

# Feature Selection Menggunakan Chi-Square Dan N-Gram Dengan Algoritma Naive Bayes Classifier Untuk Analisis Sentimen Review Produk Elektronik

<sup>1</sup>Deni Irvantoro (1410651199) <sup>2</sup>Ilham Saifudin S.Si., M.Si. <sup>3</sup>Reni Umilasari S.Si., M.Si.

Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember  
Email : [den.irtanto99@gmail.com](mailto:den.irtanto99@gmail.com)

## ABSTRAK

Analisis sentimen perlu diterapkan untuk menentukan review yang bernilai positif atau negatif. Algoritma klasifikasi seperti Naive Bayes Classifier (NBC) telah banyak di terapkan. Di samping memiliki performanya yang baik di berbagai domain, NBC memiliki sensitifitas terhadap pemilihan fitur. Chi-square merupakan salah satu metode statistik untuk pemilihan fitur dengan uji independensi dan estimasi yang bertujuan untuk mengetahui ketergantungan suatu class pada suatu fitur. Sedangkan N-Gram merupakan pemotongan kata berdasarkan nilai n. Dengan penggabungan chi-square dan model N-Gram, didapatkan hasil klasifikasi dengan akurasi, presisi, dan recall tertinggi pada uni-gram berada pada uji penggunaan 75% fitur dengan akurasi 89,00%, presisi 89,00%, dan recall 100%. Sedangkan hasil klasifikasi pada bi-gram dan uni+bi-gram cenderung memiliki hasil tertinggi yang sama dengan akurasi 89%, presisi 89%, dan recall 100% pada uji penggunaan 25% fitur. Dalam penelitian ini masing-masing teknik n-gram dengan penerapan feature selection chi-square memiliki ambang batas penggunaan fitur yang berbeda. Secara umum chi-square tidak mempengaruhi urutan hasil dari klasifikasi, tapi mampu memberikan improvisasi hasil akurasi, presisi, dan recall pada komparasi ketiga model n-gram.

**Kata Kunci :** Analisis Sentimen, Feature Selection, Chi-Square, N-Gram, Algoritma Naive Bayes Classifier

## PENDAHULUAN

Banyak situs *e-commerce* dan *market place* saat ini menyediakan penjualannya di internet. *Amazon.com, Inc.* merupakan perusahaan yang bergerak dalam bidang perdagangan berbasis daring yang berpusat di *Seattle*, Amerika Serikat. Produk elektronik merupakan salah satu produk yang memiliki banyak *review*. *Review* ini menimbulkan munculnya berbagai sentimen tentang kepuasan konsumen terhadap barang yang dibeli. Sentimen konsumen ini sangat penting dalam suatu pengambilan keputusan bagi konsumen lain untuk mengetahui kualitas suatu produk yang akan dibeli. Di sinilah peran analisis sentimen diterapkan untuk menentukan ulasan yang bernilai positif atau negatif.

Algoritma klasifikasi dengan pendekatan *supervised machine learning* telah banyak diterapkan. *Naive Bayes Classifier* (NBC) merupakan salah satu teknik pembelajaran untuk *text classification* dan memiliki performa baik di berbagai domain (Qiang Ye, Ziqiong Zhang, Rob Law, 2009). Tapi, di samping dari performanya yang baik *Naive Bayes Classifier* sangat sensitif terhadap pemilihan fitur. (J. Chen, et al, 2009). Selain itu semakin meningkatnya volume teks juga akan menimbulkan banyaknya atribut yang dipakai dalam

klasifikasi dan hal ini akan mengurangi kinerja algoritma klasifikasi (J. Ipmawati, Kusriani, Emha T. L., 2017). Maka dari itu teknik *preprocessing* data penting dalam klasifikasi sentimen (I. Hemalatha, G. P Saradhi, A. Gofardhan, 2012).

*Feature selection* merupakan teknik pemilihan fitur, menurut (Vinita Chandani, Romi Satria W, Purwanto, 2015) *feature selection* dapat digunakan untuk menyeleksi banyaknya fitur yang tidak relevan pada dataset. Beberapa algoritma *feature selection* telah diterapkan seperti *Mutual Information* (IG), *Term Frequency* (TF), *Information Gain* (IG), *Chi-Square*, *Expected Cross Entropy*, *Odds Ratio*, the *Weight of Evidence*, dan *Gini Index* banyak digunakan dan memiliki performa yang baik (Vandana Korde, C. Namrata M, 2012). Berdasarkan penelitian (J. Ling, N. E. I.Kencana, T. B. Oka, 2013) penggunaan Chi-Square disarankan dengan menggabungkan algoritma klasifikasi Naive Bayes Classifier dan berdasarkan (Novan D. P., Yuita A. S., Putra P. A., 2018) penggunaan model N-Gram disarankan untuk membantu dalam penyeleksian fitur dan juga untuk memeriksa keakuratan hasil klasifikasi yang diperoleh. Untuk itu penelitian ini mengajukan pendekatan dengan menggunakan metode *Naive Bayes Classifier* serta penggunaan *Chi-Square*

sebagai *feature selection* dengan mengkomparasikan *N-Gram* (*uni-gram*, *bi-gram*, dan gabungan *uni* dan *bi-gram*).

