

**PENERAPAN ALGORITMA DATA MINING UNTUK
KLASIFIKASI TINGKAT KESEHATAN MENTAL PEKERJA**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

MUHAMMAD AMIKATUZZAIN

22.12.2560

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**PENERAPAN ALGORITMA DATA MINING UNTUK
KLASIFIKASI TINGKAT KESEHATAN MENTAL PEKERJA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

MUHAMMAD AMIKATUZZAIN

22.12.2560

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**PENERAPAN ALGORITMA DATA MINING UNTUK KLASIFIKASI
TINGKAT KESEHATAN MENTAL PEKERJA**

yang disusun dan diajukan oleh

Muhamm'ad Amikatuzzain

22.12.2560

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Februari 2026

Dosen Pembimbing,



Afrig Aminuddin, S.Kom., M.Eng., Ph.D

NIK. 190302351

HALAMAN PENGESAHAN
SKRIPSI
PENERAPAN ALGORITMA DATA MINING UNTUK
KLASIFIKASI TINGKAT KESEHATAN MENTAL PEKERJA

yang disusun dan diajukan oleh

Muhammad Amikatuzzain

22.12.2560

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Muhammad Rudyanto Arief, S.T., M.T
NIK. 190302098

Alfie Nur Rahmi, S.Kom., M.Kom
NIK. 190302240

Afrig Aminuddin, S.Kom., M.Eng., Ph.D.
NIK. 190302351

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama pekerja : Muhammad Amikatuzzain
NIM : 22.12.2560

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan Algoritma Data Mining Untuk Klasifikasi Tingkat Kesehatan Mental Pekerja

Dosen Pembimbing: Afrig Aminuddin, S.Kom., M.Eng., Ph.D

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 Februari 2026

Yang Menyatakan,



Muhammad Amikatuzzain

HALAMAN PERSEMBAHAN

Alhamdulillah *robbil'alamin*, puji dan syukur penulis panjatkan kehadiran Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir ini sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi Universitas AMIKOM Yogyakarta. Penyusunan laporan ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan syukur, penulis menyampaikan persembahan dan ucapan terima kasih kepada :

1. Allah SWT, sembah sujud serta rasa syukur penulis panjatkan kehadiran-Mu. Atas limpahan rahmat, kasih sayang, kekuatan, dan kemudahan yang Engkau berikan, penulis dapat menyelesaikan penelitian ini dengan lancar dan baik. Segala pencapaian ini tidak lepas dari pertolongan dan izin-Mu.
2. Ibu Wakhidah dan Bapak Soetikno Inggamal, cinta pertama dan pintu surga penulis. Terima kasih atas doa yang tidak pernah terputus, kasih sayang yang tulus, serta pengorbanan yang senantiasa mengiringi setiap langkah penulis dalam menempuh pendidikan. Terima kasih telah membesarkan, mendidik, dan mengantarkan penulis hingga berada di titik ini. Semoga Allah SWT senantiasa menjaga kesehatan dan kebahagiaan kalian, serta memberi kesempatan bagi penulis untuk terus membahagiakan kalian.
3. Saudara kandung penulis, Rabbika Zain, Claudia Alvi Dora terima kasih telah menjadi bagian penting dalam perjalanan hidup penulis. Dukungan, bantuan, serta kebersamaan yang diberikan menjadi kekuatan tersendiri bagi penulis dalam menyelesaikan penelitian ini.
4. Teman-teman seperjuangan, khususnya kelas SI-05 angkatan 22 yang telah menjadi kawan diskusi, memberikan bantuan teknis, serta berbagi suka dan duka dalam proses penyelesaian skripsi ini. Terima kasih atas kebersamaan dan persahabatan yang tak ternilai harganya.

