

NEWS AGENCY INSIGHT SURVEY



MEMBER



ขวัญสวรรค์ กองประจำชุด
643020498-9



พิมพิกา ยอดศรี
643020511-3



นริศรา โยราจันทร์
643020504-0



นรีกานต์ ปัญญารักษ์
643020505-8



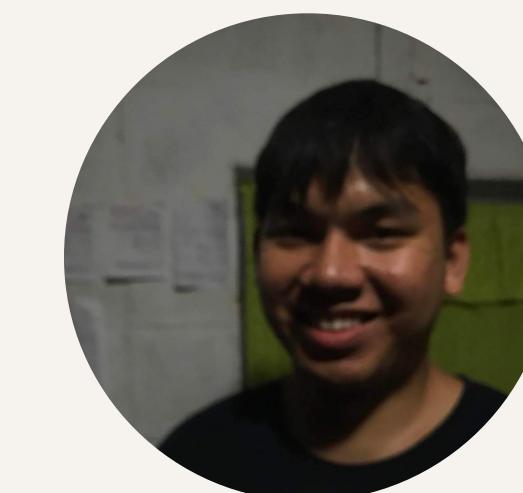
ตะวัน เบ้าหล่อเพชร
643020501-6



ธนาธิป อันตรคีรี
643021265-7

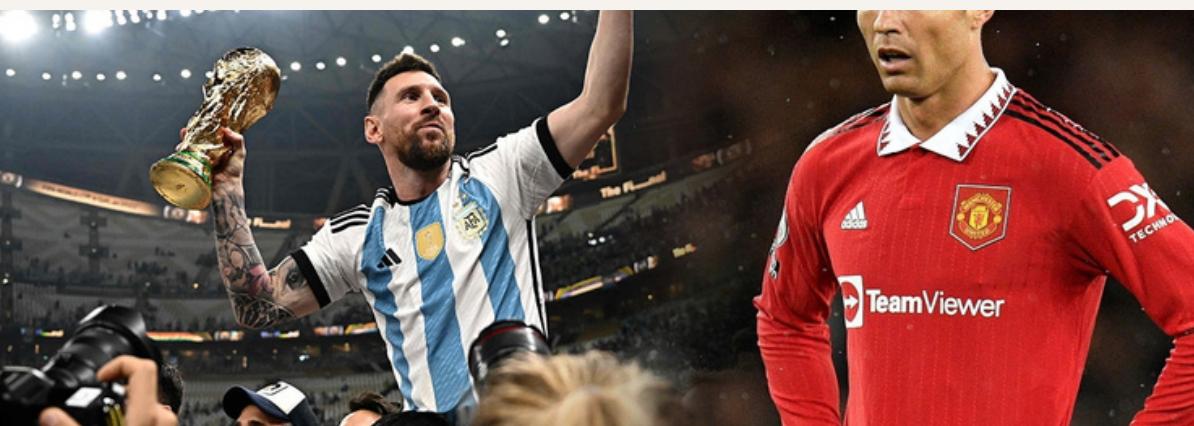


พุทธิพงศ์ ย่างนอกร
643021268-1



ศตวรรษ มูลสันเทียะ
643021273-8

STATEMENT



DATA PREPOCESSING

IMPORT DATA

SELECT COLUMN TO USE

CLEAN DATA

CHECK LOST DATA

NEW DATAFRAME TO USE

DATA PREPROCESSING

IMPORT DATA

```
[ ] survey = pd.read_excel('/content/drive/MyDrive/BSC_DPDM23/Project_data/dataset_consumer-insight-survey2019.xlsx')

[ ] dirty_data = survey.copy()
dirty_data
```

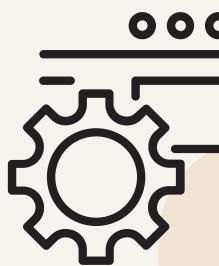
	id	a1	a2	a301	a302	a3021	a4	a5	a51	a52	...	u19new	r20new	s20new	t20new	u20new	G_genZ	G_genY	G_genX
0	1	1	2530	2	6.0	NaN	2	7	NaN	NaN	...	0	0	0	0	0	NaN	2.0	NaN
1	2	2	2529	2	6.0	NaN	2	4	NaN	NaN	...	0	0	0	0	0	NaN	2.0	NaN
2	3	2	2531	2	6.0	NaN	2	4	NaN	NaN	...	0	0	0	0	0	NaN	2.0	NaN
3	4	2	2516	2	5.0	NaN	2	2	NaN	6.0	...	0	0	0	0	0	NaN	NaN	3.0
4	5	2	2508	2	1.0	NaN	2	6	NaN	NaN	...	0	0	0	0	0	NaN	NaN	NaN
...	
2054	2044	2	2524	2	3.0	NaN	2	7	NaN	NaN	...	0	0	0	0	0	NaN	NaN	3.0
2055	2045	1	2498	2	1.0	NaN	2	6	NaN	NaN	...	0	0	0	0	0	NaN	NaN	NaN
2056	2046	2	2517	2	3.0	NaN	2	4	NaN	NaN	...	0	0	0	0	0	NaN	NaN	3.0
2057	2047	1	2511	2	6.0	NaN	2	2	NaN	NaN	...	0	0	0	0	0	NaN	NaN	3.0
2058	2048	2	2513	2	6.0	NaN	2	1	NaN	NaN	...	0	0	0	0	0	NaN	NaN	3.0

2059 rows × 200 columns

ຂ້ອມູນ 2059 RECORD ແລະ 200 COLUMNS

DATA PREPROCESSING

SELECT COLUMN TO USE



การเลือกคอลัมน์เพื่อใช้สำหรับการทำ Association Rules และ Classification

CODE

```
use_column = ['k1', 'k2', 'k3', 'k4', 'k5', 'k6', 'k7', 'k8', 'k9', 'k10', 'r1', 'r2', 'r3', 'r4', 'r5','rr1', 'a1','a5','a6', 'a81', 'a91', 'J5new', 'v1',  
'x1', 'x2','x3','x4', 'Gen_group', 'q1']
```

[] be_clean_data.describe()											
	k1	k2	k3	k4	k5	k6	k7	k8	k9	k10	
count	1849.000000	1849.000000	1849.000000	1849.000000	1849.000000	1849.000000	1849.000000	1849.000000	1849.000000	1849.000000	
mean	0.052461	0.326122	0.103840	0.164413	0.148188	0.219578	0.553272	0.243375	0.007031	0.146566	
std	0.223015	0.468919	0.305135	0.370750	0.355383	0.414073	0.497288	0.429236	0.083577	0.353768	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	
75%	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	

ใช้ 29 columns จาก 200 columns

DATA PREPROCESSING

CLEANING DATA

```
be_clean_data['r2'] = be_clean_data['r2'].replace({21: 99, 80: 99})  
be_clean_data['r3'] = be_clean_data['r3'].replace({31: 99})
```