Penelitian serupa pernah dilakukan oleh beberapa peneliti. Dari beberapa penelitian yang telah dilakukan menghasilkan akurasi yang berbeda-beda bergantung pada metode *feature selection* yang dipakai. Seperti penelitian yang dilakukan oleh (Dinda A. M, 2014), hasil menunjukkan dari 78,50% menjadi 83%. Hal ini terjadi karena pada uji pertama tidak dilakukan pemilihan fitur pada algoritma *Naïve Bayes Classifier*. Setelah kombinasi pemilihan fitur dengan kombinasi *Information Gain* dan *Genetic Algorithm* diterapkan akurasi meningkat 4% menjadi 83%. Begitu juga penelitian yang dilakukan oleh (Siti Ernawati, 2016), hasil menunjukkan dari 79,50% menjadi 86,88%. Improvisasi akurasi pada algoritma *Naïve Bayes Classifier* di dapatkan karena penerapan pemilihan fitur menggunakan *Particle Swarm Optimization*. Penerapan pemilihan fitur ini lebih efektif dibandingkan dengan kombinasi *Information Gain* dan *Genetic Algorithm* terbukti pada penelitian Siti Ernawati improvisasi akurasi lebih tinggi.

Pada penelitian Sabna Gupta dan Sabnam Parveen hasil didapatkan 78% menjadi 94% dan penelitian Linda Widiastuti didapatkan hasil 8,50 menjadi 93,50%. Dari kedua penelitian tersebut penelitian yang dilakukan oleh Sabna Gupta dan Sabnam Parveen memiliki akurasi yang tinggi dengan improvisasi 16% terhadap algoritma *Naïve Bayes Classifier*. Hal ini dikarenakan adanya optimalisasi pemilihan fitur dengan *TF-IDF* dan *Gain Ratio*. Sedangkan pada penelitian Linda Widiastuti terjadi improvisasi akurasi 12% terhadap algoritma *Naïve Bayes Classifier* karena adanya penerapan pemilihan fitur *Mutual Information* dan *Particle Swarm Optimization*.

### Analisis Sentimen

Analisis sentimen adalah proses studi komputasi pendapat, perilaku, dan emosi yang bersifat positif, negatif atau netral. Menurut (Ranjan S., Erik C., Amir H, 2017) ada dua tugas utama dalam analisis sentimen. Pertama adalah *sentiment detection* yang mengklasifikasikan teks menjadi kalimat subjektif atau objektif. Kedua adalah *sentiment classification* yang mengklasifikasikan sentimen dari potongan teks yang telah terdapat opini positif atau negatif.

### Text Mining

Menurut (Susan Koshy dan R. Padmajavalli, 2015) *text mining* diartikan sebagai penemuan informasi baru yang sebelumnya tidak dikenal dengan secara otomatis mengekstrak informasi dari sumber tertulis yang berbeda. Pengolahan *text mining* secara *high quality* biasanya mengacu pada beberapa kombinasi relevansi dan *novelty* (Jiy George, Sandya N., dan Suja George, 2014).

### Porter Stemmer

Secara khusus, *porter stemmer* memiliki lima langkah, setiap langkah mendefinisikan aturan (Wahiba Ben A. K, 2013).

1. Menangani bentuk jamak, *participles*, *present participles*, dan mengubah “y” menjadi “i”. Contoh: kata “Happy” menjadi “happi”.
2. Mengubah contoh kata “*generalization*” menjadi “*generalize*”
3. Menghapus *suffixes* ganda yang tidak ditangani pada langkah kedua. Contoh: kata “*generalize*” menjadi “*general*”.
4. Menghilangkan sisa *suffixes* seperti “*general*” yang berubah menjadi “*gener*”, dan “*oscillate*” menjadi “*oscill*”.
5. Menjadikan kata yang diakhiri dengan -e di ubah menjadi konsonan ganda. Contoh: “*attribute*” menjadi “*attribut*” dan “*oscill*” menjadi “*oscil*”.

### N-Gram Tokenizer

Dalam linguistik komputasional *N-gram* sering digunakan dalam memprediksi kata atau memprediksi karakter. *N-gram* memiliki kelebihan dalam efisiensi dan efektifitas pengkategorian teks (Kennedy Odhiambo O, 2016) dan dapat mengklasifikasikan dengan akurasi yang tinggi (Yadav & Parne, 2015). Seperti contoh yang terdapat pada kalimat “klasifikasi teks pengaduan pada sambat *online* menggunakan metode *N-Gram* dan NW-KNN” (Annisya A. P, M. Ali Fauzi, M. Tanzil. F, 2018).

