RECIPES RECOMMENDATION SYSTEM BASED ON INGREDIENTS AND USER PREFERENCES

THESIS PROPOSAL

By:

NABILAH ARGYANTI ARDYNINGRUM   NIM. 1941720083

INFORMATICS ENGINEERING STUDY PROGRAM

INFORMATION TECHNOLOGY DEPARTMENT

MALANG STATE POLYTECHNIC

2023

APPROVAL PAGE

RECIPES RECOMMENDATION SYSTEM BASED ON INGREDIENTS AND USER PREFERENCES

Arranged by:

NABILAH ARGYANTI ARDYNINGRUM.   NIM. 1941720083

This thesis proposal was tested on January 26, 2023
Approved by:

1. Main   :   Yan Watequlis Syaifudin, S.T.,
   Supervisor   M.M.T. Ph.D.   ....
                     NIP. 19810105 200501 1 005

2. Discussant I   :   Dwi Puspitasari, S.Kom., M.Kom.
                      NIP. 19791115 200501 2 002   ....

3. Discussant II :   Mamluatul Hani'ah, S.Kom.,
                      M.Kom.   ....
                      NIP. 19900206 201903 2 013

Best Regard,

Head of Major   Head of Study Program
Information Technology   Informatics Engineering

Rudy Ariyanto, S.T., M.Cs.   Imam Fahrur Rozi, S.T., M.T.
NIP. 19711110 199903 1 002   NIP. 19840610 200812 1 004

TABLE OF CONTENT

LIST OF IMAGES

LIST OF TABLES

CHAPTER I. PRELIMINARY

1.1. Background

Cooking is a daily activity that is carried out by most people to meet their daily food needs. Cooking can be done to make food for the family or to sell. Most people do cook activities every day. This activity is quite important because it can meet the nutritional needs of the body and can provide healthy and nutritious food. In addition, cooking can also be a fun activity and unites the family. Even some people think of cooking as a hobby, and there are also those who think cooking is a boring activity.

Cooking itself is the process of combining different food ingredients to produce delicious and nutritious food. The cooking process can be done in various ways, such as cooking with fire, cooking with water, cooking with a microwave, and so on. The cooking process can also use various basic ingredients such as meat, fish, vegetables, flour, sugar, salt, spices, and others. Cooking can be done using various techniques such as burning, frying, seasoning, processing, and so on. Each cooking technique can give it a different taste and appearance. In addition, cooking can also change the physical and chemical properties of food ingredients (Eufic, 2010).

People can cook by looking at recipes, instructions or guides that explain how to make a dish (Cambridge, n.d.). Recipes can be found in various sources such as cookbooks, websites, apps, or from other people. Recipes usually include a list of ingredients needed, how to cook, cooking time, and the number of servings. Seeing a recipe while cooking can help people to ensure that they are using the right ingredients and following the right steps in cooking. Recipes can also provide inspiration to try new dishes or variations on familiar dishes. Recipes can help those with less experience of cooking or those who want to try something new. However, there are also people who can cook without having to follow a recipe, they can create their own with available ingredients.

Many households have ingredients stored for cooking, such as basic ingredients like rice, sugar, salt, oil, and spices. This is done to ensure that the basic ingredients needed for cooking are always available and do not need to be

purchased continuously. In addition to the basic ingredients, they also stock other ingredients such as meat, fish, eggs, vegetables, and fruits. This is done to ensure that ingredients for cooking are always available and do not need to be purchased continuously. Some households also have good storage systems such as coolers, freezers, or special storage rooms to maintain the quality of the ingredients stored.

From the illustration above, it can be explained that many households determine the menu based on the availability of food ingredients they currently have. However, much information on recipes still refers to the title of the menu and the ingredients needed. This often makes households confused in determining what menu to choose and serve based on the availability of food ingredients they have. Existing food recipes have not been tested in terms of taste, texture, calories, health, or types of allergies. In addition, the availability of cooking tools and the choice of cooking methods are also often a consideration in determining the menu to be cooked. This reversed process becomes a daily problem that takes time to decide which menu to serve in a household.

To overcome these problems, this research proposes a new method in the culinary world that can make it easier for the community, namely by providing a system that can recommend menus and recipes based on the availability of kitchen ingredients and user preferences. The urgency of the availability of this system is:

- efficient use of food ingredients that are currently available so that they can be used immediately and avoid expiration, and reduce costs for purchasing new ingredients,
- streamline the utility of cooking utensils at home,
- makes it easier for users who want to serve food at home by providing complete cooking recipes accompanied by easy-to-follow guidelines,
- provide precise information regarding food ingredients that still need to be provided which are expected to be connected to the marketplace for instant ordering.

1.2 Problem Formulation

Based on the background above, the formulation of the problem is as follows.

1. How to provide a recipe recommendation system based on ingredient supply and user preference?

2. How to implement a recipe recommendation system based on ingredient supply and user preferences?

1.3 Problem Restriction

It is necessary to impose problem restrictions so that the research has a clear scope. Problem restrictions that will be applied in the Recipes Recommendation System Based on Ingredients and User Preference are:

1. Recipe data comes from Indonesian-language recipe websites, namely the websites https://www.masakapahariini.com/ and https://resepedia.id/ ,

2. Variables used to recommend recipes are limited to ingredient availability and user preferences.

1.4 Purpose

The purpose of doing a thesis with the title "**RECIPES RECOMMENDATION SYSTEM BASED ON INGREDIENTS AND USER PREFERENCE**", is as follows.

- Presenting a recipe recommended that is relevant to the availability of ingredients and user preferences.
- Presenting recipe recommendations that are relevant to similar users.

1.5. Benefit

The benefits derived from developing a recipe recommendation system based on ingredient supply and user preferences include:

- Users can find recipe recommendations that are relevant to the availability of ingredients and user preferences.
- Users can find recipe recommendations that are relevant to similar users.

CHAPTER II. THEORETICAL BASIS

2.1 Literature Study

Based on the problems in this study, there are several references from previous studies that are used to help determine the boundaries of the problem and determine the method to be used. The following table describes the research that has been done before, which is described in Table 2.1.

Table 2.1 Previous research by other author

Author	Title	Result
Shilpa Chaudhari, Aparna R., Vinay G Tekkur, Pavan G L., and Shreekanth R Karki	Ingredient/Recipe Algorithm using Web Mining and Web Scraping for Smart Chef	A recipe recommendation system using web scrapping can help users search for recipe recommendations based on ingredients entered in search keywords (Chaudhari et al., 2020).
Ifeoma Adaji, Czarina Sharmaine, Simone Debrowney, Kiemute Oyibo, and Julita Vassileva (2018)	Personality Based Recipe Recommendation Using Recipe Network Graphs	The recipe recommendation system using the recipe reviewer's personality type can help other users get recipe recommendations. Users with similar personalities tend to like the same recipes. However, the recipe category (vegan or meat-based) did not affect the reviewer's personality (Adaji et al., 2018).
Nilesh, Madhu Kumari, Pritom Hazarika, and Vishal Raman (2019)	Recommendation of Indian Cuisine Recipes Based on Ingredients	A recipe recommendation system using web scraping and content-based methods can provide recommendations for Indian dishes based on ingredients similar to other recipes (Nilesh et al., 2019).

2.2 Basic Theory

2.2.1 Machine Learning

Machine learning is one of the applications of Artificial Intelligent which allows the system to learn without being programmed directly. The machine learning process starts from the input data, then the system must look for a pattern to reach a conclusion. The system must learn these patterns and must adjust the necessary actions based on the conclusions that have been made before (Bi et al., 2019).

Machine learning has several types, including:

- Supervised learning: learning data patterns that have result values or labels.
- Unsupervised learning: studying relationships between data without referring to results or labels.
- Semi supervised learning: studying labeled and unlabeled data.
- Reinforcement learning: studying the system environment and obtaining data independently, then studying the data by trial and error.

2.2.2 Recommender System

The recommender system is one of the methods in unsupervised learning that provides item recommendations based on several predetermined factors. The recommendation system can assist system users in selecting items that suit their needs (Burke et al., 2011).

The recommender system has several methods, including:

2.2.2.1 Demographic filtering

The recommendation method by relying on filters on components, is commonly used to recommend the most popular items.

2.2.2.2 Content-based filtering

Content-based filtering is a method in a recommender system that utilizes data in items that have similarities as a reference to be recommended to users. Content-based recommendations are usually used to recommend content that is similar to other content. The recommendation system with content-based filtering is calculated using the cosine similarity formula to determine the level of similarity of a number of information obtained from users (Burke et al., 2011).

2.2.2.3 Collaborative filtering

Collaborative filtering is a method in a recommender system that utilizes transactions of an item made by users, with the aim that other users get item recommendations based on other users with similar interests. The recommendation system with collaborative filtering is calculated using the singular value decomposition formula to determine the level of similarity between users, by looking at preferences and recipe visits (Burke et al., 2011).

Collaborative filtering is divided into two methods, including:

- User-based collaborative filtering: recommend items by finding users who have similar interests to other users.
- Item based collaborative filtering: recommends similarities between items that are correlated with other items.

2.2.2.4 Hybrid recommendation system

Hybrid recommendations system is a combination of various methods in recommender system. Hybrid recommendation system is implemented to combine the advantages of each method (Burke, 2007).

