Republic of the Philippines
CAVITE STATE UNIVERSITY
Don Severino de las Alas Campus
Indang, Cavite

OPTIMIZING K-MEANS CLUSTERING ALGORITHM USING SILHOUETTE SCORE FOR AN INGREDIENT-BASED RECIPE RECOMMENDATION SYSTEM

Rationale/ Introduction

There has been a global attention on food loss and waste because of its tremendous growth in the last few years. Additionally, a rough estimate of ⅓ of food produced for human consumption, amounting to 1.3 billion tons yearly, is being wasted (Dou & Toth, 2020) and this number of food wasted is linked to the grand challenges being faced by many people all around the globe specifically, world hunger. A depressing number of 45% child deaths are due to hunger and its related causes. The number of food produced in the world alone can feed everybody on the planet, yet it is sad to note that food waste is still a grave challenge the world is facing. (*World Hunger Facts & Statistics*, n.d.) However, with food recommendation systems, developed to provide personalized recommendations tailored to user's preferences, it can help with supporting decision-making when it comes to coming up with food recipes based on provided ingredients, and also considering their nutritional value. These food recommendation systems utilize machine learning algorithms to analyze user data and make informed recommendations based on the limited choice of ingredients on hand, all to hopefully reduce the possibility of food waste by maximizing every ingredient and including it to produce a dish. A study by (Yuan & Luo, 2019) implemented a similar approach for a personalized diet recommendation utilizing the k-means cluster and collaborative filtering algorithm. They used the elbow method to determine the optimal number of clusters, denoted as k, which is then used to assess the effectiveness of clustering algorithms as it provides more accurate and stable clustering configuration. Then they initiate the k-means algorithm to complete the clustering process. Along with a collaborative filtering model based on user preference, using precision to measure the quality of the recommendation system, the higher the coincidence of food eaten based on the recommended food, means the higher the accuracy of the recommendation. In their testing and experiments, they achieved an average of 0.77, 0.7 higher than the recommended average of 0.70, which means the algorithm is stable and it satisfies the user's dietary needs.

While the elbow method provides a straightforward approach to estimate the optimal k value, it may not always produce clear results, especially for complex datasets.

Utilizing alternative methods or refining the elbow method by incorporating additional criteria and instead of solely relying on the elbow method, it can be combined with silhouette scores. After the “elbow” point is determined, it will then compute silhouette scores for neighboring k values to ensure that the chosen k value yields clusters with high silhouette scores, indicating better separation between clusters. The addition of determining the k value using silhouette score provides cluster results that are more optimal. This can also be applied to different datasets, with more complex data. (Mulyani et al., 2023)

By analyzing the algorithmic complexity and assessing the performance of the proposed integration of a K-means algorithm and a user-based collaborative filtering, and the implementation of silhouette score to optimize the performance of the algorithm, it should provide an output that aligns with the SDG goal of zero hunger. This would be possible because the system and its algorithms should maximize the inputs of the user to provide a recipe that not only is healthy, but also keeping in mind the limited number and availability of ingredients. This encompasses a set of problems that the system aims to provide a solution to. Reduced food waste, a wider variety of food choices based on personal preferences, and providing a meal that is healthy and nutritious.

World hunger and lack of access to basic needs such as food has been a prevalent problem being faced by many people in the world. This proposal aims to be of help in fighting this problem by developing a food recommendation system and implementing an innovative machine learning approach to produce an output that is tailored to the user’s dietary needs and the availability of ingredients which could help in reducing food waste. A deeper understanding of how the algorithm works also provides a deeper understanding of the relationship of food to things such as an individual’s wellbeing and their health.

Significance of the Study

The United Nations continues to fight a battle to eradicate world hunger, particularly in the world's poorest nations and achieve a goal of providing healthier, tastier, sustainable diets by the year 2030. This goal is what SDG 2, Zero Hunger is all about. But it is battling against a huge number of 800 million people still having problems surviving a day without enough food for the family. Yet, a whopping number of 1.3 billion tons of food waste is lost every year and it is estimated that it can provide one in three mouthfuls of food every day (Deen, 2018). With the goal of the system and its machine learning approach to provide a

personalized recommendation of recipes, it can help to reduce food waste by using every available ingredient to the user and providing it with food recommendations that includes the provided ingredients by the user. This aligns with the goal of the UN of providing a healthier, tastier, and sustainable diet as it maximizes the flexibility of many ingredients that are easily available to the pantry of most people to be combined in a healthy, cost-efficient dish.