**Uni-gram** : Klasifikasi, teks, pengaduan, pada, sambat, *online*, menggunakan, metode, *N-gram*, dan, NW-KNN.

**Bi-gram** : Klasifikasi teks, teks pengaduan, pengaduan pada, pada sambat, sambat *online*, *online* menggunakan, menggunakan metode, metode *N-Gram*, *N-Gram* dan, dan NW-KNN.

**Uni+bi-gram** : Klasifikasi, klasifikasi teks, teks, teks pengaduan, pengaduan, pengaduan sambat, sambat, sambat online, online, online menggunakan, menggunakan, menggunakan metode, metode, metode *N-Gram*, *N-Gram*, *N-Gram* dan, dan, dan NW-KNN

### Chi-Square

Dalam *feature selection*, *chi-square* digunakan untuk uji independensi dan estimasi yang bertujuan untuk mengetahui ketergantungan suatu *class* pada suatu fitur (Nachirat R, Wattana P, 2015). *chi-square* memiliki 3 uji, yaitu *observe frequency*, *expected frequency*, dan *test statistic*.

$$e_{ij} = \frac{o_i \cdot o_j}{N} \quad (1)$$

$e_{ij}$  = *expected frequency*.

$O_i$  = frekuensi kolom marginal.

$O_j$  = frekuensi baris marginal.

$N$  = Jumlah sample.

$$\chi^2 = \sum_{i=1}^n \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \quad (2)$$

$\chi^2$  = *Chi-Square*.

$o_{ij}$  = *Observed Frequency*.

$e_{ij}$  = *Expected Frequency*.

### Algoritma Naive Bayes

Asumsi teorema *Naive Bayes* bahwa nilai atribut pada suatu kelas tertentu memiliki efek *independent* (Maresha Caroline Wijanto, 2015). Bentuk klasifikasi pada *Naive Bayes* yaitu dengan menghitung derajat kecocokan. Secara umum teorema *Bayes* dapat dinotasikan sebagai berikut:

$$V_{MAP} = \arg \max_{V_c \in V} P(v_c) \prod_{i=1} P(w_i | v_c) \quad (3)$$

$P(v_c)$  digunakan untuk pelatihan data, sedangkan untuk  $P(w_i | v_c)$  digunakan sebagai uji data. Sehingga notasi dari  $P(v_c)$  dan  $P(w_i | v_c)$  adalah:

$$P(v_c) = \frac{|dt_c|}{P(n_{dt})} \quad (4)$$

$|dt_c|$  merupakan dokumen *training* yang memiliki kategori  $c$ , sedangkan untuk  $(n_{dt})$  merupakan asumsi dari jumlah keseluruhan dokumen *data training*.

Untuk nilai dari  $P(w_i | v_c)$  yang merupakan probabilitas kata ( $w_i$ ) dalam kategori  $v_c$  dapat dinotasikan sebagai berikut:

$$P(w_i | v_c) = \frac{|n_i + 1|}{|N_c + sw|} \quad (5)$$

Dari rumus di atas,  $n_i$  disumsikan sebagai frekuensi kemunculan *term* dalam kategori  $v_c$ ,  $N_c$  adalah jumlah kemunculan seluruh *term* yang berkategori  $v_c$ , sedangkan  $sw$  merupakan banyaknya seluruh *term*.

### Confussion Matrix

Pada pengukuran kinerja menggunakan *confussion matrix* terdapat 4 (empat) representasi hasil yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) yang menghasilkan nilai persamaan akurasi, presisi, dan *recall*.

1. Akurasi mempresentasikan data yang diklasifikasikan dengan benar oleh metode klasifikasi dengan semua jumlah data. Akurasi dapat dijabarkan dengan persamaan:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (6)$$

2. Presisi mempresentasikan sebagai ukuran dari ketepatan suatu klasifikasi (persentase dari data positif yang diklasifikasikan dengan benar). Presisi dapat dijabarkan dengan persamaan:

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\% \quad (7)$$

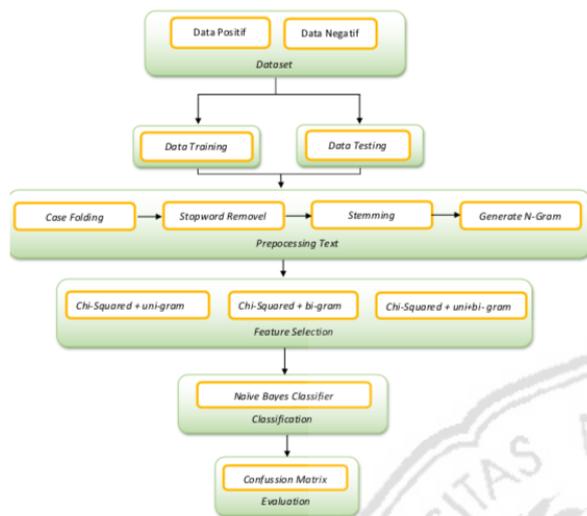
3. *Recall* mempresentasikan sebagai tolak ukur dari keberhasilan klasifikasi yang diidentifikasi dengan benar. *Recall* dapat dijabarkan dengan persamaan:

