

เลขประจำตัว
ผู้โดยสาร

ผู้โดยสารชั้น: 1, 2, 3

ชื่อ
ผู้โดยสาร

เพศ

จำนวนพี่น้องที่โดยสาร
บนเรือ

อายุ

จำนวนผู้ปกครองที่โดยสาร
บนเรือ

ราคาตัว
โดยสาร

ท่าเรือที่ขึ้น

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	C

Titanic dataset

By Napatr Sarawanangkoor

เลขตัว
โดยสาร

หมายเลขห้อง
โดยสาร

ผลลัพธ์ที่เราอยากได้
ไม่รอด = 0
รอด = 1

Source: <https://www.kaggle.com/c/titanic/data>

My GitHub: https://github.com/LonelyFriday/titanic_treeClassification



Nikhil

@niquotein



only man who could have saved
titanic 🙋



Contents

- *Machine learning*
 - *Decision tree*
- *Evaluate*
 - *Confusion Matrix*
 - *Accuracy*
- *ETC ..*
 - *MORE Algorithms !!!!!*

ปูพื้นเบา ๆ

Goal: สร้าง Model Machine learning

Regression: 1, 2, 619.55, ...

Classification: "ใช่ / ไม่ใช่"

Date Flow:

- > Collect
- > Clean
- > Preprocessing
- > Analyze
- > Evaluate
- > Deploy

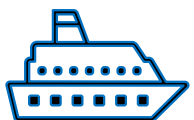
That one friend who will
be with you for lifetime



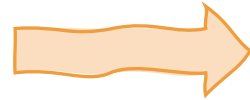


“ข้อมูล input”

- 1) index
- 2) PassengerId (เลขประจำตัวผู้โดยสาร)
- 3) Pclass (ผู้โดยสารชั้น: 1, 2, 3)
- 4) Name (ชื่อผู้โดยสาร)
- 5) Sex (เพศ)
- 6) Age (อายุ)
- 7) SibSp (จำนวนพี่น้องที่โดยสารบนเรือ)
- 8) Parch (จำนวนผู้ปกครองที่โดยสารบนเรือ)
- 9) Ticket (เลขตั๋วโดยสาร)
- 10) Fare (ราคาตั๋วโดยสาร)
- 11) Cabin (หมายเลขห้องโดยสาร)
- 12) Embark (ท่าเรือที่ขึ้น)



เลือกเฉพาะบาง Columns



1. *Pclass*

(ผู้โดยสารชั้น: 1, 2, 3)

2. *Sex*

(เพศ)

3. *Age*

(อายุ)

4. *SibSp*

(จำนวนพี่น้องที่โดยสารบนเรือ)

5. *Parch*

(จำนวนผู้ปกครองที่โดยสารบนเรือ)

6. *Fare*

(ราคาตั๋วโดยสาร)

7. *Embarked*

(ท่าเรือที่ขึ้น)

- The chosen one -





"How to หั่ง"

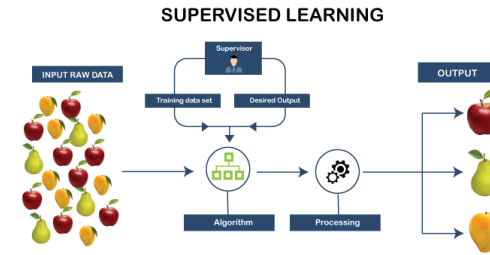
PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	0	3	male	34.5	0	0	7.8292	Q
1	1	3	female	47.0	1	0	7.0000	S
2	0	2	male	62.0	0	0	9.6875	Q
3	0	3	male	27.0	0	0	8.6625	S
4	1	3	female	22.0	1	1	12.2875	S

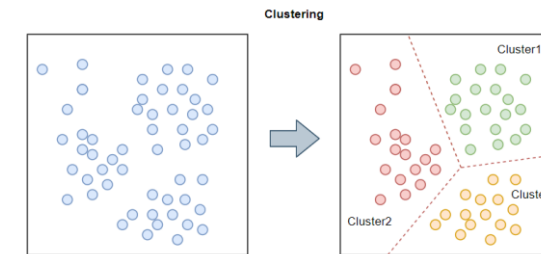


"Machine learning"

