

**PENGARUH SELEKSI CIRI *CHI SQUARE* PADA ALGORITMA *NAIVE BAYES* TERHADAP ANALISIS SENTIMEN MASYARAKAT INDONESIA
TENTANG PEMBELAJARAN TATAP MUKA
DI MASA PANDEMI COVID-19**

Tesis
untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S-2 Program Studi
Magister Sistem Informasi



AZIFA HABIBA
30000320410011

SEKOLAH PASCA SARJANA
UNIVERSITAS DIPONEGORO
SEMARANG
2022

LEMBAR PENGESAHAN

TESIS

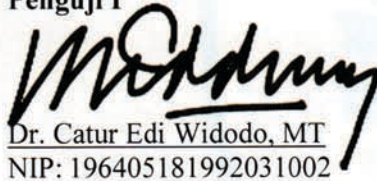
**PENGARUH SELEKSI CIRI *CHI SQUARE* PADA ALGORITMA
NAIVE BAYES TERHADAP ANALISIS SENTIMEN MASYARAKAT INDONESIA
TENTANG PEMBELAJARAN TATAP MUKA
DI MASA PANDEMI *COVID-19***

Oleh:
Azifa Habiba
30000320410011

Telah diujikan dan dinyatakan lulus ujian tesis pada tanggal 15 Juli 2022 oleh tim penguji Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Semarang, 25 Juli 2022
Mengetahui,

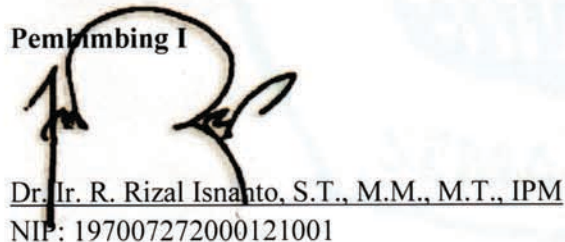
Penguji I


Dr. Catur Edi Widodo, MT
NIP: 196405181992031002

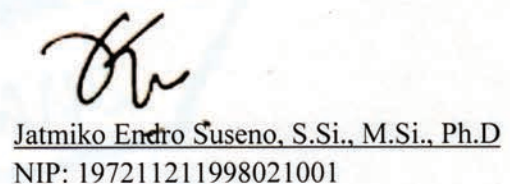
Penguji II


Dr. Adian Fatchur Rochim, S.T., M.T.
NIP: 197302261998021001

Pembimbing I

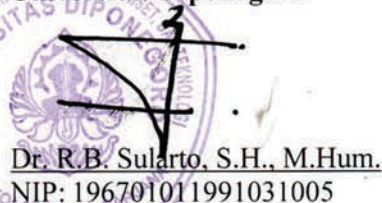

Dr. Ir. R. Rizal Isnanto, S.T., M.M., M.T., IPM
NIP: 197007272000121001

Pembimbing II

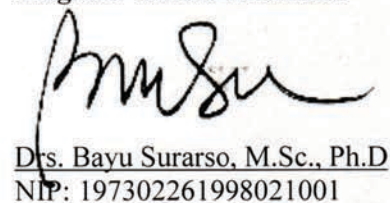

Jatmiko Endro Suseno, S.Si., M.Si., Ph.D
NIP: 197211211998021001

Mengetahui :

**Dekan Sekolah Pascasarjana
Universitas Diponegoro**


Dr. R.B. Sularto, S.H., M.Hum.
NIP: 196701011991031005

**Ketua Program Studi
Magister Sistem Informasi**


Drs. Bayu Surarso, M.Sc., Ph.D
NIP: 197302261998021001

**PERNYATAAN PERSETUJUAN
PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama : Azifa habiba
NIM : 30000320410011
Program Studi : Magister Sistem Informasi
Program : Sekolah Pascasarjana
Jenis Karya : Tesis


Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro Hak Bebas Royalti Noneksklusif atas karya ilmiah saya yang berjudul:

Pengaruh Seleksi Ciri *Chi Square* Pada Algoritma *Naive Bayes* Terhadap Analisis Sentimen Masyarakat Indonesia Tentang Pembelajaran Tatap Muka di Masa Pandemi *Covid-19*.

