Republic of Iraq
Ministry of Higher Education and Scientific Research
University of Babylon

Proposing a Restaurants Recommender System based on Implicit Feedback of Food Photos using Light Graph Convolution Network

A Thesis

Submitted to the Council of the College of Information Technology for Postgraduate Studies of University of Babylon in Partial Fulfillment of the Requirements for the Degree of Master in Information Technology-Software

By

Mustafa Saleem Manji Mahdi

Supervised by

Lecturer Dr. Wadhah Razooqi Baiee

2022 D.C. **1444 A.H.**

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

﴿إِنَّ اللَّهَ وَمَلَائِكَتَهُ يُصَلُّونَ عَلَى النَّبِيِّ ۚ يَا

أَيُّهَا الَّذِينَ آمَنُوا صَلُّوا عَلَيْهِ وَسَلِّمُوا

تَسْلِيمًا ﴾

صدق الله العظيم

سورة الأحزاب اية ٥٦

Declaration

I hereby declare that this Thesis, submitted to the University of Babylon in partial fulfillment of requirements for the degree of Master in Information Technology \ Software, has not been submitted as an exercise for a similar degree at any other University. I also certify that this work described here is entirely my own except for experts and summaries whose sources are appropriately cited in the references.

Signature:

Name: Mustafa Saleem Manji

Date: / 10 /2022

Supervisor Certification

I certify that this Thesis was prepared under my supervision at the Department of Software / Collage of Information Technology / University of Babylon by Mustafa Saleem Manji as partial fulfillment of the requirements for the Master in Information Technology degree.

Signature:

Name: Lecturer. Dr. Wadhah Razooqi Baiee

Date: / 10 / 2022

The Head of the Department Certification

In view of the available recommendation, we forward this Thesis for debate by the examining committee.

Signature:

Name: Asst.Prof. Ahmed Saleem Al-Saffar

Date: / 10 / 2022

Acknowledgment

First and foremost, praises and thanks to Allah, the Almighty, for His showers of blessings throughout my work to complete the research successfully.

I would like to express my deep and sincere gratitude to my supervisor, Lecturer. Dr. Wadhah Baiee for his invaluable guidance throughout this research. His dynamism, vision, sincerity, and motivation have deeply inspired me. He has taught me the methodology to carry out research and to present my work as best as possible. It was a great privilege and honour to work and study under his guidance. I am extremely grateful for what he has offered me. I would also like to thank him for his friendship, empathy, and great sense of humour. I am extending my heartfelt thanks to his wife and family for their acceptance and patience during my discussion with him on research work and thesis preparation.

I am extremely grateful to my parents for their love, prayers, caring, and sacrifices for educating and preparing me for my future. I am very thankful to my wife and my kids for their love, understanding, prayers, and continuing support in completing this research. Moreover, I thank my sister and brothers for their support and valuable help.

I would like to thank my friend Mohammed Baqer for his constant encouragement and help to pass this stage throughout the study period. I express my special thanks to the Head of the Software Department,in the College of Information Technology, at the University of Babylon, Assistant. Prof. Dr. Ahmed Saleem Al-Saffar, and all professors of in the Software Department for their advice that helped me and their genuine support throughout this research. I would also like to thank all staff in the Software Department/ College of Information Technology/ University of Babylon, for their kindness.

Finally, my thanks go to all people who have supported me in completing this research either directly or indirectly.

Abstract

Dietary decisions affect obesity, arterial occlusion, and heart disease. Human health care specialists such as dietitians cannot be with every user as guiding person manually toward optimal choices is a difficult process. Accordingly, the combination of the automated adaptive guidance with expert knowledge can help provide a scale health advice technologically without human intervention. This thesis focuses on recommending best restaurants to users, depending on their information. The proposed system analysis the restaurants food photos to extract new nutrition information and improve the accuracy of predicting best restaurants based on nutrition facts.

The thesis adopts the Convolutional Neural Network (CNN) technique to analyse food photos and extract a list of ingredients of meal dishes. The extracted ingredients from photos are used to calculate nutrition information for each meal. The extracted latent features from photos are combined with other users' information to check restaurant similarities to their neighbours. Then the system picks a group of the nearest restaurants to recommend similar preferences to the current user. Thus, a collaborative filtering recommendation system is adopted in this work.

The proposed system is evaluated by the recall and NDCG of the recommended products for all users during the training set period and comparing them to the most recent period of user preferences as a testing set. Recall values range between 1% and 97%, and the average is 47%, whereas the NDCG values range between 23% and 50% in which 33 % is the average of the entire system. Yelp has functioned as a data-driven application and has made a dataset containing business information, user information, and check-in data. The experiment is implemented on the Yelp dataset that was crawled by the proposed system.

Table of Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

Abbreviation	Description
A.I.	Artificial Intelligence
AGG	Aggregation
BGNN	Boundary Graph Neural Networks
BPR	Bayesian Personalized Ranking
BPR-MF	Bayesian Personalized Ranking optimizations for Matrix Factorization
CF	Collaborative Filtering
CNN	Convolutional Neural Networks
CVPR	Computer Vision and Pattern Recognition
DNN	Deep Neural Network
EXIF	Exchangeable Image File
GCN	Graph Convolution Network
GPS	Global Positioning System
HMMs	Hidden Markov Models
HMRF	Hidden Markov Random Field
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
JPEG	Joint Photographic Experts Group
KDD	Knowledge Data Discovery
LGC	Light Graph Convolution
LightGCN	Light Graph Convolutional Neural
LSTM	Long Short-Term Memory
MLP	Multilayer Perceptron
NAIS	Neural Attentive Item Similarity
NDCG	Normalized Discounted Cumulative Gain

NGCF	Neural Graph Collaborative Filtering
RS	Recommender System
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
UDD	User Data Discovery

Thesis Related Publication

• Name of Conference: 2nd ALMuthanna International Conference on Engineering Science and Technology (MICST-2022). Almuthana University, Almuthana, Iraq.

• Paper Title: Survey on Implicit Feedbacks Extraction based on Yelp Dataset using Collaborative Filtering.

• Authors: 1) Mustafa Al-Saffar 2) Wadhah R. Baiee. Software Department, College of Information Technology, Babylon University, Iraq.

DOI: 10.1109/MICEST54286.2022.9790172

Email: mustafaalsaffar.sw.msc@student.uobabylon.edu.iq

Email: wadhah.baiee@uobabylon.edu.iq

• Name of Journal: Bulletin of Electrical Engineering and Informatics (BEEI).

• Paper Title: Nutrition information estimation from food photos using machine learning based on multiple datasets.

• Authors: 1) Mustafa Al-Saffar 2) Wadhah R. Baiee. Software Department, College of Information Technology, Babylon University, Iraq.

DOI: https://doi.org/10.11591/eei.v11i5.4007

Email: mustafaalsaffar.sw.msc@student.uobabylon.edu.iq

Email: wadhah.baiee@uobabylon.edu.iq

CHAPTER ONE

General Introduction

Chapter One

General Introduction

1.1 Background

E-trade websites flourish rapidly and allow hundreds of thousands of items to be sold [1]. Selecting an item from many such options requires using an extra tool known as a recommender system [2], [3]. The recommender system (RS) enables consumers to find out things they might not have noticed otherwise. It gathers information about the user's preferred items and then suggests them [4]. Several e-commerce businesses have adopted the Collaborative Filtering (CF) strategy [4], including Amazon Prime Video, Yelp [5], marketplace, and AliExpres. The popularity of CF approaches is because of sharing the participants' interests. Calculating the typical user rating [6] is utilized to watch identical users or things. When sufficient rating information is available, CF approaches are practical [4].

In addition, food from important things that represent the quality of the place and the places with the highest values are those that can be directly tied to the variety of food available. However, one of the primary challenges is managing users' health effectively at all times and locations.

According to previous works, a recommendation system incorporating user personal information and context and food resources will shape the future of health behaviors. Explored this principle by directing users toward healthy food selections tailored to their biological and contextual characteristics [7].

Most people ignore notable foods and restaurants in their neighborhoods [8]. This is true not only for foreigners or visitors but also for the indigenous people. With the proliferation of restaurants, clients frequently become perplexed about the ideal restaurant for their interests. As a result, individuals will have difficulty locating the best location and food to eat, especially if they are new to the area [9].

Additionally, the location is critical in efficiently determining the customer's preference efficiently. In [10], almost all social networks employ location information to provide better user services. As a result, a location-aware recommender system is required.

Daily, about 3 billion images are exchanged on social media due to the widespread use of camera-equipped smartphones and the growing popularity of numerous photo-sharing platforms [1]. Consumers worldwide were anticipated to snap 1.3 trillion photographs in 2017. Consumers enjoy photographing of their dining experiences and sharing them online [11]. Moreover, restaurants provide a wealth of information about consumer ratings, restaurant attributes, and the competitive landscape. As a result, it provides with photographs containing attributes that can be extracted as features and other metadata that help predict the best restaurant in a given location.

Photographs convey a wealth of information about the restaurant (e.g., food items served, ambiance, etc.). Hypothesize that these images will assist viewers visualizing how much they could love the food items and/or their eating experience. Thus, when creating suggestions based on metadata, results can be enhanced in more accurate recommendations. The researchers use machine learning techniques to extract various attributes from 755,759 Yelp images [1].

The recommendation system for restaurants significantly impacts social media [8]. The purpose of this present thesis is to provide consumers with new upgraded recommendations. The suggested recommendation system is based on extracting new features from photographs in the Yelp dataset and their associated metadata and location tags. The suggested enhancements are based on an analysis and prediction of ingredients of food meal dishes, after a prediction of ingredients used to extract latent information to create a new rating matrix based on user preferences and restaurant features.

From the related works, the researchers found low accuracy [2] of the recommended system for the list of user preferences (preferred restaurants). So, this model is suggested to raise the RS accuracy and create a mechanism for implementing a newly merged rating matrix using multiple data threads.

The suggested method also tries to decrease the sparsity of the original rating data by enhancing the system's implicit feedback. It considers users' behavior and online shared tagged photos. Finally, Collaborative Filtering (CF) is the core of the suggested model to recommend more efficient choices a user may prefer.

1.2 Problem Statements

After reviewing related works and conducting some analysis studies on the Yelp dataset. The Proposed system focused on some significant problems that are addressed in the suggested model:

- Extraction of nutrition information from food photos, is an open problem in the recommender systems and has not been discussed clearly in previous research [18][19].
- The prediction of such food types needs to be verified from meal dishes photos, and there are problems in extracting their latent features.
- According to previous research, predicting user preferences from current row data obtained low accuracy [20]. However, rating data suffers from sparsity.
- A rating matrix that depends on an implicit feedback model suffers from degradation; such a model must depend on multiple data categories.

1.3 Related Works

Most restaurant recommendation systems focus on extracting features from metadata on social media and predicting food images. This is to help predict a good restaurant using filter algorithms. Studies explicitly solved image food prediction problems and extracted Nutrition Information (latent features).

Researchers have recently started to work on cross-prediction food photos and restaurant recommender-based preferences of users (metadata).

Existing research can be divided into three groups which are Food Photos Prediction, Nutrition Information Extraction Techniques, and Recommender Systems.

1.3.1 Food Photos Prediction

Deepak et al.(2022) proposed a model that takes an input food image and classifies it using Convolutional Neural Networks (CNN) layers according to the training list of food categories. Then, based on the selected category, it displays the estimated nutritional value of the food item. In this scenario, they recognized the food image using CNN, a Deep Learning approach, bypassing it through layers such as Dense, Dropout, Flatten, Conv2D, and Maxpooling2D. Additionally, they designed a system that uses an ever-growing and dynamic collection of culinary images. they know that food and its classes are immense and constantly growing, resulting in the disastrous loss of idealistic notions in present systems [12].

Luo et al.(2020) suggested convolutional attention based on the value of visual information and the efficacy of deep learning to examine the importance of various categories of images in restaurant suggestions. The model is fed multi-view restaurant photos (containing food, drink, inside, and outside) and a user restaurant rating matrix. A pre-trained deep convolutional neural network extracts the images' high-dimensional expressiveness [14].

Zhang et al.(2020) presented a model for restaurant's prediction and rating that incorporates multi-view visual information and implicit feedback data. The photos' visual features (visual information) are retrieved and merged into a collaborative filtering framework using a deep convolutional network. To improve personalization, the multi-view visual elements are combined using user-related weights. Weights assigned to users reflect each user's unique aesthetic taste for restaurants, and the weights are unique and autonomously amongst individuals [15].

Hashmi et al.(2019) proposed a model for semantically embedding image-specific information into the star rating and recommendation processes using this rich image data of food. A transfer learning strategy was applied to employ pre-trained CNNs (Convolutional Neural Networks) to label Yelp food photographs using the Food101 data set. Restaurant star ratings were developed by establishing a correlation between restaurant photographs and images submitted by users. An approach was suggested to identify hidden characteristics of food photos and descriptions employed in the recommendation strategy [2].

Sun et al.(2019) offered a model using the XGBoost algorithm and machine learning techniques to an image classification model for classifying food photos and extracting new features [17].

Sundermann et al.(2018) proposed a model for evaluating images to generate features specifying whether a restaurant is a Survival or not, utilizing the XGBoost technique to implement decision trees optimized for speed and performance analyzing photos uses the Clarifai API [10].

Peng et al.(2017) suggested a methodology that utilized machine learning techniques, specifically image classification algorithms to determine the context and rich information associated with various Yelp photos. Trained and tested the dataset using three distinct models [16].

Some of these studies have not considered predicting of meal food photos in Yelp dataset and extracting latent food features. Most researchers focused on predicting general photos, not meal food, and extracting general latent features. Additionally, some researchers focused on food photos, not meals but raw food. Finally, researchers predict meal food photos but with low accuracy. Therefore, predicting meal food and features extraction from photo analysis with reasonable accuracy is a work not reviewed in other related works. This technique will handle in the proposed research.

1.3.2 Nutrition Information Extraction

Kumar et al.(2021) proposed Support Vector Machine (SVM) and enhanced Multilayer Perceptron (MLP) models are used in research to perform food item recognition and calorie prediction. The suggested study utilizes a variety of preprocessing approaches, segmentation, and feature extraction to analyze a single food item. For recognition, the collected features are input into SVM and MLP classifiers [18].

Liang et al.(2018) offered deep learning neural networks that were used to create an approach for automatically predicting meal calories from ingredient pictures. The method begins by training an object recognition model to recognize all the food elements in the image, using the white dish as a reference object. It then uses polynomial linear regression to fit the relationship between the weight of the food ingredient and the area of the food ingredient in the image. Finally, it estimates the calories from the ingredients' calorific values [19].

Such papers do not estimate nutritional information from meal food photos. Some researchers focused on estimating nutritional information from general categories of food photos. Additionally, some researchers only focused on estimating nutritional facts from meal titles. Therefore, extracting the ingredients of food from meal food photos and estimating nutritional information is a work not reviewed in previous related works. This technique will be handled in the proposed system.

1.3.3 Recommender Systems

Varatharajan et al.(2022) suggested an approach that assists users in selecting the optimal food to consume based on their food preferences. The application is intended for everyone who wishes to eat at a restaurant. Collaborative filtering is used to power the restaurant suggestion system, and this model considers the user's meal preferences and ratings when making food recommendations. This collaborative filtering technique combines item- and user-based filtering [20].

V. J. C and J. S. Raj (2021) proposed a model to study the effect of location on business success through a matrix factorization model along with other location-related characteristics (postal code, latitude, longitude, etc.) [21].

Nag et al.(2017) offered a model quantifying the food ingredients on the menu and providing recommendations based on the user's food preferences [22].

Such studies make recommendations based on user metadata preferences, not food photo features. Additionally, some researchers focused on extracting latent features from food menus (text mining) to support recommendations. Therefore, extracting new features (nutritional information) from meal food photos is a work not reviewed in previous related works. This technique will support the recommendation in the proposed system. The researchers have obtained combined features from metadata or features extraction from photo analysis to support the recommender system, which is a good blend that will handle in the proposed research.

1.4 Thesis Objectives

In this thesis, the main aim is to build a restaurant recommender system with acceptable accuracy based on user preferences. Four objectives are developed through the implementation of the proposed system:

- Suggesting a model to be used for nutrition information extraction from Food Photos using the Nutrition5k dataset.
- Generating a new (photo-data) implicit feedback—model to produce newly latent features.
- Producing a newly hybrid rating matrix that can reduce sparsity.
- Using the proposed recommender system to predict best restaurants based on nutrition facts.

1.5 Thesis Organization

This thesis is structured as follows:

Chapter Two (Theoretical Background): This Chapter discusses the main concepts used in this thesis such as food ingredients estimation, nutrition information extraction, and recommender system list the main theoretical approaches of food ingredients estimation and nutrition information extraction. The proposed system used recommendation techniques like Collaborative Filtering (CF) and some similar measurement methods used in it.

Chapter Three (The Proposed System): This Chapter identifies and describes the practical aspects of the suggested algorithms and techniques used to develop the proposed system.

Chapter Four (Results and Discussions): This Chapter exhibits the system's results via tables and graphs. It presents the preliminary versions of the study and the most significant findings in this thesis.

Chapter Five (Conclusions and Future Works): This Chapter provides a comprehensive assessment of this thesis's fundamental concerns and applicable system and future work that can be built upon the suggested system or to create new techniques to improve the outcomes and the work itself.

CHAPTER TWO

Theoretical Background

Chapter Two

Theoretical Background

2.1 Overview

This chapter provides a theoretical background that forms the basis of the proposed thesis. It reviews the concepts of food photos, nutrition information extraction methods, and types of nutrition information that can be utilized. Profiling options for user interactions and data collection are also reviewed. Then, the Light Graph Convolutional Neural network (LightGCN) algorithm is examined in detail. After that, the similarity metrics utilized by the proposed system are introduced. Finally, the research datasets are discussed.

2.2 Food Photos

Sharing food-related images on social media has become fashionable, and people search for interesting culinary dishes and establishments. As a result, in-depth studies for numerous applications were linked to food recognition, eating behaviors, and dietary evaluation that have focused on detecting, classifying, and evaluating food items [49].

Food-related images have become increasingly popular due to social media, food recommendations, and dietary assessment systems. For example, sharing dining-out experiences on social media is a new trend. People are becoming increasingly interested in learning more about different parts of their food and finding and sharing new cuisines. In recent years, many efforts on food recognition based on various visual representations have been published [50]–[52]. Most of these studies are limited to a few food classes in controlled conditions [50].

Food recognition using only visual information remains a difficult task. Food things, unlike objects, are malleable and have a high degree of variability. For example, different cooking procedures and spices will result in distinct appearances of the same food. Furthermore, several meals share many ingredients, making variations within food classes challenging to discern. Another possible challenge is that the appearance and presentation of the same dish in different places add to the difficulty of recognizing it.

2.2.1 Geotagged Photos

Digital photographs with spatial information are usually referred to as geotagged photos. Photographs can be geotagged manually or automatically. A geotag saves the latitude and longitude coordinates of each Joint Photographic Experts Group (JPEG) file into the Exchangeable Image File (EXIF) data space [23].

When geotagged images are published to online sharing communities such as Facebook or Twitter, the image can be viewed on a map to determine where it was shot. In this manner, users can view images from a map, search for images from a specific area, and discover images taken by other users of the exact location. Numerous cell phones geotag their photographs, and photographers who do not wish to share their location can disable this function.

Geotagging is utilized to determine social tendencies [30]. For instance, Instagram has a considerable number of geotagged images and videos that are joined with hashtags in captions utilized by owners. This is to describe the images' content and, in some cases, their feelings and moments relevant to those photos. This can significantly improve place information such as exciting items and activities at that place. For instance, Golden Gate Park may be a good option if a user seeks a place to rest. They obtain this vital information by analyzing various hashtags based on users' experiences at that site [32]. Figure 2.1 shows a food geotagged photo example in Fawaness Restaurant in the Hilla location. This photo was posted on the owner's Instagram.

