



Classification of Responses to The Implementation of The Internship Program Using Naive Bayes Classifier Method

Klasifikasi Tanggapan Pelaksanaan Program Magang dengan Menggunakan Metode Naive Bayes Classifier

Rafly Septianarta Putra^{1*}, Iis Dewi Ratih²

^{1,2}Departemen Statistika Bisnis, Fakultas Vokasi, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
E-Mail: ¹raflyseptian@gmail.com, ²iis.dewi@statistika.its.ac.id

Received Jun 01st 2021; Revised Jul 18th 2021; Accepted Aug 22th 2021
Corresponding Author: Rafly Septianarta Putra

Abstract

An organization or individual requires an evaluation of the implementation to see the development of a given quality. Satisfaction analysis is always carried out by organizations or companies that are doing quality development. The Department of Business Statistics is a department at the Sepuluh Nopember Institute of Technology, Surabaya, which carries out the mandate to oversee statistics students in the applied field. The developments that occurred gave birth to many changes and renewal of academic learning patterns, one of which was the existence of apprenticeship-based learning. The year 2020 is the first year of internship implementation and there is no measuring tool to determine the evaluation of its implementation so that an evaluation tool is needed related to the implementation of the internship. One application to achieve this is to create an innovation by knowing student satisfaction responses using sentiment analysis with the Naive Bayes Classifier (NBC). The visualizations display the percentage of satisfaction with the implementation of the internship, 49% dissatisfied and 51% satisfied. The results of the NBC model classification accuracy on the testing data are 78.9%.

Keyword: Classification, Evaluation, Internship, Naive Bayes Classifier (NBC), Response, Sentiment Analysis

Abstrak

Suatu organisasi atau individu memerlukan evaluasi pelaksanaan untuk melihat perkembangan kualitas yang diberikan. Analisis kepuasan selalu dilakukan oleh organisasi atau perusahaan yang sedang melakukan pengembangan kualitas. Departemen Statistika Bisnis merupakan departemen yang ada di Institut Teknologi Sepuluh Nopember Surabaya yang menjalankan amanah untuk menaungi mahasiswa statistika pada bidang terapan. Perkembangan yang terjadi melahirkan banyak perubahan dan pembaharuan pola pembelajaran akademik, salah satunya adalah adanya pembelajaran berbasis magang. Tahun 2020 merupakan tahun pertama pelaksanaan magang dan belum ada alat ukur untuk mengetahui evaluasi pelaksanaannya sehingga diperlukan alat untuk evaluasi terkait dengan pelaksanaan magang. Salah satu penerapan untuk mewujudkan hal tersebut adalah menciptakan sebuah inovasi dengan mengetahui tanggapan kepuasan mahasiswa menggunakan analisis sentimen dengan *Naive Bayes Classifier (NBC)*. Hasil visualisasi menampilkan jumlah persentase kepuasan pelaksanaan magang, sebesar 49% tidak puas dan 51% puas.

Kata Kunci: Analisis Sentimen, Evaluasi, Klasifikasi, Magang, Tanggapan, *Naive Bayes Classifier (NBC)*

1. PENDAHULUAN

Evaluasi merupakan salah satu hasil penerapan dari penelitian yang digunakan untuk menentukan berhasil atau tidaknya manfaat/nilai dari program atau kebijakan yang telah dilakukan [1]. Tentunya evaluasi digunakan untuk peningkatan kualitas dari sebuah pelaksanaan terhadap suatu kegiatan. Analisis kepuasan dilakukan oleh organisasi atau perusahaan sebagai evaluasi untuk pengembangan kualitas. Departemen Statistika Bisnis (DSB) ITS merupakan salah satu departemen yang ada Fakultas Vokasi ITS yang Sejak tahun 2018 telah mengalami perubahan jenjang dari program studi Diploma 3 menjadi program studi Sarjana Terapan. Perubahan jenjang mengakibatkan terciptanya subyek-subyek pembelajaran yang berbasis terapan, seperti adanya program pembelajaran yang berbasis *internship* atau disebut magang. Magang merupakan hal

yang wajib bagi mahasiswa di Fakultas Vokasi. Sebagai penyelenggara pelaksanaan magang pertama, perlu adanya evaluasi terkait dengan pelaksanaan magang di DSB.

