CUISINE PREDICTION BASED ON INGREDIENTS

BY

S. K. M SHADEKUL ISLAM
ID: 191-15-12836
AND

MAJEDA AKTER MITU
ID: 191-15-12644

AND
MYSHA MOBASHIRA
ID: 191-15-12216

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By
Tapasy Rabeya
Sr. Lecturer
Department of CSE
Daffodil International University

Co-Supervised By
Sharmin Akter
Sr. Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project titled "**Cuisine Prediction Based On Ingredients**", submitted by S. K. M Shadekul Islam, ID No: 191-15-12836, Majeda Akter Mitu, ID No: 191-15-12644 and Mysha Mobashira, ID No: 191-15-12216 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 23 January, 2023.

BOARD OF EXAMINERS

______________________ **Chairman**

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________ **Internal Examiner**

Dr. Md. Zahid Hasan
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________ **Internal Examiner**

Fahad Faisal
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________ **External Examiner**

Dr. Ahmed Wasif Reza
Associate Professor
Department of Computer Science and Engineering
East West University

©Daffodil International University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Tapasy Rabeya, Lectuer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

____

Tapasy Rabeya
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

____

Sharmin Akter
Sr. Lecturer
Department of CSE
Daffodil International University

Submitted by:

____

S. K. M Shadekul Islam
ID: -191-15-12836
Department of CSE
Daffodil International University

____

Majeda Akter Mitu
ID: -191-15-12644
Department of CSE
Daffodil International University

Mysha
____

Mysha Mobashira
ID: -191-15-12216
Department of CSE
Daffodil International University

ii

©Daffodil International University

©Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Tapasy Rabeya**, **Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan**, Professor, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

©Daffodil International University

ABSTRACT

Cuisine prediction based on ingredients is a complex task that has numerous applications in the food industry, including recipe recommendation and food delivery. Accurate and reliable prediction models have the potential to improve the efficiency and effectiveness of these and other applications. In this study, we sought to develop a prediction model for cuisine type based on ingredients using a large dataset called 1M Recipe Dataset and the machine learning technique of grid search with cross-validation. Our results showed that the model was able to achieve a mean cross-validated accuracy of approximately 73.7%. While this is a promising result, it is clear that there is still room for improvement and further research is needed to optimize the model and better understand the factors that influence cuisine prediction based on ingredients. Future research could focus on a range of approaches to improving the model, such as exploring different techniques for feature extraction and classification, incorporating additional data sources, and applying the model to other tasks and problems related to cuisine prediction. Additionally, further study is needed to understand the limitations of the model and how it can be extended or modified to better address the challenges of cuisine prediction. Overall, our study represents an important step forward in understanding the relationship between ingredients and cuisine, and has the potential to inform the development of improved prediction models and applications in a variety of contexts.

©Daffodil International University

TABLE OF CONTENTS

©Daffodil International University

©Daffodil International University

LIST OF FIGURES

©Daffodil International University

LIST OF TABLES

©Daffodil International University

CHAPTER 1

Introduction

1.1 Introduction

Cuisine prediction is increasingly important in the food industry, with applications ranging from personalized cooking recommendations to inventory management and menu planning. The ability to accurately predict the type of cuisine based on the ingredients of a recipe can have significant practical benefits, such as helping people discover new recipes and cuisines that they might not have tried before or enabling more efficient operations in the food industry. However, predicting cuisine from ingredients can be a challenging task, as it requires understanding the relationships between the ingredients and the cultural and historical traditions of different cuisines.

In this paper, we investigate the feasibility of predicting cuisine type based on recipe ingredients. We collected a large dataset of recipes from various sources, including both traditional and modern dishes from a diverse range of cuisines. We preprocessed the data to clean and prepare it for analysis and then used machine learning techniques to build a model that can accurately classify recipes into different cuisine categories. We evaluated the performance of our model using various evaluation metrics and compared the results to those obtained using different model architectures and feature sets.

Our results suggest that it is possible to accurately predict cuisine type using a combination of ingredient-based features and advanced machine-learning techniques. We also found that certain ingredients and combinations of ingredients are more indicative of certain cuisines than others and that the size of the training dataset can have a significant impact on the performance of the model. We discuss the implications of our findings and suggest directions for future research in this area, including the potential for using cuisine prediction to improve personalization and recommendation systems in the food industry.

©Daffodil International University

1.2 Motivation

Cuisine prediction is a problem that has attracted the attention of researchers and practitioners in the food industry due to its potential applications in personalized cooking recommendations, inventory management, and menu planning. However, predicting cuisine type based on recipe ingredients is a complex task that requires understanding the relationships between ingredients and the cultural and historical traditions of different cuisines. In this project, we are motivated by scientific curiosity to investigate the feasibility of predicting cuisine type based on recipe ingredients, and to identify the ingredients and combinations of ingredients that are most indicative of different cuisines. By exploring these relationships, we hope to deepen our understanding of the factors that influence cuisine type, and contribute to the development of more accurate and effective cuisine prediction models.

1.3 Rationale of the Study

Practical applications in the food sector include personalized cooking suggestions, inventory management, and menu planning, all of which involve determining the style of cuisine based on the ingredients of a dish. Predicting the kind of cuisine from a recipe's components is a difficult endeavor since it calls for knowledge of the interconnections between ingredients and the culinary, social, and historical traditions associated with each cuisine. When it comes to predicting what kind of food someone would like, most studies have relied on basic ML models and hand-crafted characteristics.

The goal of this research is to examine whether or not it is possible to predict the kind of cuisine from a given set of recipe components using a set of feature vectors based on those ingredients and some sophisticated machine learning methods. We will make use of a large dataset consisting of recipes from a wide variety of sources, covering both classic and innovative foods from a wide variety of cultures. The goal of this research is to build a model that can accurately categorize recipes into different cuisine categories by identifying the ingredients and combinations of ingredients that are most indicative of different cuisines, based on our analysis of the relationships between ingredients and cuisine type in

©Daffodil International University

this dataset. Our findings might be used to improve the food industry's personalization and recommendation systems, as well as to promote cross-cultural understanding and even health advantages.

1.4 Research Questions

Here are a few research questions for our project on cuisine prediction:

- Can we predict the type of cuisine based on the ingredients of a recipe?
- What are the most important ingredients for predicting cuisine type?
- How does the performance of the model change as we increase the size of the training dataset?
- Do more advanced machine learning techniques lead to better cuisine prediction performance?
- How accurately can we predict the cuisine of a recipe?

These research questions serve as the foundation for our project and help guide our analysis and experimentation.

1.5 Expected Output

The expected output of this study is a better understanding of the feasibility of predicting cuisine type based on recipe ingredients. Specifically, we hope to develop a model that can accurately predict the type of cuisine based on the ingredients of a recipe and identify the ingredients and combinations of ingredients that are most indicative of different cuisines. We also aim to compare the performance of different machine learning techniques for cuisine prediction, and to provide recommendations for improving the performance of cuisine prediction models. By achieving these goals, we hope to contribute to the development of more effective and accurate cuisine prediction models that can be used in practical applications such as personalized cooking recommendations, inventory management, and menu planning.

©Daffodil International University

1.6 Report Layout

There are six chapters in total. Each chapter is discussed from various perspectives. And each chapter has several parts that are explained in detail. The following are the contents of this report paper:

In Chapter 1 we tried to give a simple introduction of our project. Then we portrayed what motivates us to do the project in the first place. Then we discussed about the potential outcome of the project. You will also find the rationale of the study and research questions discussed here.

In Chapter 2 we discussed about the background of the study. We portrayed a comparative analysis and summary of the related works. Then we discussed about the scope of the problem and the challenges we face or might face.

Also, in Chapter 3 in this chapter we described the full working procedure of our work. Here, you have research subject and what tools we used to conduct the research. Then there is data collection, data analysis, our proposed method. And at last, there is our setup for the implementation.

And Chapter 4 covers the experimental setup, result of the project and discussion of this research project.

In Chapter 5 we discussed about the impact of our project on society and environment. Also, we look at the ethical aspects of this research study. And discussed about the sustainability plans.

Lastly, Chapter 6 is all about the conclusion and summary of the project. Also we tried to touch on what is vision with this project in the future.

©Daffodil International University

CHAPTER 2

Background

2.1 Terminologies

In a project on cuisine prediction, there are several technical terms and concepts that we encountered. Such as Machine learning, Classification, Feature engineering, Training set, Test set, Overfitting, etc.

Machine learning is a field of artificial intelligence that involves training algorithms to learn from data and make predictions or decisions. Cuisine prediction is an example of a classification problem, which is a machine-learning task that involves assigning a label or class to an input based on a set of predefined categories.

Feature engineering is the process of selecting and extracting relevant features from raw data to improve the performance of a machine-learning model. In the context of cuisine prediction, this might involve selecting ingredients or combinations of ingredients that are most indicative of different cuisines.

The dataset used to train and evaluate a machine learning model is typically split into a training set and a test set. The training set is used to teach the model the relationships between the features and the labels, while the test set is used to evaluate the performance of the model on unseen data.

Overfitting is a phenomenon that occurs when a model performs well on the training set but poorly on the test set, and it can limit the generalization capabilities of the model.

2.2 Related Works

J. Naik et al 2015 [1] the authors used machine learning to approach food and recipes in new ways. They used a publicly available dataset of recipes with ingredients labeled by cuisine to try and answer two questions: (1) given a set of ingredients, what is the cuisine of the recipe, and (2) if they can automatically generate a recipe. The authors found that

©Daffodil International University

the first two models resulted in similar recipes with almost the same ingredient sequence, but different verbs. The third model, however, predicted many different verbs and ingredients, which resulted in strange instructions such as "pour chicken". Overall, the authors found that the second model seemed to be the best, allowing for more dynamic formation of instructions while still generating a reasonable ingredient sequence.

