

**LAPORAN TUGAS BESAR**  
**KECERDASANBUATAN**  
**KLASIFIKASI TINGKAT STRES MENGGUNAKAN ALGORITMA**  
***ARTIFICIAL NEURAL NETWORK (ANN)***



Disusun oleh:

Jujun Munawar – 2306148

Nazwa Mutia Salma – 2306133

Dosen Pengampu Mata Kuliah:

Leni Fitriani, S.Kom, M.Kom

**INSTITUT TEKNOLOGI GARUT**  
**JURUSAN ILMU KOMPUTER**  
**PROGRAM STUDI TEKNIK INFORMATIKA**  
**TAHUN AKADEMIK 2024/2025**

## DAFTAR ISI

DAFTAR ISI.....	2
DAFTAR GAMBAR.....	3
DAFTAR TABEL .....	4
A. PENDAHULUAN.....	5
B. BUSINESS UNDERSTANDING .....	6
C. DATA UNDERSTANDING .....	10
D. Exploratory Data Analysis(EDA) .....	12
E. Data Preparation.....	20
F. MODELING .....	25
G. EVALUATION .....	26
H. INTERPRETATION .....	29
I. KESIMPULAN .....	29
J. DAFTAR PUSTAKA .....	30
K. LAMPIRAN .....	31

## DAFTAR GAMBAR

Gambar 1 Library .....	12
Gambar 2 Load Dataset.....	12
Gambar 3 Distribusi Gender.....	13
Gambar 4 Distribusi Umur .....	13
Gambar 5 Distribusi Work Pressure .....	14
Gambar 6 Distribusi Job Satisfaction.....	14
Gambar 7 Sleep Duration.....	15
Gambar 8 Distribusi Dietary Habits .....	15
Gambar 9 Distribusi Work Hour .....	16
Gambar 10 Financial Stress.....	16
Gambar 11 Distribusi Family History.....	17
Gambar 12 Depression.....	17
Gambar 13 korelasi Antar Fitur Numerik .....	18
Gambar 14 Insight awal .....	18
Gambar 15 Info.....	19
Gambar 16 Describe .....	20
Gambar 17 Cek data .....	20
Gambar 18 Heatmap .....	21
Gambar 19 Encoding .....	21
Gambar 20 Normalisasi .....	22
Gambar 21 Smote .....	23
Gambar 22 Boxplot.....	23
Gambar 23 cek cek .....	24
Gambar 24 Model.....	25
Gambar 25 Visualisasi Model .....	26
Gambar 26 Confusion Matrix.....	27
Gambar 27 Laporan Akhir .....	28

## DAFTAR TABEL

Table 1 Ringkasan Artikel Penelitian .....	7
Table 2 Kebaharuan Project .....	7
Table 3 Kebaharuan Project Berdasarkan Jenis Data .....	8
Table 4 Kebaharuan Project Berdasarkan Tujuan Dan Pendekatan ML.....	9
Table 5 Nama dan Tipe Data Atribut.....	11
Table 6 Metrix Evaluasi Yang Digunakan .....	26
Table 7 Perbandingan Berdasarkan Aspek Penting .....	27

## A. PENDAHULUAN

Masalah stres dan gangguan kesehatan mental di kalangan pelajar dan profesional saat ini semakin mendapatkan perhatian serius dari kalangan akademisi dan praktisi kesehatan (Rika Anugrahaini, Fatchan, and Ngudi Wiyatno 2025). Dalam dunia pendidikan maupun dunia kerja, tekanan yang datang dari berbagai aspek seperti beban tugas, tuntutan sosial, kondisi lingkungan, dan keadaan psikologis sering kali menjadi pemicu meningkatnya tingkat stres seseorang (Biyantoro and Prasetyo 2024). Jika tidak ditangani dengan baik, stres dapat berkembang menjadi gangguan kesehatan mental seperti depresi, kecemasan, bahkan berdampak pada menurunnya kualitas hidup dan produktivitas individu.

Seiring dengan kemajuan teknologi di bidang kecerdasan buatan, pendekatan prediktif berbasis machine learning telah banyak digunakan untuk menganalisis kondisi psikologis individu (Rizki Agam Syahputra 2024). Salah satu metode yang terbukti efektif adalah algoritma Artificial Neural Network (ANN), yang mampu mengenali pola-pola kompleks dari data multidimensi untuk melakukan klasifikasi atau prediksi kondisi psikologis. Dalam artikel oleh Anugrahaini dkk. (2025), ANN diterapkan menggunakan perangkat lunak Orange untuk membangun model prediksi tingkat stres siswa berdasarkan 21 fitur dari berbagai faktor psikologis, fisiologis, sosial, lingkungan, dan akademis. Hasilnya menunjukkan performa model yang sangat tinggi, bahkan mencapai akurasi prediksi 100%.

Sebagai pengembangan lebih lanjut dari pendekatan tersebut, penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi kembali efektivitas model prediksi berbasis ANN dengan menggunakan Depression Professional Dataset (Muhammad Daffa Al Fahreza et al. 2024). Dataset ini memuat atribut-atribut yang relevan untuk mengidentifikasi potensi depresi dan tingkat stres dalam populasi profesional, seperti status hubungan sosial, masalah kesehatan, dan kebiasaan hidup (Rika Anugrahaini, Fatchan, and Ngudi Wiyatno 2025). Dengan mengolah data ini menggunakan pendekatan serupa, diharapkan diperoleh model prediktif yang dapat membantu dalam mendeteksi dan mengklasifikasikan risiko stres atau depresi secara lebih dini, terutama pada kalangan tenaga kerja atau profesional (Fathirachman Mahing et al. 2023).

Penelitian ini tidak hanya akan membuktikan kelayakan metode ANN dalam konteks dataset yang berbeda, tetapi juga memberikan wawasan praktis dalam pengembangan sistem pendeteksi stres yang dapat digunakan dalam dunia kerja dan lingkungan pendidikan. Dengan dukungan evaluasi metrik seperti akurasi, presisi, recall, dan MCC, performa model akan diukur secara kuantitatif untuk menilai seberapa optimal sistem dalam mengidentifikasi kondisi psikologis target.

## B. BUSINESS UNDERSTANDING

Masalah utama dalam artikel ini adalah belum tersedianya sistem yang mampu mendeteksi tingkat stres siswa secara objektif, otomatis, dan efisien. Deteksi stres selama ini cenderung dilakukan secara konvensional, melalui observasi atau wawancara langsung yang bersifat subjektif dan tidak selalu akurat. Padahal, siswa sangat rentan mengalami stres akibat berbagai tekanan akademik, sosial, dan psikologis yang dapat berdampak serius pada kesehatan mental dan prestasi belajar mereka. Tanpa adanya deteksi dini berbasis data, banyak siswa yang mengalami stres terabaikan hingga kondisinya memburuk.

Artikel ini mengusulkan pendekatan berbasis teknologi dengan memanfaatkan algoritma Artificial Neural Network (ANN) yang diterapkan menggunakan platform Orange Data Mining. Model ini dikembangkan dengan memanfaatkan StressLevelDataset.csv, yang berisi 1100 data siswa dan 21 fitur yang mencakup berbagai aspek penyebab stres. Implementasi ANN terbukti sangat efektif, dengan hasil evaluasi menunjukkan akurasi dan metrik performa lainnya mencapai 100%.

Tujuan utama dari artikel ini adalah:

1. Mengembangkan model klasifikasi tingkat stres siswa menggunakan algoritma ANN.
2. Menggunakan platform Orange untuk membangun model secara visual dan terstruktur.
3. Memanfaatkan 21 fitur dari dataset untuk mengidentifikasi tiga kategori tingkat stres (rendah, sedang, tinggi).
4. Mengevaluasi performa model menggunakan metrik akurasi, precision, recall, F1-score, AUC, dan MCC.

Sistem prediksi ini ditujukan bagi guru, konselor, dan orang tua sebagai alat bantu dalam mendeteksi kondisi stres siswa secara lebih akurat. Dengan memanfaatkan kecerdasan buatan, proses deteksi menjadi lebih cepat, efisien, dan mampu menjangkau lebih banyak siswa secara bersamaan. Selain itu, penggunaan AI mengurangi subjektivitas dalam pengambilan keputusan, serta memungkinkan intervensi psikologis dilakukan lebih awal. Proyek ini tidak hanya relevan dalam dunia pendidikan, tetapi juga memiliki potensi untuk dikembangkan di lingkungan profesional, sehingga dapat memberikan dampak yang lebih luas dalam mendukung kesehatan mental melalui teknologi.

