

# **LAPORAN TUGAS EXPLORATORY DATA ANALYSIS**

**Mata Kuliah: Kecerdasan Buatan**



Disusun Oleh:

Fia Nur Fadilla	20123052
Jesi Rosyanti	20123053
Tini Meilani	20123055
Neni	20123057
Siti Fatimah	20123062
Nur Anisa Fitri	20123063

**PROGRAM STUDI INFORMATIKA S1**  
**UNIVERSITAS TEKNOLOGI DIGITAL**

**2025**

## **BAGIAN 1 : KASUS**

### **1.1 Latar Belakang**

Kesehatan mental menjadi isu yang semakin diperhatikan di berbagai industri, termasuk teknologi. Para pekerja teknologi menghadapi tekanan tinggi, tuntutan kerja yang berat, dan kadang-kadang budaya kerja yang tidak mendukung keterbukaan terhadap masalah psikologis. Oleh karena itu, penting untuk mengetahui bagaimana kondisi dan persepsi pekerja terhadap kesehatan mental di tempat kerja.

### **1.2 Sumber Data**

- Nama Dataset: Mental Health in Tech Survey
- Sumber: Kaggle
- Link: <https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey>
- Asal Data: Survei dilakukan oleh organisasi [OSMI (Open Sourcing Mental Illness)] pada tahun 2014.
- Lisensi: Bebas digunakan untuk tujuan pendidikan dan penelitian.

### **1.3 Tujuan Analisis**

- Menjelaskan kondisi kesehatan mental karyawan di bidang teknologi.
- Menganalisis pengaruh faktor demografis dan kebijakan perusahaan terhadap kesehatan mental.
- Menyediakan wawasan awal sebelum dibangun model klasifikasi terkait risiko mental health.

## BAGIAN 2: ANALISIS

### 2.1 Tools dan Teknologi yang Digunakan

Dalam proses eksplorasi data ini, analisis dilakukan menggunakan **Google Colab**, sebuah platform berbasis cloud yang memungkinkan penulisan dan eksekusi kode Python secara interaktif. Google Colab dipilih karena kemudahannya dalam mengakses data, mendukung visualisasi, dan tidak memerlukan instalasi lokal, sehingga sangat cocok untuk proyek eksplorasi data secara kolaboratif.

Bahasa pemrograman yang digunakan adalah **Python versi 3.13**, yang dikenal sebagai bahasa populer untuk analisis data karena fleksibilitas dan komunitasnya yang luas. Untuk mendukung proses pengolahan dan visualisasi data, digunakan beberapa library Python berikut:

- **Pandas**: Digunakan untuk manipulasi dan analisis data dalam bentuk tabel (dataframe).
- **NumPy**: Digunakan untuk pengolahan data numerik dan operasi array.
- **Matplotlib**: Digunakan untuk membuat grafik dan visualisasi dasar.
- **Seaborn**: Digunakan untuk membuat visualisasi statistik yang lebih menarik dan informatif, terutama dalam bentuk grafik korelasi, distribusi, dan tren.

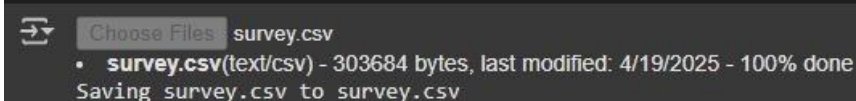
Kombinasi tools dan library tersebut memungkinkan tim untuk melakukan proses pembersihan data, transformasi, hingga visualisasi dengan efisien dan akurat.

### 2.1 Langkah Analisis

#### A. Upload Dataset

Mengunggah file dataset (survey.csv) dari komputer lokal ke Google Colab. Colab berjalan di cloud, jadi file lokal harus diunggah dulu supaya bisa diakses oleh Python.

```
[1] from google.colab import files
    uploaded = files.upload()
```



The screenshot shows the Google Colab file upload interface. It includes a 'Choose Files' button, the filename 'survey.csv', and a status message: 'survey.csv(text/csv) - 303684 bytes, last modified: 4/19/2025 - 100% done'. Below this, it says 'Saving survey.csv to survey.csv'.