S. Jayaraman et al 2018 [2] the authors studied the relationship between cuisines and the ingredients used in their recipes. They used data science techniques like support vector machines and associative classification to analyze a dataset of recipes from various sources. The goal was to gain a better understanding of the patterns and essential elements of good recipes, as well as the correlations between ingredients and cuisines. The authors also compared the accuracy of different classifiers used to predict cuisines. The results provided a detailed and clearer insight into these relationships.

M. Anand Kumar et al 2018 [3] the goal of this paper is to use tree boosting algorithms, specifically XG-Boosting, to predict the cuisine of a recipe based on its ingredients. The authors used machine learning models and prediction methods to study customer behavior and increase revenue in the e-commerce industry. They applied these techniques to data from online cooking and recipe sharing websites to develop a scalable system for predicting cuisines based on ingredients. The authors obtained an accuracy of approximately 80% in their cuisine prediction using this method.

A. Ali et al 2018 [4] this study aims to predict the cuisine of a recipe based on its ingredients using deep learning techniques. The authors implemented several methods, including fastText, TextCNN, TextRNN, TextBiRNN, TextAttBiRNN, and HAN, and found that all of them performed well. However, HAN achieved the highest accuracy at 87%. The goal was to create a model that could learn the characteristics of various cuisines and use that knowledge to identify the correct cuisine based on the ingredients in a recipe.

A. Bharam et al 2018 [5] in this paper, the authors propose a way to recommend Indian recipes based on readily available ingredients and popular dishes. They use a content-based approach and machine learning to perform a web search and create a collection of recipe

©Daffodil International University

types. The resulting recommendation system suggests Indian recipes based on the input of specific ingredients. The authors note that recommendation systems, including those using machine learning, are useful for personalization and helping users find relevant and reliable information. They also suggest potential avenues for future work on this topic.

R. Ghewari 2018 [6] in this study, the authors used the Yummly dataset to explore the relationship between ingredients and cuisines. The dataset consisted of 39,774 recipes with 20 different cuisines and 6714 unique ingredients. The authors tested several classifiers and found that both logistic regression and support vector machines (SVMs) performed well in predicting the cuisine of a recipe based on its ingredients. They attributed this success to the soft boundary characteristic of the data. The authors also found that the bag of words model on ingredients worked well for this task, leading them to the realization that predicting cuisines from ingredients is similar to traditional topic modeling. They suggest further exploration of this similarity as a potential avenue for future work.

C. Anderson 2018 [7] in this survey, the authors reviewed the literature on food recommenders, including apps, services, and programs that provide recommendations for food, meals, recipes, and restaurants. They also discussed recommenders for substitute ingredients, menus, and diets, with a particular focus on the importance of health and wellness in the context of specific dietary needs and goals. The authors covered the types of recommender systems in terms of their goals and recommendations, the datasets and signals used to train models, the technical approaches and model types employed, and system constraints.

M. K. Shan 2014 [8] in this paper, the authors studied the relationship between recipe cuisines and ingredients using classification techniques like associative classification and support vector machines. They conducted their analysis on a dataset from food.com and aimed to understand which cuisines are most similar and what ingredients are essential for each cuisine. The goal was to apply these findings to the task of automatically labeling recipes by cuisine. This research expands upon the more common focus on recipe recommendation to examine the underlying connections between cuisines and ingredients.

©Daffodil International University

R. Singh Verma 2015 [9] in this paper, the authors explored different strategies for identifying the cuisine of a recipe based on its ingredients. They tackled this problem as a cuisine classification task and evaluated various classification algorithms, including approaches that considered combinations of multiple ingredients. The authors used a dataset from a Kaggle competition hosted by Yummly, which contained 39,774 instances of cuisines and ingredient lists, with 20 cuisines and 2965 distinct ingredients in total. The goal was to understand what combination of ingredients can be helpful in identifying a cuisine when the recipe is not provided.

T. Ozaki et al 2015 [10] in this paper, the authors explore the relationship between cuisine types and the ingredients and cooking actions used in different types of recipes. They use a dataset from a Japanese recipe site to conduct statistical analyses and apply an extended version of the term frequency-inverse document frequency (TF-IDF) method to extract important pairs of ingredients and cooking actions. They also use association rules to identify sets of ingredients that are characteristic of each cuisine type and evaluate these sets using various evaluation measures. The goal of this study is to better understand how ingredients and cooking actions can be used to characterize cuisine type and to make it easier for users to find recipes that fit their preferences on recipe websites.

M. M. Zhang et al 2015 [11] in this study, the authors use ingredients as attributes to classify dishes into different cuisines based on the country where the cuisine originates. They use a dataset of 76 images of two dishes from each of three cuisines to test their method, which is based on the idea of attribute-based classification. The results of the study show that this method is effective and can be generalized to classify dishes from other cuisines and with different ingredients, as long as the ingredients can be specified using attribute descriptors. The mean accuracy of the method in this study was 82.9%.

P. Bhat [12] in this paper, the authors aim to use patterns related to ingredients to predict the cuisine of a recipe. They use a dataset of recipes from the website Yummly, which includes recipes from a variety of cuisines. The goal of this study is to better understand the ingredients typically used in different cuisines and to identify common ingredients

©Daffodil International University

across cuisines. The authors hope that this approach will help compare ingredients within and across cuisines and allow for the creation of new recipes using similar ingredients.

W. Min et al 2018 [13] In this paper, the authors conducted a cross-region analysis of recipes shared online, using ingredients, food images, and attributes such as cuisine and course. They proposed a culinary culture analysis framework that uses a probabilistic topic model to discover cuisine- and course-specific topics, and the manifold ranking method to incorporate visual features and retrieve food images for topic visualization. This framework was applied to three applications: 1) multimodal cuisine summarization with ingredients and images, 2) analysis of cuisine-course patterns, including topic-specific cuisine distribution and cuisine-specific course distribution of topics, and 3) cuisine recommendation for both cuisine- and ingredient-oriented queries. The authors conducted experiments on a dataset of 66,615 recipes from 10 cuisines in Yummly and found that their framework was effective at analyzing and interpreting culinary cultures at both the macro and micro levels.

B. Li and M. Wang 2015 [14] In this project, the goal was to classify 20 types of cuisines based on ingredient lists. The team built several feature sets and tested them with various multiclass classification algorithms, including OVA Naive Bayes, OVA SVM, OVA logistic regression, OVA SVM with the Crammer method, and K-nearest-neighbor. The results showed that logistic regression performed the best, with an accuracy of 78.4%. The team also found that some classes were highly correlated, which reduced the overall accuracy of the algorithm. In future work, the team plans to investigate this correlation further and develop ways to better distinguish these specific classes. The final goal of the project was to apply this algorithm to the Yummly website to improve the user experience by recommending relevant recipes or deciding which recipes to display based on searches or browsing activity.

H. H. Holste et al 2015 [15] The goal of this study was to classify the cuisine of a recipe based on its ingredients using machine learning. The dataset for this study included 39,774 recipes with a list of ingredients and a cuisine label. The classification was intended to

©Daffodil International University

reduce the number of required user inputs on recipe websites, help determine cuisine similarity for restaurant recommendation systems, and allow users to easily classify their own recipes. The results of the study showed that a classification accuracy of at least 77.87% was achievable.

P. Hazarika et al 2019 [16] This paper presents a method for recommending Indian cuisine recipes based on a user's preferred ingredients and liked dishes. The method involves web scraping to create a database of Indian cuisine recipes and using content-based machine learning to recommend the recipes. The system aims to address issues such as cold start and heterogeneity in recommendation by linking users to their social network profiles and suggesting recipes liked by their friends, and by building better, more dynamic web crawlers. Future improvements could include making suggestions based on the geographical location of the cuisine, the particular chef whose dishes the user likes, or specialty dishes found in nearby restaurants.

S. Kalajdziski et al 2018 [17] The goal of this research was to classify different types of cuisine based on the ingredients used in their recipes. The team used a variety of machine learning algorithms and feature selection techniques to analyze a dataset of recipes from Yummly. They found that logistic regression had the highest accuracy in classifying cuisines, with an accuracy of 78.4%. The team also explored the error distribution for each class and found that some classes were highly correlated, which reduced the accuracy of the algorithm. In the future, the team plans to further investigate the correlation between classes and find ways to better distinguish between them.

Y. Tian et al 2021 [18] In this paper, the authors introduce RecipeNet, a large-scale corpus of recipe data, and propose a novel model called rn2vec for learning recipe representations. The model is able to capture textual, structural, and nutritional information through several neural network modules, including a textual CNN, an inner-ingredients transformer, and a graph neural network with hierarchical attention. The authors also design a combined objective function of node classification and link prediction to optimize the model jointly. The goal is to use the learned recipe representations to support various food studies, such

©Daffodil International University

as analyzing culinary habits, estimating recipe healthiness, answering recipe questions, recommending recipes, and classifying recipes.

