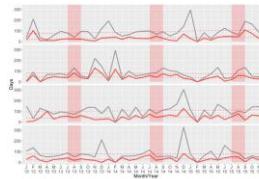
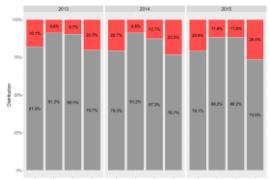


# Ứng dụng các kỹ thuật khai phá dữ liệu vào cho bộ dữ liệu đặt phòng khách sạn

## Nguyễn Đăng Tiến

### 1. Giới thiệu

#### Đặt vấn đề:



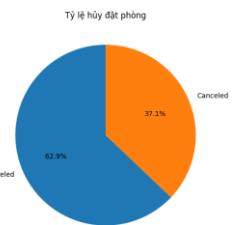
- Tỷ lệ hủy phòng cao theo từng năm
- Nhu cầu đặt phòng theo mùa
- Quyết định truyền thống theo kinh nghiệm

#### Mục tiêu:

- Dự đoán khả năng hủy phòng
- Phân nhóm hành vi khách hàng
- Phát hiện các luật dịch vụ - phòng
- Dự báo nhu cầu ngắn hạn

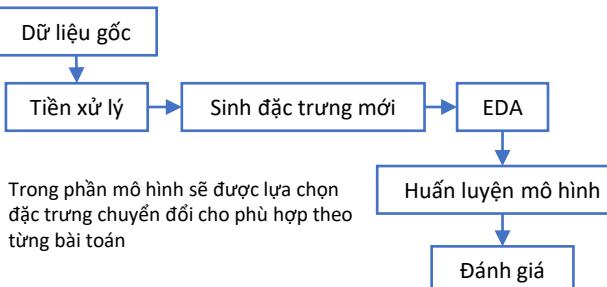
### 2. Dữ liệu

- Mẫu khoảng **119.390** xuất hiện cung **32** đặc trưng
- Loại dữ liệu:** Thời gian, giá, dịch vụ, hành vi
- Nguồn dữ liệu:** Kaggle ([Hotel booking demand | Kaggle](#))



- Biến mục tiêu:** is\_canceled  
Gồm hai nhãn **Hủy (1)** và **Không hủy (0)**. Với nhãn bằng 1 chiếm **37%** (một tỷ lệ đáng báo động)

### 3. Quy trình khai phá dữ liệu

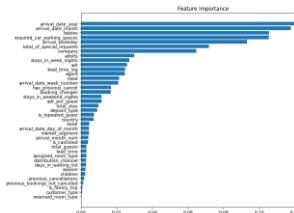


### 4. Dự đoán hủy phòng

- Mô hình:**
  - Logistic Regression
  - Decision Tree
  - Random Forest
- Xử lý mất cân bằng:**
  - Class\_weight = "balance"

Mô hình	Accuracy	Recall	F1 – Score	AUC - ROC
Logistic Regression	0.77	0.79	0.78	0.86
Random Forest	0.77	0.77	0.75	0.88
Decision Tree	0.67	0.67	0.68	0.8

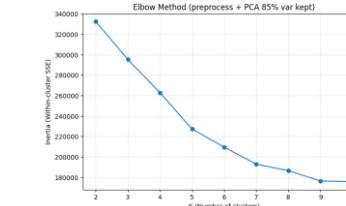
**Kết quả:** Logistic Regression được lựa chọn cho việc dự đoán khả năng hủy phòng ban đầu do có Recall cao nhất và Random Forest được sử dụng để xác định các yếu tố quan trọng (AUC - ROC)



Yếu tố thời gian và nhu cầu cụ thể của khách ảnh hưởng mạnh hơn giá phòng.

### 5. Phân cụm khách hàng

- Mô hình:**
  - K-Means
- K tốt nhất:**
  - 6
- Lựa chọn:**
  - Elbow + Silhouette
- Các đặc trưng được lựa chọn:**



### 6. Phân khúc khách hàng

#### Khách công tác ngắn ngày, ổn định quanh năm

- Thời gian lưu trú ngắn
- ít bị ảnh hưởng bởi mùa vụ

#### Khách theo tour đón sớm, tránh mùa đông

- Thời gian đặt trước dài
- Tránh mùa đông

#### Khách đại trà nhạy với marketing mùa vụ

- Đặt theo mùa cao và thấp điểm

#### Khách du lịch cao điểm

- Tập trung du lịch vào hè và lễ

#### Gia đình nhóm nhỏ nghỉ hè

- Gắn chặt với nghỉ hè
- Thời gian lưu trú dài hơn trung bình

#### Linh hoạt lắp đầy mùa thấp điểm

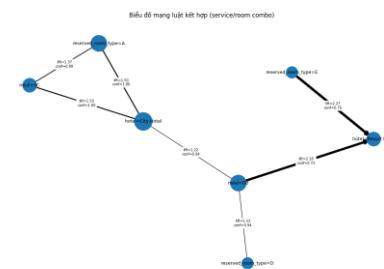
- Linh hoạt về thời gian

### 7. Luật kết hợp

#### 2 trong 7 luật mạnh:

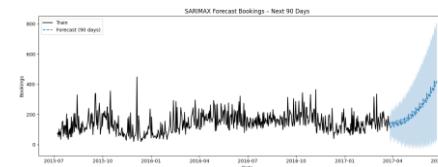
{ Hotel = City Hotel, Meal = SC -> Room = A }

{ Hotel = Resort Hotel, Meal = BB -> Room = E }



Nhu cầu đặt phòng có tính mùa vụ rõ rệt theo tháng và có thể dự báo trong ngắn hạn, nhưng mức độ bất định tăng nhanh theo thời gian.

### 8. Chuỗi thời gian



### 8. Kết luận

- Khai phá dữ liệu giúp **dự đoán – giải thích – hỗ trợ ra quyết định**
- Hiệu quả hơn phương pháp dựa trên kinh nghiệm

#### Ứng dụng:

- Giảm overbooking
- Cá nhân hóa giá & khuyến mãi
- Thiết kế combo dịch vụ
- Lập kế hoạch vận hành theo mùa

Việc kết hợp nhiều kỹ thuật khai phá dữ liệu mang lại góc nhìn toàn diện và giá trị thực tiễn trong quản lý khách sạn.