

– Tim Panda'S

Eksplorasi Masa Depan Dunia Kerja yang Berkaitan dengan Transformasi Digital

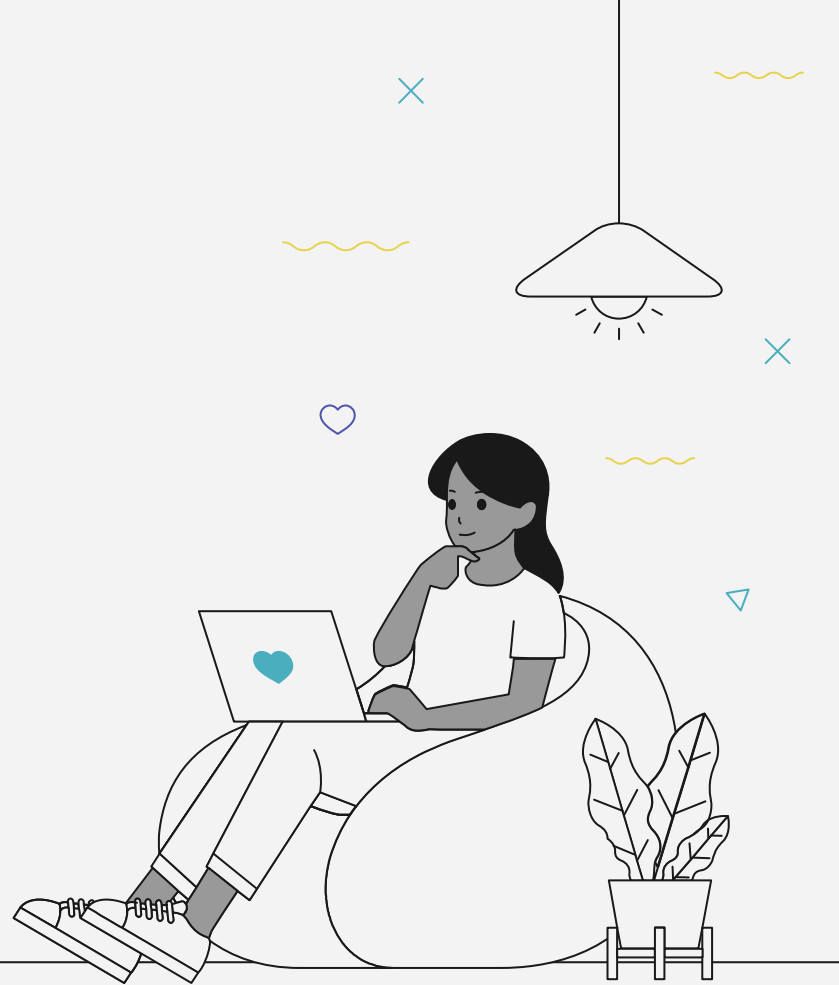




Table of contents

01

Introduction

02

Data Overview

03

EDA

04

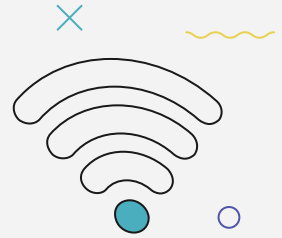
Preparation

05

Modeling

06

Evaluasi



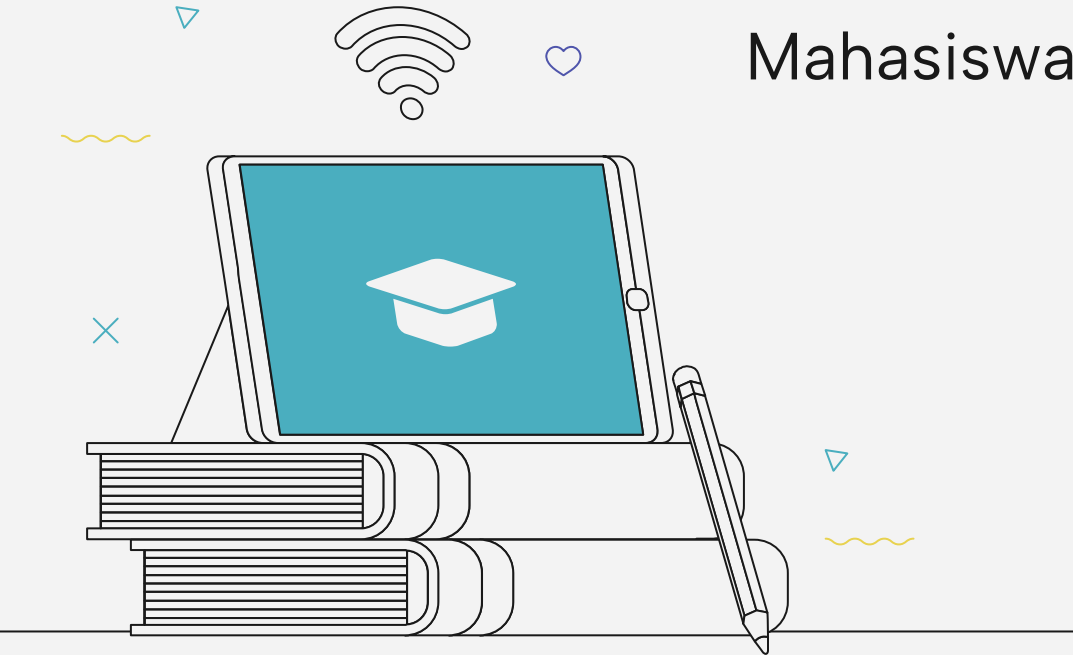
Problem

Jumlah mahasiswa yang lulus tepat waktu, menjadi salah satu parameter dalam penilaian akreditasi program studi. Pada Fakultas Informatika di sebuah kampus, diketahui terdapat 35% mahasiswa yang tidak lulus tepat waktu. Sehingga diperlukan sebuah strategi untuk meningkatkan prosentase kelulusan tepat waktu. Sayangnya, manajemen prodi tidak mengetahui apa saja faktor yang mempengaruhi waktu kelulusan mahasiswa.



