



# Tim **ODE** (Oh Data Euy)

## Institut Teknologi Bandung

Felix Fernando  
Gerend Christopher  
Jeremy





## ■ Table of Content

Business Understanding

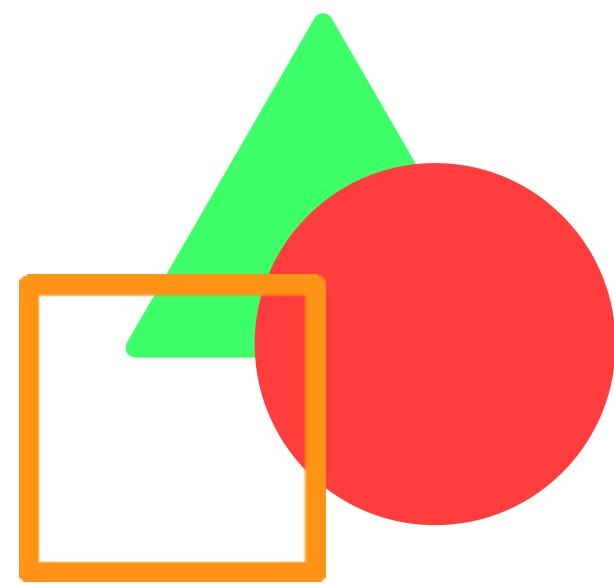
Data Understanding

Data Preparation

Exploratory Data Analysis (EDA)

Modeling and Evaluation

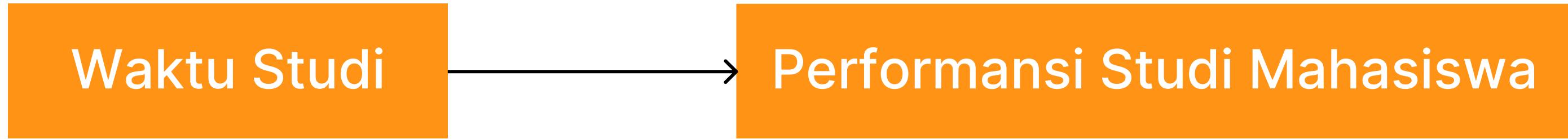
Conclusion / Suggestion





# Business Understanding

Penyelesaian studi tepat waktu → indikator prestasi belajar mahasiswa  
Penegasan masa studi dalam Permen. Menristek Pasal 16 No. 44 Tahun 2015



Tujuan: menentukan **model klasifikasi** yang **tepat** untuk dapat melakukan **prediksi kelulusan** yang **baik**

*Data mining:* data nilai mahasiswa → pola kelulusan mahasiswa yang menyelesaikan studi tepat waktu

Teknik *data mining* yang digunakan adalah **Decision Tree**

Indikator keberhasilan analisis:

Akurasi model > 80%



# Data Understanding

## ■ Dataframe 1

Data Transkrip Mahasiswa

	NIM	ANGKATAN	SEMESTER	KODE_MK	NAMA_MK	SKS	NILAI	GRADE
0	10110310002	2010	1011	EM100	EM100 Dasar-dasar Bisnis	3	57.0	C
1	10110310002	2010	1011	EM180	EM180 Matematika Bisnis	3	70.0	B
2	10110310002	2010	1011	TI100	TI100 Algoritma dan Pemrograman	4	57.0	C
3	10110310002	2010	1011	TI101	TI101 Matematika Diskrit	3	59.0	C
4	10110310002	2010	1011	TI110	TI110 Pengantar Teknologi Multimedia	3	74.0	B

Cuplikan Data **Dataframe 1**

Berdasarkan **dataframe** ini, syarat bahwa seorang mahasiswa lulus adalah memenuhi syarat sebagai berikut,

- 1 Sudah menyelesaikan **144** SKS
- 2 Tidak terdapat nilai **D**, **E**, dan **F** pada setiap mata kuliah
- 3 IPK diatas **2.5**

## Atribut :

- 1 NIM / Nomor Induk Mahasiswa - Nominal
- 2 Angkatan Tahun Mahasiswa - Ordinal
- 3 Periode Semester - Ordinal
- 4 Nama dan Kode Mata Kuliah - Nominal Kategorikal
- 5 Jumlah SKS - Numerik
- 6 Nilai - Numerik
- 7 Indeks - Ordinal



# Data Understanding

## ■ Dataframe 2

Data Evaluasi Dosen per Mata Kuliah

Atribut :

- 1 Tahun - Ordinal
- 2 Semester - Ordinal
- 3 Mata Kuliah - Nominal Kategorikal
- 4 Pertanyaan - Numerik
- 5 Keterangan - Nominal
- 6 Nilai - Numerik

TAHUN	SEMESTER	MATAKULIAH	PERTANYAAN	KETERANGAN	NILAI
0	2015	1511	IS100 Management Information Systems	1	Kesiapan memberikan perkuliahan/praktikum
1	2015	1511	IS100 Management Information Systems	2	Upaya menyampaikan materi perkuliahan/praktik...
2	2015	1511	IS100 Management Information Systems	3	Sistematis dalam menyampaikan materi perkuliahan...
3	2015	1511	IS100 Management Information Systems	4	Kemampuan memberikan contoh yang relevan dari ...
4	2015	1511	IS100 Management Information Systems	5	Penyampaian materi perkuliahan sesuai dengan k...

Cuplikan Data **Dataframe 2**



# Data Preparation

## Pengecekan Value Null

### Dataframe 1

Data Transkrip Mahasiswa

#	Column	Non-Null Count	Dtype
0	NIM	30870	non-null
1	ANGKATAN	30870	non-null
2	SEMESTER	30870	non-null
3	KODE_MK	30870	non-null
4	NAMA_MK	30847	non-null
5	SKS	30870	non-null
6	NILAI	30858	non-null
7	GRADE	30318	non-null

### Dataframe 2

Data Evaluasi Dosen per Mata Kuliah

#	Column	Non-Null Count	Dtype
0	TAHUN	2114	non-null
1	SEMESTER	2114	non-null
2	MATAKULIAH	2114	non-null
3	PERTANYAAN	2114	non-null
4	KETERANGAN	2114	non-null
5	NILAI	2114	non-null

Data tidak menunjukkan adanya data **null**

Jumlah baris data adalah **30870**  
Kolom menunjukkan adanya data **null**



# Data Preparation

## Pengecekan Value Null

### ■ Dataframe 1 - Kolom **NAMA\_MK**

Data Transkrip Mahasiswa

```
1 # check Kode_MK dari Nama_MK yang null
2 df1['KODE_MK'][df1['NAMA_MK'].isnull()].unique()

array(['SI863'], dtype=object)
```

Data **null** untuk kolom **NAMA\_MK**  
hanya pada mata kuliah dengan  
**KODE\_MK: SI863**

NIM	ANGKATAN	SEMESTER	KODE_MK	NAMA_MK	SKS	NILAI	GRADE
442	10110310011	2010	1411	SI863	<NA>	6	70.0
983	10110310022	2010	1411	SI863	<NA>	6	79.0
1528	10110310036	2010	1411	SI863	<NA>	6	56.0
1626	10110310037	2010	1411	SI863	<NA>	6	72.0
1919	10110310044	2010	1321	SI863	<NA>	6	100.0

Data **null** diisi dengan  
“SI863 Tugas Akhir”

NIM	ANGKATAN	SEMESTER	KODE_MK	NAMA_MK	SKS	NILAI	GRADE
46	10110310002	2010	1321	SI863	SI863 Tugas Akhir	6	93.0
147	10110310005	2010	1321	SI863	SI863 Tugas Akhir	6	93.0
196	10110310006	2010	1321	SI863	SI863 Tugas Akhir	6	92.0
245	10110310007	2010	1321	SI863	SI863 Tugas Akhir	6	71.0
344	10110310009	2010	1311	SI863	SI863 Tugas Akhir	6	86.0



# Data Preparation

## Pengecekan Value Null

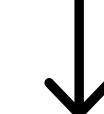
### Dataframe 1 - Kolom NILAI

Data Transkrip Mahasiswa

```
1 # check kolom NILAI yang null  
2 df1["GRADE"][df1['NILAI'].isnull()].unique()  
✓ 0.1s
```

```
array(['F'], dtype=object)
```

Data **null** untuk kolom **NILAI**  
hanya pada data dengan **GRADE**: **F**



Data null diisi dengan integer 0

### Dataframe 1 - Kolom GRADE

Data Transkrip Mahasiswa

	NIM	ANGKATAN	SEMESTER	KODE_MK	NAMA_MK	SKS	NILAI	GRADE
10	10110310002		2010	1021	SI214	SI214 Manajemen Operasional	3	71.0
16	10110310002		2010	1111	SI301	SI301 Sistem Informasi Manajemen	3	89.0
32	10110310002		2010	1211	SI527	SI527 Enterprise Resource Planning	3	73.0
53	10110310004		2010	1021	EM201	EM201 Dasar-dasar Manajemen	3	58.0
68	10110310004		2010	1121	TI403	TI403 Sistem Basis Data	4	72.0

Data **null** untuk kolom **GRADE**, diisi berdasarkan kriteria nilai yang ada.

A 85-100      A- 85-100      B+ 75-79.99

B 70-74.99      B- 65-69.99      C+ 60-64.99

C 55-59.99      D 45-54.99      E 0-44.99



# Data Preparation

## Dropping Data Duplicate

```
1 # check nim yang duplicate mk
2 def checkDuplicate():
3     nim_mhs = df1['NIM'].unique()
4     nim_for_doubled = []
5
6     for nim in nim_mhs:
7         dummy_df1 = df1.loc[df1['NIM'] == nim]
8         if dummy_df1['NAMA_MK'].duplicated().any():
9             nim_for_doubled.append(nim)
10
11     return nim_for_doubled
12
13 checkDuplicate()
✓ 3.2s
['10110310011',
 '10110310022',
 '10110310036',
 '10110310037',
 '10110310044',
 '10110310045',
```

Pada **Dataframe 1** (Data Transkrip Mahasiswa), dropping data nilai duplicate pada kolom **NIM** dan **KODE\_MK**

```
3 df1 = df1.drop_duplicates(subset=['KODE_MK', 'NIM'], keep='last')
```

Dropping dilakukan dengan menyisakan data **terakhir**.



# Data Preparation

## Dropping Data Duplicate

```
1 # check nim yang duplicate mk
2 def checkDuplicate():
3     nim_mhs = df1['NIM'].unique()
4     nim_for_doubled = []
5
6     for nim in nim_mhs:
7         dummy_df1 = df1.loc[df1['NIM'] == nim]
8         if dummy_df1['NAMA_MK'].duplicated().any():
9             nim_for_doubled.append(nim)
10
11     return nim_for_doubled
12
13 checkDuplicate()
✓ 3.2s
['10110310011',
 '10110310022',
 '10110310036',
 '10110310037',
 '10110310044',
 '10110310045',
```

Pada **Dataframe 1** (Data Transkrip Mahasiswa), dilakukan dropping data nilai duplicate pada kolom **NIM** dan **KODE\_MK**

```
3 df1 = df1.drop_duplicates(subset=['KODE_MK', 'NIM'], keep='last')
```

Dropping dilakukan dengan menyisakan data **terakhir**.