"Supervised learning"



"Unsupervised learning"



"Reinforcement learning"



Source 1: <https://www.tutorialandexample.com/wp-content/uploads/2020/11/Supervised-Machine-Learning-1.png>

Source 2: <https://www.ecloudvalley.com/wp-content/uploads/2019/09/Unsupervised-learning.png>

Source 3: https://www.mathworks.com/discovery/reinforcement-learning/_jcr_content/mainParsys3/discoverysubsection_603098216/mainParsys3/image.adapt.full.medium.png/1630398143703.png

Source 4: <https://estnn.com/wp-content/uploads/2019/04/Dota-2-Open-AI.jpg>



" Processing Data "

"Sex" : Male -> "0"
(เพศ) Female -> "1"

"Embark" : Cherbourg -> "0"
(ท่าเรือที่ขึ้น) Southampton -> "1"
Queenstown -> "2"

"Age" : 1 - 18 -> "0"
(อายุ) 19 - 35 -> "1"
36 - 60 -> "2"
60+ -> "3"

"Fare" : \$1 - \$8 -> "0"
(ราคาตั๋วโดยสาร) \$9 - \$16 -> "1"
\$17 - \$40 -> "2"
\$41+ -> "3"

** เปลี่ยนจากช่วง, คำ -> เป็นตัวเลข **

Groupby(Sex).mean() -> อัตราส่วนการรอดชีวิตของ (เพศหญิง)

" Bar plot "

" Pclass "

(ผู้โดยสารชั้น: 1, 2, 3)

" Sex "

(เพศ)

" Embark "

(ท่าเรือที่ขึ้น)

" SibSp "

(จำนวนพี่น้องที่โดยสารบนเรือ)

