Capstone: Cuisine Fusion Report

-Arjun Maniya, Swapnil Andhale, Ishan Bhusari

Introduction:

Globalization and culinary curiosity have encouraged recipe adaptation across cultural boundaries. However, substituting ingredients from one cuisine with those of another while preserving flavor compatibility and authenticity is non-trivial. In this report, we propose an algorithmic approach to cross-cuisine ingredient substitution, leveraging frequency analysis and semantic similarity derived from BERT-based embeddings. The goal of this project is to create a smart system that helps people substitute ingredients between different cuisines. If someone wants to cook an Indian recipe using Italian ingredients, the system can suggest alternatives that are commonly used in Italian cooking and are similar in taste or purpose. By understanding how ingredients are used in various regions and finding ones that are similar in meaning and flavor, the system can help make recipes more accessible, customizable, and fun—without losing their original charm. This can be useful for people who are experimenting with new cuisines, dealing with ingredient availability, or just trying to create their own fusion dishes.

Dataset Overview:

The dataset consists of structured recipe records collected from a diverse range of world cuisines. Each row in the dataset represents a single recipe, annotated with its geographical and culinary context, along with both raw and pre-processed textual fields.

The dataset contains the following fields:

A. Recipe_id: A unique identifier for each recipe
B. Region: Broad cultural origin of the recipe (e.g., "Indian Subcontinent")
C. Sub_region: A more fine-grained geographical classification (e.g., "Indian")
D. Continent: The continent to which the recipe's cuisine belongs (e.g., "North American", "Asian")
E. Ingredients: A list of raw ingredient names, often directly parsed from recipe text
F. Instructions: Step-by-step natural language description of the cooking procedure
G. cleaned_ingredients: A normalized and pre-processed version of the ingredients list

The cleaned_ingredients field is derived from the raw ingredients field through a cleaning pipeline that:

- Standardizes casing and punctuation
- Removes stop words, measurements, and irrelevant tokens
- Normalizes ingredient names to reduce sparsity (e.g., "chopped onions" → "onion")

The dataset contains 51350 recipes with the following distribution:

Cuisine	**Recipe Count**
Italian	16574
Mexican	14447
Canadian	6694
Argentine	6049
Indian	5987
Brazilian	447
Pakistani	372
Peruvian	241
Chilean	171
Colombian	151
Nepalese	88
Venezuelan	81
Ecuadorean	31
Bangladeshi	16

Algorithm:

A. Ingredient Frequency Extraction:
    - Ingredient frequencies taken from recipes of a cuisine are sorted
    - We extract the top 200 ingredients for each cuisine separately
    - The result (common list of ingredients) for a cuisine is stored as values in a dictionary with the key being the cuisine name itself

B. Ingredient Embedding using BERT:
    - We use the Sentence-BERT model all-MiniLM-L6-v2 to generate embeddings for each unique ingredient. These embeddings capture contextual and semantic information about ingredients as the BERT-based model (all-MiniLM-L6-v2) still has prior learned context from training on billions of sentences.

C. Cross-Cuisine Substitution via Similarity:
    - For each ingredient in the source cuisine, its embedding is compared to the embeddings of ingredients from the target cuisine using cosine similarity

- We return the ingredient from the target cuisine which has the highest cosine similarity with the corresponding ingredient from the source cuisine

Evaluation Metrics:

Evaluating ingredient substitution poses unique challenges due to its subjective and multi-valued nature. Unlike standard classification tasks, there is often no single "correct" substitute. Instead, the goal is to generate contextually appropriate and semantically similar alternatives that are valid within the target cuisine.

To assess the effectiveness of the substitution system, we adopt the following evaluation strategy:

A. Human Evaluation based Precision, Accuracy and Ambiguity Rate: We conduct manual validation of model-generated substitutions. For a given pair of cuisines (e.g., Indian → Italian), the model suggests the ingredient substitutions. Each substitution is manually annotated by human evaluators into one of the following categories:
    - Valid: The substitution is appropriate and commonly used in the target cuisine.
    - Invalid: The substitution is inappropriate, semantically incorrect, or alters the dish fundamentally.
    - Ambiguous: The substitutions that are context-dependent, but potentially acceptable.

    Based on this annotation, we compute

    - Precision of the system as:
      Number of Valid Substitutions / (Number of Valid Substitutions + Invalid Substitutions)
    - Ambiguity Rate as:
      Number of Ambiguous Substitutions / (Total Substitutions)
    - Accuracy as:
      (Number of Valid Substitutions + Ambiguous) / (Total Substitutions)

The inclusion of ambiguous substitutions in the accuracy formula acknowledges that not all substitutions are strictly binary when it comes to correctness, culinary substitutions often exist on a spectrum of acceptability and that ambiguous cases are a natural and acceptable part of recipe transformation tasks

To evaluate the substitutions, we considered both flavor similarity and ingredient usage in cooking. For example, ingredients were mentally grouped into broad functional or semantic categories such as meats, oils, aromatic bases (like onions or garlic), dairy, and seasonings. Substitutions were considered valid if they came from the same group and could fulfill a similar role in the recipe.