Dengan Hak Bebas Royalti Noneksklusif ini Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*) merawat, dan mempublikasikan tesis saya selama mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Dibuat di : Semarang
Pada tanggal : 25 Juli 2022

Yang menyatakan



Azifa Habiba

NIM. 30000320410011

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.



Semarang, 25 Juli 2022


Azifa Habri



KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah memberikan rahmat, karunia dan hidayah-Nya kepada penulis sehingga penulis dapat menyelesaikan penulisan Tesis ini yang berjudul “Pengaruh Seleksi Ciri Chi Square Pada Algoritma *Naïve Bayes* Terhadap Analisis Sentimen Masyarakat Indonesia Tentang Pembelajaran Tatap Muka di Masa Pandemi Covid-19’.

Penyusunan tesis ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan Program Pascasarjana Program Studi Magister Sistem Informasi Universitas Diponegoro Semarang.

Kepada pembimbing yang telah memberikan arahan, dukungan serta kesabaran dalam memberikan bimbingan kepada penulis, rasanya tiada kata yang pantas diucapkan selain terima kasih yang tak terhingga.

Tiada gading yang tak retak andai pun retak jadikanlah sebagai ukiran, begitupun dengan tesis ini masih banyak kekurangan dan jauh dari kesempurnaan. Untuk itu melalui kata pengantar ini penulis sangat terbuka menerima kritik serta saran yang membangun. Dalam penulisan tesis ini penulis menyampaikan rasa hormat, terimakasih dan penghargaan yang sebesar-besarnya kepada :

1. Dr. R.B. Sularto, S.H., M.Hum, selaku Dekan Sekolah Pascasarjana Universitas Diponegoro.
2. Drs. Bayu Surarso, M.Sc., Ph.D, selaku Ketua Program Studi Magister Sistem Informasi Universitas Diponegoro.
3. Dr. Ir. R. Rizal Isnanto, S.T., M.M., M.T., IPM, selaku Pembimbing I dalam penulisan tesis ini yang membimbing, memberikan arahan dan petunjuk demi sempurnanya tesis ini.
4. Jatmiko Endro Suseno, S.Si., M..Si., Ph.D selaku Pembimbing II dalam penulisan tesis ini yang membimbing, memberikan arahan dan petunjuk demi sempurnanya tesis ini.

5. Seluruh Guru Besar dan Dosen Program Magister (S2) Sistem Informasi yang telah mendidik selama penulis belajar pada Program Magister Sistem Informasi Universitas Diponegoro.
 6. Seluruh staf dan karyawan Pascasarjana Sistem Informasi Universitas Diponegoro.
 7. Suamiku tercinta serta anak-anakku tersayang yang telah memberikan semangat dalam penyelesaian tesis ini.
 8. Teman-teman Magister Sistem Informasi Universitas Diponegoro Angkatan 2020, bersama kalian studi ini sangat berkesan, yang saling mendukung, memotivasi demi selesainya studi ini.
 9. Semua pihak yang tidak bisa penulis sebutkan satu persatu.
- Akhirnya semoga penulisan ini dapat bermanfaat bagi penulis sendiri, maupun para pembaca yang memerlukan sebagai bahan literatur.

Semarang, 25 Juli 2022

Penulis



Azifa Habiba

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
PERNYATAAN PERSETUJUAN	iii
HALAMAN PERNYATAAN.....	iv
KATA PENGANTAR.....	v
DAFTAR ISI	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL	x
DAFTAR ARTI LAMBANG DAN SINGKATAN.....	xii
ABSTRAK	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN.....	16
1.1 Latar Belakang	16
1.2 Tujuan Penelitian	18
1.3 Manfaat Penelitian	18
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	19
2.1 Tinjauan Pustaka	19
2.2 Dasar Teori	21
2.2.1 Analisis Sentimen.....	21
2.2.2 Twitter	21
2.2.3 Python.....	22
2.2.4 <i>Text Mining</i>	22
2.2.5 Prapengolahan	23
2.2.6 <i>Chi Square</i>	25
2.2.7 <i>Naïve Bayes</i>	26
2.2.8 <i>Term Frequency Inverse Document Frequency (TF-IDF)</i>	27