ข้อมูลที่กรอกผิด แก้ไขเป็น 99

```
be_clean_data.isnull().any()
```

เช็คค่า NULL ของข้อมูล

```
clean_data = be_clean_data.dropna()
```

ทำการDROPNA

DATA PREPOCESSING

CHECK LOST DATA

DATA ກ່ອນ DROPNA

2059

DATA ກໍລັງ DROPNA

1849

ຂອນມູລມີກາຣສູງໝາຍ

10.199 %

DATA PREPROCESSING

DATAFRAME TO USE

data for association

```
▶ asso_data = can_use_data[asso_column].copy()  
asso_data
```

	k1	k2	k3	k4	k5	k6	k7	k8	k9	k10	r1	r2	r3	r4	r5	rr1
1	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	2.0	3.0	4.0	8.0	10.0	99.0
2	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	7.0	5.0	3.0	1.0	4.0	99.0
3	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	3.0	2.0	4.0	15.0	99.0
4	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	3.0	2.0	4.0	15.0	99.0
5	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	3.0	2.0	5.0	15.0	99.0
...
2054	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	3.0	15.0	18.0	99.0	99.0	99.0
2055	0.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	3.0	2.0	99.0	99.0	99.0	99.0
2056	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	3.0	15.0	1.0	18.0	2.0	99.0
2057	0.0	0.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	3.0	99.0	99.0	99.0	99.0	99.0
2058	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	2.0	15.0	18.0	19.0	7.0	99.0

1849 rows × 16 columns

DATA PREPROCESSING

DATAFRAME TO USE

data for classification

```
▶ classi_data = can_use_data[classi_column].copy()  
classi_data
```

	a1	a5	a81	a91	J5new	v1	x1	x2	x3	x4	Gen_group	q1
1	2.0	4.0	3.0	5.0	3.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0
2	2.0	4.0	1.0	5.0	3.0	2.0	1.0	1.0	0.0	0.0	2.0	1.0
3	2.0	2.0	1.0	5.0	4.0	2.0	1.0	1.0	1.0	1.0	3.0	1.0
4	2.0	6.0	4.0	5.0	3.0	2.0	1.0	0.0	0.0	0.0	4.0	1.0
5	1.0	2.0	1.0	5.0	3.0	2.0	1.0	1.0	1.0	1.0	3.0	1.0
...
2054	2.0	7.0	6.0	3.0	2.0	1.0	0.0	0.0	0.0	1.0	3.0	1.0
2055	1.0	6.0	4.0	3.0	1.0	1.0	0.0	0.0	0.0	1.0	4.0	99.0
2056	2.0	4.0	6.0	3.0	1.0	2.0	1.0	0.0	1.0	1.0	3.0	1.0
2057	1.0	2.0	4.0	3.0	1.0	1.0	1.0	1.0	1.0	1.0	3.0	1.0
2058	2.0	1.0	6.0	3.0	1.0	2.0	0.0	0.0	0.0	1.0	3.0	1.0

