Available online at www.centmapress.org

INTERNATIONAL JOURNAL ON FOOD SYSTEM DYNAMICS

Int. J. Food System Dynamics 13 (2), 2022, 165-173

DOI: https://dx.doi.org/10.18461/ijfsd.v13i2.B4

Halal Food Prediction Using the Similarity Graph Algorithms

Nur Aini Rakhmawati[ab], Girraz Karyo Utomo[a] , Rarasmaya Indraswari[a], and Irfan Rifqi Susetyo[a]

[a]Department of Information Systems, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.
[b]Halal Center, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
nur.aini@is.its.ac.id, utomogirraz@gmail.com, raras@its.ac.id, irfnrifqi@gmail.com

Received October 2021, accepted February 2022, available online April 2022

ABSTRACT

Halal food is food allowed by Islamic sharia. On the contrary, haram food is not permitted, such as alcohol, pork, blood, carrion, and meat not slaughtered according to sharia. Based on Article 39 of Law Number 33 of 2014 concerning halal product guarantees, halal certificates are issued by the Halal Product Guarantee Agency (BPJPH) in Indonesia. The halal certification guarantees that food has a composition containing halal ingredients. However, many food products still do not have a halal certificate. Therefore, it is necessary to estimate the halal status of food products that are still not certified. In this work, we predict the halal status of food using graph similarity algorithms. In this case, we acquire products from the Indomaret website. The product data contains the product name, the composition of the food product, and the manufacturer. Moreover, we crawl the halal food database on Halal MUI website. Both datasets are merged into a single dataset based on the products name. Then, the similarity algorithms such as Jaccard similarity, Approximate Nearest Neighbor, Adamic Adar and Preferential Attachment are performed amongst products in the dataset. F-measure evaluate the accuracy of each algorithm.

Keywords: *halal food; similarity algorithms; food products; graph algorithm.*

1 Introduction

Halal food is any food that meets Islamic sharia (Rezai et al., 2012), while haram food is not allowed consumed, such as alcohol, pork, blood, carrion, and meat not slaughtered according to sharia (Salman and Kamran Siddiqui, 2011). Therefore, a Muslim must look at carefully his food. As the most Muslim population globally, Halal Product Assurance Organizing Agency (BPJPH) issues a halal certificate for a food product in Indonesia (Mukidi, 2020; Svinarky and Malau, 2020). The number of halal-certified products has increased year by year (Warto and Samsuri, 2020).

The Global Islamic Economy (GIE) report for 2020/2021 shows that Indonesia's ranking, in general, rose to rank 4 after the previous period (2019/2020) and (2018/2019) was ranked 5th and 10th respectively (State of the Global Islamic Economy, 2021). The improvement of Indonesia's ranking is a positive impact of the issuance of the 2019-2024 Indonesian Islamic economic master plan issued by the Ministry of National Development Planning/National Development Planning Agency in collaboration with the National Sharia Finance Committee (KNKS) (Ministry of National Development Planning and Committee, 2018) . However, Indonesia is not in the top 10 in the categories of halal food, media and recreation, medicine and halal cosmetics. In fact, Indonesia is a country with the largest Muslim population globally. Only 10% of products are halal certified in the Indonesia (Petriella, 2019).

Not only in Indonesia, Halal food is a primary need for every Muslim in the world. GIE predicts that on a global scale, Muslims spend $2 trillion on halal food (State of the Global Islamic Economy, 2021). As a result, there is a high demand for halal certification. To deal with the lack of halal certification, we can detect the halal status of a product by looking at the ingredients (Karimah and Darwanto, 2021). One of the famous rules is the absence pork ingredients (Çetin and Dinçer, 2016). However, there are several products and its derivatives that cannot be consumed such as blood, dog , reptils, alcohol etc. (Jia and Chaozhi, 2021). Moreover, some food additives are from haram sources(GÜLTEKİN et al., 2020). It is hard to detect the halal status of food if the food ingredients are not well known, such as food additives. Rakhmawati et al. (2018) predict the halal status of uncertified-halal products using Euclidian distance and Cosine similarity. They calculate the similarity of ingredients with products in the Linked Open Data (LOD) Halal dataset (Rakhmawati et al., 2021). If an uncertified-halal product has a similarity value of greater than 0.8 with a certified-halal product, the product can be viewed as a halal product and recommended for use by muslims.