$$\text{Recall} = \frac{TP}{FN+TP} \times 100\% \quad (8)$$

### METODE PENELITIAN

Penelitian ini mengusulkan penggunaan *chi-square* dan *n-gram* sebagai *feature selection* untuk mengetahui tingkat

akurasi, presisi, dan *recall* terhadap algoritma *Naive Bayes Classifier*. Uji yang dilakukan yaitu dengan mengkomparasikan *n-gram* (*uni-gram*, *bi-gram*, dan *uni+bi-gram*) yang telah melalui proses pemilihan fitur menggunakan *chi-square*. **Gambar 1**. Menunjukkan rancangan pemikiran penelitian.



**Gambar 1** Rancangan Penelitian

Data yang digunakan dalam penelitian ini adalah *dataset review product electronic amazon.com*, yang di unduh dari <https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products>. Data yang digunakan sebanyak 600 data dengan atribut *reviews.rating* dan *reviews.text*, 500 data digunakan sebagai data latih dan 100 data digunakan sebagai data uji. *Rating* yang digunakan meliputi *rating* 1, 2, 4, dan 5 yang akan dikonversi seperti **Tabel 1**.

**Table 1.** Konversi Kategori Sentimen

Rating	Sentiment Class
★	Negative
★★	Negative
★★★★★	Positive
★★★★★	Positive

Selengkapnya dataset dapat di akses di <https://drive.google.com/open?id=1q2F7zf4McQgWQ3JiGeSA5sRFJJm9Z9tj>

Sebelum dilakukan *generate n-gram* dan pemilihan fitur menggunakan *chi-square*, data akan melalui *preprocessing*. *Preprocessing* bertujuan untuk menjadikan dokumen teks

yang tidak terstruktur menjadi terstruktur. *Preprocessing* yang dilakukan meliputi:

1. *Case folding* adalah menyamaratakan *string* menjadi *lowercase*.
2. *Remove number* adalah menghapus seluruh angka yang ada pada dokumen.
3. *Stopword removal* adalah proses menghapus kata yang sering muncul atau tidak terlalu berguna. Dalam penelitian ini teknik *stoping* menggunakan kamus *stoping* bahasa Inggris.
4. *Generate n-gram* adalah memotong kata kata berdasarkan nilai *n*. *Generate n-gram* yang dilakukan dalam penelitian ini adalah *uni-gram*, *bi-gram*, dan *uni+bi-gram*.

### Implementasi *Chi-Square*

Setelah data dilakukan *preprocessing* dan mendapatkan term *n-gram* langkah selanjutnya yaitu menghitung bobot term menggunakan *chi-square* untuk mendapatkan nilai statistik term. Diberikan contoh kemunculan kata pada **Tabel 2**.

**Table 2.** Contoh pemilihan fitur term *uni+bi-gram*

Doc	Sentimen	great	great_love	like	worst	worst_slow	Total
D1	Positif	4	7	5	0	0	16
D2	Negatif	0	0	0	5	6	11
Total		4	7	5	5	6	27

**Tabel 2.** Merupakan contoh kemunculan *term* pada dokumen atau *observed frequency*. Setelah di dapatkan nilai *observed frequency*, Langkah selanjutnya yaitu dengan menghitung nilai *expected frequency* berdasarkan bobot *term* pada **Tabel 2** dengan persamaan (1). Sebagai contoh dilakukan penghitungan *term* “*great*” dan “*great\_love*”.

$$great = (4 \cdot 16) / 27 = 2,37$$

$$great\_love = (7 \cdot 16) / 27 = 4,14$$

Setelah nilai dari *expected frequency* di dapatkan, langkah selanjutnya yaitu dengan menghitung *test statistic* pada setiap *term* berdasarkan persamaan (2). Berdasarkan nilai *expected frequency* pada *term great* dan *great\_tablet*, hasil *test statistic* yang di dapatkan adalah

$$great = (4 - 2,37)^2 / 2,37 = 1,37$$

$$great\_love = (7 - 4,14)^2 / 4,14 = 1,38$$

berdasarkan penghitungan persamaan (1) dan (2) nilai statistik  $term\ great = 1,37$  dan  $great\_love = 1,38$

### Implementasi Naive Bayes Classifier

Pada tahapan ini data yang digunakan telah melalui proses *generate n-gram* dan seleksi fitur menggunakan *chi-square*. Diberikan contoh data *training* pada **Tabel 3**.

