

Dự đoán hành khách sống sót sau vụ đắm tàu Titanic bằng thuật toán Random Forest Classifier

Lê Khánh Hoàng

Mở đầu

Input

Dữ liệu hành khách gồm thông tin cá nhân và vé tàu từ Kaggle Titanic Dataset

Output

Dự đoán trạng thái sống sót của từng hành khách. (0 hoặc 1)

Tập dữ liệu

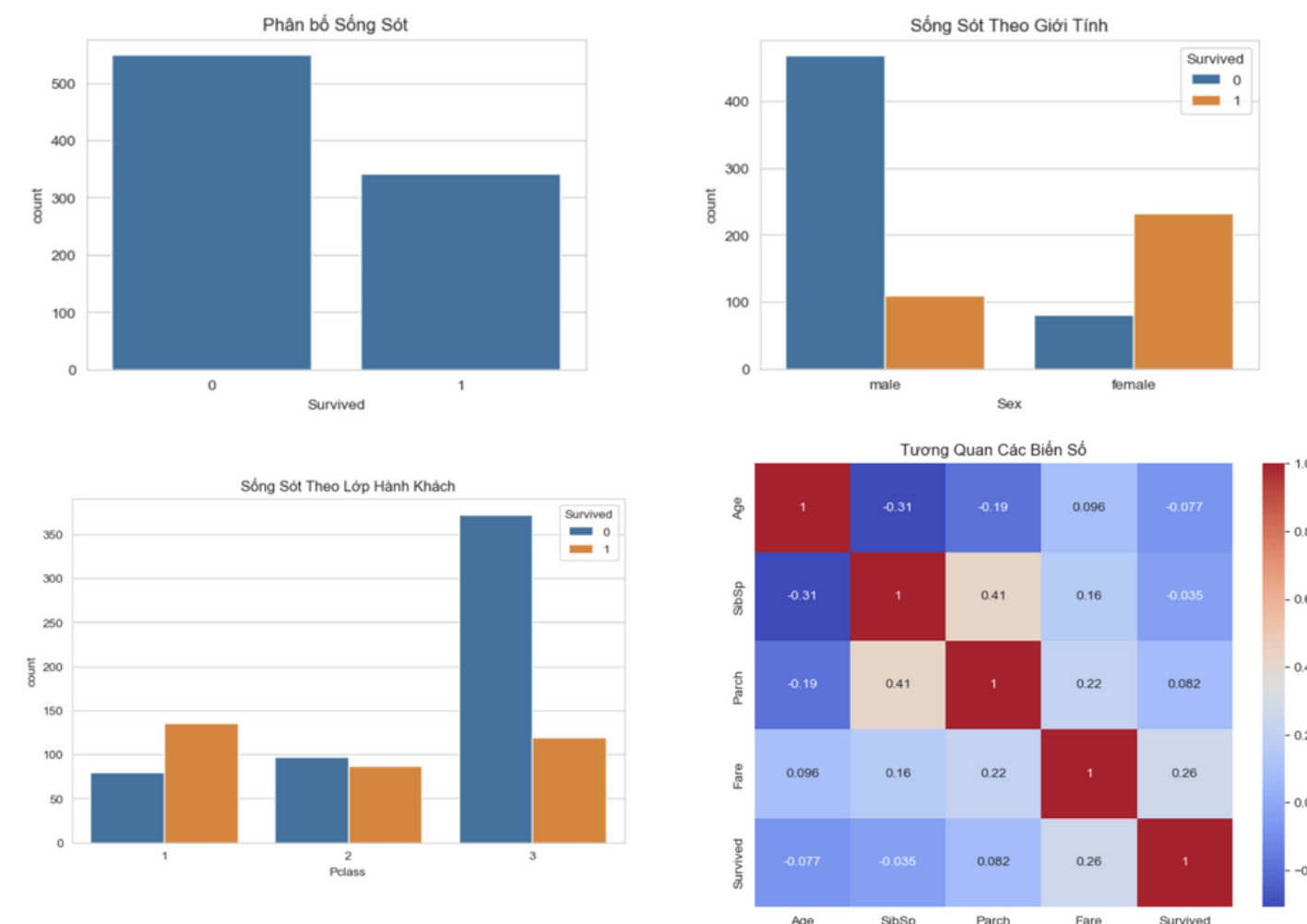
Nguồn dữ liệu:
<https://www.kaggle.com/competitions/titanic/data>

Phương pháp sử dụng

Random Forest Classifier

```
# Khởi tạo mô hình
model = RandomForestClassifier(n_estimators=100, random_state=42)
```

Exploratory Data Analysis (EDA)



Feature Engineering và Xử Lý Missing Values

Processed train shape: (891, 11)
Processed test shape: (418, 12)

First 5 rows of processed X:

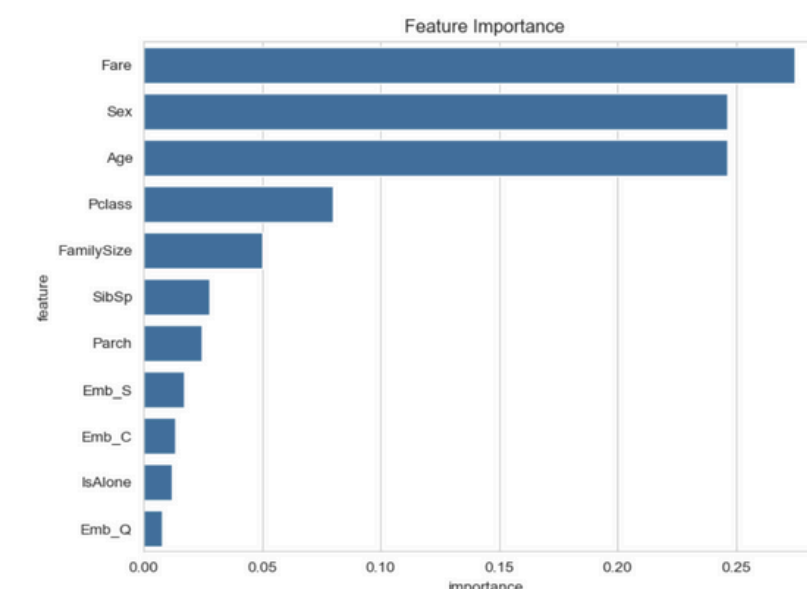
	Pclass	Sex	Age	SibSp	Parch	Fare	FamilySize	IsAlone	Emb_C	\
0	3	0	22.0	1	0	7.2500	2	0	False	
1	1	1	38.0	1	0	71.2833	2	0	True	
2	3	1	26.0	0	0	7.9250	1	1	False	
3	1	1	35.0	1	0	53.1000	2	0	False	
4	3	0	35.0	0	0	8.0500	1	1	False	

	Emb_Q	Emb_S
0	False	True
1	False	False
2	False	True
3	False	True
4	False	True

Model Building và Training

Validation Accuracy: 0.8156

Classification Report:				
	precision	recall	f1-score	support
0.0	0.85	0.85	0.85	110
1.0	0.76	0.75	0.76	69
accuracy			0.82	179
macro avg	0.81	0.80	0.80	179
weighted avg	0.82	0.82	0.82	179



Kết quả

Submission and Description		Public Score
✓ submission.csv	Complete · 4h ago	0.75837