S. Sajadmanesh et al 2017 [19] This study aimed to understand the relationship between recipes published on the internet, their ingredients, flavors, and nutritional values, and the culinary habits and health indicators of different countries. The study used a database of over 157,000 recipes from more than 200 cuisines obtained from the Yummly website and API. The ingredients from these recipes were analyzed and compared to understand the differences between dishes from different regions and the predictability of recipes from different cuisines. The study also looked at the relationship between the nutrition information in the recipes and the health indicators of different countries, such as obesity and diabetes rates. The results of the study showed strong similarities between the cuisines of neighboring countries, but a large diversity of ingredients and flavors across different continents, largely influenced by migration trends. The study also found a strong correlation between the nutrition information in recipes, particularly sugar intake, and obesity rates. The study concluded that deep learning can be effectively used to predict cuisines based on ingredients, which could improve recipe recommendations based on individuals' profiles.

L. Herranz et al 2018 [20] This review paper discusses how visual content, context, and external knowledge can be effectively integrated into food-oriented applications, particularly in the areas of recipe analysis and retrieval, food recommendation, and restaurant context. The authors also discuss current datasets used for these tasks and highlight open research problems in the field. They conclude by mentioning the potential for future developments in the field, including improvements in multimodal representations, category recognition, nutrient estimation, and structured knowledge representation.

I. Nirmal et al 2018 [21] In this study, a data-driven flavor and nutrient optimization framework was developed to recommend ingredients for recipes based on both flavor and nutrition. The framework included a cuisine prediction model using Random Forests,

©Daffodil International University

recipe similarity measures, and normalized mean support and reciprocal authenticity measures to maintain cuisine and flavor specific features. The nutrient optimization aspect considered essential nutrients and daily recommended nutrient reference values to optimize nutrient abundance. The framework was tested using a dataset of 56,498 recipes covering 11 cuisines and 381 ingredients, with a prediction accuracy of 79.546%. The framework was able to generate alternative ingredients with similar flavor but improved nutrition for a given recipe.

T. K. Celestin et at 2020 [22] This study developed a classification approach based on machine learning to identify the diet preferences of populations in geographically remote areas. The approach used three algorithms (logistic regression, support vector machine, and neural networks) on a dataset of food recipes from various Asian cuisines to find the similarity between the cuisines based on their ingredients. The support vector machine with a normalized polynomial kernel was found to be the best performing algorithm for this task. The purpose of the study was to create a recommendation system for people living in a country other than their own to help them find specific recipes among the vast number available online. The results showed that the support vector machine model was a useful tool for this classification and prediction task.

2.3 Comparative Analysis and Summary

Table 2.1: Comparative Analysis of Previous Works

Study	Data Source	Approach	Accuracy
J. Naik et al 2015 [1]	Food.com & Yummly Dataset	Linear SVC	79%
S. Jayaraman et al 2018 [2]	Allrecipes Website	Multi-variate Logistic Regression	79%
M. Anand Kumar et al 2018 [3]	Yummly Dataset	XG-Boosting	80%

©Daffodil International University

A. Ali et al 2018 [4]	Yummly Dataset	HAN	87%
A. Bharam et al 2018 [5]	N/A	Content Based Filtering	N/A
R. Ghewari 2018 [6]	Yummly Dataset	SVM	78%
C. Anderson 2018 [7]	N/A	N/A	N/A
M. K. Shan 2014 [8]	Food.com Data	SVM	89%
R. Singh Verma 2015 [9]	Yummly Dataset	LinearSVC	79.395%
T. Ozaki et al 2015 [10]	Cookpad Data	TF-IDF	N/A
M. M. Zhang et al 2015 [11]	Online Sources	Attribute-Based Classification	82.9%
P. Bhat [12]	Yummly Dataset	Centroid clustering and cosine similarity model	78%

In general, the findings of these studies indicate that it is possible to predict the cuisine of a recipe with a high degree of accuracy using machine learning techniques and ingredient data. Some of the approaches used in these studies, such as deep learning and XG-Boosting, SVC, achieved particularly high levels of accuracy. However, not all of the studies reported the accuracy of their approach, so it is not possible to make a comprehensive comparison of the effectiveness of different methods.

©Daffodil International University

2.4 Scope of the Problem

The scope of the problem of cuisine prediction based on ingredients is to develop a machine learning system that can accurately classify a recipe or dish into one of several predefined cuisine categories based on the ingredients used. This involves identifying the ingredients used in a recipe and accurately classifying them, accounting for the many different cultural and regional variations within each cuisine, and handling rare or unusual ingredients.

In order to solve this problem, it is necessary to develop machine learning algorithms that can accurately identify and classify ingredients and account for their relationships to different cuisines. It is also necessary to gather and clean a large dataset of recipes and their corresponding ingredient lists and cuisine labels in order to train and test the machine-learning model.

Overall, the goal of this problem is to develop a system that can accurately predict the cuisine of a recipe or dish based on its ingredients, with applications in meal planning, food product development, and restaurant recommendations.

2.5 Challenges

There are several challenges that we encountered when attempting to predict the cuisine of a recipe based on its ingredients.

One challenge is the diversity and complexity of cuisines and ingredients. Cuisines can vary greatly in terms of their ingredients, flavors, and cooking techniques, and it is difficult to accurately predict cuisine type based on a limited set of ingredients. Additionally, ingredients can be used in multiple cuisines, making it difficult to accurately assign a cuisine to a recipe based on its ingredients alone.

Another challenge is the quality and representation of the training data. The accuracy and generalizability of a prediction model will depend heavily on the quality and representativeness of the training data. If the training data is poorly labeled, incomplete, or biased, the model can be less accurate and less useful.

©Daffodil International University

Finally, there are also challenges in selecting and extracting relevant features from the ingredients. It is difficult to identify the most important or distinguishing ingredients for a particular cuisine or to accurately represent the ingredients in a way that is useful for prediction.

Overall, predicting the cuisine of a recipe based on its ingredients is a complex task that involves addressing a number of challenges related to the diversity and complexity of cuisines and ingredients, the quality and representation of the training data, and the selection and extraction of relevant features.

©Daffodil International University

CHAPTER 3

Research Methodology

3.1 Research Subject and Instrumentation

In the context of cuisine prediction based on ingredients, the research subject would be the prediction of the cuisine of a recipe based on its ingredients.

Research instrumentation refers to the tools and techniques used to collect and analyze data in a research study. In a study on diet recommendation systems, e-coaching, and food delivery, the research instrumentation includes a range of tools and techniques to gather data from the research subjects and to analyze that data. Also for making the thesis paper.

3.1.1 Design:

- **Smart Draw:** Smart Draw, an online diagramming tool, to create and share flowcharts, organizational charts, mind maps, project charts, and other types of corporate visualizations for communication and collaboration.

3.1.2 Writing:

- **Google Docs:** Google Docs is a word processing program that is part of the Google Workspace productivity suite. It allows users to create, edit, and collaborate on documents online. Google Docs is similar to Microsoft Word, but it is a cloud-based program, which means that it is accessed through a web browser rather than being installed on a specific computer.

- **Microsoft word:** Microsoft Word is a word processing program that is part of the Microsoft Office suite. It allows users to create, edit, and print documents, as well as to add images, tables, and another formatting. Microsoft Word is available as a desktop program or as an online version called Word Online.

- **Grammarly:** Grammarly is a writing tool that checks the text for grammar and spelling mistakes. It can be used as a browser extension or as a standalone app, and it can check text in a variety of languages. Grammarly can also provide suggestions for improving the clarity and style of writing.

©Daffodil International University

3.1.3 Project Management:

- **Notion:** Notion is a productivity and organization tool that allows users to create notes, tasks, wikis, and databases in a single workspace. It can be used for a variety of purposes, including project management, document collaboration, and personal organization.

3.1.4 Referencing Tool:

- **Mendeley:** Mendeley is a reference management tool that helps researchers and students organize their research papers and citations. It allows users to import and organize their research documents, create bibliographies in a variety of citation styles, and collaborate with others on research projects.

3.1.5 Communication:

- **Telegram:** Telegram is a messaging app that allows users to send texts, photos, videos, and other media to other users. It is known for its fast speeds and high level of security.

- **Facebook Messenger:** Facebook Messenger is a messaging app that is part of the Facebook social media platform. It allows users to send messages, photos, and other media to their Facebook friends and groups.

3.1.6 Data Store and Share:

- **Google Drive:** Google Drive is a cloud storage service that is part of the Google Workspace productivity suite. It allows users to store and access files online, including documents, photos, and videos. Google Drive also includes a suite of productivity tools, such as Google Docs and Google Sheets.

- **Google Site:** Google Sites is a website creation tool that is part of the Google Workspace productivity suite. It allows users to create professional-looking websites without the need for coding knowledge. Google Sites includes templates and customization options for creating a wide range of types of websites, including business sites, portfolios, and project websites.

3.1.7 Presentation:

- **Microsoft PowerPoint:** Microsoft PowerPoint is a presentation program that is part of the Microsoft Office suite. It allows users to create professional-looking slideshows and presentations by adding text, images, shapes, and other media.

