**BÁO CÁO CÔNG VIỆC ĐÃ THỰC HIỆN**

**Họ và tên:** Lê Thị Như Quỳnh  
**Mã số sinh viên:** 2251262632  
**Bài thực hành:** Bài thực hành 1 – Phân tích chuỗi thời gian (Time Series)  
**Đề bài:** Đề số 4

**1. Mục tiêu**

Phân tích chuỗi thời gian liên quan đến dữ liệu nghe podcast theo ngày thứ 4, tập trung vào trường thông tin “Listening\_Time\_minutes” nhằm áp dụng các kỹ thuật xử lý và dự báo chuỗi thời gian.

**2. Các bước thực hiện**

**Bước 1: Chuẩn bị dữ liệu:**

* import các thư viện cần thiết như pandas, numpy, matplotlib, seaborn để phục vụ cho việc xử lý và trực quan hóa dữ liệu.
* Dữ liệu được tải lên từ lên Google Colab.

**Bước 2: Tiền xử lý dữ liệu:**

* Lọc dữ liệu chỉ giữ lại những bản ghi có ngày phát hành là “thứ 4” và thời gian nghe (Listening\_Time\_minutes) không bị thiếu.
* Dữ liệu được sắp xếp theo id để đảm bảo thứ tự thời gian, sau đó reset chỉ số dòng để dễ theo dõi.

**Bước 3: Cài đặt thư viện**

* Đã cài đặt thư viện pykalman, dùng cho mô hình lọc Kalman, hỗ trợ xử lý và dự báo chuỗi thời gian với nhiễu.

**Bước 4: Áp dụng mô hình**

* Thực hiện mô hình lọc Kalman để làm trơn chuỗi thời gian.
* Trực quan hóa kết quả bằng biểu đồ, giúp dễ so sánh giữa dữ liệu gốc và dữ liệu đã được làm trơn

**3. Kết quả đạt được:**

- Đã xử lý thành công dữ liệu nghe nhạc theo ngày thứ 4, loại bỏ các giá trị thiếu và đảm bảo tính tuần tự.

- Mô hình Kalman đã được áp dụng để làm mượt chuỗi thời gian, thể hiện xu hướng biến đổi rõ ràng hơn so với dữ liệu gốc.

- Biểu đồ trực quan giúp nhận biết rõ sự khác biệt giữa chuỗi ban đầu và chuỗi đã xử lý.