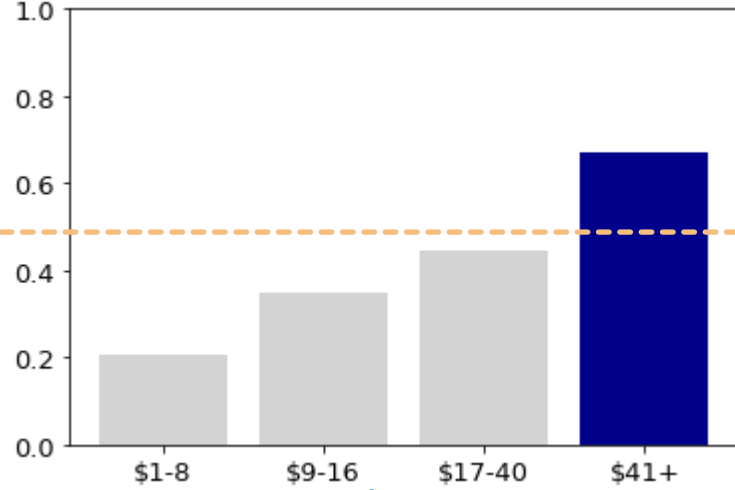
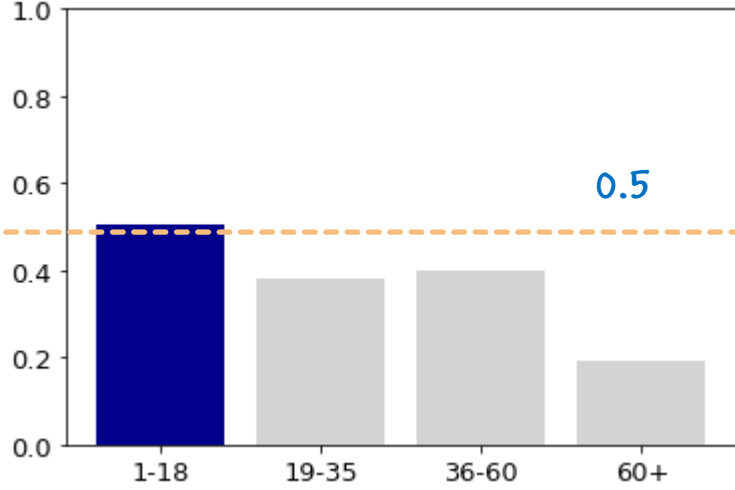
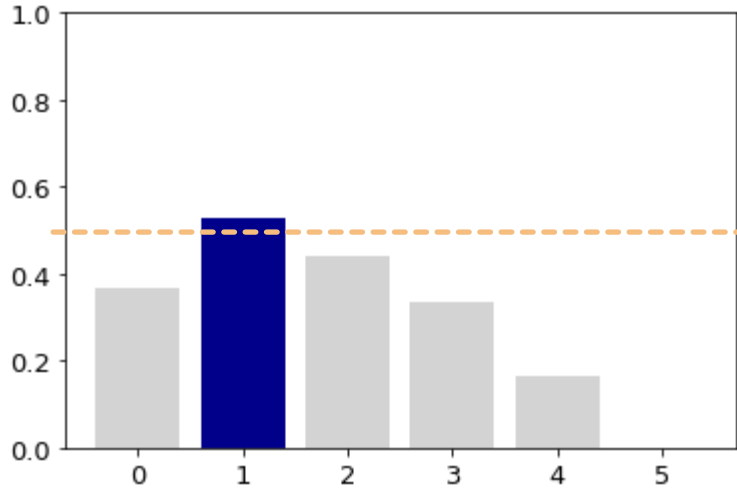
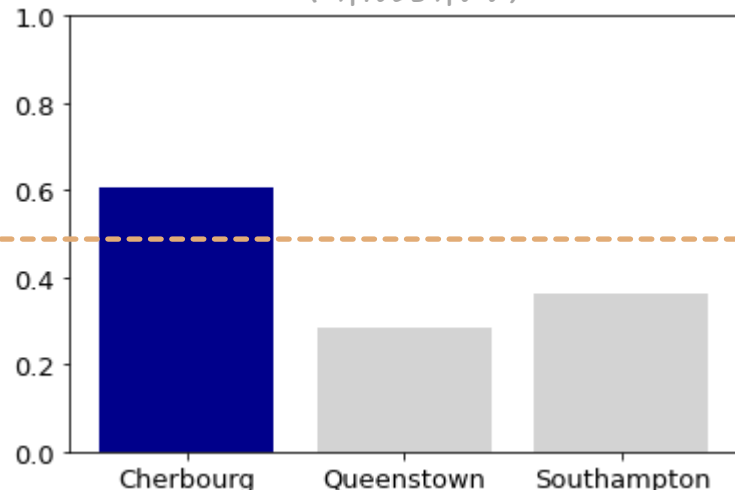
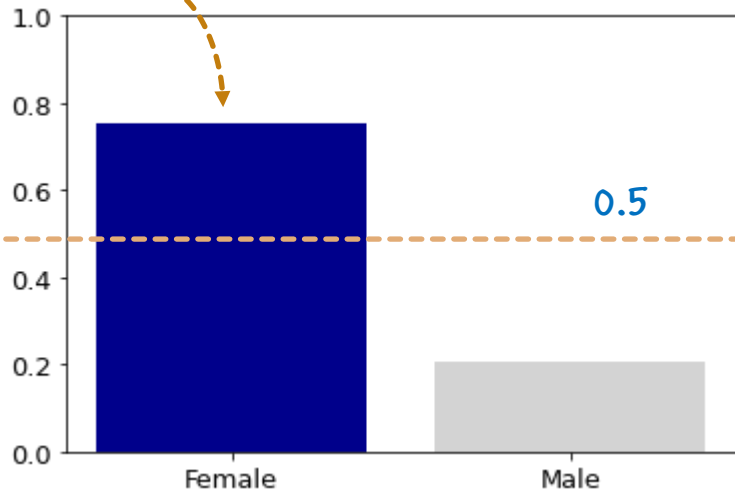
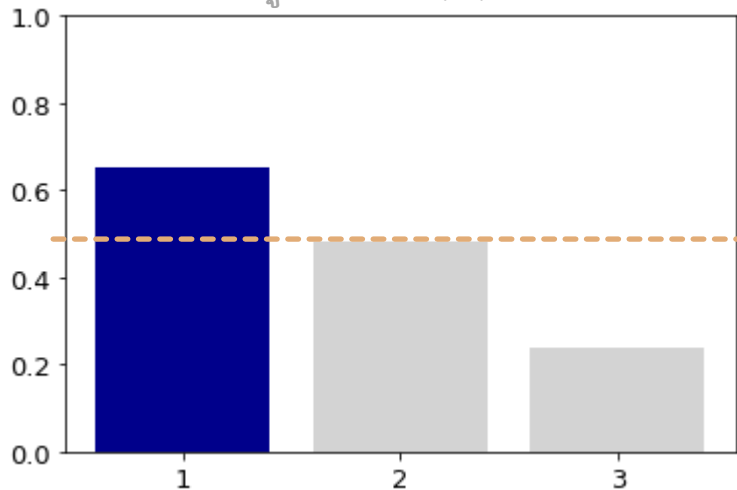
" Age "