Table 1 Ringkasan Artikel Penelitian

No	Judul Penelitian	Metode Penelitian	Dataset	Temuan Utama	Keterbatasan
1	Model Prediksi Tingkat Stres di Kalangan Siswa dengan Pendekatan Algoritma Artificial Neural Network Menggunakan Orange	Artificial Neural Network (ANN) dengan bantuan software Orange. Evaluasi menggunakan metrik: Accuracy, Precision, Recall, F1-score, AUC, dan MCC.	StressLevel Dataset.csv berisi 1100 data dan 21 fitur yang mencakup aspek psikologis, fisiologis, sosial, lingkungan, dan akademis.	Model ANN menghasilkan akurasi, presisi, recall, F1-score, dan AUC sebesar 100%, menunjukkan kemampuan klasifikasi tingkat stres yang sempurna	<ul style="list-style-type: none"> <li>- Dataset terbatas pada 1100 data dari satu konteks (siswa), sehingga generalisasi terhadap populasi yang lebih luas masih terbatas.</li> <li>- Belum dibandingkan dengan algoritma lain.</li> <li>- Tidak ada validasi data dari sumber eksternal (seperti psikolog/diagnosa klinis).</li> </ul>

Adapun beberapa perbedaan pendekatan dan kebaharuan (novelty) pada proyek kami dengan artikel sumber rujukan diatas adalah sebagai berikut :

### 1. Perbedaan pada Sumber Data (Dataset)

Table 2 Kebaharuan Project

Aspek	Artikel yang Dirujuk	Proyek yang Dilakukan
Sumber Data	Dataset yang digunakan adalah <b>StressLevelDataset.csv</b> yang diambil dari platform publik (kemungkinan besar dari Kaggle), dan <b>tidak dibuat langsung oleh peneliti</b> . Dataset ini terdiri dari <b>1100 data siswa</b> dengan berbagai variabel penyebab stres.	Menggunakan dataset eksternal berjudul “Depression Professional Dataset” (2054 entri) berisi data kuesioner sosial, psikologis, dan gaya hidup.

Fitur	Dataset memuat <b>21 fitur</b> yang mencakup lima aspek utama penyebab stres: - <i>Psikologis</i> (kecemasan, harga diri, riwayat kesehatan mental, depresi) - <i>Fisiologis</i> (sakit kepala, tekanan darah, kualitas tidur, masalah pernapasan) - <i>Sosial</i> (dukungan sosial, tekanan teman sebaya, bullying) - <i>Lingkungan</i> (kondisi tempat tinggal, keamanan, kebisingan) - <i>Akademis</i> (kinerja akademik, beban belajar, relasi dengan guru)	Data tabular: umur, tekanan kerja, durasi tidur, kebiasaan makan, pemikiran bunuh diri, jam kerja, stres finansial, dll.
Target	Kelas target adalah <b>Tingkat Stres</b> , dikategorikan ke dalam <b>tiga label klasifikasi</b> : - 0 = Tidak Stres (Low) - 1 = Stres Sedang (Medium) - 2 = Stres Tinggi (High)	Status depresi: <i>Yes/No</i>

## 2. Perbedaan pada Jenis Data

Table 3 Kebaharuan Project Berdasarkan Jenis Data

Aspek	Artikel yang Dirujuk	Proyek yang Dilakukan
Jenis Data	<b>Data tabular/numerik</b> , yaitu data survei/kuesioner yang berisi nilai-nilai atribut dalam bentuk angka atau kategori (misalnya: tingkat kecemasan, tekanan darah, kualitas tidur). Tidak menggunakan data visual seperti gambar.	Data survei/kuesioner dalam bentuk kategori dan angka (sociodemographic, lifestyle, dan stress level).
Ekstraksi Fitur	Tidak melalui proses <i>ekstraksi fitur visual</i> . Data sudah tersedia dalam bentuk fitur numerik. Setiap fitur mewakili aspek psikologis, fisiologis, sosial, lingkungan, dan akademis, sehingga tidak memerlukan proses pengolahan citra.	Label encoding dan normalisasi fitur kuesioner.



Target	<b>Tingkat stres</b> sebagai variabel target, dengan <b>tiga kelas klasifikasi</b> : - 0 = Tidak Stres (Low) - 1 = Stres Sedang (Medium) - 2 = Stres Tinggi (High)	Status depresi: <i>Yes/No</i>
--------	--	-------------------------------

### 3. Perbedaan pada Tujuan dan Pendekatan ML

Table 4 Kebaharuan Project Berdasarkan Tujuan Dan Pendekatan ML

Aspek	Artikel yang Dirujuk	Proyek yang Dilakukan
Tujuan ML	<b>Memprediksi tingkat stres siswa secara otomatis berdasarkan data numerik</b> dari faktor psikologis, fisiologis, sosial, lingkungan, dan akademis. Bukan berdasarkan ekspresi wajah, melainkan data kuesioner.	Memprediksi risiko depresi berdasarkan faktor sosial, pekerjaan, dan gaya hidup.
Algoritma yang Digunakan	<b>Artificial Neural Network (ANN)</b> digunakan sebagai algoritma utama. Model dibangun dalam lingkungan visual Orange Data Mining tanpa pengujian algoritma pembandingan lainnya seperti KNN atau Naïve Bayes.	KNN (dengan GridSearchCV untuk optimasi $k$ , serta penanganan data imbalance dengan SMOTE).
Evaluasi	Model dievaluasi menggunakan metrik standar klasifikasi: <b>Accuracy ,Precision,Recall ,F1-score ,AUC (Area Under Curve),MCC (Matthews Correlation Coefficient)</b> serta ditampilkan melalui <b>Confusion Matrix</b> .	Sama, ditambah ROC Curve & AUC, serta visualisasi korelasi fitur.

## **Kebaruan (Novelty) pada proyek yang kami buat dibandingkan dengan artikel rujukan**

### **1. Data berbasis survei/psikososial**

- Proyek yang kami buat menggunakan data dengan pendekatan berbasis gaya hidup dan riwayat pribadi, lebih representatif untuk klasifikasi depresi secara umum.

### **2. Penanganan data imbalance menggunakan SMOTE**

- Ini **tidak dilakukan** pada artikel jurnal rujukan, yang bisa menyebabkan bias model terhadap kelas mayoritas.

### **3. Optimasi Hyperparameter (GridSearchCV)**

- Proyek kami menerapkan **GridSearchCV** untuk mencari nilai k terbaik, sedangkan artikel hanya menetapkan k=5 berdasarkan eksperimen terbatas.

### **4. Visualisasi dan EDA yang mendalam**

- Proyek kami memuat heatmap, boxplot, histogram, dan ROC curve yang **tidak tersedia** dalam artikel rujukan.

### **5. Fokus pada depresi, bukan hanya stres**

- Ini merupakan cakupan yang berbeda dan lebih spesifik pada gangguan mental dibanding sekadar kondisi emosional

## **C. DATA UNDERSTANDING**

Dataset yang kami gunakan dalam proyek ini merupakan data sekunder yang diperoleh dari jurnal berjudul “Implementasi Metode Decission Tree Dalam Mengklasifikasi Depresi Menggunakan Rapidminer “ [2] yang didalamnya terdapat dataset yang kami gunakan. Dataset tersebut berasal dari Kaggle yang kemudian digunakan sebagai sumber utama dalam analisis ini. Informasi lengkap mengenai dataset dapat diakses melalui tautan berikut [Depression Professional Dataset](#).

Total data yang ada didalam dataset ini berjumlah 2054 entri, yang kemudian data dibagi menjadi dua bagian utama untuk keperluan pelatihan dan pengujian model. Sebanyak 80% digunakan sebagai data latih (training data) yang berfungsi untuk membangun dan melatih model agar mampu mengenali pola-pola yang terdapat dalam data. Sementara itu, 20% sisanya digunakan sebagai data uji (testing data) yang bertujuan untuk mengukur kinerja model terhadap data yang belum pernah dilihat sebelumnya. Sebelum training, dilakukan **penyeimbangan data (imbalanced handling)** menggunakan **SMOTE** karena data “Depression = Yes” hanya 203 dari 2054 ( $\approx 9.9\%$ ). Pembagian ini dilakukan agar evaluasi model lebih objektif dan mencerminkan kemampuannya dalam menghadapi data baru.

Adapun nama dan tipe atribut yang ada didalam dataset ini tercantum dalam tabel berikut :

Table 5 Nama dan Tipe Data Atribut

Nama Atribut	Tipe Data	Deskripsi
Gender	Kategorik	Jenis kelamin responden, seperti "Male", "Female", atau lainnya.
Age	Numerik	Usia responden dalam satuan tahun.
Work Pressure	Numerik	Tingkat tekanan kerja, kemungkinan dalam skala 1–10.
Job Satisfaction	Numerik	Tingkat kepuasan kerja responden, kemungkinan skala 1–10.
Sleep Duration	Kategorik	Lama tidur responden, seperti "< 5 jam", "5–7 jam", atau "> 7 jam".
Dietary Habits	Kategorik	Pola makan responden, seperti "Sehat", "Tidak Sehat", dll.
Have you ever had suicidal thoughts?	Kategorik	Riwayat pemikiran bunuh diri: "Yes"/"No".
Work Hours	Numerik	Jumlah jam kerja per minggu.
Financial Stress	Numerik	Tingkat stres finansial, bisa berupa skala atau skor tertentu.
Family History of Mental Illness	Kategorik	Riwayat gangguan mental dalam keluarga: "Yes"/"No".
Depression	Kategorik	Status depresi responden: "Yes", "No", atau kategori tingkat depresi.