In cases where the substitution differed slightly in naming or form but clearly retained the original intent—such as "cheddar cheese" being replaced with "cheese" or "beef tenderloin" with "beef"—we treated them as valid generalizations or in some cases ambiguous generalizations considering the target cuisine in focus.

Results

To evaluate the quality of ingredient substitutions across cuisines, we selected a controlled sample size for practical and consistent human evaluation. For each source cuisine, we randomly selected some representative recipes, say X. This sampling was intentionally limited to:

- Ensure manual evaluation feasibility by human annotators
- Allow for qualitative, in-depth analysis of substitutions
- Maintain balanced testing across all cuisines

Each selected recipe was then transformed into versions corresponding to all other target cuisines. This setup resulted in a multi-directional substitution matrix, enabling us to test how well the substitution model generalizes across diverse culinary contexts.

The manageable number of transformations (X recipes × all target cuisines) ensured that human evaluators could:

- Carefully assess each substitution for flavor compatibility and role preservation
- Maintain consistency in judgment across all comparisons
- Reflect real-world adaptability and subjective acceptability of substitutions

This sampling strategy strikes a balance between breadth (diversity of cuisines) and depth (quality of human judgment), making the evaluation both scalable and trustworthy.

The avg precision, avg ambiguity rate and avg accuracy per cuisine is given in the table below:

Cuisine	Avg Precision	Avg Ambiguity Score	Avg Accuracy
Argentina	0.5905	0.0323	0.6063
Bangladesh	0.6736	0.5940	0.6881
Brazil	0.5659	0.0350	0.5807
Canada	0.5369	0.398	0.5601
Chile	0.5884	0.1344	0.6470
Columbia	0.6378	0.0651	0.6629
Ecuador	0.5985	0.0185	0.6038
India	0.6462	0.0345	0.662
Italy	0.5432	0.05146	0.5685
Mexico	0.5231	0.1162	0.609
Nepal	0.6479	0.085	0.6768
Pakistan	0.6270	0.0604	0.6485
Peru	0.6544	0.0469	0.6648
Venezuela	0.804	0.003	0.804

The evaluation results highlight the varying effectiveness of the ingredient substitution system across different cuisines. Venezuela achieved the highest precision score of 0.804, indicating its substitutions were most frequently valid, followed by Bangladesh at 0.6736. In contrast, Mexico and Italy had the lowest precision scores (0.5231 and 0.5432, respectively), suggesting greater challenges in finding appropriate substitutions for these cuisines.

Bangladesh exhibited the highest ambiguity rate (0.594), meaning many of its substitutions were context-dependent but not necessarily invalid. On the other hand, Venezuela and Ecuador had the lowest ambiguity rates (0.003 and 0.0185), indicating clearer, more definitive substitutions.

When considering accuracy, which accounts for both valid and ambiguous substitutions, Venezuela again performed best (0.804), with Bangladesh close behind (0.6881). Canada and Italy had the lowest accuracy scores (0.5601 and 0.5685), reflecting difficulties in maintaining substitution quality for these cuisines.

These results suggest that cuisines with distinct ingredient profiles, such as Venezuelan and Bangladeshi, are better suited for the current substitution approach, while those with more versatile ingredients, like Italian and Mexican, present greater challenges. The high ambiguity rates for some cuisines also underscore the need for more nuanced, context-aware substitution strategies in future iterations of the system.

Future Work:

To improve substitution quality, we plan to explore two key enhancements:

1. Flavor-Augmented Embeddings
    - Enrich BERT embeddings with basic flavor compound data (e.g., pairing "basil" and "mint" via shared terpenes) while avoiding over-engineering.
    - Test whether simple flavor-feature integration (like grouping ingredients by dominant taste profiles) reduces ambiguity in versatile cuisines.
2. Lightweight Multimodal Approach
    - Experiment with supplementing text embeddings with minimal visual data (e.g., pre-trained image vectors for texture/color) only where clearly beneficial (e.g., cheese substitutes).
    - Focus on one practical application first—like improving dairy/meat substitutions—before broader claims.

This would be validated through small-scale user tests comparing current vs. enhanced substitutions for 2-3 high-ambiguity cuisines (e.g., Italian → Indian).