"Menurunkan Persentase
Mahasiswa yang Lulus Tidak Tepat
Waktu Menjadi 25%."

— **Goals of Tim Panda's**



Data Overview

Terdapat **803** rows dan **35** Columns



Numerical

```
data[num].columns  
  
Index(['IPS_SMT1', 'IPS_SMT2', 'IPS_SMT3', 'IPS_SMT4', 'IPS_SMT5', 'IPS_SMT6',  
       'IPS_SMT7', 'IPS_SMT8', 'IPS_SMT9', 'IPS_SMT10', 'IPS_SMT11',  
       'IPS_SMT12', 'IPS_SMT13', 'IPS_SMT14', 'SKS_AMBIL_SMT1',  
       'SKS_AMBIL_SMT2', 'SKS_AMBIL_SMT3', 'SKS_AMBIL_SMT4', 'SKS_AMBIL_SMT5',  
       'SKS_AMBIL_SMT6', 'SKS_AMBIL_SMT7', 'SKS_AMBIL_SMT8', 'SKS_AMBIL_SMT9',  
       'SKS_AMBIL_SMT10', 'SKS_AMBIL_SMT11', 'SKS_AMBIL_SMT12',  
       'SKS_AMBIL_SMT13', 'SKS_AMBIL_SMT14'],  
      dtype='object')
```



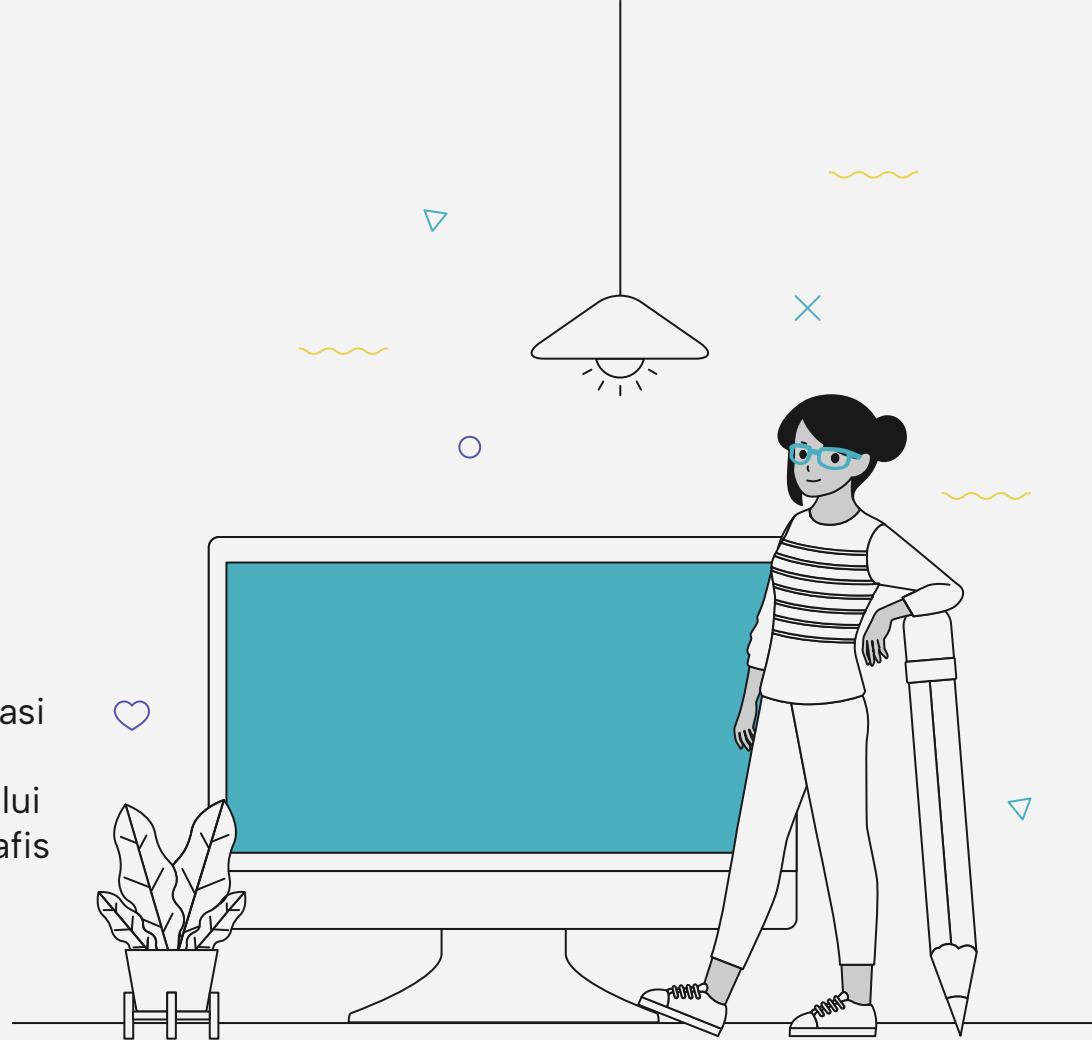
Catagorical

```
data[cat].columns  
  
Index(['Alamat', 'JENIS_SMA', 'PRODI', 'THN_MASUK', 'JENKEL', 'KELAS',  
       'THN_LULUS'],  
      dtype='object')
```