# Data Preparation

## Penambahan Kolom Baru

### Dataframe 3

Data Transkrip Mahasiswa per NIM

	NIM	ANGKATAN	TOTAL_SKS	FAILED_GRADE	TOTAL_BOBOT	WAKTU_KULIAH	IPK	TOTAL_FAILED_GRADE	HASIL
0	00000008429	2015	145	D	474.5	3.5	3.272414	1.0	Tidak Lulus
1	00000008455	2015	136	DD	416.2	3.5	3.060294	2.0	Tidak Lulus
2	00000008481	2015	127	DDDEED	313.6	3.5	2.469291	7.0	Tidak Lulus
3	00000008631	2015	145		519.3	3.0	3.581379	0.0	Lulus Tepat Waktu
4	00000008684	2015	145		512.1	3.0	3.531724	0.0	Lulus Tepat Waktu

Penambahan kolom hasil modifikasi Dataframe 1

**TOTAL\_SKS, FAILED\_GRADE, TOTAL\_BOBOT, WAKTU\_KULIAH, TOTAL\_FAILED\_GRADE, dan HASIL**



# Data Preparation

## Penambahan Kolom Baru

### Dataframe 4

Data Evaluasi Dosen per Semester dan Mata Kuliah

TAHUN		NAMA_MK	SEMESTER	KODE_MK	NILAI_RATA-RATA	LULUS	TIDAK_LULUS	Nilai per Pertanyaan														PROPORSI_LULUS
								1	2	3	...	6	7	8	9	10	11	12	13	14		
0	2015	IS100 Management Information Systems	1511	IS100	3.265000	99.0	4.0	3.28	3.25	3.24	...	3.19	3.18	3.32	3.29	3.24	3.31	3.27	3.26	3.31	3.31	0.961165
1	2015	IS110 Business Mathematics	1511	IS110	3.087143	76.0	18.0	3.19	2.98	3.08	...	3.05	2.91	3.20	3.07	3.07	2.97	3.08	3.19	3.22	3.22	0.808511
2	2015	IS201 Corporate Business Processes	1511	IS201	3.170000	2.0	1.0	3.27	3.18	3.16	...	3.20	3.02	3.20	3.14	3.14	2.98	3.09	3.20	3.33	3.33	0.666667
3	2015	IS201 Corporate Business Processes	1521	IS201	2.958571	61.0	6.0	2.96	2.83	2.93	...	2.96	3.01	2.99	3.00	2.96	2.81	2.93	3.00	3.04	3.04	0.910448
4	2015	IS220 Human and Computer Interaction	1521	IS220	3.295000	96.0	0.0	3.26	3.25	3.25	...	3.30	3.23	3.38	3.30	3.30	3.29	3.35	3.41	3.32	3.32	1.000000

$$\text{PROPORSI\_LULUS} = \frac{\text{LULUS}}{\text{LULUS} + \text{TIDAK\_LULUS}}$$

Penambahan kolom hasil modifikasi Dataframe 1 dan 2

**NILAI\_RATA-RATA, LULUS, TIDAK\_LULUS, Nilai per Pertanyaan, dan PROPORSI\_LULUS**



# Data Preparation

## Penambahan Kolom Baru

### ■ Dataframe 5

Data Kelulusan Mahasiswa per Angkatan

ANGKATAN	JUMLAH KELULUSAN	JUMLAH MAHASISWA	PROPORSI KELULUSAN
0	2010	34	0.010770
1	2011	33	0.009720
2	2012	22	0.010989
3	2013	41	0.010567
4	2014	27	0.006048
5	2015	24	0.005169

Penambahan kolom hasil modifikasi Dataframe 4  
**JUMLAH KELULUSAN**, **JUMLAH MAHASISWA**, dan **PROPORSI KELULUSAN**

# Exploratory Data Analysis

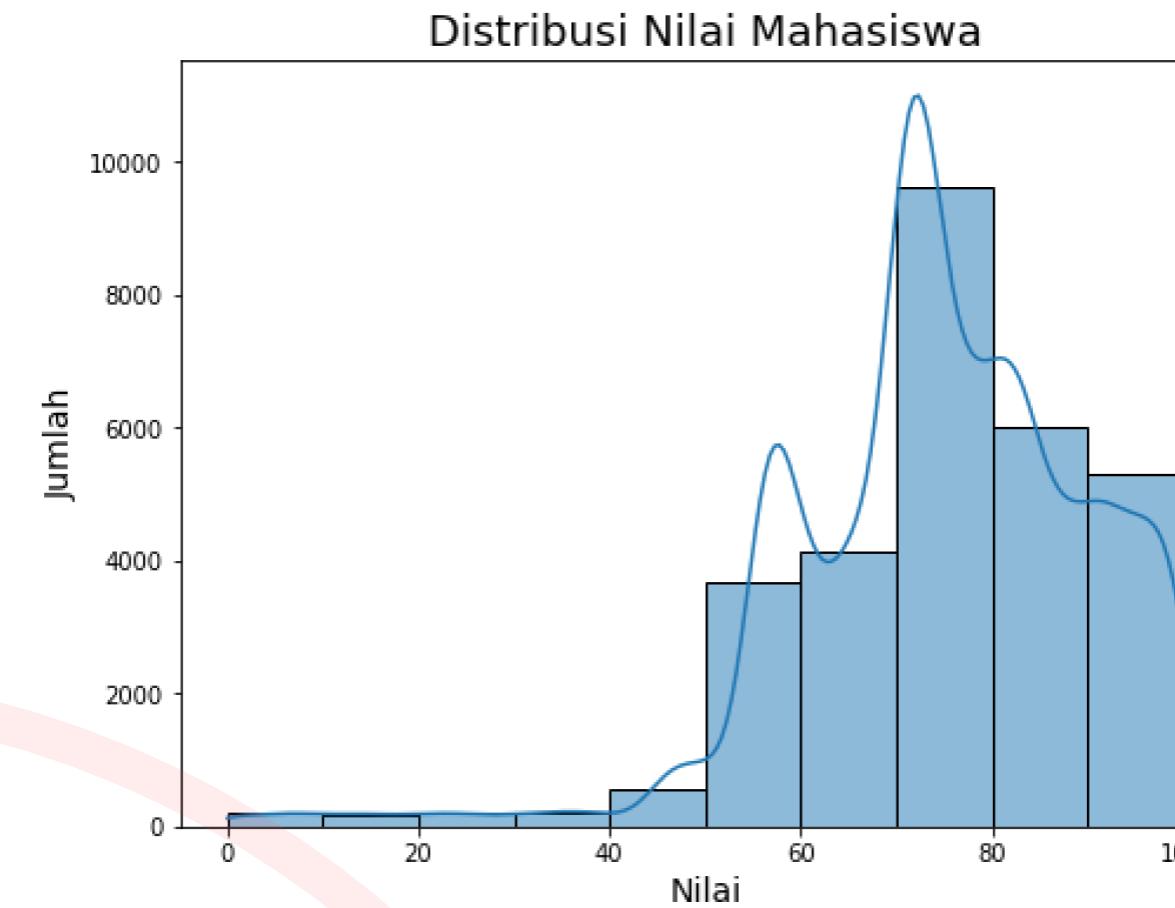


## Statistika Deskriptif & Histogram

### Dataframe 1

Data Transkrip Mahasiswa

NILAI	
count	30085.000000
mean	74.263321
std	15.751127
min	0.000000
25%	66.000000
50%	74.000000
75%	85.000000
max	100.000000



Distribusi nilai mahasiswa  
**menceng ke kiri**  
**(negatively skewed)**



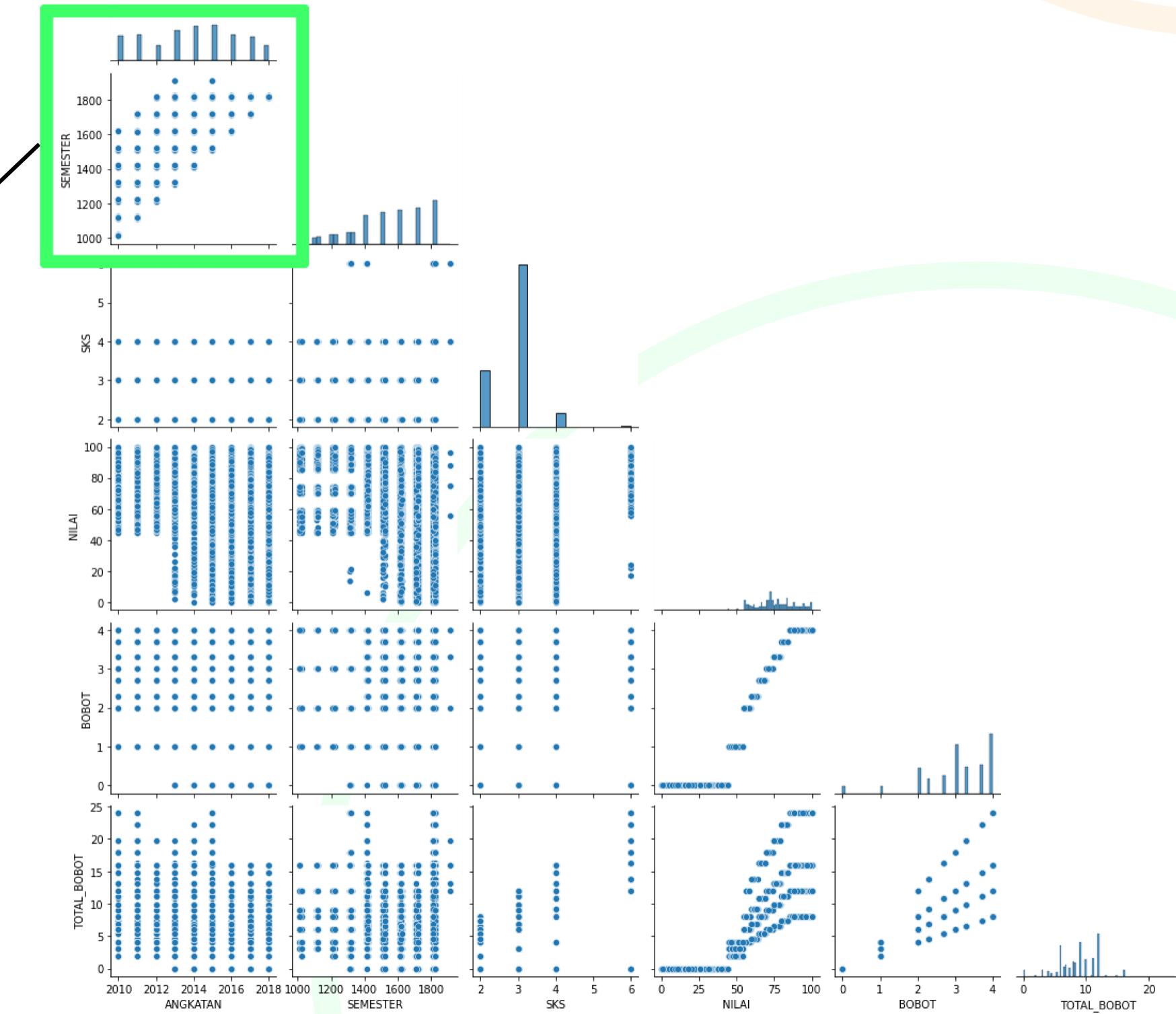
# Exploratory Data Analysis

## Scatter Matrix

### Dataframe 1

Data Transkrip Mahasiswa

**Hubungan linier** hanya terlihat pada atribut **semester dan angkatan**



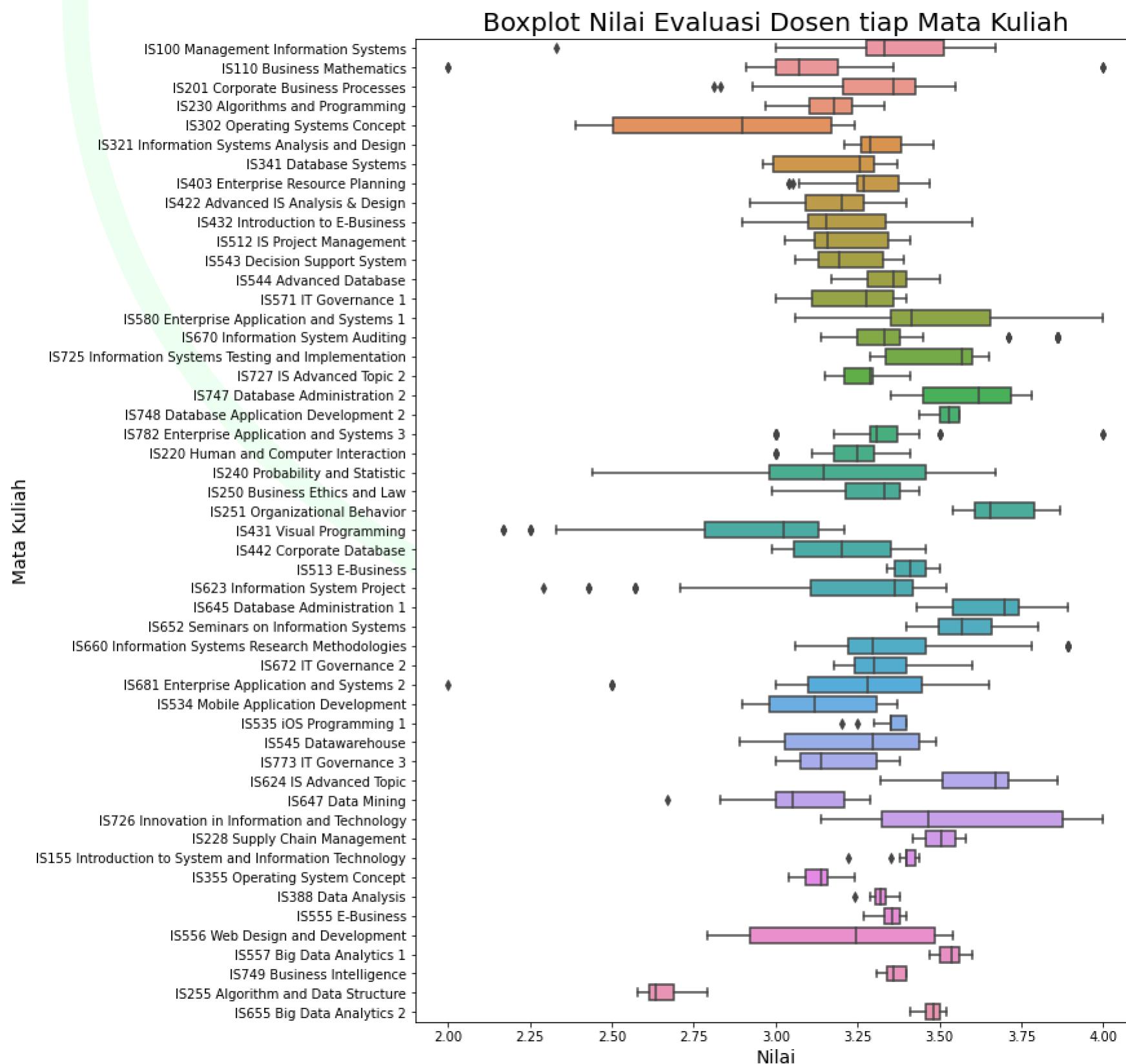
# Exploratory Data Analysis

## Box Plot per Mata Kuliah

### Dataframe 2

Evaluasi Dosen per Mata Kuliah

Hasil evaluasi dosen mata kuliah **IS255 Algorithm and Data Structure** memiliki nilai yang **cenderung kecil** dibandingkan mata kuliah lainnya