Food prediction and recommendation research have been conducted with various methods. Petković et al. (2021) created DIETHUB to predict food recipes based on diet activities. They used FoodBase (Popovski et al., 2019) as the dataset. The FoodBase corpus consists of 12,844 food entities, 2105 unique foods. The tool classifies the dataset into five categories: Appetizers and Snacks, Breakfast and Lunch, Dessert, Dinner, and Drinks. Ispirova et al.(2020) cluster the nutrient information on food consumption data in Slovenia. The dataset contains 3265 food data. They combine word embedding, graph embedding, and machine learning for clustering the dataset into 23 categories. Yunus et al.(2019) employ different deep learning models to identify food from food images. Rakhmawati and Jannah (2021) calculate the similarity of food ingredients using several methods such as Jaccard similarity, Jaro-Winkler Distance, Levenshtein Distance, Wordnet LCH Similarity, and Fuzzy String Matching. The algorithms perform over the Open Food Facta dataset (https://world.openfoodfacts.org/). Another work was done by Kautsar and Rakhmawati (2019). They integrate data from various halal certification bodies. The similarity that the seven certification bodies own as much as 11.3% have similarities with the power that already exists in Halal Nutrition Food. The best similarity method is Levenshtein similarity, with 2,685 products out of a total of 10,153. Rakhmawati and Najib (2020) investigate the similarity between food products and create product interlinks in Linked Open Data Halal. Node2Vec (Grover and Leskovec, 2016) is implemented to measure the similarity between the products.

Currently, there are about 500 halal certification institutions worldwide However, no international board manages a central database to integrate those halal certification data from various institutions (Tieman and Williams, 2019). As a result, a single product might have multiple halal certificates from different institutions, since each institution has its standards and regulations. Therefore, Rakhmawati et al. (2021) proposed Linked Open Data for halal food products. Linked Data is one of the knowledge graph constructions. The Knowledge graph is a great tool for integrating data from various sources (Rezaul Karim et al., 2019). Thus, in In our work, we will use graph algorithms including Jaccard similarity, Approximate Nearest Neighbor (ANN)(Dong et al., 2011), Adamic Adar(Adamic and Adar, 2003), and Preferential Attachment(Albert and Barabási, 2002). These algorithms also identify the similarity between the halal products. Moreover, the results of similarity calculation are presented in graph visualizations. These presentations could assist a Muslim in identifying a non-halal-certified product connecting to a halal-certified product.

The remaining paper is divided into four sections: Section 2 **Fehler! Verweisquelle konnte nicht gefunden werden.**describes the proposed methodology. Section 3 presents our experiment and the result. We conclude it in Section 4.

2 Methodology

Our methodology consists of four steps, namely data acquisition, similarity analysis, evaluation and visualization (Fig 1). The four steps can be explained as follows:

2.1 Data Acquisition

The main source of this research dataset is KlikIndomaret (https://www.klikindomaret.com/). This website is an online shopping site from the Indomaret supermarket. The data retrieved includes for food and beveages the name of the product, material, and manufacture. Data retrieval uses Scrapy Python (https://scrapy.org/). Scrapy is a Python library for extracting data from a website. There are 3274 products in 21 for food and drink groups on the website.

We also collect data from the official Halal MUI website (Majelis Ulama Indonesia, https://www.halalmui.org/). It includes 401,650 products in 36 categories. Data from the Halal MUI website will serve as complementary data for Indomaret data since the status of the halal products is unavailable at KllikIndomaret.

KlikIndomaret and MUI Halal datasets are merged into a single dataset using the Jaccard similarity algorithm. We compare the name of products and calculate the Jaccard similarity. If the similarity value is greater than 0.7, then the products in both datasets are similar.

The result of data integration is converted into a graph consisting of four types of nodes: ingredient, manufacturer, certificate and food product. The graph dataset can be found at https://github.com/utomogirraz/Paper-English-Halal.

