

Airlines Customer Satisfaction

Nhập môn Trí tuệ nhân tạo

Group 16

- Đinh Ngọc Huân
- Đỗ Minh Hiệp
- Chu Đình Đức

Nội dung

- I. Tập dữ liệu
- II. Phương pháp
- III. Khó khăn
- IV. Phân công công việc

I. Tập dữ liệu

- Tập dữ liệu: tệp csv gồm 129880 dòng x 23 cột
- Nhãn: 1 - satisfied, 0 - dissatisfied
- Mục tiêu: Dự đoán mức độ hài lòng của 1 khách hàng tương lai thông qua 1 bộ giá trị của các thuộc tính
- Và đánh giá mức độ ảnh hưởng của từng thuộc tính tới mức độ hài lòng của khách hàng
- Link:

```
satisfaction
Gender
Customer Type
Age
Type of Travel
Class
Flight Distance
Seat comfort
Departure/Arrival time convenient
Food and drink
Gate location
Inflight wifi service
Inflight entertainment
Online support
Ease of Online booking
On-board service
Leg room service
Baggage handling
Checkin service
Cleanliness
Online boarding
Departure Delay in Minutes
Arrival Delay in Minutes
```

I. Tập dữ liệu

| | | | |
|-----------------------------------|-----|-----------------------------------|--------|
| satisfaction | 0 | satisfaction | 129880 |
| Gender | 0 | Gender | 129880 |
| Customer Type | 0 | Customer Type | 129880 |
| Age | 0 | Age | 129880 |
| Type of Travel | 0 | Type of Travel | 129880 |
| Class | 0 | Class | 129880 |
| Flight Distance | 0 | Flight Distance | 129880 |
| Seat comfort | 0 | Seat comfort | 129880 |
| Departure/Arrival time convenient | 0 | Departure/Arrival time convenient | 129880 |
| Food and drink | 0 | Food and drink | 129880 |
| Gate location | 0 | Gate location | 129880 |
| Inflight wifi service | 0 | Inflight wifi service | 129880 |
| Inflight entertainment | 0 | Inflight entertainment | 129880 |
| Online support | 0 | Online support | 129880 |
| Ease of Online booking | 0 | Ease of Online booking | 129880 |
| On-board service | 0 | On-board service | 129880 |
| Leg room service | 0 | Leg room service | 129880 |
| Baggage handling | 0 | Baggage handling | 129880 |
| Checkin service | 0 | Checkin service | 129880 |
| Cleanliness | 0 | Cleanliness | 129880 |
| Online boarding | 0 | Online boarding | 129880 |
| Departure Delay in Minutes | 0 | Departure Delay in Minutes | 129880 |
| Arrival Delay in Minutes | 393 | Arrival Delay in Minutes | 129487 |

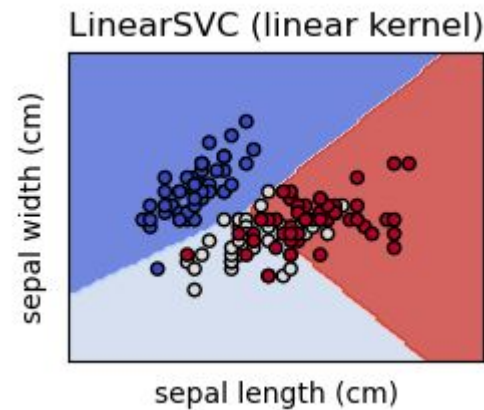
I. Tập dữ liệu

```
satisfaction      2      [satisfied, dissatisfied]
Gender            2      [Female, Male]
Customer Type     2      [Loyal Customer, disloyal Customer]
Age              75      [65, 47, 15, 60, 70, 30, 66, 10, 56, 22, 58, 3...
Type of Travel    2      [Personal Travel, Business travel]
Class            3      [Eco, Business, Eco Plus]
Flight Distance  5398    [265, 2464, 2138, 623, 354, 1894, 227, 1812, 7...
Seat comfort      6      [0, 1, 4, 5, 2, 3]
Departure/Arrival time convenient  6      [0, 1, 2, 3, 4, 5]
Food and drink    6      [0, 1, 2, 3, 4, 5]
Gate location     6      [2, 3, 4, 1, 5, 0]
Inflight wifi service  6      [2, 0, 3, 4, 5, 1]
Inflight entertainment  6      [4, 2, 0, 3, 5, 1]
Online support    6      [2, 3, 4, 5, 1, 0]
Ease of Online booking  6      [3, 2, 1, 5, 4, 0]
On-board service  6      [3, 4, 1, 2, 5, 0]
Leg room service  6      [0, 4, 3, 2, 5, 1]
Baggage handling  5      [3, 4, 1, 2, 5]
Checkin service   6      [5, 2, 4, 3, 1, 0]
Cleanliness       6      [3, 4, 1, 2, 5, 0]
Online boarding   6      [2, 3, 5, 4, 1, 0]
Departure Delay in Minutes  466    [0, 310, 17, 30, 47, 40, 5, 2, 34, 4, 13, 427,...
Arrival Delay in Minutes  472    [0.0, 305.0, 15.0, 26.0, 48.0, 23.0, 19.0, 2.0...
dtype: int64      dtype: object
```

