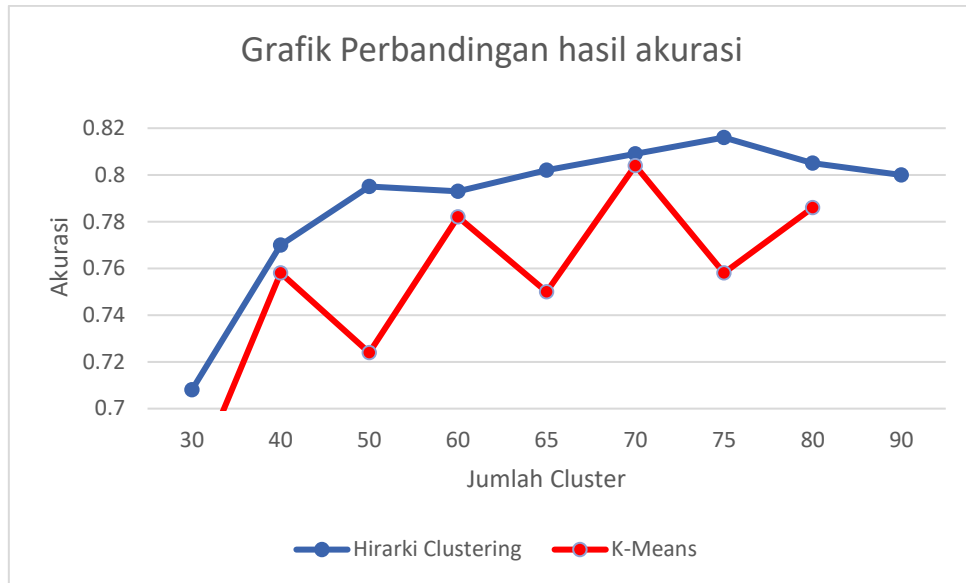
Berdasarkan tabel 5 diperoleh bahwa jumlah fitur sangat berpengaruh besar terhadap akurasi saat proses klasifikasi dan menghasilkan grafik akurasi seperti pada gambar 4.



Gambar 4. Grafik akurasi menggunakan Hirarki *Clustering*

Dengan jumlah fitur yang sedikit menghasilkan akurasi yang rendah. Maka ditambah jumlah fiturnya dan didapatkan akurasi tertinggi pada saat jumlah fitur 75. Hal ini karena pada jumlah fitur 75 kata-kata terkumpul dengan baik, sehingga pada saat *sentiment analysis* pengenalan pola kata positif dan negatif lebih mudah dipelajari. Adapun hasil perbandingan kinerja algoritma *Clustering* terdapat pada gambar 5.





Gambar 5. Perbandingan kinerja algoritma *Clustering*

### 3.1.5 Penggabungan *Average base Word2vec* dan *Bag of Centroid base Wor2vec*

Pada proses ekstraksi fitur ini data fitur *Average base Word2vec* digabungkan dengan data fitur *Bag of Centroid base Word2vec* dengan hasil akurasi terbaik. Adapun fitur terbaik dari *Bag of Centroid base Word2vec* menggunakan algoritma *K-Means* 70 dan *Hirarki Clustering* saat 75 fitur. Contoh penggabungan kedua ekstraksi fitur terdapat pada tabel 6.

Tabel 6. Data ekstrksi fitur penggabungan *Average base Word2vec* dan *Bag of Centroid base Wor2vec*

v1	v99	v100	c1	c2	c..	C69	C70	label
-0.6662942866484324	1.462.751.516.699.790	17.697.981.670.498.800	4	1	0	0	1	1
0.7452152929104426	16.522.573.049.125.400	3.556.856.980.328.170	14	2	2	0	13	0
0.2764149613391895	5.972.464.731.679.510	8.642.235.655.051.010	4	0	1	0	8	0
0.07762888669967652	1.728.857.127.428.050	21.251.153.630.018.200	4	0	0	0	1	1
0.3831836887324849	17.952.925.089.746.700	38.554.589.971.899.900	15	0	2	0	10	1
0.630289527380632	3.393.151.644.669.050	6.310.661.762.952.800	7	0	1	0	9	1
-10.120.803.117.752.0	466.977.499.127.388	5.155.191.433.429.710	1	0	1	0	0	1
0.4611934511100544	4.061.025.979.106.910	6.489.902.796.442.890	9	2	0	2	4	1
0.9527595524604504	19.967.673.039.422.000	4.416.797.601.546.230	10	0	0	0	14	1
18.061.231.261.549.000	63.037.742.853.451.200	9.194.977.505.848.950	6	0	1	0	6	0
0.3543612376919815	5.707.253.674.976.520	7.219.869.456.120.890	3	0	2	1	8	0
0.4684461589902639	29.633.056.064.136.300	58.832.329.908.385.800	7	1	1	0	9	0
0.30228009703569114	11.560.720.800.654.900	13.420.934.431.254.800	2	0	1	0	5	0
-0.1886933809992942	67.772.256.000.475.400	9.275.476.082.482.120	6	1	1	0	3	0
0.7722111276927448	28.048.856.489.005.500	5.721.197.124.038.420	11	0	5	0	5	0
0.17475399499138194	15.046.192.423.750.900	17.326.621.236.590.000	5	0	0	0	1	0
0.14292659109522554	0.3473603802949897	20.498.542.527.523.600	20	6	6	1	17	1
-0.193236369639635	4.836.442.353.576.420	4.921.978.569.030.760	1	0	0	0	1	1
0.2392844571731985	24.170.802.678.912.800	24.709.611.676.633.300	2	0	1	0	1	1

