

DETEKSI KOMENTAR *BULLYING* DI MEDIA SOSIAL DENGAN *NEURAL NETWORK*

Ladys Maopatni Sinaga¹, Ednawati Rainarli²

^{1,2} Teknik Informatika – Universitas Komputer Indonesia

^{1,2} Jl. Dipatiukur 112 – 114 Bandung

e-mail: sinagaladys@gmail.com¹, ednawati.rainarli@email.unikom.ac.id²

ABSTRAK

Media sosial Facebook merupakan salah satu media sosial yang banyak digunakan. Bebasnya pengguna dalam berkomentar menjadi salah satu alasan terjadinya konten yang bersifat melukai (*abuse*) dan berakibat *bullying* (perundungan) di media sosial. Pemanfaatan *machine learning* dapat digunakan untuk mendeteksi komentar yang mengandung *bully*. *Prerocessing* yang digunakan meliputi *case folding*, *cleaning*, *convert negation*, *stopwords removal* dan *tokenizing*. TF-IDF digunakan dalam menghitung pembobotan setiap *term*. Metode *Neural Network Backpropagation* menjadi metode yang digunakan dalam penelitian ini. Hasil pengujian yang didapatkan dengan mengubah nilai *epoch* dan *learning rate* mendapat nilai rata – rata *Precision* sebesar 0.52, *Recall* dengan nilai rata – rata sebesar 0.40, dan *f-measures* sebesar 0.5 dimana akurasi terendah adalah 0.38 dan tertinggi adalah 0.8.

Kata Kunci : *Machine learning*, Facebook, *bully*, *Neural Network Backpropagation*, *Preprocessing*, TF-IDF

1. PENDAHULUAN

1.1. Latar Belakang Masalah

Facebook merupakan salah satu media sosial yang banyak digunakan di Indonesia. Media sosial ini juga sering digunakan karena kemudahan dalam interaksi antar pengguna, cepatnya penyebaran informasi, dan kebebasan dalam menanggapi ‘status’ antar pengguna. Bebasnya pengguna dalam *posting* dan berkomentar menjadi salah satu alasan terjadinya konten yang bersifat melukai (*abuse*) dan berakibat *bullying* (perundungan) di media sosial. Untuk dapat mendeteksi *bullying* dapat dilakukan dengan memanfaatkan *machine learning*. Penggunaan *machine learning* dalam pemrosesan teks sudah pernah dilakukan sebelumnya oleh E Rainarli dan K E Dewi dalam peringkasan teks [1].

Penelitian mengenai deteksi *bullying* di media sosial sebelumnya sudah pernah dilakukan oleh Huascar Sanchez dan Shreyas Kumar [2]. Pada

penelitian tersebut media sosial yang digunakan adalah twitter dengan mendeteksi kalimat yang memiliki kata “*Gay*”, “*Homo*”, “*Dike*”, “*Queer*”. Metode klasifikasi yang digunakan adalah Naive Bayes dan hasil akurasi pada penelitian tersebut mencapai 70%.

Penelitian yang berkaitan dengan klasifikasi teks sudah dilakukan sebelumnya. Penelitian yang dilakukan oleh Maulana Aziz Assuja mengenai [3], menunjukkan bahwa berdasarkan hasil penelitian, pelatihan, dan pengujian yang dilakukan mendapatkan nilai keakuratan sebesar akurasi 78.34% dan presisi 84.21%.