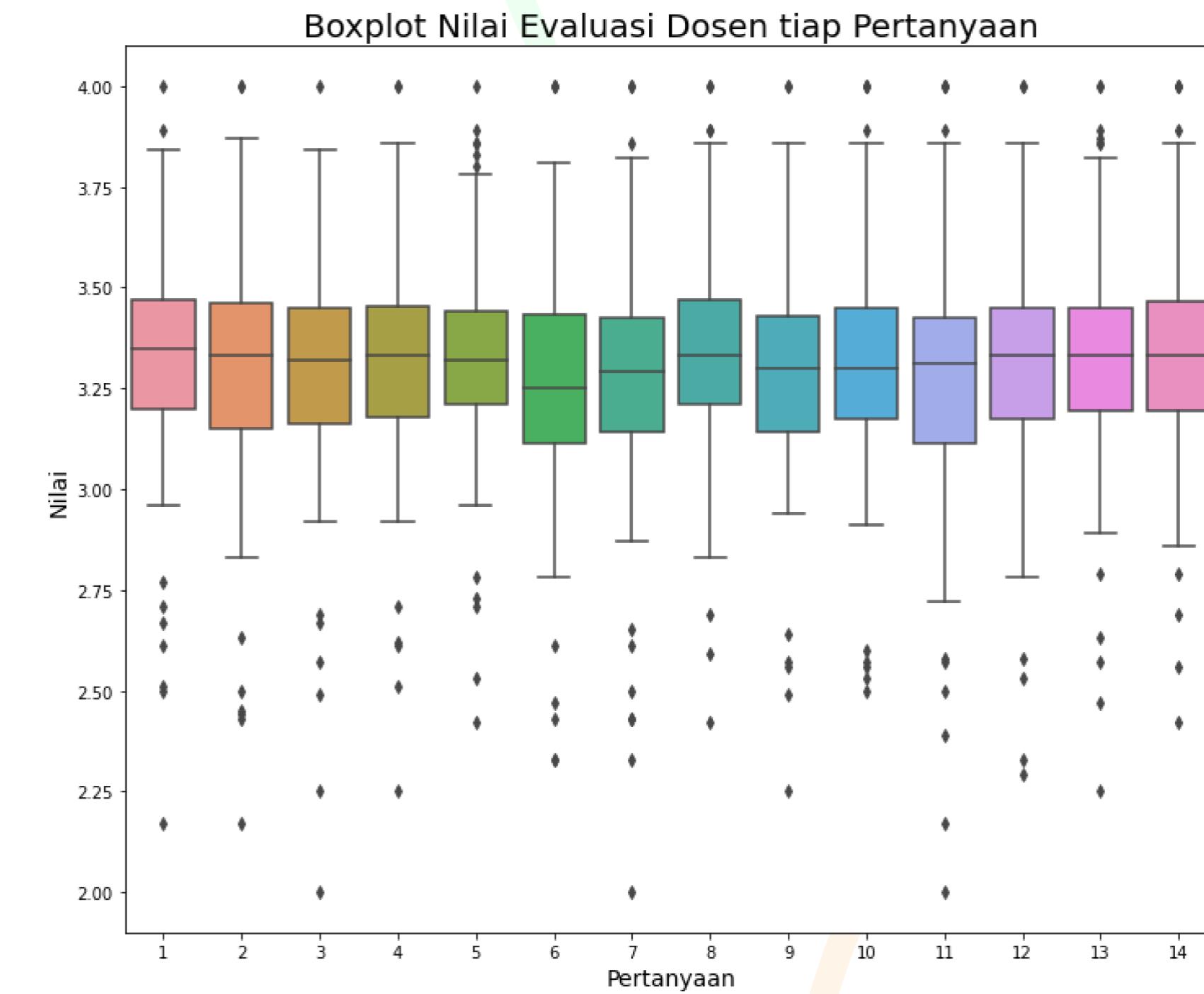
# Exploratory Data Analysis

## Box Plot per Pertanyaan

### Dataframe 2

Evaluasi Dosen per Mata Kuliah

Nilai rata-rata yang diperoleh untuk **setiap pertanyaan** dari seluruh mata kuliah **cenderung sama**





# Exploratory Data Analysis

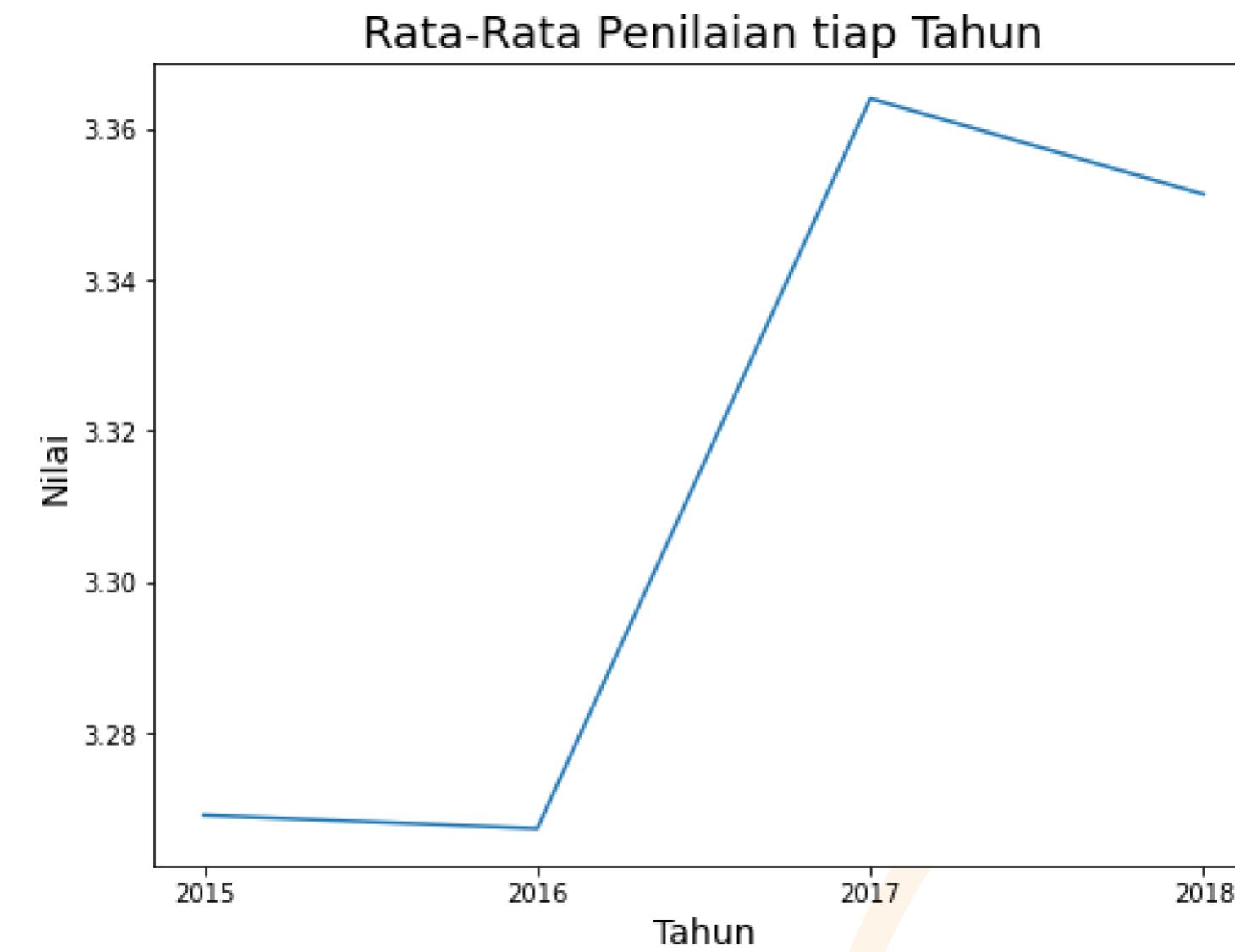
## Rata-rata nilai dosen per tahun

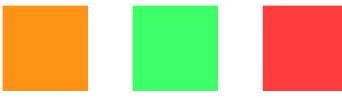
### Dataframe 2

Evaluasi Dosen per Mata Kuliah

Nilai rata-rata yang diperoleh dosen untuk **setiap tahun** dari tahun 2010-2015 untuk seluruh mata kuliah **meningkat sedikit**

TAHUN	NILAI
0	2015 3.269231
1	2016 3.267432
2	2017 3.363956
3	2018 3.351299





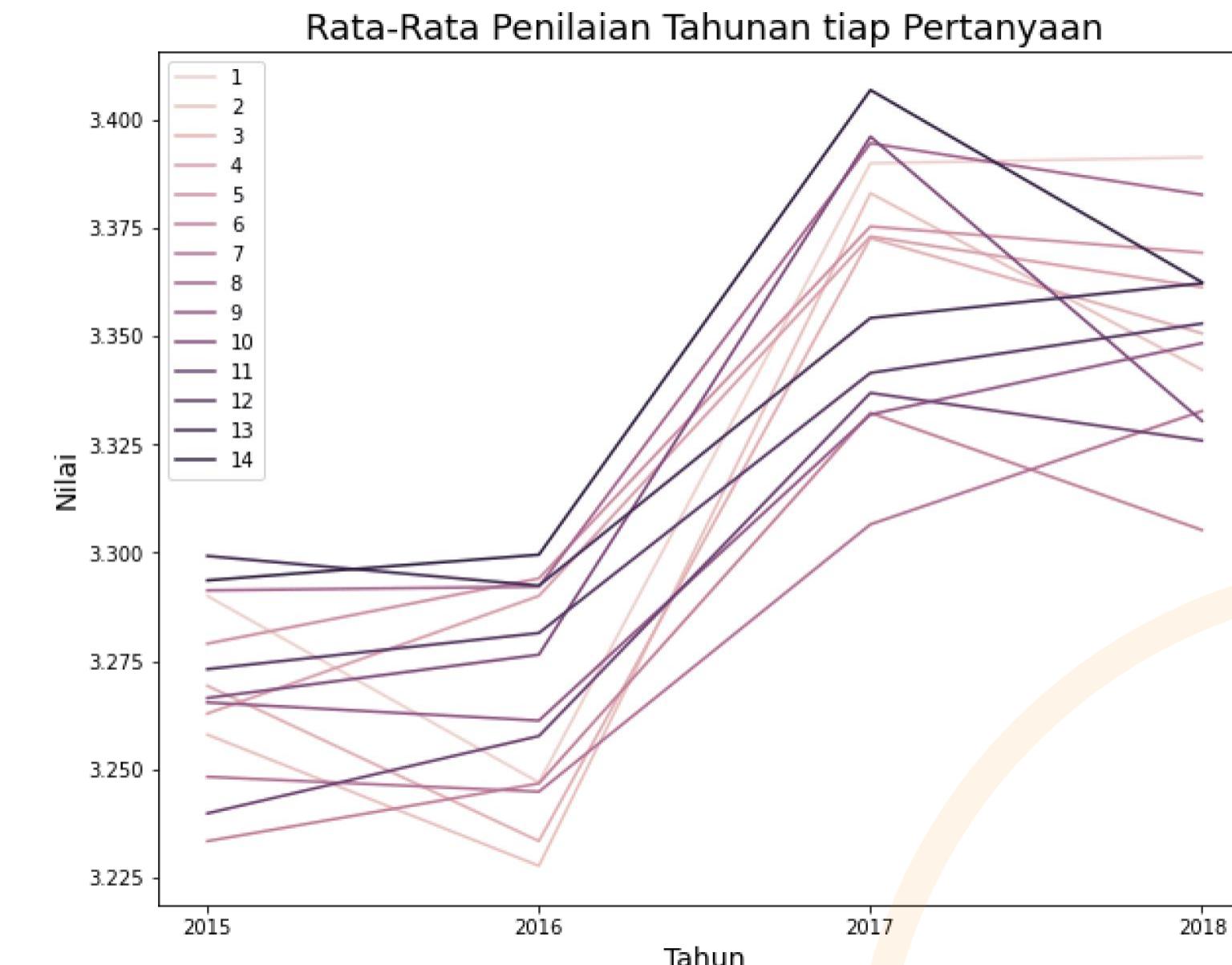
# Exploratory Data Analysis

## Rata-rata nilai dosen per pertanyaan setiap tahun

### Dataframe 2

Evaluasi Dosen per Mata Kuliah

Nilai rata-rata yang diperoleh dosen **per pertanyaan setiap tahun** dari tahun 2015-2018 untuk seluruh mata kuliah **meningkat** → kinerja dosen meningkat





