

**IMPLEMENTASI ALGORITMA *NAÏVE BAYES CLASSIFIER*
(*BERNOULLI NB, MULTINOMIAL NB, DAN GAUSSIAN NB*) UNTUK
MENGKLASIFIKASI PENGADUAN MASYARAKAT**

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



Keszya Wabang
30000320410016

SEKOLAH PASCASARJANA
UNIVERSITAS DIPONEGORO
SEMARANG

2023

HALAMAN PENGESAHAN

TESIS

IMPLEMENTASI ALGORITMA *NAÏVE BAYES CLASSIFIER*
(*BERNOULLI NB, MULTINOMIAL NB, DAN GAUSSIAN NB*) UNTUK
MENGKLASIFIKASI PENGADUAN MASYARAKAT

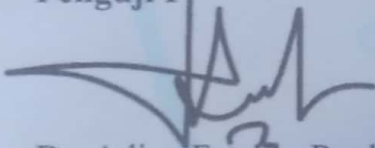
Oleh:

Keszya Wabang
30000320410016

Telah diujikan dan dinyatakan lulus ujian tesis pada tanggal 26 April 2023 oleh Tim Penguji Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

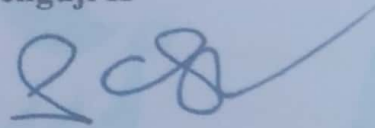
Semarang, 26 April 2023
Mengetahui,

Penguji I



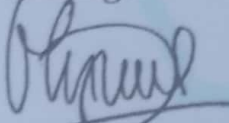
Dr. Adian Fatchur Rochim, S.T., M.T.
NIP. 197302261998021001

Penguji II



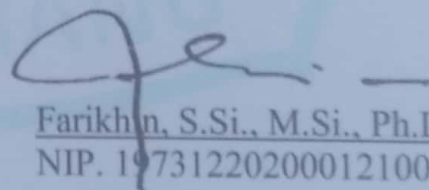
Dr. Aris Sugiharto, S.Si., M.Kom.
NIP. 197108111997021004

Pembimbing I



Dr. Oky Dwi Nurhayati, S.T., M.T.
NIP. 197910022009122001


Pembimbing II



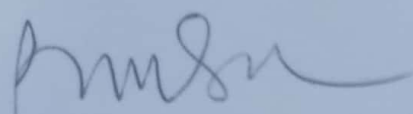
Farikh, S.Si., M.Si., Ph.D.
NIP. 197312202000121001

Mengetahui:
Dekan Sekolah Pascasarjana
Universitas Diponegoro

Ketua Program Studi
Magister Sistem Informasi



Dr. R. B. Surtarto, S.H., M.Hum.
NIP. 196701011991031005



Drs. Bayu Surarso, M.Sc., Ph.D.
NIP. 196311051988031001

PERNYATAAN PERSETUJUAN
PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS

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

Nama : Keszya Wabang
NIM : 30000320410016
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:

Implementasi Algoritma *Naïve Bayes Classifier* (*Bernoulli NB*, *Multinomial NB*, dan *Gaussian NB*) untuk Mengklasifikasi Pengaduan Masyarakat

beserta perangkat yang ada. 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 tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik hak cipta.

Dibuat di : Semarang
Pada tanggal : 10 April 2023
Yang menyatakan,



Keszya Wabang
NIM. 30000320410016

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, 10 April 2023



Keszya Wabang

KATA PENGANTAR

Segala puji syukur hanya bagi Tuhan Yang Maha Esa, karena atas kasih karunia dan rahmat-Nya sehingga tesis ini dengan judul Implementasi Algoritma *Naïve Bayes Classifier (Bernoulli NB, Multinomial NB, dan Gaussian NB)* untuk Mengklasifikasi Pengaduan Masyarakat dapat terselesaikan dengan baik. Penyusunan tesis ini telah selesai karena bantuan dari berbagai pihak. Untuk itu, Penulis mengucapkan terima kasih kepada:

1. Dr. R. B. Sularto, S.H., M.Hum., selaku Dekan Pascasarjana Universitas Diponegoro Semarang.
2. Drs. Bayu Surarso, M.Sc., Ph.D., selaku Ketua Program Studi Magister Sistem Informasi Universitas Diponegoro Semarang.
3. Dr. Oky Dwi Nurhayati, S.T., M.T., selaku Dosen Pembimbing I tesis yang telah banyak meluangkan waktu, pikiran, tenaga, berbagi ilmu, saran, nasehat, dan semangat kepada Penulis selama bimbingan tesis.
4. Farikhin, S.Si., M.Si., Ph.D., selaku Dosen Pembimbing II tesis yang telah meluangkan waktu, pikiran, tenaga, berbagi ilmu, saran, nasehat, dan semangat kepada Penulis selama bimbingan tesis.
5. Orang-orang terkasih, Bapak Christofel Wabang, Mama Martha E. Wabang-Holbala, Kakak Wanry, Kakak Windu sekeluarga, dan Suamiku Alberth Roy Kota yang telah banyak mendoakan Penulis, memberi semangat, motivasi, dan dukungan hingga tesis ini dapat terselesaikan dengan baik.
6. Segenap pihak yang terlibat dalam proses penyelesaian tesis ini.

Tesis ini masih jauh dari kesempurnaan sehingga Penulis terbuka terhadap kritik dan saran yang membangun, Akhir kata, semoga tesis ini dapat bermanfaat bagi para pembaca sebagaimana yang diharapkan.

Semarang, 10 April 2023

Penulis

DAFTAR ISI

Halaman Judul.....	i
Halaman Pengesahan	ii
Halaman Pernyataan Persetujuan Publikasi	iii
Halaman Pernyataan.....	iv
Kata Pengantar	v
Daftar Isi.....	vi
Daftar Gambar.....	viii
Daftar Tabel	ix
Daftar Lampiran	x
Daftar Arti Lambang dan Singkatan	xi
Abstrak	xii
Abstract	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Tujuan Penelitian	3
1.3. Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	4
2.1. Tinjauan Pustaka.....	4
2.2. Dasar Teori	7
2.2.1. Pengaduan/Laporan Masyarakat	7
2.2.2. <i>Preprocessing Data</i>	8
2.2.3. Klasifikasi	9
2.2.4. <i>Naïve Bayes Classifier</i>	10
2.2.5. <i>Voting</i>	14
2.2.6. <i>Python dan Jupyter Notebook</i>	15
2.2.7. <i>Confusion Matrix</i>	16
2.2.8. <i>Extreme Programming (XP)</i>	17
2.2.9. Diagram Alir Data (<i>Data Flow Diagram/ DFD</i>).....	19
2.2.10. <i>Django Framework</i>	21
BAB III METODOLOGI PENELITIAN.....	23
3.1. Bahan dan Alat Penelitian.....	23
3.1.1. Bahan Penelitian	23
3.1.2. Alat Penelitian.....	23
3.2. Prosedur Penelitian	23
3.3. Kerangka Sistem Informasi	25
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	27
4.1. Hasil Penelitian	27
4.1.1. Hasil Seleksi Data dan <i>Preprocessing Data</i>	27
4.1.2. Hasil <i>Training Data</i>	29
4.1.3. Hasil Pengujian Kinerja Algoritma.....	34
4.1.4. Hasil Perancangan Basis Data.....	45

4.1.5. Hasil Perancangan Diagram Alir Data.....	47
4.1.6. Hasil Perancangan Antarmuka Sistem.....	52
4.1.7. Hasil Implementasi Sistem.....	54
4.1.8. Hasil Perhitungan Klasifikasi dengan Algoritma.....	59
4.2. Pembahasan	63
BAB V KESIMPULAN DAN SARAN.....	68
5.1. Kesimpulan	68
5.2. Saran	68
DAFTAR PUSTAKA	69
LAMPIRAN.....	74