DSB melakukan riset kepuasan mahasiswa terhadap pelayanan kemahasiswaan setiap tahunnya. Riset kepuasan mahasiswa terhadap layanan kemahasiswaan didapat dari hasil survei yang dilaksanakan pada periode waktu tertentu, namun belum ada spesifikasi yang membawa fokus riset pada pelaksanaan magang. Selain itu, tidak disediakan pertanyaan yang bersifat terbuka untuk dapat dijawab sebagai tanggapan atau opini dari pelaksanaan. Sehingga hasil evaluasi kurang maksimal dilakukan. Salah satu alat yang dapat digunakan untuk menangani hal tersebut adalah dengan menggunakan pengolahan *text mining* dengan analisis sentimen.

Text mining merupakan variasi dari data *mining* yang berusaha menemukan pola menarik dari sekumpulan data tekstual yang berjumlah besar [2]. Selain itu, terdapat pengolahan data teks dengan menggunakan analisis sentimen. Analisis sentimen (*opinion mining*) merupakan studi komputasional dari opini, sentimen, dan emosi melalui atribut yang dimiliki yang diekspresikan dalam bentuk teks. Berdasarkan pengertian tersebut, analisis sentimen dapat digunakan sebagai alat analisis kepuasan mahasiswa terhadap pelaksanaan magang di DSB. Metode klasifikasi yang digunakan adalah *Naïve Bayes Classifier* (NBC). NBC dipilih karena memiliki beberapa kelebihan untuk diterapkan pada sentimen kepuasan mahasiswa. Penelitian yang ditulis oleh Wati tahun 2016 menjelaskan bahwa metode NBC merupakan metode paling sederhana dari pengklasifikasian probabilitas dan memiliki tingkat akurasi yang sangat tinggi ketika diaplikasikan pada *database* dan *big data* [3]. Selain itu, Penelitian oleh Liu tahun 2016 juga menyebutkan bahwa Jika NBC dibandingkan dengan metode klasifikasi yang lain, NBC memiliki tingkat kesalahan yang minimum [4]. Diharapkan dari model yang terbentuk, dapat digunakan sebagai pendukung di tahun-tahun yang akan datang dari analisis kepuasan mahasiswa terhadap pelayanan akademik terutama magang di DSB ITS yang dilakukan setiap periode waktu tertentu.

2. METODOLOGI DAN TEORI PENELITIAN

Sumber data yang digunakan merupakan data primer yang diperoleh dari hasil survei secara *online* menggunakan *typeform*. Survei dilaksanakan mulai 19 Maret 2021 hingga 11 April 2021. Unit Penelitian yang digunakan merupakan mahasiswa Departemen Statistika Bisnis ITS yang telah melaksanakan magang pada tahun 2020 sebanyak 96 responden. Variabel penelitian yang digunakan adalah menggunakan populasi dari data tanggapan pelaksanaan magang dan kategori kepuasan mahasiswa terhadap pelaksanaan magang. Langkah penelitian akan dijelaskan sebagai berikut.

1. Menyusun kuesioner dan melakukan survei penelitian sebagai alat untuk memperoleh data dari responden. dan melakukan survei kepada responden penelitian
 2. Melakukan pengecekan data tanggapan pelaksanaan magang dan kesesuaian kepuasan
 3. Melakukan *preprocessing* data (*case folding, remove punctuation, clean number, stemming, filtering, tokenisasi*)
 4. Melakukan pembobotan *TF Raw*
 5. Klasifikasi menggunakan NBC pada data tanggapan pelaksanaan magang dan kategori kepuasan pelaksanaan magang. Tahap penyusunannya adalah sebagai berikut.
 - a. Membagi data *training* dan data *testing* dengan proporsi 77 data *training* dan 19 data *testing*.
 - b. Membuat model NBC dengan data *training*
 - c. Menerapkan model klasifikasi pada data *training*
 - d. Menghitung ketepatan klasifikasi pada data *training* menggunakan *confusion matrix*.
 - e. Menerapkan model klasifikasi pada data *testing*
 - f. Menghitung ketepatan klasifikasi pada data *testing* menggunakan *confusion matrix*
- Langkah penelitian tersebut dapat disajikan pada diagram alir Gambar 1.