EDA

(Exploratory Data Analysis)

Proses melakukan uji investigasi awal pada sebuah data untuk mengidentifikasi pola, menemukan anomali, menguji hipotesis, dan memeriksa asumsi melalui statistik ringkasan dan representasi grafis (visual).



```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 863 entries, 0 to 862
Data columns (total 35 columns):
#   Column                Non-Null Count  Dtype  
---  --
0   Alamat                863 non-null   object 
1   JENIS_SMA             851 non-null   object 
2   PRODI                 863 non-null   object 
3   THN_MASUK             863 non-null   int64  
4   JENKEL               863 non-null   object 
5   KELAS                 863 non-null   object 
6   THN_LULUS            556 non-null   float64 
7   IPS_SMT1             883 non-null   float64 
8   IPS_SMT2             788 non-null   float64 
9   IPS_SMT3             709 non-null   float64 
10  IPS_SMT4             665 non-null   float64 
11  IPS_SMT5             604 non-null   float64 
12  IPS_SMT6             594 non-null   float64 
13  IPS_SMT7             556 non-null   float64 
14  IPS_SMT8             509 non-null   float64 
15  IPS_SMT9             186 non-null   float64 
16  IPS_SMT10            89 non-null    float64 
17  IPS_SMT11            46 non-null    float64 
18  IPS_SMT12            25 non-null    float64 
19  IPS_SMT13            22 non-null    float64 
20  IPS_SMT14            17 non-null    float64 
21  SKS_AMBIL_SMT1       863 non-null   int64  
22  SKS_AMBIL_SMT2       863 non-null   int64  
23  SKS_AMBIL_SMT3       863 non-null   int64  
24  SKS_AMBIL_SMT4       863 non-null   int64  
25  SKS_AMBIL_SMT5       863 non-null   int64  
26  SKS_AMBIL_SMT6       863 non-null   int64  
27  SKS_AMBIL_SMT7       863 non-null   int64  
28  SKS_AMBIL_SMT8       863 non-null   int64  
29  SKS_AMBIL_SMT9       863 non-null   int64  
30  SKS_AMBIL_SMT10      863 non-null   int64  
31  SKS_AMBIL_SMT11      863 non-null   int64  
32  SKS_AMBIL_SMT12      863 non-null   int64  
33  SKS_AMBIL_SMT13      863 non-null   int64  
34  SKS_AMBIL_SMT14      863 non-null   int64  

dtypes: float64(15), int64(15), object(5)
memory usage: 236.1+ KB
```

```
data['KELAS'].value_counts()
```

```
Reguler      644
Karyawan     219
Name: KELAS, dtype: int64
```

```
data['PRODI'].value_counts()
```

```
Teknik Informatika    454
Sistem Informasi     358
Sistem Komputer       22
Manajemen Informatika  19
Komputerisasi Akuntansi 10
Name: PRODI, dtype: int64
```

```
data['THN_MASUK'].value_counts()
```

```
2016      814
2017       49
Name: THN_MASUK, dtype: int64
```

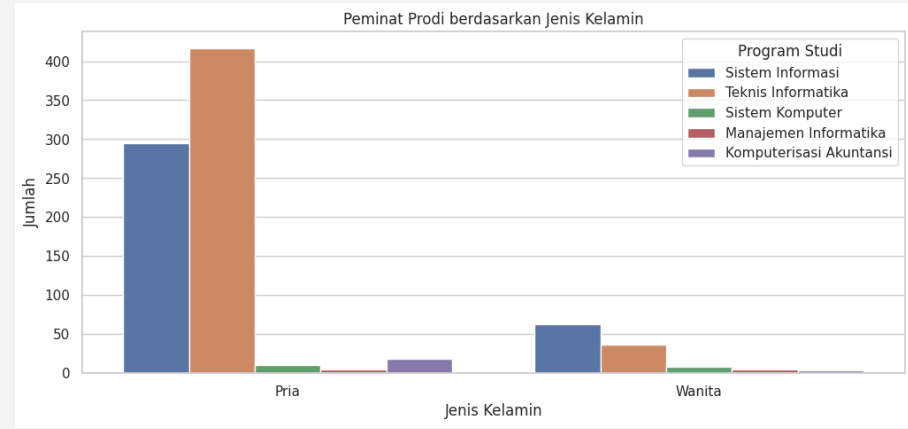
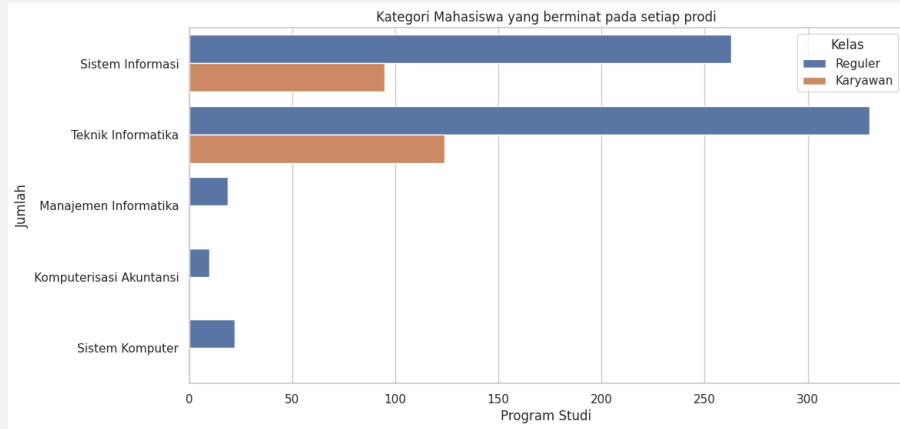
```
data['THN_LULUS'].value_counts()
```

```
2020      650
2021      122
2019       27
2022       25
2023       25
2018        7
2017        7
Name: THN_LULUS, dtype: int64
```

```
data['KELAS'].value_counts()
```

```
Reguler      644
Karyawan     219
Name: KELAS, dtype: int64
```