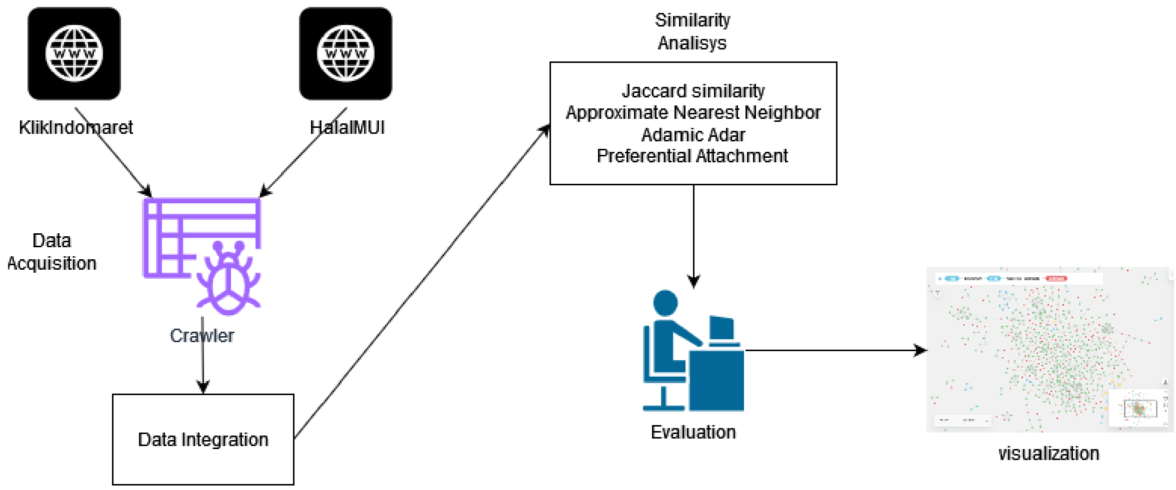

Figure 1. Methodology

2.2 Similarity Analysis

We perform the following graph similarity algorithms over the graph dataset:

- Jaccard similarity (Rakhmawati and Jannah, 2021)calculates dataset similarity, dissimilarity, and distance. It measures proximity and checks data redundancy. It is defined as the ratio of the number of common neighbours and the number of neighbours.
- Approximate Nearest Neighbor (ANN) (Dong et al., 2011) is an algorithm that builds a k-Nearest Neighbor Graph model. This algorithm calculates the similarity based on "Jaccard, Cosine, Euclidean Distance" (Liu et al., 2019), or Pearson similarity (Benesty et al., 2009).
- Adamic Adar (Adamic and Adar, 2003) is a measurement used for computing the nearest node based on shared neighbours. The Adamic Adar algorithm was first introduced in 2003 by Adamic and Adar to predict linkages in social networks.

- Preferential Attachment (Albert and Barabási, 2002) calculates the nearest node by looking at its shared neighbours. This algorithm can also be interpreted for the more related nodes, the greater the node has a connection.

Neo4J provides these four algorithms. Neo4J is a graph database platform that supports many features for data analysis.To calculate the accuracy of algorithms, we label the dataset semi-manually. First, the Jaccard similarity runs on the dataset. Then we check the accuracy of pairs of products manually. The number of pairs of the product are 282492 pairs. The average number of correct labels is 66%, where the threshold is 0.42. The results of the Jaccard label becomes the reference for other algorithms. After getting similarity values, we create an interlinked between two products that have a similarity value greater than 0.42.

2.3 Evaluation

We conduct a human evaluation for assessing the Jaccard and ANN algorithm. Three evaluators check the list of pairs of products provided in a Google Form. To reduce the number of assessments of pair products, we only pick 271 pairs of products having a similarity value higher than 0.8 both for Jaccard similarity and ANN algorithms. We calculate recall, precision and F-Measure value. The recall is the ratio of the number of true positive to the number of pairs of products, while the precision is the ratio of the number of true positives to the number of the total positive of products. F measure is the weighted average of precision and recall. We also calculate the Kappa coefficient measurement to evaluate disagreement between respondents.