2.2.3   Web Scrapping

Web scraping or web crawling is a technique of automatically extracting data from a website and storing the extracted data in a file or database (Lawson, 2015). When running code for web scraping, it will ask a request to the server and the server will send a response containing the required data. The response sent by the server will be parsed and extracted according to the data needed, according to the illustration in Figure 2.2.1.
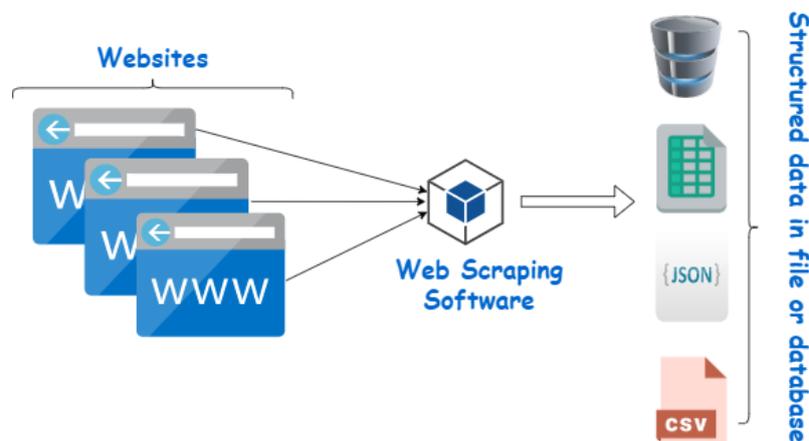

Figure 2.1 Web Scraping process

Web scraping makes the data retrieval process more optimal because some websites only display information that can be read by users and do not provide data that can be downloaded directly. Web scraping is commonly used for statistical needs, predictive systems, and training data in machine learning (Irek, 2019).

2.2.4 Python

Python is a programming language that supports several programming paradigms, such as object-oriented programming and procedural programming. Python has access to many libraries and various operating systems, such as Linux, MacOs, and Windows (Python, 2023).

Python is generally used for creating task automation systems, data analysis, and data visualization. Because it is easy to learn, Python is often used by accountants and scientists to perform certain calculations, as well as being used for basic programming lessons for beginners (Fran, 2021).

2.2.5 Pandas

Pandas is an open-source library that is used to process relational data easily. Pandas also provide various data structures, operations for manipulating numeric data, as well as processing data that is formed into an array. Pandas Python is similar to Excel, in that a dataset is a structure used by Pandas to store data (Devi, 2022).

2.2.6 MySQL

MySQL is one of the most popular Relational Database Management Systems (RDBMS) developed by Oracle using Structured Query Language (SQL) as its basis. MySQL is usually used together with Apache as a server and the PHP programming language for website-based application development. MySQL works following a client-server architectural system, where the client makes a data request in the form of a query sent to the server and the server will send a response in the form of a collection of data that matches the query (Javatpoint, NaN).

2.2.7 Django

Django is a framework for website development using the Python programming language (Django, 2022). Django uses the MVT (Model-View-

Template) architecture, which is a software design pattern for websites that has three parts (Geeks For Geeks, 2021):

- Model: to process data in the database
- View: to bridge the sending of data from the model to the template, by calling data from the model and sending it to the template
- Templates: to customize the user interface, represented by HTML/CSS/JavaScript and Jinja files

CHAPTER III. RESEARCH METHODOLOGY

3.1. Time and Place of Research

The thesis research on **COOKING RECIPE RECOMMENDATION SYSTEMS BASED ON MATERIAL SUPPLY AND USER PREFERENCES** is carried out at the State Polytechnic of Malang with hybrid method, starting from January to July 2023.

3.2. Data Collection Technique

There are some data used to run the recipe recommendation system, including recipes data with Indonesian language obtained using web scraping techniques, as well as ingredient availability data and user preference data obtained from input data obtained from the website interface. The ingredient data and preferences that have been filled in by the user will be used as parameters to provide recipe recommendations that are suitable for the user. The flow of data collection in general is illustrated in Figure 3.1.
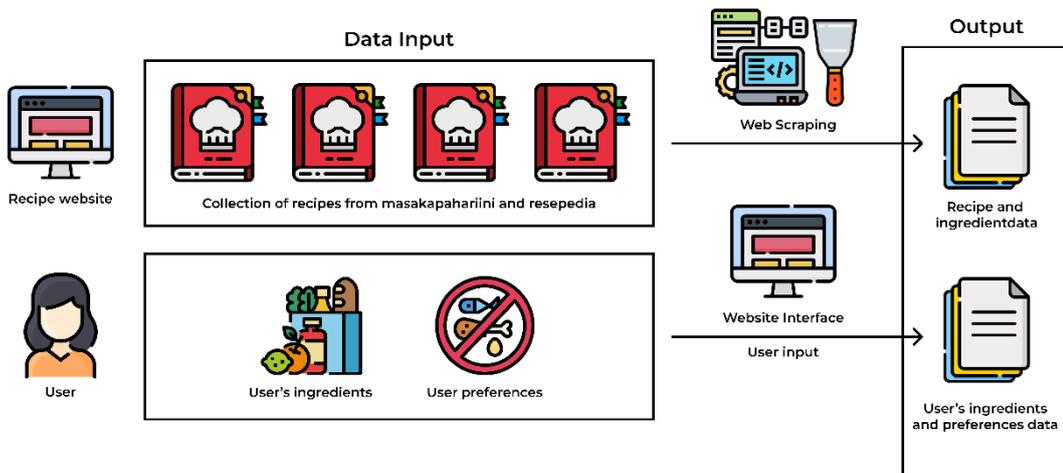

Figure 3.1 Data collection process

3.2.1 Recipe and Cooking Ingredients Data

Recipe data is used to display the steps for cooking a dish. The recipe data used in this research is Indonesian language recipe data collected from the websites https://www.masakapahariini.com/ and https://resepedia.id/ totalling 1473 recipes. Recipe data and ingredient data are collected using web scraping techniques using the Request and BeautifulSoup library from Python, with Google Colaboratory as the compiler. Collection of recipe data and ingredient data using web scraping is

broadly carried out through three stages as illustrated in Figure 3.2, namely web scraping to get recipe URLs on recipe websites, web scraping to get component data on each recipe, and defining recipe id and separating recipe data and ingredient data.
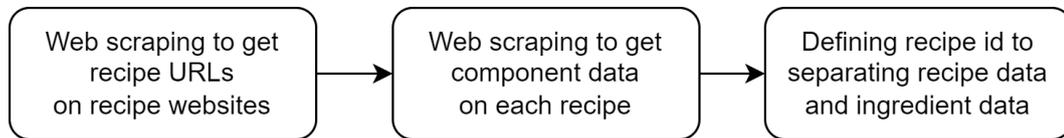

Figure 3.2 Recipe and recipe ingredient's data web scraping steps

3.2.1.1 Web scraping to get recipe URLs on recipe websites

This step is carried out to get all the URLs / links to the recipe detail pages on each website, where the URL data will be stored in an array and used as the destination page for retrieving recipe data at a next stage. The stages of web scraping to get recipe URLs on recipe websites are illustrated in Figure 3.3.

First, open a page on the recipe website and display a page containing a list of recipes. Then, look for the element that contains the URL of the recipe page. After getting these elements, open Google Colaboratory, then import BeautifulSoup and Request. Next, make a request for a page that contains a list of recipes and parse the page using BeautifulSoup. Define a variable to contain the URL, then, the data is stored in an array.
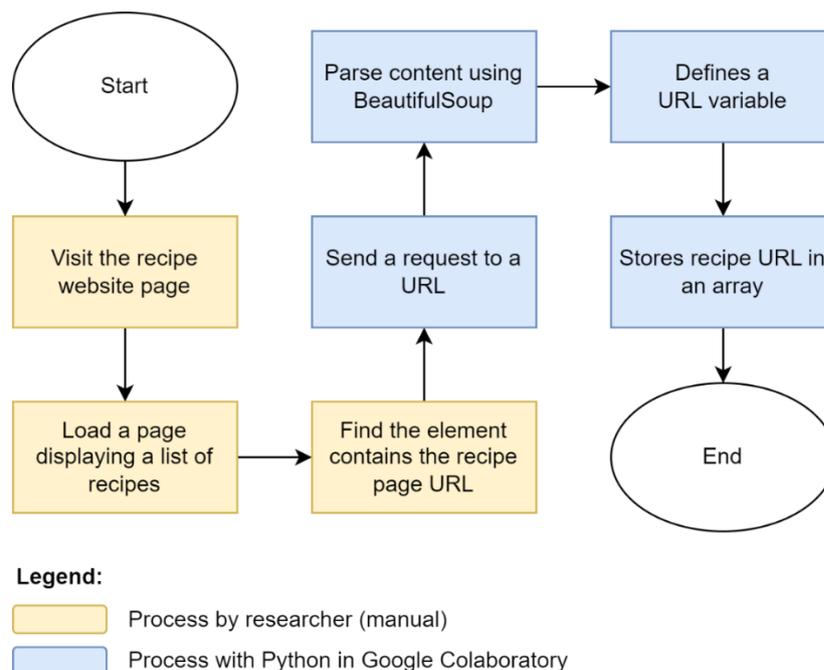
Figure 3.3 Web scraping process for getting recipe URL