1849 rows × 12 columns

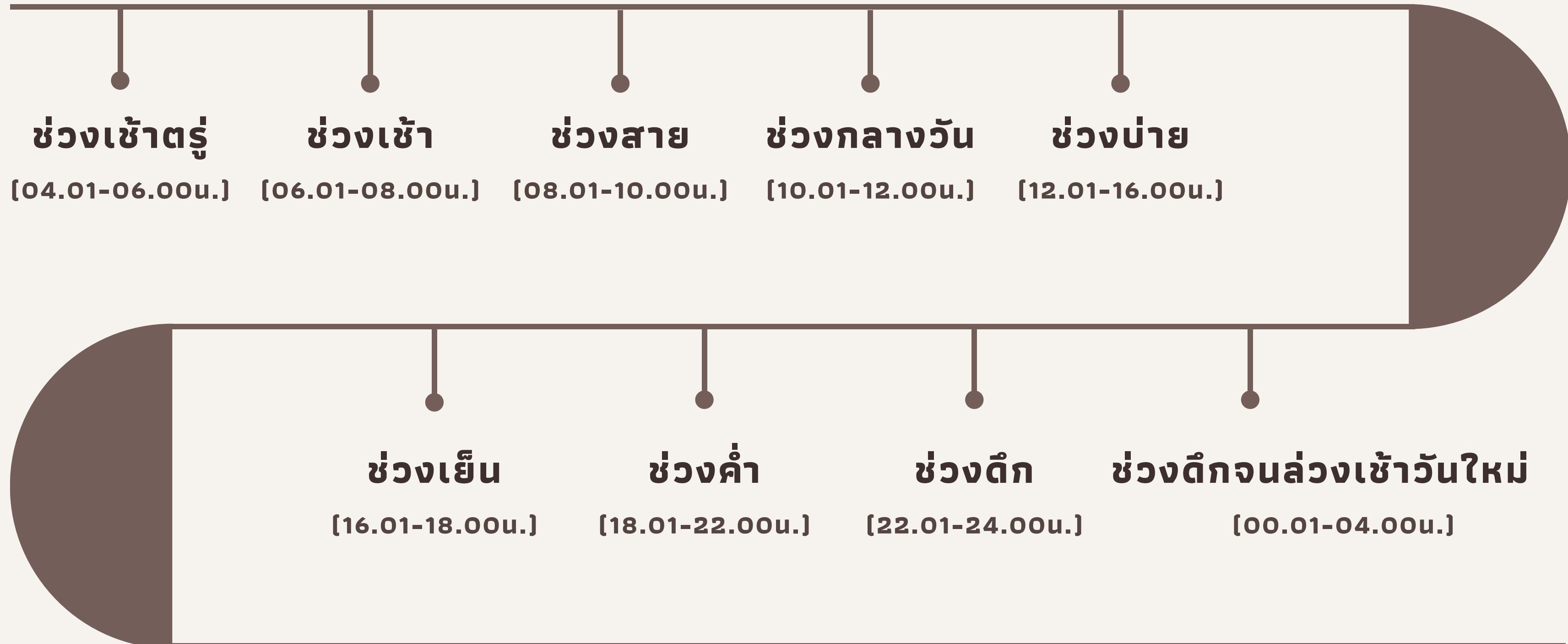


ASSOCIATION RULES

ในแต่ละช่วงเวลาของวันประจำข่าวคุ้ยได้
ถูกชื่นชอบมากที่สุด



Time



ປະເທດທ່າວ



ຂ່າວກາຮົມ



ຂ່າວເສດຖະກິຈ



ຂ່າວໃນກະແສຄວາມສັນໃຈ



ຂ່າວພະພາບສຳນັກ



ຂ່າວຕ່າງປະເທດ



ຂ່າວອົບຕິເຫດ



ຂ່າວອຸນດີເຫດ



ຂ່າວການນັນເທິງ



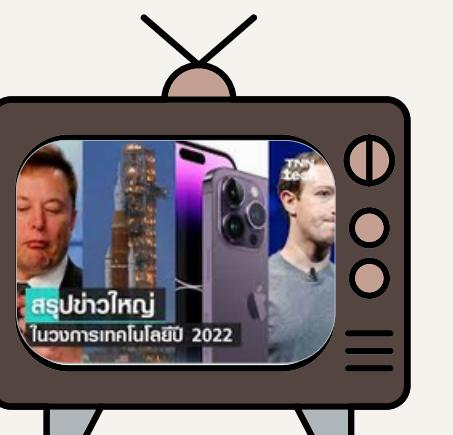
ຂ່າວສຶລປະວັນອຣນມ



ຂ່າວກີພາ



ຂ່າວເກົຫງ



ຂ່າວເທິກໂນໂລຢີ



ຂ່າວກາຮົມ



ຂ່າວກາພະຍົບ



ຂ່າວພາຍການໝາກາສ



ຂ່າວຈະຈາກ



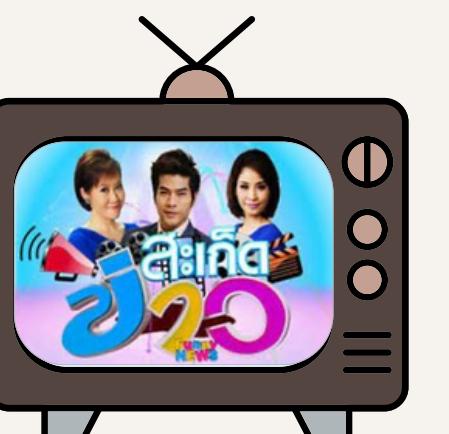
ຂ່າວສິ່ງແວດລ້ອມ



ຂ່າວຮັບກຸກົມ



ຂ່າວທີ່ປະຈຳນາຍງານ



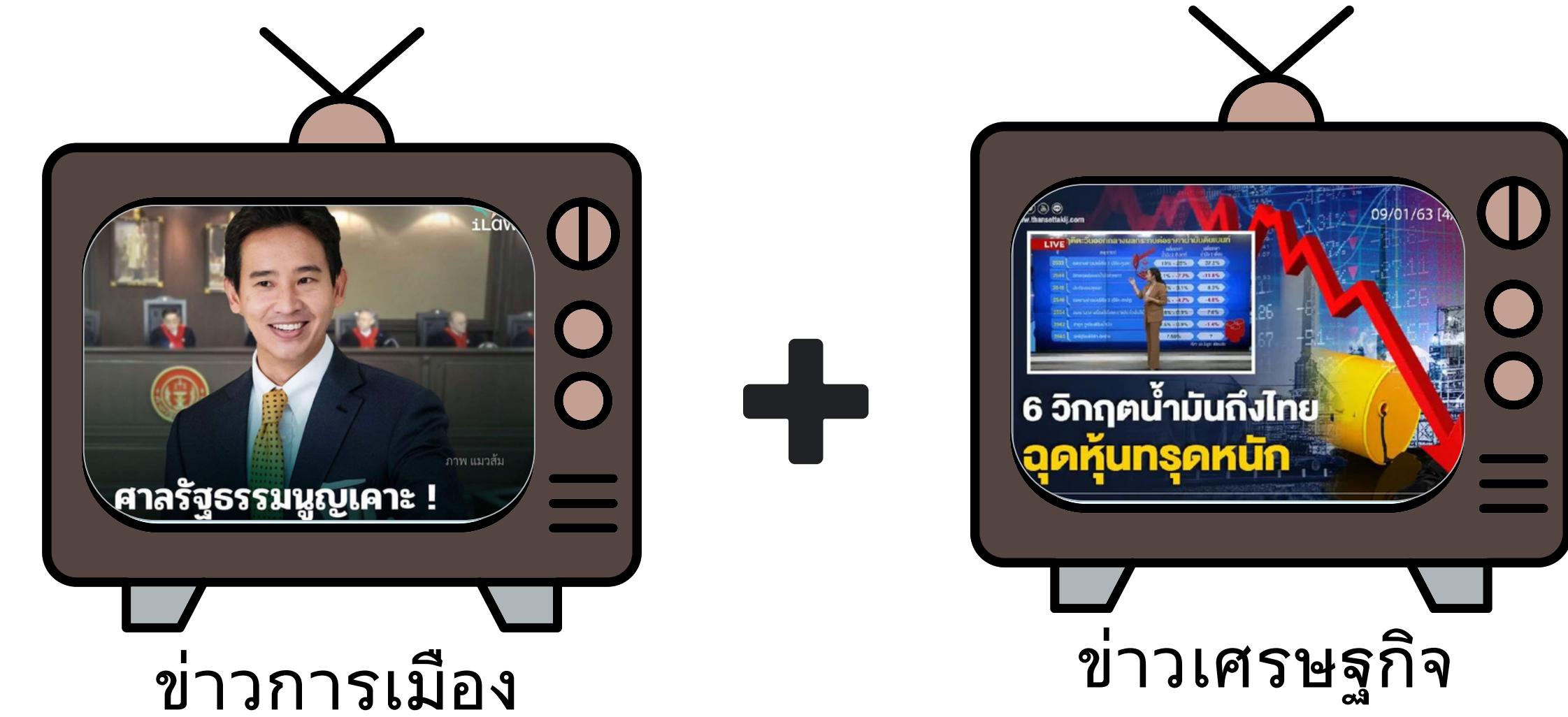
ອື່ນໆ

กลุ่ม

แหล่ง

กองประมาณการ

ช่วงเช้า
04.01-06.00 น.



Support: 0.272

ช่วงเช้า

Confidence:0.583
Lift:1.355



ข่าวการเมือง



ข่าวเศรษฐกิจ

Confidence:0.632
Lift:1.355



ข่าวเศรษฐกิจ



ข่าวการเมือง

ช่วงเวลา



ข่าว



ข่าวเศรษฐกิจ



ข่าวการเมือง

- ช่วงเช้าตรู่(04.01-06.00 u.)
- ช่วงเช้า(06.01-08.00u.)
- ช่วงสาย(08.01-10.00u.)
- ช่วงค่ำ(18.01-22.00u.)



ข่าวการเมือง



ข่าวเศรษฐกิจ

- ช่วงกลางวัน(10.01-12.00u.)
- ช่วงบ่าย(12.01-16.00u.)
- ช่วงดึก(22.01-24.00u.)



ข่าวเศรษฐกิจ



ข่าวในกระแสความสนใจ

- ช่วงเช้าตรู่(04.01-06.00 u.)
- ช่วงดึกจนล่วงเข้าวันใหม่(00.01-04.00u.)



ข่าวในกระแสความสนใจ



ข่าวการเมือง

- ช่วงเย็น(16.01-18.00u.)

กลยุทธ์ที่1 เลือกหนึ่งช่วงเวลาที่มีคนซื้อขายค่อนข้างมากที่สุด



ข่าวเศรษฐกิจ



ข่าวการเมือง

- ช่วงเช้าตรู่(04.01-06.00 น.) = 17.052
- ช่วงเช้า(06.01-08.00น.) = 147.968
- ช่วงสาย(08.01-10.00น.) = 29.016
- ช่วงค่ำ(18.01-22.00น.) = 217.74

กลยุทธ์ที่ 1

เลือกหนึ่งช่วงเวลาที่มีคนชื่นชอบข่าวคุ้นหูมากที่สุด

$$\frac{\text{จำนวนคนที่ดูข่าวคุ้นหู}}{\text{ในช่วงเวลานั้นๆ}} = \frac{\text{จำนวนคนที่ดูข่าวทั้งหมด}}{\text{ในช่วงเวลานั้นๆ}} * \text{ค่า support ข่าวคุ้นหู} \text{ ในช่วงเวลานั้น}$$

$$\begin{aligned}\text{จำนวนคนที่ดูข่าวเศรษฐกิจกับข่าวการเมือง} &= 955 * 0.228 \\ &= 217 \text{ คน}\end{aligned}$$



ข่าวเศรษฐกิจ



ข่าวการเมือง

- ช่วงเช้าตรู่(04.01-06.00 u.)
- ช่วงเช้า(06.01-08.00u.)
- ช่วงสาย(08.01-10.00u.)
- ช่วงค่ำ(18.01-22.00u.)

กลยุทธ์ที่ 2

จัดข่าวคู่นั้นๆ ไล่ในทุกช่วงเวลาที่มีคนซื่นชอบมากที่สุด



ข่าวเศรษฐกิจ



ข่าวการเมือง

- ช่วงเช้าตรู่(04.01-06.00 u.)
- ช่วงเช้า(06.01-08.00u.)
- ช่วงสาย(08.01-10.00u.)
- ช่วงค่ำ(18.01-22.00u.)



