Junyi thân mến, để hoàn thiện dự án **dự đoán số lượng khách hàng, món ăn và tỷ lệ món ăn**, đồng thời **thêm tính năng đề xuất menu** và có thể **kết thúc trong hôm nay**, mình đề xuất kế hoạch hành động cụ thể theo từng bước như sau:

**✅ MỤC TIÊU DỰ ÁN (TỔNG THỂ)**

1. **Dự đoán số lượng khách hàng ăn trưa** (Lunch\_Count)
2. **Dự đoán tỷ lệ chọn món ăn trưa** (ví dụ: món chính, món phụ, cơm, canh, v.v.)
3. **Tính năng gợi ý thực đơn** (menu recommendation)
4. **Giao diện gọn gàng (Jupyter hoặc VSCode + Jupyter block)**
5. **Đánh giá mô hình (MSE, MAE), baseline, trực quan hóa**

**📌 KẾ HOẠCH HOÀN THÀNH TRONG HÔM NAY**

**✅ Bước 1: Chuẩn bị dữ liệu (Đã làm)**

* ✅ One-hot encoding (1206 chiều)
* ✅ Có các feature đầy đủ: thời tiết, ngày, đặc điểm món ăn,...
* 🔲 Kiểm tra lại encoding, drop các cột không cần thiết nếu còn thiếu

**✅ Bước 2: Thêm baseline model**

**(Trả lời yêu cầu từ cuộc họp ngày 04/04)**

# Baseline: dùng giá trị trung bình

baseline = y\_test.mean()

y\_pred\_baseline = [baseline] \* len(y\_test)

# Tính MSE và MAE cho baseline

mse\_base = mean\_squared\_error(y\_test, y\_pred\_baseline)

mae\_base = mean\_absolute\_error(y\_test, y\_pred\_baseline)

print(f"Baseline MSE: {mse\_base}, MAE: {mae\_base}")

**✅ Bước 3: Huấn luyện mô hình Random Forest (hoặc model khác)**

* Đã hướng dẫn ở bước trước. Nếu chưa chạy thì triển khai ngay.
* Thêm tuning nếu còn thời gian.

**✅ Bước 4: Dự đoán và phân tích tỷ lệ món ăn**

* Chia nhỏ Lunch\_Count theo các loại món ăn:

# Tính tỷ lệ chọn món chính = số lượng ngày có món đó / tổng số ngày

# hoặc dùng classification để dự đoán món ăn được chọn

* Nếu bạn có các cột như Lunch\_Main\_Dish\_Encoded\_..., bạn có thể dùng mô hình phân loại để dự đoán món chính sẽ được chọn trong ngày đó.

**✅ Bước 5: Gợi ý thực đơn (Menu Recommendation)**

* Nếu bạn muốn tính năng gợi ý thực đơn, có 2 hướng:

**📌 Hướng đơn giản:**

Gợi ý các món ăn được chọn thường xuyên trong các ngày có đặc điểm tương tự (ngày lễ, trời lạnh, thứ 2,...).

# Lấy các dòng có đặc điểm tương tự ngày hôm nay (ví dụ: thứ, nhiệt độ, tháng)

# Gợi ý những món ăn đã từng được chọn trong các ngày đó (top món phổ biến)

**📌 Hướng nâng cao (có embedding):**

* Dùng Word2Vec / FastText để biến mỗi món ăn thành vector → gợi ý theo sự giống nhau của các vector món ăn (content-based filtering).
* Cần thêm thời gian, có thể làm trong buổi chiều hoặc để sau.