Multivariate Analysis



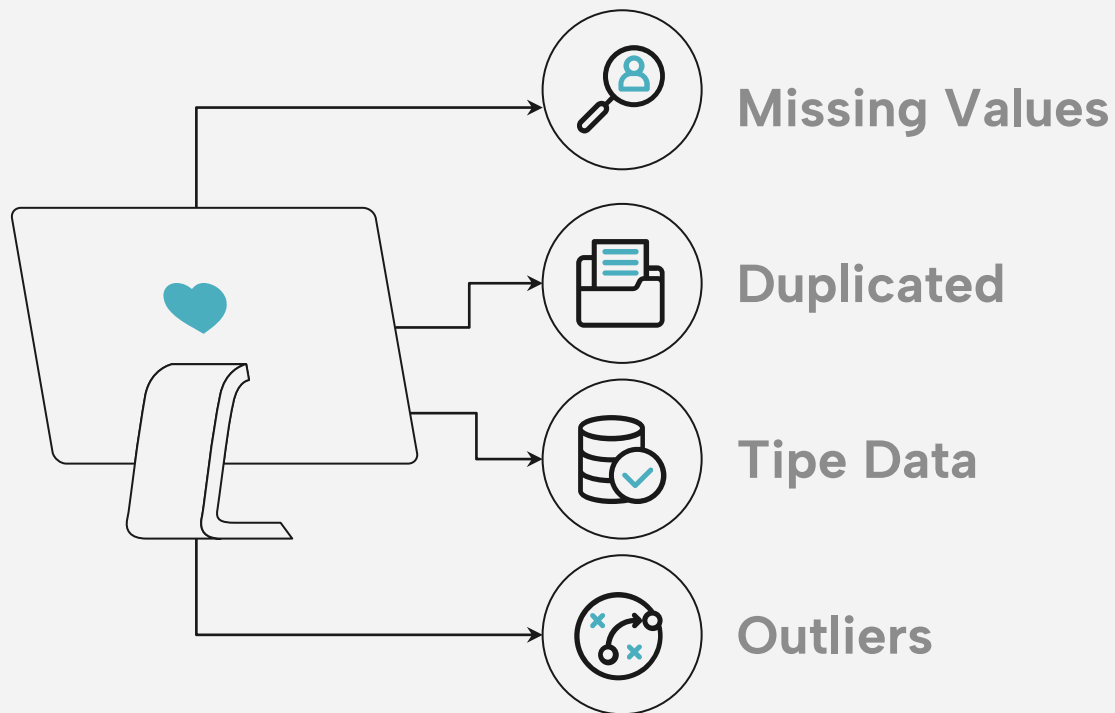
Preparation

Proses memperbaiki atau menghapus kesalahan, ketidakkonsistenan, dan ketidakakuratan dalam kumpulan data





What is problem in Dataset?



Missing Values

▶ data.isna().sum()

Alamat	0
JENIS_SMA	12
PRODI	0
THN_MASUK	0
JENKEL	0
KELAS	0
THN_LULUS	307
IPS_SMT1	60
IPS_SMT2	75
IPS_SMT3	154
IPS_SMT4	198
IPS_SMT5	259
IPS_SMT6	269
IPS_SMT7	307
IPS_SMT8	354
IPS_SMT9	677
IPS_SMT10	774
IPS_SMT11	817
IPS_SMT12	838
IPS_SMT13	841
IPS_SMT14	846