វិធីការចែកចាយក្នុងក្រសួង



- ★ ត្រូវបានចែកចាយជាប្រភពទិន្នន័យ
- ★ ត្រូវបានចែកចាយជាប្រភពទិន្នន័យ ដើម្បីបង្កើតរឹងរាល់ និងអនុវត្តន៍



CLASSIFICATION

TRAINING

80%

TEST

20%



DECISION TREE

คนแบบใดที่ใช้ facebook เป็นอันดับ 1

เป้าหมาย



DECISION TREE

คนแบบใดที่ใช้ facebook เป็นอันดับ 1



DECISION TREE



Feature ที่ใช้

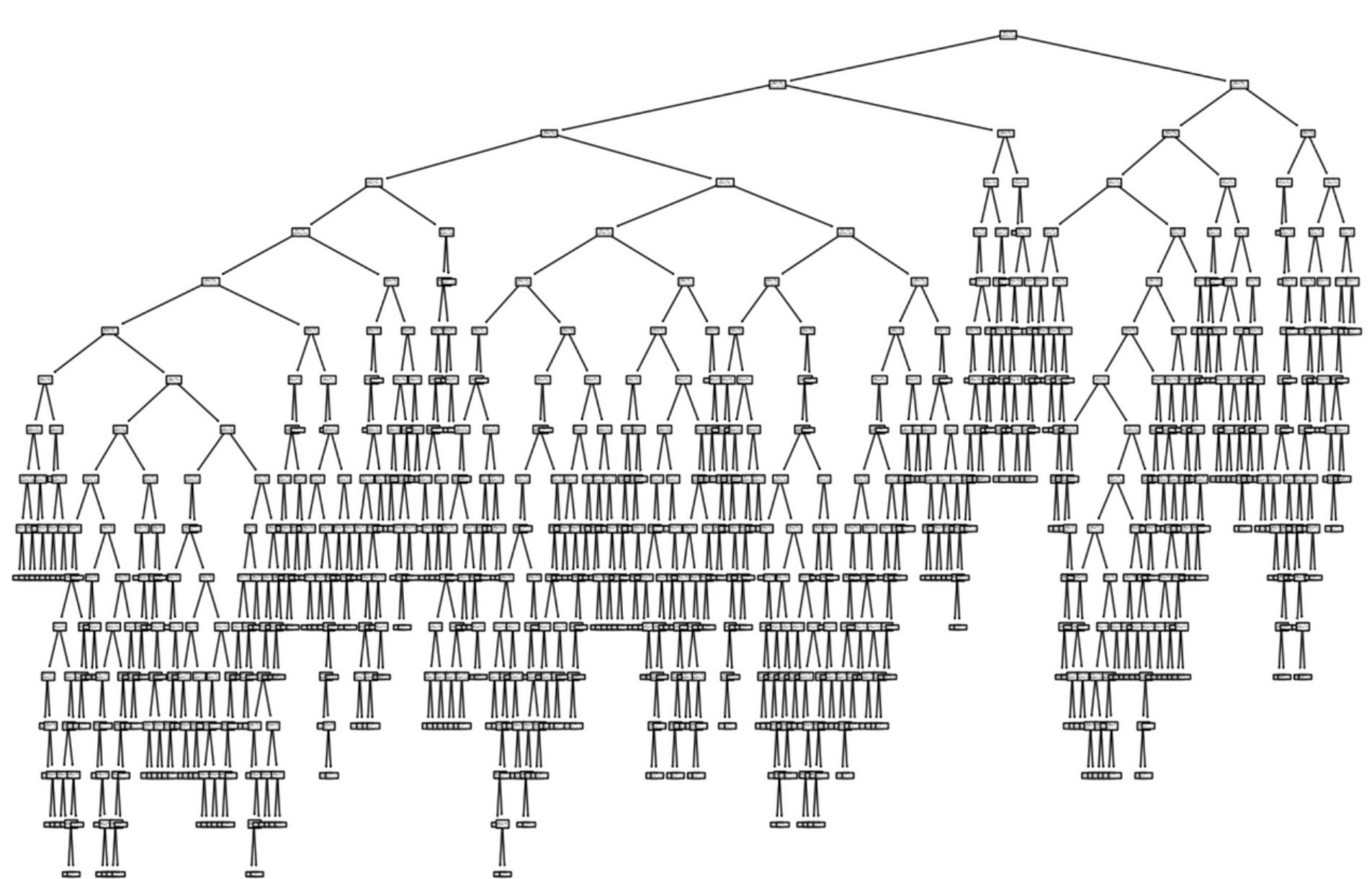
ตัวแปร X ที่ศึกษา

เพศ, อายุ, สถานภาพส่วนตัว, ภูมิภาคที่อาศัย, ติดตามข่าวผ่านมือถือ,
ติดตามข่าวแบบใดมากที่สุด, มักติดตามข่าวที่บ้าน, มักติดตามข่าวที่สถานที่ทำงาน,
มักติดตามข่าวบนยานพาหนะ, มักติดตามข่าวไม่จำกัดสถานที่, Gen_group