Figure 3.4 shows an example of a page on a https://masakapahariini.com website that contains a list of recipes, and Figure 3.5 shows the list of recipes is wrapped in a link element in the HTML code.
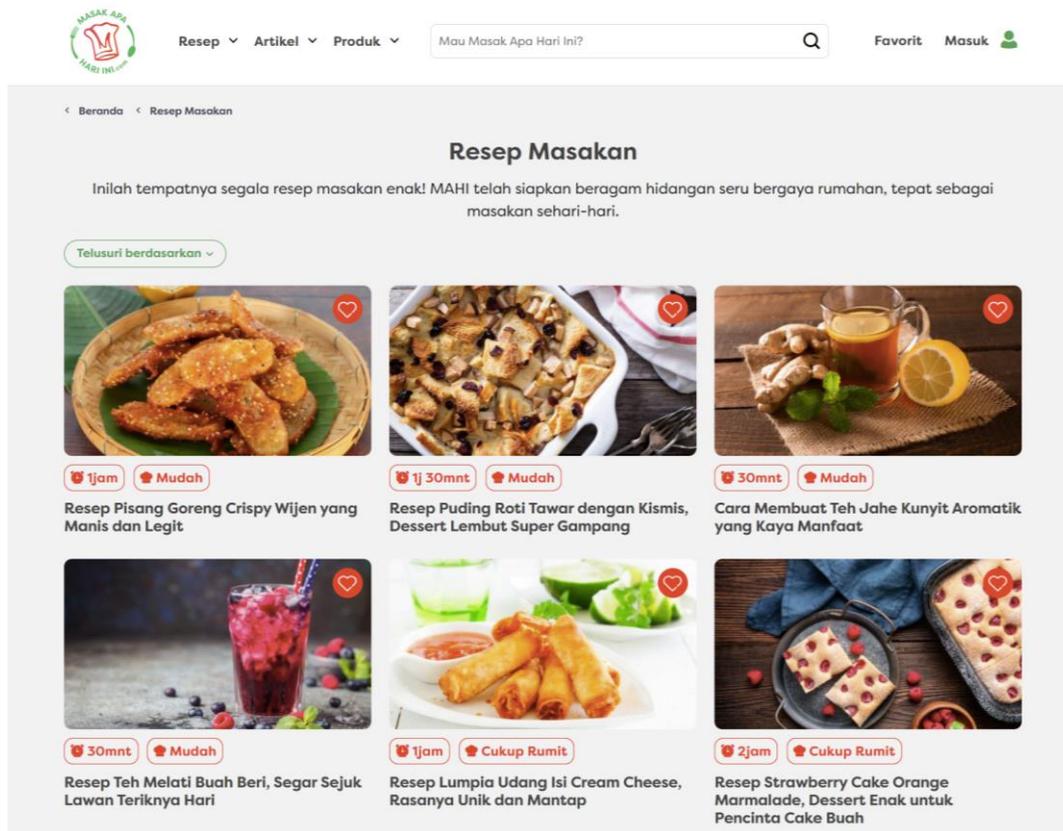

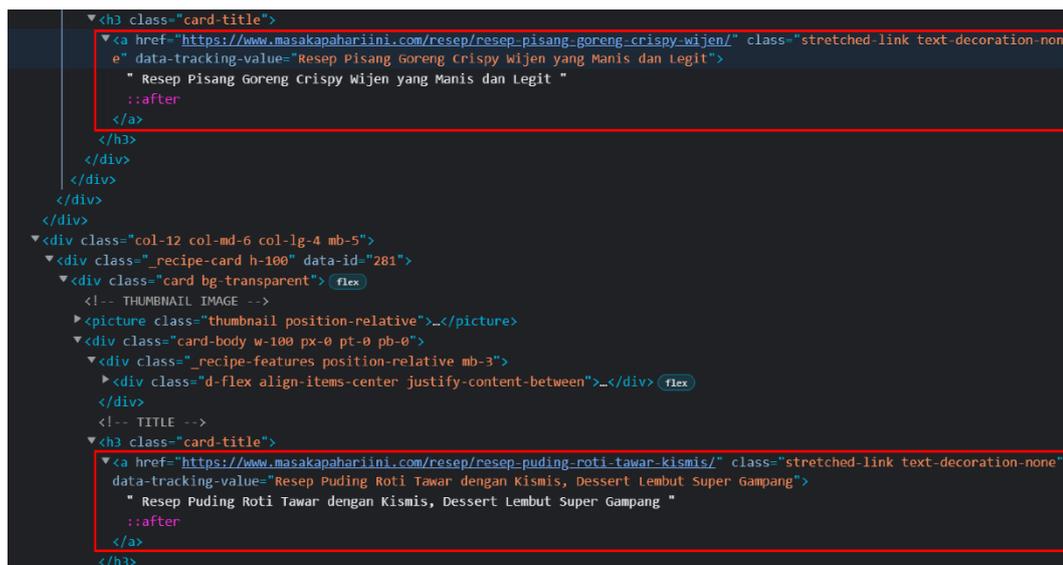Figure 3.4 Example of recipe website that contains a list of recipes

Figure 3.5 The element of recipe URL wrapped in HTML code

3.2.1.2 Web scraping to get component data on each recipe

This stage is carried out to get the data components for each recipe, where the data will be stored in a DataFrame to be processed at a next stage. Some of the component of recipe data that will be used include:

1. Title, to display the title of the recipe on the website,
2. Source URL, to display where the recipe was obtained,
3. Image URL, to display the recipe food image on the website,
4. Calories, to display the number of calories produced by the dish,
5. Servings, to display the number of portions provided by the website,
6. Time, to display the time needed to cook the food according to the recipe,
7. Difficulty level, to display the difficulty level of cooking according to the recipe,
8. Ingredients, to display a list of ingredients needed to cook according to the recipe,
9. Instructions, to display the steps that must be taken when cooking according to the recipe,
10. Rating, to display the eligibility value of the recipe to be tried by other users.

Figure 3.6 illustrates the web scraping process for getting component of recipe, where the recipe URL is loaded at the beginning of the stage. After the recipe URL is loaded, one of the recipe detail pages is opened to get the element / class that contains each component. After finding these elements, BeautifulSoup and Request are imported into the Google Colaboratory for web scraping. The web scraping stage needs to be done on each recipe URL in order to get all the recipe data, starting with making a request to one of the recipe URLs, then parsing the page using BeautifulSoup. After the page is successfully loaded, each data component must be defined, and the data stored into the recipe's DataFrame. This DataFrame will be used for the stages of separating recipe data from ingredient data.
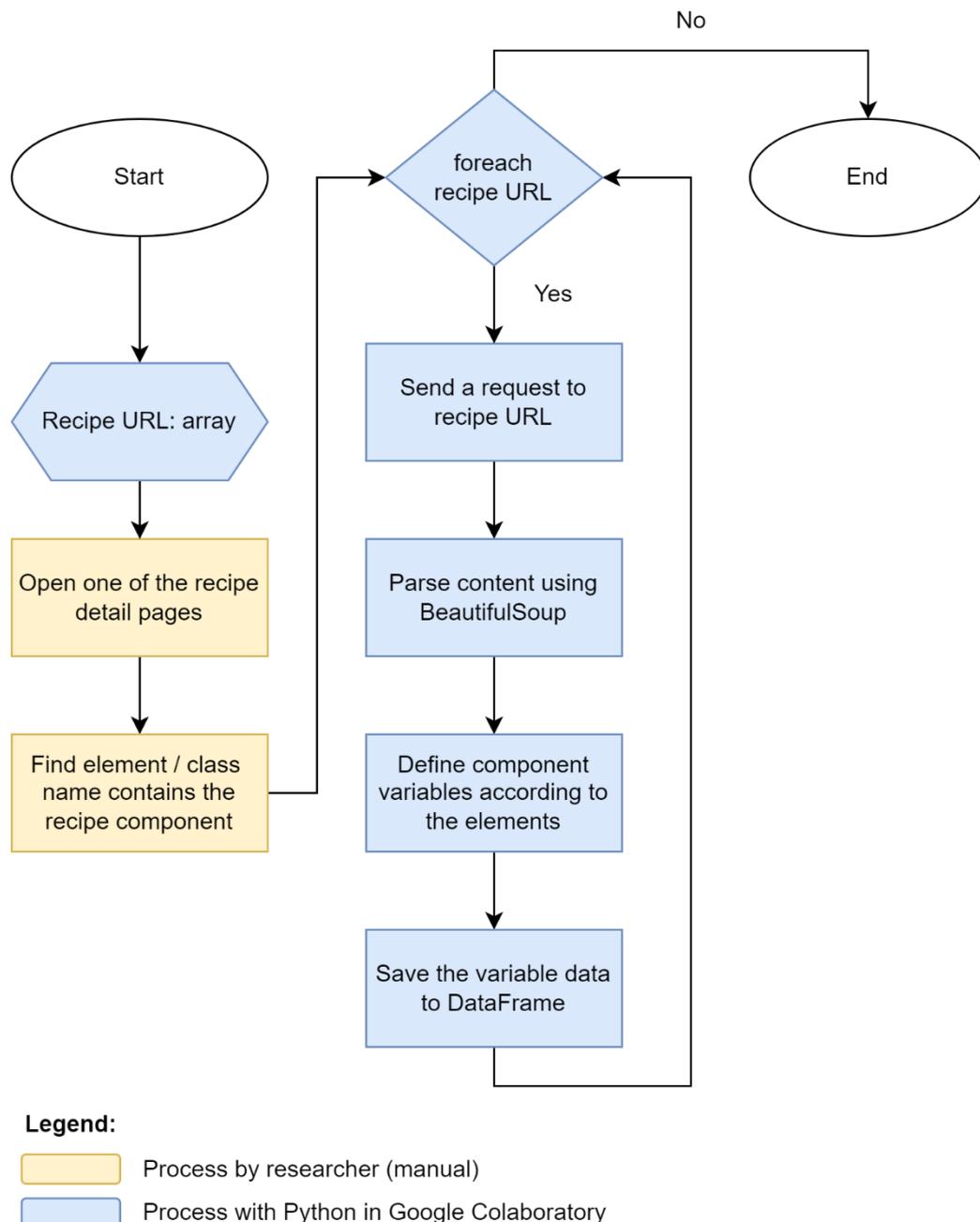
Figure 3.6 Web scraping process for getting component of recipe