2.4 Visualization

Neo4JBloom, one of the Neo4J tools for creating a code-less visualization, visualizes the interlinking between the products from the previous stage. This visualization provides results for depicting food products that have similarities in halalness. Only 21.239 out of 71.423 products are presented in this visualization since these products contain at least one ingredient. To give better insight, we calculate the page rank of the products. There are three colours for the product. The first colour is green, representing the product node with a halal certificate. Node with the yellow colour does not have a halal certificate but has a relationship with a certified product (green node). Therefore, the yellow nodes are predicted as halal products. The remaining nodes are in blue. The red nodes indicate the halal-certificate nodes. Figure 2 describes our visualization graph for the ANN similarity.

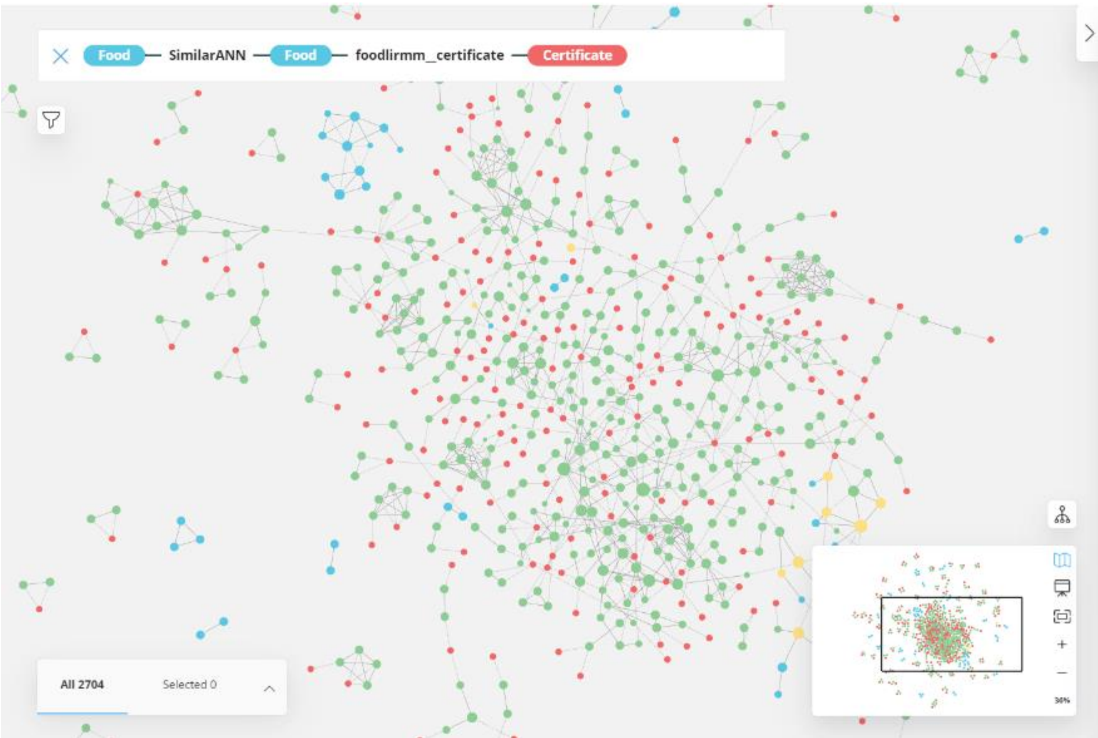

Figure 2. ANN Visualization Results

3 Results and Discussions

Table 1 describes the outcome of the human evaluation for checking the similarity of the products. Evaluators 1 and 2 have the same number of judging similar and non-similar products. In total, we have 182 agreements among the evaluator. Therefore, the coefficient kappa is 0.90, which means that the level of agreement between the respondents is very strong.

Table 1.
Evaluator Assessment Results

Result	Similar	Not Similar
Evaluator 1	140	131
Evaluator 2	140	131
Evaluator 3	160	111

As seen in Table 2, the algorithm's accuracy is 99%, where the F-measure is 96%. It can be concluded that the similarity algorithm can predict the similarity of the products.