(อายุ)

" Fare "

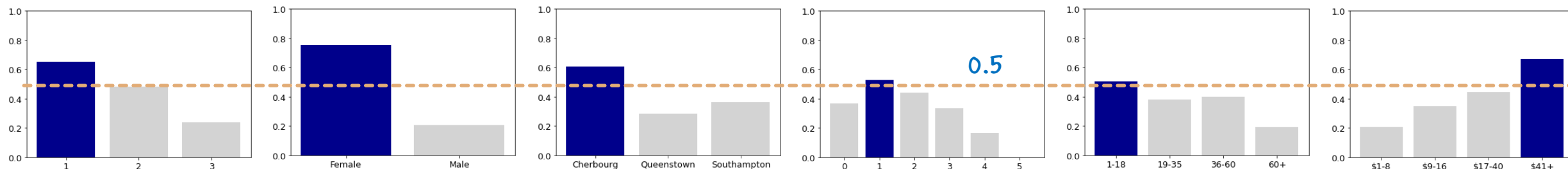
(ราคาตั๋วโดยสาร)

" Not made it = 0, Survive = 1 "



"Question ???"

"Not made it = 0, Survive = 1"



Pclass = 1

(ผู้โดยสารชั้น: 1, 2, 3)

Embarked = Southampton

(ท่าเรือที่ขึ้น)

Age = 35

(อายุ)

Sex = Male

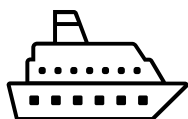
(เพศ)

SibSp = 0

(จำนวนพี่น้องที่โดยสารบนเรือ)

Fare = 35.5

(ราคาตั๋วโดยสาร)