# Exploratory Data Analysis

## Statistika Deskriptif

### ■ Dataframe 3

Data Transkip Nilai Mahasiswa per NIM

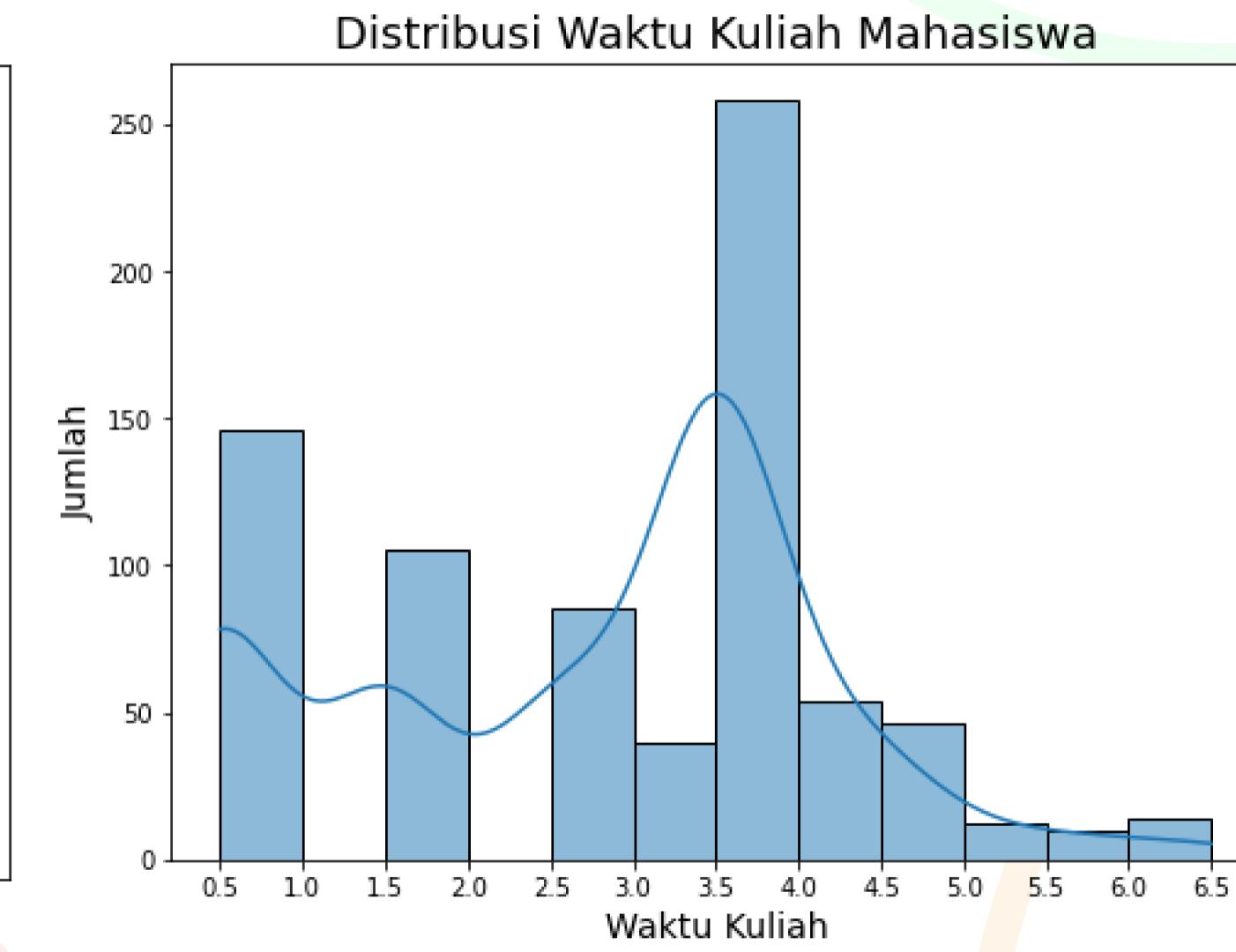
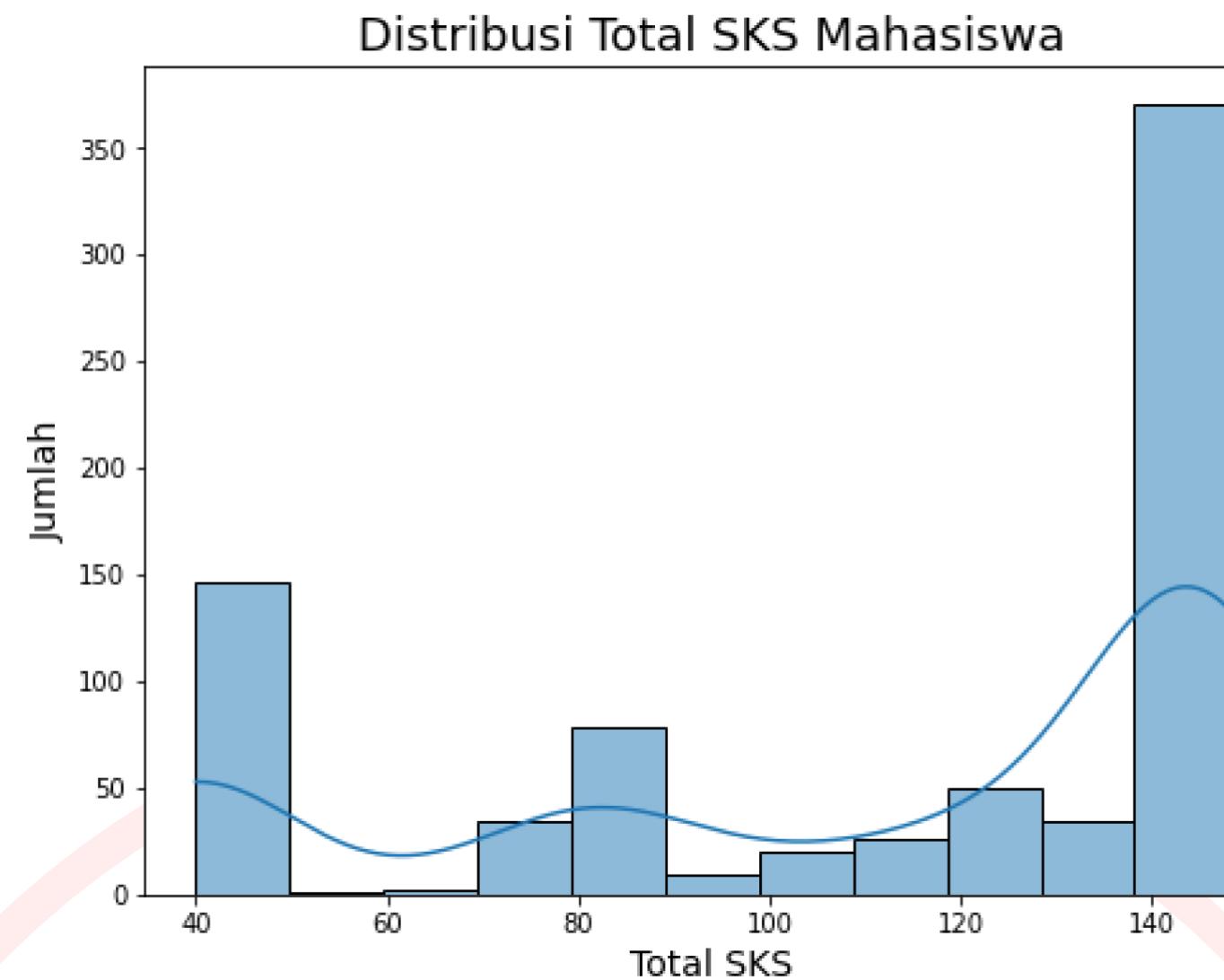
	ANGKATAN	TOTAL_SKS	TOTAL_BOBOT	WAKTU_KULIAH	IPK	TOTAL_FAILED_GRADE
count	770.000000	770.000000	770.000000	770.000000	770.000000	770.000000
mean	2014.707792	110.763636	337.295714	2.716234	2.943970	2.320779
std	2.598202	41.083773	153.354122	1.437582	0.608181	3.409191
min	2010.000000	40.000000	0.000000	0.500000	0.000000	0.000000
25%	2013.000000	82.000000	193.625000	1.500000	2.673277	0.000000
50%	2015.000000	134.000000	395.850000	3.500000	3.034729	1.000000
75%	2017.000000	145.000000	463.875000	3.500000	3.340000	3.000000
max	2018.000000	148.000000	580.000000	6.500000	4.000000	21.000000

# Exploratory Data Analysis

## Distribusi Total SKS dan Waktu Kuliah Mahasiswa

### Dataframe 3

Data Transkip Nilai Mahasiswa per NIM





# Exploratory Data Analysis

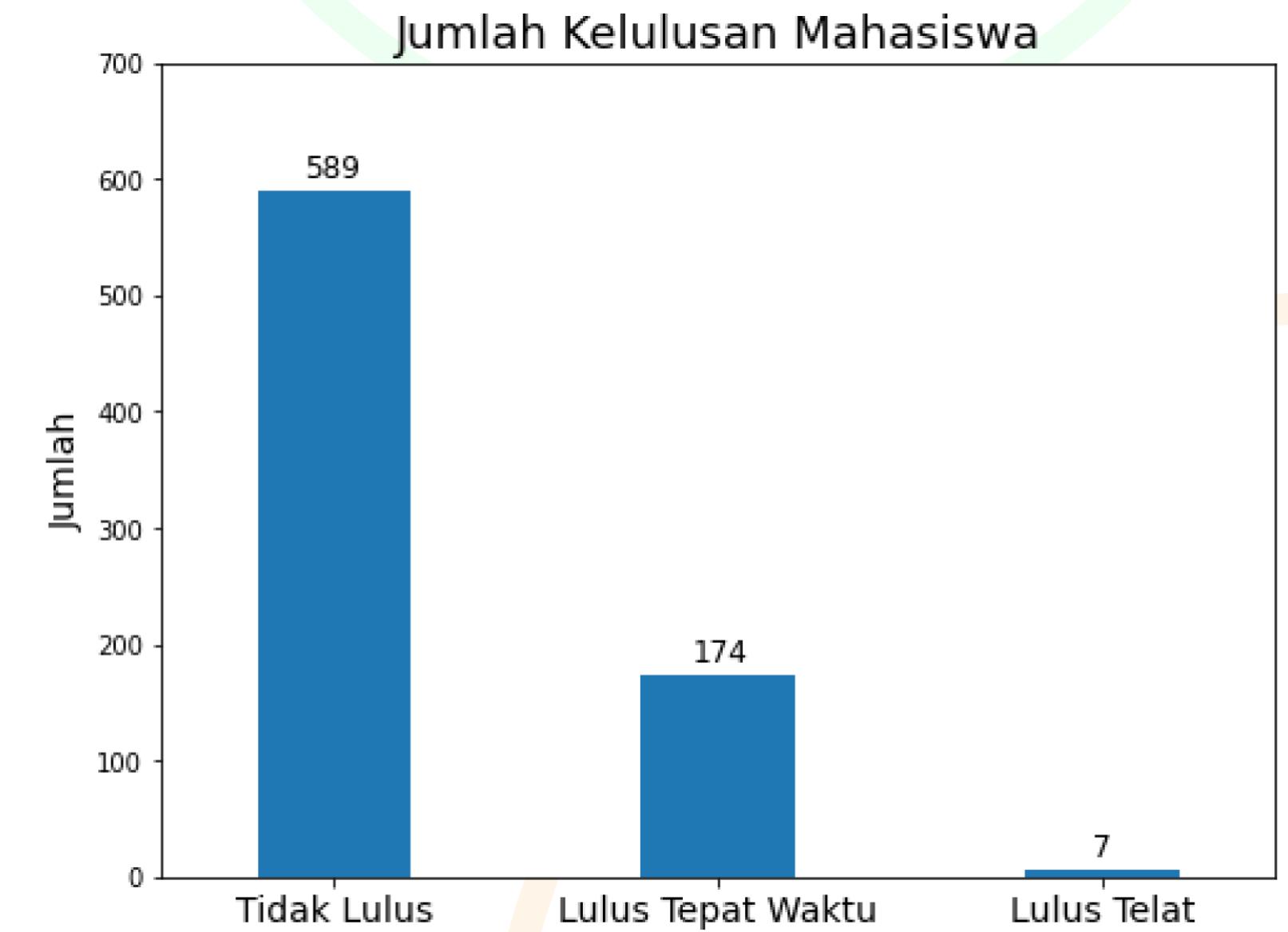
## Bar Chart Kelulusan Mahasiswa

### Dataframe 3

Data Transkip Nilai Mahasiswa per NIM

Tidak Lulus	589
Lulus Tepat Waktu	174
Lulus Telat	7

Mahasiswa yang tidak lulus masih **cukup banyak** relatif terhadap mahasiswa yang telah lulus. Hal ini disebabkan oleh banyak mahasiswa yang masih menempuh studi pada **tingkat satu, dua, dan tiga**. Selain itu, ada pula mahasiswa yang sudah memenuhi 144skt tetapi masih **belum lulus dalam suatu mata kuliah**