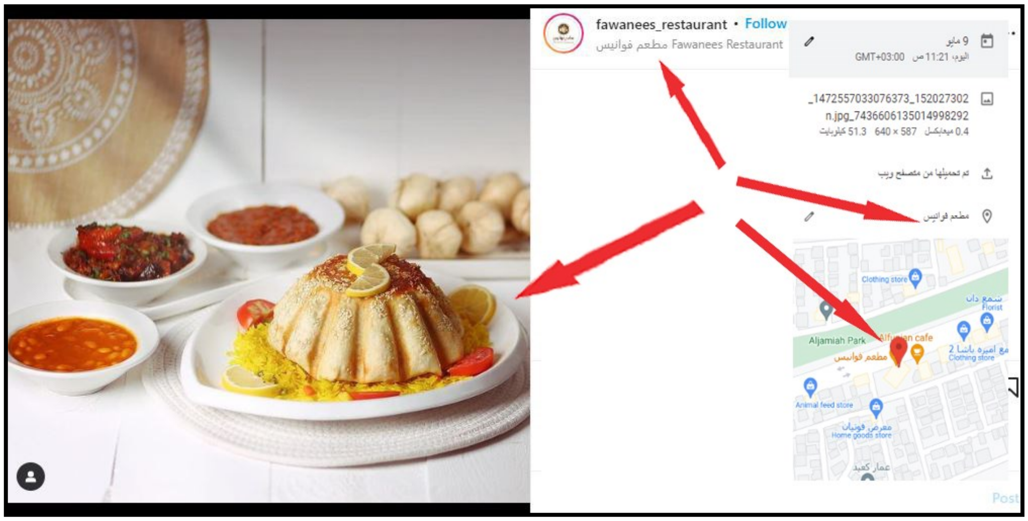

Figure 2.1: Food Geotagged Photo Example

2.2.2 Techniques for Extracting Features

Food research has fast expanded techniques for extracting features from visual contents in images since computer vision, and image processing technology have proliferated in the Artificial Intelligence field. The convolutional neural network (CNN), one of the artificial neural networks, is widely used in the analysis of the visual content of photographs because it performs well in image recognition and classification [16].

The Meta-researchers collaborate with Universitat Politecnica de Catalunya researchers to build a Facebook Inverse Cooking Algorithm that predicts a full recipe from an image better than humans. This program can expect ingredients, cooking instructions, and a recipe title based on an image alone [31].

Figure 2.2 illustrates an example of the work of this algorithm.

Tiltle: Biscuits.

Ingredients: Flour, butter, sugar, egg, milk, salt.

Instructions:

- Preheat oven to 450 degrees.
- Cream butter and sugar.
- Add egg and milk.
- Sift flour and salt together.
- Add to creamed mixture.
- Roll out on floured board to ¼ inch thickness.
- Cut with biscuit cutter.
- Place on ungreased cookie sheet.
- Bake for 10 minutes.

Figure 2.2: An Example of a Generated Recipe by the Inverse Cooking Algorithm [31]

Previously, the algorithm relies on simple recipe retrieval systems based on visual similarities in an embedding space. Currently, these methods heavily relies on the learned embedding's quality, dataset size, and variability. Thus, these methods fail when the input image and the static dataset do not match [31].

Instead of extracting a recipe straight from a picture, the inverse cooking method provides a pipeline with an intermediate stage where the list of ingredients is first obtained. This allows the instructions to be generated not just with the image but also with the components. Figure 2.3 illustrates Inverse Cooking Recipe Generation Model [31].

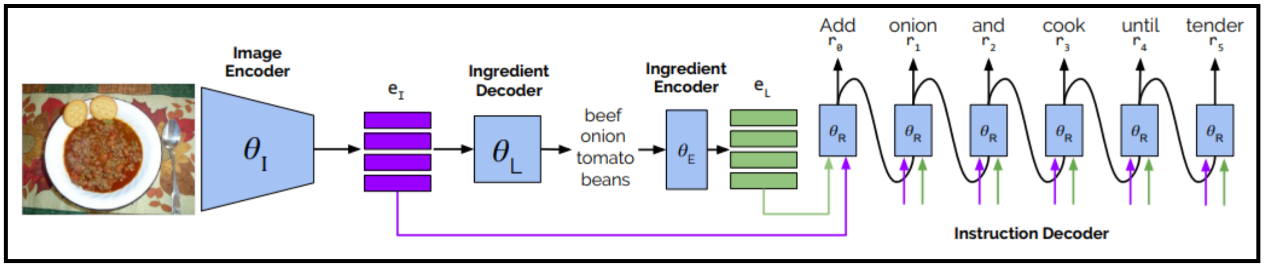

Figure 2.3: Inverse Cooking Recipe Generation Model [31]

In [31], they assess the quality of expected elements using user trials to evaluate the model's effectiveness was assessed. The study compared their model's performance to that of humans in ingredient production by randomly selecting sixteen photos from the test set and asking participants to choose up to 22 different ingredients that correspond to the presented image. To lessen the work's complexity for humans, they lowered the number of words for ingredients from 1489 to 220 by growing the frequency threshold from ten to one thousand. The paper received responses from 31 unique individuals, averaging 5.6 responses per image. The study retrain top ingredient predict algorithm by restricting several words of components to ensure a fair comparison. The paper calculates Intersection over Union (IoU) and F1 ingredient evaluations gathered from humans, the recapture baseline, and their methodology. Humans outperform the retrieval reference standard 35.25% F1 vs. 30.60%. F1 of 35.25 percent vs. 30.60 percent, respectively were obtained. Additionally, technique beats baseline human performance and retrieval-based systems, with an F1 score of 49.09 percent [31].

In [31], the proposed model uses only the ingredient list generation part from this model to generate a list of ingredients for each food photo and extract new features from these ingredients, as shown in Algorithm 2.1.

Algorithm 2.1 : Inverse Cooking Model [31]:

Inputs : Photos : JPG file // Food Photos from Yelp dataset

Outputs : IngredientsList : list[] // The Result list of ingredients

Process :

Begin

FOREACH photo in Photos DO

Begin

PhotosFeatuers ← Encode (photo) // Uses ResNet50

IngedientsList ← IngedientsPredict(PhotosFeatuers) // Uses Transformal Model

End

Return IngedientsList

End

A list of ingredients varies in size, a systematic gathering of distinct meal ingredients. Training data be composed of $\boldsymbol{M}$ photos and ingredient list pairs; the target $\left\{\left(\boldsymbol{x}^{(i)}, \boldsymbol{L}^{(i)}\right)\right\}_{i=0}^{M}$ is to predict $\hat{\boldsymbol{L}}$ predict list from a photo $\boldsymbol{x}$ by maximizing the next objective [31], as shown in Equation 2.1 [31]:

$$\underset{\theta_I,\theta_L}{\arg\max} \sum_{i=0}^{M} \log p\left(\hat{L}^{(i)} = L^{(i)} \mid x^{(i)}; \theta_I, \theta_L\right) \quad 2.1$$

θ_I: learned features extracted from food photos using Resnet50

θ_L: learned features extracted from ingredients list using Transformer model

L: list of ingredients in 1milion recipe dataset

M: food photos

x: food photo

$\hat{L}$: predicted list

A set of ingredients is a variable-sized, disorganized grouping of distinct meal ingredients. The model can gain a set of ingredients $\boldsymbol{S}$ set of ingredients by choosing $\boldsymbol{K}$ ingredients from the dictionary; $\boldsymbol{\mathcal{D}}: \boldsymbol{S} = \{\boldsymbol{s}_i\}_{i=0}^{K}$ training data be composed of $\boldsymbol{M}$ photos and ingredient set pairs: $\left\{\left(\boldsymbol{x}^{(i)}, \boldsymbol{s}^{(i)}\right)\right\}_{i=0}^{M}$ In this case, the target is to predict $\hat{\mathbf{s}}$ set of ingredients prefrom a photo $\mathbf{x}$ by maximizing the next objective [31], as shown in Equation 2.2 [31]:

$$\underset{\boldsymbol{\theta_I},\boldsymbol{\theta_L}}{\boldsymbol{arg\ max}} \sum_{i=0}^{M} \boldsymbol{log}\, \boldsymbol{p}\left(\hat{\boldsymbol{s}}^{(i)} = \boldsymbol{s}^{(i)} \mid \boldsymbol{x}^{(i)}; \boldsymbol{\theta_I}, \boldsymbol{\theta_L}\right) \qquad 2.2$$

θ_I: learned features extracted from food photos using Resnet50

θ_L: learned features extracted from ingredients list using Transformer model

s: list of ingredients in 1milion recipe dataset

M: food photos

x: food photo

$\hat{\boldsymbol{s}}$: predicted set

The researchers employed a target distribution technique $\boldsymbol{p}\left(\boldsymbol{s}^{(i)} \mid \boldsymbol{x}^{(i)}\right) = \boldsymbol{s}^{(i)}/\sum_j \boldsymbol{s}_j^{(i)}$ to model the joint distribution of set elements and train a model by minimizing the cross-entropy loss between $\boldsymbol{p}\left(\boldsymbol{s}^{(i)} \mid \boldsymbol{x}^{(i)}\right)$ and the model's output distribution $\boldsymbol{p}\left(\hat{\boldsymbol{s}}^{(i)} \mid \boldsymbol{x}^{(i)}\right)$, Despite that, it is unclear how to transform the objective distribution back to the set of items with mutable cardinality. In this case, [31] built a feed-forward network and trained it with the objective distribution cross-entropy loss. To recover the ingredient set, in [31], the researchers proposed sample element's probabilities $\boldsymbol{p}\left(\hat{\boldsymbol{s}}^{(i)} \mid \boldsymbol{x}^{(i)}\right)$ greedily and stop the sampling once the sum of probabilities $\boldsymbol{p}\left(\hat{\mathbf{s}}^{(i)} \mid \mathbf{x}^{(i)}\right)$ of chosen elements is above a threshold. Therefore, referred to this model as feed-forward (objective distribution) [31].

One of the method's key accomplishments was to outperform a baseline recipe retrieval system and the average human in predicting the ingredients from an image. Table 2.1 illustrates evaluation metrics for Inverse Cooking Model [31].

Table 2.1: Left: Intersection over Union(IoU) and F1 scores for ingredients gained with retrieval way, Facebook's methodology. Right: Recipe success rate based on human judgment [31].

	IoU	F1		Success %
Human	21.35	35.21	Real	80.32
Retrieved	18.02	30.54	Retrieved	48.82
Inverse Cooking Model	32.51	49.07	Inverse Cooking Model	55.46

The inverse Cooking algorithm was used to develop and publish food recommendation system applications. The user is given many ideas based on the expected components in the online application, such as possible ingredient combinations [31].

The first stage in the proposed methodology is pre-trained convolutional neural network model using the learning-based CNN model (Recipe Retrieval Algorithm) for ingredient prediction of food meals. This section of the model is split into two major phases. The initial stage image encoder extracts features from food photos using Resnet50 encoder and obtains the ingredient embedding. In the second phase the decoder architecture is used to predict a list of ingredients using the transformer model [31]. As shown in Figure 2.4.

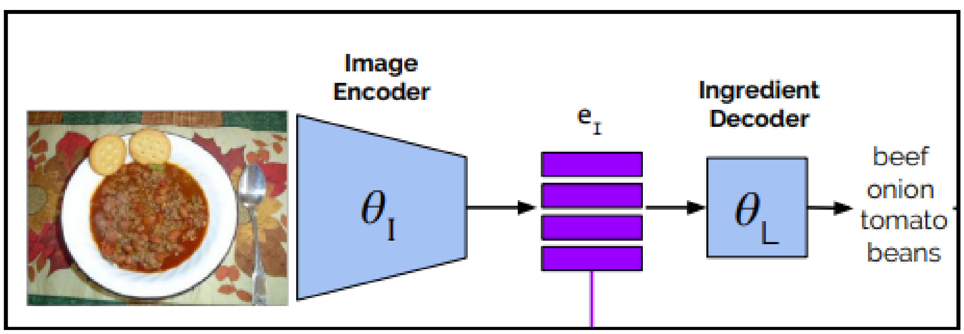

Figure 2.4: Predicted Ingredients Model [31]

A) Image Encoder

Based on the convolutional neural networks (CNN) model and introducing a deep-residual-learning framework, ResNet can successfully address the degeneration problem as network depth increases. In this thesis. ResNet50 Algorithm 2.2 was designed as the features extraction network of food meals stacked upon each other to form the primary building block of ResNet architecture as shown in Figure 2.5 [31]. Table 2.2 explain ResNet50 algorithm. The residual module has two options. It can either implement a series of operations on the input or overshoot all of them. These stacked residual modules fit a whole network.

Table 2.2: Resnet50 Details [28].

Stage	Description
	Zero-padding pads the input with a pad of (3, 3).
Stage 1	• The 2D convolution has 64 filters of shape (7, 7) • BatchNorm is applied to the channel's axis of the input • Max-pooling uses a (3, 3) window and a (2, 2) stride.
Stage 2	• The convolutional block uses three set of filters of size [64, 64, 256]. • The 2 identity blocks use three set of filters of size [64, 64, 256].
Stage 3	• The convolutional block uses three set of filters of size [128, 128, 512]. • The 3 identity blocks use three set of filters of size [128, 128, 512].
Stage 4	• The convolutional block uses three set of filters of size [256, 256, 1024]. • The 5 identity blocks use three set of filters of size [256, 256, 1024].
Stage 5	• The convolutional block uses three set of filters of size [512, 512, 2048]. • The 2 identity blocks use three set of filters of size [512, 512, 2048].
	The 2D average pooling uses a window of shape (2, 2)
	The flatten layer (learned featuers)

Algorithm 2.2 : Resnet50 Algorithm [28]:
Inputs : photos: JPG file // Food Photos

Outputs : learned_featuers : Vecter of Integer // Learned Featuers

Process :

Begin

FOREACH photo in Photos DO

Begin

ZeroPad ← ZeroPadding2D((3,3)) // add zero to input photo

X ← Conv2D (64, (7, 7), strides=(2, 2), name='conv1', kernel_initializer=glorot_uniform (seed=0))(ZeroPad) // apply convolution process
X ← BatchNormalization (axis=3, name='bn_conv1')(X) // make normalization
X ← Activation ('relu') (X)
X ← MaxPooling2D ((3, 3), strides=(2, 2))(X) // select max value
X ← convolutional_block (X, f=3, flters=[64, 64, 256], stage=2, block='a', s=1)
X ← identity_block (X, 3, [64, 64, 256], stage=2, block='b')
X ← identity_block (X, 3, [64, 64, 256], stage=2, block='c')
X ← convolutional_block (X, f = 3, flters = [128, 128, 512], stage = 3, block = 'a', s = 2)
X ← identity_block (X, 3, [128, 128, 512], stage=3,block='b')
X ← identity_block (X, 3, [128, 128, 512], stage=3,block='c')
X ← identity_block (X, 3, [128, 128, 512], stage=3,block='d')
X ← convolutional_block (X, f=3, flters = [256, 256, 1024], stage = 4, block='a', s = 2)
X ← identity_block (X, 3, [256, 256, 1024], stage=4, block='b')
X ← identity_block (X, 3, [256, 256, 1024], stage=4, block='c')
X ← identity_block (X, 3, [256, 256, 1024], stage=4, block='d')
X ← identity_block (X, 3, [256, 256, 1024], stage=4, block='e')
X ← identity_block (X, 3, [256, 256, 1024], stage=4, block='f')
X ← convolutional_block (X, f = 3, flters = [512, 512, 2048], stage = 5, block='a', s = 2)
X ← identity_block (X, 3, [512, 512, 2048], stage=5, block='b')
X ← identity_block (X, 3, [512, 512, 2048], stage=5, block='c')
X ← AveragePooling2D ((2,2), name="avg_pool")(X) // select average value
learned_featuers ← Flatten ()(X)

End

RETURN learned_featuers

End

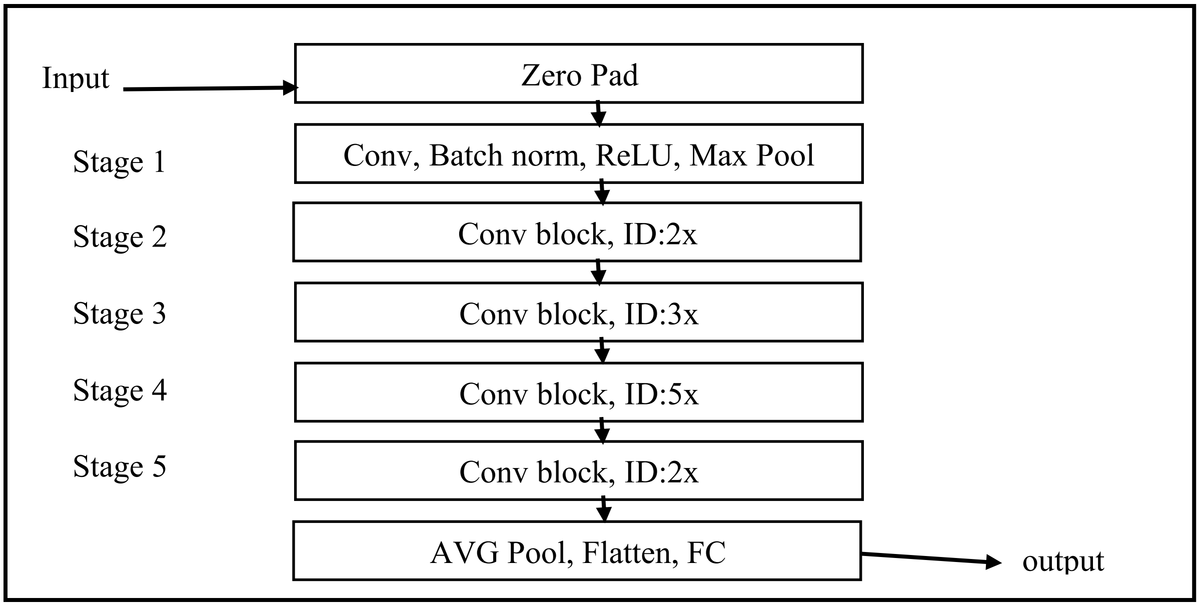

Figure 2.5: Block Diagram of the Resnet50 Model [28].

For Training, 85 percent of photos were used as training data and 15 percent of images as test data. The activation for all layers except the final layer was ReLU function. Adam was chosen for the optimization task with a learning rate (lr) of 0.0001. This network was trained through 100 training epochs, and data were transmitted to the network in batches of 16 sizes (batch-size). The length of each epoch was 28 seconds [31].

B) Ingredient Decoder

In this phase, the transformer model was used to generate a list of ingredients. The transformer uses encoder-decoder architecture. The encoder extracts features from the input (a sentence), and the decoder uses the features to produce an output list of ingredients [31].

In this case, there is no need to extract features because they are extracted using the Resnet50 network. Therefore, the transformer's decoder is needed as displayed in Figure 2.6.

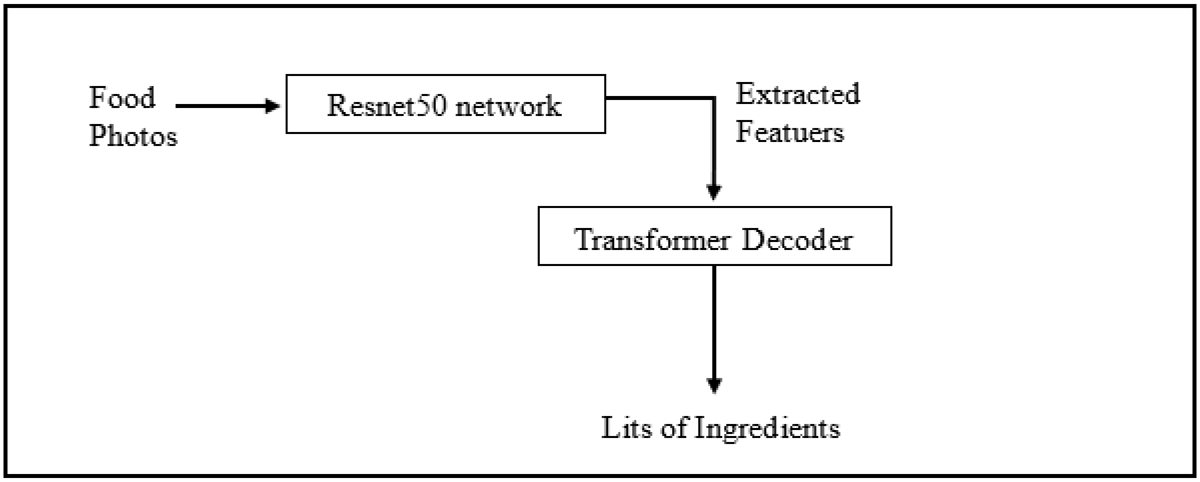

Figure 2.6: The Block Diagram of the Ingredients Decoder

Each decoder is comprised of three key elements: 1) a self-attention technique. 2) an attention technique over the encodings, and 3) a feed-forward neural network. The self-attention technique accepts input encodings from the preceding encoder and weights their relative importance to generate output encodings. The feed-forward neural network proceeds each output encoding individually and adds an attention technique that extracts relevant data from the features generated by the Resnet50. This technique is also known as the encoder-decoder attention because the decoder acts similarly to the encoder, except for the attention mechanisms [70].

Instead of encodings, the first decoder accepts positional information and embeds the output series as input. Therefore, the output sequence must be partially disguised to avoid this information flow in the opposite direction. This allows for the production of autoregressive ingredients. For each attention head, subsequent tokens cannot receive attention. The final decoder is followed by a final linear transformation and softmax layer as shown in Algorithm 2.3 to generate the output probability ingredients [70], (see Figure 2.7).