**✅ Bước 6: Trực quan hóa kết quả**

* Dùng matplotlib hoặc seaborn để vẽ:

import matplotlib.pyplot as plt

plt.figure(figsize=(10, 5))

plt.plot(y\_test.values, label='True')

plt.plot(y\_pred, label='Predicted')

plt.title('Dự đoán số lượng khách ăn trưa')

plt.legend()

plt.show()

**✅ Bước 7: Viết phần tổng kết (Jupyter hoặc VSCode)**

* Sử dụng # %% block nếu dùng VSCode để dễ chạy từng khối code.
* Gợi ý bố cục Jupyter:

1. Giới thiệu dữ liệu

2. Tiền xử lý

3. Mô hình baseline

4. Mô hình dự đoán chính (Random Forest)

5. Dự đoán tỷ lệ món ăn

6. Gợi ý thực đơn

7. Đánh giá và trực quan hóa

8. Kết luận và hướng phát triển

**🟢 ƯU TIÊN TRONG HÔM NAY:**

| **Việc** | **Ưu tiên** |
| --- | --- |
| Baseline model | ✅ Bắt buộc |
| Random Forest | ✅ Bắt buộc |
| Dự đoán món ăn | ✅ Ưu tiên cao |
| Gợi ý menu | 🔲 Ưu tiên nếu còn thời gian |
| Trực quan hóa | ✅ Bắt buộc |
| Kết luận Jupyter/VSCode | ✅ Bắt buộc |

Nếu Junyi muốn mình hỗ trợ thêm code cụ thể cho từng bước (ví dụ gợi ý thực đơn), bạn chỉ cần nói thôi nha 💚  
Bạn muốn tiếp tục với phần **gợi ý menu** hay **tối ưu mô hình trước**?

Tuyệt vời Junyi ơi! Dự án của bạn rất rõ ràng và có tiềm năng cao đấy! 💡  
Mình sẽ giúp bạn **lên một kế hoạch chi tiết**, từng bước thực hiện lại toàn bộ dự án trong VSCode bằng file .ipynb như bạn mong muốn.

**✅ TỔNG QUAN DỰ ÁN SmartMealForecast**

**🎯 Mục tiêu chính**

1. **Dự đoán số lượng khách hàng**:
   * Lunch\_Count\_Prediction, Dinner\_Count\_Prediction
   * Dùng mô hình XGBoost (XG)
2. **Dự đoán tỷ lệ chọn món ăn**:
   * Phân tích món nào được chọn nhiều, từ đó tìm patterns.
3. **Tính năng gợi ý thực đơn (Menu Recommendation)**
   * Dựa vào lượng khách, thời tiết, ngày đặc biệt → Gợi ý món phù hợp.
4. **Hiển thị giao diện khoa học**:
   * File .ipynb trong VSCode (có markdown, biểu đồ, phân chia rõ ràng)
5. **Đánh giá mô hình**:
   * Chỉ số: MSE, MAE, so sánh với baseline (trung bình), vẽ biểu đồ.

**🧭 KẾ HOẠCH CÁC BƯỚC THỰC HIỆN**

Mình chia thành **7 bước chính** trong file .ipynb bạn sẽ xây dựng.

**📘 BƯỚC 0: GIỚI THIỆU & MỤC TIÊU DỰ ÁN (Markdown Cell)**

# SmartMealForecast - Dự đoán số lượng khách và gợi ý món ăn

\*\*Mục tiêu:\*\*

- Dự đoán số lượng khách ăn trưa/tối dựa vào dữ liệu ngày, thời tiết, ngày đặc biệt.

- Dự đoán tỷ lệ chọn món ăn phổ biến.

- Đề xuất món ăn phù hợp theo từng điều kiện thời tiết/ngày lễ.

**🧹 BƯỚC 1: ĐỌC & XỬ LÝ DỮ LIỆU**

* Đọc file CSV
* Xử lý missing
* Chuyển định dạng ngày
* Tạo cột Weekday, Season,...