This research proposal is highly involved in smart engineering because of the development of an intelligent recommendation system that also utilizes advanced algorithms such the K-means clustering algorithm. It also emphasizes the use of innovative technologies and data-driven approaches to improve decision-making processes which in this context is a food recommendation system. The development of the system software and the implementation of advanced algorithms itself could also align with ICT as it demonstrates the integration of a computer science approach such as assessing the performance of the algorithm and optimizing it to provide more accurate results. And in the sense of industrial competitiveness, a system that produces personalized, ingredient based food recommendations presents an innovative and a unique solution to reduce food waste and address global issues like hunger.

Scope and Limitations of the Study

This study focuses on developing a food recommendation system that suggests recipes based on available ingredients and user preferences. The system will use a clustering technique called K-means, improved by a method called the silhouette score, to create more accurate and relevant recipe recommendations. The main goal is to help reduce food waste by maximizing ingredient use while also promoting healthier eating habits. The system will have a simple interface where users can input their available ingredients, and in return, they will receive a list of suitable recipes. By integrating machine learning, the system aims to improve decision-making in meal preparation, helping users make the most of what they have while still considering nutritional value.

However, there are certain limitations to this study. The system's effectiveness depends on the quality and variety of data used to train it, meaning that a limited dataset may result in less accurate recommendations. Additionally, while the recommendation system considers ingredient availability, it does not take into account personal allergies, cultural dietary restrictions, or specific cooking skills, which may affect the usability of the

suggested recipes. Furthermore, the system is primarily designed for individual users rather than large-scale food service providers or restaurants. Lastly, factors such as internet connectivity and user familiarity with technology may influence how accessible and practical the system is for different users.

Objectives of the Study

This proposal aims to develop an intelligent system that can provide a list of recipe recommendations based on user inputs using an optimized K-clustering algorithm with the implementation of silhouette scores. Specifically, it aims to:

1. Develop a recommendation system: Design a simple user interface that can present recommendations based on user preferences.
2. Utilize and optimize the K-means clustering algorithm: The algorithm should be utilized and optimized to provide personalized recommendations and a higher accuracy in its outputs.
3. Address food waste: Investigate how the recommendation system using the K-means algorithm can help reduce food waste by suggesting recipes that utilize ingredients already available to the users, hence promoting a more efficient use of food resources.

Expected Outputs

This proposal aims to develop a functional recommendation system tailored to its users' culinary preferences, providing a personalized recipe suggestion based on available ingredients and dietary restrictions. Initial steps involve preprocessing acquired recipes to optimize data distribution for the K-means clustering algorithm. This allows for the clustering of similar recipes according to ingredient compositions which then the system generates a curated list of potential dishes for users to explore.

The introduction of silhouette scores aims to enhance the clustering quality, which assesses cluster compactness and separation. This metric complements the traditional elbow method, resulting in more distinct and well-separated clusters. As the users interact with the system, it adapts intuitively to their preferences through a recommendation approach of user-based filtering. The system also leverages other users' data which the algorithm establishes connections between individual preferences, facilitating shared recommendations among other users with similar preferences..

The integration of the silhouette score should contribute to higher recommendation accuracy, surpassing previous studies' performance by 0.06. This enhancement ensures that the system delivers recipe suggestions that closely align with users' tastes and dietary requirements.

References

Deen, T. (2018, May 28). *Food Waste Enough to Feed World's Hungry Four Times Over - World*. ReliefWeb. Retrieved March 24, 2024, from

https://reliefweb.int/report/world/food-waste-enough-feed-world-s-hungry-four-times-o ver

Dou, Z., & Toth, J. D. (2020, December 13). Global primary data on consumer food waste: Rate and characteristics – A review. *Resources, Conservation and Recycling*, *168*. https://doi.org/10.1016/j.resconrec.2020.105332

Mulyani, H., Setiawan, R. A., & Fathi, H. (2023). Optimization of K Value in Clustering Using Silhouette Score (Case Study: Mall Customers Data). *Journal of Information Technology and Its Utilization.* https://doi.org/10.56873/jitu.6.2.5243

World Hunger Facts & Statistics. (n.d.). Action Against Hunger. Retrieved March 24, 2024, from https://www.actionagainsthunger.org/the-hunger-crisis/world-hunger-facts/ Yuan, Z., & Luo, F. (2019). Personalized Diet Recommendation Based on K-means and Collaborative Filtering Algorithm. *Journal of Physics: Conference Series*, *1213*(3). 10.1088/1742-6596/1213/3/032013