2.2.9	<i>Confusion Matrix</i>	28
BAB III	METODE PENELITIAN	30
3.1	Bahan dan Alat Penelitian	30
3.2	Prosedur Penelitian	30
3.3	Kerangka Sistem Informasi	33
3.4	Perancangan <i>Interface</i>	35
BAB IV	HASIL DAN PEMBAHASAN.....	42
4.1	Hasil Penelitian	42
4.1.1	Implementasi Perangkat lunak.....	42
4.1.3	Kerangka Kerja.....	43
4.1.4	Metode yang Digunakan.....	45
4.1.5	Hasil Kajian	63
4.2	Hasil Pembahasan	64
4.2.1	Pengambilan Data Twitter.....	64
4.2.2	Simpan Data dengan Format CSV	64
4.2.3	Prapengolahan	65
4.2.4	Seleksi <i>Chi Square</i>	67
4.2.5	Algoritma <i>Naïve Bayes</i>	70
4.2.6	<i>Confusion Matrix</i>	73
4.4.7	Hasil Analisis Sentimen	80
BAB V	PENUTUP	82
5.1	Kesimpulan	82
5.2	Saran	83
DAFTAR PUSTAKA	84

DAFTAR GAMBAR

	Halaman
Gambar 2. 1 Tahap <i>tokenizing</i>	24
Gambar 2. 2 Tahap <i>filtering</i>	24
Gambar 2. 3 Tahap <i>Stemming</i>	25
Gambar 3. 1 Prosedur Penelitian	31
Gambar 3. 2 Kerangka sistem informasi	33
Gambar 3. 3 Perancangan Halaman Data Tweet	35
Gambar 3. 4 Perancangan Halaman Data Hasil <i>Processing</i>	36
Gambar 3. 5 Perancangan Halaman Data Test	36
Gambar 3. 6 Perancangan Halaman Data Train	37
Gambar 3. 7 Perancangan Halaman Data Train (Lanjutan)	38
Gambar 3. 8 Perancangan Halaman Data Train (Lanjutan)	39
Gambar 3. 9 Perancangan Halaman <i>Naïve Bayes Chi Square</i>	40
Gambar 3. 10 Perancangan Halaman <i>Naïve Bayes Tanpa Chi Square</i>	41
Gambar 4. 1 Prosedur Penelitian	43
Gambar 4. 2 Diagram Alir <i>Naïve Bayes</i>	45
Gambar 4. 3 Diagram Alir <i>Naïve Bayes</i>	53
Gambar 4. 4 Prapengolahan.....	67
Gambar 4. 5 Hasil <i>Chi Square</i> Positif	69
Gambar 4. 6 Seleksi <i>Chi Square</i> Negatif.....	70
Gambar 4. 7 Hasil Akurasi Metode <i>Naïve Bayes Tanpa Chi Square</i>	74
Gambar 4. 8 Hasil Akurasi Metode <i>Naïve Bayes Dengan Chi Square</i>	77
Gambar 4. 9 Distribusi persentase cuitan	81