**Table 3. Contoh Data Training**

Doc	Sentimen	great	great_love	like	worst	worst_slow	unreal
D1	Positif	2	3	2	0	0	0
D2	Negatif	0	0	0	3	4	1

Langkah pertama yaitu menghitung *prior probability* dari data latih berdasarkan persamaan (4). Maka di dapatkan nilai

$$P(v_{positive}) = \frac{1}{2} = 0.5$$

$$P(v_{negative}) = \frac{1}{2} = 0.5$$

Berdasarkan data *training* pada **Tabel. 3** maka dihasilkan nilai *prior probability* kategori *positive* = 0,5 dan kategori *negative* = 0,5. Langkah selanjutnya yaitu menentukan probabilitas bersyarat dari setiap *term*. Sebagai contoh *term* “*easy\_use*” pada **Tabel 3**. Berdasarkan persamaan (5). Maka nilai probabilitas bersyarat *term* “*easy\_use*” adalah

$$P(w_{great}|v_{positif}) = \frac{|2 + 1|}{|7 + 6|} = \frac{3}{13}$$

$$P(w_{great}|v_{negatif}) = \frac{|3 + 1|}{|8 + 6|} = \frac{4}{14}$$

Sebagai implementasi uji data, diberikan contoh dokumen dengan kemunculan term sebagai berikut:

*great* : 0

*great\_love* : 1

*like* : 1

*worst* : 2

*worst\_slow* : 3

*unreal* : 2

Pada tahapan uji data penghitungan dilakukan berdasarkan persamaan (3) untuk mengetahui kemungkinan data berkategori *positive* atau *negative*. Untuk kemungkinan data berkategori *positive* dapat dilihat pada penghitungan dibawah ini.

$$V_{MAP} = 1/2 * 1/13 * 2/13 * 2/13 * 3/13 * 4/14 * 3/13$$

$$V_{MAP} = 0,5 * 0,076 * 0,153 * 0,153 * 0,230 * 0,307 * 0,230$$

$$V_{MAP} = 1,44$$

Untuk kemungkinan data berkategori *negative* dapat dilihat pada penghitungan dibawah ini.

$$V_{MAP} = 1/2 * 1/14 * 2/14 * 2/14 * 3/14 * 4/14 * 3/14$$

$$V_{MAP} = 0,5 * 0,071 * 0,142 * 0,142 * 0,214 * 0,285 * 0,214$$

$$V_{MAP} = 9,34$$

Berdasarkan penghitungan di atas untuk kemungkinan data bernilai *positive* adalah 1,44 dan untuk kemungkinan bersentimen *negative* adalah 9,34. Berdasarkan hasil propabilitas, dapat disimpulkan dokumen uji adalah berkategori *negative*.

### HASIL DAN PEMBAHASAN

Proses pengujian dalam penelitian ini menggunakan *library* klasifikasi yang disediakan oleh *RStudio 3.3.3*. Data yang digunakan sebanyak 600 data *reviews product electronic* yang memiliki *rating* 1 (satu), dua (dua), 4 (empat), dan 5 (lima). Data yang digunakan sebagai data latih sebanyak 500 data, dan 100 data digunakan sebagai data uji. Data latih meliputi 250 data *rating* 1 (satu) dan 250 data *rating* 5 (lima). Sedangkan pengambilan 100 data uji dilakukan secara acak berdasarkan *rating* tersebut. Pengujian dalam penelitian ini dilakukan sebanyak 4 kali uji dengan porsi 25% fitur, 50% fitur, 75% fitur, dan 100% fitur pada setiap model *n-gram* (*uni-gram*, *bi-gram*, dan *uni+bi-gram*). Hasil pengujian dapat dilihat pada **Table 4.**, **Table 5.**, **Table 6.**, dan **Table 7.**

**Table 4.** Hasil klasifikasi menggunakan 25% fitur

N-Gram	Fitur	Hasil		
		Akurasi	Presisi	Recall
<i>uni-gram</i>	25%	87,60%	88,84%	98,42%
<i>bi-gram</i>		89,00%	89,00%	100%
<i>uni+bi-gram</i>		89,00%	89,00%	100%

Hasil klasifikasi menggunakan 25% fitur hasil tertinggi terdapat pada proses model *bi-gram* dan *uni+bi-gram* dengan nilai akurasi 89%, presisi 89%, dan *recall* 100%.

Sedangkan untuk hasil *uni-gram* mendapatkan nilai akurasi 87,60%, presisi 88,84%, dan *recall* 98,42%. Dalam pengujian ini *uni-gram* memiliki performa lebih rendah daripada *bi-gram* dan *uni+bi-gram*, hal ini dikarenakan penggunaan fitur *bi-gram* dan *uni+bigram* lebih representatif kehadirannya dalam pengujian klasifikasi terhadap algoritma *Naive Bayes Classifier*.