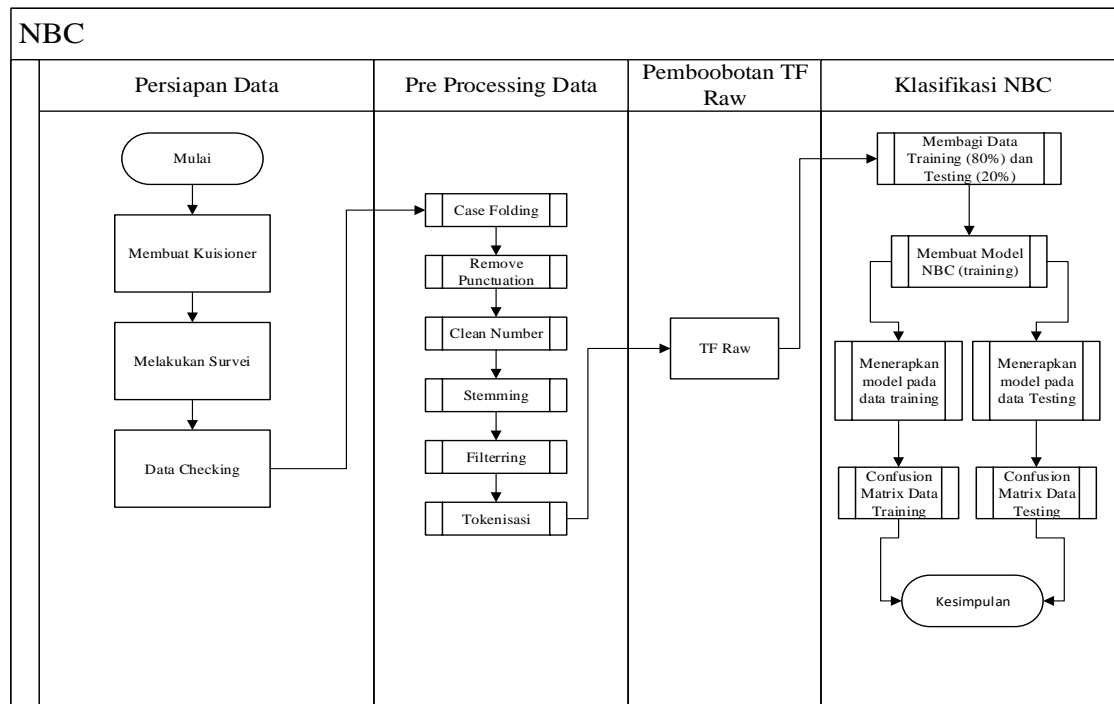
2.1 Text Mining

Text mining merupakan metode ekstraksi informasi dari pola yang tersirat, tidak diketahui, dan memiliki peluang bermakna serta berpotensi berharga yang didapat dari data tidak berstruktur seperti data teks [5]. Menurut penelitian yang dilakukan Fauziyyah tahun 2020, *text mining* dan data *mining* merupakan keilmuan yang sama. Hal tersebut dikarenakan memiliki algoritma yang sama untuk digunakan dalam *mining* [6]. Namun sebenarnya berbeda, hal tersebut dapat dijelaskan berdasarkan jenis data. Data *mining* menggunakan data terstruktur untuk pengolahannya, sedangkan *text mining* merupakan bidang yang saling terkait dengan pengolahan bahasa natural. *Text mining* digunakan sebagai ekstraksi informasi, *clustering*, klasifikasi, dan visualisasi informasi [7].

2.2 Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan teknik yang digunakan untuk analisa pendapat, sikap, evaluasi, perasaan serta emosi seseorang, tentang suatu topik, produk, layanan, individu, dan organisasi [8]. Analisis sentimen dapat digunakan pada penerapan tertentu sebagai upaya untuk mengetahui

makna dari sebuah kumpulan pendapat. Menurut penelitian yang dilakukan oleh Fink tahun 2011, Di sisi lain analisis sentimen dapat digunakan untuk mengevaluasi kinerja dan pelayanan. Jenis analisis yang paling sering digunakan adalah analisis dalam bentuk dokumen atau kalimat [9]. Hal tersebut didukung oleh penelitian yang dilakukan oleh Rozi dkk tahun 2012, menyebutkan bahwa analisis sentimen merupakan proses mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen dalam suatu kalimat opini, yang lalu akan dikelompokkan kecenderungan opini tersebut ke dalam kelompok positif, negatif, dan netral [10].