DECISION TREE

ปล่อยให้ต้นไม้โตโดยไม่เซ็ท PARAMETER

NON
PARAMETER



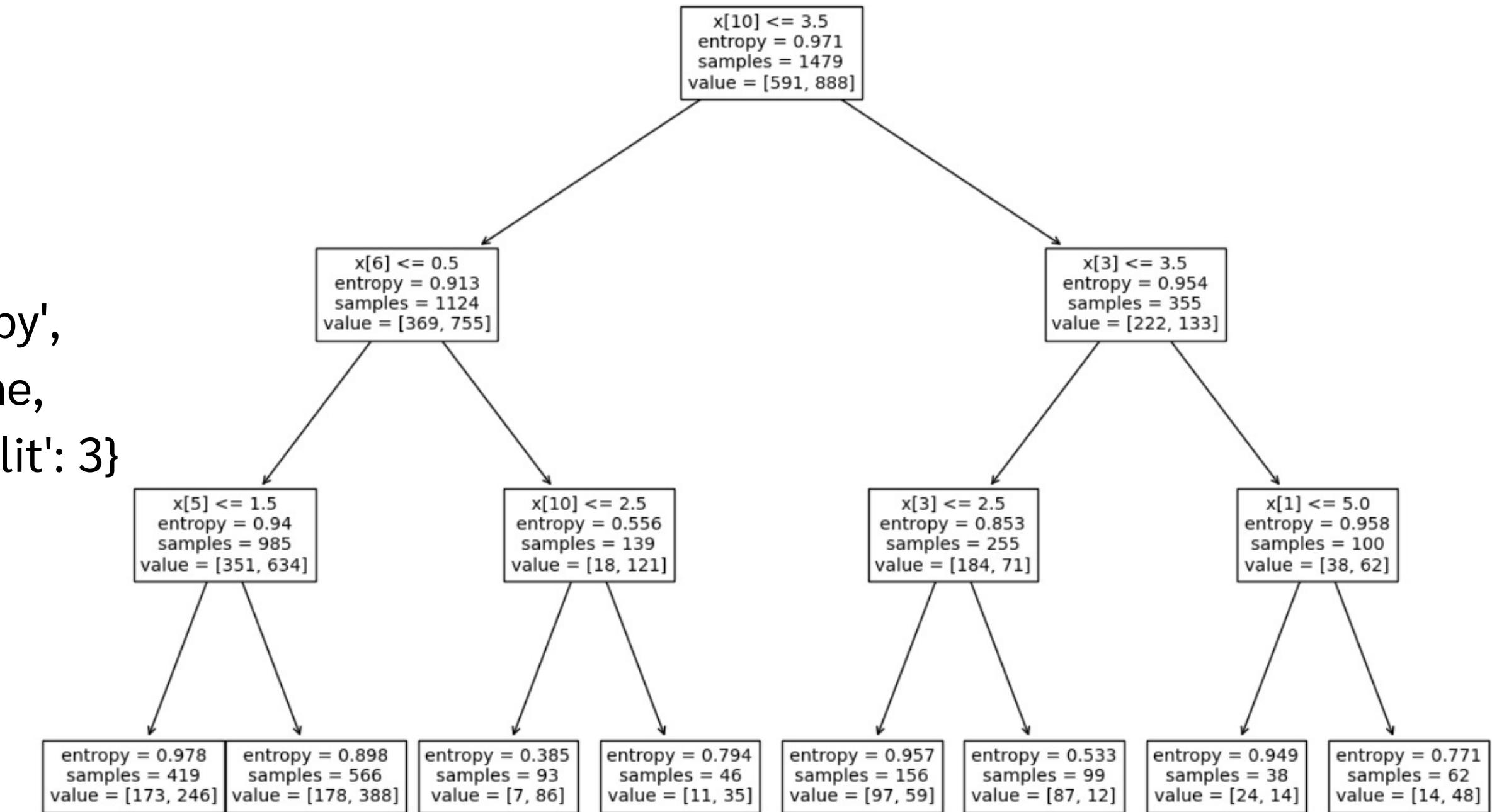
ACCURACY
68.37%

DECISION TREE

GRID SEARCH

BEST PARAMETER

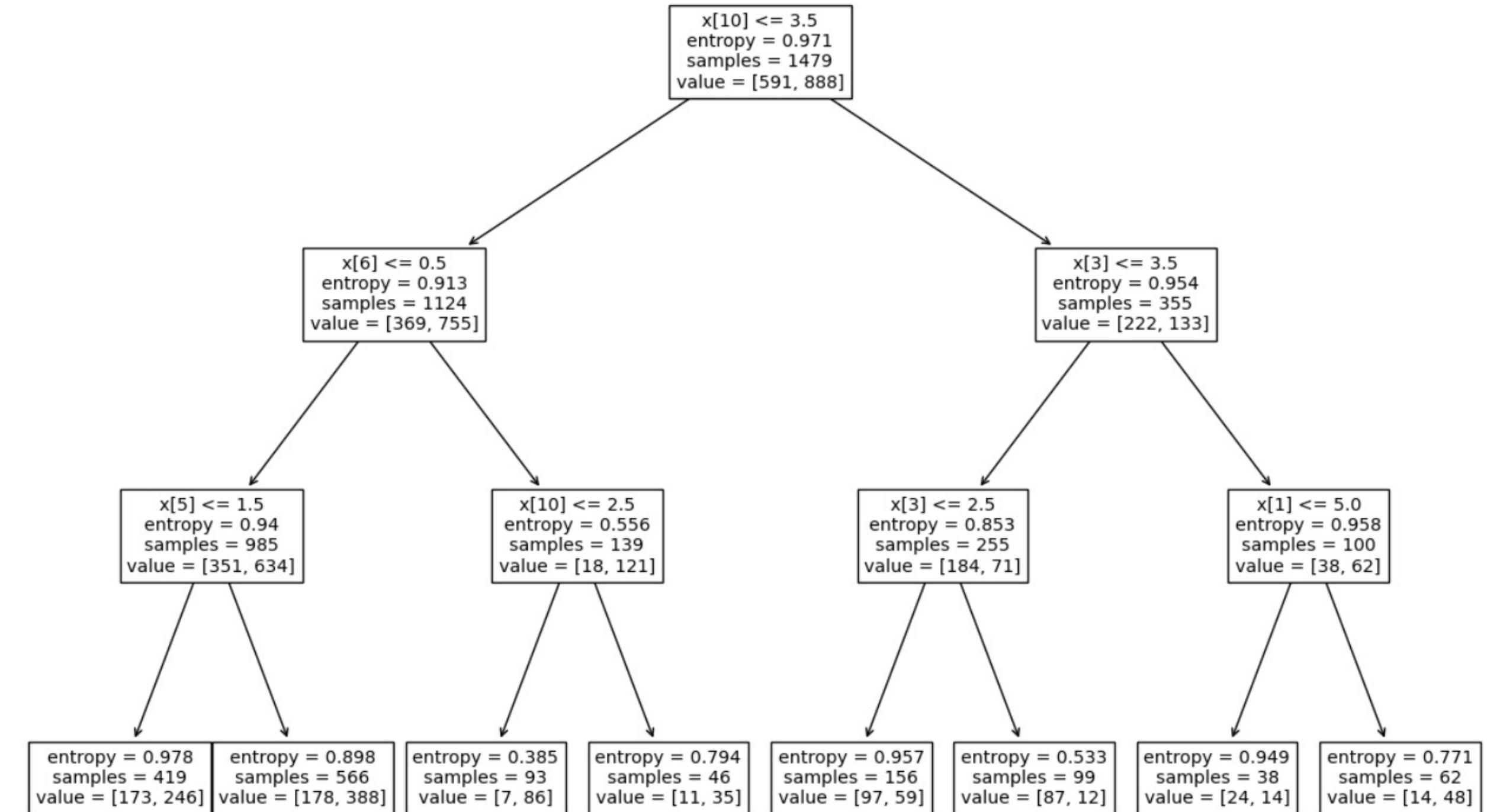
Best parameters: {'criterion': 'entropy',
'max_depth': 3, 'max_features': None,
'min_samples_leaf': 7, 'min_samples_split': 3}