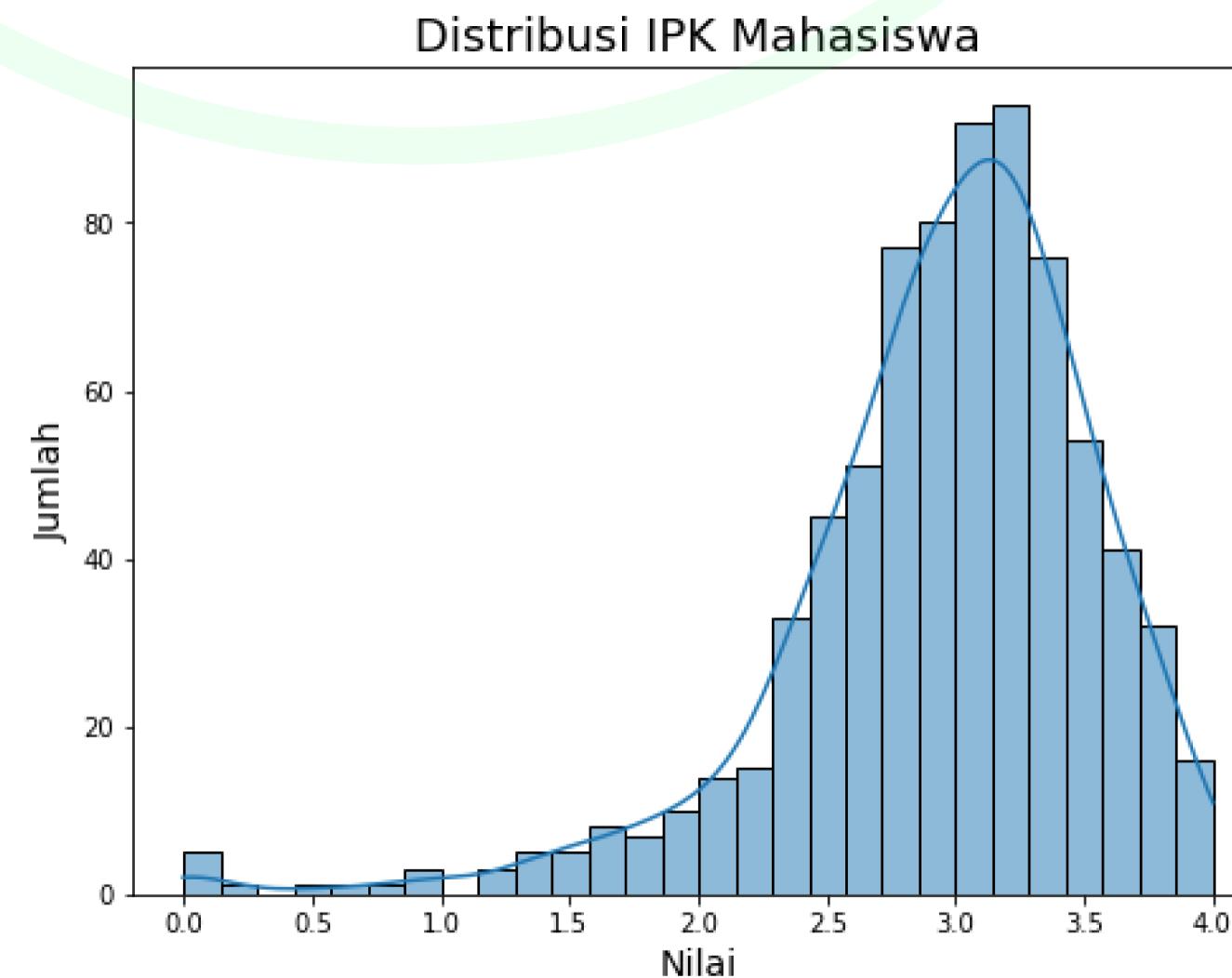


# Exploratory Data Analysis

## Distribusi IPK

### Dataframe 3

Data Transkip Nilai Mahasiswa per NIM



Skewness: -1.5035917059355948  
Kurtosis: 4.178268491050808  
IPK Avg: 2.9439700926765258

Distribusi IPK mahasiswa  
**menceng ke kiri**  
**(negatively skewed)**

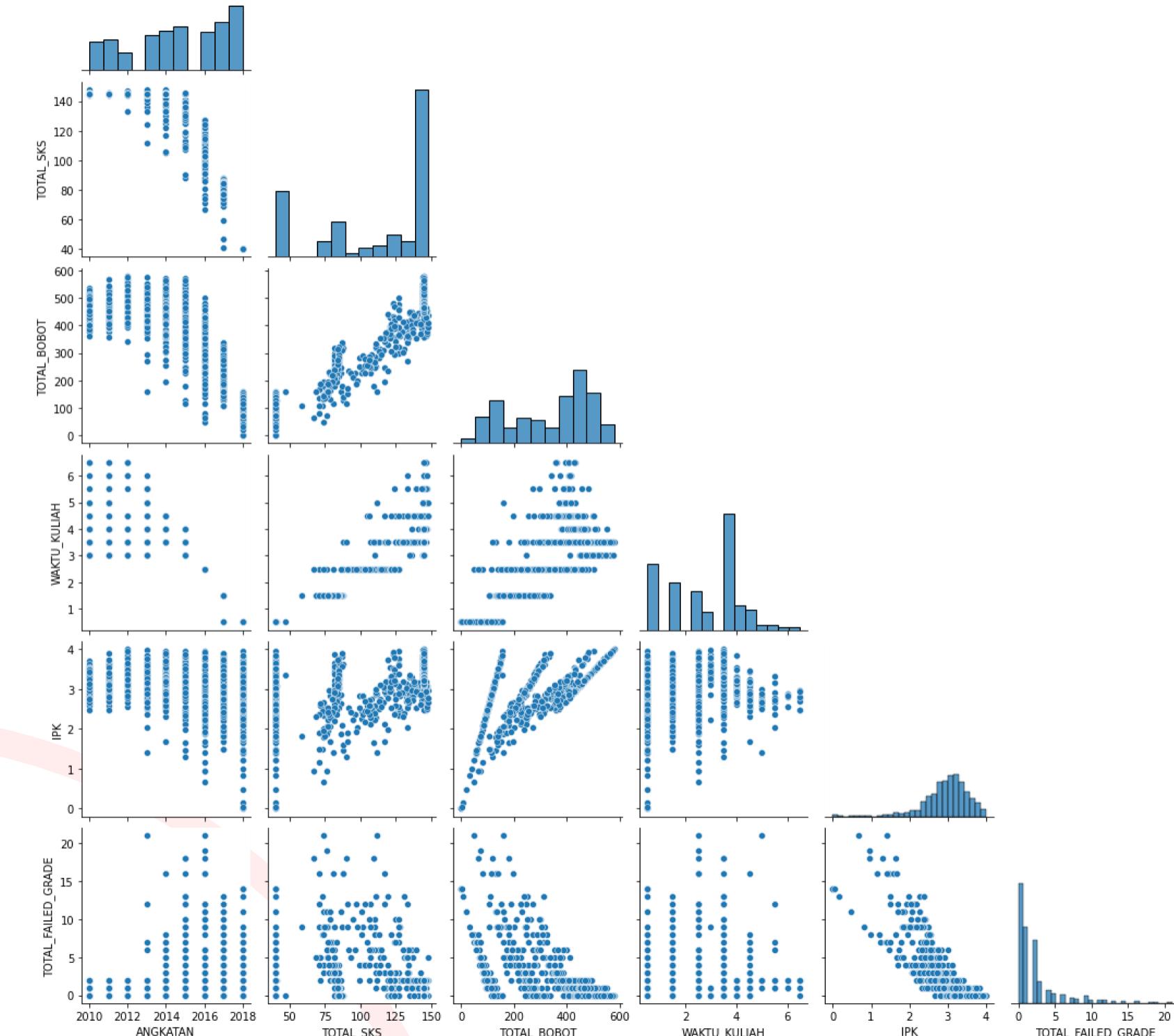
# Exploratory Data Analysis

## Scatter Matrix

### Dataframe 3

Data Transkip Nilai Mahasiswa per NIM

Atribut dataframe 3 **tidak memiliki hubungan linier**



# Exploratory Data Analysis

## Scatter Plot Proporsi Kelulusan Mahasiswa dengan Nilai Rata-Rata Evaluasi Dosen

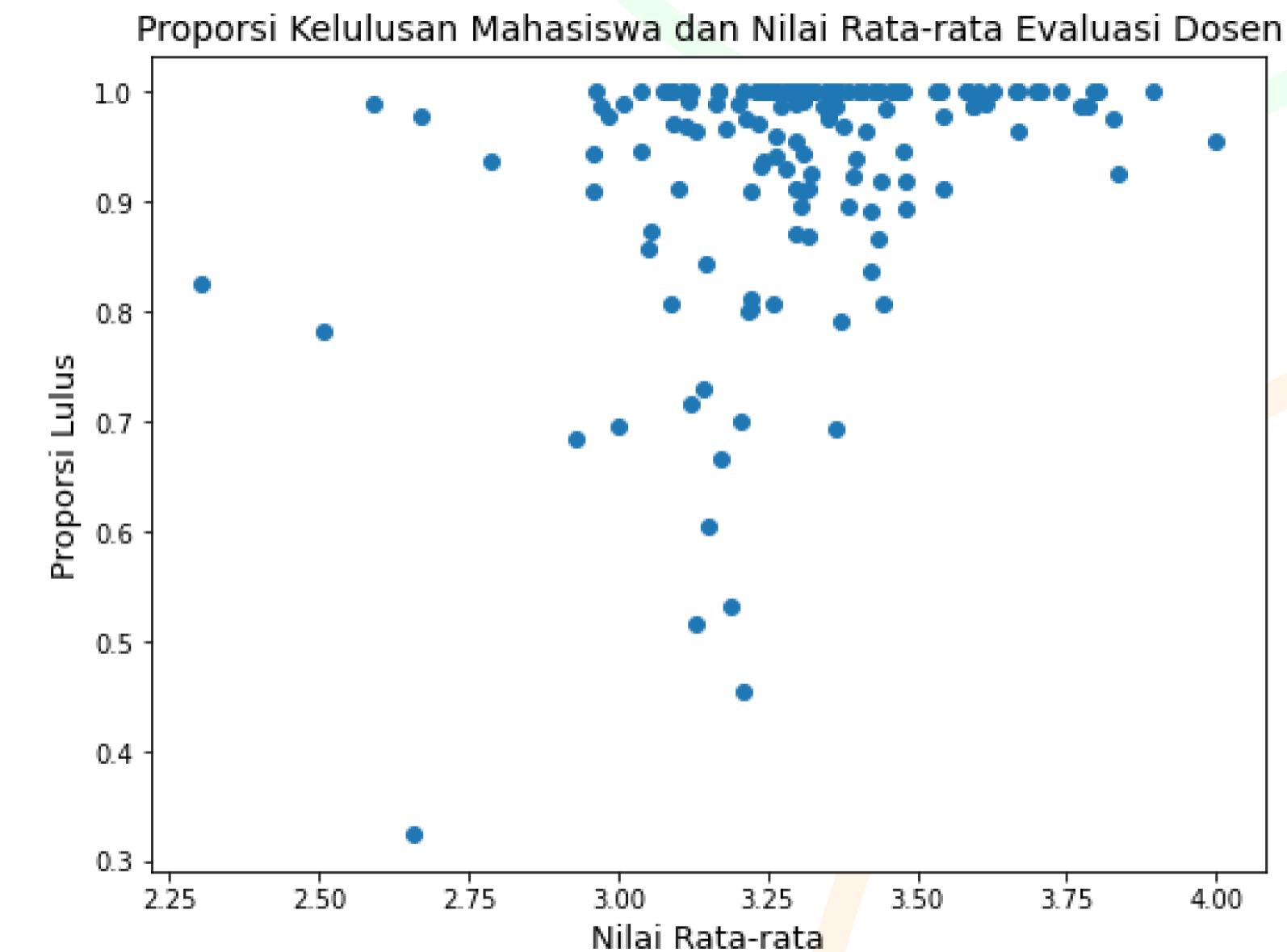
### Dataframe 4

Evaluasi Dosen per Semester dan Mata Kuliah

Terlihat bahwa nilai rata-rata evaluasi dosen **tidak memengaruhi** banyaknya mahasiswa yang lulus karena tidak adanya hubungan linier (korelasi)

#### Catatan:

Proporsi kelulusan = Jumlah mahasiswa yang lulus pada suatu angkatan / Total Mahasiswa Angkatan tersebut.