Gambar 1. Diagram Alir Penelitian

2.3 Preprocessing Data

Preprocessing merupakan langkah pertama yang penting dalam *text mining*. *Preprocessing* adalah proses pembersihan untuk menghilangkan karakter-karakter pada dokumen yang tidak dibutuhkan pada proses analisis teks [11]. Di dalam *Preprocessing* terdapat beberapa tahapan yang digunakan untuk melakukan pembersihan yaitu sebagai berikut *Case folding* bertujuan untuk mengubah teks menjadi jenis huruf kecil atau non kapital. *Remove punctuation* digunakan untuk menghapus atau menghilangkan karakter simbol atau tanda baca (*non-alphabet character*). *Clean number* berfungsi menghapus angka-angka yang ada pada data teks. Hal tersebut didasarkan pada kebiasaan seseorang dalam menuliskan pengulangan dengan menambahkan angka setelah kata ditulis. *Stemming* digunakan untuk menghapus imbuhan yang ada, sehingga kata menjadi kata dasar. *Filterring* berfungsi untuk mengambil kata penting dari hasil proses tokenisasi. Dapat menggunakan algoritma *stoplist* atau *stopword*. *Tokenisasi* berfungsi untuk membagi teks *input* yang digunakan dalam proses *preprocessing* ke dalam bentuk kata [12].

2.4 Pembobotan Term Frequency Raw (TF Raw)

TF RAW merupakan algoritma pembobotan yang didasarkan pada kemunculan *term* (istilah). Semakin sering sebuah *term* muncul, maka semakin tinggi bobot dokumen tersebut. TF RAW dapat diperoleh dengan menggunakan persamaan 1.

$$TF_{(t,d)} = f_{t,d} \tag{1}$$

Keterangan :

- $TF_{(t,d)}$: *Term Frequency*/ kemunculan kata ke-t pada dokumen ke-d
- $f_{t,d}$: Jumlah kata ke-t pada dokumen ke-d

2.5. Algoritma Naive Bayes (NBC)

Algoritma NBC merupakan pengklasifikasian yang dilakukan dengan metode probabilitas statistik yang ditemukan oleh ilmuwan inggris Thomas Bayes, yaitu melakukan prediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [13]. Keuntungan dari NBC adalah algoritmanya sederhana,

tetapi akurasi tinggi. Klasifikasi data teks menggunakan NBC dibagi menjadi dua tahap. Tahap pertama adalah melatih dokumen dengan kategori yang dikenal. Tahap kedua adalah proses klasifikasi dokumen yang tidak dikategorikan [14]. Persamaan algoritma Naïve Bayes dijelaskan pada Persamaan 2.

$$V_{map} = arg. max P(V_j) \prod_i P(a_i|V_j) \quad (2)$$

Keterangan :

$P(V_j)$: Probabilitas dokumen ke - j
$\prod_i P(a_i V_j)$: Probabilitas kemunculan term ke - i pada dokumen ke - j
V_{map}	: Probabilitas tertinggi dari semua kategori dokumen yang diujikan

Untuk dapat menentukan hasil Persamaan 2 diperlukan persamaan $P(V_j)$ dan $P(a_i|V_j)$ dapat dilihat pada persamaan 3.

$$P(V_j) = \frac{doc}{training} \quad (3)$$

Keterangan :

$P(V_j)$: Probabilitas dokumen ke - j
doc	: Jumlah dokumen yang memiliki kategori j dalam data <i>training</i>
$training$: Jumlah dokumen yang digunakan dalam <i>training</i>

$$P(a_i|V_j) = \frac{W_i}{N+kosakata} \quad (4)$$

Keterangan :

$P(a_i V_j)$: Probabilitas kemunculan term ke - i pada dokumen kategori-j
W_i	: bobot term ke - i yang memiliki kategori j
N	: Jumlah seluruh term dalam dokumen yang memiliki kategori - j
$kosakata$: Jumlah term dalam <i>data training</i>

2.6. Confusion Matrix

Confusion matrix adalah tabel yang digunakan untuk mengukur kinerja algoritma klasifikasi pada model klasifikasi [15]. Terdapat beberapa istilah pada *confusion matrix*. Di antaranya adalah sebagai berikut.