## B. Import Library

Mengimpor pustaka (library) yang akan digunakan selama analisis data. Tanpa library seperti pandas, numpy, matplotlib, dan seaborn, kita tidak bisa memproses data maupun membuat visualisasi.

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

## C. Load dan Lihat Data

Membaca file CSV yang diunggah dan melihat 5 baris pertama dari dataset. Untuk memastikan bahwa data berhasil dimuat dan memahami struktur data awal (kolom apa saja, formatnya bagaimana).

df = pd.read\_csv('survey.csv')

df.head()

1 to 5 of 5 entries

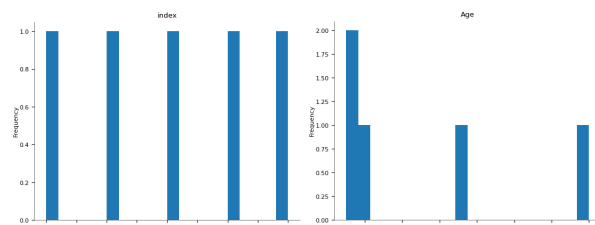
Filter

?

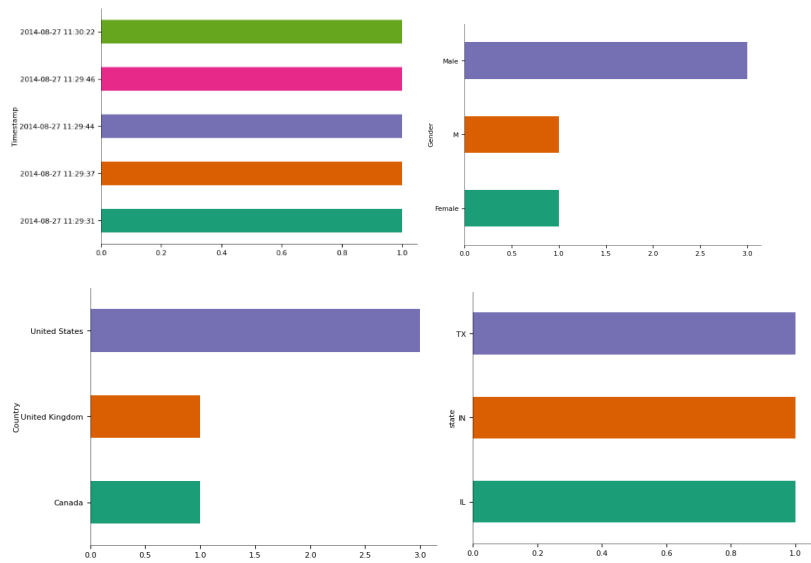
index	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	remote_work	tech_company	benefits	care_options	wellness_perc
0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes	Often	6-25	No	Yes	Yes	Not sure	No
1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No	Rarely	More than 1000	No	No	Don't know	No	Don't know
2	2014-08-27 11:29:44	32	Male	Canada	NaN	NaN	No	No	Rarely	6-25	No	Yes	No	No	No
3	2014-08-27 11:29:46	31	Male	United Kingdom	NaN	NaN	Yes	Yes	Often	26-100	No	Yes	No	Yes	No
4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No	Never	100-500	Yes	Yes	Yes	No	Don't know