# Exploratory Data Analysis

## Trendline Kelulusan Mahasiswa

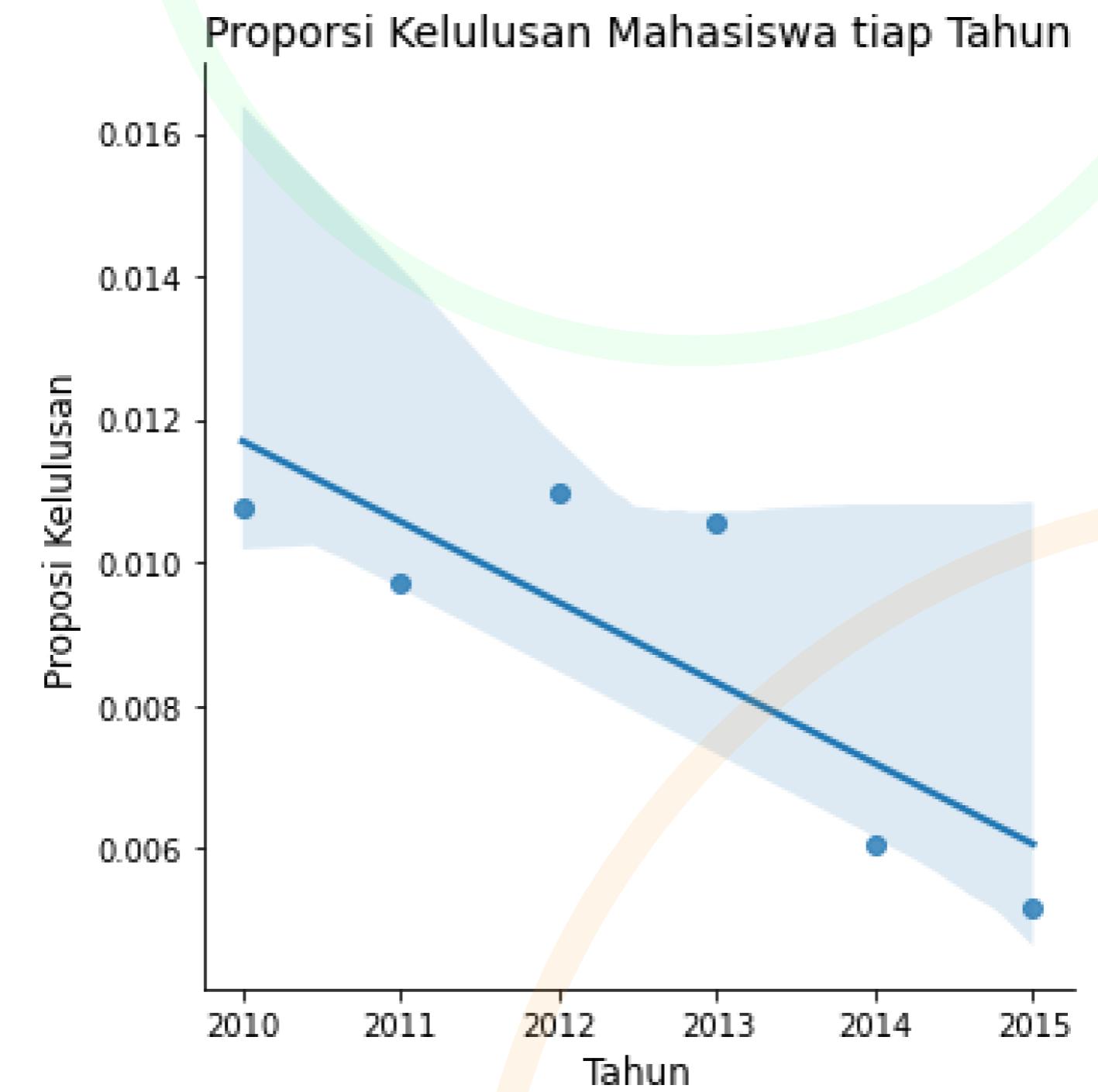
### Dataframe 5

Jumlah Kelulusan Mahasiswa Setiap Angkatan

Terlihat bahwa adanya **tren menurun** untuk **proporsi kelulusan mahasiswa** setiap tahunnya

#### Catatan:

Proporsi kelulusan = Jumlah mahasiswa yang lulus pada suatu angkatan / Total Mahasiswa Angkatan tersebut.



# Modeling and Evaluation

## Pemisahan Training dan Test Dataset

### Dataframe 6

Mahasiswa Angkatan 2010-2015

	NIM	ANGKATAN	TOTAL_SKS	FAILED_GRADE	TOTAL_BOBOT	WAKTU_KULIAH	IPK	TOTAL_FAILED_GRADE	HASIL
0	00000008429	2015	145	D	474.5	3.5	3.272414	1.0	Tidak Lulus
1	00000008455	2015	136	DD	416.2	3.5	3.060294	2.0	Tidak Lulus
2	00000008481	2015	127	DDDEEED	313.6	3.5	2.469291	7.0	Tidak Lulus
3	00000008631	2015	145		519.3	3.0	3.581379	0.0	Lulus Tepat Waktu
4	00000008684	2015	145		512.1	3.0	3.531724	0.0	Lulus Tepat Waktu

Data Transkrip Mahasiswa Angkatan 2010-2015 menjadi **Training dataset**



# Modeling and Evaluation

## Pemisahan Train dan Test Dataset

### ■ Dataframe 7

Mahasiswa Angkatan 2016-2018

	NIM	ANGKATAN	TOTAL_SKS	FAILED_GRADE	TOTAL_BOBOT	WAKTU_KULIAH	IPK	TOTAL_FAILED_GRADE
96	00000012779	2016	127		478.1	2.5	3.764567	0.0
97	00000012880	2016	123		412.7	2.5	3.355285	0.0
98	00000012905	2016	123		476.5	2.5	3.873984	0.0
99	00000012906	2016	127		460.4	2.5	3.625197	0.0
100	00000012910	2016	127		500.3	2.5	3.939370	0.0

Data Transkrip Mahasiswa Angkatan 2016-2018 menjadi **Test dataset**



# Modeling and Evaluation

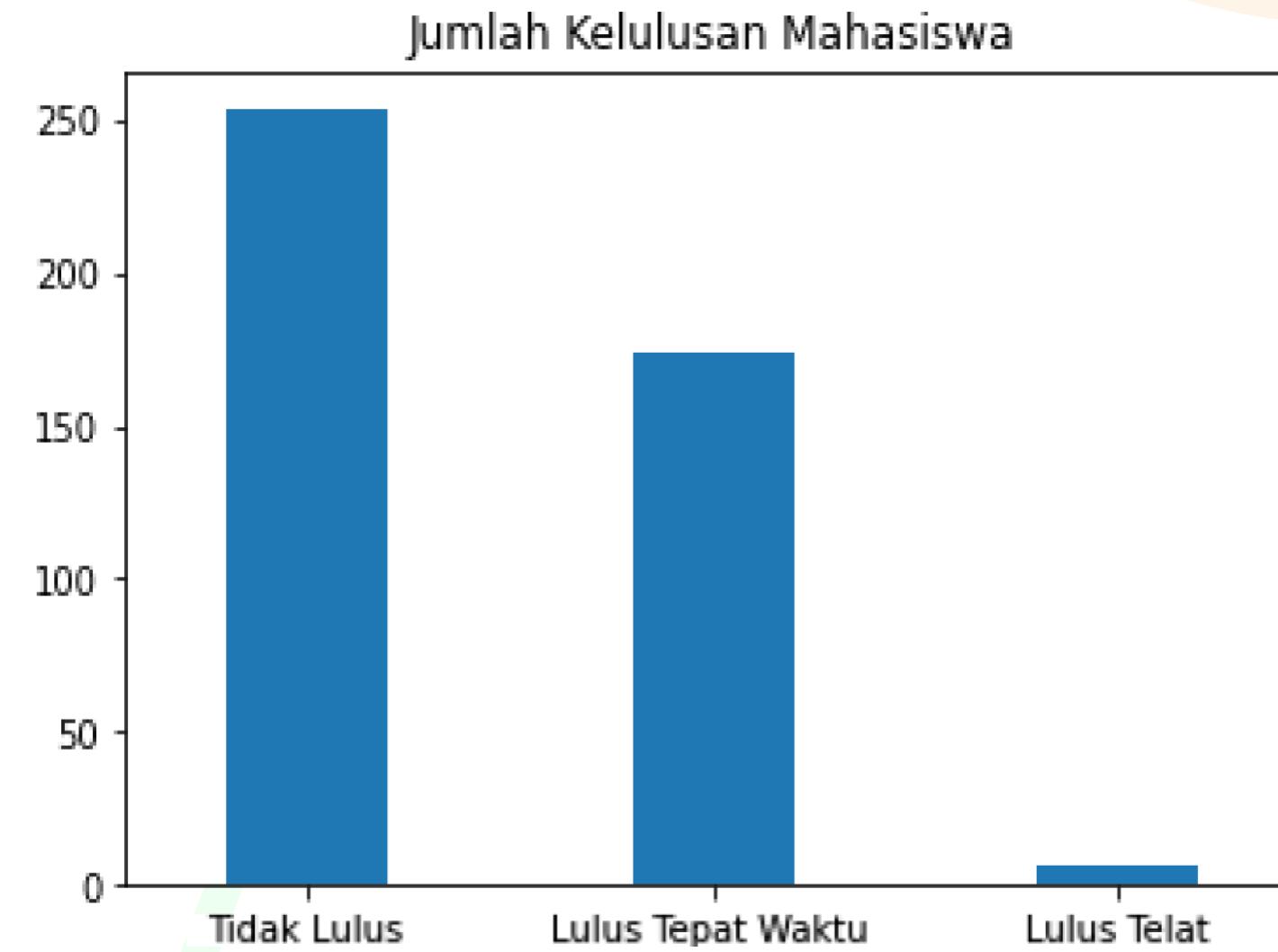
## Sampling Dataset

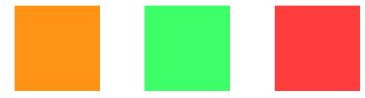
### Dataframe 6

Mahasiswa Angkatan 2010-2015

<b>Tidak Lulus</b>	253
<b>Lulus Telat</b>	174
<b>Lulus Tepat Waktu</b>	7

Pada data transkrip mahasiswa angkatan 2010-2015, masih banyak mahasiswa yang **tidak lulus** relatif terhadap mahasiswa yang lulus. Ini menyebabkan permasalahan **imbalanced classification**





# Modeling and Evaluation

## Sampling Dataset

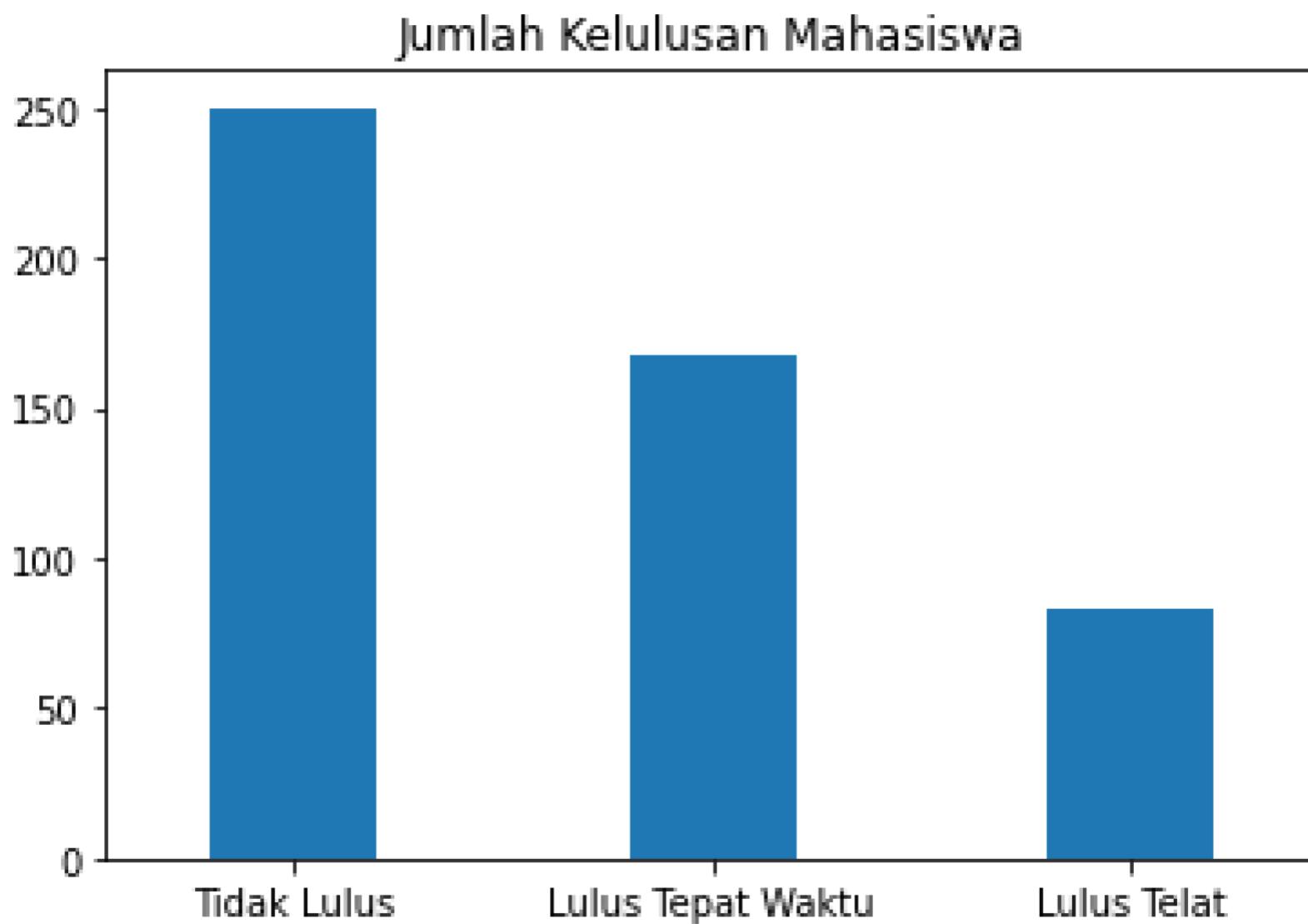
### Dataframe 6

Mahasiswa Angkatan 2010-2015

Untuk mengatasi hal ini, dilakukan sampling dengan teknik **RandomOverSampler** dan **RandomUnderSampler** sehingga perbandingannya menjadi **3:2:1**

```
ros = RandomOverSampler(sampling_strategy='all')
rus = RandomUnderSampler(sampling_strategy={'Tidak Lulus' : 250, 'Lulus Telat' : 83, 'Lulus Tepat Waktu' : 167})

X_ros, y_ros = ros.fit_resample(X, y)
X_rus, y_rus = rus.fit_resample(X_ros, y_ros)
```