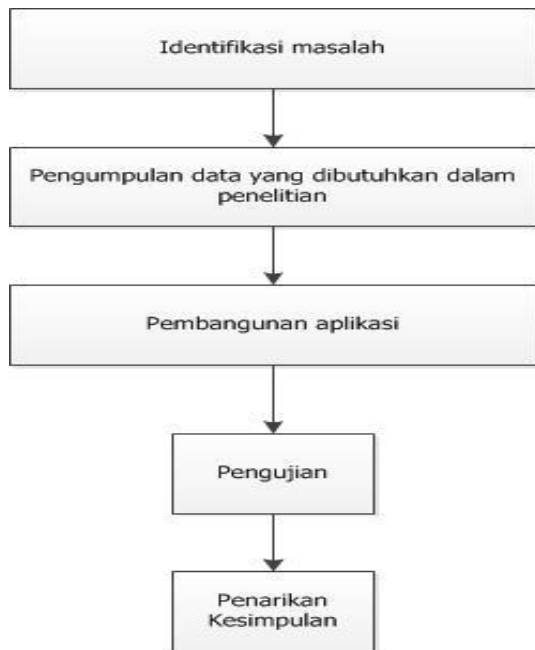
Feature yang digunakan yaitu dengan menambahkan *Contextual feature* setelah proses *preprocessing*. *Contextual feature* merupakan pendekatan kontekstual yang mencakup jumlah istilah, arti kata, dan kedekatan istilah (frasa atau kolokasi)[4]. Menurut penelitian yang dilakukan sebelumnya oleh Jensen S Lee dan Tony Martinez [3], menunjukkan bahwa ini dapat mengurangi error sebanyak 33%, namun untuk mendekati tingkat klasifikasi manusia, model pembelajaran harus dapat mengabstraksi informasi pelatihan dan menggabungkan dengan pengetahuan eksternal yang tidak tersedia dalam data pelatihan itu sendiri. Pendekatan sederhana untuk merepresentasikan hubungan kontekstual dari istilah adalah dengan menghitung berapa kali suatu istilah muncul dengan istilah lain dalam kerangka kontekstual yang diberikan. Definisi konteks ini sering disebut sebagai bigrams (konteks dua istilah) atau N-gram (konteks istilah N)

Berdasarkan apa yang sudah dipaparkan di atas, penelitian ini akan mengimplementasikan Metode klasifikasi *Neural Network Backpropagation* dengan *Contextual feature*.

1.2. Metodologi Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah Kualitatif. Metode Kualitatif merupakan suatu metode yang bersifat deskriptif dan cenderung analisis, berangkat dari data, memanfaatkan teori yang ada sebagai panduan agar fokus penelitian sesuai.

Langkah-langkah yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 1



Gambar 1. Tahapan Metodologi Penelitian

2. ISI PENELITIAN

2.1. Perundungan (Bullying)

Perundungan (*Bullying*) merupakan suatu tindakan yang bersifat melukai seseorang secara fisik ataupun mental. Dalam media sosial, perundungan merupakan salah satu dari tindakan kejahatan dunia maya (*Cyber crime*)

2.2. Media sosial

Media sosial merupakan sarana komunikasi digital yang terhubung secara online melalui jaringan Internet. Media sosial memudahkan manusia dalam berkomunikasi dengan jarak dekat maupun jauh dan penyebaran informasi yang cepat. Beberapa contoh dari media sosial adalah Facebook, Twitter, G-mail.

2.3. Facebook Automation

Facebook Automation merupakan salah satu *tool* untuk memaksimalkan kegiatan marketing secara online. Fitur-fitur yang terdapat pada *tool* yaitu dapat menambah, menghapus, mengubah ataupun mengambil (*scraping*) pesan, status, komentar, dan informasi yang terkandung pada profile individu, grup komunitas, event, halaman yang disukai di Facebook. Dalam penelitian ini Facebook Automation dijadikan sebagai *tool* untuk mengambil data set yang dibutuhkan. Hasil *scraping* disimpan dalam bentuk .CSV

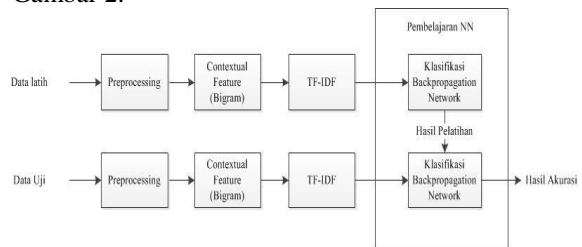
Tabel 1 Contoh Data Latih yang telah dilakukan *Scraping*

Data Latih	Kategori	Komentar
P1	<i>Bully</i>	kalo udah jelek dari awal ya kalah!!!! Ga usah bacod!!

P2	<i>Non Bully</i>	Lawan lunox sama gusion!!cukup pake harley udah cukup!!
P3	<i>Bully</i>	Jirr evos main noob udh ga berbobot ngapain main njeeng!

2.4. Analisis Proses

Proses yang dilakukan dalam penelitian ini meliputi beberapa tahap yang dapat dilihat pada Gambar 2.



Gambar 2. Tahapan Proses Yang Dilakukan

Tahap pertama yaitu *training*, load data set lalu dilakukan proses *preprocessing* setelah itu dilakukan ekstraksi fitur menggunakan *Contextual feature* dan pembobotan TF-IDF setelah itu dilakukan perhitungan *Backpropagation*.