DAFTAR TABEL

	Halaman
Tabel 2. 1 <i>Confusion Matrix</i>	28
Tabel 4. 1 Dataset Opini Publik Mengenai Vaksinasi Covid19 Pada Twitter	46
Tabel 4. 2 Hasil Perhitungan <i>Term</i> Untuk Setiap Dokumen	47
Tabel 4. 3 Hasil Pembagian Data Latih.....	48
Tabel 4. 4 Hasil Pembagian Data Uji.....	48
Tabel 4. 5 Hasil Perhitungan Multinomial Model.....	50
Tabel 4. 6 Hasil Perhitungan Posterior.....	52
Tabel 4. 7 Dataset Opini Publik Mengenai Vaksinasi Covid19 Pada Twitter	54
Tabel 4. 8 Hasil Perhitungan <i>Term</i> untuk Setiap Dokumen.....	55
Tabel 4. 9 Hasil Pembagian Data Latih.....	55
Tabel 4. 10 Hasil Pembagian Data Uji.....	56
Tabel 4. 11 Tabel Kontingensi Term dan Class	57
Tabel 4. 12 Hasil <i>Chi-Square</i> Untuk Setiap Kelas	57
Tabel 4. 13 Hasil <i>Chi-Square</i> Tunggal Untuk Setiap Term t.....	58
Tabel 4. 14 Hasil Seleksi <i>Chi-Square</i>	58
Tabel 4. 15 Hasil Perhitungan Multinomial Model.....	60
Tabel 4. 16 Hasil Perhitungan <i>Posterior</i>	62
Tabel 4. 17 Perbandingan Kelas Prediksi dengan Kelas Aktual	63
Tabel 4. 18 <i>Confusion Matriks</i>	63
Tabel 4. 19 <i>Confusion Matriks</i>	63
Tabel 4. 20 <i>Confusion Matriks Naïve Bayes</i> Tanpa <i>Chi-Square</i>	74
Tabel 4. 21 <i>Precision</i>	75
Tabel 4. 22 <i>Recall</i>	75
Tabel 4. 23 <i>F-Measure</i>	76
Tabel 4. 24 Hasil Perhitungan <i>Confusion Matrix</i>	77
Tabel 4. 25 <i>Confusion Matrix Naïve Bayes</i> dengan <i>Chi-Square</i>	78
Tabel 4. 26 <i>Precision</i>	78

Tabel 4. 27 <i>Recall</i>	79
Tabel 4. 28 <i>F-Measure</i>	80
Tabel 4. 29 Hasil Perhitungan <i>Confusion Matrix</i>	80



DAFTAR ARTI LAMBANG DAN SINGKATAN

DAFTAR ARTI LAMBANG

Lambang	Arti Lambang
T	Kata
C	Kelas/kategori
R	Data yang belum diketahui kelasnya
S	Hipotesis pada data R yang merupakan class khusus
$P(R/S)$	Nilai probabilitas pada hipotesis R yang berdasarkan kondisi S
$P(R)$	Nilai probabilitas pada hipotesis R
$P(S/R)$	Nilai probabilitas S yang berdasarkan dengan kondisi hipotesis
$R P(S)$	Nilai probabilitas S
$tf(i)$	Nilai <i>Term Frequency</i>
$Freq(t_1)$	Frekuensi kemunculan sebuah kata
$\sum freq(t)$	Jumlah keseluruhan kata dalam dokumen
$idf(i)$	Nilai <i>Inverse Document Frequency</i> sebuah kata
$ D $	Jumlah seluruh dokumen.
$ (d: t_i \in d) $	Jumlah dokumen yang mengandung kata (t)
$P(c)$	Prior dari kelas c
N_c	Banyaknya dokumen pada data latih yang menjadi kategori c
N	Banyaknya dokumen pada data latih
$count(w, c)$	Jumlah kemunculan kata w pada kelas c
$count(c)$	Jumlah total kemunculan kata pada kelas c
$ V $	Jumlah ciri
$\chi^2(t, c)$	<i>Chi-Square</i> untuk term t dan kelas
A	Jumlah term t yang termasuk dalam kelas c
B	Jumlah term t yang tidak termasuk dalam kelas c
C	Jumlah term yang bukan term t namun termasuk dalam kelas c

DAFTAR SINGKATAN

Singkatan	Kepanjangan Singkatan
TF-IDF	<i>Term Frequency Inverse Document Frequency</i>
TP	<i>True Positives</i>
FP	<i>False Positives</i>
FN	<i>False Negatives</i>
TN	<i>True Negatives</i>
NLTK	<i>Natural Language Tool Kit</i>
HTML	<i>Hypertext Markup Language</i>
CSV	<i>Comma-Separated Values</i>
SVM	<i>Support Vector Machine</i>
KNN	<i>K-Nearest Neighbor</i>