1. *True Positive (TP)*
Kasus di mana seseorang diprediksi sebagai benar dalam dugaan dan pada kenyataannya adalah sesuai dengan dugaan.
2. *True Negative (TN)*
Kasus di mana diprediksi tidak benar dalam dugaan dan pada kenyataannya adalah sesuai dengan dugaan
3. *False Positive (FP)*
Kasus di mana diprediksi tidak benar dalam dugaan namun pada kenyataannya benar
4. *False Negative (FN)*
Kasus di mana diprediksi tidak benar dengan dugaan dan pada kenyataannya memang tidak benar.

Berdasarkan nilai TP, TN, FP, dan FN dapat ditentukan ketetapan klasifikasi secara keseluruhan dengan menggunakan *accuracy*. Untuk spesifik ketepatan dapat diukur dengan *sensitivity* untuk kelas positif dan *recall/specificity* untuk kelas negatif [16].

a. Accuracy

Persentase *Classifier* benar melakukan prediksi dapat dicari dengan persamaan 5.

$$\frac{(TP+TN)}{n} \quad (5)$$

Keterangan :

TP	: <i>True positif</i>
TN	: <i>True negatif</i>
n	: Jumlah data

b. Recall

Persentase data positif yang diprediksi sebagai positif dapat dicari dengan persamaan 6.

$$\frac{FP}{TP+FN} \tag{6}$$

Keterangan :

- FP : False positif
- TP : True positif
- FN : False negatif

c. *Specificity*

Persentase data positif yang diprediksi sebagai positif dapat dicari dengan persamaan 7.

$$\frac{TN}{TN+FP} \tag{7}$$

Keterangan :

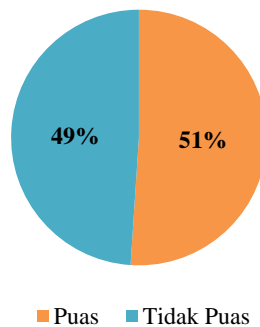
- FP : False positif
- TN : True negatif

3. HASIL DAN ANALISIS

Data tanggapan pelaksanaan magang akan diolah dengan menggunakan analisis sentimen dengan menggunakan NBC yang terdiri dari *preprocessing text*, pembobotan menggunakan TF Raw, pemodelan dengan NBC, dan terakhir evaluasi NBC untuk mengetahui nilai akurasi hasil klasifikasi.

3.1 Visualisasi Data Kepuasan Mahasiswa Terhadap Pelaksanaan Magang

Kepuasan mahasiswa merupakan kesimpulan tanggapan yang telah diberikan oleh mahasiswa atas pelaksanaan magang yang terdiri puas atau tidak puas. Kepuasan mahasiswa disajikan dalam bentuk persentase *pie chart*.



Gambar 2. *Pie Chart* Kepuasan Mahasiswa

Gambar 2 *pie chart* kepuasan mahasiswa terhadap pelaksanaan magang di DSB tahun 2020. Sebanyak 47 atau sebesar 49% mahasiswa memilih tidak puas terhadap pelaksanaan magang tahun 2020. Dan sebanyak 49 atau sebesar 51% mahasiswa memilih puas terhadap pelaksanaan magang DSB di tahun 2020.

3.2 *Preprocessing Data*

Data tanggapan responden yang telah terkumpul dilakukan *preprocessing data* yang meliputi mengubah huruf menjadi non kapital, menghapus tanda baca, menghapus angka, mengambil kata penting, dan menjadikannya kata dasar. Hasil *preprocessing* ditampilkan pada Tabel 1.

Tabel 1. Data sebelum dan Sesudah *preprocessing*

No.	Sebelum	Sesudah
1	“PENGINFORMASIAN hal terkait magang sudah dilaksanakan..”	“informasi kait magang laksana..”
2	“Bener bener sangat tidak siap ya departemen statistika ...”	“bener bener ya departemen statistika bisnis prosedur ...”
3	“Menurut saya kebijakan dan pelayanan cukup baik ...”	“turut saya bijak magang di departemen stabis tahun..”
:	:	:
94	“Waktu sosialisasi sangat terlambat sehingga kurangnya persiapan mahasiswa...”	“sosialisasi lambat mahasiswa pencarian magang..”
95	“Untuk angkatan 2017 magang terkesan terlalu	“angkat magang kesan dadak usaha terima

No.	Sebelum	Sesudah
96	mendadak. Perusahaan..” “Sosialisasi magang sangat tidak sesuai dengan kenyataan saat realisasi..”	mahasiswa magang langsung...” “sosialisasi magang nyata realisasi..”

