MDS Capstone Proposal Menu Engineering Supercharged

Capstone Proposal Report

Executive Summary

This project supports Heymate's mission to empower restaurant clients with smarter, data-driven tools. We propose a recommendation engine that identifies popular menu items based on market trends, offering a practical and flexible approach to menu optimization. While the initial version uses popularity-based scoring, future enhancements will leverage pretrained models and richer data sources to improve accuracy and impact. Integrated into Heymate's platform, this tool will help restaurants optimize offerings, reduce waste, and boost sales.

Visual overview of the recommender system workflow.

Introduction and Project Motivation

Many restaurants operate with limited insight into which cuisines are popular. This knowledge gap can lead to bloated menus, ingredient waste, and missed revenue opportunities.
Heymate, our project partner, is a Canadian tech company that offers cashback rewards and a full-suite business management platform for small and medium-sized enterprises (SMEs). Since many of Heymate's clients are restaurants, they are a primary focus for added value through smart tools.
This project helps Heymate enhance its role as a strategic partner by providing restaurant clients with a menu recommendation tool designed to answer two key questions:
1. How can menus match market trends?
2. How can market data help restaurant owners make better-informed decisions?

The solution consists of the following two components: 1. Develop a recommender system that suggests menu adjustments based on existing items and restaurant type
2. Build a data pipeline that continuously ingests new data to enhance the model's accuracy and relevance over time.

Solution Overview

The vertical flow illustrates the **Recommendation Pipeline**. In the production environment, the recommender will receive a web request from HeyMate's restaurant ERP system, which includes the restaurant type and the existing menu. In the development environment, we will simulate the process by directly using the restaurant data in the database. The recommender will generate a list of recommended items. For this pipeline, we will deliver 2 APIs where the recommender receives the input and returns the output.

The horizontal flow illustrates the **Knowledge Base Construction Pipeline**. It begins with open datasets of restaurant menus and then applies LLMs to carry out feature engineering. The LLMs standardize menu item names and their feature tags based on the cuisine name and description. All the cleaned and standardized data will be funneled into the knowledge base and assigned a unique label for each data source. As part of this pipeline, we will deliver an API that uses LLMs for feature extraction and build the initial knowledge base using a publicly available dataset.

For the recommender, we will develop one baseline algorithm based on cuisine popularity.

Data Sources

To support the development of our restaurant menu recommendation system, we aggregate and validate data for two primary use cases:

Input Data

The internal Heymate database and user-submitted forms serve as foundational inputs, providing:
- Restaurant Information: Name, Type, and Location
- Menu Information: Item Name, Category, and Price

Modelling Data

To enhance our understanding of menu popularity and strengthen the model's foundation, we incorporate external data sources:
- Kaggle dataset (Kaggle (2016)) for item frequency, restaurant names, categories, and prices
- Explore web scraping from platforms like DoorDash or Yelp to extract real-time insights such as most-liked dishes, ratings, or customer feedback

Data Validation (Initial EDA)

Our initial exploratory analysis revealed two key challenges that must be addressed before modelling.

Data Entry Challenge

Menu item entries are inconsistently formatted across restaurants. For example, the same dish may appear as “Seafood Fried Rice”, “149. Fook Chow Seafood Fried Rice”, or even in a different language. This inconsistency undermines our ability to compare and analyze menu data effectively.

To address this, we plan to leverage the capabilities of LLMs, taking inspiration from industry practices such as CloudKitchens’ approach to multi-location menu management (CloudKitchens (2024)). We will evaluate different options, such as GPT, Gemini, and some domain-specific models to:

- Normalize item names
- Extract consistent tags

```
{
  "Base Item": "Fried Rice",
  "Flavor Tag": "Seafood"
}
```

Modeling Challenge

A key challenge in our modelling process is integrating multiple datasets that use inconsistent or non-standard identifiers for restaurants and menu items.
To make these datasets work together, we need to:

- Align records using partially matching keys like restaurant and item names
- Normalize variations in naming using LLMs
- Enrich menu profiles by merging structured data from both sources, enabling robust popularity scoring

The final output is a unified and cleaned dataset enriched with relevant features such as item frequency, price, and standardized names, laying the groundwork for our Recommender Module.

Recommender Module

The recommender system operates through the following four steps:

1. **Input Data Collection**

- Restaurant Menu Data: Each restaurant provides its current list of menu items (names, categories, and optionally descriptions).

- External Dataset: We leverage the Kaggle dataset (and future datasets) to understand broader menu trends, including how often and how recently items appear across other restaurants.

2. **Matching**
   The system matches each of the restaurant's menu items to items in the knowledge base (built from the Kaggle dataset) using normalized names and categories. Matching is a key step because menu item names vary widely across restaurants (e.g., "BBQ Chicken Pizza" vs "Barbecue Chicken Flatbread").

3. **Scoring**
   Each item is assigned a popularity score based on:

   - Frequency: How often the item appears in other restaurants' menus

   - Recency: How recently it has been featured

   Scoring formula:
   `Score = 0.7 × (normalized frequency) + 0.3 × (normalized recency)`

4. **Recommendation**
   We score each menu item to identify both top and underperforming dishes. The system then returns a ranked list of menu items to guide data-informed decisions.

Example Use Case

For instance, if a restaurant uploads its menu and includes "Hawaiian Pizza," the system:
- Matches this item to the closest equivalent in the knowledge base
- Scores its market popularity and provides insight (e.g., "Low orders in the last 6 months")
- Recommends: - Updating the item if it's underperforming
- Adding trending items like "BBQ Chicken Pizza" that are missing from the current menu

Final Data Product

Our solution is a menu recommendation engine, showing the top 20 popular items tailored to each restaurant. The engine is powered by a query-based recommender trained on external datasets. To simulate real-world use, we will test the tool using historical data from Heymate's existing restaurant clients.

The anticipated business impact includes:
- Helping restaurant clients identify popular items, improve sales, and reduce waste

- Supporting smarter menu decisions and client retention
- Strengthening Heymate's value proposition as a data-driven platform partner

Project Timeline

Week	Date	Tasks
Week 1	May 5–9	Proposal, Design Knowledge Base, Database setup
Week 2–4	May 12–30	LLM research and tuning, workspace setup, API framework, core logic
Week 5	June 2–5	Integration and modularization
Week 6	June 9–13	Final review and report

Appendices

Addressing the Matching Challenge

Accurately matching menu items is one of the most challenging aspects due to variability in naming conventions.
Initially, we will use query-based techniques such as string normalization and category filtering.

To improve this process, we plan to explore pretrained models for semantic matching, including:
- Sentence Transformers (SBERT)
- spaCy similarity pipelines
- FastText embeddings

These models will improve the ability to recognize and link creatively named menu items to their standardized counterparts in the knowledge base.

References

CloudKitchens. 2024. "Multi-Channel, Multi-Location Menu Management." https://techblog.cloudkitchens.com/p/multi-channel-multi-location-menu.

Kaggle. 2016. "Restaurant Menu Dataset." https://www.kaggle.com/datasets/nypl/whats-on-the-menu.