ABSTRAK

Kritik dan komentar yang disampaikan masyarakat Indonesia terkait kebijakan pemerintah mengenai pembelajaran tatap muka di masa pandemi Covid-19 menuai pro dan kontra. Tidak sedikit orang tua yang khawatir dengan kebijakan ini dikarenakan ketakutan akan penyebaran klaster baru Covid-19 di Indonesia yang semakin berkembang, di sisi lain pembelajaran secara daring juga dinilai kurang efektif karena banyak siswa yang sulit menerima materi yang disampaikan guru serta banyaknya siswa yang belum memiliki perangkat yang memadai. Opini yang dituliskan pada *Twitter* membutuhkan pengklasifikasian sesuai sentimen yang dimiliki agar mudah untuk mendapatkan kecenderungan opini tersebut apakah cenderung beropini netral, positif maupun negatif. Analisis sentimen dalam penelitian ini dilakukan dengan menggunakan metode *Naïve Bayes* dan seleksi ciri *Chi Square* dalam melakukan klasifikasi. Dari hasil penelitian yang telah dilakukan didapatkan bahwa sentimen masyarakat tentang pembelajaran tatap muka di tengah pandemi Covid-19 mendapatkan respon yang cenderung positif sebesar 65,78% dan sentimen negatif sebesar 34,22%. Masyarakat lebih banyak memberikan sentimen positif terkait pembelajaran tatap muka di masa pandemi, sebagian dari sentimen positif berisi doa dan harapan dalam menyambut pembelajaran tatap muka. Namun tidak sedikit juga masyarakat yang memberikan sentimen negatif, beberapa sentimen negatif berisi kekecewaan yang mana salah satu penyebabnya adalah adanya wacana terkait pembelajaran tatap muka di daerah tempat tinggalnya tetapi belum terjadi. Hasil analisis dari metode *Naïve Bayes* dengan seleksi ciri *chi Square* memiliki akurasi 93% dan tanpa seleksi ciri *Chi Square* memiliki akurasi 92%, sehingga dapat disimpulkan Metode *Naïve Bayes* dengan seleksi ciri *Chi Square* memiliki tingkat akurasi yang lebih baik dibanding tanpa menggunakan seleksi ciri *Chi Square*.

Kata Kunci: Analisis Sentimen, *Chi Square*, Covid-19, *Naïve Bayes*, Seleksi Fitur.

ABSTRACT

Criticisms and comments submitted by the Indonesian people regarding government policies related to face-to-face learning during the Covid-19 pandemic reap the pros and cons. Not a few parents are worried about this policy because they are worried about the increasingly widespread spread of the new Covid-19 cluster in Indonesia, on the other hand online learning is also considered less effective because many students have difficulty accepting the material presented by the school, teachers and many students do not have adequate tools. Opinions written on Twitter require classification according to the sentiment they have so that it is easy to get the tendency of opinions whether they tend to be neutral, positive or negative. Sentiment analysis in this study was carried out using the Naïve Bayes method and Chi Square feature selection in its classification. From the results of the research that has been carried out, it was found that public sentiment towards face-to-face learning in the midst of the Covid-19 pandemic received a positive response of 65.78% and negative sentiment of 34.22%. The positive sentiment of the community is more related to face-to-face learning during the pandemic, some of these positive sentiments contain prayers and hopes in welcoming face-to-face learning. However, not a few also gave negative sentiments, some of these negative sentiments contained disappointment, one of which was the discourse related to face-to-face learning in the area where they lived but it had not yet happened. The results of the analysis of the Nave Bayes method with chi Square feature selection has an accuracy of 93% and without Chi Square feature selection has an accuracy of 92%, so it can be concluded that the Naïve Bayes method with Chi Square feature selection has a better accuracy rate than without using Chi Square characteristic selection.

Keywords: *Sentiment Analysis, Chi Square, Covid-19, Naïve Bayes, Feature Selection.*