Tabel 1 merupakan perbandingan antara data sebelum dan sesudah *preprocessing*. Data tanggapan pelaksanaan magang yang belum dilakukan *preprocessing* masih mengandung karakter kapital, angka, tanda baca dan kata hubung. Setelah dilakukan *preprocessing* data teks berubah menjadi karakter dasar dari kata yang sudah terbentuk.

3.3 Pembobotan TF Raw

Data tanggapan pelaksanaan magang DSB tahun 2020 yang telah dilakukan *preprocessing* data akan dilakukan pembobotan dengan menggunakan TF Raw, hasil ditampilkan pada Tabel 2.

Tabel 2. Pembobotan TF Raw

No	Terms	Dokumen ke-									
		1	2	3	4	5	6	...	94	95	96
1	informasi	1	0	0	0	0	0	...	0	0	0
2	kait	1	1	0	1	1	0	...	0	0	0
3	laksana	1	1	1	0	1	0	...	0	0	0
4	magang	1	4	2	1	1	0	...	2	7	1
5	prosedur	1	2	0	0	0	0	...	1	0	0
6	sosialisasi	1	0	0	0	0	0	...	1	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
351	peraturan	0	0	0	0	0	0	...	0	1	0
352	terima	0	0	0	0	0	0	...	0	2	0
353	klaim	0	0	0	0	0	0	...	0	0	1

Tabel 2 merupakan pembobotan menggunakan *TF Raw*. *Term* yang terbentuk adalah sebanyak 353 *terms*, *Term* diberi bobot berdasarkan kemunculan pada setiap dokumen yang ada.

3.4 Pembentukan Model NBC

Hasil pembobotan menggunakan *TF Raw* dilanjutkan dengan klasifikasi dengan menggunakan NBC. Data dibagi dengan porsi 80% data training atau sebanyak 77 data tanggapan. Sedangkan 20% merupakan data testing atau sebanyak 19 data tanggapan. Model NBC yang dibentuk merupakan model dengan menggunakan data *training*. Hasil probabilitas tanggapan dengan kategori puas dan tidak puas menggunakan Persamaan 3 disajikan pada Tabel 3.

Tabel 3. Probabilitas masing-masing kategori

Kategori	Probabilitas
Puas	0.52
Tidak Puas	0.48

Tabel 3 menjelaskan bahwa kategori tanggapan puas memiliki probabilitas sebesar 0.52 dan kategori tanggapan tidak puas adalah sebesar 0.48. Berdasarkan Persamaan 4 Hasil model yang terbentuk disajikan pada Tabel 4.

Tabel 4. Pembobotan TF Raw

No	Terms	$P(t_i V_{puas})$	$P(t_i V_{tidak\ puas})$
1	informasi	0.025	0.972
2	kait	0.200	0.783
3	laksana	0.175	0.810
4	magang	0.750	0.162
5	adakan	0.05	0.891
6	agustus	0.05	0.972
⋮	⋮	⋮	⋮
308	biar	0.025	1.00
309	aspek	0.025	1.00

Tabel 4 menjelaskan model probabilitas yang terbentuk adalah sebanyak 309 *terms* dengan masing-masing *term* memiliki probabilitas pada setiap kategori. Model diterapkan pada data testing sebanyak 19

data. Menggunakan Persamaan 2 dan nilai yang terbesar kategori merupakan hasil keputusan kategori berdasarkan NBC, hasil disajikan pada Tabel 5.

Tabel 5. Hasil NBC pada data *testing*

Dokumen ke-	Puas	Tidak Puas	Keputusan
78	0,009672	0,003270	Puas
79	0,008624	0,001375	Puas
80	0,001987	0,009801	Tidak Puas
81	0,003739	0,006260	Tidak Puas
82	0,009992	0,007778	Puas
83	0,003643	0,009963	Tidak Puas
84	0,009999	0,002028	Puas
85	0,002464	0,009753	Tidak Puas
86	0,001067	0,008932	Tidak Puas
87	0,009999	0,001106	Puas
88	0,001353	0,008646	Tidak Puas
89	0,009796	0,002039	Puas
90	0,009203	0,007965	Puas
91	0,008338	0,001661	Puas
92	0,008770	0,009991	Tidak Puas
93	0,001220	0,008779	Tidak Puas
94	0,001095	0,008904	Tidak Puas
95	0,009787	0,009999	Tidak Puas
96	0,009813	0,001867	Puas

Tabel 5 merupakan hasil keputusan dari pemodelan NBC menggunakan data *training* yang diterapkan pada data *testing*. Terdapat 9 tanggapan yang diklasifikasikan puas dan 10 tanggapan diklasifikasikan tidak puas.