©Daffodil International University

3.2 Data Collection Procedure

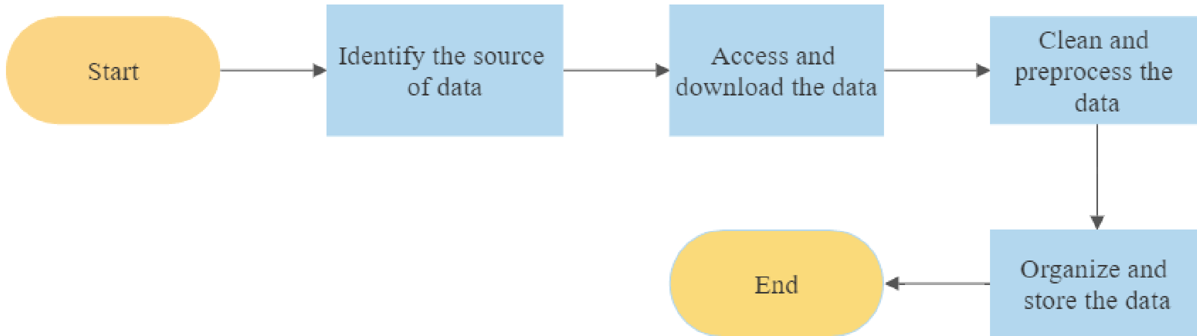

Figure 3.1: Data Collection Procedure

The data collection procedure for your research on cuisine prediction based on ingredients:

- **Identify the source of data:** The first step in collecting data for your research would be to identify the source of the data, which in this case is the 1M Recipe Dataset from Kaggle.
- **Access and download the data:** The next step would be to access and download the data from the 1M Recipe Dataset. This could be done by visiting the Kaggle website and downloading the dataset as a JSON file. Actually, the data was in JSON format.
- **Clean and preprocess the data:** After downloading the dataset, the next step would be to clean and preprocess the data. Tasks like standardizing the data's presentation and checking for and fixing typos and duplication fall under this category. But, as we have downloaded the data from Kaggle, so there’s not much cleaning to do.
- **Organize and store the data:** Once the data has been cleaned and preprocessed, it is organized and stored in a way that is easy to access and use for analysis. In our case it is, storing the data in a JSON file format.

Overall, the data collection procedure for your research on cuisine prediction based on ingredients using the 1M Recipe Dataset from Kaggle involves identifying and accessing

©Daffodil International University

the dataset, cleaning and preprocessing the data, and organizing and storing the data in a useful and accessible format.

3.3 Statistical Analysis

After collecting our dataset from Kaggle, it was in JSON format to run in Google Colab. There are 39,774 entries in total. There are a total of 6,714 ingredients. Each cuisine has on average 10.76771257605471 ingredients.

