

بخش پیاده‌سازی

سوال (۶)

ابتدا داده‌ها را می‌خوانیم و پس از آن ویژگی‌های بی‌اهمیت زیر را از داده‌ها حذف می‌کنیم:

```
X_train = drop_column(ship_df, ["Name", "Ticket", "Cabin", "Fare", "PassengerId"])
```

سپس لازم داریم که ویژگی‌های غیر عددی را با دسته‌بندی آن‌ها به ویژگی‌های عددی تبدیل کنیم.

در نهایت دادگان تمیز ما به این صورت خواهند بود:

In [8]: X_train

Out[8]:

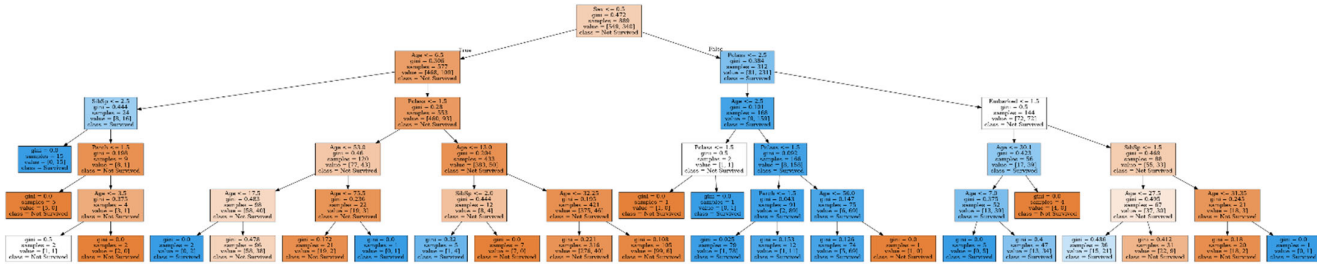
	Pclass	Sex	Age	SibSp	Parch	Embarked
0	3	0	22.000000	1	0	2
1	1	1	38.000000	1	0	0
2	3	1	26.000000	0	0	2
3	1	1	35.000000	1	0	2
4	3	0	35.000000	0	0	2
...
886	2	0	27.000000	0	0	2
887	1	1	19.000000	0	0	2
888	3	1	29.699118	1	2	2
889	1	0	26.000000	0	0	0
890	3	0	32.000000	0	0	1

889 rows × 6 columns

با استفاده از معیارهای جینی و آنتروپی و همچنین عمق‌های مختلف درخت تصمیم، نتیجه‌های مختلفی گرفته می‌شد. یکی از نتایج خوبی که من گرفتم با روش جینی و در حداکثر عمق ۵ اتفاق افتاد که نتیجه‌ای که از سایت Kaggle گرفتم به صورت زیر شد:

10 submissions for Ali Nazari		Sort by	Most recent
All	Successful	Selected	
Submission and Description		Public Score	
result4.csv 21 hours ago by Ali Nazari 'gini', 5		0.78468	

درخت تصمیم ما هم به شکل زیر در می‌آید:



سوال (۷)

در این سوال‌ها پس از دریافت دادگان، مشاهده کردیم که داده‌ها به این صورت هستند:

```
In [3]: heart_df = read_csv("./heart.csv")
heart_df
```

Out[3]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

303 rows × 14 columns

داده‌ها را به ۲ دسته‌ی تست و آموزش تقسیم کردیم و پس از آموزش مدل با استفاده از الگوریتم‌های نایو بیز و KNN، داده‌های تست را به این دو الگوریتم دادیم تا پیش‌بینی آن‌ها را ببینیم. که امتیازی که این دو مدل گرفتند را با هم می‌بینیم:

KNN:

```
In [9]: classifier1, model1, prediction1, result1 = train_knn_model(30, X_train, y_train, X_test, y_test)
score(model1, X_test, y_test)
```

Out[9]: 0.6557377049180327

نایو بیز:

```
In [11]: classifier2, model2, prediction2, result2 = train_GaussianNB_model(X_train, y_train, X_test, y_test)
score(model2, X_test, y_test)
```

Out[11]: 0.7868852459016393