2.5. Preprocessing

Preprocessing adalah proses pengolahan data untuk mempersiapkan data yang tidak terstruktur menjadi data yang terstruktur dan siap untuk proses selanjutnya[5]. Tahapan – tahapan yang dilakukan dalam penelitian ini yaitu *Case Folding*, *Cleaning*, *Convert Negation*, *Tokenizing*, dan *Stopword Removal*. Blok diagram *preprocessing* dapat dilihat pada gambar 3.



Gambar 3. Blok diagram *preprocessing*

a. Case folding

Case folding merupakan salah satu proses *preprocessing*. Tahapan ini bertujuan untuk menyeragamkan setiap huruf dengan mengubah semua huruf dalam dokumen menjadi *lowercase* (huruf kecil)

Tabel 2 Hasil Menggunakan *Case Folding*

Pernyataan	Hasil Case Folding
kalo udah jelek dari awal ya kalah!!!! Ga usah bacod!!	kalo udah jelek dari awal ya kalah!!!! ga usah bacod!!
Lawan lunox sama gusion !!cukup pake harley udah cukup!!	lawan lunox sama gusion!!cukup pake harley udah cukup!!
Jirr evos main noob udh ga berbobot ngapain main njeeng!	jirr evos main noob udh ga berbobot ngapain main njeeng!

b. Cleaning

Cleaning adalah tahapan untuk menghapus yang tidak perlu seperti tanda baca. Berikut hasil *cleaning* dapat dilihat pada gambar 3.

Tabel 3. Hasil Menggunakan *Cleaning*

Hasil Case Folding	Hasil Cleaning
kalo udah jelek dari awal ya kalah!!!! ga usah bacod!!	kalo udah jelek dari awal ya kalah ga usah bacod
lawan lunox sama gusion!!cukup pake harley udah cukup!!	lawan lunox sama gusion cukup pake harley udah cukup
jirr evos main noob udh ga berbobot ngapain main njeeng!	jirr evos main noob udh ga berbobot ngapain main njeeng

c. *Convert Negation*

Convert Negation adalah tahapan untuk menggabungkan kata negasi dengan kata setelahnya. Berikut hasil *Convert Negation* dapat dilihat pada gambar 4.

Tabel 4. Hasil Menggunakan *Convert Negation*

Hasil Cleaning	Hasil Convert Negation
kalo udah jelek dari awal ya kalah ga usah bacod	kalo udah jelek dari awal ya kalah gausah bacod
lawan lunox sama gusion cukup pake harley udah cukup	lawan lunox sama gusion cukup pake harley udah cukup
jirr evos main noob udh ga berbobot ngapain main njeeng!	jirr evos main noob udh gaberbobot ngapain main njeeng!

d. *Tokenizing*

Tokenizing adalah tahapan untuk memotong *string input* berdasarkan tiap kata yang menyusunnya. Berikut hasil *Tokenizing* dapat dilihat pada gambar 5.

Tabel 5. Hasil Menggunakan *Tokenize*

Hasil Convert Negation	Hasil Tokenize
kalo udah jelek dari awal ya kalah gausah bacod	kalo udah jelek dari awal ya kalah gausah bacod
lawan lunox sama gusion cukup pake harley udah cukup	lawan lunox sama gusion cukup pake harley

	udah cukup
jirr evos main noob udh gaberbobot ngapain main njeeng	jirr evos main noob udh gaberbobot ngapain main njeeng

e. *Stopword Removal*

Stopword Removal tahapan ini untuk menghapus kata yang tidak deskriptif. Berikut hasil *Stopword Removal* dapat dilihat pada gambar 6.

