



**IMPLEMENTASI ALGORITMA *NAÏVE BAYES* UNTUK
ANALISIS KLASIFIKASI SURVEI KESEHATAN MENTAL
(STUDI KASUS: *OPEN SOURCING MENTAL ILLNESS*)**

SKRIPSI

Reza Alfaresy Chaerudin

1810512082

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI S1 SISTEM INFORMASI

2022



**IMPLEMENTASI ALGORITMA *NAÏVE BAYES* UNTUK
ANALISIS KLASIFIKASI SURVEI KESEHATAN MENTAL
(STUDI KASUS: *OPEN SOURCING MENTAL ILLNESS*)**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
Sarjana Komputer**

Reza Alfaresy Chaerudin

1810512082

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI S1 SISTEM INFORMASI

2022

LEMBAR ORISINALITAS

LEMBAR ORISINALITAS

Skripsi ini adalah hasil karya sendiri dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar:

Nama : Reza Alfarezy Chaerudin
NIM : 1810512082
Program Studi : S1 Sistem Informasi
Judul : Implementasi Algoritma Naive Bayes Untuk Prediksi
Analisis Survei Kesehatan Mental (Studi Kasus: Open
Sourcing Mental Illness)

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 22 Juni 2022

Yang membuat pernyataan,


Reza Alfarezy

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Reza Alfarezy Chaerudin
NIM : 1810512082
Program Studi : S1 Sistem Informasi
Fakultas : Ilmu Komputer

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional "Veteran" Jakarta Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**Implementasi Algoritma *Naive Bayes* Untuk Analisis Klasifikasi Survei Kesehatan Mental
(Studi Kasus: *Open Sourcing Mental Illness*)**

Beserta Perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional "Veteran" Jakarta berhak menyimpan, mengalih data/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Jakarta, 22 Juni 2022

Yang membuat pernyataan,



Reza Alfarezy

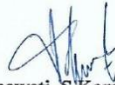
LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut:

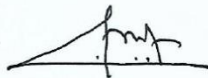
Nama : Reza Alfaresy Chaerudin
NIM : 1801512082
Program Studi : Sistem Informasi
Judul Tugas Akhir : IMPLEMENTASI ALGORITMA NAÏVE BAYES
UNTUK ANALISIS KLASIFIKASI SURVEI
KESEHATAN MENTAL (STUDI KASUS: OPEN
SOURCING MENTAL ILLNESS)

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer. Universitas Pembangunan Nasional Veteran Jakarta.



Iin Ernawati, S.Kom., M.Si.

Penguji I



Dr. Ermatita M.Kom.

Pembimbing I



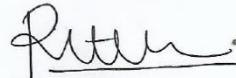
Dr. Ermatita M.Kom.

Dekan



Helena Nurramdhani Irmanda, S.Pd.,
M.Kom.

Penguji II



Ruth Mariana Bunga Wadu, S.Kom.,
MMSI

Pembimbing II



Helena Nurramdhani Irmanda, S.Pd.,
M.Kom.

Ketua Program Studi



Ditetapkan di : Jakarta

Tanggal Ujian : 8 Juli 2022

**IMPLEMENTASI ALGORITMA *NAÏVE BAYES* UNTUK ANALISIS
KLASIFIKASI SURVEI KESEHATAN MENTAL INDUSTRI TEKNOLOGI
(STUDI KASUS: *OPEN SOURCING MENTAL ILLNESS*)**

Reza Alfaresy Chaerudin

ABSTRAK

Kesehatan mental telah menjadi sorotan penting dalam kehidupan masyarakat sekarang, dan tidak luput dari berbagai industri dalam dunia kerja, termasuk industri teknologi. Kesadaran akan kepentingan kesehatan mental pekerja masih sering dianggap rendah, dan hal ini juga tidak luput dalam industri teknologi, oleh karenanya *Open Source Mental Illness* (OSMI), sebagai lembaga yang bergerak di bidang kesehatan mental, mengadakan survei untuk mengetahui kesadaran mengenai kesehatan mental pada pekerja dalam industri teknologi. Hasil dari survei ini telah dirilis sebagai dataset, di mana dataset ini kemudian dapat dianalisis lebih lanjut menggunakan *data mining* dengan metode klasifikasi sebagai analisis kesadaran kesehatan mental berdasarkan data pada survei. Algoritma klasifikasi yang digunakan adalah *Naïve Bayes*, yang mana hasil klasifikasi ini dapat digunakan lebih dalam untuk analisis lanjut mengenai kesadaran pengaruh kesehatan mental pada pekerja industri teknologi, dalam bentuk model prediksi. *Dataset* yang digunakan awalnya terdiri dari 1259 *record* data, dimana setelah dilakukan praproses didapatkan 1254 *record* data. Penelitian ini dilakukan ujicoba dengan pembagian data uji sebesar 30% dan data latih sebesar 70%, dimana didapatkan hasil akurasi sebesar 72%. Analisis *data mining* ini kemudian dilaksanakan dalam bahasa pemrograman *Python*, untuk mendapatkan suatu model prediksi sederhana yang kemudian digunakan untuk sistem prediksi sederhana berbasis *website*.