Figure 3.7 contains an example of a page on the website https://masakapahariini.com that contains detailed recipes, and Figure 3.8 shows the ingredient wrapped in a div element at class "item" in HTML code.

Figure 3.7 Example of recipe detail on https://masakapahariini.com

```
▼<div class="_recipe-ingredients ">
  ▶<header class="_section-title d-flex align-items-center ">…</header> flex
  ▶<div class="portions d-inline-flex align-itens-center text-white p-2 ps-4 mb-5">…</div> flex
  ▼<div class="d-flex"> flex
      <div class="part fw-bold me-3" data-value="125" data-base-measurement-unit="g " data-base-quantity="500"> 500 </div>
      <div class="item"> g kikil yang sudah dibersihkan </div> flex
    </div>
  ▼<div class="d-flex"> flex
      <div class="part fw-bold me-3" data-value="0.38" data-base-measurement-unit="l " data-base-quantity="1.5"> 1.5 </div>
      <div class="item"> l air, untuk merebus </div> flex
    </div>
  ▼<div class="d-flex"> flex
      <div class="part fw-bold me-3" data-value="0.5" data-base-measurement-unit="lembar " data-base-quantity="2"> 2 </div>
      <div class="item"> lembar daun salam </div> flex
    </div>
  ▼<div class="d-flex"> flex
      <div class="part fw-bold me-3" data-value="0.75" data-base-measurement-unit="cm " data-base-quantity="3"> 3 </div>
      <div class="item"> cm lengkuas, memarkan </div> flex
    </div>
```

Figure 3.8 The element of recipe detail wrapped in HTML code

3.2.1.3 Defining recipe id and separating recipe data and ingredient data

This step is carried out to separate the recipe data and ingredient data, with the aim that the recipe data is not duplicated because the ingredient data is still stored in the recipe data and the ingredient data can be processed without disturbing the recipe data. These stages are illustrated in Figure 3.9, starting with adding a random recipe id with a UUID number code using Python's Random and UUID library. Then, create a new DataFrame to store the ingredients data along with the recipe id. Next, create a new DataFrame to store the recipe data that contains all recipe components except ingredients. Once the recipe DataFrame is created, remove duplicate data from the recipe. Recipe data and ingredient data can be saved to a file in .csv format to be entered into a database or processed again at a next stage.

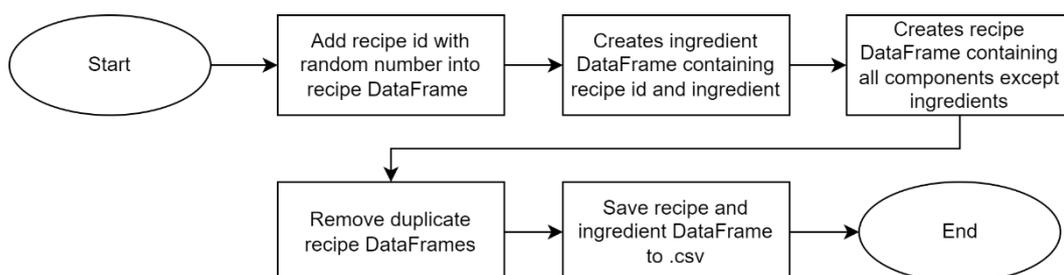Figure 3.9 Define recipe id and separating recipe data and ingredient data process

Figure 3.10 contains an example of recipe data before separating data, where recipe data is still duplicated because each recipe data contains different ingredients in same recipe. Figure 3.11 contains an example of recipe data and ingredient data after data separation, where recipe data and ingredient data are interconnected via recipe id.

title	source_url	img_url	servings	time	difficulty	ingredient	instructions
Resep Ikan Pecah Kulit Bakar, Unik dan Tak Kalah Lezatnya	https://www.masakapahariini.com/resep/resep-ikan-pecah-kulit-bakar/	https://www.masakapahariini.com/wp-content/uploads/2019/11/resep-ikan-pecah-kulit-bakar-620x440.jpg	4 Porsi	1jam	Cukup rumit	750 g ikan pecah kulit, bersihkan	Lumuri ikan dengan air jeruk nipis dan garam....
Resep Ikan Pecah Kulit Bakar, Unik dan Tak Kalah Lezatnya	https://www.masakapahariini.com/resep/resep-ikan-pecah-kulit-bakar/	https://www.masakapahariini.com/wp-content/uploads/2019/11/resep-ikan-pecah-kulit-bakar-620x440.jpg	5 Porsi	1jam	Cukup rumit	2 sdm air jeruk nipis	Lumuri ikan dengan air jeruk nipis dan garam....
Resep Ikan Pecah Kulit Bakar, Unik dan Tak Kalah Lezatnya	https://www.masakapahariini.com/resep/resep-ikan-pecah-kulit-bakar/	https://www.masakapahariini.com/wp-content/uploads/2019/11/resep-ikan-pecah-kulit-bakar-620x440.jpg	6 Porsi	1jam	Cukup rumit	1 sdt garam	Lumuri ikan dengan air jeruk nipis dan garam....
Resep Tumis Buncis Daging Giling, Nikmat Tersaji Sebagai Hidangan Pendamping	https://www.masakapahariini.com/resep/resep-tumis-buncis-daging-giling/	https://www.masakapahariini.com/wp-content/uploads/2019/04/tumis-buncis-daging-giling-780x440.jpg	7 Porsi	30mnt	Mudah	1 ikat buncis, potong memanjang selebar 2.5 cm	Tumis bawang putih dan cabai hingga wangi...
Resep Tumis Buncis Daging Giling, Nikmat Tersaji Sebagai Hidangan Pendamping	https://www.masakapahariini.com/resep/resep-tumis-buncis-daging-giling/	https://www.masakapahariini.com/wp-content/uploads/2019/04/tumis-buncis-daging-giling-780x440.jpg	8 Porsi	30mnt	Mudah	200 g daging sapi giling	Tumis bawang putih dan cabai hingga wangi...
Resep Tumis Buncis Daging Giling, Nikmat Tersaji Sebagai Hidangan Pendamping	https://www.masakapahariini.com/resep/resep-tumis-buncis-daging-giling/	https://www.masakapahariini.com/wp-content/uploads/2019/04/tumis-buncis-daging-giling-780x440.jpg	9 Porsi	30mnt	Mudah	2 siung bawang putih	Tumis bawang putih dan cabai hingga wangi...

Figure 3.10 Example of recipe data before separating the ingredient with recipe

Recipe Data

id	title	source_url	img_url	servings	time	difficulty	ingredient	instructions
10370393470	Resep Ikan Pecah Kulit Bakar, Unik dan Tak Kalah Lezatnya	https://www.masakapahariini.com/resep/resep-ikan-pecah-kulit-bakar/	https://www.masakapahariini.com/wp-content/uploads/2019/11/resep-ikan-pecah-kulit-bakar-620x440.jpg	4 Porsi	1jam	Cukup rumit	750 g ikan pecah kulit, bersihkan	Lumuri ikan dengan air jeruk nipis dan garam....
10523610475	Resep Tumis Buncis Daging Giling, Nikmat Tersaji Sebagai Hidangan Pendamping	https://www.masakapahariini.com/resep/resep-tumis-buncis-daging-giling/	https://www.masakapahariini.com/wp-content/uploads/2019/04/tumis-buncis-daging-giling-780x440.jpg	7 Porsi	30mnt	Mudah	1 ikat buncis, potong memanjang selebar 2.5 cm	Tumis bawang putih dan cabai hingga wangi...

Ingredient Data

id	ingredient
10370393470	750 g ikan pecah kulit, bersihkan
10370393470	2 sdm air jeruk nipis
10370393470	1 sdt garam
10523610475	1 ikat buncis, potong memanjang selebar 2.5 cm
10523610475	200 g daging sapi giling
10523610475	2 siung bawang putih

Figure 3.11 Example of recipe and ingredient data after define recipe id and separating the ingredient with recipe

3.2.2 Ingredient Availability Data

Ingredient availability data is material data that is owned and stored by the user. Ingredient availability data will be used as one of the parameters in determining cooking recommendations, to display the percentage of ingredients that are owned in each recipe and display a list of ingredients that are owned or not owned in a recipe. Ingredient availability data is used as a parameter for determining food recommendations because users tend to choose dishes that require pre-available ingredients (Simmons & Chapman, 2012).