Table 2.
Algorithm Measurement Value

Result	Value
Accuracy	0.9998
F1 Score	0.9645
Precision	1.0
Recall	0.9314

The number of nodes, relationships and certificate nodes generated by Jaccard, ANN and Preferential Attachment are precisely the same (Table 3). However, the average scores are different since the similarity calculations are different.

Table 3.
Algorithm Comparison Results

Algorithm	Number of Product Nodes	Number of Certificate Nodes	Number of Relationships	Number of Certificate Relationships	Average Score
Jaccard	603	248	1332	521	0.621
ANN	603	248	1332	521	0.621
Adamic Adar	590	238	1279	508	2.803
Preferential Attachment	603	248	1332	521	152.67

Figure 3 shows a visualization for products that do not have a relationship with certificates or halal-certified products in Jaccard similarity. Coated peanuts (garlic powder) product has no relation with *Garuda Pilus Pedas*. These products have no connection with products that have certificates.

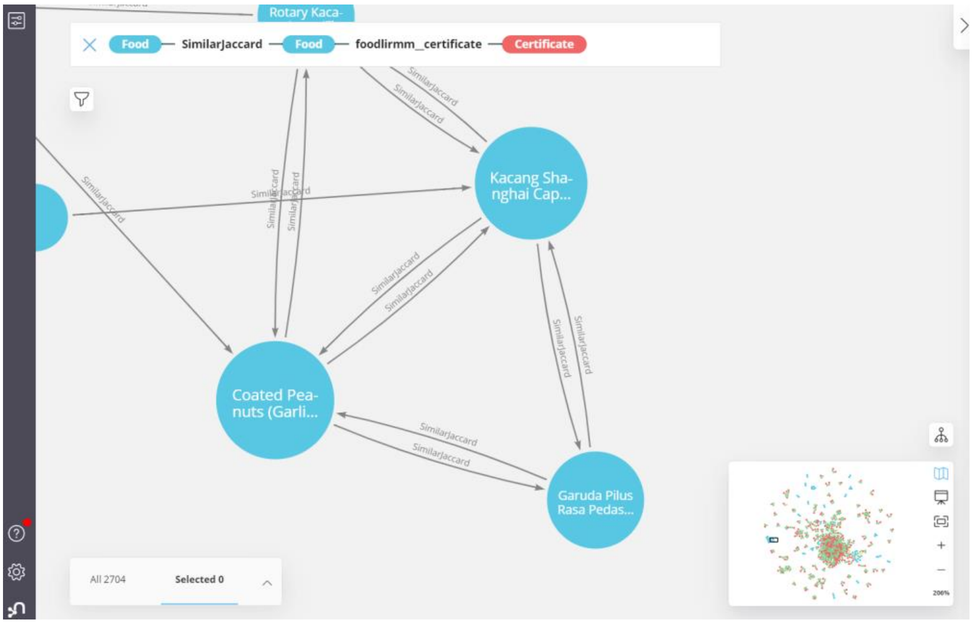

Figure 3. Visualization for Products without halal-certified (Jaccard similarity)

A product does not have a halal certificate but is related to a halal-certified product. Thus, the product is considered halal and given a yellow colour. As seen in Figure 4, *Tango Susu Vanilla* and *Wafer Selamat* connect to Dua Kelinci Deka Wafer Rolls, which has a halal certificate.