Algorithm 2.3 : Transformer Decoder Algorithm [70] :

Inputs : ingredient_recipe: list of ingredient DS // 1 milion recipe dataset

learned_featuers: vecter of featuers // featuers extracted from ResNet50

Outputs : ingredient_list: list of ingredient // list of ingredient extracted from model

Process :

Begin

X ← Embadding (ingredient_recipe) // embedding to convert list into vector
X ← SelfAttention(X) // weighted position of ingredient in list
X ← Attention(X, learned_featuers) // use learned featuers to select from ingredients
X ← FeedForward(X)
X ← Linear(X)
ingredient_list ← Dense (activation='sofmax')(X) // select best ingredient represent food meal

End

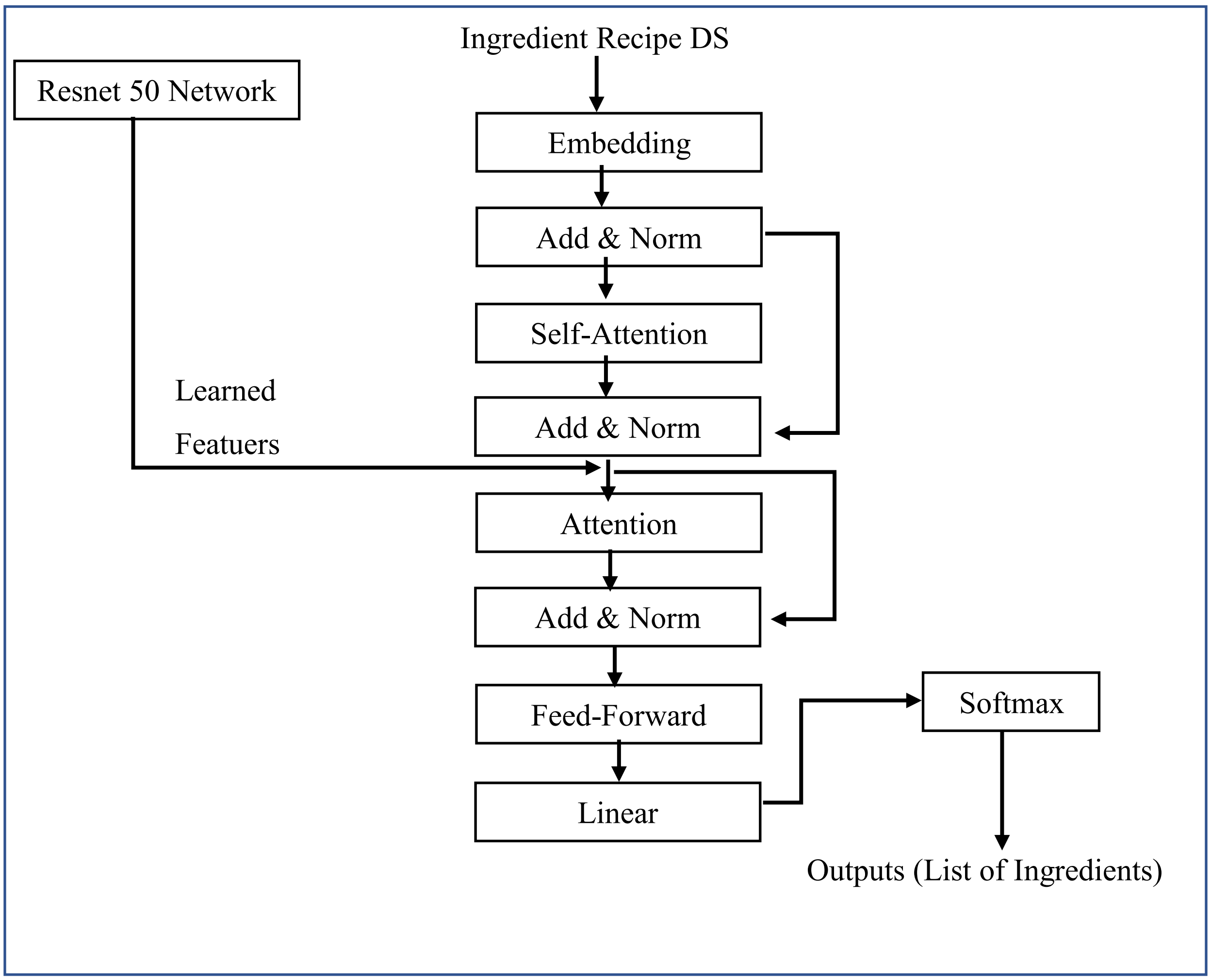

Figure 2.7: The Block Diagram of the Transformer Decoder [70]

Nutrition information refers to related nutritional information on food, food products, or dietary supplements that are intended for one or more healthy population groups; the calorie value and composition of protein, fat, and carbs, as well as vitamins and minerals, are based on scientific data, reports, and research [32]. Food and nutrition information significantly affect humans' quality of life, health, and happiness [33]. The number of overweight or obese persons is increasing. According to the WHO, In 2016, more than 1.9 billion adults (39%) were overweight, and more than 650 million (13%) were obese [35]. Obesity is a significant cause of diseases. For these reasons, food-related research [36][37] has consistently been a hot topic and garnered significant attention from a variety of sectors.

Nutrients are chemical compounds in food that the body needs to generate energy, structure, and regulate chemical processes. Nutrients are classified into six categories which are: Carbohydrates, Lipids, Proteins, Water, Vitamins, Minerals [38].

Nutrients are classified into macronutrients and micronutrients, organic and inorganic, and whether or not they supply energy to the body (energy-yielding) [38]. Figure 2.8 describes significant categories of nutrient information.

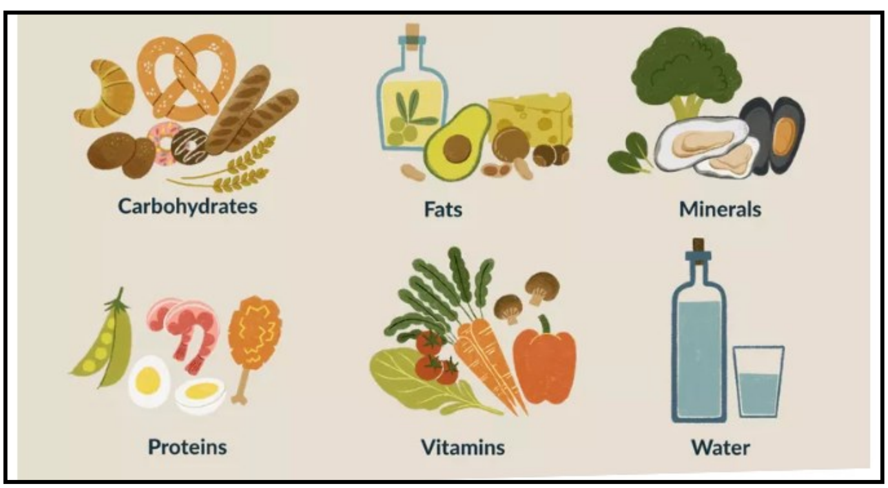

Figure 2.8: Major Categories of Nutrients [38]

2.3 User Profiling

Personalization of information has propelled recommender systems to new heights. Personalization allows these systems to deliver precise and effective user-specific recommendations. Personalization is facilitated by user profiling, which retrieves the information to personalize a situation while preserving a unique user profile for each user [39].

In addition, personalization has gained substantial importance in computer science, particularly in the applications of Recommendation Systems. A recommender system must accommodate several users with varied preference criteria. Recommender systems must fulfill the demands of all users, either by recommending user-specific goods or adapting to their needs. Consequently, the user profile aids the recommender system in comprehending the user's needs and acting appropriately [39].

Therefore, user profiling is identifying information about a user's benefit field. The system can use this information to learn more about the user, improving the retrieval process to the user's delight. two crucial features of user profiling are efficiently knowing the user and offering items of interest based on this knowledge [39].

2.3.1 Explicit User Profiling

Static profiling is the investigation of fixed and predictable user traits. This method predicts user behavior by examining available facts about the user. Typically, this data is gathered through online forms, questionnaires, etc. This technique is often referred to as static profiling. When researchers rely only on explicit profiling, they run into problems because users are reluctant to reveal their information out of privacy concerns or because the form-filling process is boring, and the user attempts to avoid it. Therefore, the accuracy of this form of profile reduces with time [40].

2.3.2 Implicit User Profiling

Instead of relying on the present information about the user, implicit profiling depends more on what will discover in the future, i.e., systems attempt to learn more about the user. Consequently, this type of system is often referred to as Behavioral Profiling, Adaptive Profiling, and, more commonly, the Ontological Profiling of the user. Various filtering approaches are also applied in such profiling. A wealth of academic literature describes different filtering strategies, including Rule-based filtering, Collaborative filtering, and Content-based filtering [41].

2.3.3 Hybrid User Profiling

This strategy combines the advantages of both implicit and explicit user profiling. It takes a user's static and behavioral features into account, i.e., retrieving behavioral information about a user. This strategy improves the efficiency of profiling and maintains the correctness of temporal data by updating information in a temporal manner [42].

2.4 Profile Techniques

By analyzing multiple experiments, the researchers can determine that a user profiling system can be broken down into subtasks such as profile extraction and profile integration. Lastly, benefit detection. Now, the numerous approaches for each of these responsibilities are discussed.

2.4.1 Profile Extraction

Profile extraction differs from extracting meaningful information about a user from multiple sources, and researchers have utilized numerous techniques and models. These strategies include web data extraction, social media data extraction, and user behavior-based techniques that assist user profiling systems in collecting valuable information about users [43].

Some of the primary operations may be brief as follows:

- Content tracing: Movable the mouse pointer over a paragraph while reading [43].
- Link pointing: This entails positioning the mouse cursor over a link without clicking it [45].
- Link clicking: Utilizing a link to navigate to a new page [45].
- Content selection: choosing text by dragging with the mouse [46].
- Scrolling: The movement of a window at a defined rate [45].
- Registration of bookmarks: marking a page as a bookmark [39].
- Saving: Saving an HTML page [43].
- Printing: Printing a document [39].
- Window movement: Repositioning a browser window [42].
- Window resizing: Changing the window size in a web browser [42].

2.4.2 Profile Integration

After extracting essential information, there may occasionally be issues with data cleaning. There is a possibility that some of the acquired data are duplicates or appear to be duplicated but are unique. It is necessary to identify singular and duplicate data to overcome this issue, which will assist the profiling system with the following subtask.

2.4.3 Interest Discovery

After collecting information about the users, it is required to split them into distinct classes to provide the system with feedback. This can be accomplished by classifying users based on their behavior. This approach is known as filtering, and considerable research has been conducted [47].

2.5 Recommender Systems

A recommender system or recommendation system is a subclass of information filtering systems that predict the "rating" or "preferred" a user would assign to an item [48].

In recent years, the explosion of data has intensified due to the rapid development of sensor technology, storage technology, computing technology, and network technology. As the volume of data grows, the problem of too much information occurs. This makes it harder for a user to choose the best decision. The term for this phenomenon is information overload. The use of artificial intelligence to extract abstract information from vast amounts of data and translate it into practical knowledge is one of the core challenges of large-scale data analysis. To counteract the problem of information overload, a recommender system emerges as the need arises. Its primary goal is to evaluate users' historical behavior and preference data, construct a model, automatically offer valuable items or services to the user, and then generate a tailored list. In addition to recommending items with comparable preferences based on the user's preferences, the recommender system can also recommend obscure objects of interest to the user. These challenges can be mitigated by recommender systems rapidly recognizing users' future demands and selecting desirable products from a massive quantity of candidate data [49].

Recommender systems can be classified into three classes:

- Content based Recommender Systems
- Collaborative Filtering based Recommender Systems
- Hybrid Recommender Systems

2.5.1 Content based Recommender Systems

Content-based recommender systems identify commonalities between things based on their contents. After examining many things a user has previously favored, a profile of the user's preferences is produced. Based on this profile, the recommender system might search the database and select the right things [50].

These techniques are complicated due to their inability to discern user preferences based on the contents of objects. Numerous solutions to this problem have been developed via data mining and machine learning. For example, to recommend articles to a particular reader, a recommender system must first obtain and analyze the contents of every book the reader has previously read. Text mining techniques, such as the well-known Term Frequency-Inverse Document Frequency (TF-IDF) analysis, can extract keywords from text (TF-IDF). After incorporating a book's keywords with their respective weights, a multidimensional vector can represent it. The reader's interests can be represented by the centers of these vectors, which can be found using specialized clustering techniques [50].

2.5.2 Collaborative Filtering(CF) Recommender Systems

Collaborative filtering (CF) has been proven as one of the most efficient algorithmic techniques for generating recommendations. CF, in contrast to content-based approaches, relies only on each user's item ratings. It is founded on the concept that individuals who have ranked comparable items similarly are likely to share similar preferences. Collaborative filtering suggests items based on the preferences of other users with similar tastes or identifies items equivalent to those previously rated by the target user. It employs statistical techniques to establish the similarity between the user and item vectors. Memory-based and model-based CF approaches are two separate categories [49].

2.5.3 Hybrid Recommender Systems

There are three types of hybrid recommendation systems: monolithic hybrid, parallel hybrid, and hybrid pipeline. The monolithic hybrid recommendation is a mixed technique incorporating numerous recommendation systems into a single algorithm [51]. Combining at least two different recommendation systems is required for the remaining two-hybrid suggestions. Second, the parallel hybrid recommendation is autonomous, independently generating a list of requests for each input and then integrating the output data to build the final recommendation set. The pipeline hybrid recommendation integrates multiple recommender systems in a pipelined architecture, with the output of one recommender system acting as the input to the next. The subsequent recommendation unit may utilize a portion of the initial input data. Hybrid recommender systems take advantage the power of multiple data sources or improve the effectiveness of existing recommender systems within a particular data modality [52].

2.6 Deep Learning

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign or to distinguish a pedestrian from a lamppost. The key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately, and for good reason. It's achieving results that were not possible before [75].

In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers [75].

2.6.1 Graph Neural Networks (GNN)

Neural Networks have gained massive success in the last decade. However, early variants of Neural Networks could only be implemented using regular or Euclidean data, while a lot of data in the real world have underlying graph structures which are non-Euclidean. The non-regularity of data structures has led to recent advancements in Graph Neural Networks. In the past few years, different variants of Graph Neural Networks have been developed, with Graph Convolutional Networks (GCN) being one of them. GCN is also considered as one of the basic Graph Neural Networks variants [76].

As in the convolution layers of the Convolutional Neural Networks, 'convolution' in GCN has the same operation. It refers to multiplying the input neurons with a set of weights that are commonly known as filters or kernels. The filters act as a sliding window across the whole image and enable Convolution Neural Network (CNN) to learn features from neighboring cells. Within the same layer, the same filter will be used throughout an image, referred to as weight sharing.

For example, using Convolution Neural Network (CNN) to classify images of cats versus non-cats, the same filter will be used in the same layer to detect the nose and the ears of the cat [76].

GCN perform similar operations where the model learns the features by inspecting neighboring nodes. The major difference between CNN and Graph Neural Network (GNN) is that CNN are specially built to operate on regular (Euclidean) structured data, while GNN are the generalized version of CNN where the numbers of nodes connections vary, and the nodes are unordered (irregular on non-Euclidean structured data) [76]. In this thesis, Light Graph Convolutional Network (LightGCN) model is adopted for many reasons. First, the LightGCN model was used in previous work [58]. Second, it removed some operations from the GCN model, like self-connection, feature transformation, and nonlinear activation, to simplify and reduce calculation time without affecting the quality of the model. Finally, It obtained the best accuracy compared with other techniques, Table 4.5 illustrate compared LightGCN with other models.

2.6.2 Light Graph Convolutional Network (LightGCN)

LightGCN is model that incorporates the most fundamental component of GCN, and neighborhood aggregation for collaborative filtering. After assigning each user (item) an I.D. embedding, in paper [53] the researchers propagate the embedding on the user-item interaction graph to refine them. Next, employ a weighted sum to combine the embedding collected at different propagation layers to construct the final embedding for prediction. The entire model is beautiful and essential, making it easier to train and more effective experimentally than Neural Graph Collaborative Filtering (NGCF) and other methods such as Mult-VAE [53].

Graph Convolutional Network (GCN) is based on the principle of learning node representations by smoothing graph features [54]. To do this, graph convolution is conducted iteratively, i.e., the new representation of a target node is produced by merging its neighbors' attributes. This form of neighborhood aggregation may be summed up, as shown in Equation 2.3 [59].

$$\boldsymbol{e_u^{(k+1)}} = \boldsymbol{AGG}\ (\boldsymbol{e_u^k}, \{\boldsymbol{e_i^k} : \boldsymbol{i} \in \ \boldsymbol{N_u}\}) \qquad 2.3$$

e_u: embedding of user

e_i: embedding of item

k: length of a recommended list

AGG: convolution process between users embedding and item embedding

N_u: number of user neighbors

AGG is an aggregation function that examines the representation of the target node and its neighbors in the kth layer. Numerous papers, including the weighted sum aggregator in Graph Isomorphism Networks (GIN), have detailed the AGG [55], the Long Short-Term Memory (LSTM) aggregator in GraphSAGE [56], and the bilinear interaction aggregator in Boundary Graph Neural Networks (BGNN) [57], among others. However, most research associates the AGG function with feature transformation or nonlinear activation. Perform well on node or graph classification tasks that include semantic input data [53] .

The secret to LightGCN, as shown in Figure 2.9, resides in two basic designs: (1) Light Graph Convolution (LGC) and (2) Layer Combination.

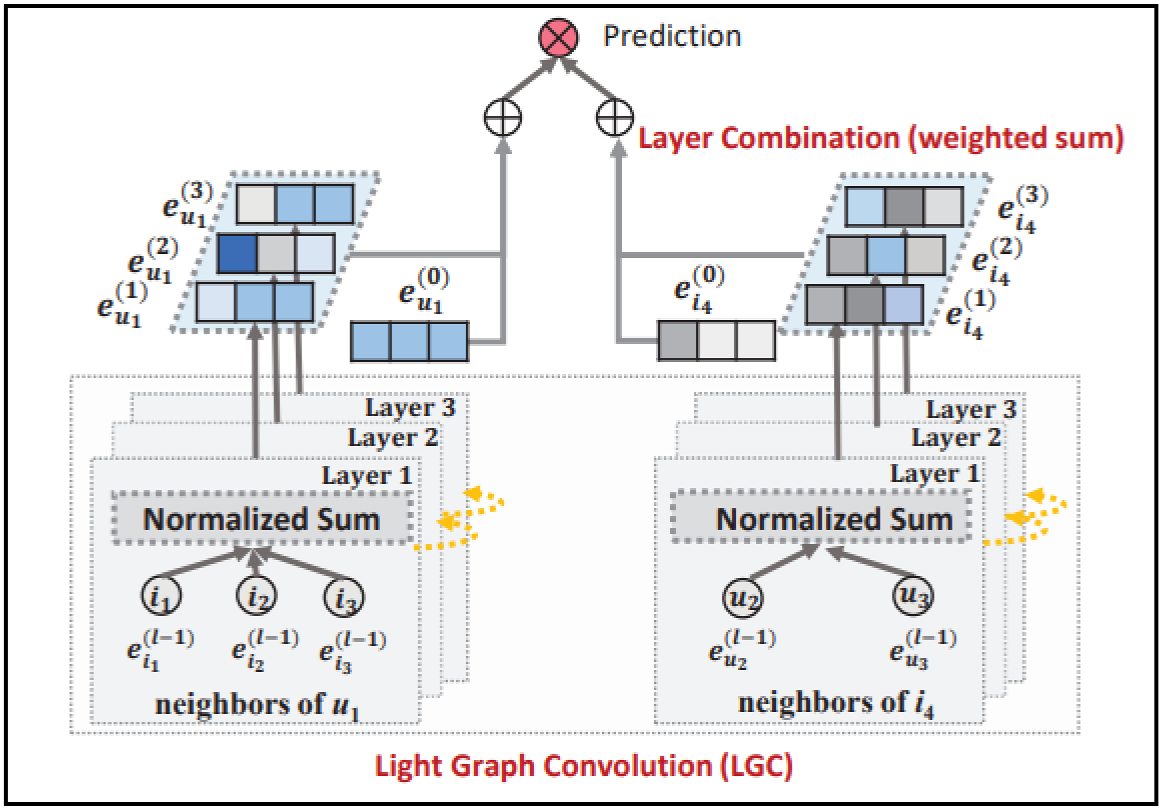

Figure 2.9: LightGCN Model Architecture [58]

A) Layer Neighborhood Aggregation

In this stage, within each layer, for each user in the graph, embedding is the weighted sum of all its neighboring items' embedding.