Ingredient availability data is obtained from input data entered by the user on the website, where the user is required to register an account to store material availability data permanently in the MySQL database. Ingredient availability data can be added or deleted according to the availability of materials owned by the user.

The following is an example of material availability data stored in the database, which is illustrated in Figure 3.12.

User Table

id	username
1	Andi
2	Budi

UserIngredient Table

id	user_id	ingredient
1	1	Garam
2	1	Jeruk nipis
3	2	Air
4	2	Minyak

Figure 3.12 Example of user that have ingredients stored in database

3.2.3 User Preference Data

User preference data is data that contains specific preferences for each user, where the preferences of each user is different. User preference data will be used as one of the parameters in determining food recommendations, to determine the priority of food recommendations according to the preferences selected and to determine similarities between users.

Based on research conducted by Yu Chen, Ananya Subburathinam, Ching-Hua Chen, and Mohammed J. Zaki, user personalization using allergies is important so that users can avoid foods that contain allergens. In addition, the nutrition in food also needs to be displayed so that users can also measure the health level of the food to be cooked (Chen et al., 2021). Roehr et al defined some of the most common allergies people experience, such as chicken eggs, milk, peanuts, chocolate, and fish (Roehr et al., 2004). In addition, the number of Indonesian citizens who are Muslim reaches 237.56 million (Rizaty, 2022). Thus, the several studies are used as a reference in determining user preference parameters. The categories of user preferences that will be used in the recipe recommendation system include:

1. Food restrictions: halal, diet.
2. Allergies: chicken eggs, milk, peanuts, chocolate, and fish.

User preference data is obtained from input data entered by the user on the website, where the user is required to register an account to store user preference data permanently in the MySQL database. User preference data can be changed according to user conditions. The following is an example of user preference data stored in the database, which is illustrated in Figure 3.13.

Preferences Table

id	category	name
1	Restriction	Halal
2	Restriction	Diet
3	Allergy	Chicken
4	Allergy	Egg
5	Allergy	Milk
6	Allergy	Peanut
7	Allergy	Chocolate
8	Allergy	Fish

User Table

id	username
1	Andi
2	Budi

UserPreferences

id	user_id	preference_id
1	1	1
2	1	2
3	2	1
4	2	8

Figure 3.13 Example of user preferences stored in database

3.3 Data Processing Technique

3.3.1 Recipe Data Pre-processing

Before the recipe data is entered into the database for processing, it is necessary to carry out a data cleaning process, to improve the data format as needed. Determining the required format in data cleaning is done by looking at the writing patterns on each data manually, then determining the actions that need to be taken so that the data format is aligned with system requirements. Some of the data that needs to be done data cleaning include:

1. Title: delete unnecessary words so the system can recommend dishes according to the title keywords

2. Time: change the time format which was originally in the form of a String to an integer, so that the estimated cooking time can be processed

3. Portion: change the portion format which was originally in the form of a String to an integer, where the integer format makes it easier for the system to process data

4. Level of difficulty: change the sentences in the category of difficulty level into categories that are easy to understand, to make it easier for the system to sort the level of difficulty in recipes.

The data cleaning process is carried out using the Pandas library from Python and Google Colaboratory as a compiler, where the results of this data cleaning will be saved in .csv format. The data cleaning process for cooking recipe data is illustrated in Figure 3.14.
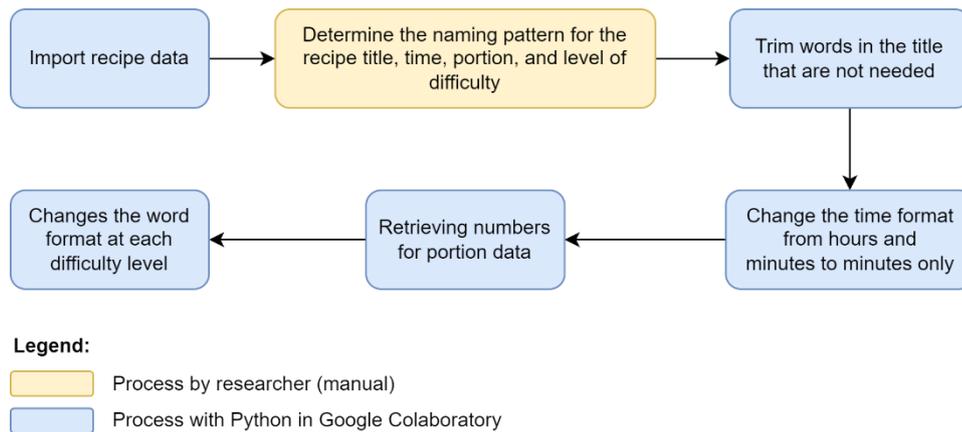
Figure 3.14 Data cleaning process for recipe data

Figure 3.15 illustrates an example of recipe data before data cleaning and after data cleaning. In the initial data, some data such as title, time, portion, and level of difficulty have different formats, making it difficult to process the data if is not corrected. Based on the example, some of the things that need to be fixed include:

- Title: delete the word "Recipe" at the beginning of the text and delete the sentence after the ","
- Time: change the format which was originally in the form of sentences into minutes in integer units
- Portion: delete the word "portion" and change the data format to integer
- Difficulty level: change the sentence "Difficult enough" to be difficult

After determining the data repair pattern, data cleaning is executed using the previously defined patterns, so the results will look like Figure 3.15.

Recipe data before data cleaning

Title	image_url	Time	Servings	Difficulty	Instructions	source_url
Resep Ayam Bakar Bumbu Rujak	https://www.masakapahariini.com/wp-content/upl...	1jam 15mnt	4 Porsi	Mudah	Bumbui ayam dengan sendok, ratakan.;Goreng hingga matang.;Sajikan.	https://www.masakapahariini.com/resep/resep-ayam-bakar-bumbu-rujak/
Resep Martabak Mie Telur Kornet, Enak dan Praktis	https://www.masakapahariini.com/wp-content/upl...	25mnt	4 Porsi	Cukup sulit	Campurkan bahan ke dalam mangkok.;Goreng hingga matang.;Tiriskan.	https://www.masakapahariini.com/resep/resep-martabak-mie-telur/
Bubur Ayam Hainan, Resep Makanan Rumahan yang Enak dan Praktis	https://www.masakapahariini.com/wp-content/upl...	45mnt	4 Porsi	Mudah	Masukkan bumbu ke panci.;Masak nasi hingga menjadi bubur.; Sajikan.	https://www.masakapahariini.com/resep/resep-bubur-ayam-hainan/

Recipe data after data cleaning

Title	image_url	Time	Servings	Difficulty	Instructions	source_url
Ayam Bakar Bumbu Rujak	https://www.masakapahariini.com/wp-content/upl...	75	4	Mudah	Bumbui ayam dengan sendok, ratakan.;Goreng hingga matang.;Sajikan.	https://www.masakapahariini.com/resep/resep-ayam-bakar-bumbu-rujak/
Martabak Mie Telur Kornet	https://www.masakapahariini.com/wp-content/upl...	25	4	Sulit	Campurkan bahan ke dalam mangkok.;Goreng hingga matang.;Tiriskan.	https://www.masakapahariini.com/resep/resep-martabak-mie-telur/
Bubur Ayam Hainan	https://www.masakapahariini.com/wp-content/upl...	45	4	Mudah	Masukkan bumbu ke panci.;Masak nasi hingga menjadi bubur.; Sajikan.	https://www.masakapahariini.com/resep/resep-bubur-ayam-hainan/

Figure 3.15 Example of recipe data before and after data cleaning

3.3.2 Recipe's Ingredients Data Processing

Before the ingredient data for the recipe is entered into the database, it is necessary to do data splitting. The data splitting process on recipe ingredients data aims to simplify data processing. Determination of the pattern of data separation is done by looking at the pattern of writing materials manually, and the data separation is determined based on that pattern. Some of the data needed for data splitting include:

1. Material name: to display the material name
2. Quantity: to display the amount of material needed
3. Unit: to display the unit of material used
4. State: to display the treatment of ingredients before cooking