Tabel 6. Hasil Menggunakan *Stopword Removal*

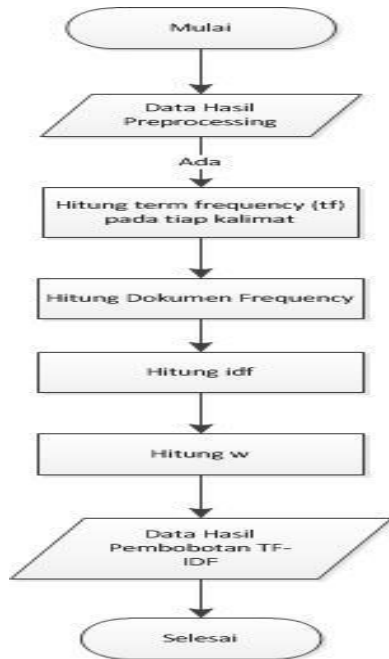
Hasil Tokenizing	Hasil Stopword Removal
kalo udah jelek dari awal ya kalah gausah bacod	jelek awal kalah gausah bacod
lawan lunox sama gusion cukup pake harley udah cukup	lawan lunox gusion harley
Jirr evos main noob udh gaberbobot ngapain main njeeng	jirr evos main noob gaberbobot main njeeng

2.6. Analisis TF-IDF

TF-IDF adalah teknik pembobotan pada setiap kata atau *term*, menerapkan pembobotan kombinasi yaitu perkalian bobot lokal (tf) dan bobot global (idf). Nilai idf sebuah *term* (kata) dapat dihitung menggunakan

$$\text{IDF} = \log\left(\frac{D}{df}\right)$$

Alur proses TF-IDF yaitu menghitung per-term yang dapat dilihat pada gambar 4.



Gambar 4. Flowchart perhitungan TF-IDF

- Menghitung Term Frekuensi (tf)
- Menghitung Document Frekuensi (df)
- Menghitung Invers Document Frekuensi

$$\text{IDF} = \log\left(\frac{D}{df}\right) = \log\left(\frac{3}{1}\right) = 0.477$$

- Menghitung bobot / weight.

Menghitung bobot (w) kata “jelek” pada setiap pernyataan dapat dihitung dengan berikut ini:

$$W_{p1} = \text{tf}_{p1} * \text{idf} = 1 * 0.477 = 0.477$$

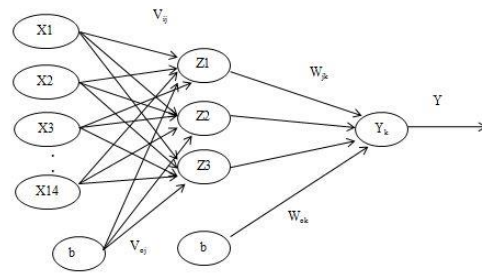
$$W_{p2} = \text{tf}_{p2} * \text{idf} = 1 * 0.477 = 0$$

$$W_{p3} = \text{tf}_{p3} * \text{idf} = 0 * 0.477 = 0$$

2.7. Backpropagation

Algoritma *neural network backpropagation* melakukan dua tahap perhitungan yaitu perhitungan maju untuk menghitung *error* antara keluaran *actual* dan *target*, dan perhitungan mundur untuk mempropagasikan baik *error* untuk memperbaiki bobot pada semua *neuron* yang ada.

Backpropagation terdiri dari lapisan-lapisan diantaranya lapisan masukan (*input layer*), lapisan yang terdiri dari neuron input baik dari *neuron* pertama sampai *neuron network*. Lapisan tersembunyi (*hidden layer*), terdiri dari minimal satu lapis layer mulai dari neuron tersembunyi pertama dan seterusnya, dan lapisan keluaran (*output*).



Gambar 5. Arsitektur Backpropagation (maju)

Keterangan :

X : Nilai Masukan (*Input layer*)

V_{ij} : Nilai bobot sambungan pada lapisan *input* X ke Unit Z

V_{oj} : Bias pada lapisan tersembunyi X ke unit Z

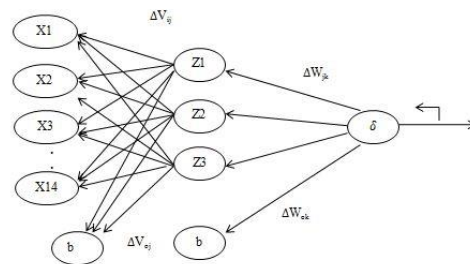
Z : Bobot pada lapisan tersembunyi (*hidden layer*)

W_{jk} : Nilai perbaikan sambungan pada lapisan tersembunyi Z ke unit Y

W_{ok} : Keluaran (*output*)

b : Bias

Y : Keluaran (menggunakan satu output)



Gambar 6. Arsitektur Backpropagation (mundur)

Keterangan :

X : Nilai Masukan (*Input layer*)

ΔV_{ij} : Nilai bobot sambungan pada lapisan *input* X ke Unit Z

ΔV_{oj} : Bias sambungan pada lapisan *input* X ke Unit Z

Z_i : Bobot pada lapisan tersembunyi (*hidden layer*)

ΔW_{jk} : Nilai bobot sambungan dari Z ke unit Y

ΔW_{ok} : Bias pada sambungan dari Z ke Y

b : Bias pada lapisan tersembunyi dan lapisan keluaran

δ : Faktor pengaturan nilai penimbang sambungan pada lapisan keluaran.