**Table 5.** Hasil klasifikasi menggunakan 50% fitur.

N-Gram	Fitur	Hasil		
		Akurasi	Presisi	Recall
<i>uni-gram</i>	50%	88,80%	88,97%	99,97%
<i>bi-gram</i>		88,80%	88,97%	99,97%
<i>uni+bi-gram</i>		88,80%	88,97%	99,97%

Dari ke-tiga model *N-Gram* dengan pengujian 50% fitur *uni-gram*, *bi-gram*, dan *uni+bi-gram* memiliki nilai akurasi 88,80%, Presisi 88,97%, dan *Recall* 99,97%. Kesamaan hasil tersebut merupakan akibat dari frekuensi kemunculan fitur yang sama antara *uni-gram* dan *bi-gram* sehingga perhitungan probabilitas pada ke-tiga fitur *N-Gram* memiliki hasil yang sama.

**Table 6.** Hasil klasifikasi menggunakan 75% fitur.

N-Gram	Fitur	Hasil		
		Akurasi	Presisi	Recall
<i>uni-gram</i>	75%	89,00%	89,00%	100%
<i>bi-gram</i>		88,80%	88,97%	99,97%
<i>uni+bi-gram</i>		88,80%	88,97%	99,97%

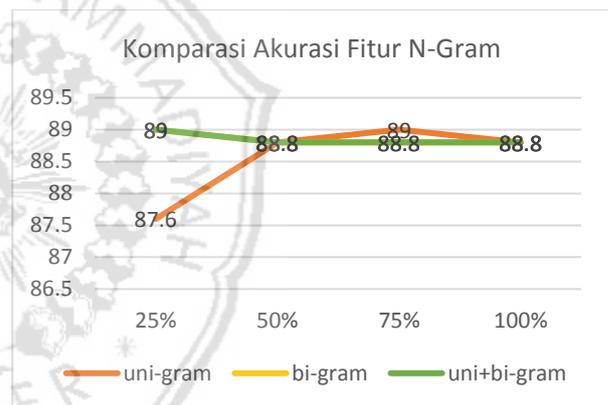
Dengan menggunakan 75% fitur klasifikasi, *uni-gram* memiliki hasil tertinggi dibandingkan dengan *bi-gram* dan *uni+bi-gram* dengan akurasi 89%, presisi 89%, dan 100% *recall*. Hasil tersebut menunjukkan performa *uni-gram* lebih

baik daripada performa *bi-gram* dan *uni+bi-gram* pada penggunaan 75% fitur.

**Table 7** Hasil klasifikasi menggunakan 100% fitur.

N-Gram	Fitur	Hasil		
		Akurasi	Presisi	Recall
<i>uni-gram</i>	100%	88,80%	88,97%	99,77%
<i>bi-gram</i>		88,80%	88,97%	99,77%
<i>uni+bi-gram</i>		88,80%	88,97%	99,77%

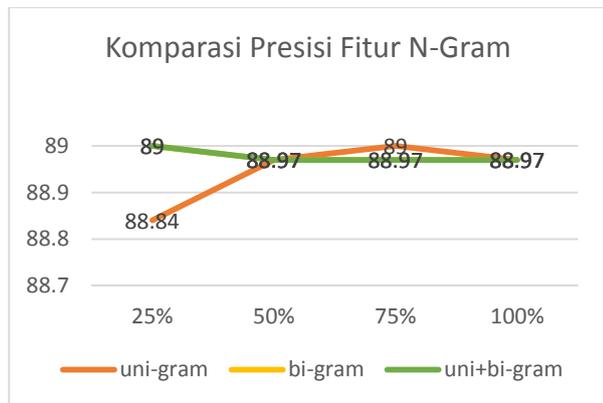
Hasil ini memiliki kesamaan pada hasil pengujian menggunakan 50% fitur. Hal tersebut disebabkan pada kasus frekuensi kemunculan fitur *uni-gram* dan *bi-gram* yang sama pada klasifikasi sehingga hasil pada *uni+bi-gram* juga memiliki hasil yang sama.



**Gambar 2.** Komparasi Akurasi Fitur *N-Gram*

Hasil komparasi fitur *bi-gram* dan *uni+bigram* lebih tinggi daripada fitur *uni-gram* pada uji 25%. Pada uji 50% dan 100% hasil akurasi menunjukkan hasil yang sama yaitu dengan akurasi 88,8% untuk ke-tiga model *N-Gram*. Hasil tersebut disebabkan prediksi pada fitur ketiga model *N-Gram* memiliki fitur negatif dan positif yang terdeteksi sama. Tapi perbedaan hasil terjadi kembali pada uji 75% fitur dengan hasil akurasi *uni-gram* lebih tinggi daripada fitur *bi-gram* dan *uni+bigram*. Untuk hasil performa setiap model *N-Gram* menunjukkan peningkatan pada uji 25% ke 75% fitur *uni-gram* dari 87,6% menjadi 89%, tapi mengalami penurunan pada uji 100% fitur menjadi 88,8%. Sedangkan untuk *bi-gram* dan *uni+bi-gram* memiliki akurasi tertinggi pada uji 25%