#### Relasi antar atribut & potensi noise/outlier

- **Relasi antar atribut** (berdasarkan heatmap korelasi dan pemahaman domain):
  - *Work Pressure*, *Job Satisfaction*, dan *Financial Stress* memiliki korelasi potensial terhadap *Depression*.
  - *Sleep Duration*, *Dietary Habits*, dan *Suicidal Thoughts* juga menjadi indikator risiko depresi.
- **Potensi Noise/Outlier:**
  - Tidak ditemukan missing value maupun data duplikat.
  - Telah dilakukan visualisasi dengan boxplot dan histogram untuk memantau distribusi dan outlier.

## D. Exploratory Data Analysis(EDA)

### 1) Import Library

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt


from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.preprocessing import LabelEncoder, MinMaxScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from imblearn.over_sampling import SMOTE
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
```

Gambar 1 Library

**Penjelasan :** Memuat semua library yang dibutuhkan untuk analisis data, visualisasi, preprocessing, modeling, dan evaluasi.

### 2) Load Dataset

```
[ ] from google.colab import drive
    drive.mount('/content/drive')
    path = "/content/drive/MyDrive/Colab Notebooks/Dataset/Depression Professional Dataset .csv"
    df = pd.read_csv(path)
```

 Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

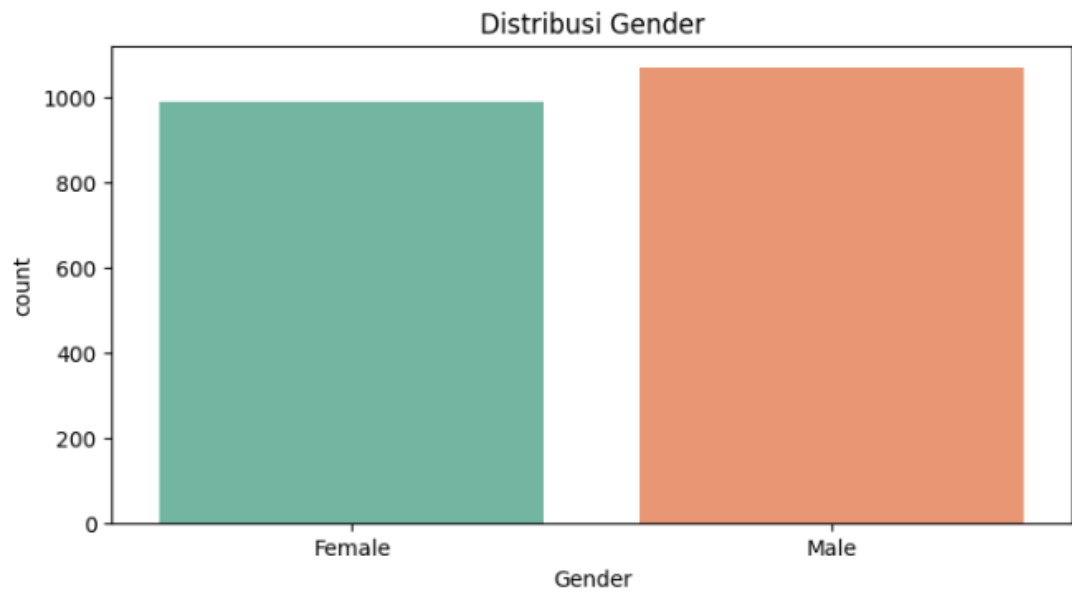
Gambar 2 Load Dataset

**Penjelasan :** Membaca file dataset yang berisi informasi terkait faktor penyebab depresi.

### 3) Visualisasi distribusi data

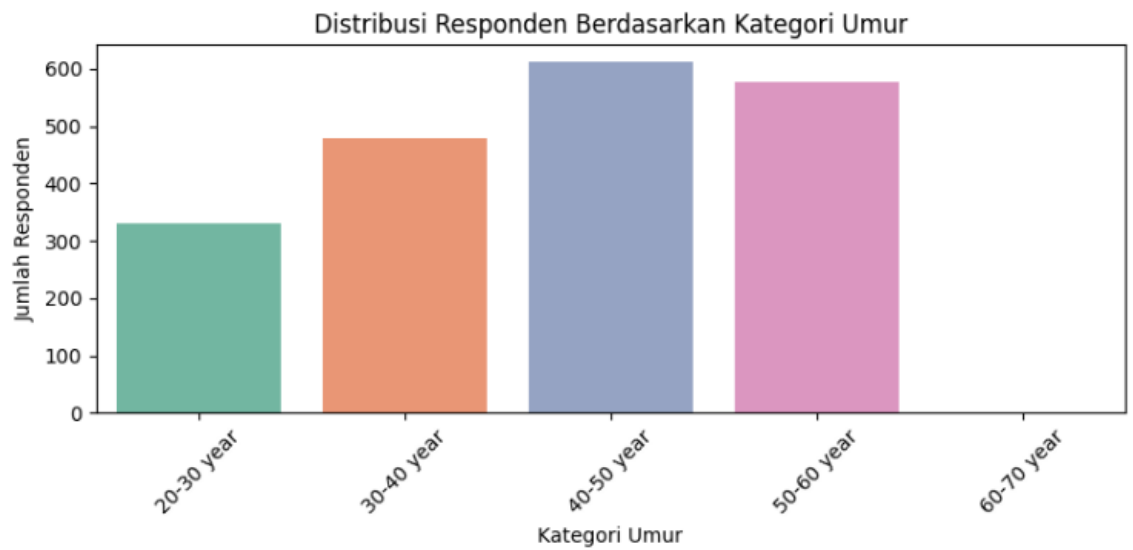
Visualisasi distribusi data pada tahap EDA bertujuan untuk memahami pola dan sebaran data, mendeteksi outlier, serta melihat keseimbangan kelas pada variabel target. Dengan visualisasi seperti histogram dan boxplot, peneliti dapat mengetahui karakteristik tiap fitur, menentukan kebutuhan preprocessing seperti normalisasi atau SMOTE, serta mengidentifikasi fitur yang relevan untuk model. Ini membantu memastikan data siap digunakan dalam proses pemodelan secara akurat dan efisien.

- **Gender**



Gambar 3 Distribusi Gender

- **Age**



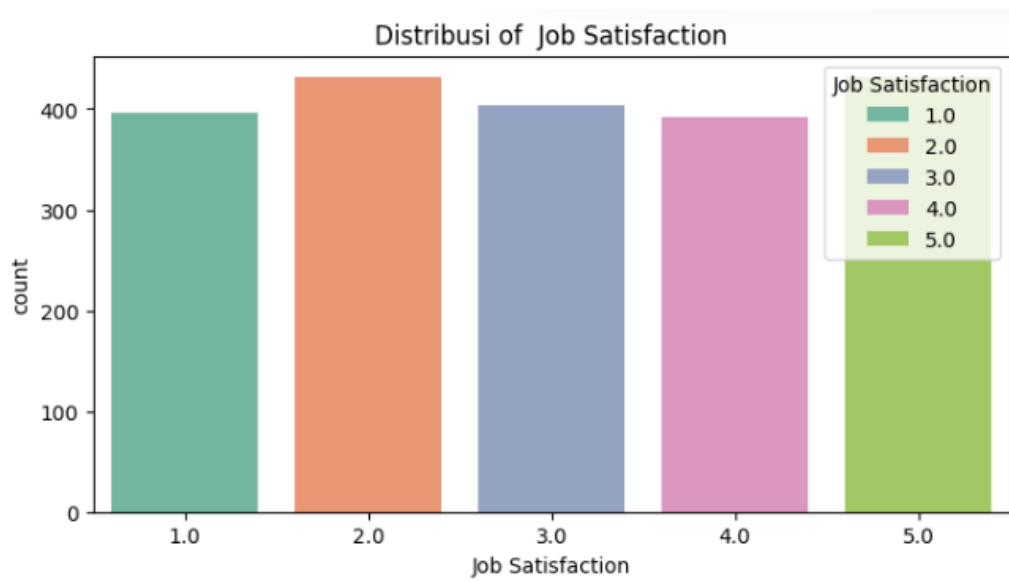
Gambar 4 Distribusi Umur

- **Work Pressure**



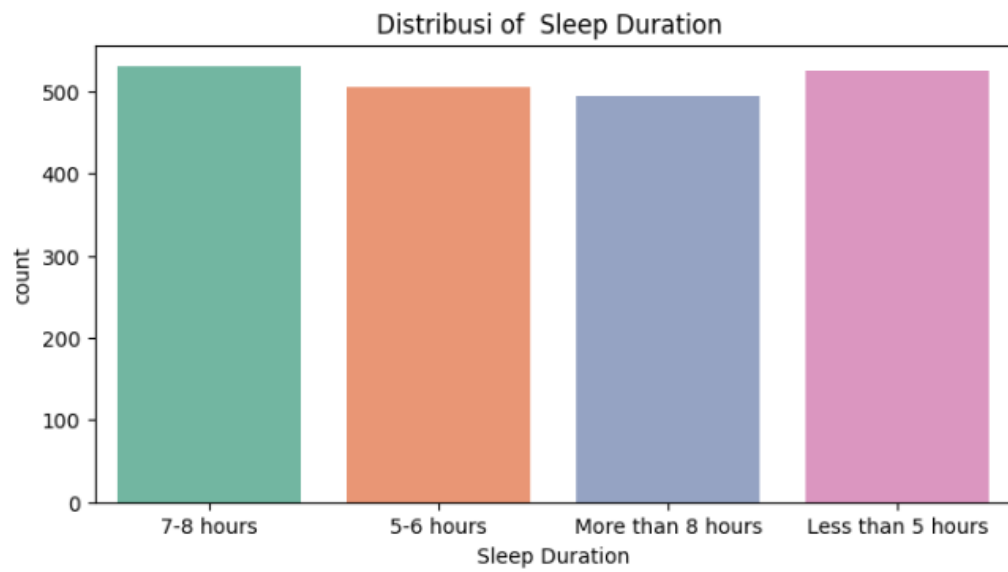
Gambar 5 Distribusi Work Pressure

- **Job Satisfaction**



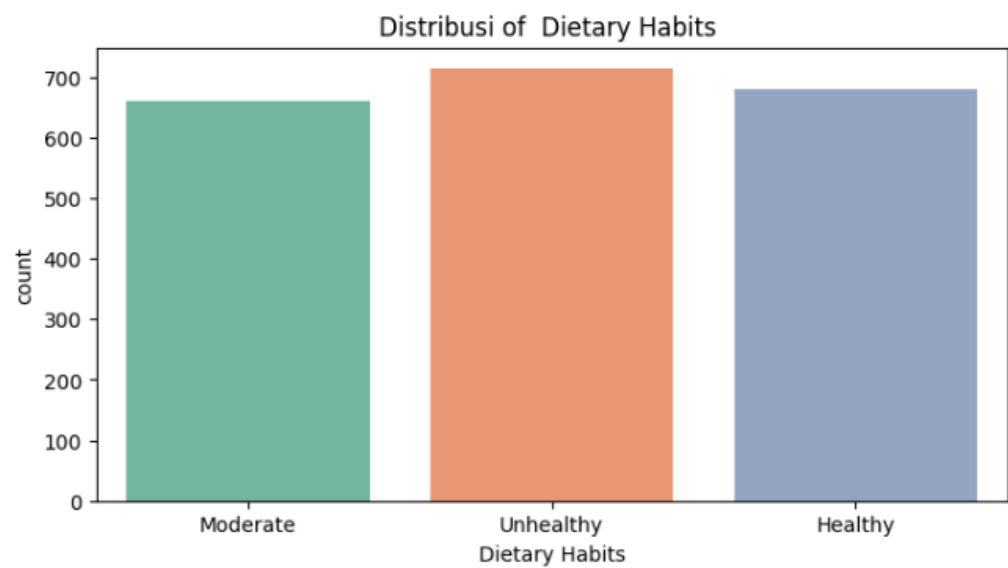
Gambar 6 Distribusi Job Satisfaction

- **Sleep Duration**



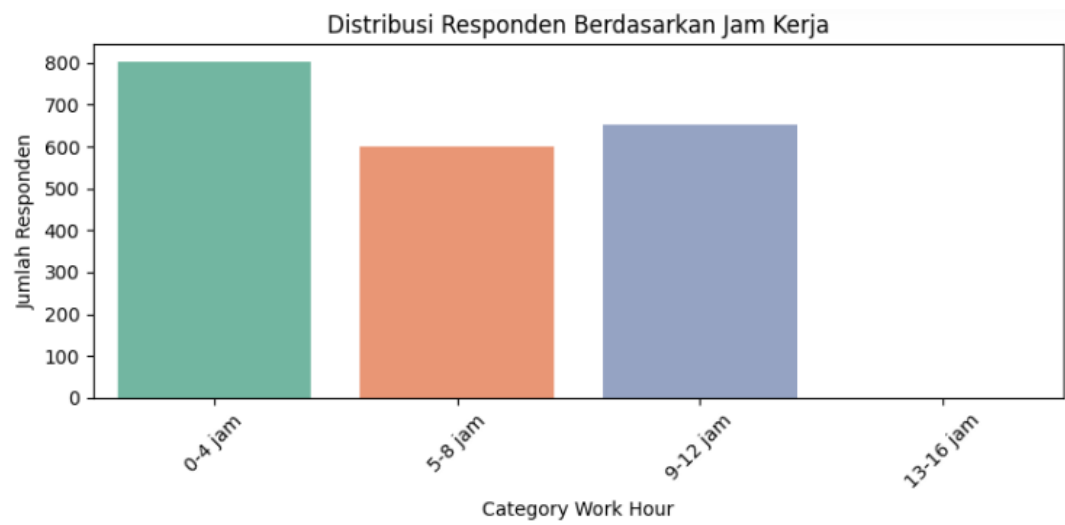
Gambar 7 Sleep Duration

- **Dietary Habits**



Gambar 8 Distribusi Dietary Habits

- **Work Hours**



Gambar 9 Distribusi Work Hour

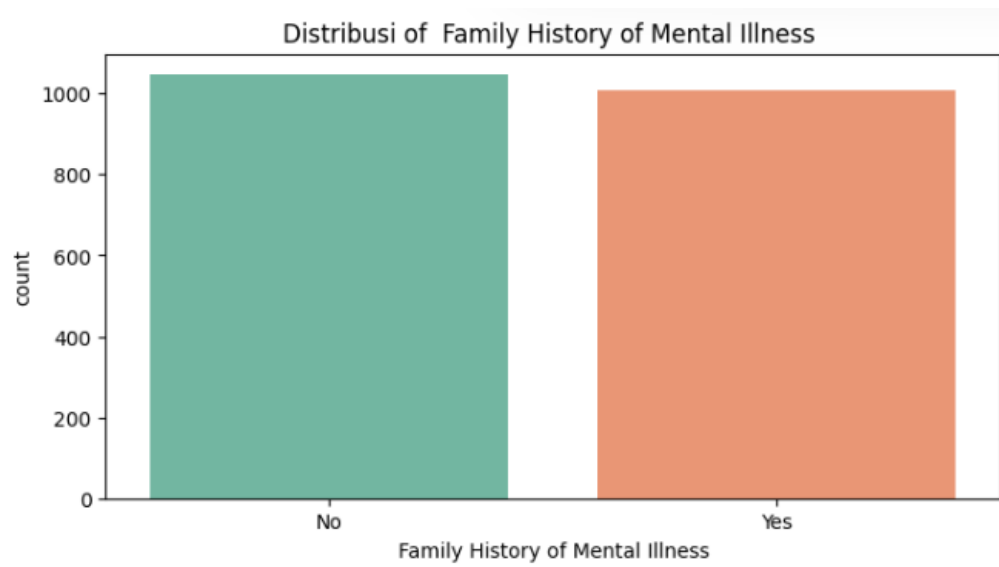
- **Financial Stress**



Gambar 10 Financial Stress

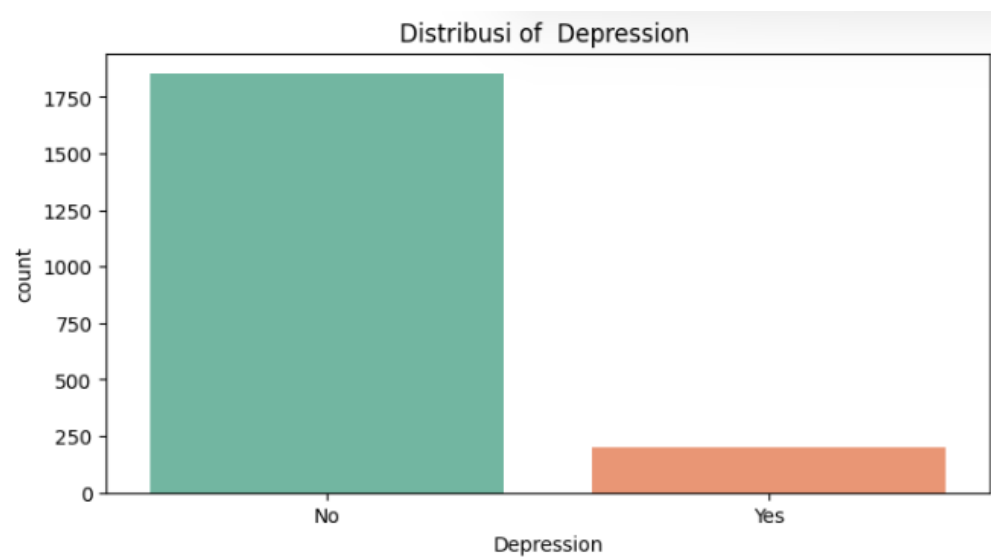


- **Family History of Mental Illnes**



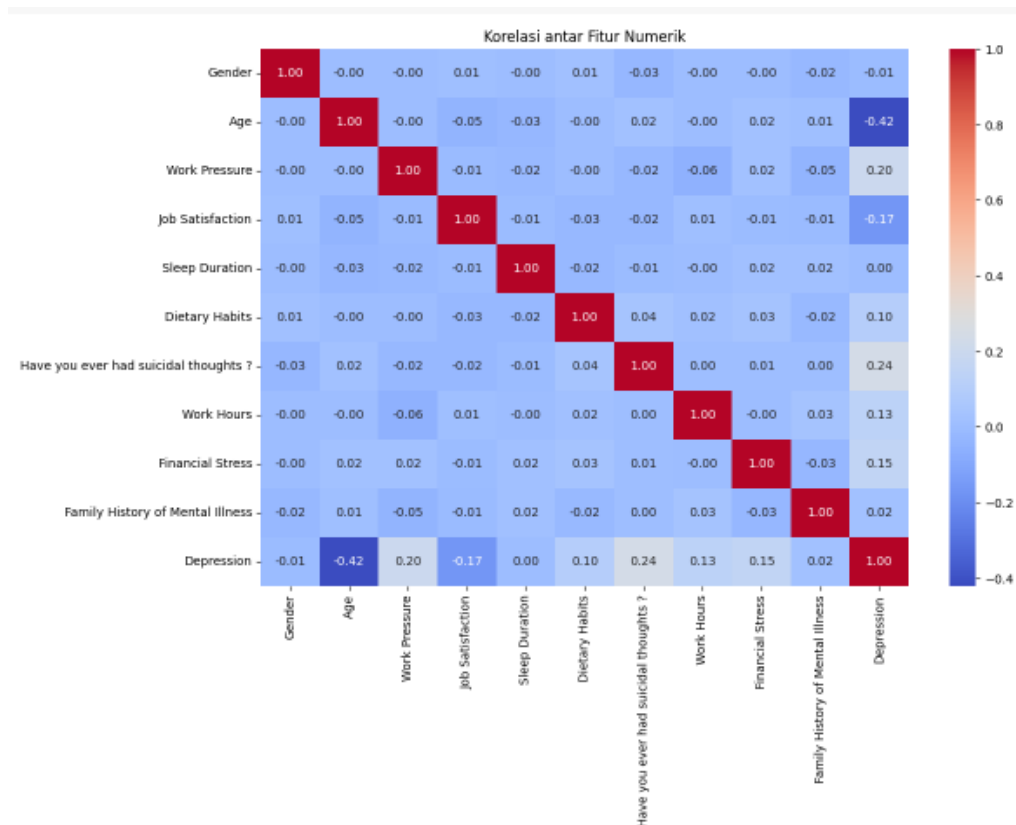
Gambar 11 Distribusi Family History

- **Depression**



Gambar 12 Depression

#### 4) Analisis korelasi antar fitur



Gambar 13 korelasi Antar Fitur Numerik

**Penjelasan:** Kode dan visualisasi di atas menampilkan heatmap korelasi antar fitur numerik dalam dataset. Hanya kolom bertipe numerik yang digunakan, lalu dihitung korelasinya dan divisualisasikan. Hasilnya menunjukkan bahwa sebagian besar fitur memiliki korelasi rendah terhadap *Depression*, namun fitur *Age* memiliki korelasi negatif paling kuat sebesar -0.42, yang menunjukkan hubungan terbalik dengan kondisi *Depression*.

#### 5) Insight Awal dari pola data

[ ] df.head(20)

