



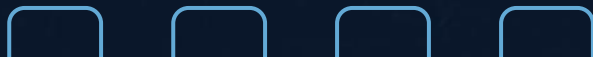
ADULT income CLASSIFICATION

01418496 Introduction to Machine Learning



OBJECTIVE

ทำนายรายได้ของแต่ละบุคคล โดยใช้ **K-NN** ในการทำนาย โดยเปรียบเทียบข้อมูลส่วนบุคคล
เช่น เพศ อายุ การศึกษา อาชีพ เป็นต้น



data set

ADULT income

500 Row

AGE
Workclass
fnlwgt
Education-num
marital-status
Occupation
Relationship

Race
Gender
capital-gain
capital-loss
hours-per-week
native-country
income

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	age	workclass	fnlwgt	education	educational-na	marital-status	occupation	relationship	race	gender	capital-gain	capital-loss	hours-per-week	native-country	income
2		25 Private	226802	11th		7 Never-married	Machine-op	Own-child	Black	Male	0	0	40	United-States	<=50K
3		38 Private	89814	HS-grad		9 Married-civ	Farming-fish	Husband	White	Male	0	0	50	United-States	<=50K
4		28 Local-gov	336951	Assoc-acadm		12 Married-civ	Protective-serv	Husband	White	Male	0	0	40	United-States	>50K
5		44 Private	160323	Some-college		10 Married-civ	Machine-op	Husband	Black	Male	7688	0	40	United-States	>50K
6		18 ?	103497	Some-college		10 Never-married	?	Own-child	White	Female	0	0	30	United-States	<=50K
7		34 Private	198693	10th		6 Never-married	Other-serv	Not-in-family	White	Male	0	0	30	United-States	<=50K
8		29 ?	227026	HS-grad		9 Never-married	?	Unmarried	Black	Male	0	0	40	United-States	<=50K
9		63 Self-employed	104626	Prof-school		15 Married-civ	Prof-specialty	Husband	White	Male	3103	0	32	United-States	>50K
10		24 Private	369667	Some-college		10 Never-married	Other-serv	Unmarried	White	Female	0	0	40	United-States	<=50K
11		55 Private	104996	7th-8th		4 Married-civ	Craft-repair	Husband	White	Male	0	0	10	United-States	<=50K
12		65 Private	184454	HS-grad		9 Married-civ	Machine-op	Husband	White	Male	6418	0	40	United-States	>50K
13		36 Federal-gov	212465	Bachelors		13 Married-civ	Adm-clerical	Husband	White	Male	0	0	40	United-States	<=50K
14		26 Private	82091	HS-grad		9 Never-married	Adm-clerical	Not-in-family	White	Female	0	0	39	United-States	<=50K
15		58 ?	299831	HS-grad		9 Married-civ	?	Husband	White	Male	0	0	35	United-States	<=50K
16		48 Private	279724	HS-grad		9 Married-civ	Machine-op	Husband	White	Male	3103	0	48	United-States	>50K
17		43 Private	346189	Masters		14 Married-civ	Exec-managerial	Husband	White	Male	0	0	50	United-States	>50K
18		20 State-gov	444554	Some-college		10 Never-married	Other-serv	Own-child	White	Male	0	0	25	United-States	<=50K
19		43 Private	128354	HS-grad		9 Married-civ	Adm-clerical	Wife	White	Female	0	0	30	United-States	<=50K
20		37 Private	60548	HS-grad		9 Widowed	Machine-op	Unmarried	White	Female	0	0	20	United-States	<=50K
21		40 Private	85019	Doctorate		16 Married-civ	Prof-specialty	Husband	Asian-Pac-Islander	Male	0	0	45	?	>50K
22		34 Private	107914	Bachelors		13 Married-civ	Tech-support	Husband	White	Male	0	0	47	United-States	>50K
23		34 Private	238588	Some-college		10 Never-married	Other-serv	Own-child	Black	Female	0	0	35	United-States	<=50K
24		72 ?	132015	7th-8th		4 Divorced	?	Not-in-family	White	Female	0	0	6	United-States	<=50K
25		25 Private	220931	Bachelors		13 Never-married	Prof-specialty	Not-in-family	White	Male	0	0	43	Peru	<=50K

ข้อมูลจาก: <https://www.kaggle.com/wenruihu/adult-income-dataset>

MODEL

ใช้ **K-NN** หรือ **K-NearestNeighbors** เป็นอัลกอริธึมในการทำนายผล โดยจะทำการ
เปรียบเทียบข้อมูลต่างๆของบุคคลนั้นๆ เช่น อายุ ระดับการศึกษา เพศ สถานะ สันติภาพที่มี
ที่ใช้ **K-NN** เพราะว่า ชุดข้อมูลสามารถใช้การเปรียบเทียบได้ โดยนำข้อมูลจาก ตัวที่ใช้ **Train**
มาเทียบกับตัว **Test** โดยเลือกตัวที่ใกล้เคียงที่สุดตามหลักการ ของ **K-NN**