DAFTAR GAMBAR

Gambar 2.1. <i>Hard voting</i>	14
Gambar 2.2. <i>Soft voting</i>	15
Gambar 2.3. <i>Confusion Matrix</i>	16
Gambar 2.4. Tahapan proses <i>Extreme Programming (XP)</i>	18
Gambar 3.1. Kerangka Sistem Informasi.....	25
Gambar 4.1. Alur Proses Data	27
Gambar 4.2. Cuplikan data hasil seleksi data	28
Gambar 4.3. Hasil pengecekan <i>missing value</i>	29
Gambar 4.4. Hasil transformasi data pada 5 atribut.....	29
Gambar 4.5. Hasil transformasi data pada label/ kelas	29
Gambar 4.6. Hasil <i>training</i> data dengan <i>Bernoulli NB</i>	30
Gambar 4.7. Hasil <i>training</i> data dengan <i>Multinomial NB</i>	31
Gambar 4.8. Hasil <i>training</i> data dengan <i>Gaussian NB</i>	32
Gambar 4.9. Hasil <i>training</i> data dengan <i>Hard Voting</i>	32
Gambar 4.10. Hasil <i>training</i> data dengan <i>Soft Voting</i>	33
Gambar 4.11. Perbandingan Skor Hasil <i>Training Data</i>	34
Gambar 4.12. <i>Confusion matrix</i> untuk algoritma <i>Bernoulli NB</i>	35
Gambar 4.13. <i>Classification report</i> untuk algoritma <i>Bernoulli NB</i>	35
Gambar 4.14. <i>Confusion matrix</i> untuk algoritma <i>Multinomial NB</i>	37
Gambar 4.15. <i>Classification report</i> untuk algoritma <i>Multinomial NB</i>	37
Gambar 4.16. <i>Confusion matrix</i> untuk algoritma <i>Gaussian NB</i>	38
Gambar 4.17. <i>Classification report</i> untuk algoritma <i>Gaussian NB</i>	39
Gambar 4.18. <i>Confusion matrix</i> untuk <i>hard voting</i>	40
Gambar 4.19. <i>Classification report</i> untuk <i>hard voting</i>	41
Gambar 4.20. <i>Confusion matrix</i> untuk <i>soft voting</i>	42
Gambar 4.21. <i>Classification report</i> untuk <i>soft voting</i>	43
Gambar 4.22. Perbandingan skor kinerja algoritma.....	44
Gambar 4.23. Diagram Konteks Sistem.....	48
Gambar 4.24. DFD Level 1.....	49
Gambar 4.25. DFD Level 2 Proses 2	50
Gambar 4.26. DFD Level 2 Proses 3	51
Gambar 4.27. DFD Level 2 Proses 4	52
Gambar 4.28. Rancangan antarmuka halaman <i>home</i>	52
Gambar 4.29. Rancangan antarmuka halaman <i>input</i> data dan klasifikasi.....	53
Gambar 4.30. Rancangan antarmuka halaman data pengaduan.....	54
Gambar 4.31. Antarmuka halaman <i>login</i>	55
Gambar 4.32. Antarmuka halaman <i>home</i>	55
Gambar 4.33. Antarmuka halaman <i>input</i> data dan klasifikasi	56
Gambar 4.34. Antarmuka halaman data pengaduan	57
Gambar 4.35. Antarmuka halaman <i>Django Administration</i>	57
Gambar 4.36. Antarmuka halaman alur penelitian	58
Gambar 4.37. Antarmuka halaman setelah <i>logout</i>	59

DAFTAR TABEL

Tabel 2.1. Daftar simbol pada diagram alir data	20
Tabel 4.1. Skor kinerja tiga tipe algoritma <i>Naïve Bayes Classifier</i>	44
Tabel 4.2. Rancangan basis data pada tabel klasifikasi_data.....	45
Tabel 4.3. Rancangan basis data pada tabel auth_user	47
Tabel 4.4. Data pengaduan yang <i>diinput</i> pada sistem	59

DAFTAR LAMPIRAN

Lampiran 1. Surat keterangan pengambilan data	74
Lampiran 2. <i>Source Code</i> pada <i>Jupyter Notebook</i>	75
Lampiran 3. Cuplikan data pengaduan masyarakat	78
Lampiran 4. Daftar data uji tidak tepat diklasifikasi.....	79

DAFTAR ARTI LAMBANG DAN SINGKATAN

DAFTAR ARTI LAMBANG

Lambang	Arti Lambang
X	Data <i>tuple</i>
A	Data atribut
n	Indeks data
P	Peluang atau probabilitas
H	Data hipotesis
C_i	Data pada indeks i
C_j	Data pada indeks j
C_m	Data pada indeks m
D	Jumlah data
x_k	Data pada indeks k
t	Waktu

DAFTAR SINGKATAN

Singkatan	Kepanjangan Singkatan
NB	<i>Naïve Bayes</i>
NBC	<i>Naïve Bayes Classifier</i>
BNB	<i>Bernoulli Naïve Bayes</i>
MNB	<i>Multinomial Naïve Bayes</i>
GNB	<i>Gaussian Naïve Bayes</i>
XP	<i>Extreme Programming</i>
DFD	<i>Data Flow Diagram</i>
BSD	<i>Berkeley Software Distribution</i>
SQL	<i>Structured Query Language</i>
HTML	<i>HyperText Markup Language</i>
CSS	<i>Cascading Style Sheet</i>