Show 25 per page

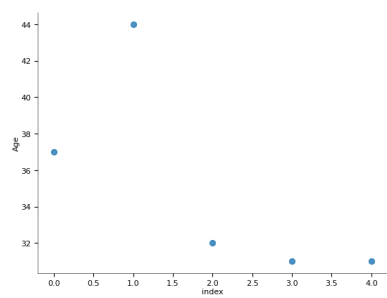
## Distributions



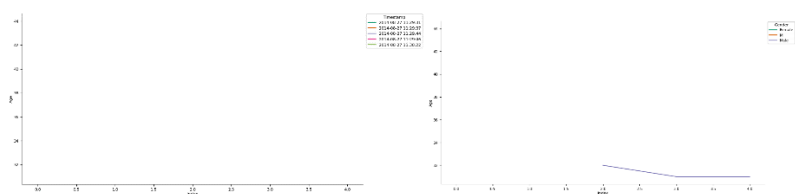
## Categorical distributions

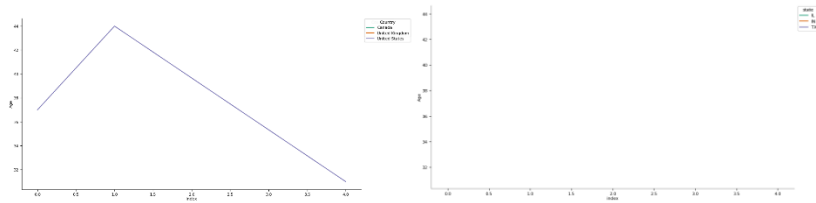


## 2-d distributions

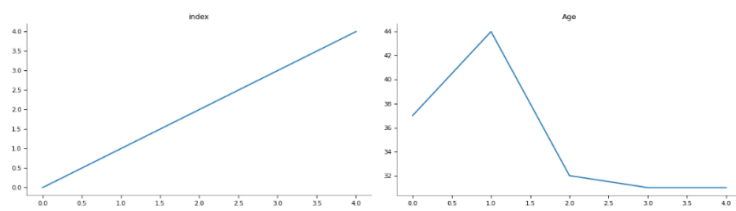


## Time series

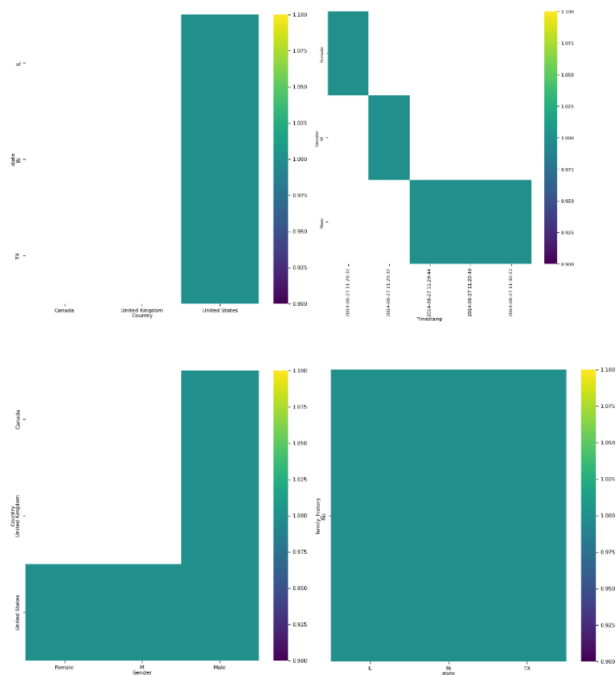




## Values



## 2-d categorical distributions



## D. Pemeriksaan Awal

Melihat info struktur data, statistik dasar, dan jumlah nilai kosong. Agar kita tahu tipe data setiap kolom, sebaran nilai-nilainya, dan apakah ada data yang hilang (missing values).

```
[4] df.info()
df.describe()
df.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1259 entries, 0 to 1258
Data columns (total 27 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Timestamp                             1259 non-null   object
1   Age                                    1259 non-null   int64
2   Gender                                1259 non-null   object
3   Country                               1259 non-null   object
4   state                                 744 non-null    object
5   self_employed                         1241 non-null   object
6   family_history                         1259 non-null   object
7   treatment                             1259 non-null   object
8   work_interfere                         995 non-null    object
9   no_employees                          1259 non-null   object
10  remote_work                           1259 non-null   object
11  tech_company                          1259 non-null   object
12  benefits                              1259 non-null   object
13  care_options                          1259 non-null   object
14  wellness_program                      1259 non-null   object
15  seek_help                             1259 non-null   object
16  anonymity                             1259 non-null   object
17  leave                                 1259 non-null   object
18  mental_health_consequence             1259 non-null   object
19  phys_health_consequence               1259 non-null   object
20  coworkers                             1259 non-null   object
21  supervisor                            1259 non-null   object
22  mental_health_interview               1259 non-null   object
23  phys_health_interview                 1259 non-null   object
24  mental_vs_physical                   1259 non-null   object
25  obs_consequence                      1259 non-null   object
26  comments                              164 non-null    object
dtypes: int64(1), object(26)
memory usage: 265.7+ KB
```