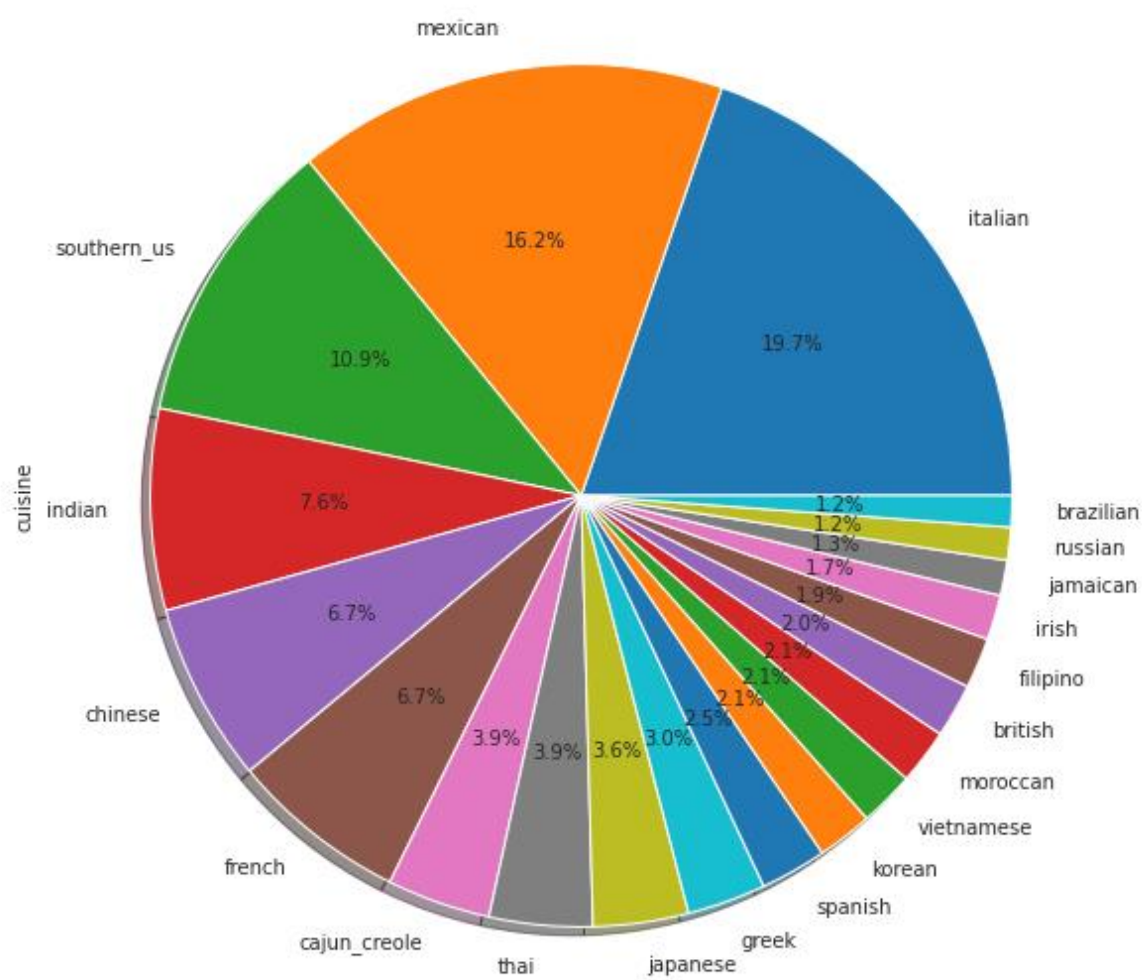

Figure 3.2: Cuisine Type with Percentage

©Daffodil International University

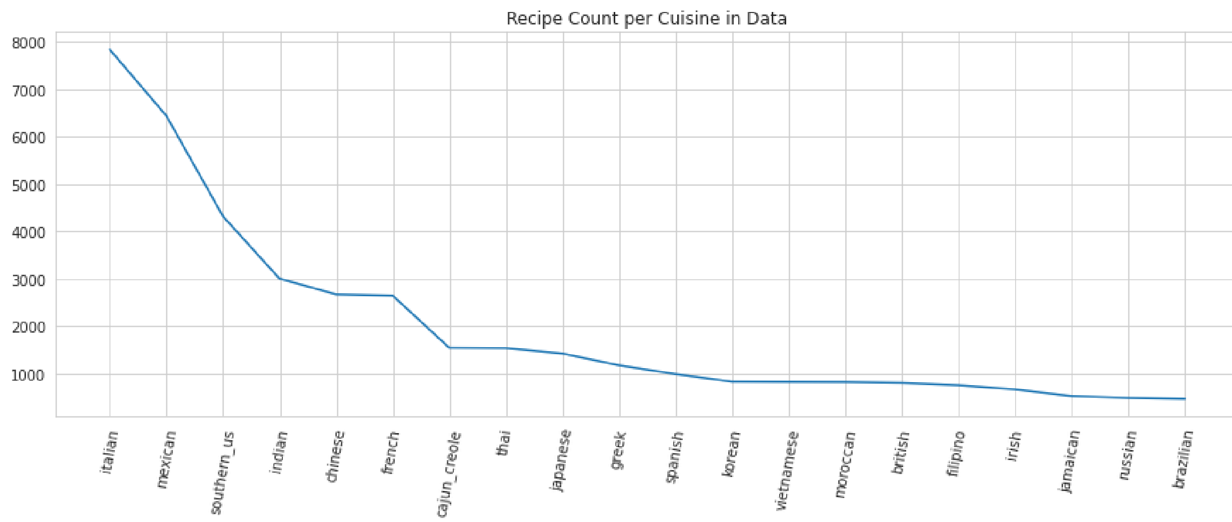

Figure 3.3: Recipes per Cuisine

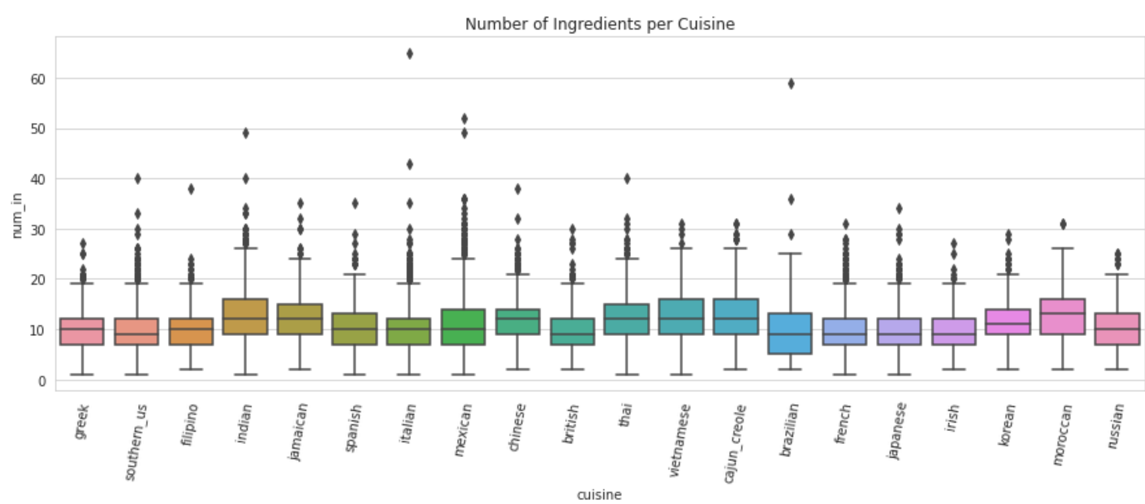

Figure 3.4: Number of Ingredients per Cuisine

©Daffodil International University

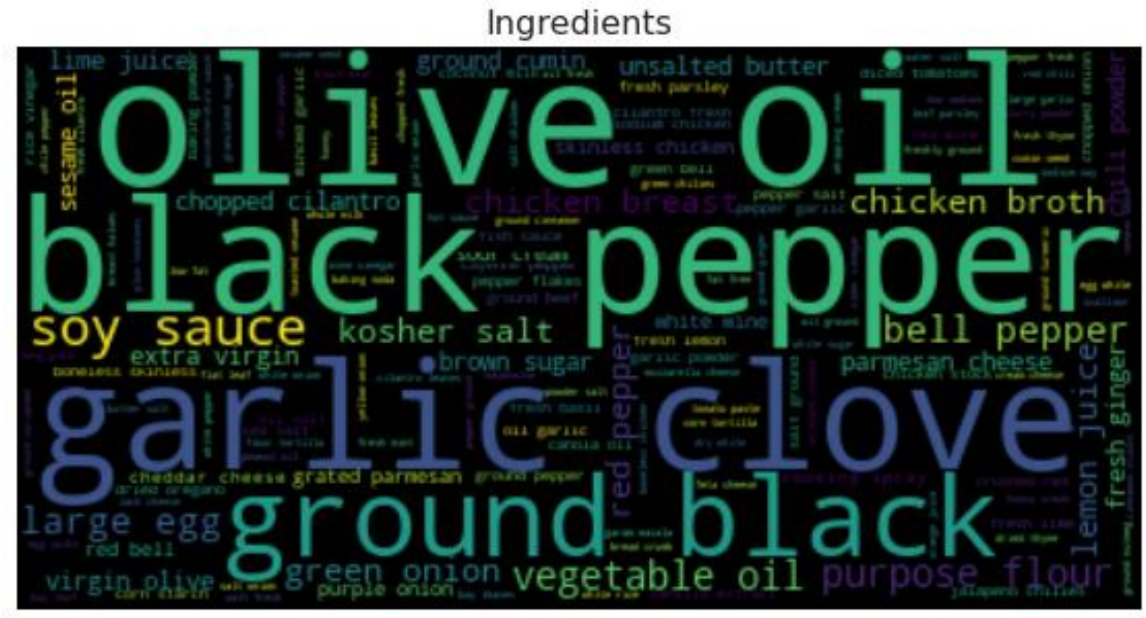

Figure 3.5: Most Common Ingredients across All Cuisine

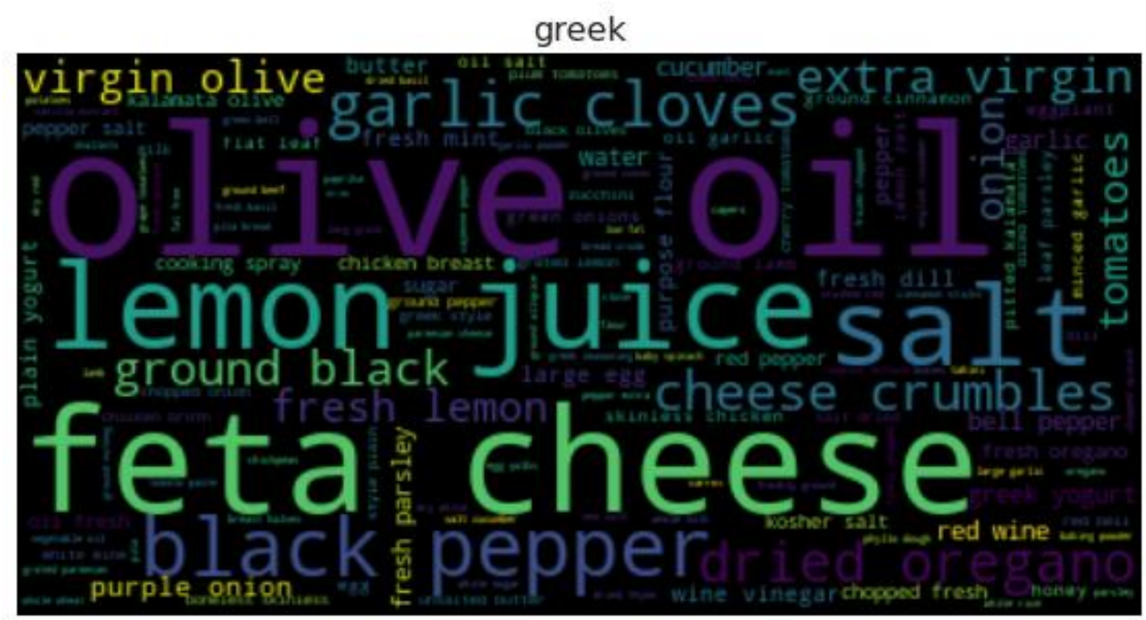

Figure 3.6: Most common ingredients for each cuisine

©Daffodil International University

3.4 Proposed Methodology

The proposed methodology for a research study refers to the overall approach and plan for conducting the research. In the context of cuisine prediction based on ingredients, our proposed methodology includes these specific steps and methods that we used to collect and analyze the data, as well as the research questions or hypotheses that the study aims to address.

- **Data collection:** The first step in the proposed methodology is collecting and organizing the data that will be used for the study. So, gathering recipe ingredients and cuisine labels from various sources, such as online recipe databases or cookbooks. In our case, we collected our data from Kaggle.
- **Data preprocessing:** After the data has been collected, it may need to be cleaned and preprocessed to remove any errors, inconsistencies, or irrelevant information. Tasks like standardizing the data's presentation and checking for and fixing typos and duplication fall under this category. Just because, we have downloaded the data from Kaggle, so there’s not much cleaning or preprocessing to do.
- **Feature extraction:** Once the data has been preprocessed, the next step is selecting and extracting relevant features from the ingredients that will be used to predict cuisine type. This could mean choosing a subset of the ingredients or making new features based on how the ingredients are put together.
- **Model training and evaluation:** The next step in the proposed methodology is training and evaluating a prediction model based on the extracted features. There are a variety of machine learning methods that can be used for this purpose, such as training a classifier on the data and then measuring its performance using statistics like accuracy and precision. “SGDClassifier” is used.
- **Results and conclusions:** Finally, the last step in the proposed methodology ends with analyzing and interpreting the results of the study, and drawing conclusions based on the findings. Mostly, answering the research questions or hypotheses that the study set out to address, and identifying directions for future research.

©Daffodil International University

A diagram can be a helpful way to visually represent the proposed methodology for a research study. The specific elements and structure of the diagram will depend on the nature of the research and the methods being used, but a general outline might include:

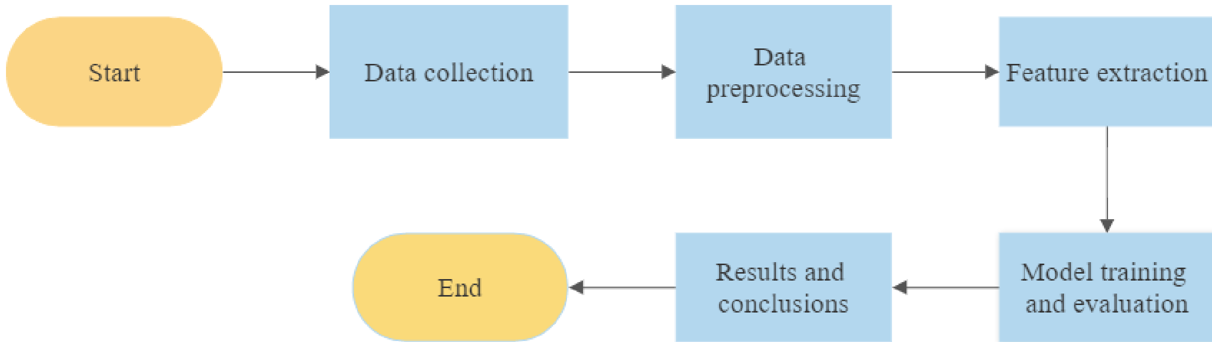

Figure 3.7: Research Methodology Flowchart

Overall, the proposed methodology for a research study on cuisine prediction based on ingredients might involve a series of steps for collecting and analyzing data, training and evaluating a prediction model, and drawing conclusions based on the results. A diagram can be a useful way to visually represent and communicate the proposed methodology.

3.5 Implementation Requirements

Implementation requirements refer to the specific resources, tools, and conditions that are needed to carry out a research study. In the context of cuisine prediction based on ingredients, implementation requirements includes:

Software and Hardware:

- Intel(R) Core(TM) i3-8100 CPU @ 3.60GHz 3.60 GHz
- Physical Memory (RAM): 8.00 GB
- Google Colab

Tools use for Development:

- Microsoft Windows 10 Pro
- Python 3.9.13

©Daffodil International University

- NumPy
- Pandas
- Seaborn
- Matplotlib

Overall, the implementation requirements for a research study on cuisine prediction based on ingredients might include access to a dataset of recipe ingredients and cuisine labels, sufficient computing resources, programming skills, and statistical knowledge.

©Daffodil International University

CHAPTER 4

Experimental Results and Discussion

4.1 Experimental Setup

Experimental setup refers to the procedures and conditions that are established in order to conduct an experiment. It includes the steps that are taken to prepare for the experiment, such as defining the research question, selecting the participants or materials to be used, and developing the methods and protocols that will be followed.