2.8 Pelatihan Backpropagation

Pada tahap ini terdapat variabel pelatihan yang ditentukan yaitu : *input layer*, jumlah *neuron* pada *hidden layer*, *learning rate*, batas epoch, dan batas iterasi (*error*). Variabel pelatihan pada penelitian ini berdasarkan penelitian Maulana Aziz Assuja[2]. Berikut adalah variabel pelatihan yang digunakan.

- Jumlah *term* yang diekstraksi dengan bigrams dan TF-IDF (N) = 14
- Input layer* $\rightarrow 14(+1 \text{ Bias Unit}) = 15 \text{ perceptron}$
- Jumlah *neuron* pada *hidden layer* $N*3 = 14*3 (+1 \text{ Bias Unit}) = 103 \text{ perceptron}$
- Learning rate* = 0.1
- Batas epoch = 100
- Batas iterasi (*error*) = 10%

Berikut merupakan hasil perbaikan bias dan bobot *input layer* ke *hidden layer*.

Tabel 7. Hasil Perbaikan Bias Dan Bobot *input layer* ke *hidden layer*.

	vi.1	vi.2	vi.3
X1	0,1 * 0,00003 * 1,87055 = 0,00019	0,1 * 0,00008 * 1,87055= 0,000238	0,1 * 0,00007 * 1,87055 = 0,00018
X2	0,1 * 0,00003 * 1,97064 = 0,00020	0,1 * -0,00008 * 1,97064 = 0,000251	0,1 * 0,00007 * 1,97064 = 0,00018
X3	0,1 * 0,00003 * 1,82892 = 0,00019	0,1 * 0,00008 * 1,82892=0,00023 3	0,1*0,00007 * 1,82892 = 0,00017
Bias	0,1 * - 0,00003 = 0,0001	0,1 * -0,00008 = 0,000127	0,1 * 0,00007 = 0,00009

Berikut merupakan perbaikan bias menggunakan dan bobot menggunakan *hidden layer* ke *output layer* dan *input layer* ke *hidden layer* dapat dilihat pada tabel 8

Tabel 8 Hasil perbaikan bias dan bobot *hidden layer* ke *output layer*.

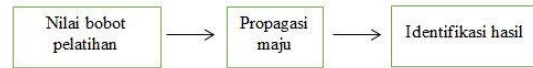
	Vi.1baru	Vi.2baru	Vi.3baru
X1	0,86851 +(- 0,00019) = 0,86831	0,18106+ (- 0,000238) = 0,18082	0,71324+(- 0,00018) = 0,71307
X2	0,14882 +(- 0,00020) = 0,14862	0,23276+(- 0,000251) = 0,23251	0,69606+(- 0,00018) = 0,69588
X3	0,49377+(- 0,00019) = 0,49358	0,65987+(- 0,000233) = 0,65963	0,46591+(- 0,00017) = 0,46574
Z	0,33131+0,00027=	0,70253 +	0,56664+(-

	0,33194	(- 0,000127) = 0,70320	0,00009) = 0,56726
--	---------	---------------------------------	-----------------------

Nilai bias dan bobot diatas merupakan nilai bias dan bobot baru yang akan digunakan sebagai iterasi selanjutnya

2.9 Pengujian Backpropagation

Pengujian backpropagation merupakan tahapan pengujian menggunakan metode backpropagation. Pada tahap ini hanya melalui proses propagasi maju untuk indentifikasi pola yang sudah ditentukan.



Gambar 7. Block Diagram Pengujian Backpropagation

a. Hasil Pengujian Model

Hasil pengujian model dengan data *training* 50 dan data *testing* 50. Pengujian dilakukan dengan menggunakan nilai *Epoch* 100 dan 600 dan nilai *learning rate* 0.1, 0.01 dan 0.001 Hasil pengujian dapat dilihat pada table 9.