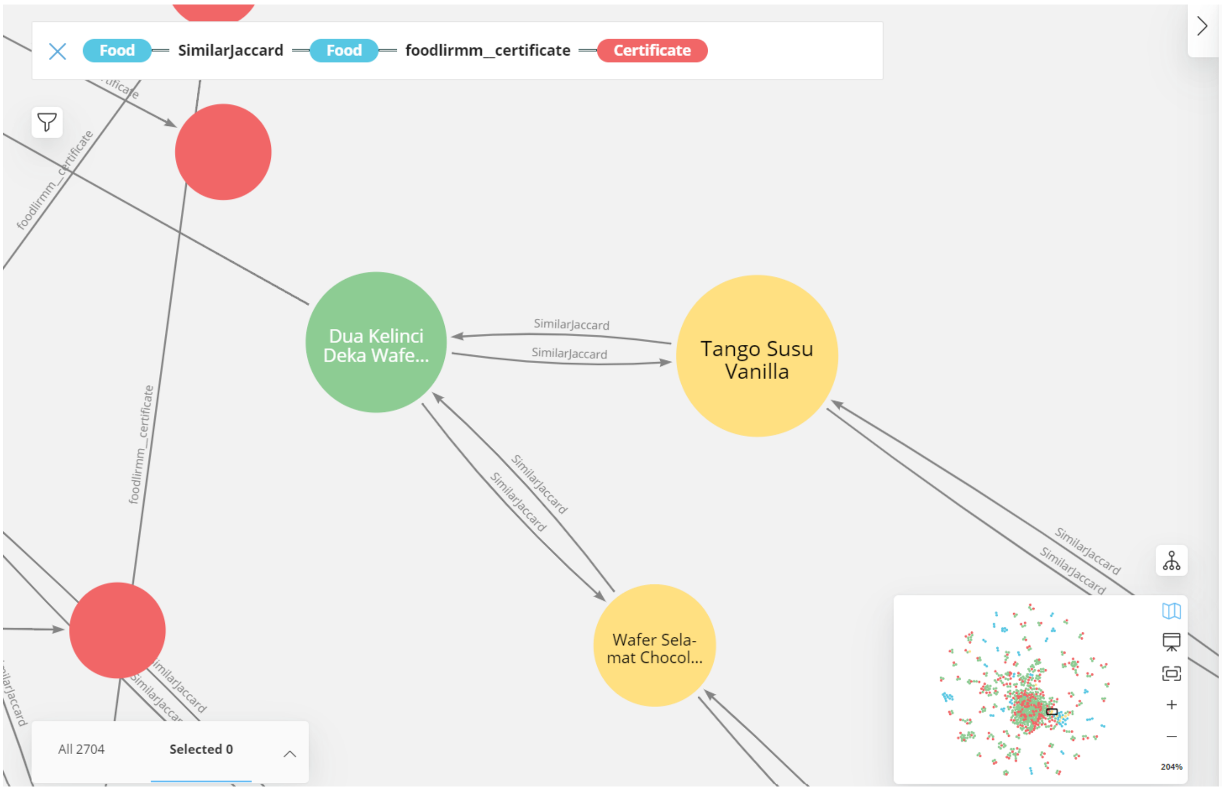

Figure 4. Two products connect to a halal-certified product.

4 Conclusions

We have presented our effort to link halal-certified products and non-halal certified products. These products are from the KlikIndomaret and Halal MUI websites. Four algorithms were utilized in analysing the dataset, such as Jaccard, ANN, Adamic Adar and Preferential Attachment. Three algorithms, Jaccard, ANN and Preferential

Attachment, generate the same number of relationships between the products. The human evaluator assesses the algorithm's performance. The visualization could assist people, especially Muslims, in predicting the halal status of the food.

Based on our experience with the dataset from KlikIndomaret, one of the popular grocery retail chains in Indonesia, the number of non-halal-certified products is greater than the number of halal-certified products. Our results can recommend substitute halal products as halal confirmed alternatives. The product analysis could be helpful for countries in judging a halal status of a product where a traceability of products is not in place.

In addition, the results provide the basis for the development of a halal database that can assist Muslims all over the world, even when living in non-muslim countries. Such a halal database would not only contain products that have been certified but also products that have not been officially certified. However, as the majority of world halal certification bodies may only show the brand name, company and expiration date of a product there is a lack of data on the composition of halal food, necessary for evaluating non-certified products (Aini Rakhmawati and Choirun Najib, 2020). It is, therefore, necessary to match product and company names from halal certification institutions with data from open food databases such as OpenFoodFacts. According to DataWorld (data world, 2022), there are 266 food datasets available which could be assessed. Further problems may arise in situations where halal certificate institutions do not share halal-certified product names. As an example, Majelis Ulama Indonesia accredits 44 world halal certification bodies, but only 22% have disclosed their data, and only two institutions have listed food compositions (Kautsar and Rakhmawati, 2019). Thus, we will integrate the dataset with the LOD dataset (Rakhmawati et al., 2021).