**IMPLEMENTASI ALGORITMA NAÏVE BAYES CLASSIFIER
(BERNOULLI NB, MULTINOMIAL NB, DAN GAUSSIAN NB) UNTUK
MENGKLASIFIKASI PENGADUAN MASYARAKAT**

ABSTRAK

Adanya pelayanan publik yang kurang memuaskan mendorong masyarakat untuk menyampaikan pengaduan/ laporan kepada penyelenggara pelayanan publik agar melakukan perbaikan layanannya. Masalah layanan publik yang dialami setiap masyarakat berbeda-beda sehingga setiap pengaduan/ laporan yang disampaikan bervariasi. Langkah awal dari proses penyelesaian pengaduan masyarakat adalah melakukan klasifikasi terhadap setiap pengaduan masyarakat yang masuk agar penanganan pengaduan menjadi lebih efektif dan efisien. Lembaga Ombudsman RI sebagai Lembaga Negara Pengawas Pelayanan Publik di Indonesia yang setiap tahunnya menerima minimal 10.000 pengaduan dengan rata-rata 300-500 laporan per provinsi per tahun melakukan klasifikasi terhadap data-data pengaduan/ laporan masyarakat untuk membaginya ke dalam tiga kelas, yaitu laporan sederhana, laporan sedang, dan laporan berat. Proses klasifikasi yang dilakukan menggunakan penilaian bobot terhadap setiap pengaduan/ laporan dengan menggunakan 5 (lima) atribut. Klasifikasi terhadap setiap pengaduan/ laporan yang masuk menjadi suatu pekerjaan besar apabila dikerjakan satu per satu secara manual apalagi jika dalam satu waktu jumlah pengaduan yang masuk sangat banyak. Hal ini berdampak pada kurang efisiennya waktu kinerja petugas pengelola pengaduan. Sebagai alternatif solusi, dalam penelitian ini diterapkan metode pembelajaran mesin (*machine learning*) dengan tiga tipe algoritma *Naïve Bayes Classifier*, yaitu *Bernoulli Naïve Bayes*, *Multinomial Naïve Bayes*, dan *Gaussian Naïve Bayes* untuk memudahkan proses klasifikasi pengaduan/ laporan masyarakat secara otomatis agar lebih efektif dan efisien. Hasil yang diperoleh menunjukkan, bahwa klasifikasi pengaduan/ laporan masyarakat dengan mengkombinasikan ketiga tipe algoritma *Naïve Bayes Classifier* menggunakan *soft voting* memberikan nilai akurasi yang tinggi sebesar 93,41%. Selain itu, rata-rata nilai presisi, *recall*, dan *f1-score* secara berturut-turut adalah 92,12%, 92,52%, dan 92,21%.

Kata kunci: Klasifikasi, *Naïve Bayes Classifier*, Pengaduan/ Laporan Masyarakat, *Voting*

**IMPLEMENTATION OF NAÏVE BAYES CLASSIFIER ALGORITHM
(BERNOULLI NB, MULTINOMIAL NB, AND GAUSSIAN NB) TO
CLASSIFY COMMUNITY COMPLAINTS**

ABSTRACT

Unsatisfactory public services encourage the public to submit complaints/ reports to public service providers to improve their services. However, each complaint/ report submitted varies. Therefore, the first step of the community complaint resolution process is to classify every incoming community complaint. The Ombudsman of The Republic of Indonesia annually receives a minimum of 10,000 complaints with an average of 300-500 reports per province per year, classifies complaints/ community reports to divide them into three classes, namely simple reports, medium reports, and heavy reports. The classification process is carried out using a weight assessment of each complaint/ report using 5 (five) attributes. It becomes a big job if done manually. This impacts the inefficiency of the performance time of complaint management officers. As an alternative solution, in this study, machine learning methods with the types of Naïve Bayes Classifier algorithm which are Bernoulli Naïve Bayes, Multinomial Naïve Bayes, and Gaussian Naïve Bayes were applied to facilitate the process of automatically classifying complaints/ community reports to be more effective and efficient. The results showed that by combining three types of Naïve Bayes Classifier algorithm using soft voting technique gives a high accuracy value of 93,41%. In addition, the average precision, recall, and f1-score values, respectively, are 92,12%, 92,52%, and 92,21%.

Keywords: Classification, Complaints/ Community Reports, Naïve Bayes Classifier, Voting