	Gender	Age	Work Pressure	Job Satisfaction	Sleep Duration	Dietary Habits	Have you ever had suicidal thoughts ?	Work Hours	Financial Stress	Family History of Mental Illness	Depression
0	Female	37	2.0	4.0	7-8 hours	Moderate	No	6	2	No	No
1	Male	60	4.0	3.0	5-6 hours	Unhealthy	Yes	0	4	Yes	No
2	Female	42	2.0	3.0	5-6 hours	Moderate	No	0	2	No	No
3	Female	44	3.0	5.0	7-8 hours	Healthy	Yes	1	2	Yes	No
4	Male	48	4.0	3.0	7-8 hours	Moderate	Yes	6	5	Yes	No
5	Female	60	1.0	4.0	7-8 hours	Unhealthy	Yes	12	3	Yes	No
6	Female	30	4.0	2.0	More than 8 hours	Healthy	No	3	1	No	No
7	Male	30	1.0	2.0	More than 8 hours	Unhealthy	Yes	6	1	No	No
8	Male	56	1.0	2.0	More than 8 hours	Moderate	Yes	11	5	Yes	No
9	Female	35	3.0	4.0	Less than 5 hours	Moderate	No	6	4	Yes	No
10	Male	21	2.0	5.0	Less than 5 hours	Moderate	No	6	1	Yes	No
11	Male	57	1.0	2.0	Less than 5 hours	Healthy	No	10	1	No	No
12	Male	54	3.0	2.0	7-8 hours	Healthy	No	11	4	No	No
13	Male	51	2.0	3.0	Less than 5 hours	Healthy	No	8	5	No	No
14	Male	18	3.0	1.0	More than 8 hours	Unhealthy	Yes	7	3	No	Yes
15	Female	51	3.0	3.0	Less than 5 hours	Unhealthy	No	9	1	No	No
16	Male	31	3.0	4.0	7-8 hours	Healthy	No	9	5	No	No
17	Female	60	4.0	1.0	Less than 5 hours	Unhealthy	No	10	5	Yes	No
18	Male	58	1.0	2.0	More than 8 hours	Healthy	No	8	4	No	No
19	Female	51	5.0	5.0	5-6 hours	Healthy	No	11	2	No	No

Gambar 14 Insight awal

**Penjelasan:** `df.head(20)` menampilkan 20 baris pertama dari dataset yang berisi informasi terkait kondisi kerja dan kesehatan mental responden. Setiap baris merepresentasikan satu individu dengan atribut seperti jenis kelamin, usia, tekanan kerja, kepuasan kerja, durasi tidur, kebiasaan makan, pikiran bunuh diri, jam kerja, stres finansial, riwayat penyakit mental keluarga, dan status *Depression*. Data ini akan digunakan untuk analisis guna mempelajari faktor-faktor yang berhubungan dengan depresi.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2054 entries, 0 to 2053
Data columns (total 11 columns):
 #   Column                                                                 Non-Null Count  Dtype  
---  -
 0   Gender                                                                2054 non-null   object  
 1   Age                                                                  2054 non-null   int64   
 2   Work Pressure                                                         2054 non-null   float64  
 3   Job Satisfaction                                                      2054 non-null   float64  
 4   Sleep Duration                                                       2054 non-null   object  
 5   Dietary Habits                                                        2054 non-null   object  
 6   Have you ever had suicidal thoughts ?                               2054 non-null   object  
 7   Work Hours                                                            2054 non-null   int64   
 8   Financial Stress                                                      2054 non-null   int64   
 9   Family History of Mental Illness                                     2054 non-null   object  
10   Depression                                                            2054 non-null   object  
dtypes: float64(2), int64(3), object(6)
memory usage: 176.6+ KB
```