Kata Kunci: *Data Mining*, klasifikasi, *Naïve Bayes*, kesehatan mental, kesadaran kesehatan mental, industri teknologi.

**NAÏVE BAYES ALGORITHM IMPLEMENTATION FOR MENTAL HEALTH
IN TECHNOLOGY INDUSTRY SURVEY CLASSIFICATION ANALYSIS
(STUDI KASUS: OPEN SOURCING MENTAL ILLNESS)**

Reza Alfaresy Chaerudin

ABSTRACT

Mental health has become a highlight in our society's lives now, and it appears in every side of workplaces, including the technology industry. The consciousness regarding the importance of mental health among workers are still considered low, and this doesn't preclude the tech industry themselves, which is why Open Source Mental Illness (OSMI), as an organization that is based on mental health works, commences survey to understand regarding the awareness of mental health issues amongst the tech industries' workers. The result of this survey has been released as a dataset, where this dataset then can be analyzed further with data mining using classification method as analysis of awareness regarding mental health based on the survey's data. The classification algorithm chosen for the analysis is Naïve Bayes, where the result of this classification then can be used further in deeper analysis regarding awareness of mental health issues in a form of prediction model. The dataset used for the analysis initially consisted of 1259 data record, where after preprocessing process, the dataset ends with 1254 record data. In this research, an experiment is done with the split of 30% of testing data and 70% of training data, where it is then obtained the accuracy result of 72%. The data mining analysis then continues in form of Python programming language, to achieve a simplistic prediction model which is then implemented in a simplistic website-based prediction system.

Keywords: *Data Mining, classification, Naïve Bayes, mental health, mental health awareness, technology industry.*

KATA PENGANTAR

Puji dan syukur penulis panjatkan pada kehadirat Allah SWT atas segala berkah dan karunia yang Allah telah berikan sehingga penulisan Tugas Akhir ini yang berjudul “Implementasi Algoritma *Naïve Bayes* Untuk Analisis Klasifikasi Survei Kesehatan Mental (Studi Kasus: *Open Sourcing Mental Illness*)” dapat diselesaikan.

Penulisan ini ditujukan untuk memenuhi kebutuhan atas salah satu syarat untuk memperoleh gelar Sarjana S1 Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta. Untuk itu, penulis ingin mencurahkan rasa terima kasih untuk:

1. Ibu Dr. Ermatita, M.Kom selaku Dekan Fakultas Ilmu Komputer dan juga dosen pembimbing pertama yang telah menuntun penulis dalam penulisan Tugas Akhir ini dengan baik dan tepat waktu.
2. Ibu Helena Nurramdhani Irmanda selaku Ketua Program Studi S1 Sistem Informasi
3. Ibu Ruth Mariana Bunga Wadu, S.Kom., MMSI. selaku dosen pembimbing kedua yang juga telah menuntun dan memberi saran penulisan pada Tugas Akhir ini.
4. Bapak Rio Wirawan, S.Kom., MMSI. selaku dosen pembimbing akademik
5. Ibu dan Bapak Dosen S1 Sistem Informasi UPNVJ atas segala ilmu dan pengajaran yang telah diberikan.
6. Teman-teman Jurusan S1 Sistem Informasi UPNVJ yang telah saling mendukung dan memberikan diskusi, terutama Arji, Rendy, Soleh, Andru, dan teman-teman lain.
7. Mama, Ayah, Kakak, dan Adik yang telah memberikan dukungan sepenuhnya untuk penulis agar dapat menyelesaikan penulisan ini dengan sekuat tenaga.