The data splitting process is carried out using the Python and Pandas programming languages, where the processed recipe ingredient data will be stored in a .csv file to be imported into the database. The data splitting process for ingredient data is illustrated in Figure 3.16. After importing the data, it is necessary to perform a pattern search for quantity, ingredient name, unit, and state. Next, split the data for each column and save the results to the DataFrame. After all data splitting is done, the DataFrame is exported to .csv.
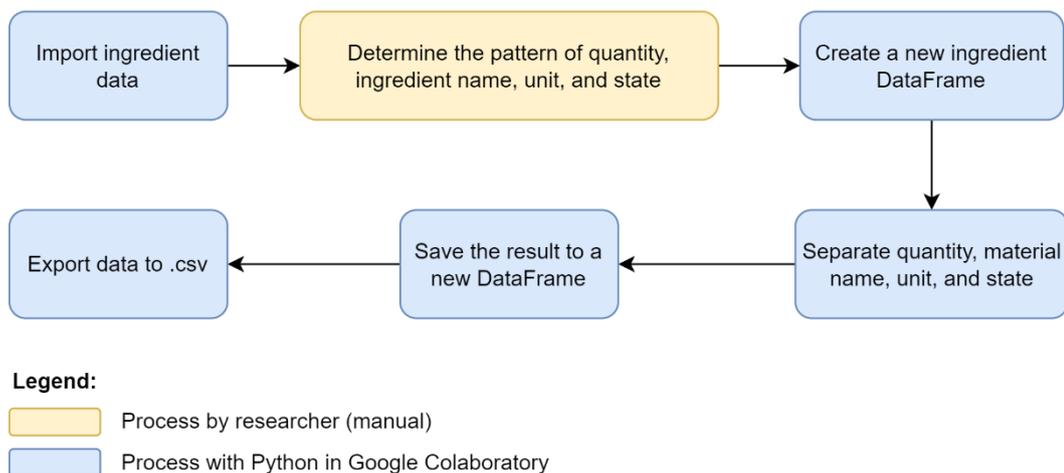
Figure 3.16 Data splitting process

The data splitting example is illustrated in Figure 3.17, where the material data which was originally in the form of a sentence is split into components of material name, quantity, unit, and state. Based on the example, the pattern of data splitting are:

- Quantity: get the first word of the ingredient phrase
- Unit: get the second word of the ingredient phrase
- Name: get the third word until the word before the comma
- State: get all word after commas

Ingredients

50 gram gula, haluskan
50 ml air
1 kg dada ayam

Components

| Name | Quantity | Unit | State |

Results

Ingredient Phrase	Name	Quantity	Unit	State
50 gram gula, haluskan	Gula	50	gram	haluskan
50 ml air	Air	50	ml	
1 kg dada ayam	Dada ayam	1	kg	rebus
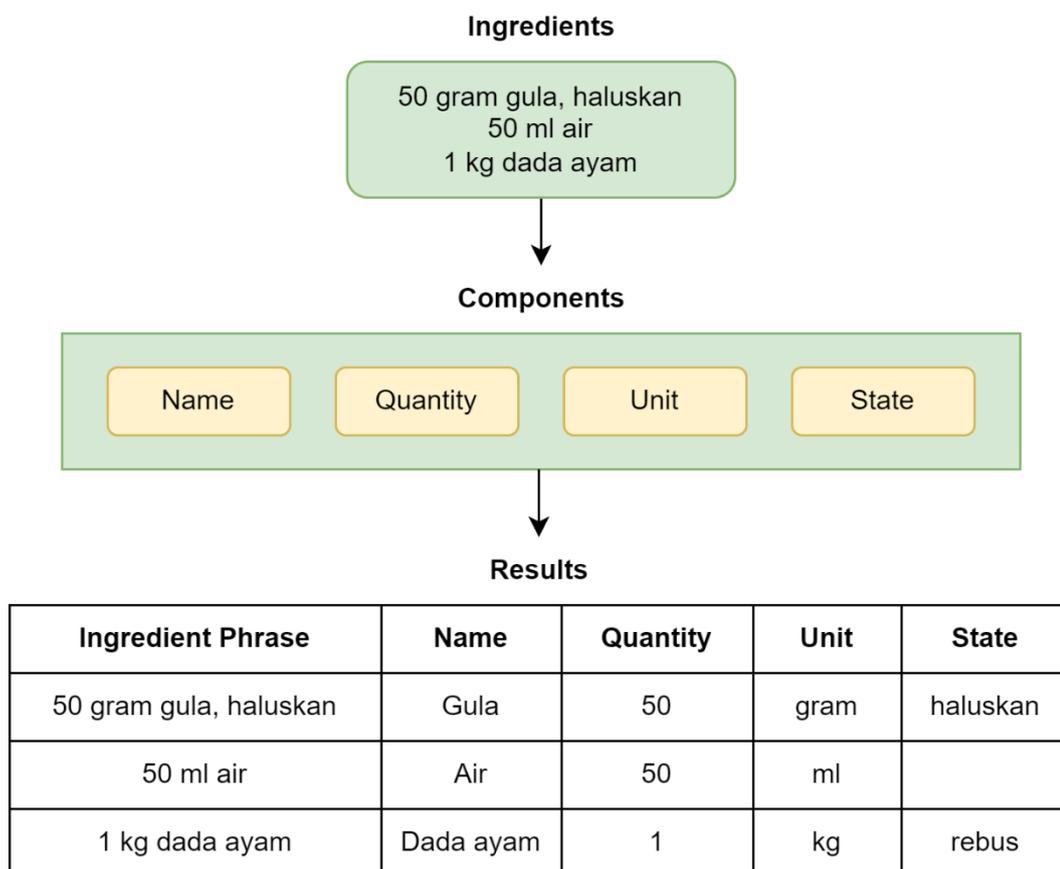
Figure 3.17 Example of ingredient data before and after data splitting

3.3.3 Recipe Recommendation System

The recipe recommendation system in this study aims to provide food recommendations to users based on ingredients and user preferences. This recipe recommendation system uses the recommender system method in Machine Learning, which is used to recommend a list of recipes according to certain criteria. The recommender system methods that will be used include demographic filtering, content-based filtering, and collaborative filtering, which depend on certain features, as illustrated in Figure 3.18.
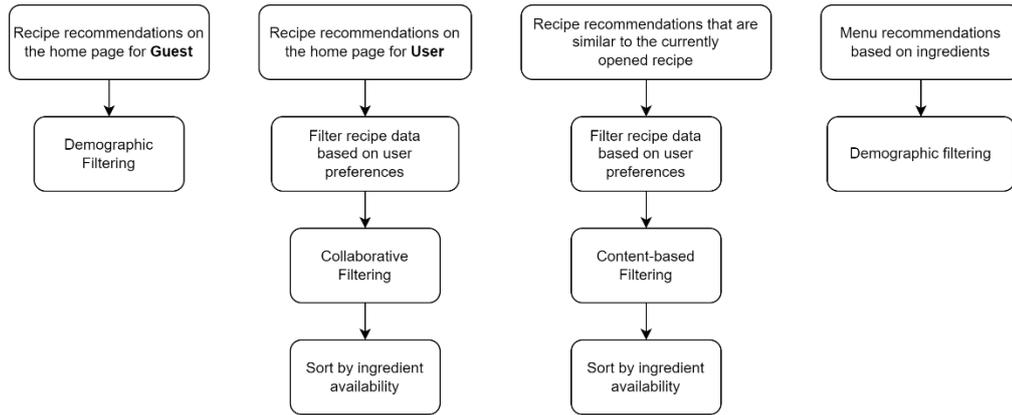
Figure 3.18 The division of the recommender system method on several features

3.3.3.1 Recipe recommendation on the home page for Guest and User

      Recipe recommendations on the home page for Guests are used to display recipe recommendations for Guests who do not have ingredient data and user preferences data stored in the database. Recipe recommendations on the home page for User are used to display recipe recommendations for users who have ingredient data and user preferences data stored in the database. This stage is carried out by demographic filtering which is illustrated in Figure 3.19.
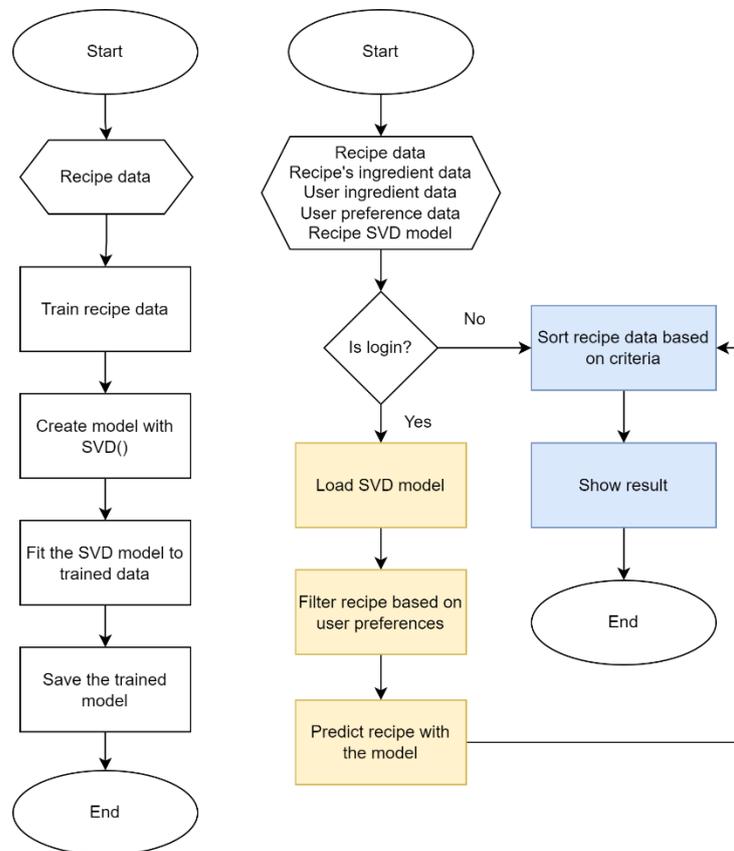
Figure 3.19 Recipe recommendation on home page