The exact equation for aggregators for users graph convolution operations, as shown in Equation 2.4 [58]:

$$\boldsymbol{e}_{\boldsymbol{u}}^{(\boldsymbol{k}+1)} = \sum_{i \in \mathcal{N}_u} \frac{1}{\sqrt{|\mathcal{N}_u|}\sqrt{|\mathcal{N}_i|}} \boldsymbol{e}_{\boldsymbol{i}}^{(\boldsymbol{k})} \qquad 2.4$$

e_u: embedding of user

e_i: embedding of item

k: length of a recommended list

N_u: number of user neighbors

N_i: number of item neighbors

The symmetric normalization term $\frac{\mathbf{1}}{\sqrt{|\boldsymbol{\mathcal{N}_u}|}\sqrt{|\boldsymbol{\mathcal{N}_i}|}}$ follows the design of standard GCN, which can avoid the scale of embeddings increasing with graph convolution operations [58].

The exact equation for aggregators for items graph convolution operations, as shown in Equation 2.5 [58]:

$$\boldsymbol{e}_{\boldsymbol{i}}^{(\boldsymbol{k+1})} = \sum_{\boldsymbol{u}\in\boldsymbol{\mathcal{N}_i}} \frac{\mathbf{1}}{\sqrt{|\boldsymbol{\mathcal{N}_i}|}\sqrt{|\boldsymbol{\mathcal{N}_u}|}} \boldsymbol{e}_{\boldsymbol{u}}^{(\boldsymbol{k})} \qquad 2.5$$

e_u: embedding of user

e_i: embedding of item

k: length of a recommended list

N_u: number of user neighbors

N_i: number of item neighbors

The symmetric normalization term $\frac{\mathbf{1}}{\sqrt{|\boldsymbol{\mathcal{N}_i}|}\sqrt{|\boldsymbol{\mathcal{N}_u}|}}$ follows the design of standard GCN, which can avoid the scale of embedding increasing with graph convolution operations [58].

Where $\boldsymbol{e}_{\boldsymbol{u}}^{(\boldsymbol{k})}$ and $\boldsymbol{e}_{\boldsymbol{i}}^{(\boldsymbol{k})}$ are the users and item node embeddings at the k-th layer, $|\boldsymbol{\mathcal{N}_u}|$ and $|\boldsymbol{\mathcal{N}_i}|$ are the user and item nodes' number of neighbors, as shown in Figure 2.9.

Similarly, for each item, the updated embedding is computed using a weighted sum of its neighboring users.

B) Layer Combination

At layer combination, instead of taking the embedding of the last layer, LightGCN computes a weighted sum of the embeddings at distinct layers, as shown in Figure 2.9:

Weighted summation can be calculated, as shown in Equation 2.6 [58]:

$$\boldsymbol{e_u = \sum_{k=0}^{K} \alpha_k e_u^{(k)}; \; e_i = \sum_{k=0}^{K} \alpha_k e_i^{(k)}} \qquad 2.6$$

e_u: embedding of user

e_i: embedding of item

k: length of a recommended list

α_k: learning rate

where $\alpha \geq 0$. Here, alpha values may be learned as network parameters or established as empirical hyperparameters. It has been discovered the Equation 2.7 [58] is effective way to specify better learning rate.

$$\boldsymbol{\alpha = \frac{1}{k+1}} \qquad 2.7$$

Based on the inner product of the final user and item embeddings, LightGCN predicts, as shown in Equation 2.8 [58] :

$$\boldsymbol{\hat{y}_{ui} = e_u^T e_i^T.} \qquad 2.8$$

e_u^T: final embedding of user

e_i^T: final embedding of item

$\hat{y}_{ui}$: final prediction

This inner product (cosine similarity) examines the similarity between the user and the items and this, in turn, allows determining the likelihood that the user will enjoy the item.

2.7 Similarity Measures

This section examines the concept of distance measurement. Similarities play a significant role in recommendation systems that employ neighborhood-based approaches:

- Similarity metrics inside RS choose neighbors with ratings that can be utilized in prediction systems.
- In the prediction computation methods provide extra or low-importance attributes for the neighbors.

The calculation of similarities is one of the most crucial components of developing neighborhood-based recommendation systems. This can significantly affect accuracy and performance. Many methods are proposed to calculate similarities, such as Euclidean distance [77], Minkowski distance [77], and Cosine distance. This thesis used cosine distance because this measure was used by previous work and depends on the angle between two vectors projected in a multi-dimensional space [62].

In data analysis, cosine similarity is the cosine of the angle between two sequences to evaluate the similarity. It is the dot product of the vectors divided by the product of the vectors' lengths. The cosine similarity is, therefore, independent of the magnitudes of the vectors and only reliant on their angles. The interval [-1,1] always contains the cosine similarity. For example, the cosine similarity of two proportional vectors is 1, that of two orthogonal vectors is 0, and that of two opposing vectors is -1. Specifically, cosine similarity is utilized in positive space, where the result is cleanly by [0,1] [63], as shown in Equation 2.9 [63].

$$\boldsymbol{cos}(\boldsymbol{x},\boldsymbol{y}) = \frac{\sum_i^n x_i y_i}{\sqrt{\sum_i^n x_i^2 \sum_i^n y_i^2}} \qquad 2.9$$

x: first vector

y: second vector

n: length of vectors

This form of remote measurement can be used to determine users' likenesses by implying a user $\boldsymbol{u}$ as a vector $\mathbf{x_u} \in \mathbf{R}^{|\mathbf{I}|}$, So $\mathbf{x_u} = \mathbf{r_i}$ Suppose the user u has rated the item $\boldsymbol{i}$ and 0 otherwise” [63].

This distance metrics shortcoming is that it does not consider the differences in the means and variances of the rating values supplied by users x and y [63].

2.8 Evaluation Metrics

Evaluation metrics are employed to quantify the quality of the statistical or machine learning model. Evaluation of machine learning models and algorithms is fundamental for every project, and numerous evaluation metrics are available for testing models. These are precision, recall, Normalized Discounted Cumulative Gain (NDCG), and others [44].

2.8.1 Precision

Precision is the actual correct prediction divided by total prediction made by model. It is used to answer with the number of things from all correct recommendations, as shown in Equation 2.10 [26].

$$\boldsymbol{p} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \qquad 2.10$$

p: precision

TP: true positive

FP: false positive

In recommender system, Equation 2.11 is used [25]:

$$\boldsymbol{P} = \frac{\boldsymbol{\#\ of\ our\ recommendations\ that\ are\ relevant}}{\boldsymbol{\#\ of\ items\ we\ recommended}} \qquad 2.11$$

2.8.2 Recall

Recall is the actual correct prediction divided by total correct made by model. In order to answer the coverage issue, it is necessary to determine how many of the deemed-relevant elements are covered by the recommendations. as shown in Equation 2.12 [26].

$$\boldsymbol{R} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \qquad 2.12$$

R: recall

TP: true positive

FN: false negative

In recommender system, Equation 2.13 is used [25] :

$$\boldsymbol{R} = \frac{\boldsymbol{\#\ of\ our\ recommendations\ that\ are\ relevant}}{\boldsymbol{\#\ of\ all\ the\ possible\ relevant\ items}} \qquad 2.13$$

2.8.3 Normalized Discounted Cumulative Gain (NDCG)

Gain for an item is essentially equivalent to its relevance score, which can be a numeric rating such as Google's search results, which can be evaluated on a scale from one to five, or binary in the case of implicit data, in which we only know if a user has consumed a given item or not.

Cumulative Gain is defined as the aggregate of gains up to recommendation list position k, as shown in Equation 2.14 [24].

$$\boldsymbol{CG(k)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{G_i} \qquad 2.14$$

CG: Cumulative Gain

k: length of list

G: Gain

Ordering is not considered, it is a shortcoming of CG. The CG would not be changed by rearranging the relative order of any two items. When ranking order is crucial, this is troublesome. For instance, users would not want most relevant web pages to appear at the bottom of Google's search results.

By dividing the Gain by rank, we encourage the algorithm to place the most relevant items at the top to achieve the maximum Discounted Cumulative Gain (DCG) score, as shown in Equation 2.15 [24].

$$\boldsymbol{DCG(k) = \sum_{i=1}^{k} \frac{G_i}{log_2(i+1)}} \qquad 2.15$$

DCG: Discounted Cumulative Ggain

k: length of list

G: Gain

A problem remains with the DCG score. The relation between the DCG score and the number of recommendations on the list. Therefore, cannot compare the DCG score of a system that suggests the top five things to that of a system that recommends the top ten, as the latter will have a higher score not due to the quality of its recommendations but due to its length.

This problem is resolved by introducing the ideal DCG (IDCG). IDCG is the DCG score for the ideal ranking, which ranks items from the most relevant to the least relevant up to position k, as presented in Equation 2.16 [24].

$$\boldsymbol{IDCG(k) = \sum_{i=1}^{|I(k)|} \frac{G_i}{log_2(i+1)}} \qquad 2.16$$

$IDCG$: Ideal Discounted Cumulative Ggain

k: length of list

G: Gain

Moreover, Normalized Discounted Cumulative Gain (NDCG) merely normalizes the DCG score by IDCG so that its value is always between zero and one, irrespective of the period, as presented in Equation 2.17 [24].

$$\boldsymbol{NDCG(k) = \frac{DCG(k)}{IDCG(k)}} \qquad 2.17$$

$NDCG$: Normalized Discounted Cumulative Gain

k: length of list

DCG: Discounted Cumulative Ggain

$IDCG$: Ideal Discounted Cumulative Ggain

2.9 Datasets

In this thesis, two datasets are used which are Yelp dataset [64] and Nutrition5K dataset [65].

2.9.1 Yelp Dataset

The primary resource in this model is the Yelp dataset, which will estimate nutrition facts in this application. A subset of Yelp's businesses, reviews, and user data is available for personal, educational, and research use. In the form of JSON files. It contains a directory of 10000 restaurants, with addresses, menu selections, and ranking. Yelp provides around 200000 images from over 1999 establishments, with the dataset available on the Yelp Data Challenge website. This dataset contains interior, exterior, beverage, and food photographs. Yelp has designated the four categories above, but there are no subcategories for specific types of cuisine. Customers or business owners upload the majority of photographs. These photographs may contain good captions which accurately describe the photographs they accompany. Numerous photographs have lack captions or contain inaccurate captions, as shown in Figure 2.10 [64].

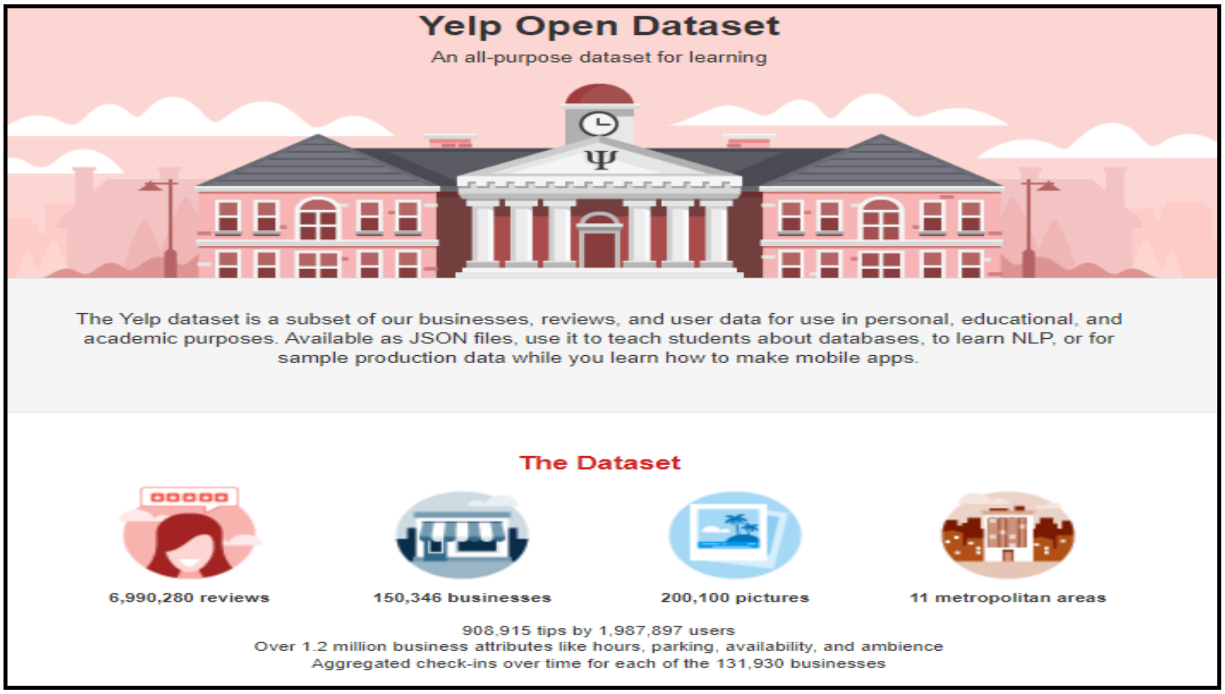

Figure 2.10: Yelp Open Dataset [64]

2.9.2 Nutrition5k

Nutrition5k is a visual and nutritional dataset for ~5k realistic plates of a food captured from Google cafeterias using a custom scanning rig [65].

Key Features

Scans data for 5,006 plates of food, each containing:

- rotating side-angle videos
- Overhead RGB-D images (when available)
- Fine-grained list of ingredients
- Per-ingredient mass
- Total dish mass and calories
- Fat, protein, and carbohydrate macronutrient masses

CHAPTER THREE

The Proposed System

Chapter Three

The Proposed System

3.1 Overview

This chapter explains the practical aspects of the suggested system pertaining to this thesis. In addition, illustrate preprocess phase on Yelp dataset to extract implicit feedback. Also, in the preprocessing phase, explain the technique for creating user profiles from metadata, generating a rating matrix, manipulating dataset photos to extract nutrition information, and generating a new implicit rating matrix. Further, mixed two rating matrices to generate an explicit, implicit rating matrix and inserted the LightGCN model to recommend the best restaurants.

3.2 System Design

This section describes the proposed system phases and algorithms used in this thesis. The first step is the preprocessing of data gathered from the Yelp dataset. It includes generating a new profile for each user from metadata by selecting suggested features related to the recommender system.

Generating implicit feedback is the second task for the proposed system, by processing food photos using Facebook Inverse Cooking Algorithm and the Nutrition5k dataset to extract new features and estimate nutrition information for each Ingredient extracted from the previous algorithm.

The system goes through rating matrix implementation based on user history data information and generates implicit features; as mentioned above, it calculates implicit features from food photos. The rating matrix will be implemented using a Collaborative filtering mechanism for each user's relations with other users' history. Normalization is applied to these values for users' information and saved into the rating matrix as implicit feedback inducted from user histories.

Deep learning is used to extract new implicit feedback characteristics from food photographs to improve the rating matrix. The Facebook Inverse Cooking model is used to implement this technique, as will be detailed in better detail below.

This chapter concludes with a discussion of the nearest neighbor mechanism utilized in this framework to obtain final findings and the best-predicted restaurant that the user will visit next, based on the user's dietary preferences and other metadata to improve the proposed system's outcomes. Figure 3.1 illustrate the primary stages of the proposed system, and each level will be explored in detail in the following sections.

3.3 Data Gathering

The suggested system utilizes a subset of businesses, reviews, user information, and geotagged food photos for businesses (restaurants).

The dataset consists of information about 150.346 businesses (business_id, checkin_count, name). Table 3.1 displays an example of one business record across eleven metropolitan areas in the USA and Canada of 200.000 photos (photo_id, business_id, caption, label). Table 3.2 displays an example of one record of photo, and 6.990.280 reviews (review_id, user_id, business_id, stars). Table 3.3 displays an example of one record of review posted on Yelp from 1.987.897 users (user_id, name). Table 3.4 displays an example of one record of a user [64].

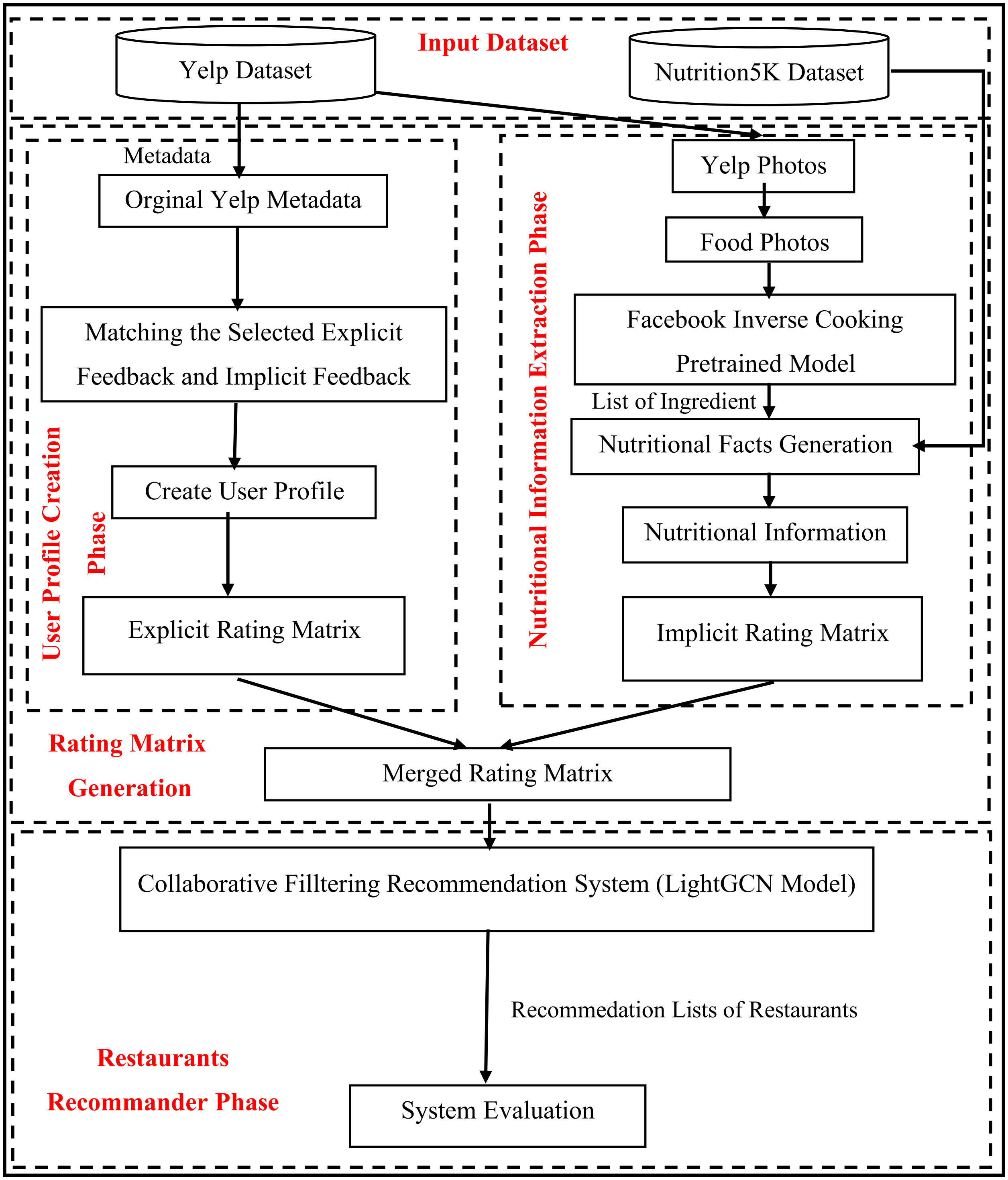

Figure 3.1: Block Diagram for the Main Phases of the Proposed System

Table 3.1: Business Metadata Example

business_id	name	checkin_count
2bncbx08BFs_IO6H-yWBxw	Bubor Cha Cha	325

Table 3.2: Photo Metadata Example

photo_id	business_id	caption	label
_nN_DhLXkfwEkwPNxne9hw	tnhfDv5Il8EaGSXZGiuQGg	carne asada fries	food

Table 3.3: Review Metadata Example

review_id	user_id	business_id	stars
lWC-xP3rd6obsecCYsGZRg	ak0TdVmGKo4pwqdJSTLwWw	buF9druCkbuXLX526sGELQ	4

Table 3.4: User Metadata Example

user_id	name
q_QQ5kBBwlCcbL1s4NVK3g	Jane

The second dataset utilized by the proposed system is a collection of ingredient metadata from a Comprehensive Nutrition Dataset. It consists of a list of all ingredients covered in the dataset's dishes, their unique IDs, and important per-gram nutritional information sourced from the USDA [66] Food and Nutrient Database. Table 3.5 shows an example of one record of strawberry's nutrition information [64].