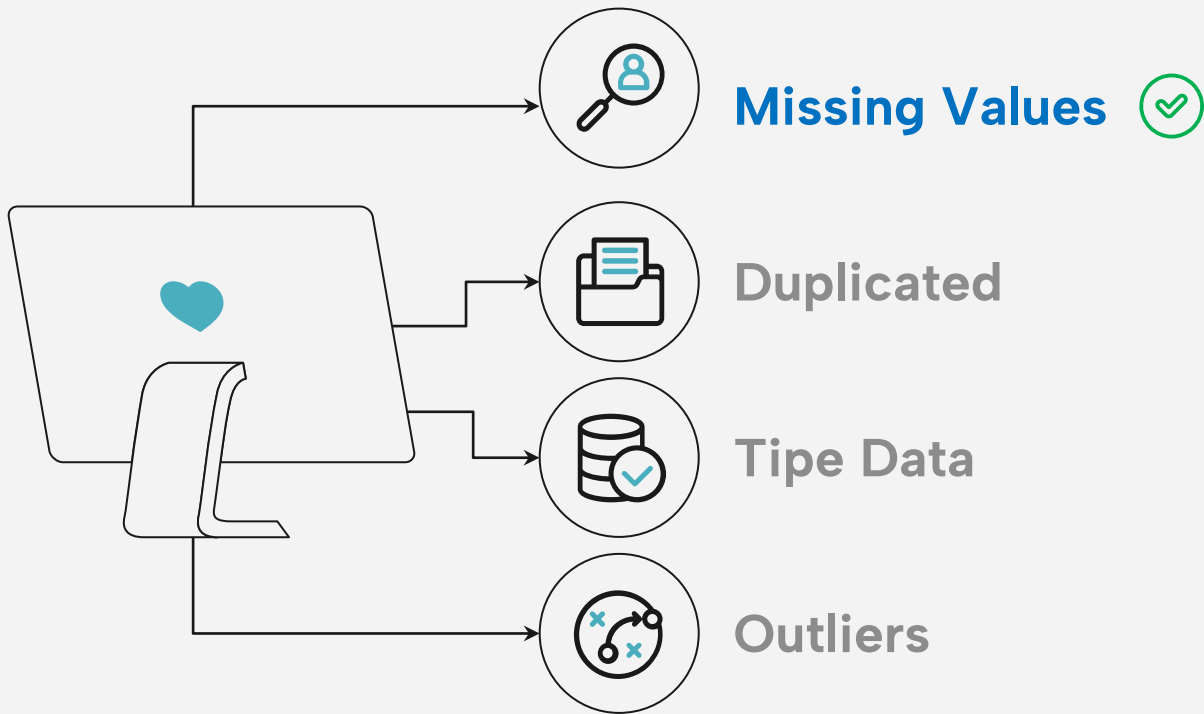


▶ data.isna().sum()

Alamat	0
JENIS_SMA	0
PRODI	0
THN_MASUK	0
JENKEL	0
KELAS	0
THN_LULUS	0
IPS_SMT1	0
IPS_SMT2	0
IPS_SMT3	0
IPS_SMT4	0
IPS_SMT5	0
IPS_SMT6	0
IPS_SMT7	0
IPS_SMT8	0
IPS_SMT9	0
IPS_SMT10	0
IPS_SMT11	0
IPS_SMT12	0
IPS_SMT13	0
IPS_SMT14	0

Before

After





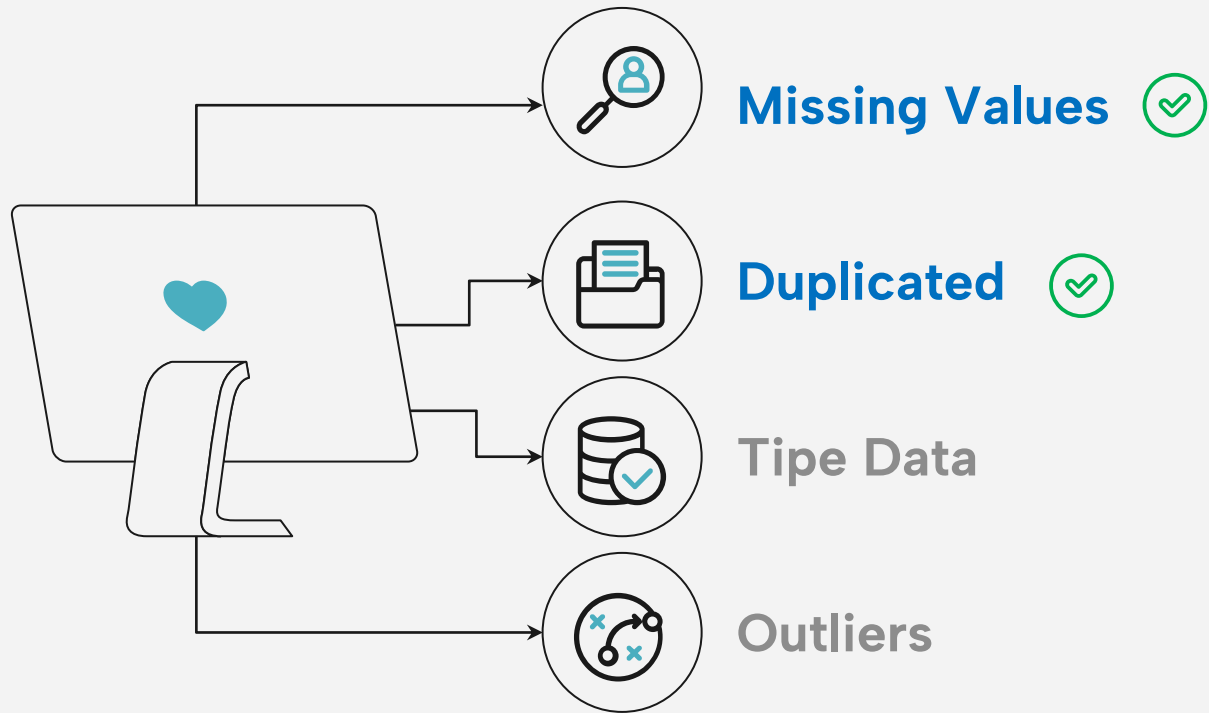
Duplicated



```
▶ data.duplicated().sum()
```

```
⇒ 0
```





Tipe Data

```
data_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 863 entries, 0 to 862  
Data columns (total 35 columns):  
#   Column      Non-Null Count  Dtype  
---  ---  
0   Alamat      863 non-null    object  
1   JENIS_SMA    863 non-null    object  
2   PRODI        863 non-null    object  
3   THN_MASUK    863 non-null    int64  
4   JENKEL       863 non-null    object  
5   KELAS        863 non-null    object
```