Jakarta, 20 Juni 2022

Penulis

Reza Alfarezy

DAFTAR ISI

LEMBAR ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iv
LEMBAR PERSETUJUAN	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan.....	Error! Bookmark not defined.
1.4. Manfaat.....	Error! Bookmark not defined.
1.5. Ruang Lingkup Penelitian:	3
1.6. Luaran yang Diharapkan:	4
1.7. Sistematika Penulisan:.....	4
BAB II LANDASAN TEORI.....	6
2.1. Kesehatan Mental	6
2.2. Data Mining.....	6
2.3. Tahapan Data Mining.....	7
2.4. Teknik Data Mining	11
2.5. Klasifikasi.....	12
2.6. Algoritma Naïve Bayes	14
2.7. Evaluasi Akurasi.....	16

2.8.	Sistem	19
2.9.	Python.....	19
2.10.	Blackbox Testing	20
2.11.	Penelitian Terdahulu	21
BAB III METODOLOGI PENELITIAN		25
3.1.	Kerangka Berpikir	25
3.2.	Studi Pustaka	26
3.3.	Pengumpulan Data	26
3.4.	Pemahaman Data.....	27
3.5.	Data Praproses	31
3.6.	<i>Data Mining</i> dengan Metode Naïve Bayes	31
3.7.	Evaluasi	36
3.8.	Hasil.....	37
3.9.	Rancangan dan Pembangunan Sistem.....	37
3.10.	Implementasi dan Pengujian Sistem.....	37
3.11.	Perangkat Penelitian	38
3.12.	Jadwal Penelitian	38
BAB IV HASIL DAN PEMBAHASAN		39
4.1.	Penjelasan <i>Dataset</i>	39
4.2.	Praproses Data.....	41
4.2.1	Data Cleaning.....	41
4.2.2	Data Selection	44
4.2.3	Data Transformation	46
4.3.	Implementasi Algoritma Naïve Bayes.....	49
4.3.1	Persiapan Data Uji dan Data Latih.....	50
4.3.2.	Perhitungan Probabilitas dari Jumlah Data	54
4.3.3	Perhitungan <i>Prior Probability</i>	55
4.3.4.	Pembuatan Hasil Evaluasi.....	81
4.4.	Praktik <i>Data Mining Python</i> dengan Naïve Bayes.....	87
4.4.1	Import <i>Dataset</i> Dengan <i>Library Pandas</i>	88

4.4.2	Merapihkan Data Dengan Melakukan Encoding Data.....	89
4.4.3	Menentukan Data Training Dan Data Uji (x_train, x_test, y_train, y_test)	90
4.4.4	Membuat Model Klasifikasi Algoritma <i>Naïve Bayes Classifier</i> Dengan Library Scikit-Learn.....	93
4.4.5	Membuat Confusion Matrix Dan Classification Report Untuk Menampilkan Hasil Akurasi Untuk Klasifikasi Yang Dilakukan.....	94
4.4.6	Melakukan Contoh Klasifikasi Dengan Model Prediksi Yang Telah Dibuat	96
4.5	Perancangan dan Pembangunan Sistem	97
4.5.1	Use Case Diagram	98
4.6	Implementasi dan Pengujian Sistem.....	99
4.6.1	Halaman Utama	100
4.6.2	Halaman Prediksi.....	103
4.6.3	Pengujian dengan Blackbox Testing.....	104
BAB V PENUTUP		106
5.1.	Kesimpulan.....	106
5.2.	Saran.....	106
DAFTAR PUSTAKA		108
DAFTAR RIWAYAT HIDUP		111
LAMPIRAN.....		113

DAFTAR GAMBAR

Gambar. 1 Tahapan dalam Data Mining (Han et al., 2012).....	8
Gambar. 2 Alur Proses Klasifikasi (Sharma et al., 2018).....	13
Gambar. 3 Kerangka Berpikir Penelitian.....	25
Gambar. 4 Website Resmi <i>Dataset</i> OSMI.....	27
Gambar. 5. Grafik Perbandingan Data Kelas.....	41
Gambar. 6. <i>Missing Value</i> dari <i>Dataset</i>	42
Gambar. 7. Pembersihan Data untuk <i>Missing Value</i>	43
Gambar. 8. Pembersihan Data pada Kolom <i>Age</i>	44
Gambar. 9. Variabel Setelah <i>Data Selection</i>	45
Gambar. 10. Transformasi <i>Data Age</i>	46
Gambar. 11. Transformasi Data untuk Kolom <i>Gender</i>	48
Gambar. 12. <i>Import Dataset</i>	88
Gambar. 13. <i>Encoding Dataset</i>	89
Gambar. 14. <i>Data Training</i> dan <i>Data Testing</i> dengan <i>train_test_split</i>	90
Gambar. 15. Perhitungan Data <i>Uji</i> dan Data <i>Latih</i> Sebelum dan Setelah <i>Resampling</i> SMOTE.....	91
Gambar. 16. Hasil Pembagian Data Uji.....	92
Gambar. 17. Hasil Pembagian Data Latih.....	92
Gambar. 18. Implementasi Algoritma <i>Naïve Bayes</i> pada Analisis.....	93
Gambar. 19. Hasil Implementasi <i>Naïve Bayes</i> pada Analisis.....	93
Gambar. 20. <i>Accuracy Score</i> Analisis.....	94
Gambar. 21. <i>Confusion Matrix</i> Analisis.....	95
Gambar. 22. <i>Classification Report</i> Analisis.....	95
Gambar. 23. Contoh Klasifikasi dengan Model Hasil Klasifikasi.....	96
Gambar. 24 Pembuatan <i>File Model</i>	97
Gambar. 25. <i>Use Case Diagram</i> untuk Sistem.....	98
Gambar. 26. Pembukaan Sistem dengan IDE <i>Spyder</i>	100
Gambar. 27 Tampilan Halaman Utama.....	100