Table 3.5: Ingredients Information Example

Ingr	cal/g	fat(g)	carb(g)	protein(g)
strawberries	0.33	0.003	0.08	0.007

3.3.1 Files Format of Dataset

Some Data are ignored in this dataset. Used data are described in the following lines, one for each line. Table 3.6 and Table 3.7 show a structure of Yelp dataset.

Table 3.6: JSON File Yelp Dataset

Column Name	Datatype	Description
Column Name	Datatype	Description
Bussiness_id	Str	22 char string
Name	Str	Business name
ReviewId	Str	22 char
User_Id	Str	22 char user-id
Name	Str	user first name
Photo_id	Str	22 char photo id
label	Str	Category of the photo belongs to

Table 3.7: TXT File Nutrition5K

Column Name	Datatype	Description
Ingr	Str	Ingredient name
Cal/g	Float	Amount of calories in one gram
Fat/g	Float	Amount of fat in one gram
Carb/g	Float	Amount of carbohydrate in one gram
Protein/g	Float	Amount of protein in one gram

3.3.2 Data Structure and Data Flow

User's, Restaurant information, and reviews(metadata) are examined by scanning all files for users and restaurants to generate a data frame that represents whole users’ histories, restaurants' histories, and reviews.

Food photos are collected from users' images shared on Yelp social media for business. Foods' photos generate ingredient nutrition information (implicit feedback), as illustrated in Figure 3.2.

The collected metadata from the users' histories and restaurant histories and other metadata will be combined with ingredients nutrition information generated from food photos analysis to be used later in the proposed system to create a new rating matrix that is used as the base data for the system as illustrated in Figure 3.1.

The rating data section offers rating values that are derived and reinforced from an ingredient's nutrition information extracted from food images.

The collaborative filtering data section is the final data in the suggested system. This section of data is represented as relevant features between users' data and restaurants' data after embedding. Thus, each entry in that matrices is one significative value of an entire vector that represents the user’s feature in constructed dimensional representation of data for predicting and recommending relevant information to users based on their history and restaurant data, as illustrated in Figure 3.2.

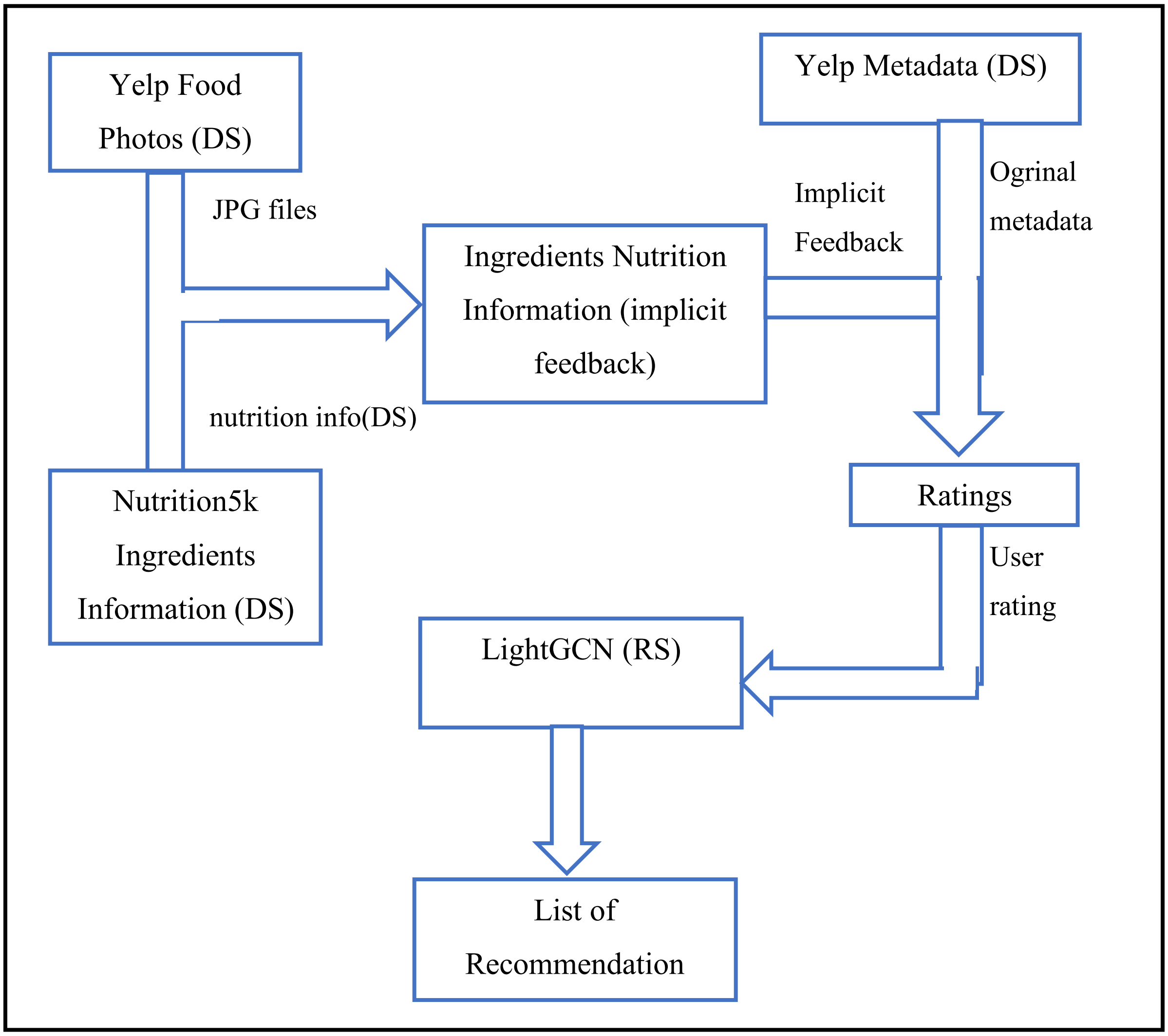

Figure 3.2:The Proposed System Data Flow

3.4 Data Preprocessing

The raw data of the Yelp dataset is vast and has high dimensionality. Therefore, must make preprocessed to prepare for the rating matrix.

The first stage is cleaning data by dropping all unnecessary data for all data frames after taking one million records as samples from the data, and then combining users' data with review data based on ("user_id") and then getting the check-in count from the restaurant visited to date in a check-in data frame, and then combines photos data and nutrition information extracted from food photos analysis based on ("photo_id"), and then combines the results data with a business data frame based on ("business_id").

The second stage makes a filter to minimize the number of records data resulting from the first stage based on the city name (for instance: "Orlando"). to limit data resulting (businesses) to restaurants, only make another filter in ("categories").

The third stage in preprocessing is obtaining the records average of businesses that have multiple photos taken from the same user. The final result matrix represents a new profile for each user review for each restaurant.

The proposed model suggests processing food images on Yelp dataset to extract the contents of each food type and nutrition information for each image food meal as shown in Figure 3.1. This is used as healthy implicit feedback to take advantage of the rapid accumulation of rich photos on social media. The proposed methodology is a machine-learning framework for predicting food ingredients. Also compute critical health metrics for each ingredient and combine them to obtain nutrition data for the food.

The proposed model aims to prepare a machine-learning framework for predicting food ingredients. Additionally, essential health metrics were computed for each ingredient and combined them to obtain the food's nutritional information [63]. The findings demonstrated a promising method for collecting food components and nutrition information from them, allowing developers and architects to leverage this model when designing food and health systems and systems of recommendation.

Figure 3.3 illustrates an example of the proposed model. The model has the main module of the CNN-based pre-trained model. After completing the prediction procedure using a pre-trained model, the estimated nutrition information model gives the item's name and ingredients. Then, nutritional data is derived from the Nutrition5k Dataset.

Inverse Cooking Model

Title	Ingredients
Bbq Chicken Salad	Onion, Chicken, Pepper, Lettuce, Barbecue_sauce

Nutrition Information Extraction

ingr	cal/g	fat(g)	carb(g)	protein(g)
Onion	0.4	0.001	0.09	0.011
Chicken	1.65	0.036	0	0.31
Pepper	0.4	0.002	0.093	0.02
Lettuce	0.15	0.002	0.029	0.014
Barbecue_sauce	1.72	0.006	0.41	0.008
Total	**4.32**	**0.047**	**0.622**	**0.363**

Figure 3.3: The Proposed Model Example

Figure 3.4 illustrates that before predicting the dataset's photos, it must be preprocess stage to the dataset to extract food photos only from the Yelp dataset. The next stage is predicting food ingredients.

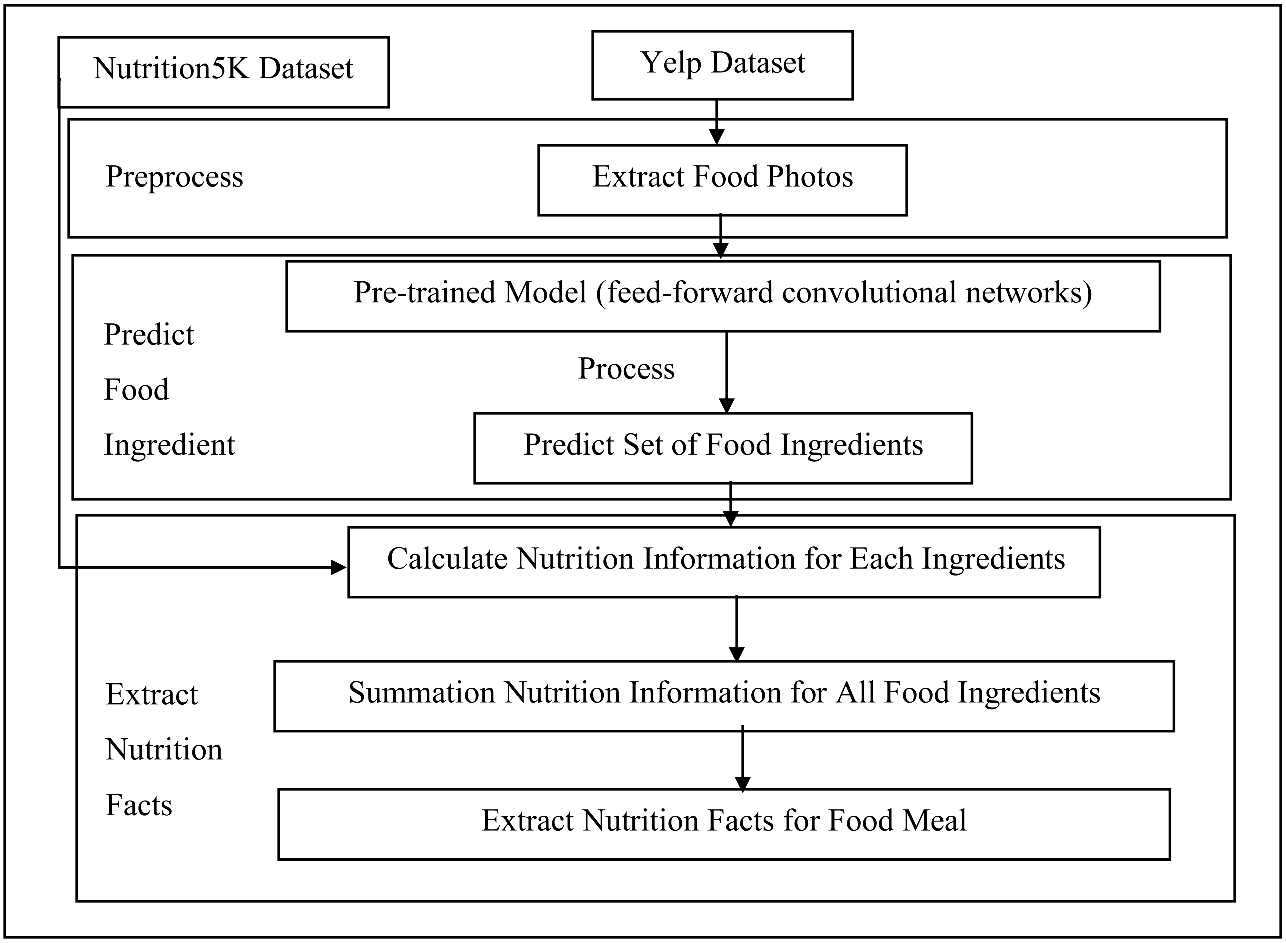

Figure 3.4: The Block Diagram of the Proposed Model

In the final stage, the nutritional information was extracted. After obtaining the name of the food and its components, it was sent to extract the nutritional components. After taking the predicted components, the nutritional information for each component was calculated from the Nutrition5k dataset, as shown in Algorithm 3.1. All the ingredients were collected together to extract the essential nutritional information (calories, carbohydrates, fat, protein), according to the United States Department of Agriculture Food and Nutrition Information Center [66] for one gram of the food. The mass and size of the food were not considered due to the difficulty of predicting the size of the food and its components from the photo.

Algorithm 3.1 : Nutrition Information Extraction :

Inputs : photo , photo id , photo category , // Food Photos , Photo ID , Photo Category
ingredient name , calories , fat , // Nutrition5k Dataset
carbohydrate , protein

Outputs : result cal , result fat , result carb , // Result of Nutrition Information Extracted
result prot

Process :

```
Begin
    FOREACH photo category is "food" DO
            ResultIngedientsList ← PredictFoodIngedients(photo)
    FOREACH ResultIngedient in ResultIngedientsList DO
    Begin
            IF ResultIngedient within IngedientsList(DS) THEN
            Begin
                    Summation result Cal with Calories
                    Summation result fat with Fat
                    Summation result carb with Carbohydrate
                    Summation result prot with Protein
            End
            ELSE
            Begin
                    Summation result Cal with zero
                    Summation result fat with zero
                    Summation result carb with zero
                    Summation result prot with zero
            End
    End
End
```

3.5 The Methodology

The methodology is divided into two distinct ways. The first is the use of learning-based CNN models for ingredient prediction, whereas the second generates a list of predicted ingredients and nutrition facts extraction.

3.5.1 Pre-Trained Recipe Retrieval Model

The first stage in the proposed model is using Recipe Retrieval Model to predict a list of ingredients from food meals. This model is split into two main phases:

In the initial phase, the Resnet50 model extracts learned features from food photos this phase is called the image encoder as shown in Algorithm 2.2 (Table 2.2 explain the Resnet50 model). The second phase, use learned features extracted in the previous phase to generate a list of ingredients using a transformer model as shown in Figure 2.7 this phase is called the ingredient decoder. Figure 2.4 illustrates the Recipe Retrieval Model.

3.5.2 Creating the Proposed Rating Matrix

To accomplish the primary objective of the suggested system, a rating matrix that serves as an input to the following processing stage is required. The proposed rating matrix is a combined rating matrix from nutrition information extracted from food photos and a rating matrix from user review profiles to be a new rating matrix. The outcome matrix must be normalized (see Algorithm 3.2) because the nutrition information has a large and scattered rating; thus algorithm of normalization technique is applied for each user rating on the values of the matrix to be turned into a rating matrix with given values of the domain [1....5]. The hybrid rating matrix represent as an input to the recommender system as shown in Figure 3.1.

Algorithm 3.2 : Ratings Normalization [74]:

Inputs : Ratings, MinValue, MaxValue,

Outputs : Norm_Ratings

Process :

Begin

```
OldRange ← (MaxValue – MinValue)
NewRange ← (5 - 1)                    // because orginal rate in scope (1..5)
FOREACH value IN Ratings DO
newValue ← (((value – MinValue) * NewRange) / OldRange) + 1
add newValue to Norm_Ratings
```

End

The algorithm of the implicit feedback technique (see Algorithm 3.3) is proposed to create and construct rating matrices. The system relies on implicit feedback to construct a rating matrix collected in the user review profile by explicit feedback. The suggested system has a unique technique that generates ratings from scratch without needing user intervention. This concept will increase systems' abilities to be more independent from users' natural choices and decisions.

Algorithm 3.3 : Proposed Rating Matrices Generate :

Inputs : UserProfile, NutritionInformation // implicit and explicit rating matrix

Outputs : FinalRatings // hybrid rating matrix

Process :

Begin

```
FOREACH photo_id IN NutritionInformation
    IF photo_id in UserProfile DO   // matching
        add NutritionInformation and UserProfile to FinalRatings
```

End

3.5.3 Recommendation

The primary function of the proposed system is to suggest restaurants to the user. The user-restaurant relationships should be represented as a bipartite graph in the proposed system that applies the LightGCN [58] model to the new rating matrix. Users and restaurants can be imagined as two nodes on opposite sides of a graph with edges reflecting user-restaurant interactions as shown in Figure 3.5.

In this case, it is aimed to predict whether a user likes a restaurant, so one way to summarize the data is using a bipartite graph [61]. As displayed in Figure 3.6.

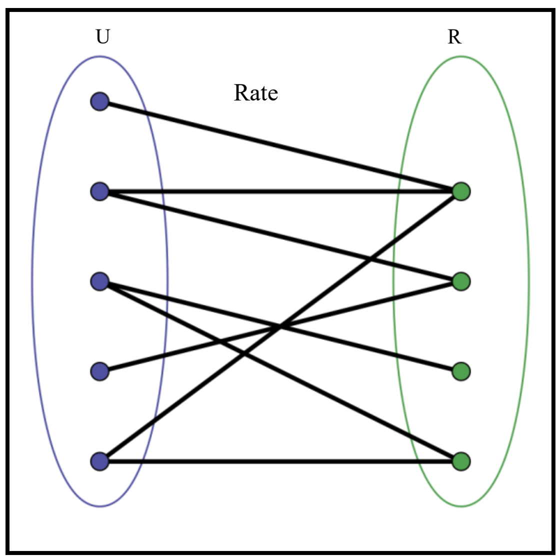

Figure 3.5: Bipartite Graph [61]

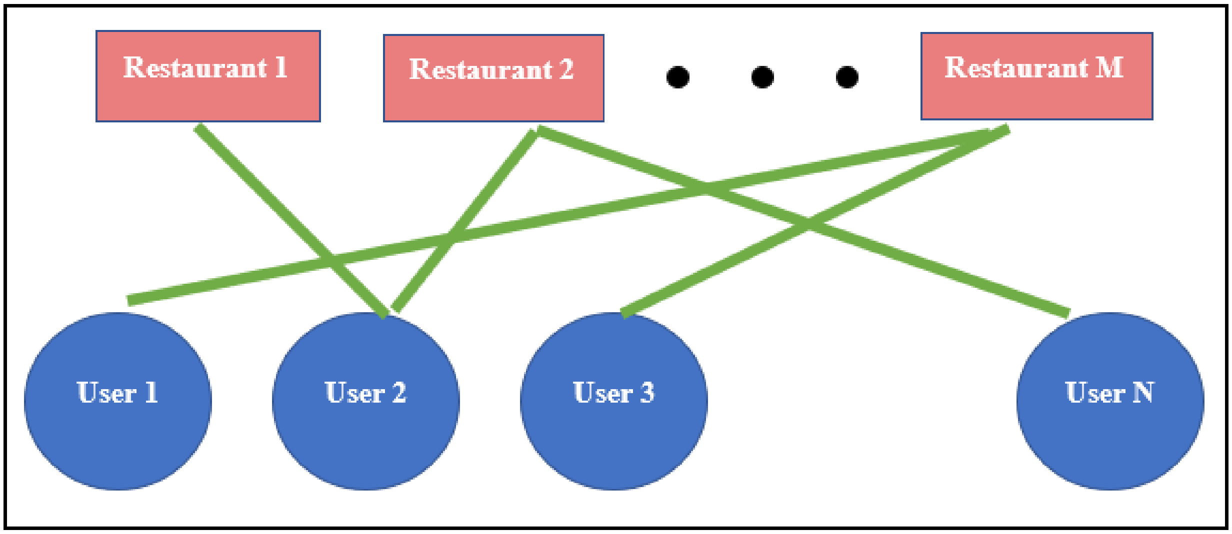

Figure 3.6: Bipartite Graph (User – Restaurant)

The secret to LightGCN, as shown in Figure 3.7, resides in two basic designs: (1) intra-layer neighborhood aggregation and (2) inter-layer combination.