KATA PENGANTAR

Segala Puji serta Syukur penulis panjatkan kehadiran Allah SWT, Tuhan Semesta Alam, yang karena rahmat dan karunia-Nya, penulis telah diberikan kemudahan dan kelancaran dalam menyelesaikan laporan kelulusan ini. Dalam menyelesaikan penelitian dan penyusunan laporan ini, penulis menyadari bahwa keberhasilan ini tidak akan terwujud tanpa adanya bimbingan, dukungan, dan bantuan tulus dari berbagai pihak. Oleh karena itu, pada kesempatan yang berharga ini, penulis ingin menyampaikan ucapan terima kasih yang mendalam kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusrini, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Bapak Anggit Dwi Hartanto, M.Kom., selaku Ketua Program Studi S1 Sistem Informasi Universitas AMIKOM Yogyakarta.
4. Bapak Afrig Aminuddin, S.Kom., M.Eng., Ph.D., selaku Dosen Pembimbing, yang telah memberikan bimbingan, arahan, serta masukan kepada penulis selama proses penyusunan tugas akhir ini.
5. Ibu Wakhidah dan Bapak Soetikno selaku orang tua penulis, yang senantiasa memberikan cinta, kasih sayang, doa, serta dukungan yang tulus sehingga penulis dapat menyelesaikan pendidikan ini dengan baik.

Penulis menyadari bahwa laporan ini masih jauh dari kata sempurna dan memiliki keterbatasan. Oleh karena itu, segala bentuk kritik dan saran yang membangun sangat diharapkan bagi perbaikan penelitian selanjutnya.

Yogyakarta, 15 Februari 2026

Penulis

DAFTAR ISI

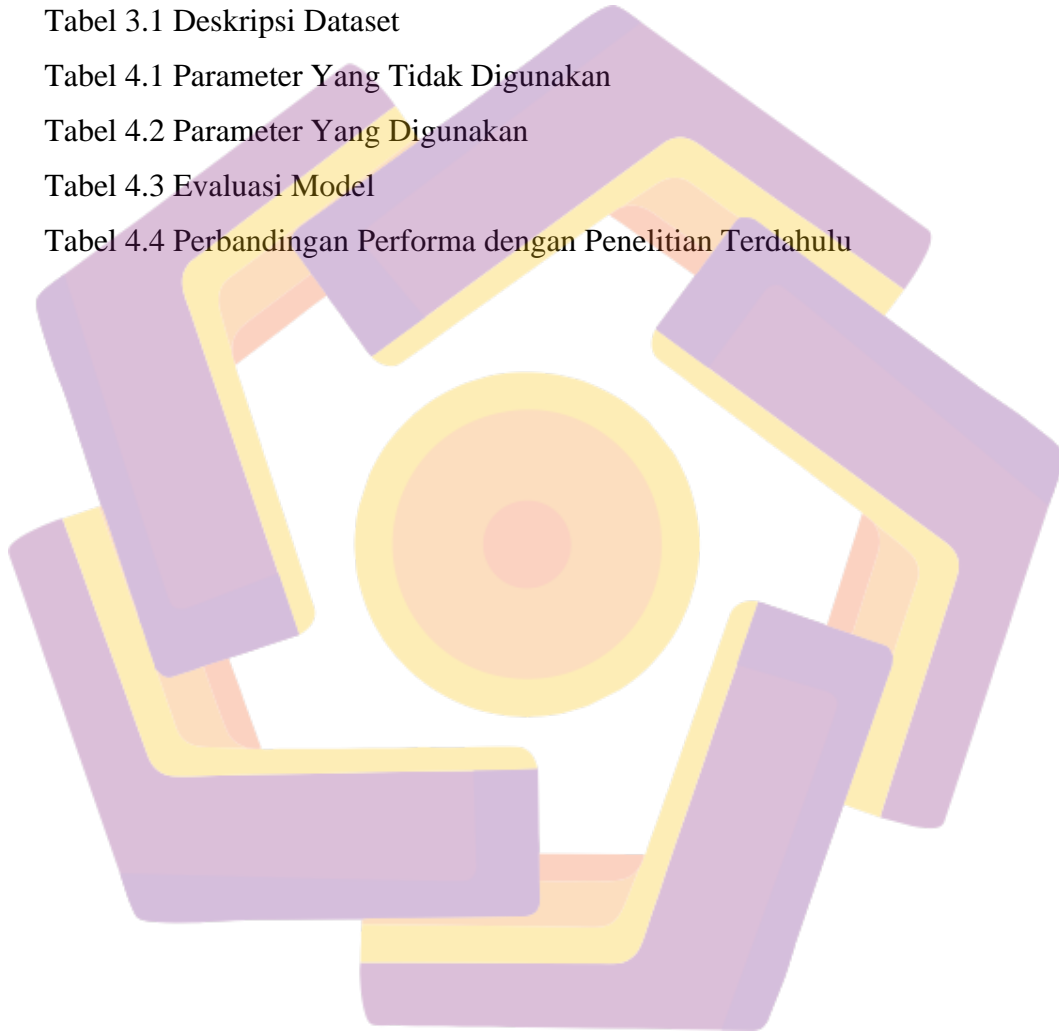
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7