3.3.3.2 Recipe recommendation that is similar to the recipe opened

Recipe recommendations on the recipe detail page that is being opened serves to display recommendations similar to the recipe being opened, starting from the name of the dish, calories, portion, cooking time, and ingredients used. This stage is carried out with content-based filtering which is illustrated in Figure 3.20.
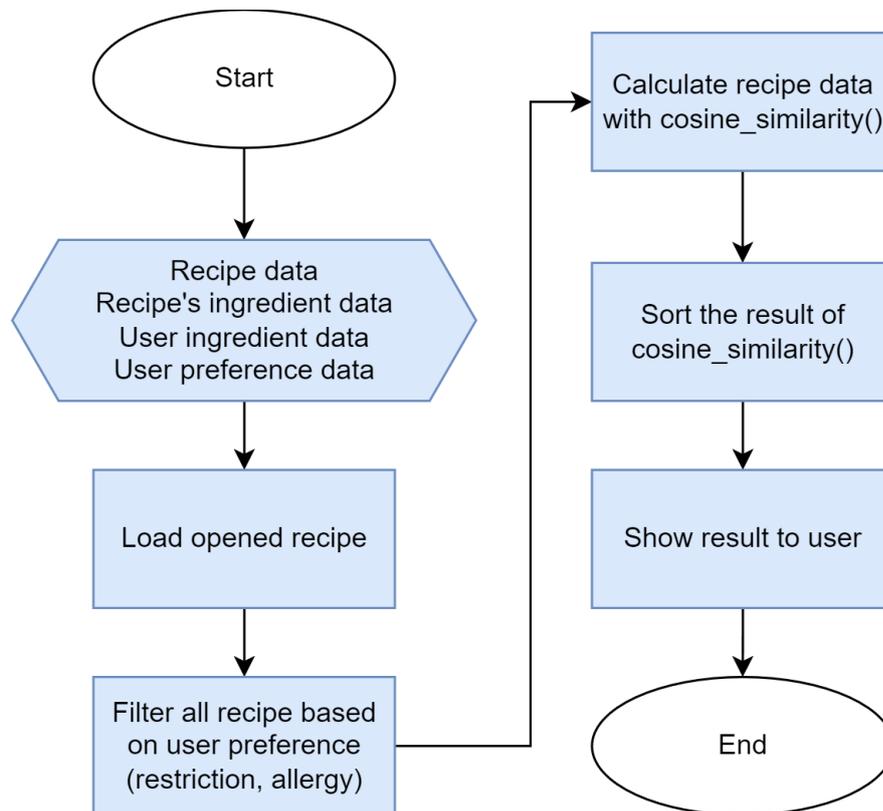
Figure 3.20 Recipe recommendation process based on opened recipe

3.3.3.3 Menu recommendations based on ingredients

Recipe recommendations on the ingredient filling page function to display recipe recommendations based on the number of ingredients and user preferences, where user preferences will filter the dishes that are most suitable for the user. Calculation of the percentage of ingredient availability using the formula:

$$percentage\ of\ user\ ingredient = \frac{count(user\ ingredient)}{count(recipe\ ingredient)} \times 100\%$$

The stages of recipe recommendations based on ingredients is carried out with demographic filtering, which is illustrated in Figure 3.21.
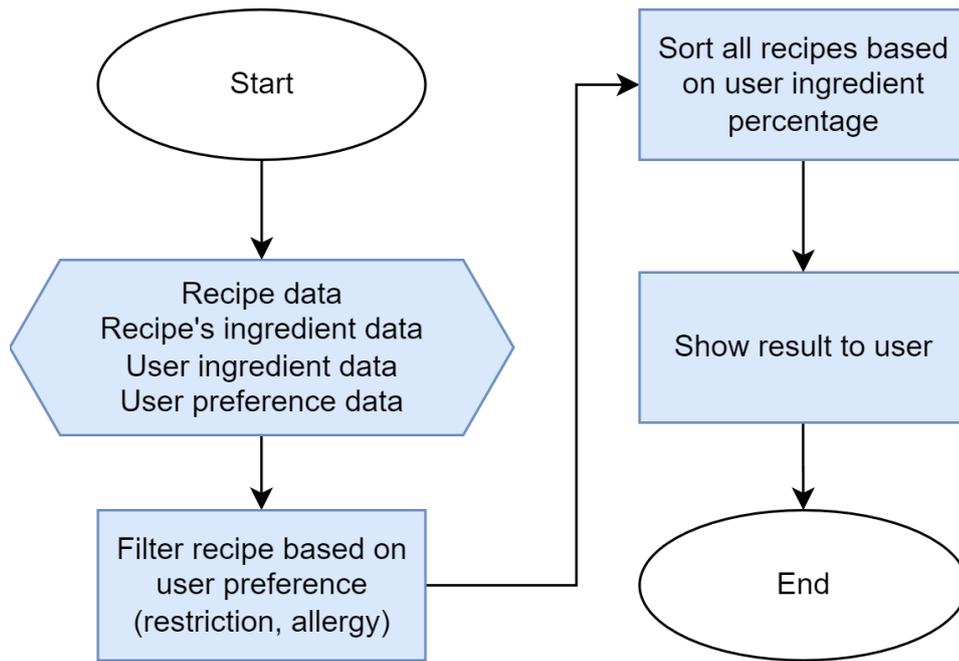
Figure 3.21 Recipe recommendation on fill ingredient page

3.4 System Design

3.4.1 System Development Process

There are several stages of developing a food recommendation system, from data collection to data testing, which is illustrated in Figure 3.22. The system development process starts from collecting recipe data and recipe ingredient data using web scraping. Then, data cleaning is performed for recipe data and data splitting for recipe ingredient data. After the data is obtained, a user interface design is carried out for the website display and a system database design to accommodate the data to be processed. After designing the interface and database, a website-based system was developed so that users can access the system via a computer or smartphone. Website development is carried out simultaneously with the development of a recipe recommendation algorithm. After the website is developed, the testing phase is carried out to determine the level of accuracy of recipe recommendations, namely by carrying out functional system testing using the Selenium IDE as well as direct experiments through input from users.
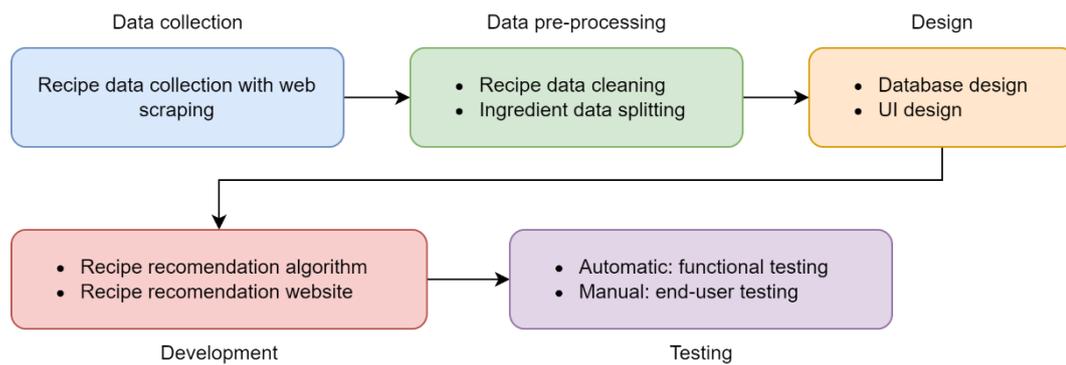
Figure 3.22 System development process for recipe recommendation system

3.4.2   System Architecture

A recipe recommendation system based on material availability and user preferences will be developed using Python programming language and use the Django framework to create website-based user interfaces, with MySQL as the data storage in the database. The system architecture is illustrated in Figure 3.23.
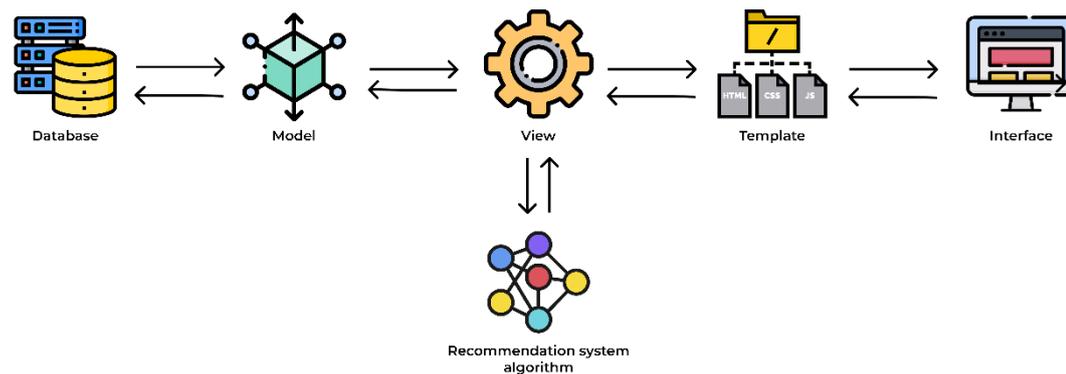
Figure 3.23 System architecture of recipe recommendation system