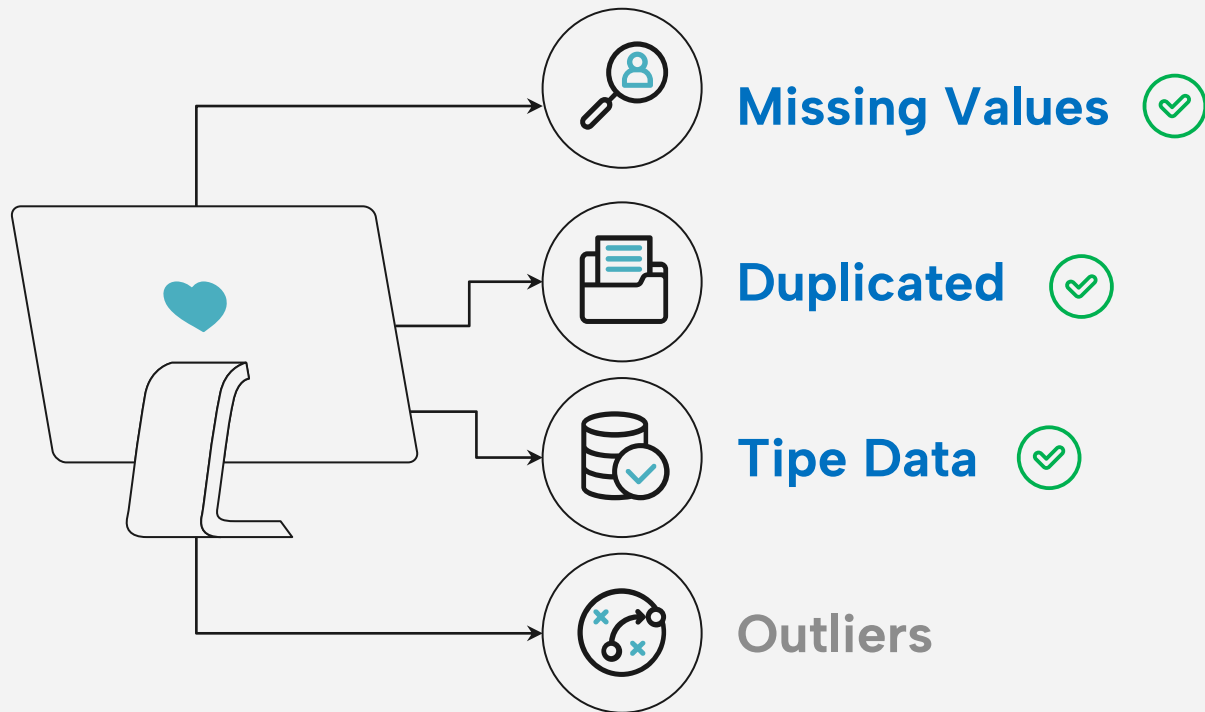
Before



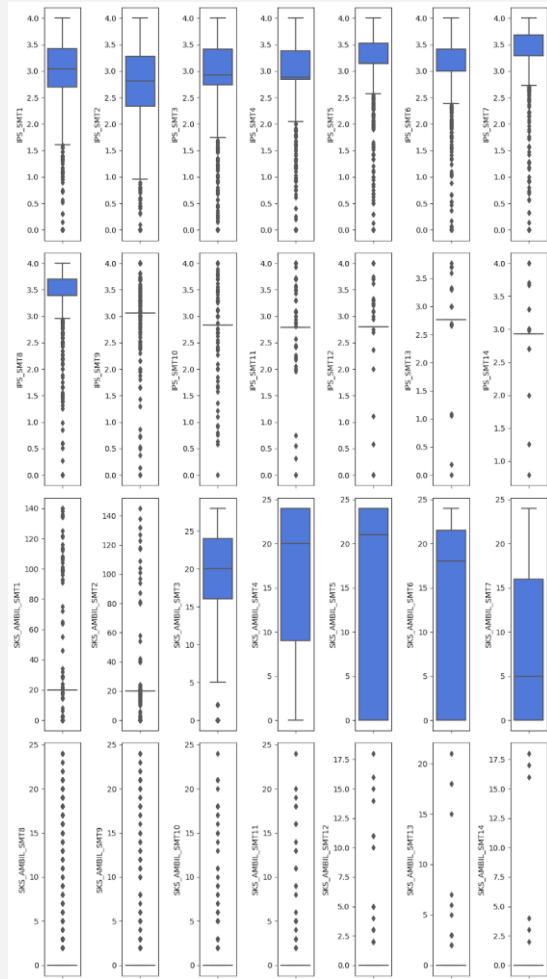
```
data_clean.info()
```

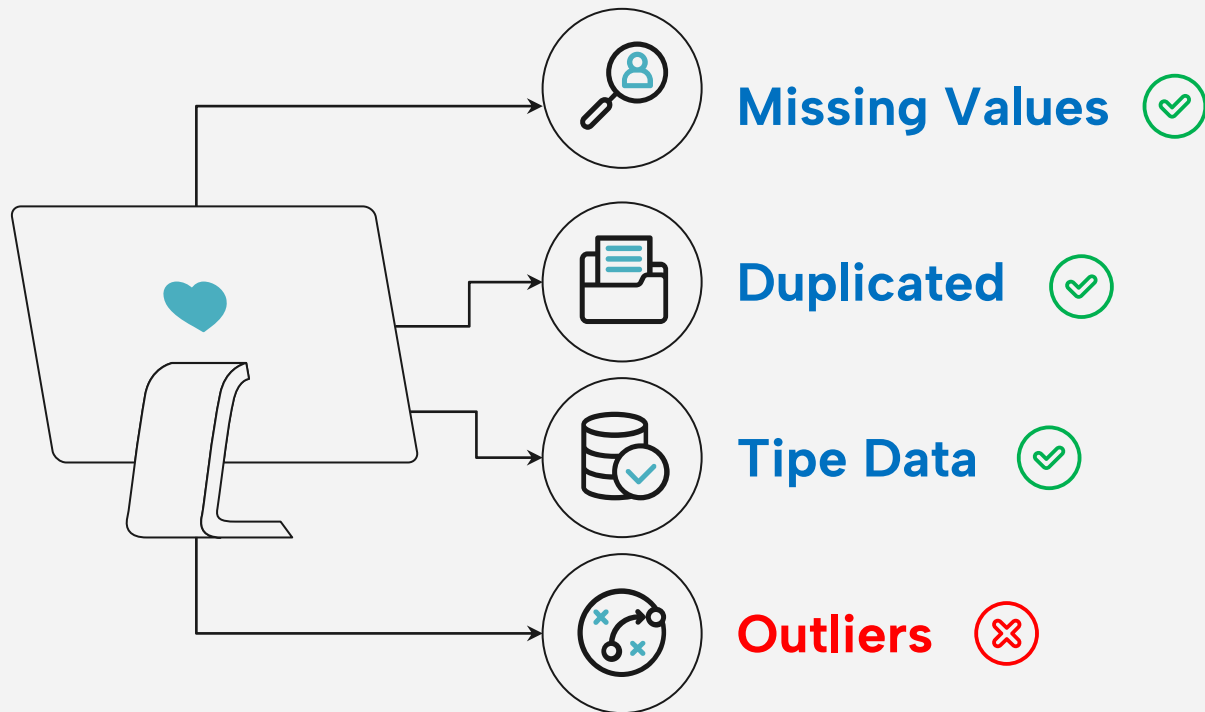
```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 863 entries, 0 to 862  
Data columns (total 35 columns):  
#   Column      Non-Null Count  Dtype  
---  ---  
0   Alamat      863 non-null    object  
1   JENIS_SMA    863 non-null    object  
2   PRODI        863 non-null    object  
3   THN_MASUK    863 non-null    datetime64[ns]  
4   JENKEL       863 non-null    object  
5   KELAS        863 non-null    object  
6   THN_LULUS    863 non-null    datetime64[ns]
```

After



Outliers







Feature Encoding

```
label_encoder = LabelEncoder()

data_encoding['JENIS_SMA'] = label_encoder.fit_transform(data_encoding['JENIS_SMA'])
data_encoding['PRODI'] = label_encoder.fit_transform(data_encoding['PRODI'])
data_encoding['JENKEL'] = label_encoder.fit_transform(data_encoding['JENKEL'])
data_encoding['KELAS'] = label_encoder.fit_transform(data_encoding['KELAS'])
```

	Alamat	JENIS_SMA	PRODI	THN_MASUK	JENKEL	KELAS	THN_LULUS	IPS_SMT1	IPS_SMT2	IPS_SMT3	...	SKS_AMBIL_SMT5
0	Kel. Larangan Kec. Ciledug - Tangerang	4	2	2016-01- 01	0	1	2020-01- 01	3.11	3.080000	3.130000	...	23
1	Kp. Bintaro, Pesanggrahan	4	4	2016-01- 01	0	1	2022-01- 01	3.51	2.260000	2.910000	...	21