2.2	Dasar Teori	13
2.2.1	Kesehatan Mental.....	13
2.2.2	Distribusi Fitur	13
2.2.3	Machine Learning	14
2.2.4	Pra-pemrosesan Data	15
2.2.5	KDD	16
2.2.6	SMOTE	18
2.2.7	Support Vector Machine	19
2.2.8	Random Forest	20
2.2.9	Naive Bayes	20
2.2.10	XGBoost	21
2.2.11	Logistic Regression.....	21
2.2.12	AdaBoost.....	22
2.2.13	Evaluasi Kinerja.....	22
BAB III METODE PENELITIAN		25
3.1	Objek Penelitian	25
3.2	Metodologi Penelitian	26
3.2.1	Data Collection	27
3.2.2	Pre-processing.....	28
3.2.3	Transformation.....	30
3.2.4	Data Mining	31
3.2.5	Evaluation	32
3.2.6	Deployment.....	33
3.3	Alat dan Bahan	34
3.3.1	Data Penelitian	34

3.3.2	Perangkat Lunak	36
3.3.3	Perangkat Keras	36
BAB IV HASIL DAN PEMBAHASAN		38
4.1	Data Collection.....	38
4.2	Pre-Processing	39
4.2.1	Data Cleaning.....	39
4.2.2	Fitur Selection.....	43
4.2.3	Integrasi.....	45
4.3	Transformation	46
4.3.1	Label Encoding	46
4.3.2	Feature Scaling	47
4.3.3	SMOTE	48
4.4	Data Mining.....	49
4.4.1	Splitting Data	49
4.4.2	Algoritma	50
4.5	Evaluation.....	56
4.5.1	Kurva ROC	59
4.5.2	Confusion Matrix	61
4.6	Pembahasan	63
4.7	Deployment	66
BAB V PENUTUP		70
5.1	Kesimpulan.....	70
5.2	Saran.....	71
REFERENSI		72
LAMPIRAN.....		75

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	10
Tabel 2.2 Confusion Matrix	23
Tabel 3.1 Deskripsi Dataset	34
Tabel 4.1 Parameter Yang Tidak Digunakan	43
Tabel 4.2 Parameter Yang Digunakan	44
Tabel 4.3 Evaluasi Model	56
Tabel 4.4 Perbandingan Performa dengan Penelitian Terdahulu	64