# Modeling and Evaluation

## Pengembangan Model

### ■ Split Dataset

Split dataset hasil sampling  
menjadi **training dataset (75%)**  
dan **validation dataset (25%)**

```
# split dataset
X_train, X_val, y_train, y_val = train_test_split(X_rus, y_rus, test_size=0.25, shuffle=True, random_state=42)
```



# Modeling and Evaluation

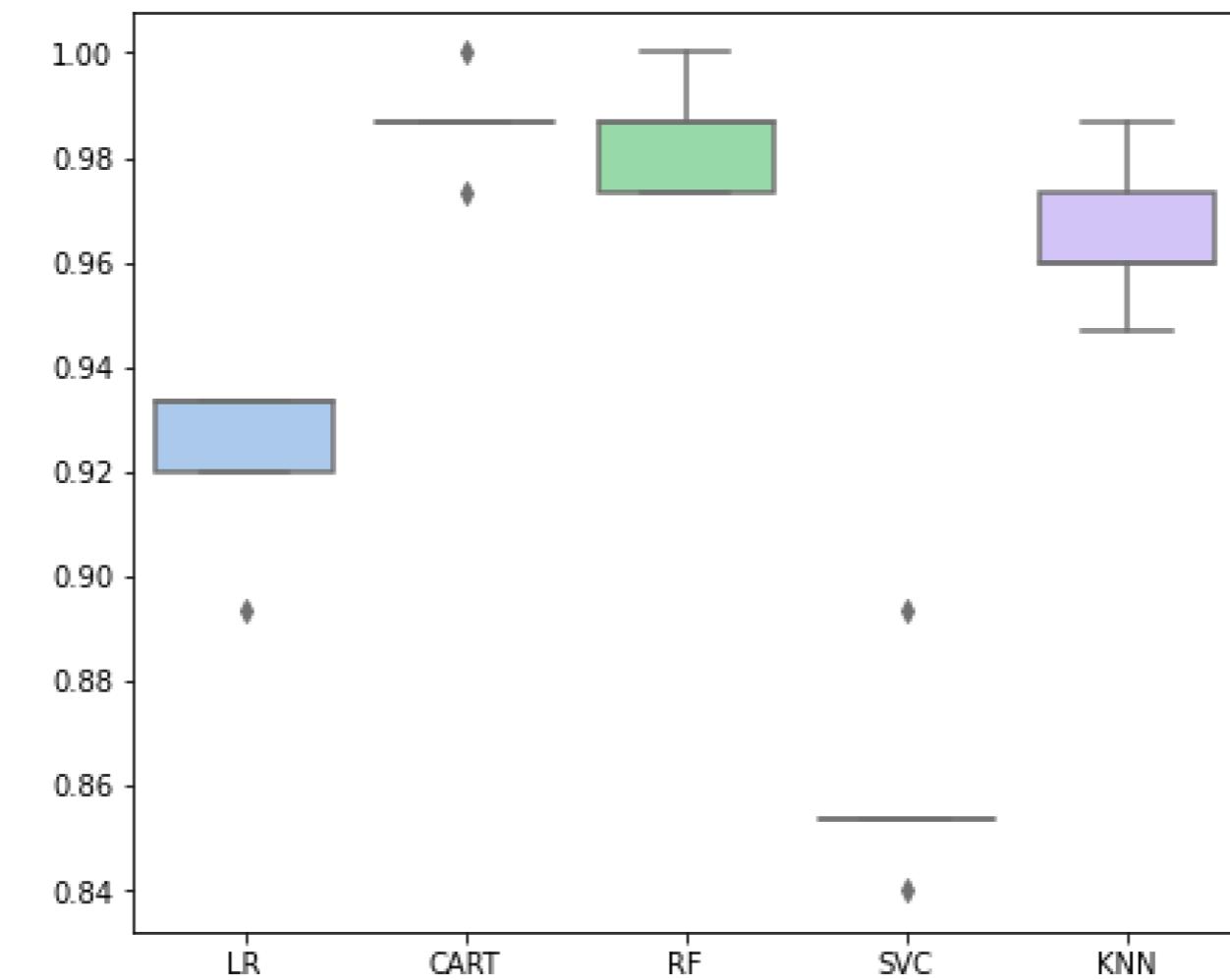
## Pengembangan Model

### ■ Pemeriksaan Algoritma

Memeriksa rata-rata dan simpangan baku akurasi dari 5-fold terhadap 5 algoritma

```
models = ['LR', LogisticRegression(),  
         ('CART', DecisionTreeClassifier()),  
         ('RF', RandomForestClassifier()),  
         ('SVC', SVC()),  
         ('KNN', KNeighborsClassifier())]  
  
models_name = []  
results = []  
  
for name, model in models:  
    kfold = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)  
    cv_score = cross_val_score(model, X_train, y_train, cv=kfold, scoring='accuracy')  
    models_name.append(name)  
    results.append(cv_score)  
    print(f'{name} \t | cv_score_mean: {cv_score.mean()} \t | cv_score_std: {cv_score.std()}')
```

LR	cv_score_mean: 0.9226666666666669	cv_score_std: 0.015549205052920814
CART	cv_score_mean: 0.9866666666666667	cv_score_std: 0.008432740427115663
RF	cv_score_mean: 0.984	cv_score_std: 0.009977753031397158
SVC	cv_score_mean: 0.8586666666666666	cv_score_std: 0.018086213288334037
KNN	cv_score_mean: 0.9653333333333333	cv_score_std: 0.013597385369580781



# Modeling and Evaluation

## Pengembangan Model

### ■ Hyperparameter Tuning

Memeriksa tinggi atau kedalaman Decision Tree yang memberikan akurasi terbaik

```
# check maximum depth of the tree
acc_score = []

def checkMaxDepth(X, y):
    depth = np.arange(2, 21, 1)

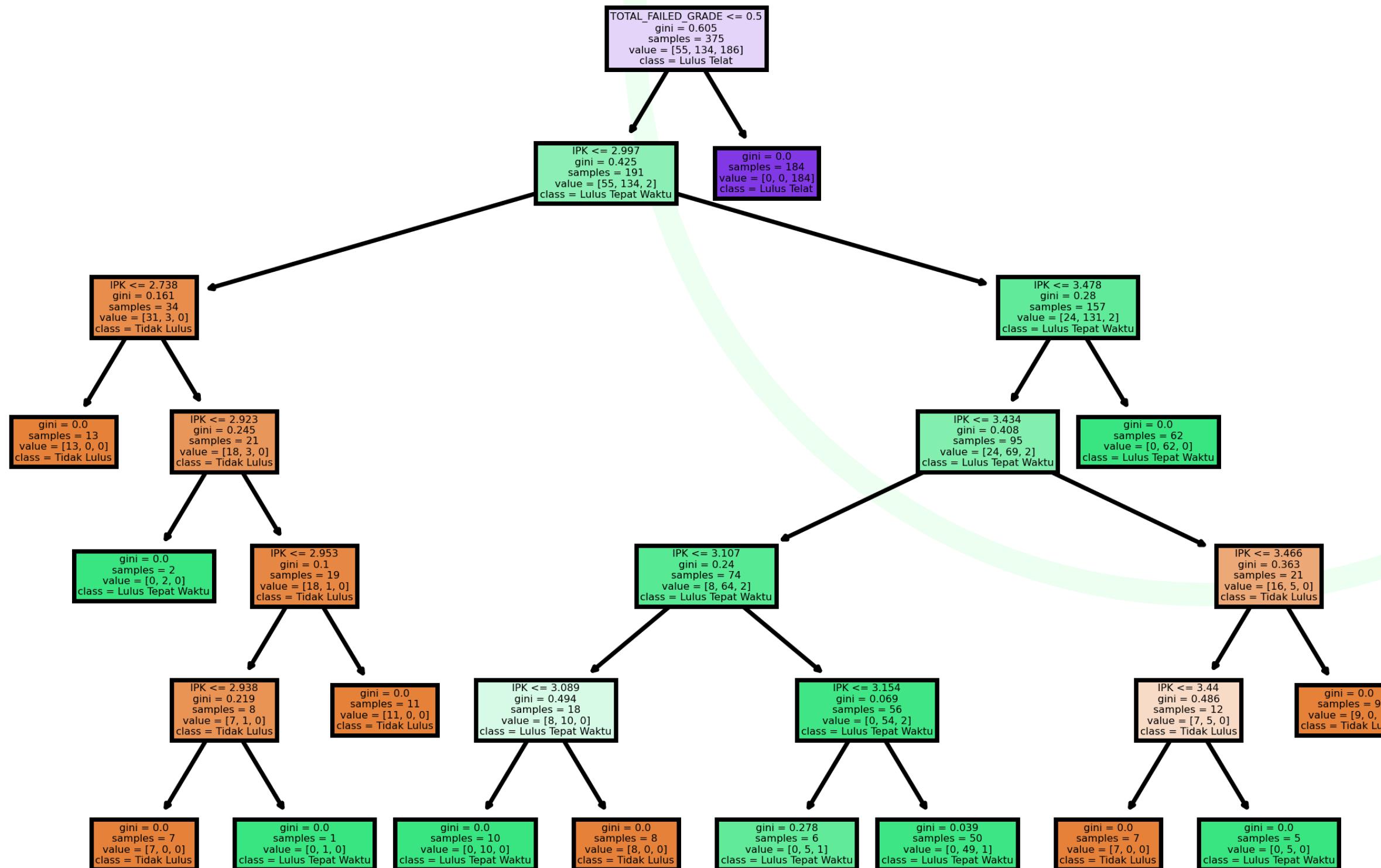
    for i in depth:
        cart = DecisionTreeClassifier(max_depth=i)
        cart.fit(X_train, y_train)
        pred = cart.predict(X_val)
        accuracy = accuracy_score(y_val, pred)
        acc_score.append(accuracy)
        print(f'max_depth: {i} | accuracy score: {accuracy}')

checkMaxDepth(X=X_train, y=y_train)
print("Depth with highest accuracy score :", acc_score.index(max(acc_score))+2)
```

```
max_depth: 2      | accuracy score: 0.912
max_depth: 3      | accuracy score: 0.912
max_depth: 4      | accuracy score: 0.944
max_depth: 5      | accuracy score: 0.944
max_depth: 6      | accuracy score: 0.984
max_depth: 7      | accuracy score: 0.984
max_depth: 8      | accuracy score: 0.984
max_depth: 9      | accuracy score: 0.984
max_depth: 10     | accuracy score: 0.984
max_depth: 11     | accuracy score: 0.984
max_depth: 12     | accuracy score: 0.984
max_depth: 13     | accuracy score: 0.984
max_depth: 14     | accuracy score: 0.984
max_depth: 15     | accuracy score: 0.984
max_depth: 16     | accuracy score: 0.984
max_depth: 17     | accuracy score: 0.984
max_depth: 18     | accuracy score: 0.984
max_depth: 19     | accuracy score: 0.984
max_depth: 20     | accuracy score: 0.984
Depth with highest accuracy score : 6
```

Dengan demikian, **tinggi pohon 6** memberikan akurasi terbaik.