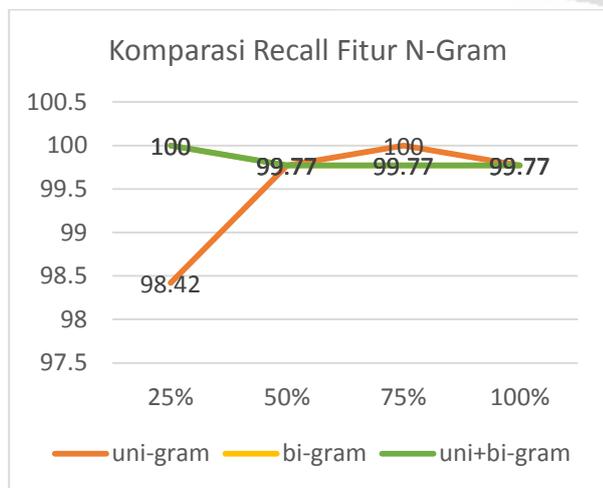
dengan akurasi 89%, tapi mengalami penurunan pada uji 50% sampai 100% dengan akurasi 88,8%.



**Gambar 3.** Komparasi Presisi Fitur *N-Gram*

Pada **Gambar 4.2**, menunjukkan tingkat presisi fitur *uni-gram* pada uji 25% lebih rendah dari pada fitur *bi-gram* dan *uni+bi-gram*. Sedangkan pada uji 50% dan 100% fitur menunjukkan hasil yang sama yaitu dengan presisi 88,97%, hal ini disebabkan fitur negatif dan positif dari ketiga model *N-Gram* memiliki hasil prediksi yang sama. Namun perbedaan hasil kembali terjadi pada uji 75% fitur dengan presisi fitur *uni-gram* lebih tinggi daripada fitur *bi-gram* dan *uni+bi-gram* dengan nilai presisi *uni-gram* 89% dan presisi *bi-gram* dan *uni+bi-gram* 88,97%.

Untuk hasil performa setiap model *N-Gram*, *uni-gram* memiliki hasil presisi tertinggi pada uji 75% fitur sedangkan untuk uji fitur pada *bi-gram* dan *uni+bi-gram* hasil presisi tertinggi terdapat pada uji 25% fitur dengan presisi 89%



**Gambar 4.** Komparasi Recall Fitur *N-Gram*.

Pada komparasi *recall* dari fitur *uni-gram* cenderung memiliki peningkatan yang sama dengan akurasi dan presisi pada *uni-gram*, *bi-gram*, *uni+bi-gram* yaitu pada uji 25%. Pada uji fitur *uni-gram* 25%, *recall* mengalami peningkatan dari 98,42% menjadi 100% dan mengalami penurunan pada uji fitur 100% dengan *recall* 99,77%. Sedangkan pada fitur gabungan *uni+bi-gram*, hasil uji tertinggi pada uji 25% fitur dengan *recall* 100% dan mengalami penurunan pada uji 50% sampai 100% fitur dengan *recall* 99,77%.

## KESIMPULAN DAN SARAN

### Kesimpulan

Penelitian ini telah melalui beberapa pengujian yang dimaksudkan untuk mengetahui pengaruh penggunaan fitur pada setiap model *N-Gram* terhadap algoritma *Naive Bayes Classifier*. Pengujian yang telah dilakukan mendapatkan hasil klasifikasi dengan akurasi, presisi, dan *recall* tertinggi pada *uni-gram* berada pada uji penggunaan 75% fitur dengan akurasi 89,00%, presisi 89,00%, dan *recall* 100%. Sedangkan hasil klasifikasi pada *bi-gram* dan *uni+bi-gram* cenderung memiliki hasil tertinggi yang sama dengan akurasi 89%, presisi 89%, dan *recall* 100% pada uji penggunaan 25% fitur.

Dalam penelitian ini pengujian klasifikasi pada setiap model *N-Gram* dengan penerapan *feature selection chi-square* memiliki ambang batas penggunaan fitur yang berbeda. Secara umum *chi-square* tidak mempengaruhi urutan hasil dari klasifikasi, tapi mampu memberikan improvisasi hasil akurasi, presisi, dan *recall* pada komparasi ketiga model *N-Gram*. Hasil tersebut dibuktikan pada uji 25% fitur bahwa model *bi-gram* dan *uni+bi-gram* lebih baik performanya daripada *uni-gram*, sedangkan pada uji 75% fitur *uni-gram* lebih baik performanya daripada *bi-gram* dan *uni+bi-gram* dan sedangkan pada pengujian 50% dan 100% fitur ketiga model *N-Gram* memiliki hasil performa yang sama.