3.5 Evaluasi Algoritma NBC

Evaluasi data dibagi menjadi dua yaitu penerapan dengan data *training* sendiri dan penerapan dengan data *testing* berdasarkan pembahasan sebelumnya.

a. Penerapan Model Pada Data *Training*

Data dengan pemodelan menggunakan data *training* diterapkan dengan data *training* sebanyak 77 data dengan cara yang sama seperti Tabel 5, hasil ditampilkan pada Tabel 6.

Tabel 6. Perbandingan Prediksi dan Aktual Tanggapan Data *Training*

Prediksi	Aktual	
	Puas	Tidak Puas
Puas	39	3
Tidak Puas	1	34

Tabel 6 merupakan jumlah tanggapan pada data *training* yang berhasil diprediksi dengan metode NBC. Sebanyak 39 mahasiswa yang puas tepat diprediksi puas, 34 mahasiswa yang memang tidak puas tepat diprediksi tidak puas. Sementara 1 mahasiswa puas diprediksi tidak puas. Hasil evaluasi yang didapatkan dari Persamaan 5-7 disajikan pada Tabel 7.

Tabel 7. Evaluasi NBC Data *Training*

<i>Accuracy</i>	<i>Specificity</i>	<i>Recall</i>
0.948	0.918	0.975

Tabel 7 merupakan hasil evaluasi dari penerapan menggunakan data *training*. Hasil *accuracy* adalah 0.948 yang artinya tanggapan yang benar diprediksi puas dan tidak puas memiliki persentase 94.8%. Hasil *specificity* adalah 0.918 yang artinya persentase data tanggapan tidak puas dan aktualnya tidak puas (sesuai) adalah sebesar 91.8%. Hasil *recall* sebesar 0.975 menunjukkan bahwa persentase klasifikasi data tanggapan yang diprediksi puas dan aktualnya puas adalah sebesar 97.5%.

b. Penerapan Model Pada Data *Testing*

Data dengan pemodelan menggunakan data *training* diterapkan dengan data *testing* sejumlah 19 tanggapan. Hasil ditampilkan pada Tabel 8.

Tabel 8. Perbandingan Prediksi dan Aktual Tanggapan Data *Testing*

Prediksi	Aktual	
	Puas	Tidak Puas
Puas	7	2
Tidak Puas	2	8

Tabel 8 merupakan jumlah tanggapan pada data *testing* yang berhasil diprediksi dengan metode NBC dengan merujuk pada Tabel 5. Sebanyak 7 mahasiswa yang puas tepat diprediksi puas. 8 mahasiswa yang memang tidak puas tepat diprediksi tidak puas. Sementara 2 mahasiswa puas diprediksi tidak puas. Hasil evaluasi yang didapatkan dari Persamaan 5-7 ditampilkan pada Tabel 9.

Tabel 9. Evaluasi NBC Data *Testing*

<i>Accuracy</i>	<i>Specificity</i>	<i>Recall</i>
0.789	0.800	0.778

Tabel 9 merupakan hasil evaluasi dari penerapan menggunakan data *testing*. Hasil *accuracy* adalah 0.789 yang artinya persentase klasifikasi benar melakukan prediksi adalah 78.9%. Hasil *specificity* adalah 0.80 yang artinya persentase data tanggapan tidak puas dan aktualnya tidak puas (sesuai) adalah sebesar 80%. Hasil *recall* sebesar 0.778 menunjukkan bahwa persentase klasifikasi data tanggapan yang diprediksi puas dan aktualnya puas adalah sebesar 77.8%.