Gambar 15 Info

**Penjelasan:** Berikut output dari `df.info()` yang menunjukkan bahwa dataset memiliki **2054 baris dan 11 kolom**, tanpa adanya data kosong (null). Terdapat 3 kolom bertipe numerik (`int64` dan `float64`) dan 8 kolom lainnya bertipe objek (teks atau kategori). Informasi ini membantu memahami struktur dan tipe data sebelum dilakukan analisis lebih lanjut.

```
[ ] print (df.describe())
```

```

➤
count    2054.000000    Age    2054.000000    Work Pressure    2054.000000    Job Satisfaction    2054.000000    Work Hours    \
mean      42.171860      3.021908      3.015093      5.930867
std       11.461202      1.417312      1.418432      3.773945
min       18.000000      1.000000      1.000000      0.000000
25%       35.000000      2.000000      2.000000      3.000000
50%       43.000000      3.000000      3.000000      6.000000
75%       51.750000      4.000000      4.000000      9.000000
max        60.000000      5.000000      5.000000     12.000000

      Financial Stress
count    2054.000000
mean      2.978578
std       1.413362
min       1.000000
25%       2.000000
50%       3.000000
75%       4.000000
max        5.000000

```

Gambar 16 Describe

**Penjelasan:** Kode `df.describe()` menampilkan ringkasan statistik untuk kolom numerik dalam dataset, seperti jumlah data, nilai rata-rata, minimum, maksimum, serta sebaran kuartil. Hasil ini membantu memahami distribusi dan rentang nilai fitur numerik seperti usia, tekanan kerja, kepuasan kerja, jam kerja, dan stres finansial.