Hasil penggabungan kedua ekstraksi fitur tersebut memperoleh akurasi sebesar 0,846. Ekstraksi fitur penggabungan *Average base Word2vec* dan *Bag of Centroid base Word2vec* memiliki hasil yang tidak lebih baik dari *Average base Word2vec*. Hasil dari penggabungan kedua ekstraksi fitur dengan data ini memperoleh akurasi sebesar 0,853. Dari hasil tersebut maka dapat disimpulkan bahwa penggabungan ekstraksi fitur menggunakan *K-Means* mengurangi hasil akurasi, berbeda dengan penggabungan menggunakan Hirarki *Clustering* meningkatkan akurasi dari *Average base Word2vec*.

### 3.2 Pembahasan

Data yang digunakan pada penelitian ini merupakan data kolom komentar evaluasi dosen semester genap tahun ajaran 2017/2018 FMIPA ULM dengan jumlah data 9154 data. Untuk membuat model *Word2vec* digunakan data komentar dan juga data artikel Wikipedia bahasa Indonesia yang berjumlah 36000 artikel sebagai korpus. Setelah melewati proses *preprocessing* yang mana menghapus file redundansi, menghapus komentar yang tidak menunjukkan sentimen, menghapus karakter selain huruf, didapatkan 2823 data dengan 1242 data berlabel positif dan 1581 data berlabel negatif.

Pada model *Word2vec* setiap kata memiliki nilai *vector* yang merepresentasikan kata tersebut. Nilai *vector* tersebut kita gunakan untuk proses ekstraksi fitur. Pada penelitian ini kita gunakan *Average base Word2vec*, *Bag of Centroid base Word2vec* dan penggabungan keduanya untuk proses ekstraksi fitur yang mana menggunakan model *Word2vec* yang telah kita buat. Cara kerja ekstraksi fitur menggunakan *Average base Word2vec* pada dokumen yang kita klasifikasi terdiri dari kata-kata, Setiap kata memiliki nilai *vector* pada model *Word2vec*. Adapun jumlah fitur yang digunakan pada *Average base Word2vec* mengikuti jumlah fitur model *Word2vec* yang kita latih, yang mana pada penelitian ini digunakan model *Word2vec* dengan jumlah fitur 100 sehingga jumlah fitur pada ekstraksi fitur *Average base Word2vec* juga 100 fitur. Ekstraksi fitur menggunakan *Average base Word2vec* menghasilkan akurasi sebesar 0,848.

Cara kerja ekstraksi fitur menggunakan *Bag of Centroid base Word2vec*, pertama model *Word2vec* diklaster menggunakan algoritma *clustering* dengan metode ward, yang mana pada penelitian ini digunakan algoritma *K-Means* dan juga Hirarki *Clustering* dengan metode ward. Kemudian kata-kata pada model terkumpul berdasarkan klasternya. Setelah itu kata-kata pada setiap komentar dikelompokkan berdasarkan label *cluster*, kemudian hasil pengelompokan kata pada komentar menjadi fitur untuk dilakukan proses klasifikasi.

Dari hasil menggunakan *K-Means* dengan beberapa jumlah *cluster* atau *centroid* tidak berpengaruh banyak dalam hasil kinerja klasifikasi, hal ini kemungkinan karena pada algoritma *K-Means* dalam penentuan nilai K awal acak sehingga nilai *cluster* pada setiap prosesnya berbeda-beda. Nilai *cluster* pada proses ini sangat berpengaruh besar dalam menentukan akurasi klasifikasi. Dalam *Bag of Centroid base Word2vec* menggunakan algoritma *K-Means* hasil akurasi terbaik pada saat jumlah fitur 70 dengan akurasi 0,804.