3.4.3   Use Case Diagram

Use case diagrams are used to define the features that will be used in the system. Users or recipe recommendation system actors are recipe seekers, namely the user who accesses all features on the system, including searching for recipes or accessing recipe details. Recipe seekers can register, log in, access the home page, access recipe details, and fill in the availability of ingredients and user preferences. The description of the features using a use case diagram is illustrated in Figure 3.24.
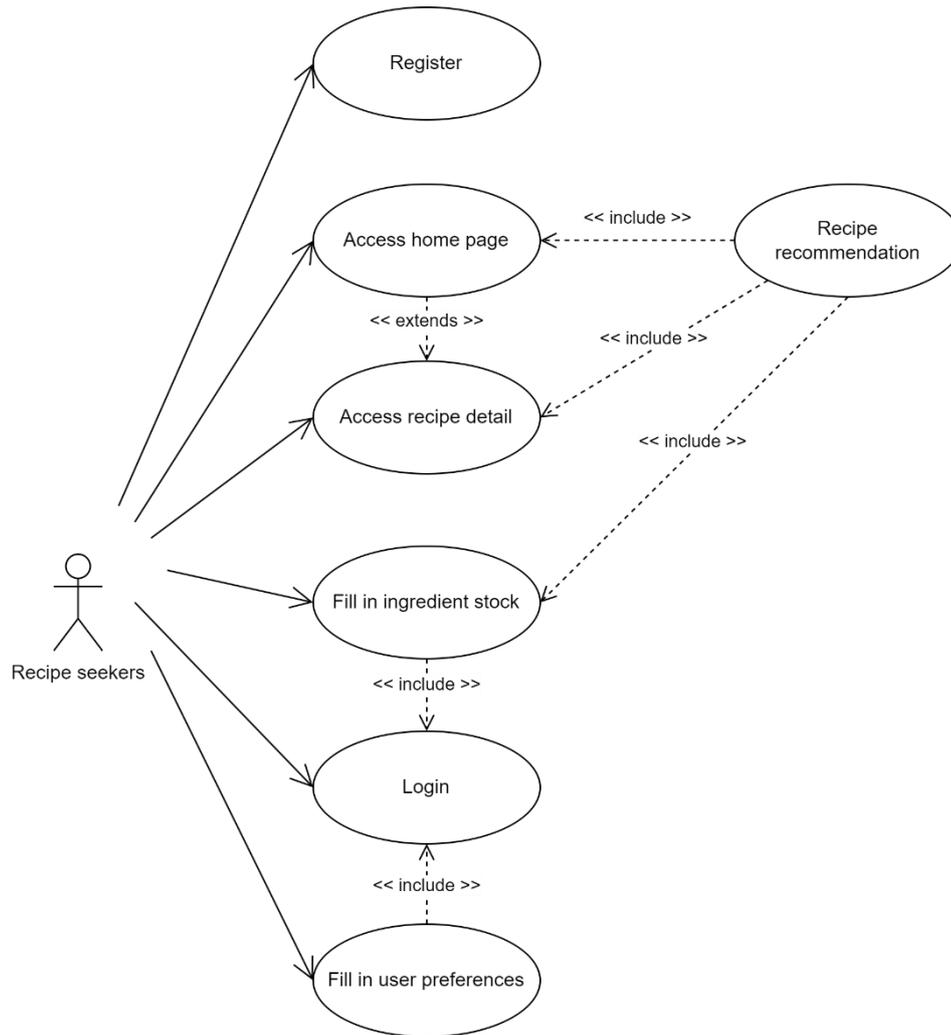
Figure 3.24 Use case diagram of recipe recommendation system

3.4.3.1 Actor Definition

Table 3.1 describes the roles of actors in the recipe recommendation system based on ingredient availability and user preference.

Table 3.1 Actor definition table

Actor	Definition
Recipe seekers	The user who accesses all features on the system, including searching for recipes or accessing recipe details. Recipe seekers can register, log in, access the home page, access recipe details, and fill in the availability of ingredients and user preferences.

3.4.3.2 Use Case Definition

Table 3.2 describes the definition of each use case in the recipe recommendation system based on the availability of ingredients and user preferences.

Table 3. 2 Use case definition table

Use Case	Definition
Login	Displays the login page so that the user can change the actor's status, starting from Guest to User
Register	Page to create an account
Access home page	Pages that display recipe recommendations based on certain criteria, such as the most popular recipe or the most popular recipe
Access recipe details	Pages that display recipe detail, such as title of recipe, ingredients, instruction, and related recipe
Recipe recommendation	The process of recommending recipes that appear on the home page, recipe detail page, and ingredient filling page
Fill in ingredient stock	Page to fill in the availability of user's ingredients
Fill in user preferences	Page to set user preferences

3.5 System Testing

3.5.1 Website Functional Testing

Website functional testing is carried out to ensure that the website can run according to its function. Website functional testing is carried out using Selenium IDE, which is a automated testing application to test website.

3.5.2 Recipe Recommendation System Algorithm Testing

Testing the recipe recommendation system algorithm is carried out to ensure that the system can display recipe recommendations according to user interests and profiles. Testing the recipe recommendation system algorithm is carried out by involving users directly, where users are asked to fill in a list of ingredients they have and user preferences. After the system displays recipe recommendations, the user is asked to choose the recipe that best suits the user. The results of tests involving users will be displayed in statistical form to find out how accurate the recipe recommendation algorithm is for users.

CHAPTER IV. RESEARCH SCHEDULE

The following is a research schedule for RECIPE RECOMMENDATION SYSTEMS BASED ON INGREDIENTS AND USER PREFERENCES, starting from topic submission to completion.

No	Activity	December				January				February				March				April				May				June				July			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	**Title submission**																																
	Idea submission		■	■																													
	Supervisor submission			■	■																												
2	**Study of literature**																																
	Studying the literature according to the topic					■	■	■																									
	Studying system concepts					■	■	■																									
	Studying the method used in the topic					■	■	■																									
3	**Pre proposal**																																
	Proposal guidance					■	■	■	■																								
	Proposal seminars								■																								
4	**Analysis**																																
	Data collection									■	■																						
	Data processing										■	■	■																				
5	**System design**																																
	Database design													■	■	■	■																
	Interface design													■	■	■	■																
6	**System implementation**																																
	Algorithm development																		■	■	■	■	■										
	Website development																	■	■	■	■	■	■	■									
7	**System Testing**																																
	Functional system testing																										■	■	■				
	User testing																									■	■	■					
8	**Results and Conclusions**																																
	Create a report document																												■	■	■	■	
	Thesis seminars																																■

BIBLIOGRAPHY

Adaji, I., Sharmaine, C., Debrowney, S., Oyibo, K., & Vassileva, J. (2018). Personality Based Recipe Recommendation Using Recipe Network Graphs. *SCSM 2018: Social Computing and Social Media. Technologies and Analytics* (hal. 161-170). Springer International Publishing.

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology, 188*(12), 2222-2239.

Cambridge. (t.thn.). *Recipe*. Dipetik January 08, 2023, dari Cambridge Dictionary: https://dictionary.cambridge.org/dictionary/english/recipe

Caudhari, S., R., A., Tekkur, V. G., Pavan, G. L., & Karki, S. R. (2020). Ingredient/Recipe Algorithm using Web Mining and Web Scraping for Smart Chef. *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* (hal. 1-4). Bangalore: IEEE.

Devi, N. (2022). *What is Pandas in Python?* Dipetik January 09, 2023, dari Scaler: https://www.scaler.com/topics/pandas/what-is-pandas-in-python/

Django. (2022). *Django Documentation FAQ: General*. Dipetik January 08, 2023, dari Django: https://docs.djangoproject.com/en/4.1/faq/general/

Eufic. (2010). *Why Do We Cook Our Food and What Happens When We Do?* Retrieved January 08, 2023, from Eufic: https://www.eufic.org/en/food-safety/article/the-why-how-and-consequences-of-cooking-our-food

Fran. (2021). *What is Python used for? 10 practical Python uses*. Dipetik January 15, 2023, dari FutureLearn: https://www.futurelearn.com/info/blog/what-is-python-used-for

Geeks For Geeks. (2021). *Django Project MVT Structure*. Dipetik January 15, 2023, dari Geeks For Geeks: https://www.geeksforgeeks.org/django-project-mvt-structure/

Irek, J. (2019). Web scraping for food price research. *British Food Journal*, 3350–3361.

Lawson, R. (2015). *Web Scraping with Python.* Birmingham: Packt.

Myhrvold, N. (n.d.). *Cooking*. Retrieved January 08, 2023, from Britannica: https://www.britannica.com/topic/cooking

Nilesh, Kumari, M., Hazarika, P., & Raman, V. (2019). Recommendation of Indian Cuisine Recipes based on Ingredients. *2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)* (hal. 96-99). Macao, China: IEEE.

Python. (2023). *General Python FAQ*. Dipetik January 09, 2023, dari Python Documentation: https://docs.python.org/3.11/faq/general.html

Samuel, A. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development, 3*(3), 210-229.

WebHarvy. (t.thn.). *What is Web Scraping? | Practical Uses & Methods*. Dipetik January 15, 2023, dari WebHarvy: https://www.webharvy.com/articles/what-is-web-scraping.html