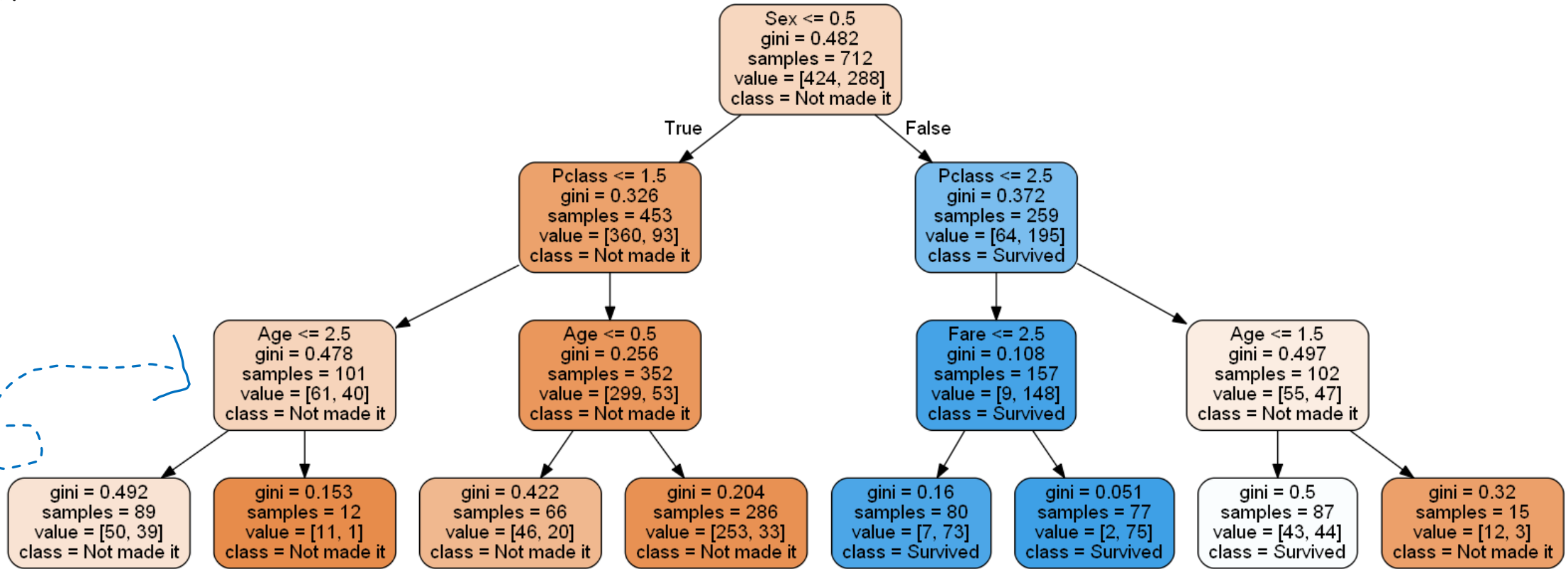


PassengerID = 24

Survived ???



" Decision Tree "

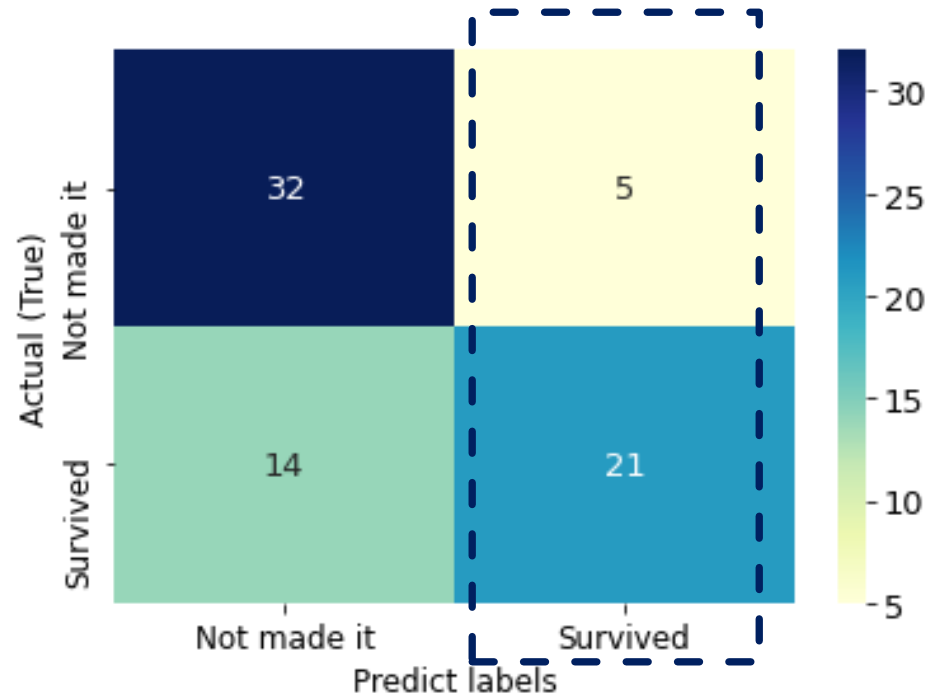


เงื่อนไข Age <= "36-60", Gini: ถ้าเป็น 0 คือ มีกลุ่มเดียว ↗

Sample (จำนวน) = 101 คน , Value (ผลการทำนาย): ['Not make it'=61, 'Survived'=40]