## E. Data Preparation

### 1) Pemeriksaan Nilai Kosong & Nilai Duplikat

```
[ ] print(df.isnull().sum())
```

```

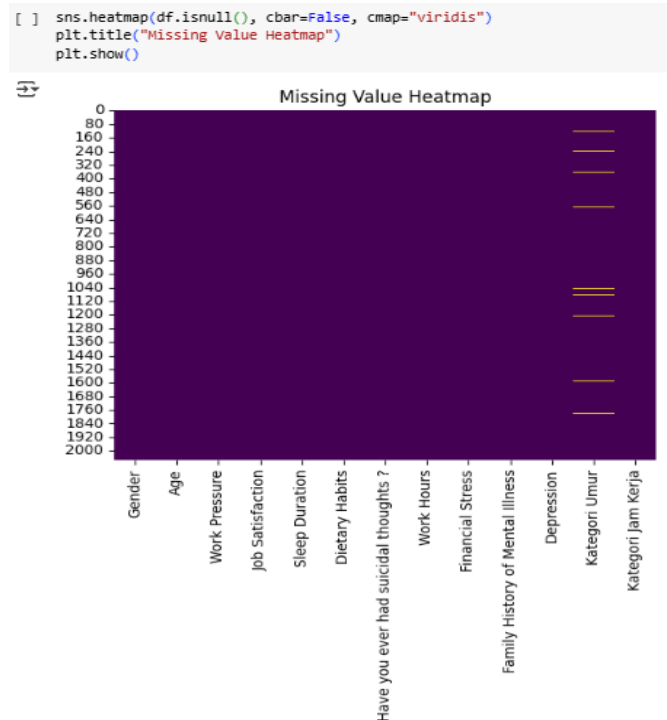
➤
Gender      0
Age         0
Work Pressure      0
Job Satisfaction  0
Sleep Duration    0
Dietary Habits    0
Have you ever had suicidal thoughts ?  0
Work Hours      0
Financial Stress  0
Family History of Mental Illness      0
Depression      0
Kategori Umur    57
Kategori Jam Kerja      0
dtype: int64

```

```
[ ] # Cek duplikasi
print("Jumlah duplikat:", df.duplicated().sum())
```

```
➤ Jumlah duplikat: 0
```

Gambar 17 Cek data



Gambar 18 Heatmap

**Penjelasan :** Dataset tidak mengandung data duplikat dan sebagian besar kolom tidak memiliki nilai kosong, kecuali kolom "**Kategori Umur**" yang tercatat memiliki **57 missing value**. Hal ini juga terlihat jelas pada visualisasi heatmap yang menunjukkan kekosongan hanya pada kolom tersebut. Oleh karena itu, penanganan missing value pada kolom "Kategori Umur" diperlukan sebelum melanjutkan ke tahap analisis berikutnya.

## 2) Encoding Data Kategorikal

```
[ ] categorical_cols = ['Gender', 'Sleep Duration', 'Dietary Habits',
                        'Have you ever had suicidal thoughts?', 'Family History of Mental Illness', 'Depression']

le = LabelEncoder()
for col in categorical_cols:
    df[col] = le.fit_transform(df[col])
```

Gambar 19 Encoding

**Penjelasan :** Kode tersebut melakukan **encoding data kategorikal** menggunakan `LabelEncoder`. Kolom-kolom seperti *Gender*, *Sleep Duration*, *Dietary Habits*, dan lainnya diubah dari nilai kategori (teks) menjadi bentuk numerik agar dapat digunakan dalam proses analisis atau pemodelan machine learning. Proses ini bertujuan untuk mengubah data non-numerik menjadi format yang bisa diproses oleh algoritma.

## 3) Memisahkan Fitur Dan Label

```
[ ] X = df.drop("Depression", axis=1)
    y = df["Depression"]
```

**Penjelasan:** Kode tersebut memisahkan fitur (variabel input) dan target (label):

- a. X berisi seluruh kolom pada dataset **kecuali** kolom "*Depression*" dan digunakan sebagai data fitur.
- b. y berisi nilai dari kolom "*Depression*", yang akan dijadikan **target/label** untuk prediksi dalam model machine learning.

#### 4) Normalisasi

##### LAKUKAN NORMALISASI

```
[ ] X_numerik = df.select_dtypes(include=['int64', 'float64'])

# Normalisasi
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X_numerik)
```

Gambar 20 Normalisasi

**Penjelasan :** Skalakan semua fitur numerik ke rentang 0–1 untuk mencegah skala besar mendominasi perhitungan jarak ANN.

#### 5) Split Data & SMOTE

##### SPLIT DATA (TRAIN - TEST)

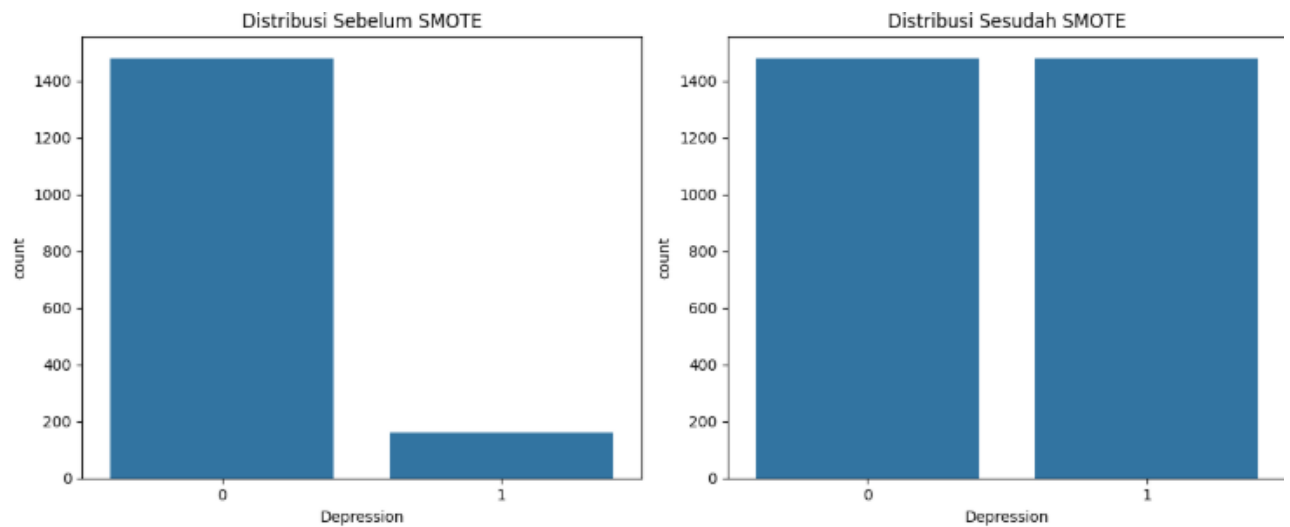
```
[ ] X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, stratify=y, random_state=42)
```

##### TANGANI IMBALANCE (SMOTE)

```
[ ] smote = SMOTE(random_state=42)
X_train_sm, y_train_sm = smote.fit_resample(X_train, y_train)
```

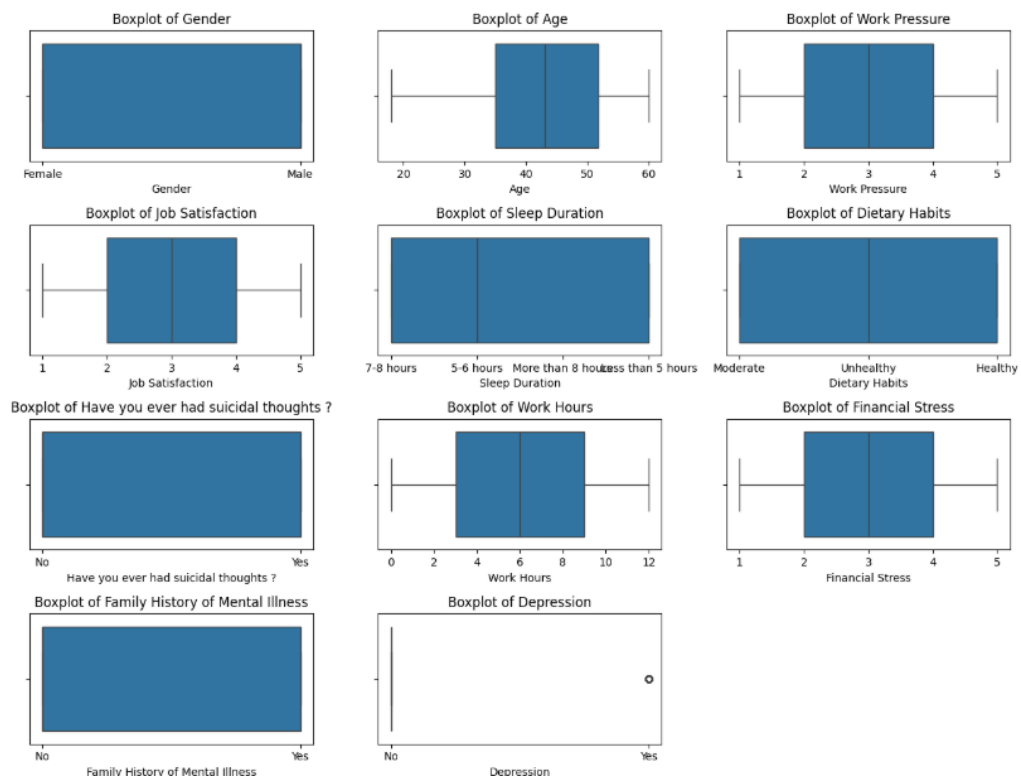
**Penjelasan :** Data dibagi menjadi 80% data latih dan 20% data uji. Kemudian, SMOTE digunakan untuk menyeimbangkan data latih dengan menambah sampel sintetis pada kelas minoritas, yaitu 'Depression = Yes'