2	Poris Plawad Indah Cipondoh Tangerang Tangera...	2	2	2016-01- 01	0	1	2020-01- 01	0.00	2.641992	2.811509	...	0
3	Pesanggrahan, Jakarta Selatan	2	1	2016-01- 01	0	1	2019-01- 01	3.27	3.330000	3.430000	...	20
4	Pondok Aren, Tangerang Selatan	2	2	2016-01- 01	0	1	2022-01- 01	2.82	3.170000	2.220000	...	20





Feature Selection

```
cat = data_new.select_dtypes(include=['object']).columns.tolist()
num = data_new.select_dtypes(include=['int','float64']).columns.tolist()

data_new[num] = data_new[num].round(2)

batas_waktu = pd.to_timedelta(4 * 365, unit='D')
data_new['LULUS_TEPAT_WAKTU'] = (data_new['THN_MASUK'] - data_new['THN_LULUS']) <= batas_waktu
```

SKS_AMBIL_SMT9	SKS_AMBIL_SMT10	SKS_AMBIL_SMT11	SKS_AMBIL_SMT12	SKS_AMBIL_SMT13	SKS_AMBIL_SMT14	LULUS_TEPAT_WAKTU
0	0	0	0	0	0	True
0	0	0	0	0	0	True
0	0	0	0	0	0	True
0	0	0	0	0	0	True
2	0	0	0	0	0	True





Feature Transformation

```
from sklearn.preprocessing import MinMaxScaler

feature = ['PRODI', 'JENKEL', 'KELAS', 'IPS_SMT1', 'IPS_SMT2', 'IPS_SMT3', 'IPS_SMT4',
           'IPS_SMT9', 'IPS_SMT10', 'IPS_SMT11', 'IPS_SMT12', 'IPS_SMT13', 'IPS_SMT14',
           'SKS_AMBIL_SMT4', 'SKS_AMBIL_SMT5', 'SKS_AMBIL_SMT6', 'SKS_AMBIL_SMT7', 'SKS_AMBIL_SMT8',
           'SKS_AMBIL_SMT11', 'SKS_AMBIL_SMT12', 'SKS_AMBIL_SMT13', 'SKS_AMBIL_SMT14']

scaler = MinMaxScaler()
data_scaler[feature] = scaler.fit_transform(data_scaler[feature])
```