"Confusion Matrix"



```
tree_pred = decision_tree.predict(X_test)
value, counts = np.unique(tree_pred, return_counts=True)
pd.DataFrame(counts, index=['Not made it', 'Survived'], columns=['People'])
```

People	
Not made it	46
Survived	26



Accuracy =

21

32

True Positives + True Negatives

True Positives + True Negatives + False Positives + False Negatives

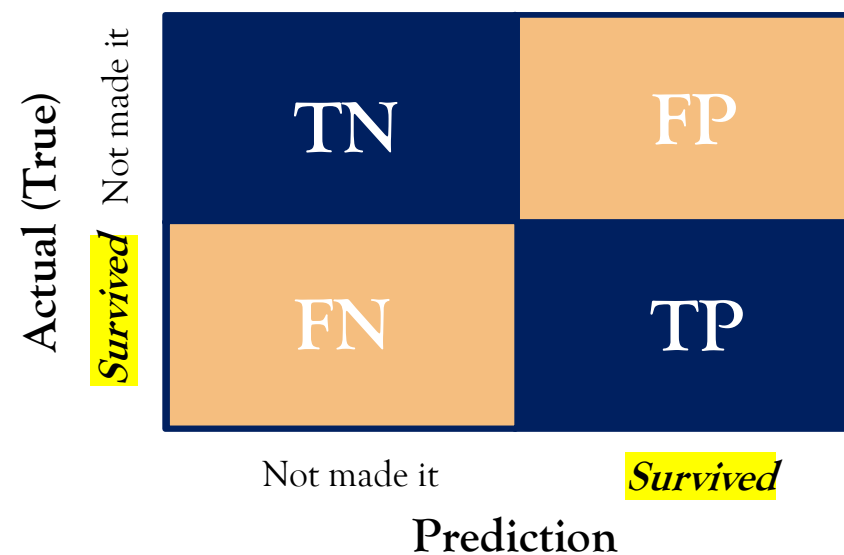
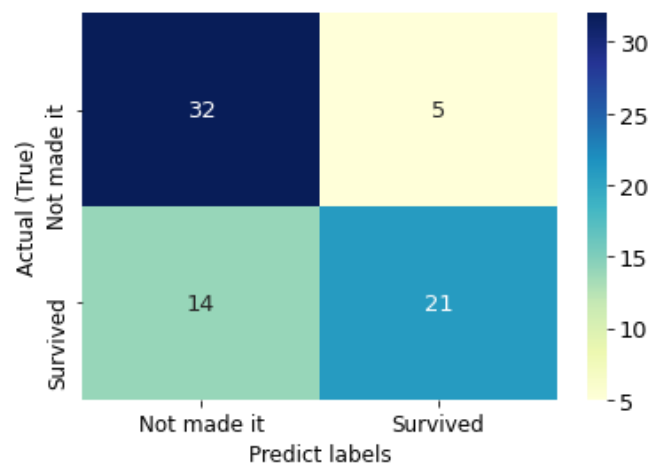
21

32

5

14

Confusion Matrix





"Confusion Matrix"

10

Actual (True)			
Survived	Not made it		
Prediction	Not made it	TN	FP
	Survived	FN	TP

=

Predicted Class

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN



Accuracy =

$$\frac{\overset{21}{\text{True Positives}} + \overset{32}{\text{True Negatives}}}{\underset{21}{\text{True Positives}} + \underset{32}{\text{True Negatives}} + \underset{5}{\text{False Positives}} + \underset{14}{\text{False Negatives}}}$$