# Modeling and Evaluation

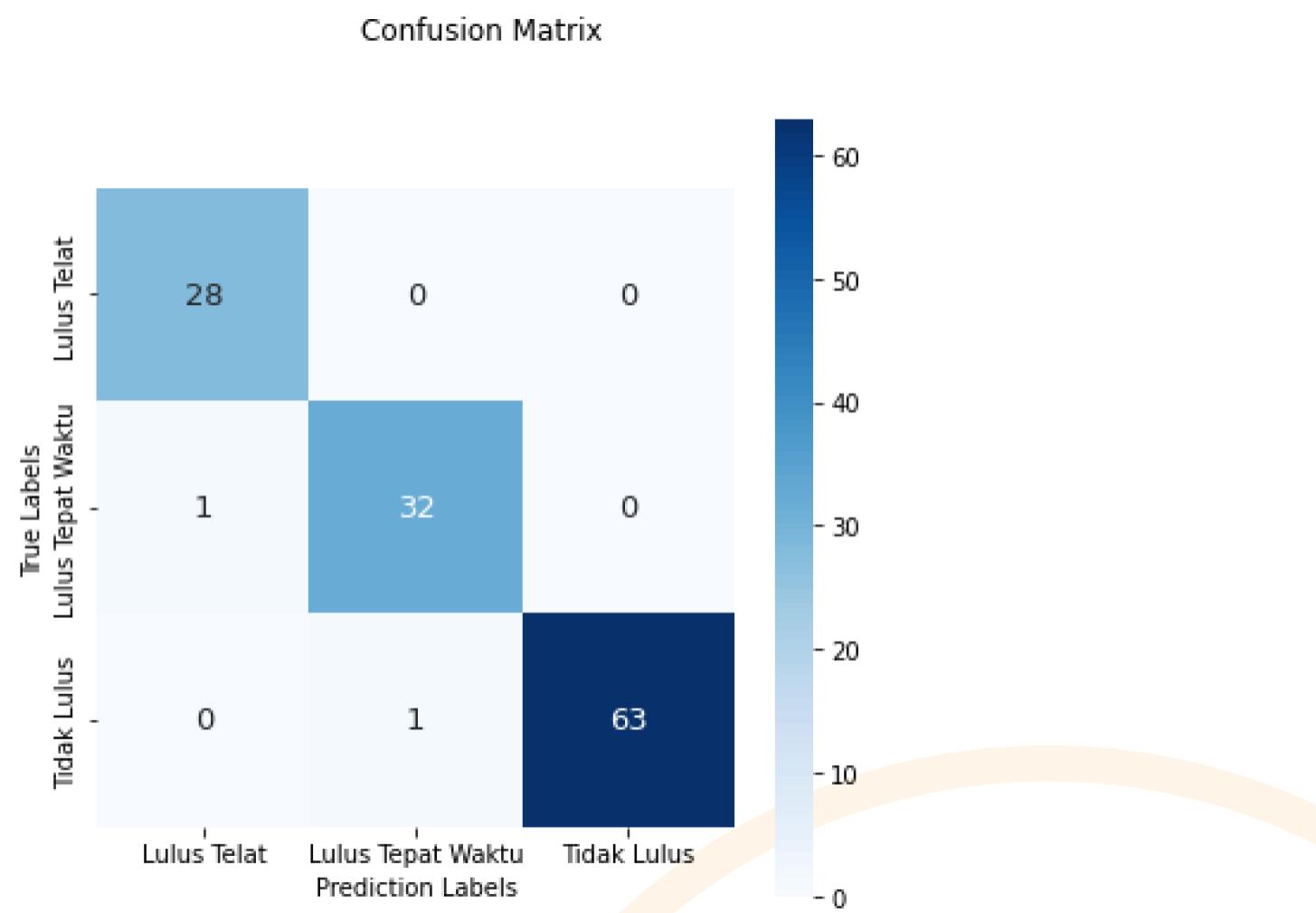




# Modeling and Evaluation

## Evaluasi Model

```
DecisionTreeClassifier(max_depth=6) Evaluation  
-----  
accuracy score: 0.984  
  
classification report:  
precision recall f1-score support  
  
Lulus Telat      0.97   1.00   0.98     28  
Lulus Tepat Waktu 0.97   0.97   0.97     33  
Tidak Lulus      1.00   0.98   0.99     64  
  
accuracy          0.98   0.98   0.98    125  
macro avg         0.98   0.98   0.98    125  
weighted avg      0.98   0.98   0.98    125
```



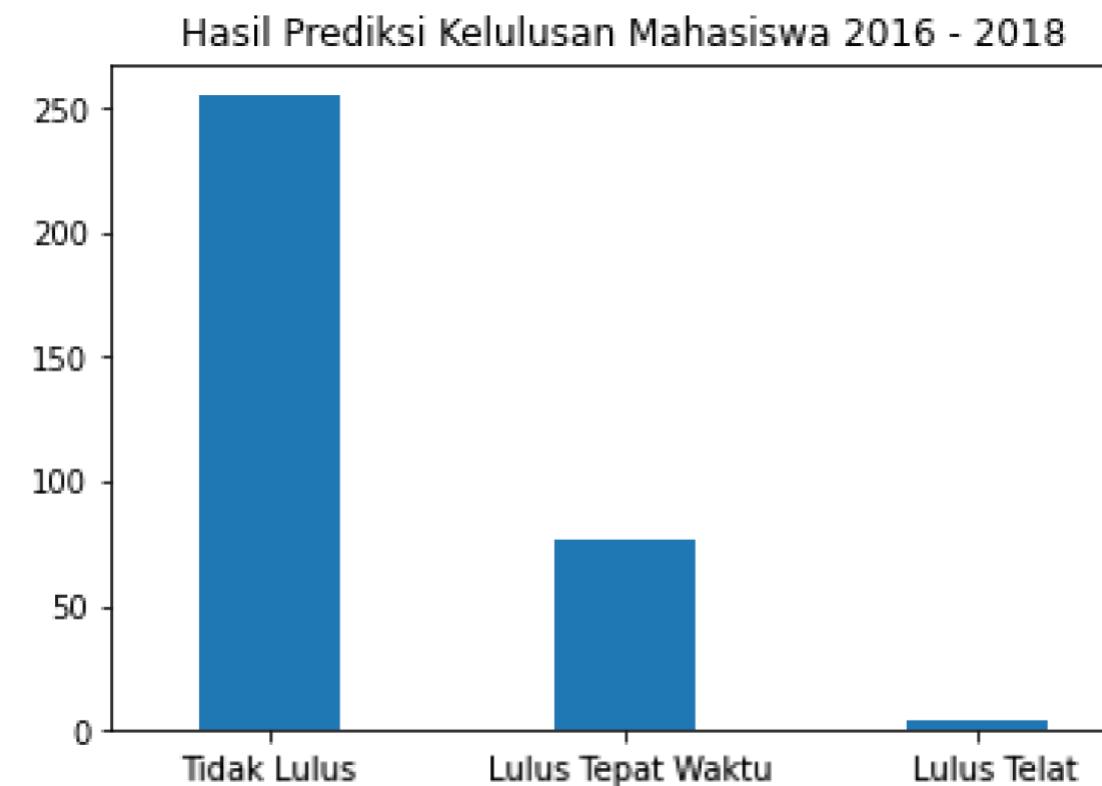
Hasil evaluasi menunjukkan akurasi sebesar 0.984

Dengan kesalahan 2 dari 125 hasil prediksi



# Modeling and Evaluation

## Prediksi Kelulusan Mahasiswa Angkatan 2016-2018



Berdasarkan diagram batang di samping, hasil prediksi menunjukkan **masih banyak mahasiswa yang Tidak Lulus** dibandingkan dengan mahasiswa yang Lulus

# Modeling and Evaluation

## Prediksi Kelulusan Mahasiswa Angkatan 2016-2018

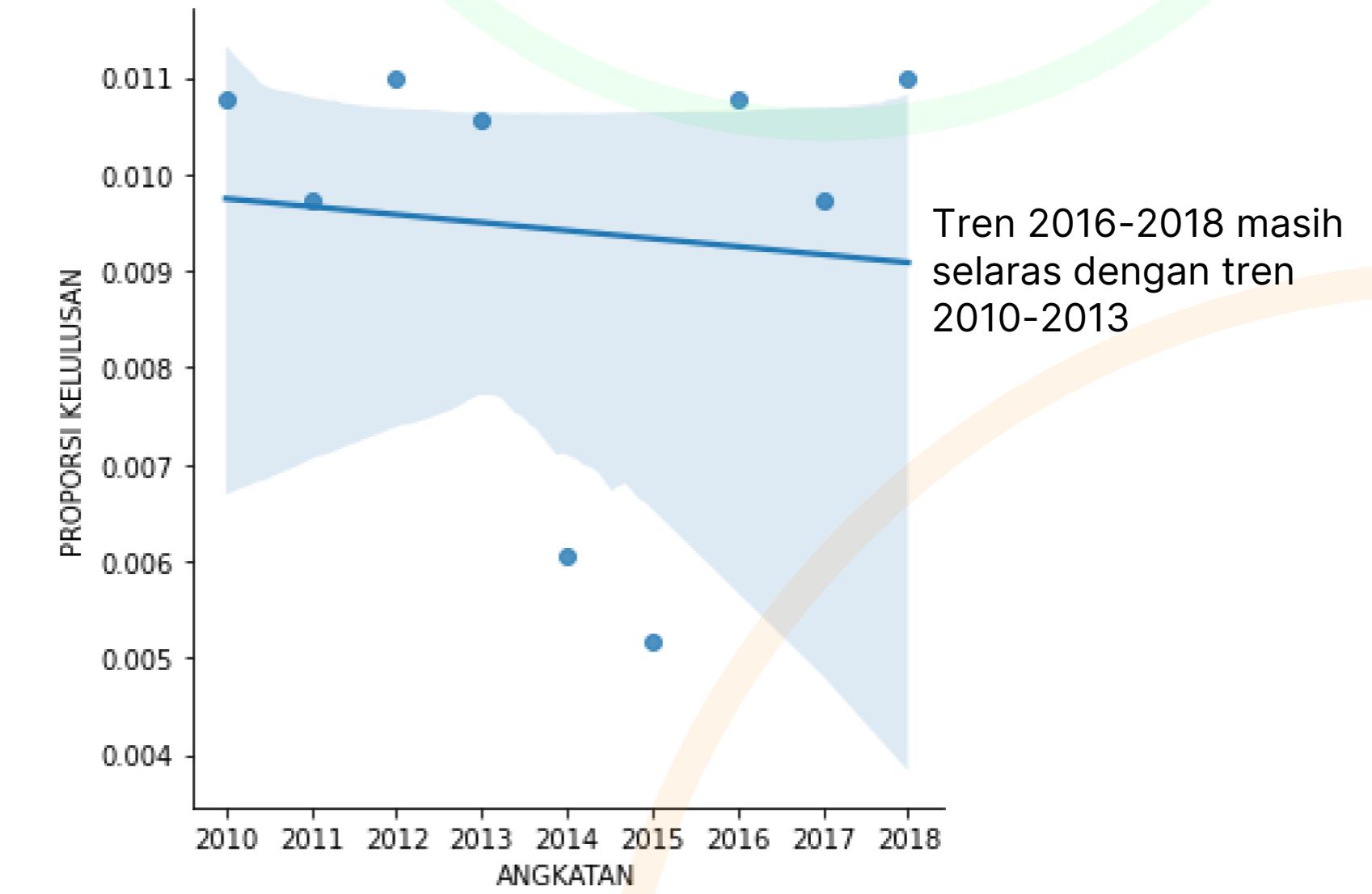
### Proporsi Kelulusan tiap Tahun

ANGKATAN	JUMLAH KELULUSAN	JUMLAH MAHASISWA	PROPORSI KELULUSAN
0	2016	15	0.010770
1	2017	25	0.009720
2	2018	41	0.010989

### Catatan:

Proporsi kelulusan = Jumlah mahasiswa yang lulus pada suatu angkatan / Total Mahasiswa Angkatan tersebut.

### Tren Kelulusan





# Conclusion / Suggestion

- Kualitas dari dosen atau hasil dari evaluasi dosen tidak memengaruhi performa dari mahasiswa yang mengikuti kelasnya.
- Terdapat 2 buah data yang memiliki pengaruh besar terhadap waktu kelulusan dari mahasiswa, yaitu data IPK mahasiswa dan banyaknya mata kuliah yang tidak lulus (Mata kuliah dengan nilai dibawah C).
- Berdasarkan tren kelulusan yang telah diplot dari angkatan 2010 hingga 2015, terlihat tren menurun pada proporsi kelulusan mahasiswanya. Akan tetapi, berdasarkan prediksi, tren dari kelulusan mahasiswa sama dengan tren dari tahun 2010 hingga 2013. Dari analisis yang telah dilakukan, hal ini terjadi karena pada mahasiswa angkatan 2014 dan 2015 terdapat banyak mahasiswa yang masih “belum lulus”.



**TERIMA  
KASIH**

The text "TERIMA KASIH" is centered in bold black capital letters. It is framed by two curved lines: a light green curve on the left and a light orange curve on the right, both of which start and end at the top and bottom corners of the text area respectively.