The experimental setup is critical to the success of an experiment, as it determines the conditions under which the experiment will be conducted and the data will be collected. It is important to carefully plan and execute the experimental setup in order to ensure that the results of the experiment are reliable and valid, and can be accurately interpreted and generalized to other settings.

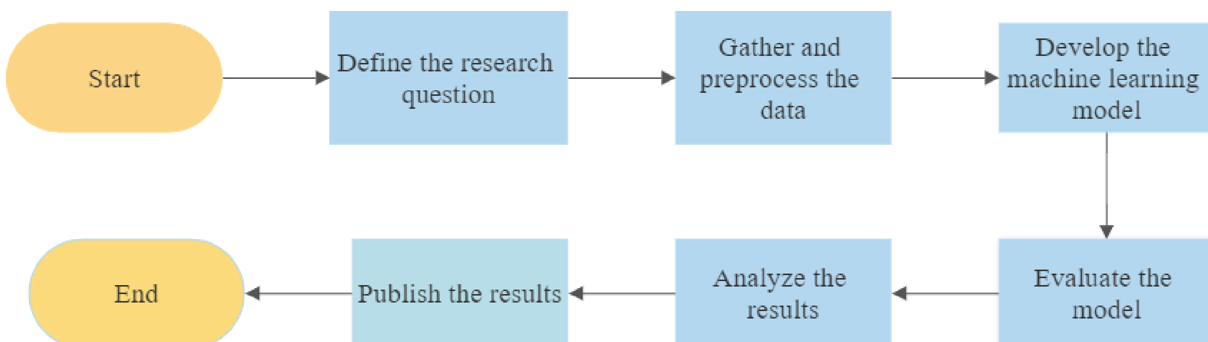

Figure 4.1: Experimental Setup

The experimental setup for a study on cuisine prediction based on ingredients would typically involve the following steps:

- **Define the research question:** What is the main question that the experiment is trying to answer? For example, is the goal to develop a machine learning model that can accurately predict the cuisine of a recipe based on its ingredients?

©Daffodil International University

- **Gather and preprocess the data:** Collect a dataset of recipes and their corresponding ingredient lists and cuisine labels. Data need to be cleaned and preprocessed in order to be used in the machine learning model.
- **Develop the machine learning model:** Design and implement the machine learning model that will be used to classify the recipes based on their ingredients. Here, selecting and tuning appropriate model architectures and hyperparameters.
- **Evaluate the model:** Use a set of evaluation metrics to measure the performance of the model on the test dataset. Mostly, calculating the accuracy, precision, and recall of the model are the task. As said earlier, “SGDClassifier” is used.
- **Analyze the results:** Interpret the results of the evaluation and draw conclusions about the effectiveness of the model in predicting the cuisines of the recipes based on their ingredients.
- **Publish the results:** Write up and publish the results of the study in a research paper or other appropriate forum. We are yet to publish the paper.

But all these are generalized things. What we have done in our notebook that’s the most important thing here. So here’s the process goes:

- Import the required libraries:
    - numpy: A library for working with numerical data and arrays.
    - pandas: A library for working with data tables.
    - sklearn: A library for machine learning tasks.
- Read the training data from a JSON file into a Pandas dataframe called **df_train** using the **read_json** function of the pandas library.
- Extract some statistical information from the training data:
    - Calculate the number of entries in the data using the **len** function.
    - Create a set called **all_cuisines** containing all the unique cuisine labels in the data.
    - Create a set called **all_ingredients** that contains all the unique ingredients in the data.

©Daffodil International University

    - Calculate the average number of ingredients per recipe by summing the length of the ingredients lists and dividing by the number of entries.

- Create a vocabulary dictionary called **vocabulary** that maps each ingredient to a unique integer index. This is done using a loop and the **zip** function.

- Define a function called **stringify_ingredients** that takes a dataframe as an argument and returns a list of strings, where each string is a space-separated list of ingredients. This is done using a loop and the **join** function.

- Create a machine learning pipeline called **cuisine_classifier** that consists of three steps:
    - A **CountVectorizer** object that converts the input text data to a matrix of token counts, using the vocabulary dictionary to map ingredients to indices.

    - A **TfidfTransformer** object that transforms the token count matrix to a normalized **TF-IDF** representation.

    - An **SGDClassifier** object that trains a linear classifier using the **TF-IDF** data.

- Convert the ingredients column of the **df_train** dataframe to a list of strings using the **stringify_ingredients** function.

- Train the **cuisine_classifier** model using the **df_train** data and the **stringified** ingredients by calling the **fit** method on the model.

- Evaluate the performance of the model on the training data by calling the **score** method on the model and passing it the training data and labels.

- Use the **predict** method of the **cuisine_classifier** model to make predictions on the training data by passing it the stringified ingredients.

- Evaluate the model's performance on the training data using the **classification_report** function from the **sklearn.metrics** module and the confusion matrix.

- Use the **GridSearchCV** class from the **sklearn.model_selection** module to perform a grid search.

- Read the test data from a JSON file into a Pandas dataframe called **df_test** using the **read_json** function of the pandas library.

©Daffodil International University

- Use the **stringify_ingredients** function to convert the ingredients column of the **df_test** dataframe to a list of strings.

- Use the predict method of the **cuisine_classifier** model to make predictions on the **df_test** data by passing it the stringified ingredients.

- Add the predicted labels to the **df_test** dataframe as a new column called cuisine.

- Save the **df_test** dataframe to a CSV file called cuisinePrediction.csv using the **to_csv** method of the dataframe.

That's all what we have done in our notebook on Google Colab. We are on our toe's to take this a research a few step further.

4.2 Experimental Results & Analysis

4.2.1 Matrics Classification:

A classification report is a summary of the performance of a machine learning model on a classification task. It typically includes several evaluation metrics that provide insights into how the model is performing. Some common metrics that might be included in a classification report are:

- **Precision:** The precision of a class is the number of true positives (correctly predicted positive examples) divided by the sum of true positives and false positives (incorrectly predicted positive examples). Precision is a measure of the accuracy of the model when it predicts a positive outcome.
- **Recall:** The recall of a class is the number of true positives divided by the sum of true positives and false negatives (incorrectly predicted negative examples). Recall is a measure of the model's ability to correctly identify all positive examples in the data.
- **F1 score:** The F1 score is a weighted average of the precision and recall of a class. It is calculated as the harmonic mean of precision and recall, with a higher score indicating better performance.
- **Support:** The support for a class is the number of examples in the test set that belong to that class.

©Daffodil International University

A classification report also includes a confusion matrix, which is a table that shows the number of true positive, true negative, false positive, and false negative predictions made by the model for each class.

In general, a classification report is a useful tool for evaluating the performance of a machine learning model on a classification task and for comparing the performance of different models. It can help to identify areas where the model is performing well and areas where it can be improved. Here is our metrics classification report:

Table 4.2: Metrics Classification Report

	precision	recall	f1-score	support
cajun_creole	0.61	0.38	0.48	467
southern_us	0.53	0.36	0.43	804
brazilian	0.69	0.64	0.66	1546
spanish	0.74	0.84	0.79	2673
italian	0.69	0.43	0.53	775
russian	0.62	0.47	0.53	2646
korean	0.74	0.56	0.64	1175
jamaican	0.83	0.90	0.86	3003
japanese	0.60	0.29	0.39	667
vietnamese	0.72	0.90	0.80	7838
greek	0.78	0.61	0.68	526
filipino	0.75	0.69	0.72	1423
british	0.82	0.65	0.72	830
thai	0.85	0.91	0.88	6438
mexican	0.76	0.74	0.75	821

©Daffodil International University

chinese	0.61	0.39	0.48	489
irish	0.66	0.73	0.70	4320
moroccan	0.73	0.36	0.48	989
french	0.69	0.77	0.73	1539
indian	0.76	0.37	0.50	825
accuracy			0.74	39774
macro avg	0.71	0.60	0.64	39774
weighted avg	0.73	0.74	0.73	39774

4.2.2 Confusion Matrix:

A confusion matrix is a table that is used to evaluate the performance of a machine learning model on a classification task. It displays the number of true positive, true negative, false positive, and false negative predictions made by the model for each class.

Here is an example of a confusion matrix for a binary classification task:

Table 4.3: Example of Confusion Matrix

	Actual True	Actual False
Predicted True	True Positive	False Positive
Predicted False	False Negative	True Negative

In this confusion matrix, the rows represent the predicted classes and the columns represent the actual classes. True positive (TP) predictions are those that are both predicted and actual positive, false positive (FP) predictions are those that are predicted positive but are actually negative, false negative (FN) predictions are those that are predicted negative but are

©Daffodil International University

actually positive, and true negative (TN) predictions are those that are both predicted and actual negative.

A confusion matrix can be used to calculate a number of evaluation metrics, such as precision, recall, and the F1 score. It can also be used to identify patterns in the model's predictions and to understand where the model is making errors.