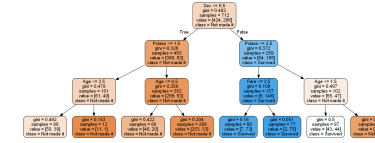
$$= \frac{21 + 32 = 53}{21 + 32 + 5 + 14 = 72}$$

$$= 0.7361$$

$$= \mathbf{73.61\%}$$

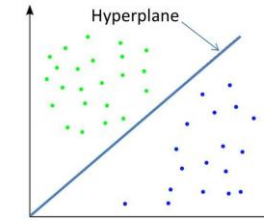


“Another algorithm” Decision Tree



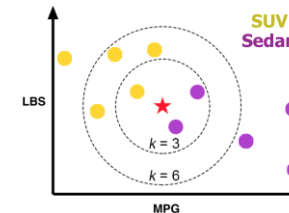
= **73.61** %

Support Vector Machine



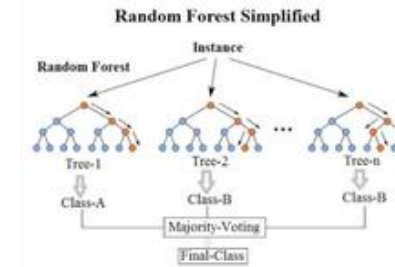
= **73.61** %

K-neighbors



= **77.78** %

Random Forest



= **79.17** %

Naïve bayes

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Likelihood of the Evidence given that the Hypothesis is True

Prior Probability of the Hypothesis

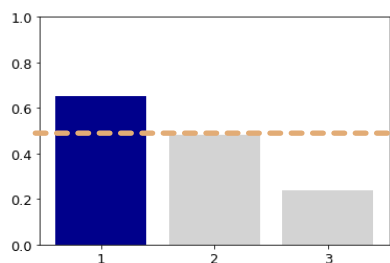
Posterior Probability of the Hypothesis given that the Evidence is True

Prior Probability that the evidence is True

= **76.39** %

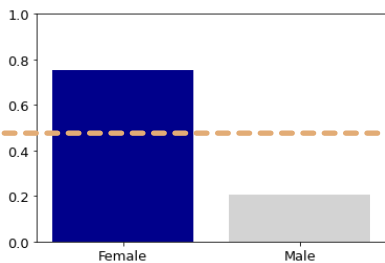
" โมเดลพยากรณ์ว่า .. Not make it "

" Not made it = 0, Survive = 1 "



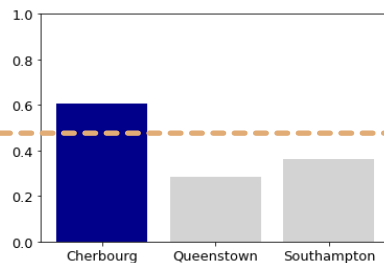
Pclass = 1

(ผู้โดยสารชั้น: 1, 2, 3)



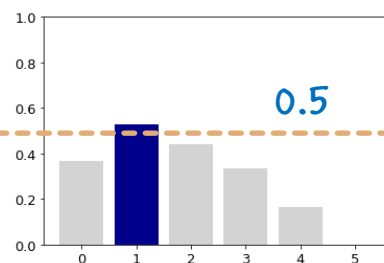
Sex = Male

(เพศ)



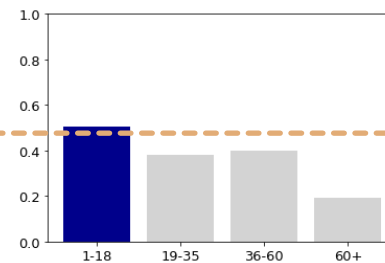
Embarked = Southampton

(ท่าเรือที่ขึ้น)



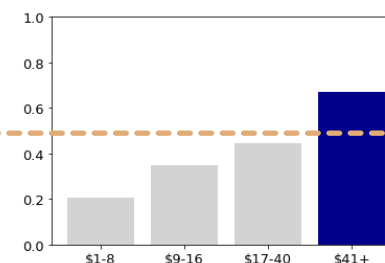
SibSp = 0

(จำนวนพี่น้องที่โดยสารบนเรือ)



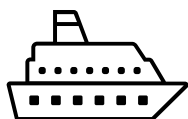
Age = 35

(อายุ)



Fare = 35.5

(ราคาตั๋วโดยสาร)



PassengerID = 24

But He is Survived

ไททานิก

 Data
TALK