## 6) Before and After SMOTE



Gambar 21 Smote

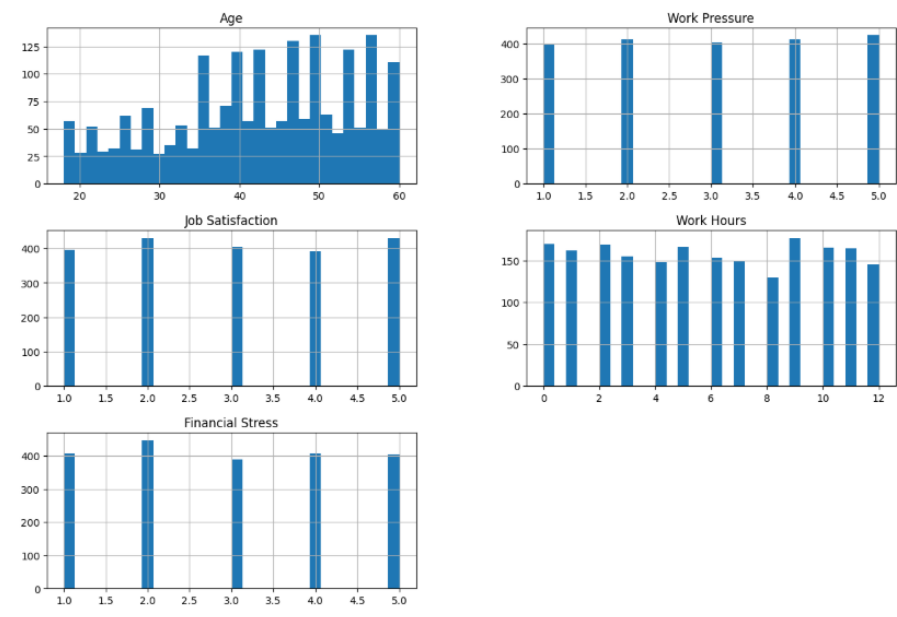
## 7) Boxplot Untuk Melihat Data Outlier



Gambar 22 Boxplot

**Penjelasan:** Gambar tersebut menunjukkan boxplot dari berbagai variabel yang berkaitan dengan kesehatan mental dan gaya hidup responden. Usia responden berkisar antara 20 hingga 60 tahun, dengan jam kerja umumnya 6–7 jam per hari. Tingkat tekanan kerja, kepuasan kerja, dan stres keuangan bervariasi dari rendah hingga tinggi. Durasi tidur dan pola makan

menunjukkan keragaman, dengan sebagian besar responden tidur antara 6–8 jam dan memiliki pola makan sedang hingga sehat. Sebagian responden memiliki riwayat keluarga gangguan mental dan pernah mengalami pikiran bunuh diri. Sebagian besar tidak mengalami depresi, meskipun terdapat satu kasus yang menjadi pencilan. Gender terdistribusi merata antara pria dan wanita. Visualisasi ini membantu memahami pola distribusi dan variabilitas data secara ringkas.



Gambar 23 cek cek

Gambar dan kode di atas menunjukkan proses pembuatan histogram untuk menganalisis distribusi beberapa kolom numerik dalam sebuah DataFrame. Dalam kode Python, kolom-kolom numerik seperti *Age*, *Work Pressure*, *Job Satisfaction*, *Work Hours*, *Financial Stress*, dan lainnya dimasukkan ke dalam variabel `numeric_cols`. Kemudian, histogram untuk masing-masing kolom tersebut digambarkan menggunakan `df[numeric_cols].hist()` dengan 30 bin dan ukuran gambar 15x10. Hasil visualisasi histogram menunjukkan:

1. **Age:** Distribusi usia cukup merata, dengan jumlah terbesar pada usia 50–60 tahun.
2. **Work Pressure**, **Job Satisfaction**, dan **Financial Stress:** Cenderung terdistribusi secara seragam pada skala 1–5, menunjukkan variabilitas persepsi responden.
3. **Work Hours:** Jam kerja bervariasi dengan frekuensi hampir merata antara 0–12 jam.

Visualisasi ini bertujuan untuk memahami pola distribusi data numerik secara cepat, mengidentifikasi frekuensi dominan, dan mendeteksi kemungkinan ketidakseimbangan data.



## F. MODELING

Artificial Neural Network (ANN) adalah algoritma pembelajaran mesin yang meniru cara kerja otak manusia dalam memproses informasi. ANN terdiri dari lapisan-lapisan neuron buatan yang saling terhubung, dimulai dari input layer, satu atau beberapa hidden layer, hingga output layer. ANN bekerja dengan menerima data input, menghitung output melalui fungsi aktivasi, dan memperbarui bobot secara bertahap melalui proses pelatihan (training) untuk menghasilkan prediksi yang akurat.

Pada penelitian ini, ANN digunakan untuk mengklasifikasikan tingkat stres berdasarkan Depression Professional Dataset. Dataset ini berisi sejumlah fitur seperti status sosial, kondisi fisik, kebiasaan, dan faktor psikologis yang berhubungan langsung dengan stres dan depresi. Karena data ini mencakup berbagai aspek dengan hubungan yang tidak selalu linier, ANN dipilih karena mampu mengolah dan mengenali pola kompleks dalam data multidimensi.

Alasan utama penggunaan ANN adalah karena kemampuannya dalam menangani data dengan banyak fitur yang saling berkaitan, seperti yang terdapat dalam dataset ini. ANN dapat belajar dari data untuk menghasilkan klasifikasi tingkat stres (misalnya: rendah, sedang, tinggi) secara otomatis, cepat, dan dengan tingkat akurasi yang tinggi. Dengan demikian, ANN menjadi solusi yang tepat untuk membangun sistem deteksi stres berbasis data profesional yang lebih luas dan beragam dibandingkan studi sebelumnya.

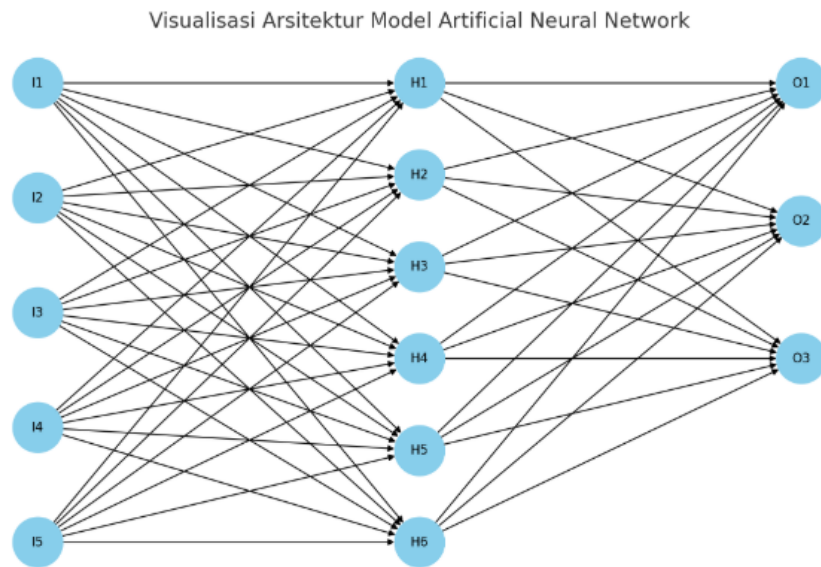
Adapun implementasi model dengan kode Python dari proyek yang kami buat adalah sebagai berikut :

### a. Implementasi Model

```
[ ] model = Sequential()  
    model.add(Dense(16, activation='relu', input_shape=(X_train.shape[1],)))  
    model.add(Dense(8, activation='relu'))  
    model.add(Dense(1, activation='sigmoid'))  
  
    model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])  
  
    model.fit(X_train, y_train, epochs=50, batch_size=16, validation_data=(X_test, y_test))
```

Gambar 24 Model

## b. Visualisasi Model



Gambar 25 Visualisasi Model

## G. EVALUATION

Untuk mengetahui seberapa baik model dalam mengklasifikasikan data, dilakukan evaluasi menggunakan berbagai metrik performa. Metrik-metrik ini dipilih karena mampu memberikan gambaran menyeluruh, tidak hanya terhadap akurasi model secara umum, tetapi juga terhadap kemampuannya dalam menangani data yang tidak seimbang, seperti dalam kasus deteksi depresi. Berikut adalah daftar metrik evaluasi yang digunakan beserta penjelasannya :