We have used four algorithms in analysing the dataset. However, there are other similarity algorithms that could be used and might deliver different weights and values.

In the end, the suggestion for halal certification institutions can be listed as follows:

1. Sharing their dataset includes product name, company name and expired date, and the list of ingredients.
2. The dataset should be in a machine-readable format such as CSV. Most of the datasets are in HTML and PDF, which needs further pre-processing before using the datasets.
3. The halal certification institutions should provide an API for accessing the dataset in real-time. Thus, we could analyze the data in run time
4. Several halal certification bodies list the data, not in English name or non-Latin character such as in Arabic, Urdu, etc. They should show the dataset in two alphabets: Latin and non-Latin alphabets.

Acknowledgements

This work was supported by the Department of Information Systems, Institut Teknologi Sepuluh Nopember with grant number 1838/PKS/ITS/2021.

References

Adamic, L. A., Adar, E. (2003). Friends and neighbors on the Web. *Social Networks*, **25**(3): 211–230. https://doi.org/10.1016/S0378-8733(03)00009-1.

Aini Rakhmawati, N., Choirun Najib, A. (2020). Discovering Entity Profiles Candidate for Entity Resolution on Linked Open Data Halal Food Products. *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*. https://doi.org/10.1109/BigData50022.2020.9378411.

Albert, R., Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, **74**(1): 47. https://doi.org/10.1103/RevModPhys.74.47.

Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). Pearson Correlation Coefficient. *Springer Topics in Signal Processing*, **2**: 1–4. https://doi.org/10.1007/978-3-642-00296-0_5.

Çetin, G., Dinçer, M. Z. (2016). Muslim Friendly Tourism (MFT): A Discussion. *Journal of Tourismology*, **2**(1). https://doi.org/10.26650/jot.2016.2.1.0005.

Data world (2022). *Food data on data.world*. https://data.world/datasets/food.

Dong, W., Charikar, M., and Li, K. (2011). Efficient K-nearest neighbor graph construction for generic similarity measures. *Proceedings of the 20th International Conference on World Wide Web, WWW 2011*: 577–586. https://doi.org/10.1145/1963405.1963487.

Grover, A., Leskovec, J. (2016). Node2vec: Scalable feature learning for networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*: 855–864. https://doi.org/10.1145/2939672.2939754.

Gültekin, F., Çeker, O., Elgün, A., Gürbilek, M., Şimşek, M., Oral, R., Köse, S., Türker, S., Döndüren, H., Ünaldi, M., Küçüköner, E., Yetim, H., Aydin, A., Dede, B., Özmen, İ., Başayiğit, L., Cengiz, M., Tilki, T., and Akin, S. (2020). Evaluation of Food Additives in Terms of Istihalah. *Akademik Platform Helal Yaşam Dergisi*, **2**(1): 1–13. https://dergipark.org.tr/tr/pub/apjhls/issue/55626/713197.

Ispirova, G., Eftimov, T., and Seljak, B. K. (2020). P-nut: Predicting nutrient content from short text descriptions. *Mathematics*, **8**(10). https://doi.org/10.3390/math8101811.

Jia, X., Chaozhi, Z. (2021). Turning impediment into attraction: A supplier perspective on Halal food in non-Islamic destinations. *Journal of Destination Marketing and Management*, **19**: 100517. https://doi.org/10.1016/J.JDMM.2020.100517.

Karimah, N., Darwanto, D. (2021). Effect of Halal Awareness, Halal Certification, Food Ingredients Composition, Halal Marketing and Religiosity on Purchase Intention of Imported Food Products. *Al-Kharaj : Jurnal Ekonomi, Keuangan and Bisnis Syariah*, **3**(2): 177–191. https://doi.org/10.47467/alkharaj.v3i2.319.

Kautsar, A., Rakhmawati, N. A. (2019). *Integrasi Data Lembaga Sertifikasi Halal Di Dunia Dengan Data Halal Nutrition Food Integration Of Halal Certification Bodies Data From Around The World With Halal Nutrition Food Data*. Institut Teknologi Sepuluh, November.