IPS_SMT1	IPS_SMT2	IPS_SMT3	IPS_SMT4	IPS_SMT5	...	SKS_AMBIL_SMT6	SKS_AMBIL_SMT7	SKS_AMBIL_SMT8	SKS_AMBIL_SMT9
0.7775	0.7700	0.7825	0.8375	0.8175	...	0.708333	0.708333	0.125000	0.000000
0.8775	0.5650	0.7275	0.6600	0.7175	...	0.750000	0.875000	0.291667	0.000000
0.0000	0.6600	0.7025	0.7200	0.7850	...	0.000000	0.000000	0.000000	0.000000
0.8175	0.8325	0.8575	0.8725	0.8650	...	0.250000	0.000000	0.000000	0.000000
0.7050	0.7925	0.5550	0.7225	0.8750	...	0.875000	0.833333	0.625000	0.083333





Split Data

```
# pemisahan features vs target
X = data_scaler[feature]
y = data_scaler['LULUS_TEPAT_WAKTU'].values
print(X.shape)
print(y.shape)
```

```
(863, 31)
(863,)
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
print(X_train.shape)
print(X_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(690, 31)
(173, 31)
(690,)
(173,)
```





Modeling

Pada tahap ini kami melatih model (Training model), mengatur dan menyesuaikan parameter-parameter model, memvalidasi model, mengembangkan dan menguji model, memilih algoritma yang tepat, dan optimasi model.



Random Forest

```
rf_model = RandomForestClassifier()  
rf_model.fit(X_train, y_train)
```

▼ RandomForestClassifier

RandomForestClassifier()

```
eval_classification(rf_model)
```

Accuracy (Test Set): 1.00
Precision (Test Set): 1.00
Recall (Test Set): 1.00
F1-Score (Test Set): 1.00

KNN

```
knn_model = KNeighborsClassifier(n_neighbors=5)  
knn_model.fit(X_train, y_train)
```

▼ KNeighborsClassifier

KNeighborsClassifier()

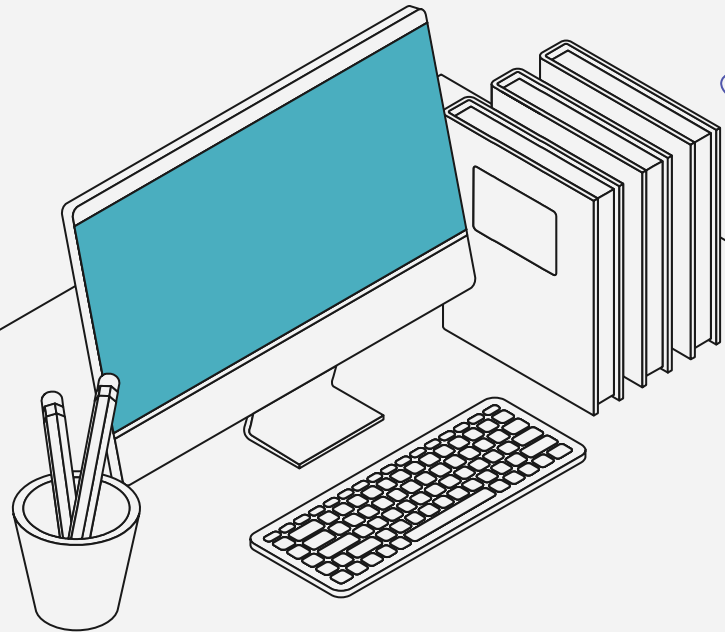
```
eval_classification(knn_model)
```

Accuracy (Test Set): 1.00
Precision (Test Set): 1.00
Recall (Test Set): 1.00
F1-Score (Test Set): 1.00



Evaluasi

Karena dari hasil model menentukan nilai yang sempurna (Accuracy, Precision, Recall, dan F1-Score semua sebesar 1.00). sehingga dapat disimpulkan bahwa model tersebut sangat baik dalam melakukan klasifikasi.



Rekomendasi

Rekomendasi yang dapat kami sampaikan untuk mencegah terjadinya mahasiswa tidak lulus tepat waktu :

- **Pengembangan Kebijakan Kampus**

Dengan mempertimbangkan hasil model, kampus dapat mengembangkan kebijakan atau strategi untuk meningkatkan tingkat kelulusan tepat waktu. Ini bisa termasuk program bimbingan akademis, monitoring proaktif terhadap mahasiswa yang berisiko, atau peningkatan akses ke sumber daya pendukung.

- **Analisis Karakteristik Mahasiswa**

Identifikasi faktor-faktor yang mempengaruhi kinerja mahasiswa dan kelulusan tepat waktu. Ini mungkin melibatkan analisis lebih lanjut terhadap atribut mahasiswa seperti kelas-kelas yang diambil, partisipasi dalam kegiatan ekstrakurikuler, atau faktor-faktor personal.

- **Pemantauan Berkelanjutan**

Terus memantau dan mengevaluasi keefektifan kebijakan dan strategi yang diimplementasikan. Dengan pemantauan yang berkelanjutan, kampus dapat menyesuaikan pendekatan mereka sesuai dengan perubahan kondisi dan kebutuhan mahasiswa.