DECISION TREE

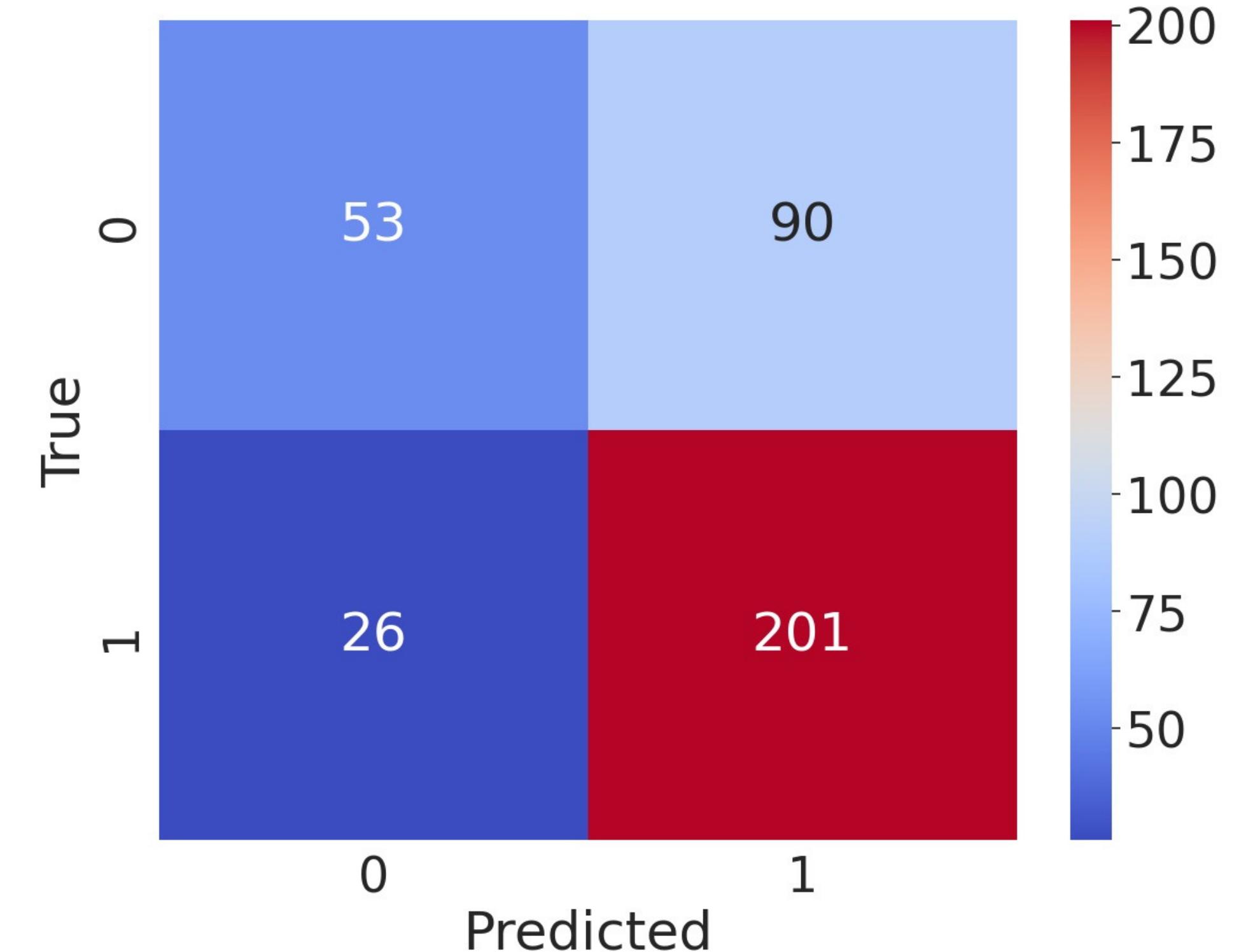
GRID SEARCH

ACCURACY
68.65 %



CONFUSION MATRIX

DECISION TREE



DECISION TREE

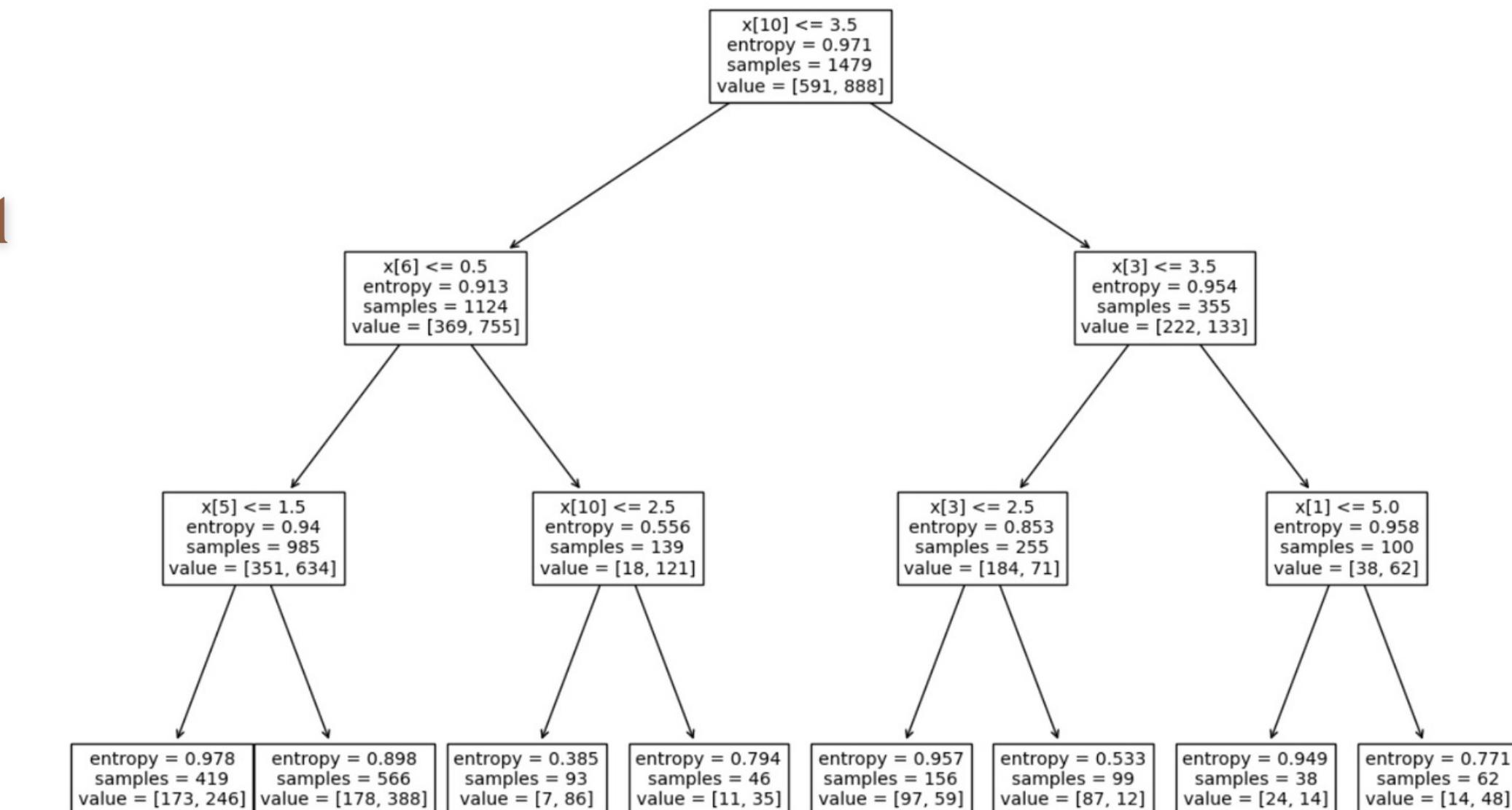
CROSS - VALIDATION

CLASS 0

คนที่ไม่ได้ใช้FACEBOOKเป็นอันดับ1

CLASS 1

คนที่ใช้FACEBOOKเป็นอันดับ1



DECISION TREE

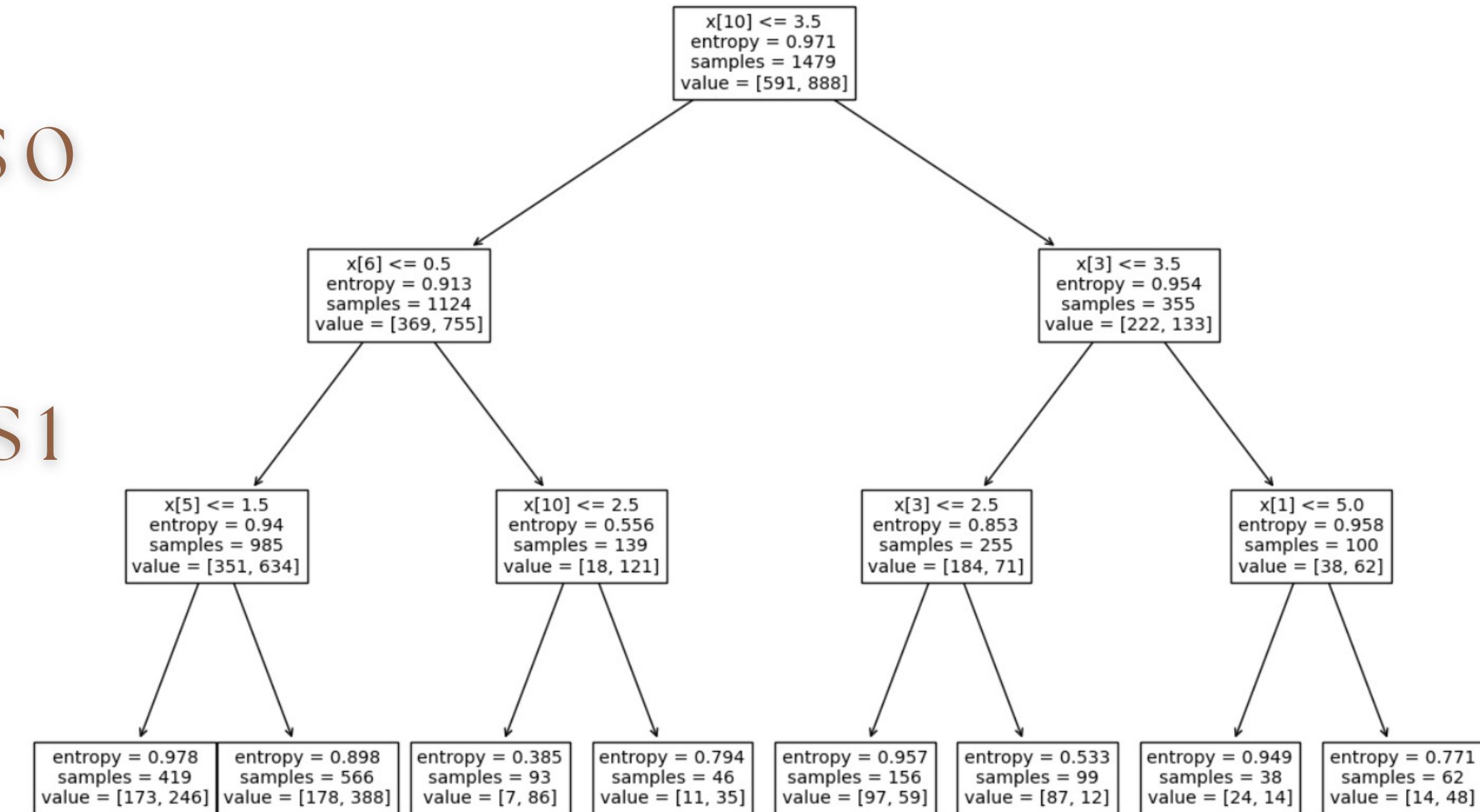
PRECISION

PRECISION CLASS 0

