

Airlines Customer Satisfaction

Nhập môn Trí tuệ nhân tạo

Group 16

Thành viên:

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

Nội dung

1. Giới thiệu đề tài
2. Phương pháp áp dụng
3. Xây dựng chương trình
4. Kết luận
5. Thông tin khác

1. Giới thiệu đề tài

- Đặt vấn đề: Sự hài lòng của khách hàng đóng vai trò chủ chốt đối với thành công của các hãng hàng không
- Đề tài: Đánh giá mức độ hài lòng của hành khách đi máy bay
- Mô tả đề tài: Bài toán dự đoán khách hàng tương lai sẽ hài lòng với dịch vụ của hãng hàng không này hay không dựa vào dữ liệu của chuyến bay và phản hồi của khách hàng

Attributes	Gender	Customer Type	Age	Type of Travel	...	satisfaction
Example	Female	Loyal Customer	65	Personal Travel	...	satisfied

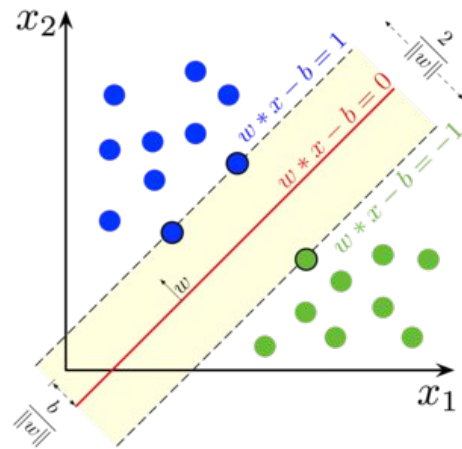
x y

2. Phương pháp áp dụng

Phương pháp 1: LinearSVC - mô hình SVM gốc

- LinearSVC học ra một siêu phẳng từ tập dữ liệu mà có thể phân chia tốt tập dữ liệu thành 2 phần
- Cụ thể là sử dụng lý thuyết đối ngẫu trong tối ưu để học ra siêu phẳng H mà đủ với mức lề lớn nhất

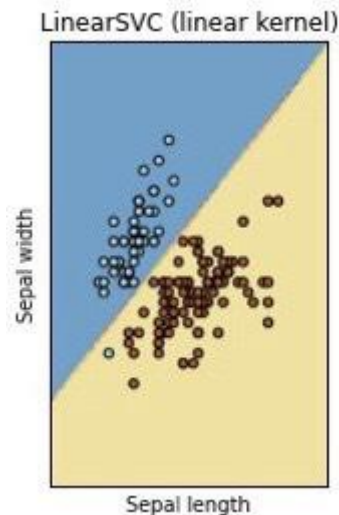
$$\max \frac{2}{\|w\|} \Leftrightarrow \min \frac{\|w\|}{2} \text{ conditioned on: } y_i(\langle w, x_i \rangle + b) \geq 1$$



2. Phương pháp áp dụng

Phương pháp 1: LinearSVC - mô hình SVM gốc

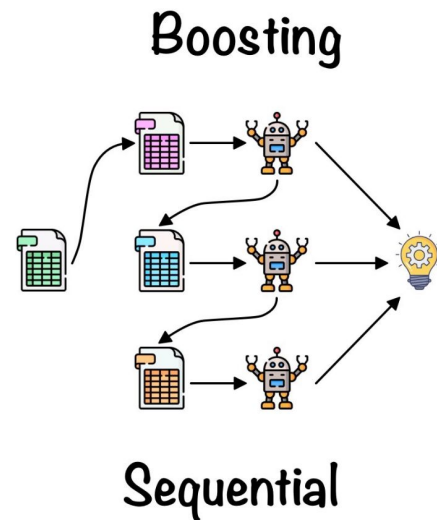
- Ưu điểm:
 - Hiệu quả trong không gian chiều cao
 - Hiệu quả với cả bài toán có số chiều lớn hơn số mẫu quan sát
 - Hiệu quả về bộ nhớ do chỉ sử dụng một tập nhỏ các mẫu quan sát (support vectors) làm bộ phân loại
 - Sử dụng lý thuyết đối ngẫu mạnh mẽ trong tối ưu
- Nhược điểm:
 - Khó khăn trong việc lựa chọn regulation để tránh overfit



2. Phương pháp áp dụng

Phương pháp 2: CatBoost - biến thể mới của Gradient Boosting

- Boosting là một dạng ensemble learning, lớp phương pháp này tổng hợp nhiều model (cây quyết định) yếu thành một model mạnh hơn
- Cây phía sau sẽ được học tiếp từ cây phía trước với mong muốn sẽ khắc phục được yếu điểm của cây trước đó
- Cuối cùng chúng ta tổng hợp kết quả của tất cả các cây để thu được kết quả cuối cùng



2. Phương pháp áp dụng

Phương pháp 2: CatBoost - biến thể mới của Gradient Boosting

Gradient boosting



2. Phương pháp áp dụng

Phương pháp 2: CatBoost - biến thể mới của Gradient Boosting

Cải tiến của CatBoost so với Gradient Boosting:

- Cây cân bằng
- Boosting có thứ tự
- Hỗ trợ thuộc tính gốc

2. Phương pháp áp dụng

Lý do lựa chọn phương pháp: Do hiện tại chưa biết có thể tìm được một siêu phẳng phân chia tốt tập dữ liệu thành 2 phần (hài lòng và không hài lòng) hay không, chúng em lựa chọn mô hình LinearSVC trong trường hợp tìm được siêu phẳng như vậy và trong trường hợp ngược lại chúng em lựa chọn CatBoost.

3. Xây dựng chương trình

Chương trình được xây dựng bằng framework Streamlit - framework chuyên dùng cho các model học máy

- Input: Giá trị của các thuộc tính đầu vào
- Output: Kết quả dự đoán của 2 model

Airlines Customer Satisfaction

INPUT

Gender Female	Customer Type Loyal Customer	Age 0 65 130	Type of Travel Personal Travel
Class Eco	Flight Distance 265 - +	Seat comfort 0	De/Ar time convenient 0
Food and drink 0	Gate location 2	Inflight wifi service 2	Inflight entertainment 4
Online support 2	Ease of Online booking 3	On-board service 3	Leg room service 0
Baggage handling 3	Checkin service 5	Cleanliness 3	Online boarding 2
Departure Delay in Minutes 0 - +	Arrival Delay in Minutes 0 - +		

OUTPUT

LinearSVC

Satisfied



CatBoost

Satisfied



4. Kết quả và hướng phát triển

Kết quả:

- Huấn luyện 2 model LinearSVC, CatBoost để so sánh kết quả
- Độ chính xác trên tập test của LinearSVC là 0.83 và CatBoost là 0.93
- Xây dựng thành công chương trình với giao diện đẹp mắt, dễ hiểu

Hướng phát triển:

- Thử nghiệm thêm các phương pháp khác
- Chuyển từ localhost thành host
- Thu thập, tăng cường dữ liệu

4. Kết quả và hướng phát triển

- Mô hình 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			

4. Kết quả đạt được

- Mô hình CatBoost

	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			

5. Thông tin khác

- Phân công công việc

Thành viên	Công việc	Mức độ hoàn thành
Đinh Ngọc Huân	Tiền xử lý dữ liệu Viết báo cáo	100%
Đỗ Minh Hiệp	Triển khai mô hình LinearSVC Viết báo cáo	100%
Chu Đình Đức	Triển khai mô hình CatBoost Viết báo cáo, xây dựng chương trình	100%

THANK YOU