

Môn: Trực Quan Hóa Dữ Liệu

Câu Hỏi Cá Nhân

30/04/2022

Dữ liệu

Tập dữ liệu [FlightDelay.xlsx](#) là lịch sử các chuyến bay. Tập dữ liệu bao gồm 10 biến

Biến	Mô tả
Day of Week	Coded as 1 = Monday, 2 = Tuesday,..., 7 = Sunday
Origin	Three airport codes: DCA (Reagan National), IAD (Dulles), BWI (Baltimore–Washington Int'l)
Destination	Three airport codes: JFK (Kennedy), LGA (LaGuardia), EWR (Newark)
Carrier	Eight airline codes: CO (Continental), DH (Atlantic Coast), DL (Delta), MQ (American Eagle), OH (Comair), RU (Continental Express), UA (United), and US (USAirways)
Weather	Coded as 1 if there was a weather-related delay
Departure Time	Coded as “hhmm” or “hmm”
Distance	
FL_DATE	Flight date
FL_NUM	Flight number
Flight status	ontime or delayed

Yêu cầu

1. Hãy trực quan hóa các thông tin thống kê mô tả cho các biến.
2. Tìm và trực quan mối quan hệ *tương quan* giữa các cặp biến (nếu có).
3. Hãy trực quan hóa tỉ lệ trễ máy bay theo các biến

- Day of Week
- Origin
- Destination
- Carrier
- Weather

4. Hãy đưa ra mô hình dự báo về khả năng máy bay bị trễ (có thể sử dụng mô hình hồi quy logistic hoặc mô hình học máy bất kỳ).

Lưu ý

Sinh viên làm

- Bằng **notebook** (colab hoặc jupyter, ...)
- Lưu thành tập tin [Student ID].pdf
- Nộp tập tin [Student ID].pdf qua Moodle