**🍱 BƯỚC 2: VECTOR HÓA DỮ LIỆU MÓN ĂN**

* Tách các cột món ăn
* One-hot encoding món ăn (1206 vector)
* Ghép lại vào X

**🌦️ BƯỚC 3: TẠO BIẾN ĐẦU VÀO (X) VÀ BIẾN DỰ ĐOÁN (y)**

* X: thời tiết + ngày đặc biệt + món ăn
* y:
  + Lunch\_Count → y\_lunch
  + Dinner\_Count → y\_dinner

**🤖 BƯỚC 4: DỰ ĐOÁN VỚI XGBOOST**

* Dự đoán Lunch\_Count và Dinner\_Count
* Baseline = trung bình trong quá khứ
* Tính **MSE**, **MAE**
* So sánh mô hình XGBoost với baseline

**📊 BƯỚC 5: TRỰC QUAN HÓA KẾT QUẢ**

* Biểu đồ đường so sánh y\_true vs y\_pred
* Biểu đồ scatter giữa thời tiết và lượng khách

**📌 BƯỚC 6: DỰ ĐOÁN TỶ LỆ CHỌN MÓN ĂN**

* Với mỗi món chính (main dish) → tính tần suất được chọn
* Có thể dự đoán xác suất chọn trong ngày mới

**🍽️ BƯỚC 7: GỢI Ý MENU (RECOMMENDATION)**

* Dựa trên:
  + Số khách dự đoán
  + Ngày lễ / không lễ
  + Thời tiết
* Gợi ý menu phù hợp (dùng top món được chọn nhiều trong điều kiện tương tự)

**✍️ BẠN CẦN CHUẨN BỊ:**

1. Tên file dữ liệu (CSV) đầy đủ
2. Xác nhận bạn muốn mình:
   * Gửi **mẫu file SmartMealForecast.ipynb** mẫu không?
   * Hay bạn tự xây dựng từng phần theo hướng dẫn ở trên?

👉 Nếu bạn **gửi mình khoảng 5-10 dòng đầu trong file dữ liệu thật (.csv)**, mình sẽ:

* Làm sẵn 1 file mẫu .ipynb với tất cả các bước
* Gửi bạn để chạy trực tiếp trong VSCode

Bạn muốn mình hỗ trợ theo cách nào nhỉ?

Dựa vào cấu trúc dữ liệu bạn cung cấp, mình đã điều chỉnh lại các tính năng cho hợp lý hơn, kèm phần ghi chú về khả năng thực hiện bằng Python (✅ = có thể làm được, ⚠️ = cần thêm dữ liệu hoặc xử lý phức tạp):

| **Tên tính năng** | **Mục đích** | **Có thể làm không?** | **Chi tiết thực hiện (Gợi ý)** |
| --- | --- | --- | --- |
| **1. Dự đoán theo thời gian chỉ định** | Dự đoán cụ thể theo tháng/ngày người dùng chọn (ví dụ: dự đoán cho tháng 8) | ✅ | Sử dụng dữ liệu lịch sử tháng tương ứng (cùng tháng ở các năm trước hoặc tháng trước đó), dùng mô hình RandomForest/Linear Regression |
| **2. Phân tích món ăn theo mùa** | Gợi ý thực đơn theo mùa | ✅ | Dùng cột Season và phân tích món ăn hay xuất hiện theo mùa (xuất hiện nhiều trong mùa đó) |
| **3. Hiệu quả món ăn** | Đánh giá món hút khách | ✅ | So sánh số lượng người ăn hôm đó với trung bình các ngày bình thường. Nếu cao hơn → món hút khách |
| **4. Gợi ý món ăn theo mùa/dữ liệu cũ** | Tăng độ hấp dẫn thực đơn | ✅ | Dựa vào món được ưa chuộng trong mùa hoặc trong tháng trước đó có lượng người ăn cao |
| **5. Tính calories cho từng món ăn** | Giúp khách hàng quản lý sức khỏe | ⚠️ | Cần bạn cung cấp danh sách món ăn kèm calories tương ứng → mình sẽ tạo thêm cột Calories ứng với từng món (sau khi xử lý trùng tên, tách món) |

Nếu bạn đồng ý với các tính năng trên, mình sẽ giúp bạn tạo **file Python riêng biệt** cho từng nhóm tính năng sau:

1. prediction\_features.py → cho tính năng (1)
2. menu\_analysis.py → cho tính năng (2, 3, 4)
3. calorie\_counter.py → cho tính năng (5)

👉 Bạn muốn mình bắt đầu từ file nào trước? Hoặc nếu bạn muốn gộp tất cả vào một file, cũng được luôn nhé!