II. Phương pháp

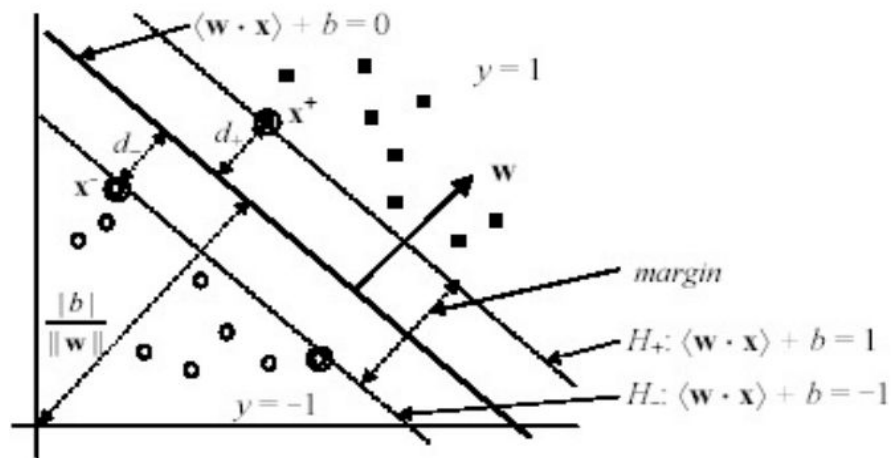
1. LinearSVC

| | precision | recall | f1-score | support |
|-----------------------------------|-----------|--------|----------|---------|
| 0 | 0.82 | 0.82 | 0.82 | 11749 |
| 1 | 0.85 | 0.85 | 0.85 | 14149 |
| accuracy | | | 0.83 | 25898 |
| macro avg | 0.83 | 0.83 | 0.83 | 25898 |
| weighted avg | 0.83 | 0.83 | 0.83 | 25898 |
| accuracy_score 0.8348134991119005 | | | | |



II. Phương pháp

1. LinearSVC



- Bài toán quy hoạch lồi
- Giải bằng định lý đối ngẫu + điều kiện KKT

Minimize

$$\frac{\langle \mathbf{w} \cdot \mathbf{w} \rangle}{2}$$

Conditioned on

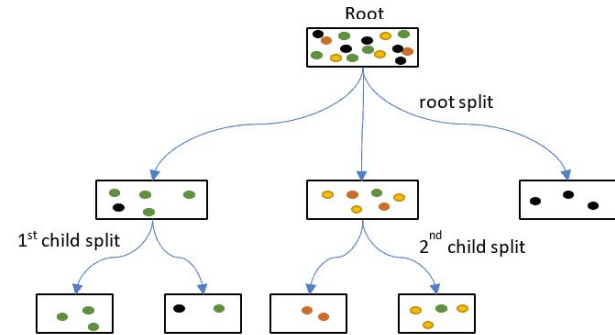
$$y_i(\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) \geq 1, \quad \forall i = 1..r$$

II. Phương pháp

2. CatBoost

- 2017

| | precision | recall | f1-score | support |
|----------------|--------------------|--------|----------|---------|
| 0 | 0.92 | 0.93 | 0.92 | 11548 |
| 1 | 0.94 | 0.93 | 0.94 | 14350 |
| accuracy | | | 0.93 | 25898 |
| macro avg | 0.93 | 0.93 | 0.93 | 25898 |
| weighted avg | 0.93 | 0.93 | 0.93 | 25898 |
| accuracy_score | 0.9315391149895745 | | | |



III. Khó khăn

- Lựa chọn tham số phù hợp với tập dữ liệu

IV. Phân công công việc

- Đinh Ngọc Huân: tiền xử lý dữ liệu
- Đỗ Minh Hiệp: triển khai model LinearSVC
- Chu Đình Đức: triển khai model CatBoost