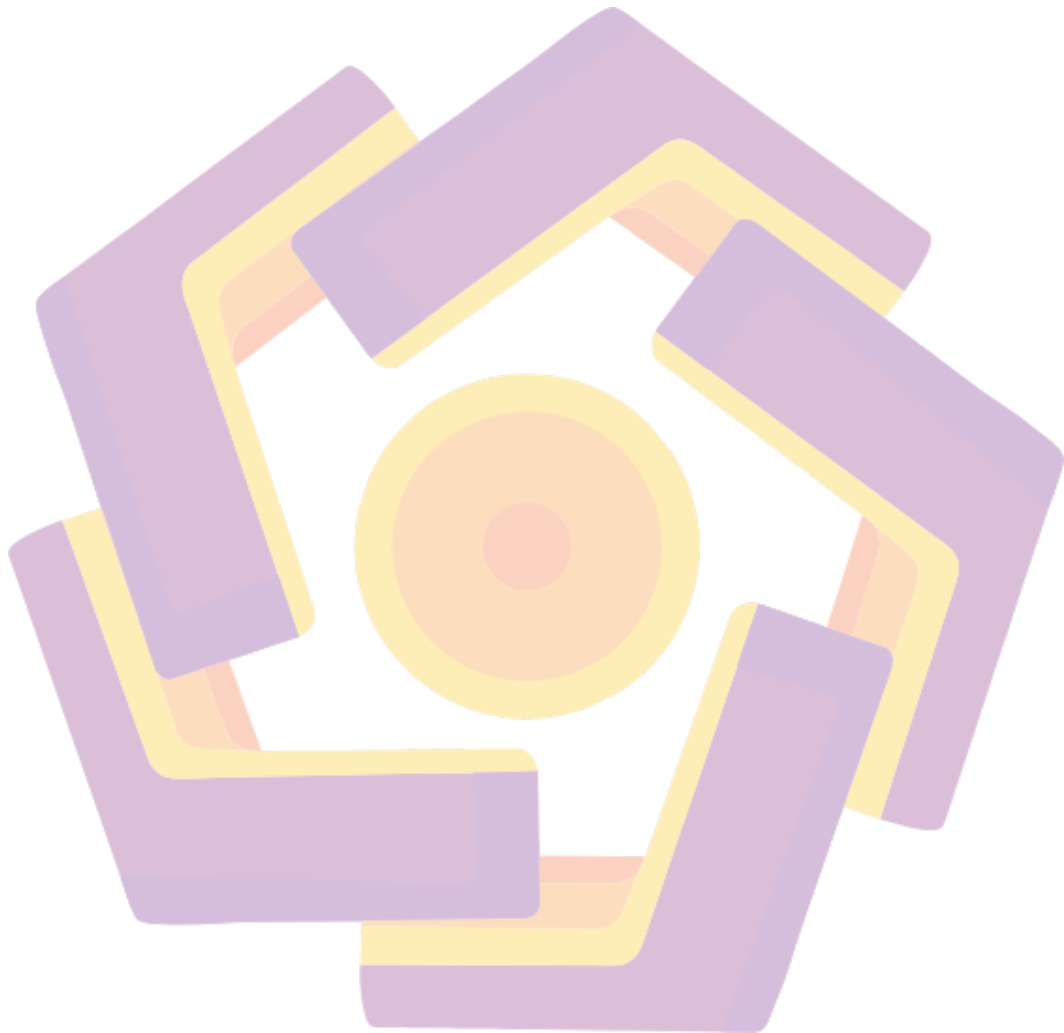


DAFTAR GAMBAR


Gambar 2.1 Tahapan Metode KDD	16
Gambar 2.2 SMOTE	18
Gambar 3.1 Alur Penelitian	27
Gambar 4.1 Sebelum Penanganan Nilai Yang Hilang	41
Gambar 4.2 Setelah Penanganan Nilai Yang Hilang	42
Gambar 4.3 SMOTE	48
Gambar 4.4 Splitting Data	49
Gambar 4.5 Fitur Yang Digunakan	50
Gambar 4.6 Hasil SVM	51
Gambar 4.7 Hasil Random Forest	52
Gambar 4.8 Hasil Naïve Bayes	52
Gambar 4.9 Hasil Logistic Regression	53
Gambar 4.10 Hasil AdaBoost	54
Gambar 4.11 Hasil XGBoost	55
Gambar 4.12 Akurasi Model	58
Gambar 4.13 Kurva ROC Seluruh Model Klasifikasi	60
Gambar 4.14 Confusion Matrix Seluruh Model Klasifikasi	62
Gambar 4.15 Hasil Prediksi Pekerja Tidak Memerlukan Perawatan	67
Gambar 4.16 Hasil Prediksi Pekerja Memerlukan Perawatan	68

DAFTAR LAMPIRAN

Lampiran 1. Dataset	75
Lampiran 2. Code	75



DAFTAR LAMBANG DAN SINGKATAN



SVM	Support Vector Machines
OSMI	Open Sourcing Mental Illness
ANN	Artificial Neural Network
SMOTE	Synthetic Minority Over-sampling Technique
AUC	Area Under Curve
KDD	Knowledge Discovery in Databases
ML	Machine Learning
RBF	Radial Basis Function
TP	True Positif
FP	False Positif
FN	False Negatif
TN	True Negatif
XGBoost	Extreme Gradient Boosting
ROC	Receiver Operating Characteristic