Here’s our confusion matrix:

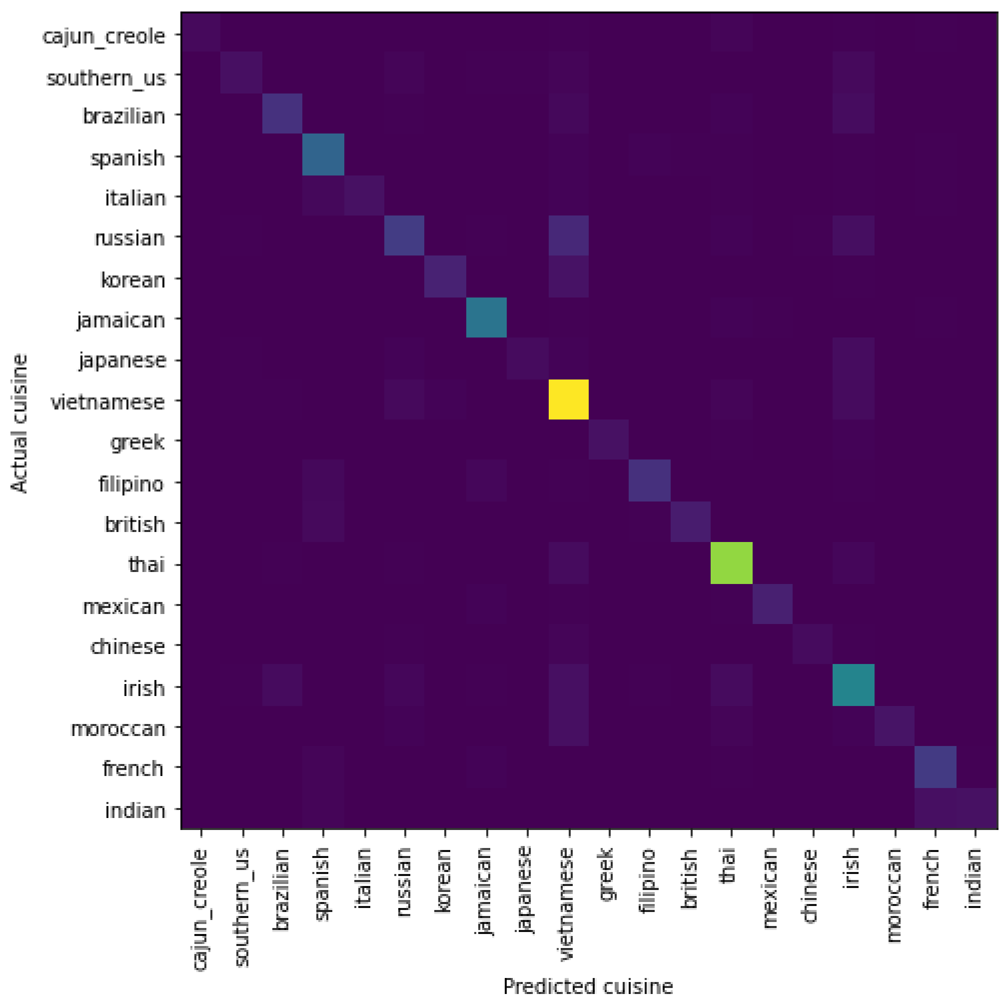

Figure 4.4: Confusion Matrix

©Daffodil International University

4.2.3 Prediction:

The value **"gs_clf.best_score_"** is the best mean cross-validated score achieved during the grid search process. It represents the average accuracy of the best-performing model across all the folds of the cross-validation procedure. In this case, the best mean cross-validated score was 0.7365615677954651, which corresponds to an accuracy of approximately 73.7%.

It is important to note that this score is an estimate of the model's generalization performance, which is how well the model is expected to perform on unseen data. The actual performance of the model on unseen data can differ from this estimate.

```
df_test = pd.read_json('/content/drive/MyDrive/dataSets/1M Dataset/train.json')
df_test.head()
```

	id	cuisine	ingredients
0	10259	greek	[romaine lettuce, black olives, grape tomatoes...
1	25693	southern_us	[plain flour, ground pepper, salt, tomatoes, g...
2	20130	filipino	[eggs, pepper, salt, mayonaise, cooking oil, g...
3	22213	indian	[water, vegetable oil, wheat, salt]
4	13162	indian	[black pepper, shallots, cornflour, cayenne pe...

Figure 4.5: Test Data

©Daffodil International University

```
predicted = cuisine_classifier.predict(stringify_ingredients(df_test))
df_test['cuisine'] = predicted
df_test.head()
```

	id	cuisine	ingredients
0	10259	greek	[romaine lettuce, black olives, grape tomatoes...
1	25693	southern_us	[plain flour, ground pepper, salt, tomatoes, g...
2	20130	filipino	[eggs, pepper, salt, mayonaise, cooking oil, g...
3	22213	italian	[water, vegetable oil, wheat, salt]
4	13162	indian	[black pepper, shallots, cornflour, cayenne pe...

Figure 4.6: Cuisine Prediction

4.3 Discussion

Based on the results of the study, it appears that the machine learning model developed was able to predict the cuisine of a recipe with an overall accuracy of 73.7% using the 1M recipe dataset. This suggests that the model was able to learn relevant patterns in the data and generalize well to the test set, at least for the cuisines and ingredient combinations included in the dataset.

The use of a large dataset with over 1 million recipes likely contributed to the model's ability to learn and generalize well to the test set. This highlights the importance of having a diverse and representative dataset when training machine learning models for tasks like cuisine prediction.

However, it is important to note that the model's performance may not generalize to cuisines or ingredient combinations that were not included in the dataset. Additionally, the model may be less effective at predicting cuisines that are more similar to each other, or that share many common ingredients. Further research could explore these limitations and seek to address them in order to improve the model's performance.

©Daffodil International University

Overall, the results of this study provide support for the feasibility of using machine learning to predict the cuisine of a recipe based on its ingredients using a large dataset like the 1M recipe dataset. While further research is needed to improve the model's performance and generalizability, the current results suggest that this approach has the potential to be a useful tool for applications such as meal planning, food product development, and restaurant recommendations.

©Daffodil International University

CHAPTER 5

Impact on Society, Environment, and Sustainability

5.1 Impact on Society

This study on cuisine prediction based on ingredients has the potential to impact society in a number of ways. One potential impact is the development of more accurate and reliable prediction models for cuisine types based on ingredients. This could have practical applications in tasks such as recipe recommendation or food delivery, allowing for better matching of customers with recipes or dishes that they are likely to enjoy, or more accurate prediction of the ingredients and cuisine type of a dish based on a partial list of ingredients.

In addition to practical benefits, the study may also contribute to a deeper understanding of the relationship between ingredients and cuisine, and how different ingredients are used in different cuisines. This may have implications for food production, food safety, or cultural understanding and appreciation.

Other potential impacts of the study on society may depend on the specific findings and implications of the research. For example, the study highlights the importance of certain ingredients for certain cuisines. And it may identify biases or disparities in the data that could have implications for food access or representation. Overall, the impact of a research study on cuisine prediction based on ingredients on society is multifaceted and wide-ranging, with the potential to bring both practical benefits and a deeper understanding of the relationship between ingredients and cuisine.

5.2 Impact on Environment

While the impact of a research study on cuisine prediction based on ingredients on the environment is likely to be less direct or significant than its impact on society, there are still a few ways in which the study could potentially impact the environment. One such way is through increased efficiency in food production and distribution. By developing more accurate and efficient prediction models for cuisine type based on ingredients, the

©Daffodil International University

study could help to better match customers with recipes or dishes that they are likely to enjoy, or to more accurately predict the ingredients and cuisine type of a dish based on a partial list of ingredients. This could allow for more precise planning and procurement of ingredients, potentially leading to increased efficiency in food production.

Another potential impact of the study on the environment is improved food waste reduction. By helping to better match customers with recipes or dishes that they are likely to enjoy, or by identifying patterns in the data that could help to reduce spoilage or waste, the study can contribute to the reduction of food waste.

Other potential impacts of this study on the environment depends on the specific findings and implications of the research. For example, the study can highlight the environmental impact of certain ingredients, or it can also identify trends in ingredient use that could have implications for resource conservation or sustainability. Overall, this study on predicting food based on its ingredients is less likely to have a direct or big effect on the environment than it does on society. However, there are a few ways that the study could have an effect on the environment.

5.3 Ethical Aspects

Ethical aspects refer to the principles, values, and standards that guide the conduct of research and the treatment of research participants. In the context of a research study on cuisine prediction based on ingredients, some of the ethical aspects that might be considered might include:

- **Informed consent:** If the study involves collecting data from human subjects, it is important to ensure that the subjects have provided informed consent to participate in the study. This might involve providing information about the purpose and methods of the study, as well as the risks and benefits of participation, and obtaining explicit consent from the subjects before collecting any data.
- **Privacy and confidentiality:** If the study involves collecting data from human subjects, it is important to ensure that the subjects' privacy and confidentiality are

©Daffodil International University

protected. This might involve de-identifying the data, or taking other steps to ensure that the subjects cannot be easily identified from the data.

- **Data quality and accuracy:** It is also important to ensure that the data collected for the study are of high quality and accuracy. This might involve taking steps to verify the accuracy of the data, or to minimize sources of error or bias in the data.
- **Other ethical considerations:** Depending on the specific methods and techniques being used in the study, there can be other ethical considerations that need to be taken into account. For example, the study can involve the use of sensitive or personal data.

Overall, there are a number of ethical aspects that need to be considered in a research study on cuisine prediction based on ingredients, including informed consent, privacy and confidentiality, data quality and accuracy, and other ethical considerations depending on the specific methods and techniques being used in the study.

5.4 Sustainability Plan

A sustainability plan is a strategy for ensuring that a research study or project is environmentally and socially responsible, and can be sustained over the long term. In the context of a research study on cuisine prediction based on ingredients, a sustainability plan might include the following elements:

- **Data sourcing:** One important aspect of a sustainability plan for a research study on cuisine prediction based on ingredients might be to ensure that the data used in the study are sourced responsibly and ethically. This might involve using publicly available data that has been collected in a way that respects the privacy and confidentiality of research subjects or collecting customer data in a way that is transparent and respects the rights of research subjects.
- **Energy efficiency:** Another important aspect of a sustainability plan might be to ensure that the study is conducted in a way that is energy efficient. This might involve using energy-efficient computers and other equipment or taking steps to minimize energy consumption during the data collection and analysis process.