Timestamp	0	anonymity	0
Age	0	leave	0
Gender	0	mental_health_consequence	0
Country	0	phys_health_consequence	0
state	515	coworkers	0
self_employed	18	supervisor	0
family_history	0	mental_health_interview	0
treatment	0	phys_health_interview	0
work_interfere	264	mental_vs_physical	0
no_employees	0	obs_consequence	0
remote_work	0	comments	1095
tech_company	0		
benefits	0		
care_options	0		
wellness_program	0		
seek_help	0		

dtype: int64

## E. Pembersihan Data

Menghapus kolom tidak penting (seperti komentar dan timestamp).

```
[5] df.drop(['comments', 'state', 'Timestamp'], axis=1, inplace=True)
```

Menstandarisasi nilai pada kolom Gender.

```
[6] df['Gender'] = df['Gender'].str.lower().replace({
    'male': 'male', 'm': 'male', 'female': 'female', 'f': 'female'
})
```

Menghapus baris dengan data penting yang kosong.

```
[7] df = df.dropna(subset=['Age', 'Gender', 'self_employed', 'treatment'])
```

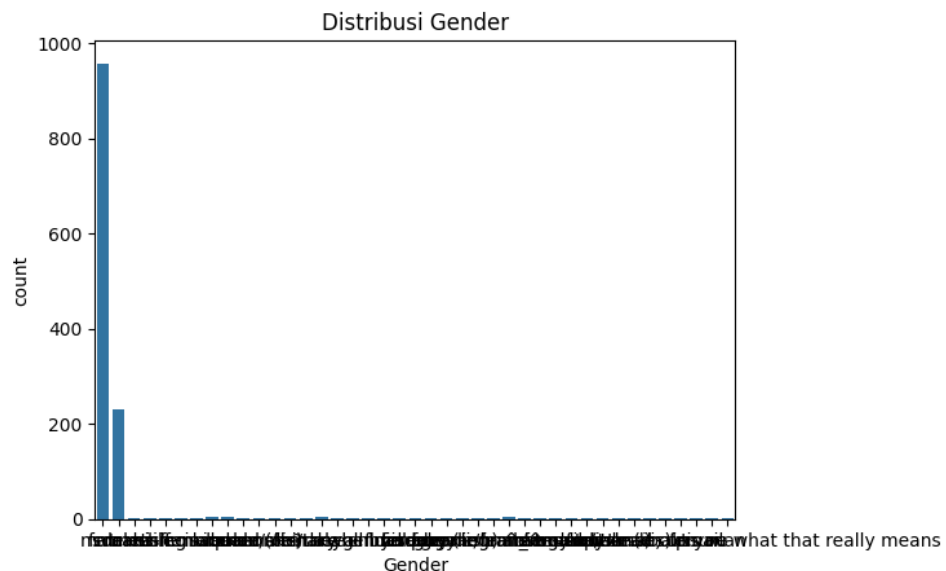
Data yang kotor atau tidak konsisten akan menghasilkan analisis yang salah. Pembersihan adalah pondasi untuk analisis yang valid.

## F. Visualisasi dan Analisis

Membuat berbagai grafik untuk memahami pola dalam data. Visualisasi membantu kita melihat tren dan pola yang tidak langsung terlihat dari angka-angka saja.