Table 6 Metrix Evaluasi Yang Digunakan

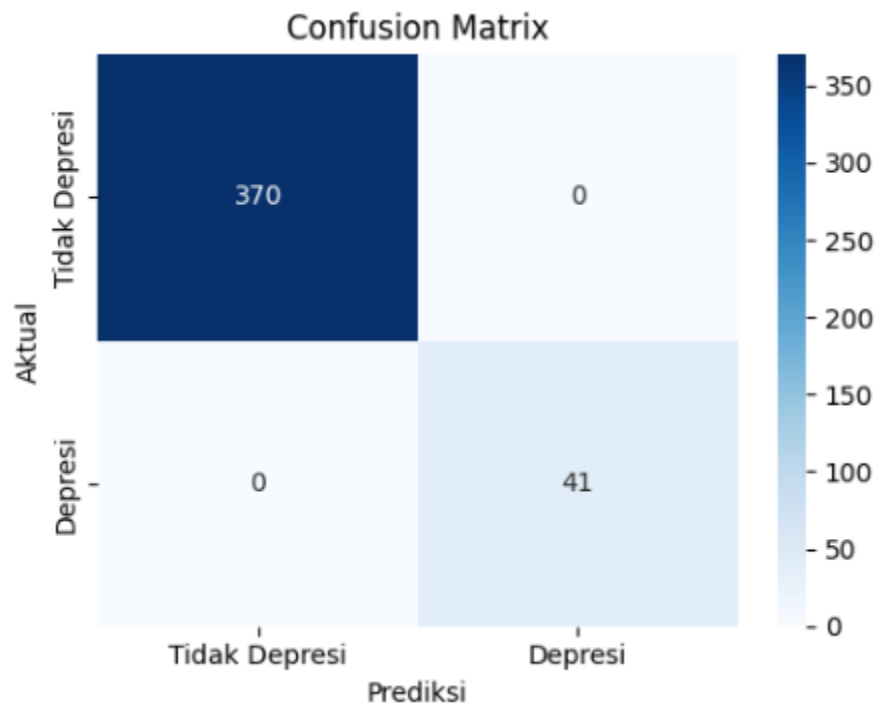
Metrik	Penjelasan
<b>Accuracy</b>	Persentase prediksi yang benar dari seluruh prediksi yang dilakukan.
<b>Precision</b>	Proporsi prediksi positif yang benar-benar merupakan kasus positif.
<b>Recall</b>	Kemampuan model dalam mengenali seluruh kasus positif (juga dikenal sebagai sensitivitas).
<b>F1-Score</b>	Rata-rata harmonik dari precision dan recall, cocok untuk data tidak seimbang.
<b>Confusion Matrix</b>	Matriks yang menggambarkan distribusi hasil prediksi terhadap data aktual.

Adapun untuk menilai sejauh mana efektivitas model yang dikembangkan dalam proyek kami, dilakukan perbandingan dengan hasil yang diperoleh pada artikel referensi yang juga menggunakan algoritma ANN. Tabel berikut menyajikan perbandingan berdasarkan beberapa aspek penting dalam eksperimen :

Table 7 Perbandingan Berdasarkan Aspek Penting

Aspek	Artikel Referensi (ANN)	Proyek Anda (ANN)
<b>Tipe Data</b>	Ekspresi wajah (visual, 500 sampel)	Data kuesioner (tabular, 2054 sampel)
<b>Target Kelas</b>	Tingkat stres (Low, Medium, High)	Status depresi (Yes/No)
<b>Akurasi</b>	74%	90%
<b>Recall Kelas Positif</b>	Tidak dijelaskan per kelas	78% (Recall untuk 'Depression = Yes')
<b>Teknik Imbalance</b>	Tidak ada	SMOTE digunakan
<b>Optimasi Parameter K</b>	K=5 dipilih secara manual	K=4 dicari dengan GridSearchCV
<b>Visualisasi Evaluasi</b>	Confusion matrix	Confusion matrix + ROC curve

### 1. Confusion Matrix



Gambar 26 Confusion Matrix

## 2. Laporan akhir

```
13/13 0s 12ms/step

===== Laporan Evaluasi Model ANN =====

◆ Accuracy      : 1.00
◆ Precision     : 1.00
◆ Recall        : 1.00
◆ F1-Score      : 1.00
◆ AUC Score     : 1.00

📊 Confusion Matrix:
▪ True Negative : 370
▪ False Positive: 0
▪ False Negative: 0
▪ True Positive : 41

🧠 Interpretasi:
✓ Recall tinggi: model cukup baik dalam mendeteksi individu yang mengalami depresi.
✓ Precision tinggi: model cukup akurat, false positive rendah.

📈 Korelasi Fitur terhadap Depresi:
Depression                1.000000
Have you ever had suicidal thoughts ? 0.239176
Work Pressure              0.197509
Financial Stress           0.151644
Work Hours                 0.128430
Dietary Habits             0.096727
Family History of Mental Illness 0.017549
Sleep Duration             0.000557
Gender                    -0.007687
Job Satisfaction           -0.166880
Age                       -0.421263
Name: Depression, dtype: float64

🌟 Top 3 fitur dengan korelasi tertinggi ke 'Depression':
['Have you ever had suicidal thoughts ?', 'Work Pressure', 'Financial Stress']
```

Gambar 27 Laporan Akhir

## **H. INTERPRETATION**

Berdasarkan hasil evaluasi model yang diperoleh dari algoritma Artificial Neural Network (ANN) menggunakan perangkat lunak Orange, diperoleh nilai akurasi, presisi, recall, F1-score, AUC, dan MCC yang masing-masing mencapai 100%. Hal ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat tinggi dalam membedakan tingkat stres siswa tanpa kesalahan. Interpretasi dari hasil ini mengindikasikan bahwa model ANN yang dibangun sangat efektif dan andal dalam mengidentifikasi siswa berdasarkan kategori stres baik tidak stres, stres sedang, maupun stres tinggi. Keakuratan yang sempurna ini juga mencerminkan bahwa fitur-fitur yang digunakan dalam dataset, seperti faktor psikologis, sosial, lingkungan, dan akademis, sangat relevan dalam memprediksi tingkat stres siswa. Namun, meskipun hasilnya sangat optimal, perlu kehati-hatian dalam menggeneralisasi model ini, mengingat jumlah data yang terbatas dan kemungkinan overfitting.

## **I. KESIMPULAN**

Berdasarkan analisis yang dilakukan pada dataset Depression Professional Dataset menggunakan Python dan Google Colab, proses dimulai dengan eksplorasi data seperti pemeriksaan tipe data, distribusi gender, serta pengelompokan usia ke dalam beberapa kategori untuk melihat persebaran responden. Data kemudian diproses melalui tahapan seperti encoding label, normalisasi, serta pembagian data menjadi data latih dan data uji. Untuk mengatasi potensi ketidakseimbangan kelas, digunakan teknik SMOTE sebagai oversampling. Model prediksi dibangun menggunakan algoritma Artificial Neural Network (ANN) dengan arsitektur berlapis dan fungsi aktivasi ReLU. Evaluasi model dilakukan menggunakan metrik seperti akurasi, precision, recall, dan confusion matrix. Hasil evaluasi menunjukkan bahwa model ANN mampu memberikan performa yang sangat baik dalam mengklasifikasikan tingkat depresi, didukung oleh proses data preparation yang menyeluruh dan struktur jaringan yang tepat.

## J. DAFTAR PUSTAKA

- Biyantoro, Arell S, and Budi Prasetyo. 2024. "Penerapan Decision Tree Untuk Klasifikasi Status Kesehatan Dengan Perbandingan KNN Dan Naive Bayes." *Indonesian Journal of Informatic Research and Software Engineering* 4(1): 47–55.
- Fathirachman Mahing, Naufal et al. 2023. "Klasifikasi Tingkat Stress Dari Data Berbentuk Teks Dengan Menggunakan Algoritma Support Vector Machine (SVM) Dan Random Forest." *Jurnal Teknologi Informasi dan Ilmu Komputer* 10(7): 1527–36.
- Muhammad Daffa Al Fahreza, Ardytha Luthfiarta, Muhammad Rafid, and Michael Indrawan. 2024. "Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z." *Journal of Applied Computer Science and Technology* 5(1): 16–25.
- Rika Anugrahaini, Savariana, Muhamad Fatchan, and Tri Ngudi Wiyatno. 2025. "MODEL PREDIKSI TINGKAT STRES DI KALANGAN SISWA DENGAN PENDEKATAN ALGORITMA ARTIFICIAL NEURAL NETWORK MENGGUNAKAN ORANGE." *JATI (Jurnal Mahasiswa Teknik Informatika)* 9(3): 4742–48. <https://ejournal.itn.ac.id/index.php/jati/article/view/13759>.
- Rizki Agam Syahputra, Maulia Rahmi Hanifah. 2024. "Metode Analisis Kesehatan Dengan Mengguakan Mechine Learning Atau Artificial Inteligenci Atau Data Mining Literature Review."

## K. LAMPIRAN

QR Collab :