Kemudian untuk ekstraksi fitur *Bag of Centroid base Word2vec* menggunakan algoritma Hirarki *Clustering* juga digunakan jumlah fitur yang bervariasi. Ketika jumlah fitur meningkat maka hasil akurasi juga meningkat. Adapun akurasi tertinggi pada saat jumlah fitur berjumlah 75 dengan hasil akurasi sebesar 0,816. Dengan hasil ini maka algoritma Hirarki *Clustering* memiliki performa yang lebih baik dari pada algoritma *K-Means*. Algoritma Hirarki *Clustering* dalam hal ini lebih baik dalam mengelompokkan kata-kata karena algoritma ini menyusun kata dengan *vector* yang berdekatan kemudian memotong dalam beberapa bagian sesuai jumlah *centroid* yang kita tentukan. Berbeda dengan algoritma *K-Means* yang mana pada algoritma tersebut menentukan nilai *K* awal dengan acak sehingga *cluster* pun terkumpul menjadi kurang terstruktur.

Kemudian untuk ekstraksi fitur penggabungan *Average base Word2vec* dan *Bag of Centroid base Word2vec*, yang mana 100 fitur *Average base Word2vec* digabung dengan fitur terbaik *Bag of Centroid base Word2vec* dalam hal ini menggunakan algoritma *Clustering K-Means* dengan jumlah fitur 70 memperoleh akurasi 0,846. Dalam hal ini hasil penggabungan ekstraksi fitur tidak lebih baik dari ekstraksi fitur *Average base Word2vec*. Karena hasil ekstraksi fitur *Bag of Centroid base Word2vec* menggunakan *K-Means* menghasilkan fitur yang kurang baik.

Ekstraksi fitur *Bag of Centroid base Word2vec* menggunakan algoritma Hirarki *Clustering* diperoleh kinerja tertinggi pada saat jumlah fitur 75 dengan akurasi 0,816 maka dilakukan penggabungan fitur dengan *Average base Word2vec* yang mana hasil penggabungan ini menghasilkan akurasi sebesar 0,863. Dari hasil penggabungan ini memperoleh akurasi yang lebih baik dibandingkan dengan *Average base Word2vec* karena fitur yang dihasilkan algoritma Hirarki *Clustering* bagus dan juga memberikan informasi terhadap fitur yang lebih lengkap.

Dari hasil penelitian ini didapatkan hasil terbaik ekstraksi fitur menggunakan penggabungan ekstraksi fitur *Average base Word2vec* dengan ekstraksi fitur *Bag of Centroid base Word2vec* menggunakan algoritma Hirarki *Clustering*.

#### 4 SIMPULAN

Kesimpulan yang dapat diambil dari hasil penelitian yang telah dilakukan bahwa perbandingan ekstraksi fitur *Average base Word2vec* dan *Bag of Centroid base Word2vec*, pada pengujian menghasilkan *Average base Word2vec* memiliki kinerja yang lebih baik. Penerapan *Average base Word2vec* pada data komentar evaluasi dosen memiliki akurasi 0,848, Sedangkan *Bag of Centroid base Word2vec* menggunakan data komentar evaluasi dosen menghasilkan kinerja terbaik dengan akurasi 0,816. Dan penggabungan data ekstraksi fitur *Average base Word2vec* dan *Bag of Centroid base Word2vec* menggunakan data komentar evaluasi dosen menghasilkan akurasi 0,853. Untuk penelitian lebih lanjut dapat menggunakan *Average base Wordvec* dengan jumlah fitur yang bervariasi, dapat menggunakan ekstraksi fitur menggunakan *Word2vec* yang mana menggabung beberapa kata menjadi satu makna, dengan contoh kata “tidak” dan kata “baik” menjadi “tidak baik” dalam satu makna atau dapat gunakan *Stopword Removal* agar kata-kata lebih jelas.

**DAFTAR PUSTAKA**

- [1] Fauzi, M. Ali, 2019. ***Word2vec model for Sentiment Analysis of product reviews in Indonesian language***, Brawijaya University. Vol. 9, No. 1. page. 525-530.
- [2] Rossiello, Gaetano dkk, 2017. ***Centroid-based Text Summarization through Compositionality of Word Embeddings***, University of Bari. Proceedings of the MultiLing 2017 Workshop on Summarization and Summary Evaluation Across Source Types and Genres, pages 12-21