4. KESIMPULAN

Kesimpulan pada penelitian ini adalah hasil evaluasi NBC menggunakan model data *training* yang diterapkan pada data *training* memiliki tingkat akurasi sebesar 94.8%. Sedangkan model yang diterapkan pada data *testing* memiliki akurasi sebesar 78.9%. Akurasi evaluasi model yang terbentuk menggunakan data *training* dapat ditingkatkan dengan menambah jumlah responden atau melakukan pembaharuan data peserta magang. Semakin banyak data *training* yang digunakan maka semakin tepat model dalam mengklasifikasikan tanggapan pelaksanaan magang

REFERENSI

- [1] S. Kantun, "Penelitian Evaluatif Sebagai Satu Model Penelitian Dalam Bidang Pendidikan," *Maj. Ilm. Din.*, vol. 37, no. 1, p. 15.
- [2] B. Kurniawan, S. Effendi, and O. S. Sitompul, "Klasifikasi Konten Berita Dengan Metode *Text Mining*," *J. Dunia Teknol. Inf.*, vol. 1, no. 1, pp. 14–19, 2012, [Online]. Available: <http://download.portalgaruda.org/article.php?article=58993&val=4123>.
- [3] R. Wati, "Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan," *J. Evolusi*, vol. 4, no. 1, pp. 25–31, 2016.
- [4] Y. Lin, X. Wang, and A. Zhou, "Opinion spam detection," *Opin. Anal. Online Rev.*, no. May, pp. 79–94, 2016, doi: 10.1142/9789813100459_0007.
- [5] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, "Text mining in big data analytics," *Big Data Cogn. Comput.*, vol. 4, no. 1, pp. 1–34, 2020, doi: 10.3390/bdcc4010001.
- [6] A. K. Fauziyyah, "Analisis Sentimen Pandemi Covid19 Pada *Streaming* Twitter Dengan *Text Mining Python*," *J. Ilm. SINUS*, vol. 18, no. 2, p. 31, 2020, doi: 10.30646/sinus.v18i2.491.
- [7] Y. N and M. S, "A Review on Text Mining in Data Mining," *Int. J. Soft Comput.*, vol. 7, no. 2/3, pp. 01–08, 2016, doi: 10.5121/ijsc.2016.7301.
- [8] F. Romadoni, Y. Umidah, and B. N. Sari, "Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma *Support Vector Machine*," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 2, p. 247, 2020, doi: 10.32736/sisfokom.v9i2.903.
- [9] C. R. Fink, D. S. Chou, J. J. Kopecky, and A. J. Llorens, "Coarse- and fine-grained sentiment analysis of social media text," *Johns Hopkins APL Tech. Dig. (Applied Phys. Lab.)*, vol. 30, no. 1, pp. 22–30, 2011.
- [10] I. Rozi, S. Pramono, and E. Dahlan, "Implementasi *Opinion Mining* (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi," *J. EECCIS*, vol. 6, no. 1, pp. 37–43, 2012.
- [11] L. Hermawan and M. Bellanier Ismiati, "Pembelajaran *Text Preprocessing* berbasis Simulator Untuk Mata Kuliah *Information Retrieval*," *J. Transform.*, vol. 17, no. 2, p. 188, 2020, doi: 10.26623/transformatika.v17i2.1705.
- [12] I. Hemalatha and A. Govardhan, "Preprocessing the Informal Text for efficient ALGORITHM FOR," *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 1, no. 2, p. 4, 2012.
- [13] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *J. Inform. Ahmad Dahlan*, vol. 8, no. 1, p. 102632, 2014, doi: 10.26555/jifo.v8i1.a2086.

- [14] Falahah and D. Dwiki Adriadi Nur, "PENGEMBANGAN APLIKASI SENTIMENT ANALYSIS MENGGUNAKAN METODE NAÏVE BAYES (Studi Kasus Sentiment Analysis dari media Twitter)," *Semin. Nas. Sist. Inf. Indones.*, no. November, pp. 2–3, 2015.
- [15] R. M. Faisal and D. T. Nugrahadi, *Belajar Data Science dengan Pemrograman R*, no. February. 2019.
- [16] M. R. Dewi, "Klasifikasi Akses Internet Oleh Anak-Anak dan Remaja Dewasa di Jawa Timur Menggunakan Support Vector Machine," *J. Ris. dan Apl. Mat.*, vol. 4, no. 1, p. 17, 2020, doi: 10.26740/jram.v4n1.p17-27.