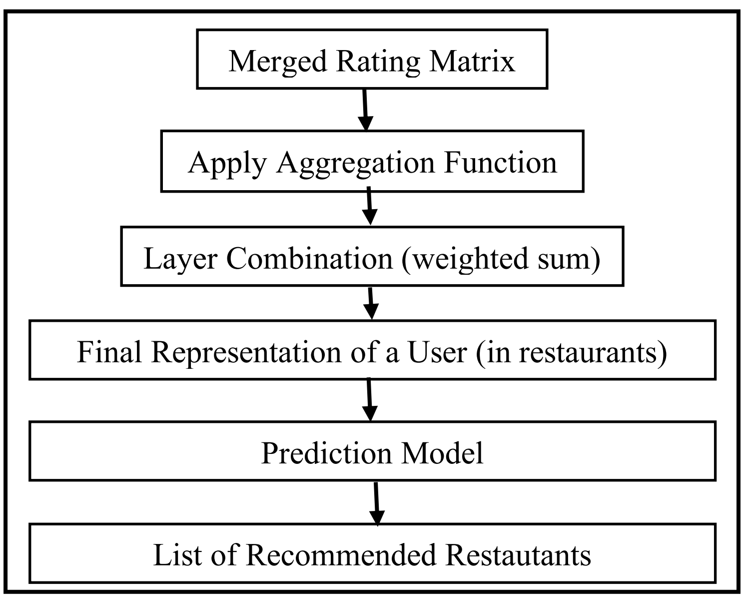

Figure 3.7: The Block Diagram of the LightGCN Model

A) Layer Neighborhood Aggregation

For each layer, each user in the graph is embedded as the weighted sum of all its neighboring items' embeddings (restaurants), as presented in Figure 3.8.

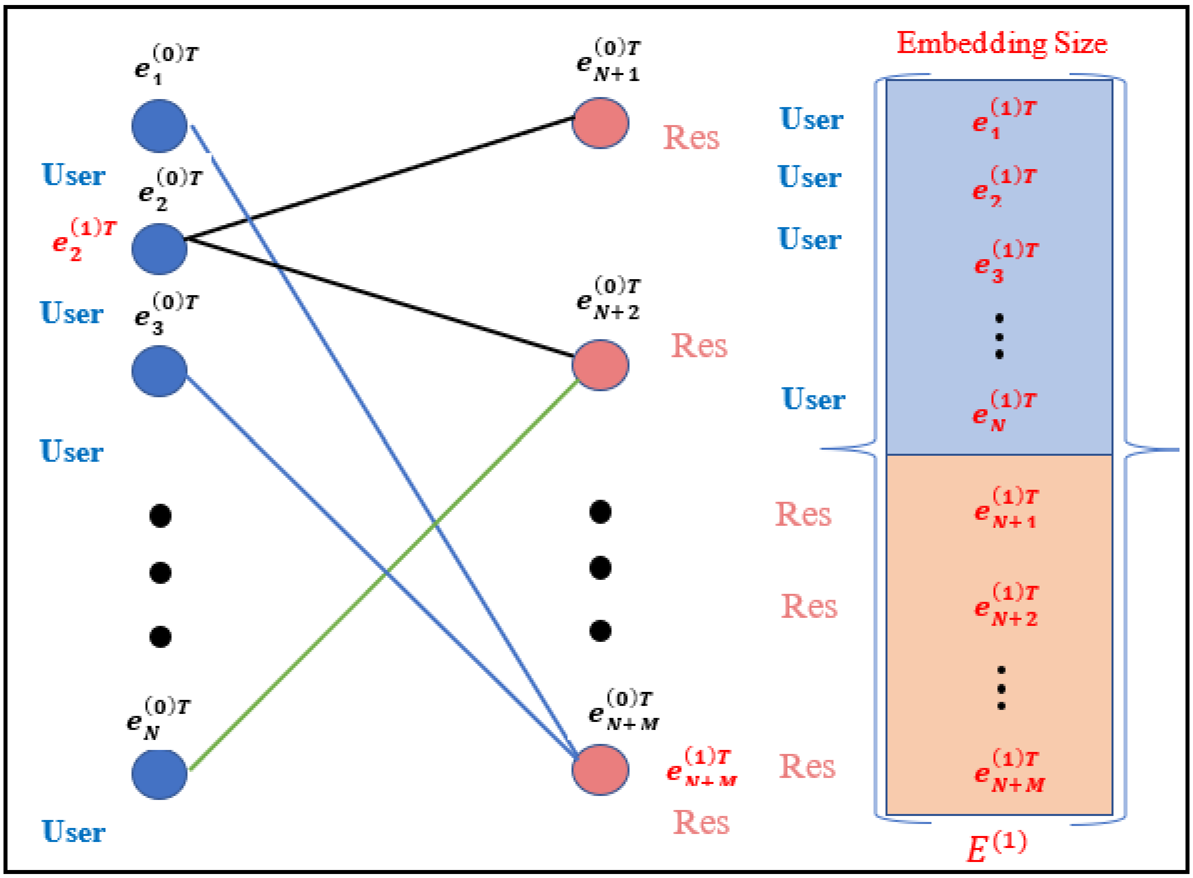

Figure 3.8: Weighted Sum of Embeddings

where $e_u^{(k)}$ and $e_i^{(k)}$ are the user and item (restaurant) node embeddings at the k-th layer, $|N_u|$ and $|N_i|$ are the user and item nodes' number of neighbors.

Similarly, the updated embedding is computed using a weighted sum of its neighboring users for each item.

After K iterations over all the nodes, derive the K-th layer embeddings is derived, $E^{(K)}$ as shown in Figure 3.9 and Algorithm 3.4 .

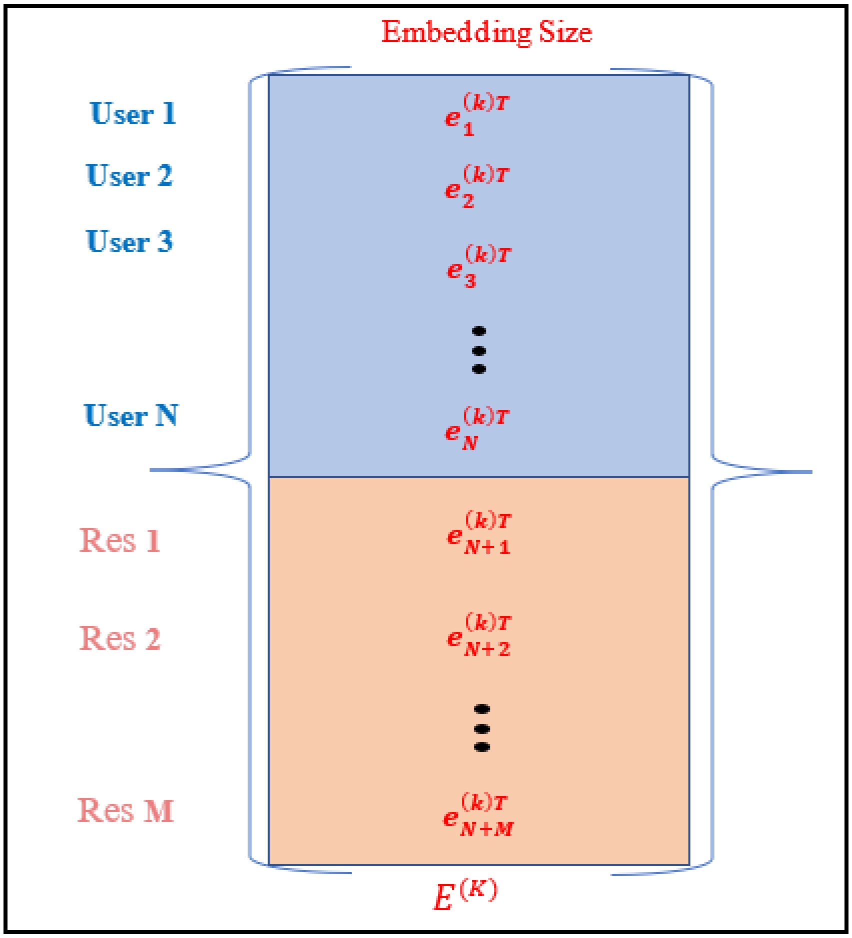

Figure 3.9: K-th Layer Embeddings

Algorithm 3.4 : Layer Neighborhood Convolution:

Inputs : Hybrid Rating Matrix

Outputs : K-th layer embeddings

Process :

Begin

FOREACH node IN bipartite_graph

E ← ComputeEmbedding(node) // compute embedding for each node using Eq 2.4

RETURN $E^{(K)}$ // derive the K-th layer embeddings

End

B) Layer Combination

At layer combination, instead of taking the embedding of the last layer, LightGCN computes a weighted sum of the embeddings at distinct layers, as depicted in Figure 3.10 :

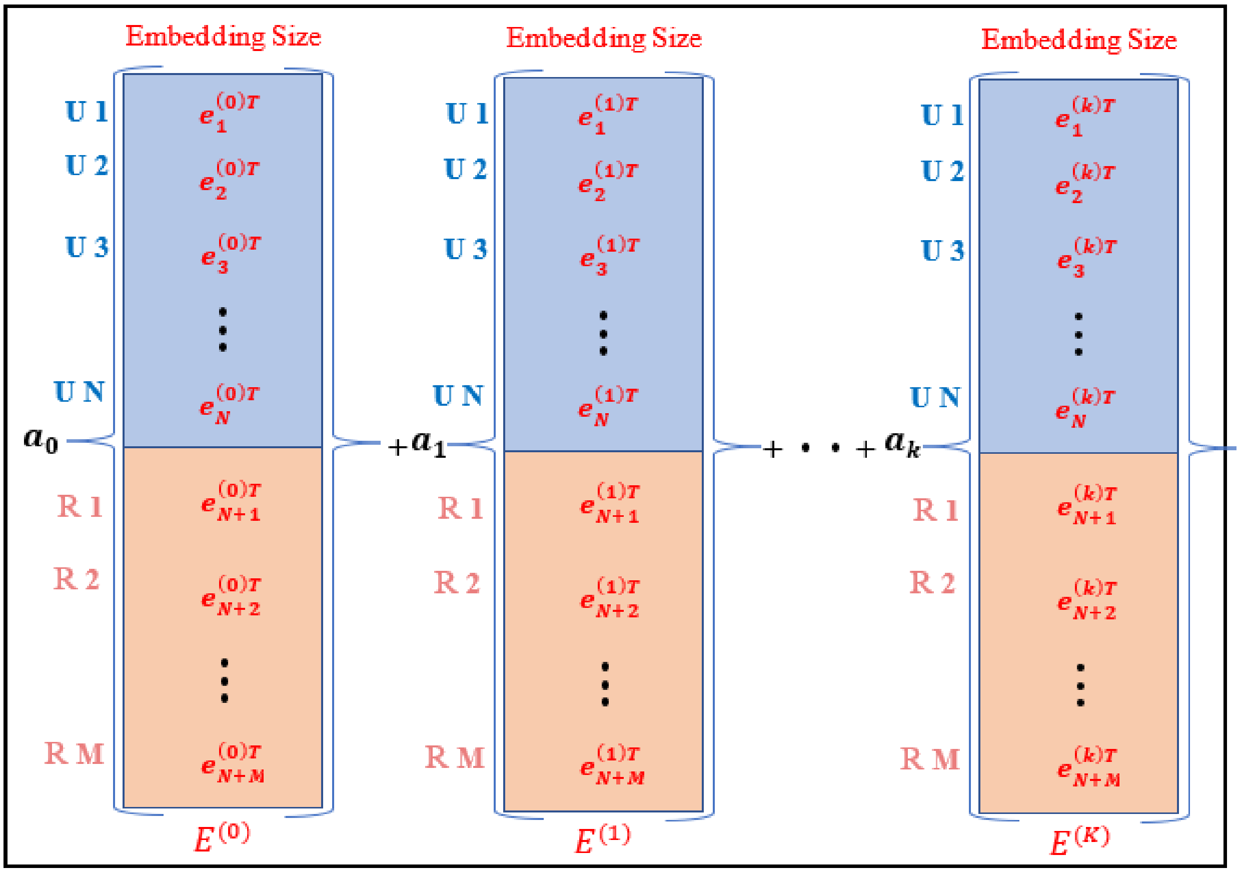

Figure 3.10: Layer Combination

where α ≥ 0. Here, alpha values may be learned as network parameters or established as empirical hyperparameters.

This is based on the inner product of the final user and item (restaurant) embeddings.

This inner product examines the similarity between the user and the restaurant. This allows determining, the likelihood that the user will enjoy the restaurant.

Train the LightGCN model; an objective function congruent with restaurant recommendation objective is required. The Bayesian Personalized Ranking (BPR) loss was used, which encourages observed user-item predictions to have higher values than unobserved user-item predictions. This is illustrated in Algorithm 3.5.

Algorithm 3.5 : Calculate Total Weights:

Inputs : K-th layer embeddings , a_i // alpha values using Eq 2.7

Outputs : result_list // List of recommended restaurants

Process :

Begin

FOREACH weight_emb IN layers_embeddings

Begin

$e_i \leftarrow$ Sum(weight_emb$_i$,a_i) // weighted summation using Eq. 2.6

$e_u \leftarrow$ Sum(weight_emb$_u$,a_u) // weighted summation using Eq. 2.6

End

y $\leftarrow$ InnerProduct(e_i , e_u) // similarity between the user and restaurantembeddings using Eq.2.8

result_list $\leftarrow$ Train_model(y) // using Bayesian Personalized Ranking (BPR) as objective function

End

CHAPTER FOUR

Results and Discussions

Chapter Four

Results and Discussions

4.1 Overview

This Chapter presents the outcomes of the proposed system stages mentioned in Chapter three. It begins with user profiling results, nutrition information estimation results, and the recommended system results, with a few examples and comparisons with the original recommender system and some others.

The rating matrices are the primary data space introduced here, and the nutrition information estimation model inducts implicit feedback generation to produce a new rating matrix. Finally, it shows the recommended lists of restaurants that the suggested system recommends based on the history of users' preferences.

4.2 Nutrition Information Extraction

In the first stage, the suggested system extracted implicit feedback from food photos as represented nutrition information to be used later in the proposed recommender system. As described in Chapter 3 a machine-learning framework (Facebook inverse cooking Model) for predicting food ingredients.

The framework's result is the food meal's title with the ingredients. Table 4.1 shows the framework results of a sample of food photos was used.

Table 4.1: A Sample of Framework Results of a Sample of Food Photos

Meal Food Photo	**Title**	**Ingredients Production**
	Chicken tacos	tortilla, pepper, oil, onion, cilantro, salt, corn, cumin, clove
	Chicken and vegetable soup	pepper, onion, oil, chicken, salt, broth, celery
	Blt & cheese sandwich	lettuce, bun, cheese, tomato, cream, bacon, pepper

Table 4.2 illustrates the model's test accuracy examples depend on the correct prediction of components. The more correct the estimate of the components, the more accurate and better the result of calculating nutritional information.

Table 4.2: Ingredient Prediction Examples

Meal Food Photo	Ingredients Production	Ingredients Real	Extracted Nutritional Information/g	Real Nutritional Information/g
	Onion, chicken, pepper, lettuce, barbecue_sauce	Chicken, tomatoes, onion, perpper, lettuce, beans, corn, BBQ sauce	Cal/g = 4.32 Fat/g = 0.047 Carb/g = 0.62 Prot/g = 0.36	Cal/g = 7.14 Fat/g = 0.11 Carb/g = 1.11 Prot/g = 0.46
	Oil, pepper, onion, seeds, spinach, vinegar, salt, egg, sugar	Spinach, tahini, miso pasta, sesame, oil, caster suger, lime juice, toasted sesame seeds	Cal/g = 20.91 Fat/g = 1.604 Carb/g = 1.33 Prot/g = 0.49	Cal/g = 28.48 Fat/g = ٢,٠٨٩ Carb/g = ٢,٠٩ Prot/g = ٠,٦٨

Figure 4.1 shows the high predictive accuracy rate of the nutritional estimation model of the food images after taking 53 samples, inserting them into the proposed model, and comparing the estimated ingredients from the proposed model with the actual ingredients of the food. Worthy of attention is that the precision of the suggested model is primarily dependent on the precision of the ingredients prediction model.

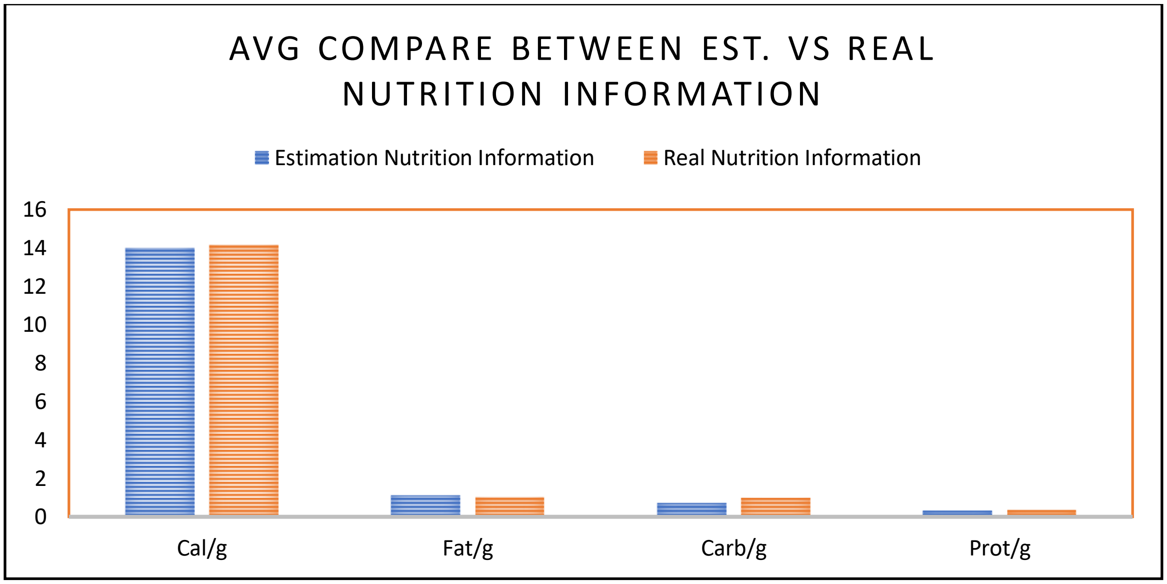

Figure 4.1: Accuracy Rate Ingredient Prediction for Proposed Model

4.3 User Profile Creation

The second stage is creating a user profile to be used later in the recommender system. As described in chapter three, dropping unnecessary fields, adding new implicit feedback extraction from geotagged photos (nutrition information) and multiple filters to the Yelp dataset were applied to create a user profile for related features of restaurants to be used later in the recommendation as shown in Table 4.3.

Table 4.3: User Profile Example

Columns	Data	DataType
Index	0	int
user_id	5	Int
business_id	1031	Int
Name	Porch Light Latin Kitchen	Str

Stars	5	Int
CalRating/g	4	Int
FatRating/g	2	Int
ProtRating/g	1	Int
checkin_count	314	int

4.4 The Proposed Hybrid Rating Matrix

The most critical phase in Recommender System is the rating matrix phase. The proposed rating matrix is a hybrid of original metadata and implicit feedback features extracted as shown in Table 4.3. Table 4.4 illustrates the rating for each restaurant after adding nutrition information to the original features in the proposed rating matrix. Some ratings increased because their food had high nutrition information, others were decreased because their food had low nutrition information. Finally, some restaurants have moderate nutrition information, which is still rated without change.

Table 4.4: The Proposed Rating Matrix Samples

UserID	RestaurantID	Old Rating	Proposed Rating
-7402470122358638537	9217269774983615800	5	٤
-6765064205309550796	-3459707240469909404	5	٤
6060192195421711078	-2645252880351862891	١	٢
-1053509608820923404	-4858826198152703627	٣	٤
-3561313031674806888	-257564561779987369	٤	٤

4.5 The Proposed Recommender System

LightGCN was adopted in this thesis to compare its findings with previous researchers as shown in Table 4.5.

Table 4.5: Compared LightGCN with Other Models

Dataset	**Yelp**	
Method	**Recall**	**NDCG**
GRMF [69] – 2015	0.0571	0.0462
NGCF [53] – 2019	0.0579	0.0477
GRMF-norm [71] –2019	0.0561	0.0454
Mult-VAE [72] – 2020	0.0584	0.0450
LightGCN [58] – 2020	0.0649	0.0530

4.5.1 System Training

After creating a hybrid, the proposed rating matrix must describe the results of training the proposed system using the LightGCN model to note the loss function development of the model through increased epochs iteration to get the best weights to use in system evaluation.

The Model configuration used in the training phase is as follows:

n_layers=64,

batch_size=20000,

epochs=1000,

learning_rate=0.005,

eval_epoch=5,

top_k=10,

Figure 4.2 describes the training phase through 1000 epochs.