DAFTAR ISTILAH

MinMaxScaler	Teknik normalisasi yang mentransformasikan fitur numerik ke dalam rentang baku (biasanya 0 hingga 1)
Confusion Matrix	Instrumen evaluasi berupa tabel yang membandingkan nilai aktual dengan nilai prediksi untuk mengukur kinerja model secara mendalam.
Accuracy	Persentase total prediksi yang benar dari keseluruhan dataset.
Precision	Ukuran ketepatan model dalam memprediksi kelas positif, bertujuan meminimalkan kesalahan false positive
Recall	Kemampuan model dalam menemukan kembali semua instansi kelas positif
ROC Curve	Grafik yang memvisualisasikan hubungan antara sensitivitas dan spesifisitas pada berbagai ambang batas klasifikasi.
Machine Learning	Pendekatan komputasi yang memungkinkan sistem belajar dari data untuk melakukan prediksi
Preprocessing	Tahap awal pembersihan dan penyiapan data sebelum diolah.
Klasifikasi	Proses pengelompokan data ke dalam kelas tertentu
Feature	Variabel atau atribut yang digunakan sebagai input model
Dataset	Kumpulan data terstruktur yang digunakan untuk analisis sistem.
Deployment	Penerapan model yang sudah jadi ke dalam aplikasi nyata.
Streamlit	Kerangka kerja Python untuk membuat aplikasi web data secara cepat.

INTISARI

Penelitian ini bertujuan untuk menerapkan teknik data mining dalam mengklasifikasikan tingkat kesehatan mental pekerja guna memberikan instrumen deteksi dini yang objektif. Permasalahan utama yang diangkat adalah tingginya risiko gangguan psikososial di lingkungan perusahaan yang sering kali terlambat ditangani akibat keterbatasan diagnosa awal. Metodologi yang digunakan dalam penelitian ini mengikuti kerangka kerja *Knowledge Discovery in Databases* (KDD) yang mencakup tahap pra-pemrosesan data, seleksi fitur, hingga implementasi enam algoritma klasifikasi, yaitu Naive Bayes, XGBoost, Random Forest, SVM, Logistic Regression, dan AdaBoost. Untuk mengatasi ketidakseimbangan kelas pada dataset survei yang berjumlah 1.259 responden, diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). Hasil penelitian menunjukkan bahwa penggunaan delapan fitur kunci, seperti riwayat keluarga dan interferensi kerja, sangat efektif dalam memprediksi kebutuhan penanganan kesehatan mental. Berdasarkan pengujian metrik evaluasi, algoritma Naive Bayes dan XGBoost menunjukkan performa paling unggul dengan tingkat akurasi mencapai 0,8413. Secara khusus, XGBoost mencatatkan nilai recall sebesar 0,9055, yang mengindikasikan sensitivitas tinggi dalam mendeteksi individu yang benar-benar membutuhkan bantuan profesional. Perbandingan dengan penelitian terdahulu mengonfirmasi bahwa integrasi teknik SMOTE dan algoritma ensemble dalam penelitian ini mampu melampaui capaian akurasi model-model sebelumnya. Kesimpulannya, model klasifikasi yang dihasilkan memiliki reliabilitas yang kuat untuk diintegrasikan ke dalam sistem pendukung keputusan bagi perusahaan dalam melakukan intervensi dini terhadap kesehatan mental pekerja.

Kata kunci: *Data Mining*, Kesehatan Mental, Klasifikasi, Naive Bayes, XGBoost, SMOTE.

ABSTRACT

This study aims to apply data mining techniques to classify the mental health levels of workers to provide an objective early detection instrument. The primary issue addressed is the high risk of psychosocial disorders in corporate environments, which are often treated too late due to limitations in early diagnosis. The methodology used in this research follows the Knowledge Discovery in Databases (KDD) framework, encompassing data pre-processing, feature selection, and the implementation of six classification algorithms: Naive Bayes, XGBoost, Random Forest, SVM, Logistic Regression, and AdaBoost. To address class imbalance in the survey dataset consisting of 1,259 respondents, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. The results indicate that the use of eight key features, such as family history and work interference, is highly effective in predicting the need for mental health treatment. Based on evaluation metric testing, the Naive Bayes and XGBoost algorithms demonstrated the most superior performance, achieving an accuracy rate of 0.8413. Specifically, XGBoost recorded a recall value of 0.9055, indicating high sensitivity in detecting individuals who genuinely require professional assistance. Comparisons with previous studies confirm that the integration of SMOTE techniques and ensemble algorithms in this research outperforms the accuracy levels of prior models. In conclusion, the resulting classification model possesses strong reliability for integration into decision support systems for organizations to conduct early interventions regarding workers' mental health.

Keyword: Data Mining, Mental Health, Classification, Naive Bayes, XGBoost, SMOTE