Tabel 9. Hasil Pengujian Model

<i>Epoch</i>	<i>Learning rate</i>	Akura si	<i>Recall</i>	<i>Precision</i>	<i>F-measure</i>
100	0.1	0.7	0.70	0.69	0.69
100	0.01	0.68	0.68	0.69	0.63
100	0.001	0.74	0.74	0.75	0.72
600	0.1	0.72	0.72	0.73	0.69
600	0.01	0.74	0.74	0.78	0.71

b. Hasil pengujian performa

Hasil pengujian performa dengan data *training* 100 dan data *testing* 50. Hasil skenario dapat dilihat pada tabel 10

Tabel 10. Hasil Pengujian Performa

<i>Epoch</i>	<i>Learning rate</i>	Akura si	<i>Recall</i>	<i>Precision</i>	<i>F-measure</i>
100	0.1	0.62	0.62	0.38	0.47
100	0.01	0.72	0.72	0.76	0.68
100	0.001	0.78	0.78	0.79	0.78
600	0.1	0.72	0.72	0.81	0.66
600	0.01	0.78	0.78	0.81	0.76

2.9 Evaluasi Pengujian

Evaluasi pengujian terhadap deteksi teks *bullying* yang telah dilakukan dengan membandingkan hasil deteksi teks. Pengujian yang dilakukan perhitungan dengan persamaan 2.19 (*recall*), persamaan 2.20 (*precision*) dan persamaan 2.21 (*f-measure*).

3. PENUTUP

3.1. Kesimpulan

Berdasarkan hasil evaluasi pengujian dengan perhitungan nilai rata – rata *precision*, *recall* dan *F-measures* pada tabel 4.7. Nilai rata – rata yang didapat yaitu *Precision* sebesar 0.73 yang menunjukkan kata relevan yang dihasilkan oleh sistem. *Recall* dengan nilai rata – rata sebesar 0.73 yang menunjukkan kata relevan pada deteksi teks manual sehingga dapat disimpulkan nilai rata – rata hasil deteksi teks manual dan hasil deteksi teks sistem dengan rata – rata *f-measures* sebesar **0.70** dimana akurasi terendah adalah 0.62 dan tertinggi adalah 0.82.

3.2. Saran

Saran untuk pengembangan deteksi teks bullying di media sosial berbahasa Indonesia dari penelitian ini adalah :

1. Data set yang digunakan bisa ditambahkan untuk mengetahui apakah adanya peningkatan dalam pembelajaran machine learning. Data set dapat diganti dari media sosial lain dan bahasa typo dapat di atasi sehingga meminimalisir kesalahan pada tahap preprosesing.
2. Mengubah metode yang digunakan untuk dapat membandingkan kinerja metode dalam mendeteksi teks bullying bahasa Indonesia.

DAFTAR PUSTAKA

- [1] Rainarli, E., and K. E. Dewi. “Relevance Vector Machine for Summarization.” *IOP Conference Series: Materials Science and Engineering*. Vol. 407. No. 1. IOP Publishing, 2018.
- [2] Huascar Sanchez., Shreyas Kumar, “*Twitter Bullying Detection*” , Journal ser. NSDI, Vol.12, UC Santa Cruz.
- [3] Maulana Aziz Assuja., Saniati, “Analisis Sentimen *Tweet* Menggunakan *Backpropagation Neural Network*”, Jurnal TEKNOINFO, Vol.10, No.2, ISSN: 1693-0010.
- [4] Lee S. Jensen., Martinez, Tony., “Improving Text Classification by Using Conceptual and Contextual Features”, Computer Science Department, Brigham Young University.
- [5] Olweus, D., “Bullying at school: What we know and what we can do”, New York: Blackwell, 1993.
- [6] Siti Mujilahwati, “Pre-Processing Text Mining Pada Data Twitter”, Seminar Nasional Teknologi Informasi dan Komunikasi, Yogyakarta, 2016.
- [7] Irma Pujadayanti., Moch. Ali Fauzi., Yuita Arum Sari, “Prediksi Rating Otomatis pada Ulasan Produk Kecantikan dengan Metode Naïve Bayes dan N-gram” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 2, No. 11, pp 4421-

4427.

- [8] Dr. Suyanto, S.T., M.Sc., “Data Mining Untuk Klasifikasi Dan Klusterisasi Data”, Informatika Bandung, Bandung, 2017.