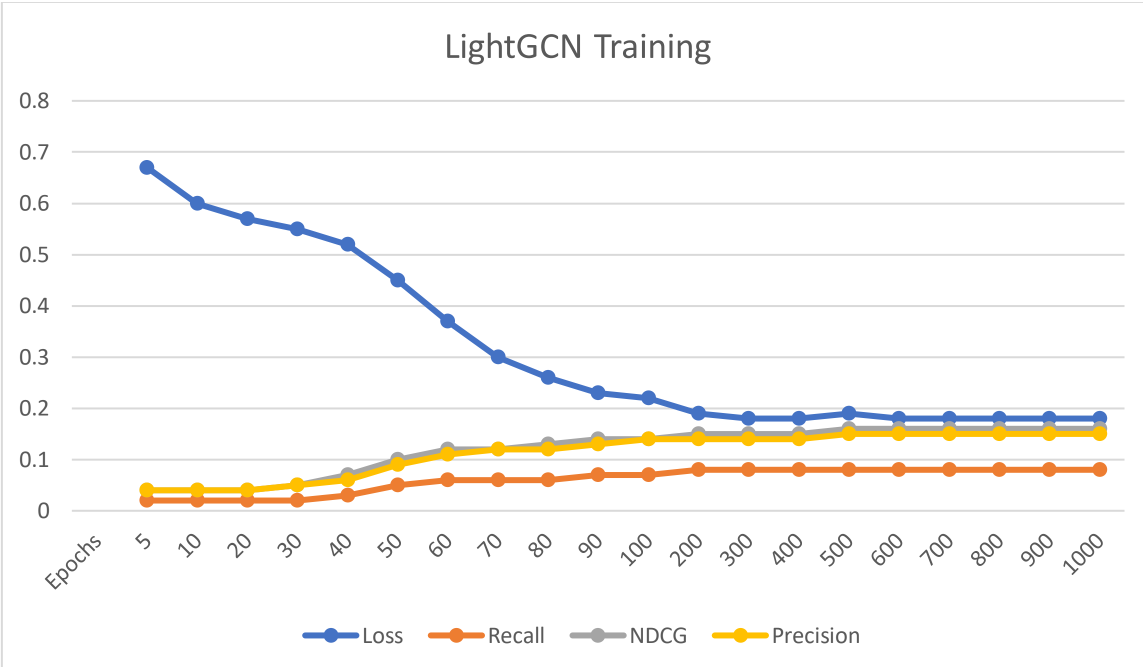

Figure 4.2: The Evaluation Metrics and Loss Through Training Phase

4.5.2 System Evaluation

The results show the superiority of the LightGCN model over other models to verify the system. After training LightGCN Model multiple times in different numbers on top of K, the result is shown in Table 4.6.

Table 4.6: Results of the Recommender System for Multiple k

Recommended Items Count	Avg Recall System Users	Avg NDCG System Users	Avg Precision System Users
500	0.97467	0.50187	0.03596
400	0.96321	0.4976	0.04444
300	0.92845	0.48678	0.0572
200	0.79245	0.4403	0.07359
100	0.5138	0.33457	0.09578
80	0.43742	0.30259	0.10184
50	0.31367	0.24682	0.11702
25	0.18365	0.18353	0.1364
10	**0.09673**	**0.19156**	**0.1789**

5	0.05481	0.20987	0.2014
1	0.01296	0.237	0.237
Average	**0.479256364**	**0.330226364**	**0.116320909**

The evaluation metrics used for measuring the performance of the system (Recall, precision, Normalized Discounted Cumulative Gain (NDCG)) depend on the number of top K (number of recommender restaurants).

The result showed that when the top K =1, the highest precision and the lowest recall were obtained. The NDCG is the avg between recall and precision, as shown in Figure 4.3.

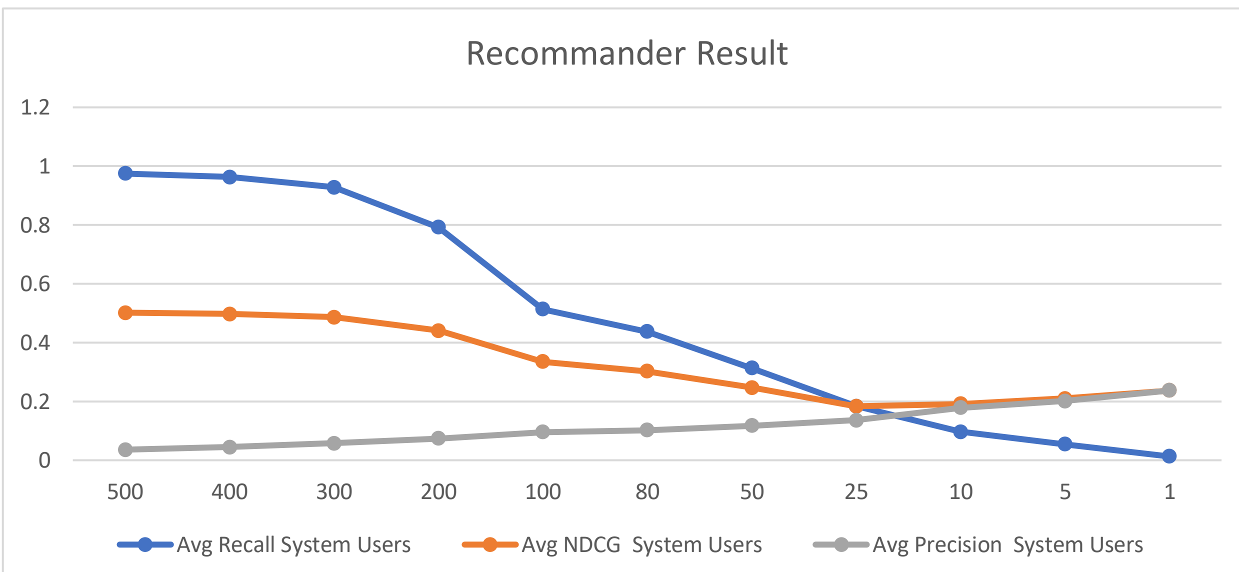

Figure 4.3: Avg Recommander Results

Precision and recall are the two most popular and practical criteria for assessing the quality of relevant predictions. However, recommender systems have a life of their own. If precision is chosen, The measurements will focus on how well relevant items are retrieved (already rated as good). Nevertheless, greater precision reduces the potential to provide original and recommendations (false positives)—for instance, a recommender system for cancer [73].

The result in Table 4.7 illustrates the proposed systems result's superiority using the same model on the proposed rating matrix with the same experiment parameters (records = 61.437) compared with the benchmark paper results.

Table 4.7: Comparing the Results with Other Works

LightGCN Model	Recommended Items Count	Recall System Users	NDCG System Users
Results of the proposed system	10	0.09673	0.19156
Benchmark Paper Results	10	0.0649	0.0530

Figure 4.4 illustrates the superiority of the proposed system's results compared to other models and benchmark paper results.

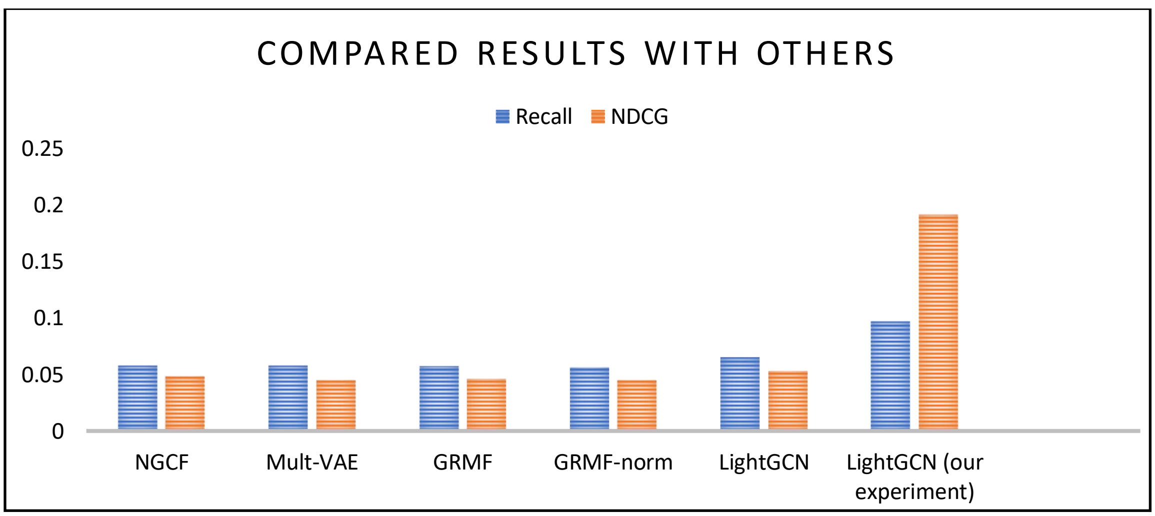

Figure 4.4: The Flowchart of Compare Results with Others methods

The key challenge with this dataset is the sparsity that causes these results.

CHAPTER FIVE

Conclusions and Future Works

Chapter Five

Conclusions and Future Works

5.1 Conclusions

Based on the suggested system, the following conclusions can be drawn:

At the moment, obesity is a severe worry of human existence, and people have developed interested methods in monitoring their weight and eating habits to avoid becoming obese. Thus, this research presents an approach to show the type of food people consume and its characteristics. This thesis has developed a system for nutritional assessment that takes an image of a meal and generates a title and ingredients list. This is a study from rare studies to estimate food ingredients from food photographs and then evaluate nutritional information based on the predicted ingredients.

The system proposed implicit feedback extractions to recommend the best restaurants, and it depends on generating new features from geotagged food photos shared by users on Yelp business social media.

The applied LightGCN model predicts the best restaurants based on nutritional information and other related features in the proposed system. The results show that the proposed system outperformed other models. The overall findings are avg recall of 47% and avg NDCG of 33% with different experiments on different numbers of K.

The suggested rating matrix generated from implicit features with other metadata minimized sparsity in the original dataset based on implicit feedback.

The system cannot add new users because of the “cold start” recommendation problem, the system training is offline, and the pre-trained model is not compatible with photos taken by users that causing some problems in predicting meal ingredients that are not accurate nutritional information estimation.

5.2 Future Works

Further improvement is still required according to the concepts and models utilized to develop this system. Numerous concepts might be used to enhance and improve the system and its models such as:

- Creating a new Geo-Social recommender system by combining a separate social network with the proposed system.
- Adding new users to recommender systems is one of the objectives of such systems; therefore, developing a mechanism to add new users is one of the future efforts.
- Instead of Top-N suggestions, different forms of distribution can be used to determine the list of nearby restaurants and/or the list of recommended items.
- Utilizing or enhancing machine learning model to predict food ingredients from food photos to improve the accuracy of nutritional information.
- Real-time applications can also be used to access online databases to improve the proposed system.
- In the future, the proposed system can be applied to Iraqi restaurants after collecting a complete dataset on Iraqi restaurants.

References

References

[1] M. Srifi, A. Oussous, A. Ait Lahcen, and S. Mouline, “Recommender Systems Based on Collaborative Filtering Using Review Texts—A Survey.” Information, vol. 11, no. 6, pp. 317-338, Jun. 2020, doi: 10.3390/info11060317.

[2] H. M. Z. Hasan, H. Khan, T. Asif, S. Hashmi, and M. Rafi, “Towards a transfer learning approach to food recommendations through food images,” in Proceedings of the 3rd International Conference on Machine Learning and Soft Computing - ICMLSC 2019, Jan. 2019, pp. 99–105, doi: 10.1145/3310986.3310990

[3] K. Bauman, B. Liu, and A. Tuzhilin, “Estimating Customer Reviews in Recommender Systems Using Sentiment Analysis Methods,” New York University, New York University, 2014.

[4] De, Clara, and Paolis Kaluza. “Recommender System for Yelp Dataset CS6220 Data Mining Northeastern University.” Https://Www.khoury.northeastern.edu/, Khoury College of Computer Sciences, 2016, pp. 1–10.

[5] “Yelp: Restaurants, Dentists, Bars, Beauty Salons, Doctors.” Yelp. https://www.yelp.com/ (accessed: Sep. 16, 2022).

[6] N. Carrillo et al., Recommender Systems Designed for Yelp.com. University of California: Department of Mathematics, 2015, pp. 1–8.

[7] X. Lei, X. Qian, and G. Zhao, “Rating Prediction Based on Social Sentiment From Textual Reviews,” IEEE Transactions on Multimedia, vol. 18, no. 9, IEEE, IEEE Transactions on Multimedia, pp. 1910–1921, Sep. 2016. doi: 10.1109/tmm.2016.2575738.

[8] R. Duan, R. S. M. Goh, F. Yang, Y. K. Tan, and J. F. B. Valenzuela, “Towards building and evaluating a personalized location-based recommender system,” in IEEE Xplore, Washington, DC, USA, Oct. 2014, pp. 43–48. doi: 10.1109/BigData.2014.7004211.

[9] O. Tal and Y. Liu, “TCENR: A Hybrid Neural Recommender for Location Based Social Networks,” in 2018 IEEE International Conference on Data Mining Workshops (ICDMW), Singapore, Nov. 2018, pp. 1186–1191. doi: 10.1109/ICDMW.2018.00170.

[10] C. Sundermann, J. Antunes, M. Domingues, and S. Rezende, “Exploration of Word Embedding Model to Improve Context-Aware Recommender Systems,” in 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Santiago, Chile, Dec. 2018, pp. 383–388. doi: 10.1109/WI.2018.00-64.

[11] X. Wang, M. Nguyen, J. Carr, L. Cui, and K. Lim, “A group preference-based privacy-preserving POI recommender system,” ICT Express, vol. 6, no. 3, pp. 204–208, Sep. 2020, doi: 10.1016/j.icte.2020.05.005.

[12] D. N R, D. S. G K, and D. P. Kumar Pareek, “A Framework for Food recognition and predicting its Nutritional value through Convolution neural network,” SSRN, Rochester, NY, pp. 1–6, Feb. 22, 2022.

[13] M. Zhang and L. Luo, “Can Consumer-Posted Photos Serve as a Leading Indicator of Restaurant Survival? Evidence from Yelp.” Management Science, vol. 26, no. 3, pp. 1-83, Apr. 2022, doi: 10.1287/mnsc.2022.4359..

[14] H. Luo, X. Zhang, and G. Guoy, “Convolutional Attention Model For Restaurant Recommendation With Multi-View Visual Features,” in 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, Oct. 2020, pp. 838–842. doi: 10.1109/ICIP40778.2020.9190768.

[15] X. Zhang, H. Luo, B. Chen, and G. Guo, “Multi-view visual Bayesian personalized ranking for restaurant recommendation,” Applied Intelligence, vol. 50, no. 9, pp. 2901–2915, Apr. 2020, doi: 10.1007/s10489-020-01703-6.

[16] H. Peng, D. Yiqing, and B. Huang, “Capstone Project CSC 591 -Algorithms for Data Guided Business Intelligence Yelp Restaurant Photo Classification,” OmkarAcharya, 2017.

[17] F. Sun, Z. Gu, and B. Feng, “Yelp Food Identification via Image Feature Extraction and Classification,” arXiv:1902.05413 [cs, stat], vol. 25, no. 4, pp. 1–5, Feb. 2019, doi: 10.48550/arXiv.1902.05413.

[18] R. D. Kumar, E. G. Julie, Y. H. Robinson, S. Vimal, and S. Seo, “Recognition of food type and calorie estimation using neural network,” The Journal of Supercomputing, vol. 77, no. 8, pp. 8172–8193, Jan. 2021, doi: 10.1007/s11227-021-03622-w.

[19] H. Liang, Y. Gao, Y. Sun, and X. Sun, “CEP: calories estimation from food photos,” International Journal of Computers and Applications, vol. 42, no. 6, pp. 569–577, Jun. 2018, doi: 10.1080/1206212x.2018.1486558.

[20] N. Varatharajan, J. Guruprasad, and K. Mathumitha, “RESTAURANT RECOMMENDATION SYSTEM USING MACHINE LEARNING,” INTERNATIONAL EDUCATIONAL APPLIED RESEARCH JOURNAL (IEARJ), vol. 4, no. 3, pp. 1–4, Mar. 2022,

[21] V. J. C and J. S. Raj, “Location-based Orientation Context Dependent Recommender System for Users,” March 2021, vol. 3, no. 1, pp. 14–23, Apr. 2021, doi: 10.36548/jtcsst.2021.1.002.

[22] N. Nag, V. Pandey, and R. Jain, “Live Personalized Nutrition Recommendation Engine,” Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care, vol. 25, no. 7, Oct. 2017, doi: 10.1145/3132635.3132643.

[23] A. Sima and P. Loudjani, “Use of geotagged photographs in the frame of Common Agriculture Policy checks,” pdf, Joint Research Centre, 2020.

[24] P. Chandekar, “Evaluate your Recommendation Engine using NDCG,” Medium, Oct. 25, 2021. https://towardsdatascience.com/evaluate-your-recommendation-engine-using-ndcg-759a851452d1 (accessed Aug. 14, 2022).

[25] Maher Malaeb, “Recall and Precision at k for Recommender Systems,” Medium, Aug. 13, 2017. https://medium.com/@m_n_malaeb/recall-and-precision-at-k-for-recommender-systems-618483226c54 (accessed Aug. 15, 2022).

[26] Wikipedia Contributors, “Precision and recall,” Wikipedia, Apr. 19, 2019. https://en.wikipedia.org/wiki/Precision_and_recall (accessed Aug. 17, 2022).

[27] C. Barros, B. Moya-Gómez, and J. C. García-Palomares, “Identifying Temporal Patterns of Visitors to National Parks through Geotagged Photographs,” Sustainability, vol. 11, no. 24, p. 6983, Dec. 2019, doi: 10.3390/su11246983.

[28] “Understanding ResNet50 architecture,” OpenGenus IQ: Learn Computer Science, Mar. 30, 2020. https://iq.opengenus.org/resnet50-architecture/ (accessed Aug. 19, 2022).

[29] M. Memarzadeh and A. Kamandi, “Model-Based Location Recommender System Using Geotagged Photos On Instagram,” in 2020 6th International Conference on Web Research (ICWR), Tehran, Iran, Apr. 2020, pp. 203–208. doi: 10.1109/ICWR49608.2020.9122274.

[30] Y. Kang, N. Cho, J. Yoon, S. Park, and J. Kim, “Transfer Learning of a Deep Learning Model for Exploring Tourists’ Urban Image Using Geotagged Photos,” ISPRS International Journal of Geo-Information, vol. 10, no. 3, p. 137, Mar. 2021, doi: 10.3390/ijgi10030137.

[31] A. Salvador, M. Drozdzal, X. Giro-i-Nieto, and A. Romero, “Inverse Cooking: Recipe Generation From Food Images,” in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 10445–10454. doi: 10.1109/CVPR.2019.01070.

[32] N. Kliemann et al., “Serving Size and Nutrition Labelling: Implications for Nutrition Information and Nutrition Claims on Packaged Foods,” Nutrients, vol. 10, no. 7, p. 891, Jul. 2018, doi: 10.3390/nu10070891.

[33] P. Achananuparp, E.-P. Lim, and V. Abhishek, “Does Journaling Encourage Healthier Choices?,” Proceedings of the 2018 International Conference on Digital Health - DH ’18, vol. 7, no. 21, pp. 35–44, Apr. 2018, doi: 10.1145/3194658.3194663.

[34] “Home,” www.who.int. https://www.who.int/en (accessed Sep. 16, 2022).

[35] “Obesity: Facts and statistics,” www.medicalnewstoday.com, Feb. 26, 2020. https://www.medicalnewstoday.com/articles/319902#age (accessed Oct. 21, 2022).

[36] J. Chung, J. Chung, W. Oh, Y. Yoo, W. G. Lee, and H. Bang, “A glasses-type wearable device for monitoring the patterns of food intake and facial activity,” Scientific Reports, vol. 7, no. 1, Jan. 2017, doi: 10.1038/srep41690.

[37] S. Sajadmanesh et al., “Kissing Cuisines,” in Proceedings of the 26th International Conference on World Wide Web Companion - WWW ’17 Companion, Apr. 2017, vol. 22, no. 8, pp. 1013–1021. doi: 10.1145/3041021.3055137.

[38] A. Callahan, H. Leonard, and T. Powell, "Classification of Nutrients," openoregon.pressbooks.pub, Oct. 14, 2019. https://openoregon.pressbooks.pub/nutritionscience/chapter/1c-classification-of-nutrients/ (accessed Aug. 16, 2022).

[39] S. Kanoje, S. Girase, and D. Mukhopadhyay, "User Profiling Trends, Techniques and Applications," International Journal of Advance Foundation and Research in Computer (IJAFRC), vol. 1, no. 1, pp. 1–6, Mar. 2015.

[40] P. K. Singh, E. Othman, R. Ahmed, A. Mahmood, H. Dhahri, and P. Choudhury, "Optimized recommendations by user profiling using apriori algorithm," Applied Soft Computing, vol. 106, no. 15, p. 107272, Jul. 2021, doi: 10.1016/j.asoc.2021.107272.

[41] C. I. Eke, A. A. Norman, L. Shuib, and H. F. Nweke, "A Survey of User Profiling: State-of-the-Art, Challenges, and Solutions," IEEE Access, vol. 7, pp. 144907–144924, Sep. 2019, doi: 10.1109/access.2019.2944243.