### Saran

1. Perlu percobaan variasi pada pada teknik *n-gram*, terutama pada *bi-gram* dan gabungan *uni+bi-gram* dengan penggunaan data uji yang lebih banyak.
2. Perlu dilakukan komparasi *chi-square* dengan algoritma *feature selection* lain untuk memberikan hasil *feature selection* yang lebih baik.

## DAFTAR PUSTAKA

- Ali, M. F., D. C. Utomo, B. D. Setiawan. *Automatic Essay Scoring System Using N-Gram and Cosine Similarity for Ga mification Based E-Learning*. ICAIP 2017, Agustus 25-27, 2017, Bangkok, Thailand.
- Caroline, M. W., 2015. *Sistem Pendeteksi Pengirim Tweet dengan Metode Klasifikasi Naïve Bayes*. Jurnal Teknik Informatika dan Sistem Informasi. Vol.1, No.2, Agustus 2015.
- Chandani, Vinita., Romi Satria W., Purwanto., 2015. *Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada analisis Sentiment Review Film*. *Journal of Intelligent Systems*, Vol. 1, No.1, February 2015.
- Chen, Jingnian., et al., 2009. *Feature Selection for Text Classification with Naive Bayes*. *Expert Sistem with Applications*, 36 (2009), 5432-5435.
- Hall, Mark., 2018. *Amazon.com American Company*. *Encyclopedia Britannica*. <https://www.britannica.com/topic/Amazoncom>. Di akses pada 10 Oktober 2018.
- Hemalata, I., G. P Saradhi Varma., A. Govardhan., 2012. *Preprocessing the Informal Text For Efficient Sentiment Analysis*. *Internasional Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, Vol. 1, Issues 2, July-August 2012.
- Ipmawati, J., Kusriani., Emha T. L., 2017. *Komparasi Teknik Klasiikasi Teks Mining Pada Analisis Sentimen*. *Indonesian Journal on Networking and Security*, Vol. 6, No. 1, Thn. 2017.
- Karaa, Wahiba. B. A., 2013. *A New Stemmer to Improve Information Retrieval*. *International Journal of Network Security and Its Applications (IJNSA)*, Vol.5, No.4, July 2013.
- Korde, Vandana., C. Namrata Mahender. *Text Classification and Classifiers: Survey*. *International Journal of Artificial Intelligence and Application (IJAIA)*, Vol.3, No.2, March 2012.
- Koshy, Susan., R. Padmavalli., 2015. *Text Categorization of Multi-Label Document Text Mining*. *International Journal of Data Mining Techniques and Applications*, Vol. 4, Issue: 2 Desember 2015 Page No.52-58. ISSN 2278-2419.
- Ling, J., I. P. E. N. Kencana., 2014. T. Bagus Oka. *Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi-Square*. *E-Jurnal Matematika*. Vol.3, No.3, Agustus 2014, pp. 92-99.
- Ogadha, K. O., 2016. *N-Gram for Text Classification Using Supervised Machine Learning Algorithms*. *A thesis submitted in fulfillment for the degree of Doctor of Philosophy in Information Technology*. *University of Agriculture and Technology*.
- Prasanti, A. A., M. A. Fauzi., M. Tanzil F. *Klasifikasi teks Pengaduan Pada Sambat Online Menggunakan metode N-Gram dan Neighbor Weighted K-Nearest Neighbor (NW-KNN)*. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. Vol.2, No.2, Februari 2018, hlm. 594-601.
- Pratama, N. D., Yuita A. S., Putra P. A., 2018. *Analisis Sentimen Pada Review Konsumen Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Chi-Square Untuk Rekomendasi Lokasi Makanan Tradisional*. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol.2, No.9, September 2018, hlm. 2982-2988.
- Rachburee, Nachirat., Wattana Punlumjeak, 2015. *A Comparison of Feature Selection Approach Between Greedy, IG, Chi-Square, and mRMR in Educational Mining*. 2015 7<sup>th</sup> International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thailand.
- Satapaty, R., Cambria E., Hussain A., 2017. *Sentiment Analysis in the Bio-Medical Domain*. Springer International Publishing AG, Gwerbestrasse 11, 6630 Cham, Switzerland. <https://doi.org/10.1007/978-3-319-68468-0>.
- Yadav, S. H., Balu L. Pame., 2015. *A Survey on Different Text Categorization Techniques for Text Filtration*. *IEEE Sponsored 9<sup>th</sup> International Conference on Intelligent System and Control (ISCO)*, 2013.
- Ye, Qiang., Ziqiong Zhang., Rob Law. *Sentiment Classification of Online Reviews To Travel*

*Destinations By Supervised Machine Learning Approaches*. Expert System with Applications, 36 (2009), 6527-6535.

Yusuf, 2010. *Chi-Square Independence Test*. <http://www.spss-tutorials.com/chi-square-independent-test/>. Di akses pada 10 Oktober 2018.