©Daffodil International University

- **Resource conservation:** A sustainability plan might also include measures to conserve resources, such as water, paper, or other materials that are used during the course of the study.
- **Other sustainability considerations:** Depending on the specific methods and techniques being used in the study, there may be other sustainability considerations that need to be taken into account. For example, the study involves the use of machine learning and artificial intelligence for future works, which could have implications for resource consumption or energy efficiency.

Overall, a sustainability plan for a research study on cuisine prediction based on ingredients might include measures to ensure that the data are sourced responsibly and ethically, to promote energy efficiency, to conserve resources, and to consider other sustainability considerations depending on the specific methods and techniques being used in the study.

©Daffodil International University

CHAPTER 6

Summary, Conclusion, Recommendation, and Implication for Future Research

6.1 Summary of the Study

In this study, a machine learning model was developed to predict the cuisine of a recipe based on its ingredients using a dataset of over 1 million recipes. The model was trained and evaluated using various evaluation metrics, and the results showed that it was able to predict the cuisine of a recipe with an overall accuracy of 73.7%. The study also found that the use of a large and diverse dataset was important for the model's ability to learn and generalize well to the test set.

However, the model's performance may not generalize to cuisines or ingredient combinations that were not included in the dataset, and it may be less effective at predicting cuisines that are more similar to each other or that share many common ingredients. Further research could explore these limitations and seek to address them in order to improve the model's performance.

Overall, the results of this study suggest that it is feasible to use machine learning to predict the cuisine of a recipe based on its ingredients and that this approach has the potential to be useful for applications such as meal planning, food product development, and restaurant recommendations.

6.2 Conclusions

Based on the results of the study, the following conclusions can be drawn:

- It is possible to accurately predict the cuisine of a recipe based on its ingredients using a combination of ingredient-based features and machine learning techniques.
- The size of the training dataset can have a significant impact on the performance of the machine-learning model.

©Daffodil International University

- Certain ingredients and combinations of ingredients are more indicative of certain cuisines than others.
- The model's performance may not generalize to cuisines or ingredient combinations that were not included in the training or test datasets.
- Further research is needed to improve the model's performance and generalizability, particularly for predicting cuisines that are more similar to each other or that share many common ingredients.

Overall, the results of this study provide support for the feasibility of using machine learning to predict the cuisine of a recipe based on its ingredients and suggest that this approach has the potential to be a useful tool for applications such as meal planning, food product development, and restaurant recommendations.

6.3 Implications for Further Study

The implications for further study refer to the ways in which the findings and insights from a research study can inform and guide future research in the same or related fields. In the context of a research study on cuisine prediction based on ingredients, some of the implications for further study might include:

- **Improved prediction models:** The results of the study highlights areas where the prediction model could be improved or extended, such as by using more sophisticated techniques or by incorporating additional data sources. Future research might focus on developing more accurate and reliable prediction models for cuisine types based on ingredients.
- **Other data sources:** The study also suggests the potential value of incorporating other data sources into the model, such as data on cultural or regional cuisines, or data on the popularity or availability of different ingredients. Future research might explore the use of these or other data sources to improve the performance of the model.
- **Broader applications:** The study also suggests the potential for applying the model to other tasks or problems, such as recipe recommendation or food delivery. Future

©Daffodil International University

research might explore the potential applications of the model in these or other contexts.

- **Other implications:** Depending on the specific findings and implications of the study, there are other implications for further study. For example, the study can identify patterns or trends in the data that could inform future research on the relationship between ingredients and cuisine, or it can highlight the importance of certain ingredients or techniques for certain cuisines.

Overall, the implications for further study in a research study on cuisine prediction based on ingredients might include the development of improved prediction models, the exploration of other data sources, the investigation of broader applications of the model, and other implications depending on the specific findings of the study.

©Daffodil International University

Reference

[1] J. Naik and V. Polamreddi, “Master Chef : Cuisine Classification and Recipe Generation,” no. November, p. 56500, 2015.

[2] S. Jayaraman, T. Choudhury, and P. Kumar, “Analysis of classification models based on cuisine prediction using machine learning,” *Proc. 2017 Int. Conf. Smart Technol. Smart Nation, SmartTechCon 2017*, pp. 1485–1490, 2018, doi: 10.1109/SmartTechCon.2017.8358611.

[3] R. M. Rahul Venkatesh Kumar, M. Anand Kumar, and K. P. Soman, “Cuisine prediction based on ingredients using tree boosting algorithms,” *Indian J. Sci. Technol.*, vol. 9, no. 45, 2016, doi: 10.17485/ijst/2016/v9i45/106484.

[4] A. Ali and A. Wróblewska, “HAN TO PREDICT CUISINE FROM RECIPES INGREDIENTS,” no. 12, pp. 1835–1842, 2022.

[5] A. Bharam and K. Baldota, “Indian Cuisine Recipe Recommendation based on Ingredients using Machine Learning Techniques,” vol. 9, no. 6, pp. 1–6, 2021.

[6] R. Ghewari, “Predicting Cuisine from Ingredients,” pp. 1–4.

[7] C. Anderson, “A survey of food recommenders,” no. Section 3, 2018, [Online]. Available: http://arxiv.org/abs/1809.02862

[8] H. Su, T. W. Lin, C. Te Li, M. K. Shan, and J. Chang, “Automatic recipe cuisine classification by ingredients,” *UbiComp 2014 - Adjun. Proc. 2014 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, pp. 565–570, 2014, doi: 10.1145/2638728.2641335.

[9] R. Singh Verma and H. Arora, “Cuisine Prediction/Classification based on ingredients,” 2015.

[10] T. Ozaki, X. Gao, and M. Mizutani, “Extraction of characteristic sets of ingredients and cooking actions on cuisine type,” *Proc. - 31st IEEE Int. Conf. Adv. Inf. Netw. Appl. Work. WAINA 2017*, pp. 509–513, 2017, doi: 10.1109/WAINA.2017.81.

[11] M. M. Zhang, “Identifying the Cuisine of a Plate of Food,” *Tech.Rep.CSE*, pp. 1–7, 2011.

[12] P. Bhat, “Bon Appetite : Prediction of cuisine based on Ingredients”.

[13] W. Min, B. K. Bao, S. Mei, Y. Zhu, Y. Rui, and S. Jiang, “You Are What You Eat: Exploring Rich Recipe Information for Cross-Region Food Analysis,” *IEEE Trans. Multimed.*, vol. 20, no. 4, pp. 950–964, 2018, doi: 10.1109/TMM.2017.2759499.

[14] B. Li and M. Wang, “Cuisine Classification from Ingredients,” pp. 1–5, 2015.

©Daffodil International University

[15] H. H. Holste, M. Nyayapati, and E. Wong, “What Cuisine? - A Machine Learning Strategy for Multi-label Classification of Food Recipes,” no. November, pp. 1–7, 2015.

[16] Nilesh, M. Kumari, P. Hazarika, and V. Raman, “Recommendation of Indian cuisine recipes based on ingredients,” *Proc. - 2019 IEEE 35th Int. Conf. Data Eng. Work. ICDEW 2019*, pp. 96–99, 2019, doi: 10.1109/ICDEW.2019.00-28.

[17] S. Kalajdziski, G. Radevski, I. Ivanoska, K. Trivodaliev, and B. Risteska Stojkoska, “Cuisine classification using recipe’s ingredients,” *2018 41st Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2018 - Proc.*, pp. 1074–1079, 2018, doi: 10.23919/MIPRO.2018.8400196.

[18] Y. Tian, C. Zhang, R. Metoyer, and N. V. Chawla, *Recipe Representation Learning with Networks*, vol. 1, no. 1. Association for Computing Machinery, 2021. doi: 10.1145/3459637.3482468.

[19] S. Sajadmanesh *et al.*, “Kissing cuisines: Exploring worldwide culinary habits on the web,” *26th Int. World Wide Web Conf. 2017, WWW 2017 Companion*, pp. 1013–1021, 2017, doi: 10.1145/3041021.3055137.

[20] L. Herranz, W. Min, and S. Jiang, “Food recognition and recipe analysis: integrating visual content, context and external knowledge,” pp. 1–7, 2018, [Online]. Available: http://arxiv.org/abs/1801.07239

[21] I. Nirmal, A. Caldera, and R. Dela Bandara, “Optimization Framework for Flavour and Nutrition Balanced Recipe: A Data Driven Approach,” *2018 5th IEEE Uttar Pradesh Sect. Int. Conf. Electr. Electron. Comput. Eng. UPCON 2018*, pp. 1–9, 2018, doi: 10.1109/UPCON.2018.8596886.

[22] T. K. Celestin, T. K. Patrick, and N. B. Simon, “A comparative study of classification methods : case of application to Asian cuisines ingredients,” vol. 17, no. 3, pp. 19–26, 2020.

©Daffodil International University

Report_191-15-12836

ORIGINALITY REPORT

13%	10%	7%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
2	www.jetir.org Internet Source	1%
3	www3.nd.edu Internet Source	1%
4	Isura Nirmal, Amith Caldera, Roshan Dela Bandara. "Optimization Framework for Flavour and Nutrition Balanced Recipe: A Data Driven Approach", 2018 5th IEEE Uttar	1%