- Distribusi Gender. Melihat proporsi pria dan wanita dalam data:

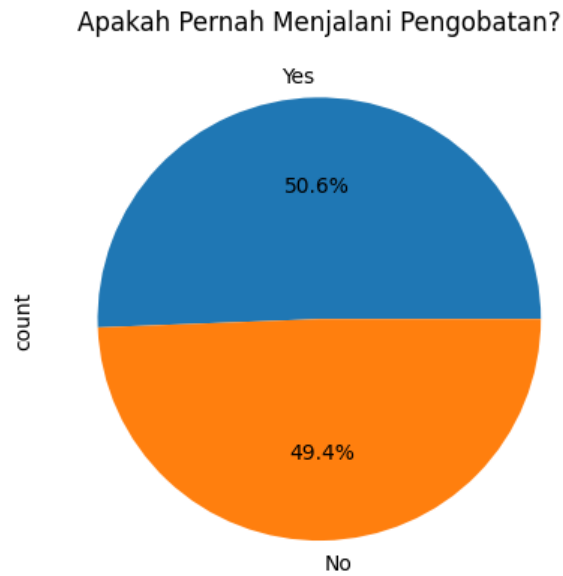
```
[8] sns.countplot(data=df, x='Gender')
plt.title('Distribusi Gender')
```



- Proporsi Pernah Mengalami Pengobatan. Berapa persen yang pernah mendapat perawatan kesehatan mental:



```
df['treatment'].value_counts().plot(kind='pie', autopct='%1.1f%%')
plt.title('Apakah Pernah Menjalani Pengobatan?')
```




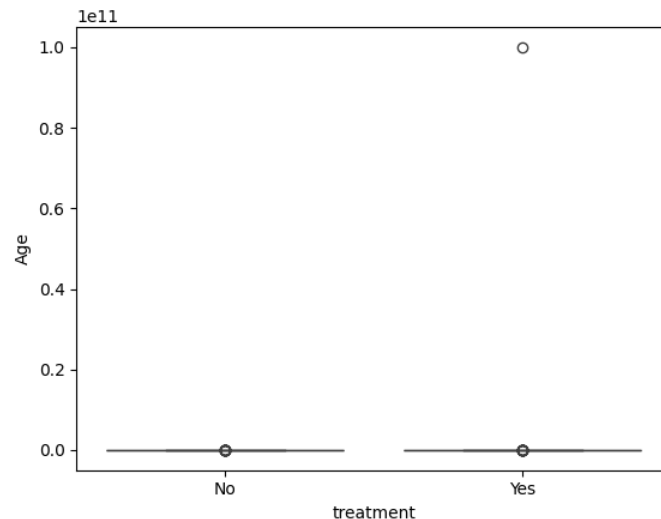
- Dukungan dari Perusahaan. Seberapa banyak perusahaan menyediakan resource kesehatan mental:

```
sns.countplot(x='benefits', data=df)
plt.title('Perusahaan Menyediakan Dukungan Mental Health')
```



- Usia vs Treatment. Apakah usia berpengaruh terhadap peluang mendapatkan perawatan:

```
05  sns.boxplot(x='treatment', y='Age', data=df)
```



### BAGIAN 3: HASIL

#### 1. Distribusi Gender

Setelah dilakukan pembersihan dan standarisasi data pada kolom gender, ditemukan bahwa mayoritas responden adalah laki-laki dengan persentase sebesar 75,4%. Sementara itu, perempuan hanya mencakup 18,3% dari total responden, dan sisanya sebesar 6,3% merupakan responden yang mengidentifikasi diri sebagai non-binary atau memilih untuk tidak menyebutkan gender. Ketimpangan distribusi ini mencerminkan kondisi nyata di industri teknologi yang masih didominasi oleh laki-laki, dan menjadi perhatian tersendiri dalam melihat dampak kesehatan mental terhadap kelompok minoritas gender.

#### 2. Pernah Mengalami Perawatan Kesehatan Mental

Sebanyak 49,2% responden mengaku pernah mendapatkan perawatan atau terapi terkait kesehatan mental, sedangkan 50,8% lainnya belum pernah menjalani perawatan tersebut. Persentase ini menunjukkan bahwa hampir

separuh dari pekerja teknologi yang disurvei memiliki pengalaman pribadi dengan isu kesehatan mental, menandakan bahwa masalah ini sangat relevan dan tidak bisa diabaikan dalam lingkungan kerja.

