Chapter- V
Summary and Conclusion

Chapter V

5. Summary and Conclusion

5.1 Summary

This study was initiated with a focus to examine the statistical characteristics of the food recipes of the regional cuisine prevailing in the Northeast part of India. The motivation was to explore and deduce some criteria for food preference in the region based on an data science based analysis, with an emphasis on its potential to contribute in developing foods meeting the consumer preference. To our knowledge, this is the first study on, identification and characterization of the culinary practices of northeast regional cuisine. By making use of data-driven analysis of culinary patterns, applied identification algorithms detect the degree of regional cuisine style preferences. Through the application of network theory and statistical measures to a large number of recipes from the region, we were able to discover the ingredient combinations in recipes as a characteristic of the regional cuisine

For the fulfilment of objective-1 we collected 702 recipes of Northeast regional cuisines through an extensive data curation process. The collected data were subjected to rank-frequency analysis at cuisine level. Tools from statistical analysis were employed for the identification of authentic and prevalent ingredients, ingredient pairing patterns, flavour pairing behaviours, ingredients responsible to attribute a flavour pairing behaviour, etc. The study carried out at the level of recipes, ingredients, and cuisines, revealed the role of ingredients and ingredient categories in determining the general patterns of culinary practice. The frequency of ingredient usage showed a generic culinary pattern conforming to the unique taste palate of the region. The majority of the authentic ingredients in Northeast regional cuisines were found to be from the category of plant derivatives, spices and vegetables such as black mustard seed oil, green bell pepper, ginger, cayenne, bay laurel, garlic and turmeric. In addition, the ingredient usage pattern is almost similar across the Northeast regional cuisine. The choice of ingredient combination in a recipe strongly determines the factor leading to the bias in food pairing as it depends on the number of flavours compound each ingredient shares. In terms of food pairing behaviour, the difference comes down to the question of how closely the flavour compounds are shared, i.e., either relatively low or high (negative and positive). Our study reported that

the Northeast regional cuisine exhibits a uniform negative food pairing behaviour. The majority of the ingredients that made a substantial impact on the food pairing were from the spice category, such as cayenne, bay laurel and turmeric.

For the fulfilment of objective-2 similarity analysis and clustering algorithm are used for making a comparison of the recipes – intra-cuisine and with recipes from beyond the region and recipes from western countries. Considering how close the Northeast states are to each other, geographical proximity is partly responsible for the similarity of regional cuisine. Further, the flavour network reveals that the affinities toward ingredient pairs based on the number of shared flavour compounds for the authentic pairs and triplets reveals the links between the ingredient pairs and ingredient triplets are not very significant, indicating that the ingredients do not share many flavour compounds. To determine the degree of similarity and dissimilarity among the recipes within Northeast regional cuisines, a cosine similarity analysis was conducted. It was found that recipes from the same region have the highest degree of similarity. The recipes that are found to be similar have a difference of at least one ingredient or none at all. We observed in the case of a recipe where the main ingredient belongs to the meat or fish category, the replacement of ingredients happens within the ingredients from the same category itself. Our analysis revealed that Assamese recipes share many similarities with other regional cuisines and Tripura recipes as the least similar followed by Mizoram. This result reflects the uniqueness of regional cuisine in terms of the ingredients used in their recipes. The comparative study of Northeast cuisine with the other Indian regional cuisines to determine the extent of similarities and dissimilarities between the regional cuisines. We analysed the similarity of Northeast regional cuisine across the Indian regional cuisines in terms of their ingredient usage. It was observed that Assam regional cuisine shares the most affinities with Indian regional cuisine with a total of 519 similar recipes out of 2916 recipes of the other Indian regional cuisine. Additionally, Tripura recipes were found to be the least similar as only 4 recipes are found to be similar.

The comparative study carried out to estimate the similarity of the Northeast regional cuisine with a western (positively paired) regional cuisine and east Asian (negatively paired) regional cuisine revealed that the cosine similarity values of the similar recipes were comparatively low in western cuisine as compared to east Asian cuisine. There is less similarity between the recipes of the Northeast regional cuisines and other regional cuisines except for the recipe within the Northeast region. Consequently, the Northeast

regional cuisines have a distinct identity in terms of their uniqueness. Based on the comparative study done using cosine similarity we can conclude that the proposed algorithm can be used as an effective tool in determining the similarity across various regional cuisines. Additionally, the results obtained from the cosine similarity can be used as a dish recommender system to find similar dishes across various regional cuisines which is based on flavour similarity. An analysis based on t-SNE clustering of the Northeast regional cuisine was carried out to visualize multi-dimensional data and to examine whether there is any overlap in the choice of ingredients and the flavour profile within the regional cuisines. We observed that the distinct clusters are formed mostly in the case of the ingredient profile but not in the flavour profile. The similarity in flavour profile further highlights the possibility of using the ingredient as a substitute in situations where ingredients are not readily available for a recipe or product development.