0.67

PRECISION CLASS 1

0.69



DECISION TREE

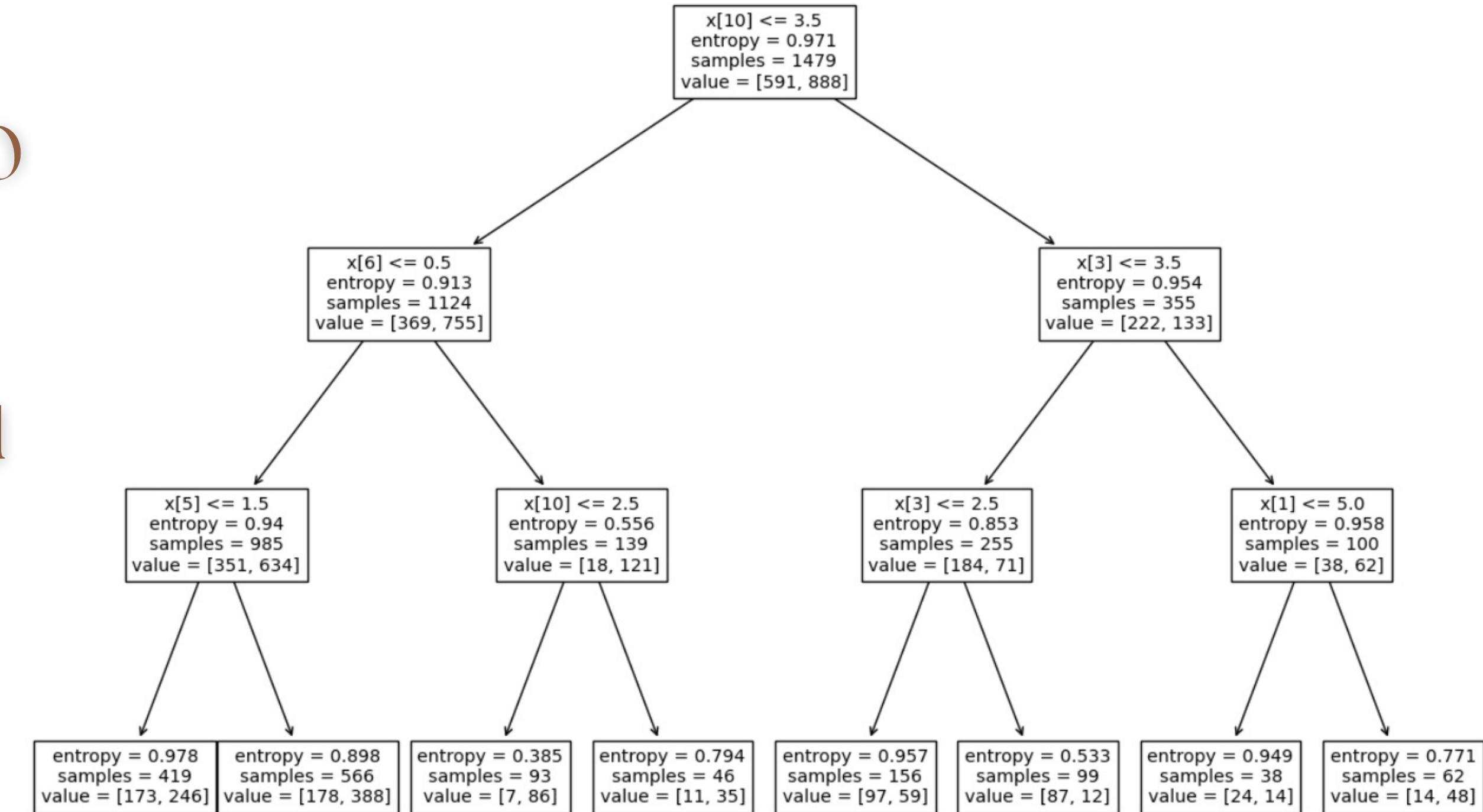
RECALL

RECALL CLASS 0

0.37

RECALL CLASS 1

0.89



DECISION TREE

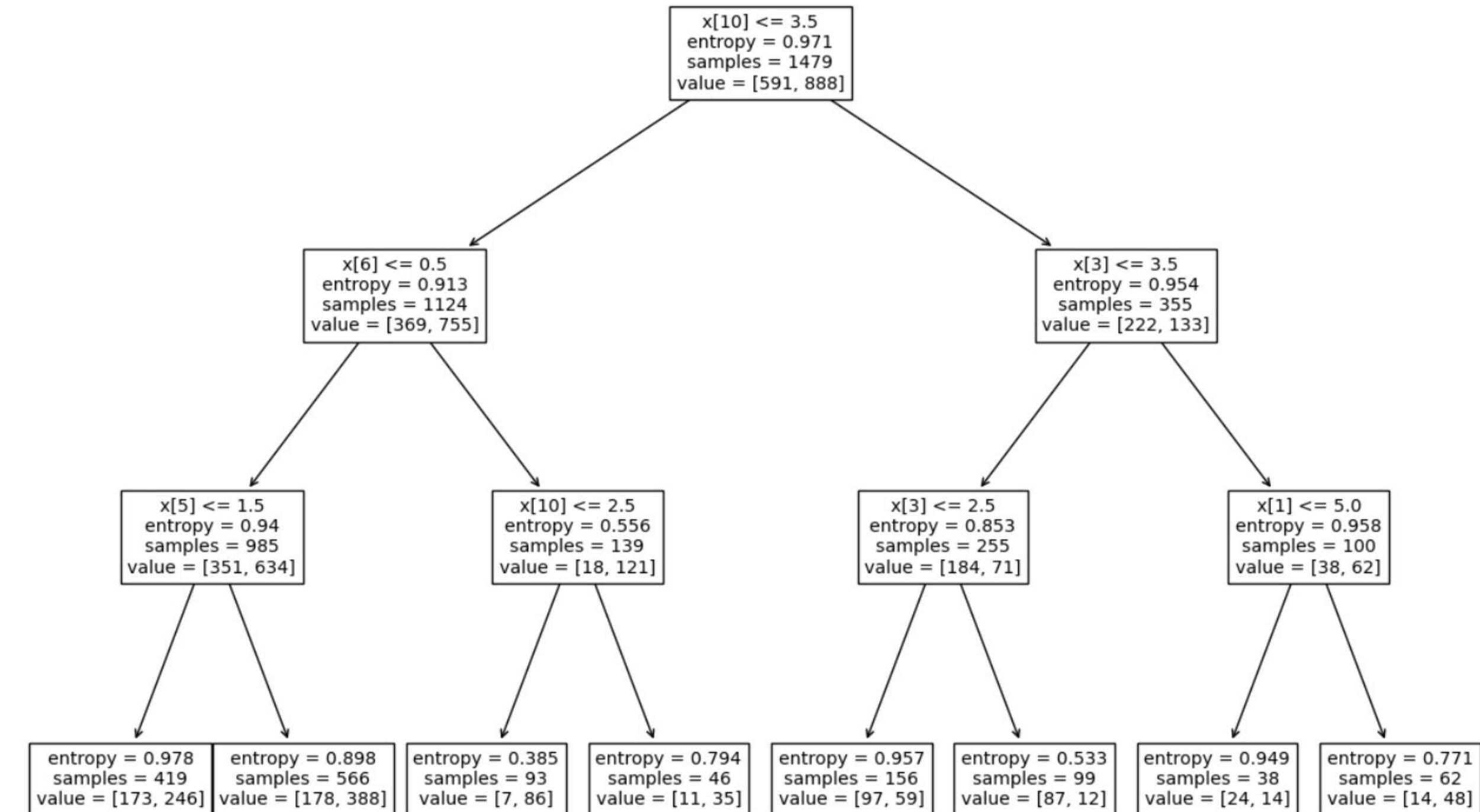
F1-SCORE

CLASS 0

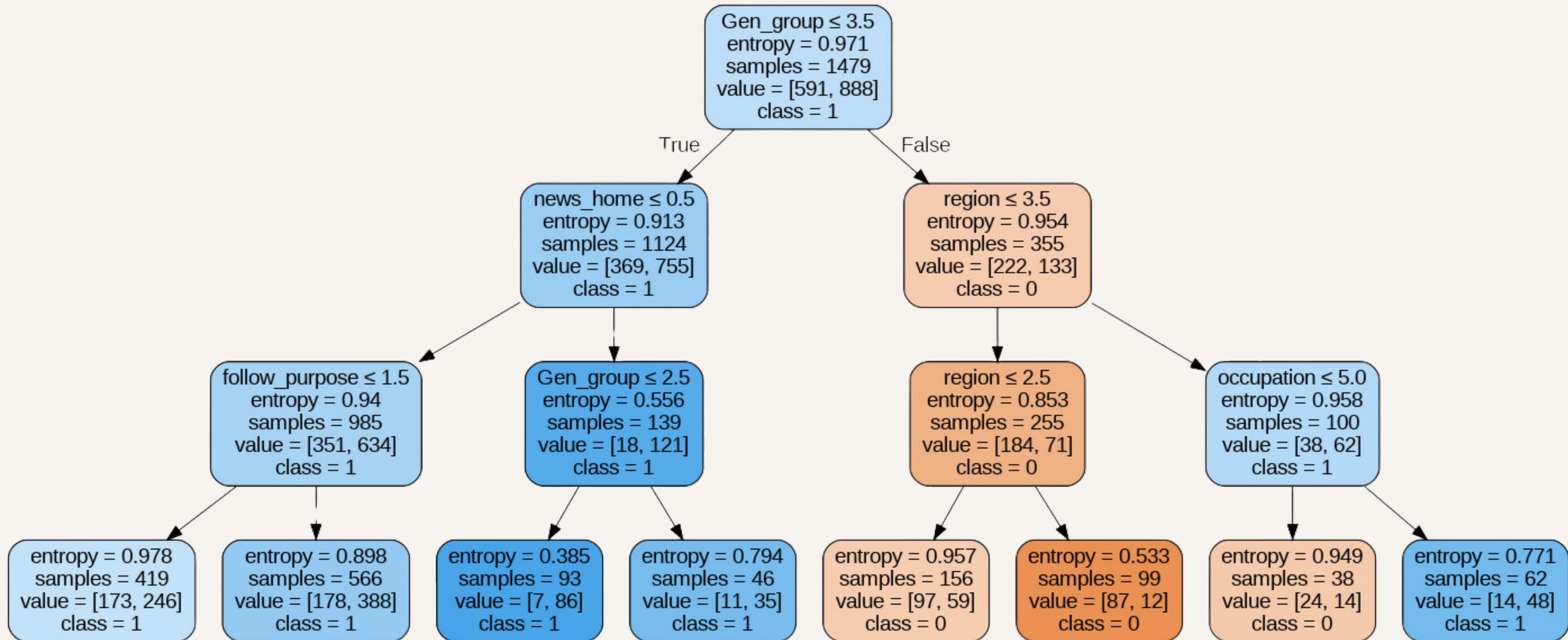
0.48

CLASS 1

0.78



DECISION TREE GRID SEARCH





DECISION TREE

GRID SEARCH

คนที่เลือก facebook เป็นอันดับ 1

คนที่อยู่เจน
X,Y หรือ Z

→ คนที่ติดตามหัววกีบ้าน → คนที่ติดตามหัว

คนที่อยู่เจน B → คนอยู่ภาคกลางหรือใต้ → อาชีพแม่บ้านหรือ
เกษตรกร



DECISION TREE

GRID SEARCH

คนที่ไม่ได้เลือก facebook เป็นอันดับ 1

คนที่อยู่ใน B → คนอยู่กรุงเทพ, ภาคเหนือ หรือภาคตะวันออกเฉียงเหนือ

คนที่อยู่ใน B → อาชีพรับราชการ, พนักงานเอกชน, เจ้าของธุรกิจ, รับจ้าง
หรือนักเรียนนักศึกษา

THANK YOU