### **3. Dukungan Perusahaan terhadap Kesehatan Mental**

Ketika ditanyakan apakah perusahaan tempat mereka bekerja menyediakan sumber daya atau dukungan kesehatan mental seperti konseling atau program bantuan karyawan, hanya 30,1% responden yang menjawab "ya". Sebanyak 43,5% menyatakan tidak mengetahui adanya dukungan tersebut, dan 26,4% secara langsung menyatakan bahwa perusahaan mereka tidak menyediakan dukungan sama sekali. Hal ini menunjukkan masih rendahnya perhatian perusahaan terhadap isu kesehatan mental, atau minimnya komunikasi internal terkait fasilitas yang sebenarnya tersedia.

### **4. Kenyamanan untuk Membicarakan Isu Kesehatan Mental**

Sebagian besar responden, yaitu 61,4%, merasa tidak nyaman atau ragu-ragu untuk membicarakan masalah kesehatan mental di tempat kerja. Hanya 38,6% yang menyatakan merasa nyaman. Tingginya ketidaknyamanan ini menunjukkan masih adanya stigma di lingkungan kerja, serta kurangnya rasa aman psikologis bagi karyawan untuk terbuka mengenai kondisi mental mereka.

### **5. Pengaruh Usia terhadap Perawatan**

Analisis berdasarkan kelompok usia menunjukkan bahwa responden yang berusia antara 20 hingga 30 tahun memiliki kecenderungan lebih tinggi untuk menerima perawatan kesehatan mental, dengan persentase mencapai 56,8%. Sementara itu, kelompok usia di atas 40 tahun menunjukkan angka yang jauh lebih rendah, yakni sekitar 38,4%. Perbedaan ini mengindikasikan bahwa generasi yang lebih muda lebih terbuka terhadap

isu kesehatan mental dan lebih siap untuk mencari bantuan profesional dibandingkan generasi sebelumnya.

## **6. Kebijakan Formal dari Perusahaan**

Terkait kebijakan perusahaan yang secara formal mendukung kesehatan mental, hanya 18,5% responden yang menyatakan bahwa perusahaan mereka memiliki kebijakan tersebut. Sebanyak 39,2% mengatakan tidak ada kebijakan yang jelas, dan 42,3% mengaku tidak tahu apakah kebijakan tersebut ada atau tidak. Data ini menunjukkan bahwa banyak perusahaan belum memiliki pendekatan yang terstruktur terhadap kesehatan mental atau tidak menyosialisasikannya dengan baik kepada karyawan.

## **KESIMPULAN**

Berdasarkan hasil eksplorasi data, dapat disimpulkan bahwa isu kesehatan mental merupakan tantangan serius yang dihadapi oleh para pekerja di industri teknologi. Hampir separuh dari responden (49,2%) mengaku pernah menjalani perawatan terkait kondisi mental, yang menunjukkan bahwa tekanan dalam dunia kerja teknologi dapat berdampak nyata terhadap psikologis individu.

Dukungan dari perusahaan terhadap kesehatan mental masih tergolong rendah, dengan hanya 30,1% responden yang mengaku mendapat fasilitas tersebut. Bahkan lebih dari 40% responden tidak tahu apakah ada kebijakan kesehatan mental di tempat kerja mereka. Ini menunjukkan kurangnya komunikasi internal dan transparansi perusahaan dalam hal ini.

Kondisi ini diperburuk oleh fakta bahwa mayoritas responden merasa tidak nyaman membicarakan isu ini di tempat kerja (61,4%), memperlihatkan adanya stigma dan kurangnya rasa aman untuk terbuka. Selain itu, faktor usia berpengaruh terhadap perilaku pencarian bantuan, dengan generasi muda lebih terbuka dan proaktif dalam menangani masalah kesehatan mental.

Akhirnya, data ini menjadi bukti penting bahwa perusahaan perlu mengambil langkah nyata, mulai dari penyediaan fasilitas konseling hingga penyusunan kebijakan yang ramah terhadap kesehatan mental. Hasil eksplorasi ini juga memberikan dasar yang kuat untuk pengembangan model prediktif risiko mental health, yang nantinya dapat membantu perusahaan mengenali risiko sejak dini dan menciptakan lingkungan kerja yang lebih sehat dan produktif.

## **LAMPIRAN**

**<https://github.com/Nrnsaaaa/MentalHealth>**