For the fulfilment of objective-3 generative models were applied to generate new combinations of ingredients based on the consumer preference. Based on canonical correlation analysis of recipe and flavour data, data-driven models were able to suggest the ingredients that are best combined with a particular set of ingredients. The data models can complete or suggest ingredients in a recipe, which can be seen as recipe completion. It helps us to identify ingredients which can be best paired with the desired set of ingredients where the ingredients generated are considered to be the best candidate to use for recipe completion purposes. The recommended list of ingredients suggested appears to be acceptable as most of the recommended ingredients are from the same categories itself. Additionally, new recipe generation with the application of a flavour network was carried out considering the ingredients with the least shared flavour compounds in preparation for a new dish. The recipe generated when validated for similarity using cosine similarity with pre-existing recipe data showed a similarity score close to 1. Further, it can also be used as a tool for new product development to estimate the best-paired ingredients and in developing alternative dishes for people who are habituated of using similar ingredients.

For the fulfilment of objective-4 the development of an algorithm for alternative ingredients considering customer specifications was used successfully to create new ingredient combinations that mimic the flavour characteristics of the initial item being replaced initially generating a profile of flavour similar to that of a selected item for people with dietary restrictions. The study showed that the recipe generated is similar to

a pre-existing recipe as the recipe generated through the algorithm when compared to the pre-existing recipe considering the flavour properties were closely related. As a result, we can consider the proposed algorithm to be used for future recipe development purposes for recommending alternative ingredients for any target recipe.

5.2 Conclusion

1. Few ingredients appear in a greater number of recipes, and analysis revealed most of them to be from spice category which is validated by rank frequency plots and comparison of recipes with random recipes and frequency preserving random recipes. These ingredients are identifiable from prevalence studies.
2. Food ingredients from spice category are more prevalent than the other categories in the cuisines from the Northeast region. Ingredients such as black mustard seed oil, onion, cayenne, ginger, green bell pepper, garlic, turmeric, bay laurel, pork, rice and tomato are the most prevalent ingredients across the cuisines. Whereas ingredients such as black mustard seed oil, green bell pepper, onion, cayenne, ginger, garlic, rice, turmeric, bay laurel, tomato and pork were found to be the most authentic ingredients across the regional cuisines.
3. Occurrence of black mustard seed oil as the most prevalent ingredient in the region as a whole and individually in 07 of the state cuisines; and pork meat is one of the authentic ingredients in 04 state cuisines characterizes the Northeast regional cuisine which is not the case for other Indian regional cuisines.
4. Ingredients from spice category are contributing to the negative pairing of ingredients, consistently for all the eight regional cuisines.
5. Ingredients from spice category from the region are sharing very less compounds within the ingredients leading to a flavour network with very thin edges among the spice category ingredients as compared to the other categories ingredients.
6. Among the recipes from Northeast region, flavour-based distinction between recipes is insignificant in many recipes with meat or fish as the main ingredient as these recipes differ with a single ingredient of meat or fish being replaced with another ingredient from the same category.
7. Computational algorithm-based clustering highlights a distinct preference, both in terms of ingredients and flavour, in the Northeast regional cuisines as compared to cuisines of other geographic regions from India.

8. Assamese regional cuisine shares the most affinities with other Indian regional cuisines with Tripura cuisines having least similarity with other Indian regional cuisines.
9. Higher prevalence of spice category ingredients like cayenne and ginger, and rice, the Northeast regional cuisine has more similarity with East Asian cuisines as compared to the western cuisines.
10. Most of the Northeast recipes which have similarity to western recipes have major ingredients derived from the dairy category.
11. Application of the RLS algorithm to the best paired ingredient sets have resulted in recipe completion based on shared flavour compounds.
12. Computational algorithm-based search for alternative ingredients, with the criteria of mimicking the flavour characteristics of the replaced ingredient facilitate generation of newer recipes.

In summary, our work presents a scientific validation of the existing trend of ingredient combination which forms a food recipe for a cuisine used in a limited geographical location. This highlights the prospect for food developers in developing new food products to cater to the preferences of consumers who have more preferences for spicy-flavoured products than for dairy flavoured products. Further, understanding the nature of ingredient usage in recipes in the cuisines of concern would aid in the development of recipes or products that would assure customer acceptability. Additionally, the ingredients used to create substitutes or supplements for certain lifestyle diseases, which often entail dietary restrictions must possess a balance between flavour, taste, and nutritional properties based on the list of foods that may be consumed. The development of healthy food with good sensory qualities is an area of application of the food pairing hypothesis and the findings of this work will facilitate the same for the Northeast regional cuisines.

5.3 Limitations and scope for future works

Present study has a few important limitations.

1. There are limited data in public domain on Northeast regional cuisines, resulting in collection of relatively smaller data for recipes. In future more such data are expected to be available in public domain and our proposed system can be applied on a larger recipe data set.

2. The cooking methods (e.g., deep frying, boiling, and baking) and the relative quantities of ingredients were not considered in the analysis. Inclusion of these characteristics would make such an analysis more useful in generation of knowledge for application in food product development.

However, the work demonstrated how the data-driven network analysis can be applied to other areas, such as food science. Food companies or professional chefs might find it beneficial to develop data-driven systems that can convert recipes into any regional cuisine style to create new recipes in response to the increasing diversity in food preferences and regional cuisine styles, for taking up as separate research work.