Gambar. 28. Tampilan Halaman Pengisian Data	101
Gambar. 29 Tampilan untuk Melakukan <i>Submit Data</i>	102
Gambar. 30. Tampilan Halaman <i>About Us</i>	102
Gambar. 31. Tampilan Halaman <i>Prediction</i>	103

DAFTAR TABEL





Table 1 Tabel <i>Confusion Matrix</i>	17
Table 2. Tabel Penelitian Terdahulu	21
Table 3. Dataset Hasil Survei Kesehatan Mental.....	28
Table 4. Tabel <i>Conditional Probability remote_work</i> (simulasi perhitungan).....	34
Table 5. Tabel <i>Confusion Matrix</i> (simulasi perhitungan)	35
Table 6. <i>Dataset Awal</i>	39
Table 7. Transformasi Data <i>Age</i>	46
Table 8. Ringkasan <i>Dataset</i> Setelah Praproses Data	48
Table 9. Tabel Jumlah Data	50
Table 10. Contoh Data #1 Untuk SMOTE.....	51
Table 11. Contoh Data #2 Untuk SMOTE.....	51
Table 12. . Contoh Data #3 Untuk SMOTE.....	52
Table 13. Hasil SMOTE dari Contoh Data	52
Table 14. Jumlah Data Latih Setelah Resampling	53
Table 15. Ringkasan Spesifikasi Data Latih	53
Table 16. Tabel <i>Conditional Probability Age</i>	57
Table 17. Tabel <i>Conditional Probability Gender</i>	59
Table 18, Tabel <i>Conditional Probability family_history</i>	61
Table 19. Tabel <i>Conditional Probability treatment</i>	62
Table 20. Tabel <i>Conditional Probability work_interfere</i>	65
Table 21. Tabel <i>Conditional Probability mental_health_consequence</i>	68
Table 22. Tabel <i>Conditional Probability coworkers</i>	70
Table 23. Tabel <i>Conditional Probability supervisor</i>	73
Table 24. Tabel <i>Conditional Probability mental_health_interview</i>	76
Table 25. Tabel <i>Conditional Probability mental_vs_physical</i>	78
Table 26. Contoh Data Uji Perhitungan.....	79
Table 27. <i>Confusion Matrix</i> Hasil Perhitungan.....	81
Table 28. <i>Confusion Matrix</i> (Pembagian 60% Data Latih 40% Data Uji).....	82

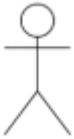


Table 29. <i>Confusion Matrix</i> (Pembagian 80% Data Latih 20% Data Uji).....	83
Table 30. Evaluasi Hasil Akurasi Percobaan	84
Table 31. <i>Confusion Matrix</i> Perhitungan Data Sebelum <i>Resampling</i>	85
Table 32. Tabel Perbandingan Evaluasi untuk Data Kelas Minoritas Sebelum dan Sesudah <i>Resampling</i>	86
Table 33. Contoh Data Uji Percobaan.....	96
Table 34. Ujicoba Blackbox Testing pada Sistem	104

DAFTAR LAMPIRAN

Lampiran 1 Dataset Awal Survei	113
Lampiran 2 Source Code app.py	186
Lampiran 3 Python Code Perbandingan Berdasarkan Pembagian Data 20% Uji 80% Latih	187
Lampiran 4 Python Code Perbandingan Berdasarkan Pembagian Data 40% Uji 60% Latih	188
Lampiran 5 Hasil Turnitin.....	189

DAFTAR SIMBOL

Simbol (<i>Flowchart</i>)	Nama Simbol	Arti Simbol
	Proses	Penanda proses yang terjadi
	Dokumen	Penanda dokumen yang digunakan dalam proses
	Arus Data	Penanda alur arah dari data yang berjalan pada proses
	<i>Terminator</i>	Penanda permulaan dan pengakhiran proses

Simbol (<i>Use Case</i>)	Nama Simbol	Arti Simbol
	<i>Actor</i>	Penanda aktor atau pengguna sistem yang berinteraksi
	Asosiasi	Penanda hubungan antara aktor dan <i>use case</i>
	<i>Use Case</i>	Penanda interaksi proses yang terjadi dalam sistem