[42] T. Kulkarni, M. Kabra, and R. Shankarmani, "User Profiling Based Recommendation System for E-Learning," in 2019 IEEE 16th India Council International Conference (INDICON), Rajkot, India, Dec. 2019, pp. 1–4. doi: 10.1109/INDICON47234.2019.9028982.

[43] K. Gatziolis and A. C. Boucouvalas, "User profile extraction engine," in the 20th Pan-Hellenic Conference, New York , NY , United States, Nov. 2016, pp. 1–6.

[44] Rahul Agarwal, "The 5 Classification Evaluation metrics every Data Scientist must know," Medium, Sep. 17, 2019. https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226 (accessed Aug. 17, 2022).

[45] Ndjiongue, A. R., Ngatched, T. M., Dobre, O. A., Armada, A. G., & Haas, H. (2021). Performance Analysis of RIS-Based nT-FSO Link Over G-G Turbulence With Pointing Errors. arXiv. https://doi.org/10.48550/arXiv.2102.03654

[46] B. Kelly, A. Manela, and A. Moreira, "Text Selection," Journal of Business & Economic Statistics, vol. 39, no. 4, pp. 859–879, Jul. 2021, doi: 10.1080/07350015.2021.1947843.

[47] M. A. Siddiqui, A. Fern, T. G. Dietterich, R. Wright, A. Theriault, and D. W. Archer, "Feedback-Guided Anomaly Discovery via Online Optimization," in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Jul. 2018, p. Pages 2200–2209. doi: 10.1145/3219819.3220083.

[48] Gao, C., Lei, W., He, X., de Rijke, M., & Chua, T. (2021). Advances and Challenges in Conversational Recommender Systems: A Survey. arXiv. https://doi.org/10.1016/j.aiopen.2021.06.002

[49] R. Mu, "A Survey of Recommender Systems Based on Deep Learning," IEEE Access, vol. 6, pp. 69009–69022, 2018, doi: 10.1109/access.2018.2880197.

[50] T. Di Noia, R. Mirizzi, V. C. Ostuni, D. Romito, and M. Zanker, "Linked open data to support content-based recommender systems," in Proceedings of the 8th International

Conference on Semantic Systems - I-SEMANTICS '12, 2012, pp. 1–8. doi: 10.1145/2362499.2362501.

[51] T. Hussein, T. Linder, W. Gaulke, and J. Ziegler, "Hybreed: A software framework for developing context-aware hybrid recommender systems," User Modeling and User-Adapted Interaction, vol. 24, no. 1–2, pp. 121–174, Dec. 2012, doi: 10.1007/s11257-012-9134-z.

[52] X. Li, M. Wang, and T.-P. . Liang, "A multi-theoretical kernel-based approach to social network-based recommendation," Decision Support Systems, vol. 65, pp. 95–104, Sep. 2014, doi: 10.1016/j.dss.2014.05.006.

[53] Wang, X., He, X., Wang, M., Feng, F., & Chua, T. (2019). Neural Graph Collaborative Filtering. arXiv. https://doi.org/10.1145/3331184.3331267

[54] W. Shi, L. Wang, and J. Qin, "User Embedding for Rating Prediction in SVD++-Based Collaborative Filtering," Symmetry, vol. 12, no. 1, p. 121, Jan. 2020, doi: 10.3390/sym12010121.

[55] X. He, Z. He, J. Song, Z. Liu, Y.-G. Jiang, and T.-S. Chua, "NAIS: Neural Attentive Item Similarity Model for Recommendation," IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 12, pp. 2354–2366, Dec. 2018, doi: 10.1109/tkde.2018.2831682.

[56] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN," in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Jul. 2020, pp. 639–648. doi: 10.1145/3397271.3401063.

[57] Devooght, R., & Bersini, H. (2016). Collaborative Filtering with Recurrent Neural Networks. arXiv. https://doi.org/10.48550/arXiv.1608.07400

[58] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "LightGCN: Simplifying Graph Convolutional Networks," in Proceedings of the 36th International Conference on Machine Learning, May 2019, vol. 97, pp. 6861–6871.

[59] L. Zhang and H. Lu, "A Feature-Importance-Aware and Robust Aggregator for GCN," in Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Oct. 2020, pp. 1813–1822. doi: 10.1145/3340531.3411983.

[60] J. Liu, G. P. Ong, and X. Chen, "GraphSAGE-Based Traffic Speed Forecasting for Segment Network With Sparse Data," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 3, pp. 1–12, 2020, doi: 10.1109/tits.2020.3026025.

[61] Zhu, H., Feng, F., He, X., Wang, X., Li, Y., Zheng, K., & Zhang, Y. (2020). Bilinear Graph Neural Network with Neighbor Interactions. arXiv. https://doi.org/10.48550/arXiv.2002.03575

[62] T. Alasadi and W. Baiee, "A New Technique for Generating Implicit Feedbacks from Users Movements History," in International Conference on Change, Innovation, Informatics and Disruptive Technology 2016, London - U.K, Oct. 2016, pp. 393–406.

[63] M. Abdel-Basset, M. Mohamed, M. Elhoseny, L. H. Son, F. Chiclana, and A. E.-N. H. Zaied, "Cosine similarity measures of bipolar neutrosophic set for diagnosis of bipolar disorder diseases," Artificial Intelligence in Medicine, vol. 101, p. 101735, Nov. 2019, doi: 10.1016/j.artmed.2019.101735.

[64] "Yelp Dataset," Yelp.com, 2019. https://www.yelp.com/dataset (accessed Aug. 17, 2022).

[65] Thames, Q., Karpur, A., Norris, W., Xia, F., Panait, L., Weyand, T., & Sim, J. (2021). Nutrition5k: Towards Automatic Nutritional Understanding of Generic Food. arXiv. https://doi.org/10.48550/arXiv.2103.03375

[66] "USDA," Usda.gov, 2018. https://www.usda.gov (accessed Sep. 17, 2022).

[67] T. Ege and K. Yanai, "Multi-task learning of dish detection and calorie estimation," in Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management, Jul. 2018, pp. 53–58. doi: 10.1145/3230519.3230594.

[68] H. Zha, X. He, C. Ding, H. Simon, and M. Gu, "Bipartite graph partitioning and data clustering," in Proceedings of the tenth international conference on Information and knowledge management - CIKM'01, Oct. 2001, pp. 25–32. doi: 10.1145/502585.502591.

[69] N. Rao, H.-F. Yu, P. K. Ravikumar, and I. S. Dhillon, "Collaborative Filtering with Graph Information: Consistency and Scalable Methods," Neural Information Processing Systems, 2015. https://proceedings.neurips.cc/paper/2015/hash/f4573fc71c731d5c362f0d7860945b88-Abstract.html (accessed Oct. 16, 2022).

[70] S. Cristina, "The Transformer Model," Machine Learning Mastery, Nov. 03, 2021. https://machinelearningmastery.com/the-transformer-model/ (accessed Aug. 20, 2022).

[71] Z. Cui, Y.-L. Gao, J.-X. Liu, L.-Y. Dai, and S.-S. Yuan, "L2,1-GRMF: an improved graph regularized matrix factorization method to predict drug-target interactions," BMC Bioinformatics, vol. 20, no. S8, Jun. 2019, doi: 10.1186/s12859-019-2768-7.

[72] M.-D. Nguyen and Y.-S. Cho, "A Variational Autoencoder Mixture Model for Online Behavior Recommendation," IEEE Access, vol. 8, pp. 132736–132747, 2020, doi: 10.1109/access.2020.3010508.

[73] Maher Malaeb, "Recall and Precision at k for Recommender Systems," Medium, Aug. 13, 2017. https://medium.com/@m_n_malaeb/recall-and-precision-at-k-for-recommender-systems-618483226c54 (accessed Aug. 17, 2022).

[74] "Matrix Grid Reports - Normalized Ratings, Custom Weights and Average Rating - SAP SuccessFactors - Support Wiki," wiki.scn.sap.com. https://wiki.scn.sap.com/wiki/display/SAPSF/Matrix+Grid+Reports+-+Normalized+Ratings%2C+Custom+Weights+and+Average+Rating (accessed Sep. 17, 2022).

[75] “What Is Deep Learning? | How It Works, Techniques & Applications,” Mathworks.com, 2019. https://www.mathworks.com/discovery/deep-learning.html (accessed Oct. 20, 2022).

[76] I. Mayachita, “Understanding Graph Convolutional Networks for Node Classification,” Medium, Aug. 18, 2020. https://towardsdatascience.com/understanding-graph-convolutional-networks-for-node-classification-a2bfdb7aba7b (accessed Oct. 20, 2022).

[77] M. Harmouch, “17 types of similarity and dissimilarity measures used in data science.,” Medium, Apr. 02, 2021. https://towardsdatascience.com/17-types-of-similarity-and-dissimilarity-measures-used-in-data-science-3eb914d2681 (accessed Oct. 20, 2022).

Appendix A

The Published Papers

2nd AL-Muthanna International Conference on Engineering Science and Technology MICEST 2022

To:

12/03/2023

Mustafa Al-Saffar

University of Babylon,

Email: mustafaalsaffar.sw.msc@student.uobabylon.edu.iq

We are pleased to inform you that the peer reviewed manuscript code **MIC-22228** entitled **(Survey on Implicit Feedbacks Extraction based on Yelp Dataset using Collaborative Filtering)** has been accepted for oral presentation as well as inclusion in the conference proceedings of the 2nd Al-Muthanna International Conference on Engineering Science and Technology MICEST-2022 to be held at Al-Muthanna University, Samawah, Iraq during March, 16-17, 2022.

All the IEEE track accepted papers will also be considered for publication in the Scopus-indexed conference proceeding at IEEE Xplore

We are looking forward to welcoming you in the MICEST 2022: 2nd Al-Muthanna International Conference on Engineering Science and Technology.

Sincerely,
Conference co-chair, Prof. Dr. Ahmed Hasan Ali
MICEST 2022
https://micest.org/

2nd ALMuthanna International Conference on Engineering Science and Technology (MICST-2022) , Almuthana University, Almuthana, Iraq

Survey on Implicit Feedbacks Extraction based on Yelp Dataset using Collaborative Filtering

Mustafa Al-Saffar
Software Department
University of Babylon
Hilla, Iraq
mustafaalsaffar.sw.msc@student.uobabylon.edu.iq

Wadhah Baiee
Software Department
University of Babylon
Hilla, Iraq
wadhah.baiee@uobabylon.edu.iq

Abstract— **In e-commerce websites, associated micro-blogs, and business social media, users provide online feedback demonstrating their preferences for different items. These studies are usually found in textual comments, reviews, geo-tagged photos, and other contextual data and account for essential user preferences. Several factories have recently utilized review texts and the amount of information associated with them, such as review words, review subjects, and review moods. They also employed social photographs and other contextual information to improve collaborative filtering recommender systems based on ratings. These efforts employ review texts, geo-tagged photographs, and other contextual information to determine user preferences. This study gives a targeted survey of the most recent studies that mix review texts, photographs, and other contextual information and explores how these metadata and visual information are used to solve some of the most critical topics in Algorithms for collaborative filtering.**

Keywords— Collaborative Filtering; Contextual Information Geotagged Photos; Recommender Systems; Survey; User Reviews; Yelp Dataset.

I. INTRODUCTION

Nowadays, e-commerce websites flourish fast and allow hundreds of thousands of selling items [1]. The preference of an item from this large quantity of items produces essentially the usage of an additional tool referred to as a recommender system [16,17]. The Recommender System (R.S.) provides a way for consumers to discover things they might not have discovered on their own. It collects information about users' favorite things and then suggests them [20]. The Collaborative Filtering (C.F.) The approach is employed by a variety of e-commerce companies [17], including Amazon Prime Video (www.primevideo.com/), Yelp (www.yelp.com/),marketplace(www.facebook.com/market place/), and AliExpress (/www.aliexpress.com/). The popularity of C.F. techniques is based on the participants' shared interests. Calculating the users' standard ratings [18] is used to observe similar users or things. When there is sufficient rating information, C.F. techniques work well [16].

Nevertheless, their power is in pain when the rating sparsity problem happens because users frequently have restricted standard rating numbers [1]. Further, C.F. does not capture the rationale behind user ratings. As a result, it is impossible to record a target user's choice [13] precisely. To address these issues, a variety of content-based solutions have been created to act on behalf of users and things using various types of data, such as tags [24], item descriptions [10], geotagged images, and social aspects [12]. After all, these strategies are useless when rating sparsity is high, or the target user has a small number of previous ratings [20]. With the Web's flow scenario, users have more options. Textual evaluations [12], location photographs, locations, and buddy networks have all evolved. People are more comfortable expressing themselves and debating their points of view on e-platforms. As a result, customer reviews have become an integral aspect of e-commerce. Websites like Booking(www.Booking.com/) and Yelp and online point websites like Amazon are examples of forum websites amassing vast quantities of online reviews [17]. Both organizations and clients gain extensively from the treasured and wealthy knowledge of studies [13]and photos. More semantic information can be found in rating information, written evaluations, and images, giving a recommender system that is more fine-grained, nuanced, and trustworthy when it comes to user preferences. [18].

Simultaneously, The approach can create a thorough preference depiction for the user that is impossible to obtain from international rating scores [21]. Not long ago, many attempts were made to capture For User interest information from review texts used for rating prediction [22]. As a result of these studies, review texts have been shown to positively impact the overall performance of available rating-based systems [6],[12],[14]. As a result, this work focuses on extracting implicit user feedback and surveys of recent studies on recommender systems based on the Yelp Dataset. The rich data available in reviews, photos, and other contextual data is used to address the critical problems with traditional rating-based systems, such as sparsity and forecast accuracy. The remainder of this article is divided into the following sections.

II. YELP DATASET

"The Yelp dataset is a subset of our businesses, reviews, and user data for use for personal, educational, and academic purposes. Available as JavaScript Object Notation (JSON) files, use it to teach students about databases, learn Natural Language Processing (NLP), or sample production data while learning how to make mobile apps. It contains 10,000 No of restaurants with restaurant location, menu items, and ratings. Yelp publishes over 280,000 pictures from over 2000 businesses, the dataset of which can be found on the Yelp Data Challenge Website. This dataset consists of inside, outside, drink, and food photos" [26]; we will survey recent work relevant to this dataset.

Intelektual Pustaka
Media Utama

Institute of Advanced Engineering and Science

Scopus® indexed
CiteScoreTracker 2022: 2.3 | h-index: 16

SJR Q3 0.357 0.730 2.4

CERTIFICATE

No. 4007/BEEI/A/08/2022

Bulletin of Electrical Engineering and Informatics (BEEI)

is hereby awarding this certificate to

Mustafa Al-Saffar, Wadhah R. Baiee

in recognition of his/her contribution in this scientific journal

as *Authors* for paper entitled:

Nutrition information estimation from food photos using machine learning based on multiple datasets

in Vol. 11, No. 5, October 2022

ISSN 2089-3191
http://beei.org

Yogyakarta, August 19, 2022

Tole Sutikno
Managing Director, Journals

Bulletin of Electrical Engineering and Informatics
Vol. 11, No. 5, October 2022, pp. 2922~2929
ISSN: 2302-9285, DOI: 10.11591/eei.v11i5.4007

Nutrition information estimation from food photos using machine learning based on multiple datasets

Mustafa Al-Saffar, Wadhah Baiee
Department of Software, Information Technology, University of Babylon, Babylon, Iraq

Article Info

Article history:

Received Apr 26, 2022
Revised Jul 10, 2022
Accepted Jul 25, 2022

Keywords:

Food ingredients
Food photos
Machine learning
Nutrition information
Yelp dataset

ABSTRACT

Bodyweight, blood pressure, and cholesterol are all risk variables that can aid people in making educated decisions regarding their health promotion activities. Food choices are among the most effective methods for preventing chronic illnesses, including heart disease, diabetes, stroke, and some malignancies. Because various meals give varying amounts of energy and minerals, good eating necessitates keeping track of the nutrients we ingest. Furthermore, there is a paucity of information on whether understanding food constituents might aid in more accurate nutrition calculations. Therefore, this research suggests processing food images on social media to anticipate the contents of each food and extracting nutrition information for each food image to serve as healthy implicit feedback to take advantage of the rapid accumulation of rich photos on social media. The proposed methodology is a framework based on a machine-learning model for predicting food ingredients. We also compute critical health metrics for each ingredient and combine them to obtain nutrition data for the food. The result revealed a promising way of extracting food components and nutrition information. Compared with other researchs, our proposed prediction and attribute extraction strategy achieves a remarkable accuracy of 85%.

Corresponding Author:

Mustafa Al-Saffar
Department of Software, Information Technology, University of Babylon
80-street, Hilla, Babil, Iraq
Email: mustafaalsaffar.sw.msc@student.uobabylon.edu.iq

1. INTRODUCTION

Food significantly affects humans' quality of life [1], health, and happiness [2]. The number of overweight or obese persons is increasing. According to the WHO [3], over 1.9 billion obese adults between the ages of 18 and over 650 million obese adults, obesity is a significant cause of diseases. For these reasons, food-related research [4]–[6] has consistently been a hot topic and garnered significant attention from various sectors. Previously, food-related research focused on various topics, including food selection [7] and food perception [8]. These studies, however, were undertaken before the web transformed research in various fields.

Additionally, most approaches rely on small data sets, such as questionnaires, cookbooks, and recipes [9]. Nowadays, with the rapid rise of multiple networks such as social networks, mobile networks, and the Internet of Things, people can easily share food images, recipes, cooking videos, and meal diaries, resulting in vast food databases [10]. These food data suggest a wealth of knowledge and hence present significant prospects for food-related research, including the discovery of food perception principles [11], the analysis of culinary habits [6], and diet monitoring [5]. Additionally, network analysis, computer vision, machine learning, and data mining offer various unique data analysis tools. Recent breakthroughs in artificial intelligence (AI), notably in deep learning [8], have sparked renewed interest in large-scale food-related

المستخلص

القرارات الغذائية هي المؤشرات الرئيسية للسمنة وانسداد الشرايين وأمراض القلب. لا يمكن لمتخصصي الرعاية الصحية البشرية ، أن يكونوا دائمًا بجانب كل مستخدم ، لتوجيههم يدويًا نحو الخيارات المثلى. الإرشادات التكيفية المؤتمتة جنبًا إلى جنب مع معرفة الخبراء بإمكانية توسيع نطاق المشورة الصحية تقنيًا دون تدخل بشري.

تركز هذه الرسالة على توصية المستخدمين بأفضل المطاعم ؛ اعتمادا على معلومات المستخدمين. يحلل النظام المقترح صور الطعام للمطعم لاستخراج معلومات غذائية جديدة ويحسن دقة تنبؤ المطاعم اعتمادًا على حقائق التغذية.

تقدم الرسالة تقنيات الشبكة العصبية Convolutional Neural Network (CNN) لتحليل صور الطعام لاستخراج قائمة مكونات أطباق الوجبات. تُستخدم المكونات المستخرجة من الصور لحساب المعلومات الغذائية لكل وجبة. يتم دمج الميزات الكامنة المستخرجة من الصور مع معلومات المستخدمين الآخرى لبناء مصفوفة تقييم جديدة مقترحه للتحقق من تشابه المطاعم مع جيرانهم. ثم يختار النظام مجموعة أقرب المطاعم للتوصية بتفضيلات مماثلة للمستخدم الحالي. وبالتالي ، تم اعتماد نظام ترشيح تعاوني للتوصية في هذا العمل.

يتم تقييم النظام من خلال Recall و Normalized Discounted Cumulative (NDCG) Gain للمطاعم الموصى بها لجميع المستخدمين خلال فترة مجموعة التدريب ومقارنتها بالفترة الأخيرة لتفضيلات المستخدم كمجموعة اختبار. تتراوح قيم ال Recall بين [١٪ -٩٧٪] ، والمتوسط ٤٧٪ ، بينما تتراوح قيم NDCG بين [٢٣٪ -٥٠٪] و ٣٣٪ للنظام بأكمله. يتم عمل جميع التجارب على مجموعة بيانات Yelp التي تم استخدامها كقاعدة بيانات أساسية في نظامنا المقترح.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل
كلية تكنولوجيا المعلومات
قسم البرامجيات

اقتراح نظام توصية المطاعم على اساس التغذية الراجعة الضمنية لصور الطعام باستخدام شبكة التفاف الرسوم البيانية الخفيفة

رسالة

تُقدم بها الى مجلس كلية تكنولوجيا المعلومات - جامعة بابل

جزءاً من متطلبات نيل درجة الماجستير في تكنولوجيا المعلومات - برمجيات.

الطالب

مصطفى سليم منجي مهدي

بإشراف

م.د وضاح رزوقي عبود حسن بيعي

٢٠٢٢ م   ١٤٤٤ هـ