Liu, D., Chen, X., and Peng, D. (2019). Some cosine similarity measures and distance measures between q-rung orthopair fuzzy sets. *International Journal of Intelligent Systems*, **34**(7): 1572–1587. https://doi.org/10.1002/INT.22108.

Ministry of National Development Planning, and Committee. (2018). *Masterplan Ekonomi Syariahindonesia 2019-202*. https://knks.go.id/storage/upload/1573459280-Masterplan Eksyar_Preview.pdf.

Mukidi, M. (2020). Prosedur Pemberian Sertifikat Label Halal Terhadap Produk Makanan Di Restoran Hotel Syariah Untuk Mewujudkan Hak Kenyamanan Konsumen Muslim (Studi Di Kementerian Agama Provinsi Sumatera Utara). *Jurnal Hukum Kaidah: Media Komunikasi Dan Informasi Hukum Dan Masyarakat*, **19**(3). https://doi.org/10.30743/jhk.v19i3.2674.

Petković, M., Popovski, G., Seljak, B. K., Kocev, D., and Eftimov, T. (2021). DIETHUB: Dietary habits analysis through understanding the content of recipes. *Trends in Food Science and Technology*, **107**. https://doi.org/10.1016/j.tifs.2020.10.017.

Petriella, Y. (2019). *LPPOM UI: Baru 668.615 Produk di Indonesia Tersertifikasi Halal*. Bisnis.Com. https://ekonomi.bisnis.com/read/20190116/12/879298/lppom-ui-baru-668.615-produk-di-indonesia-tersertifikasi-halal.

Popovski, G., Seljak, B. K., and Eftimov, T. (2019). FoodBase corpus: A new resource of annotated food entities. *Database*, **2019**(1). https://doi.org/10.1093/database/baz121.

Rakhmawati, N.A., Firmansyah, A. A., Effendi, P. M., Abdillah, R., and Cahyono, T. A. (2018). Auto Halal detection products based on euclidian distance and cosine similarity. *International Journal on Advanced Science, Engineering and Information Technology*, **8**(4-2).

Rakhmawati, N.A., Fatawi, J., Najib, A. C., and Firmansyah, A. A. (2019). Linked open data for halal food products. *Journal of King Saud University - Computer and Information Sciences*, **33**(6): 728-739. https://doi.org/10.1016/j.jksuci.2019.04.004.

Rakhmawati, Nur Aini, Jannah, M. (2021). Food Ingredients Similarity Based on Conceptual and Textual Similarity. *Halal Research Journal*, **1**(2): 87–95.

Rezai, G., Mohamed, Z., and Shamsudin, M. N. (2012). Assessment of Consumers' confidence on Halal labelled manufactured food in Malaysia. *Pertanika Journal of Social Science and Humanities*, **20**(1): 33-42.

Rezaul Karim, M., Cochez, M., Jares, J. B., Uddin, M., Beyan, O., and Decker, S. (2019). Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-LSTM network. *ACM-BCB 2019 - Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. https://doi.org/10.1145/3307339.3342161.

Salman, F., Kamran Siddiqui. (2011). An exploratory study for measuring consumers awareness and perceptions towards halal food in Pakistan. *Interdisciplinary Journal of Contemporary Research in Business*, **3**(2): 639-651.

State of the Global Islamic Economy. (2021). State of the Global Islamic Economy Report 2020/2021. In *Dubai International Financial Centre*.

Svinarky, I., Malau, P. (2020). Penerbitan Sertifikat Halal Berdasarkan Prespektif Hukum Di Indonesia. *Jurnal Cahaya Keadilan*, **8**(1): 71-85. https://doi.org/10.33884/jck.v8i1.1896.

Tieman, M., Williams, G. (2019). Creative Destruction of Halal Certification (Bodies) By Blockchain Technology? *ICR Journal*, **10**(1): 127–131. https://doi.org/10.52282/ICR.V10I1.79.

Warto, W., Samsuri, S. (2020). Sertifikasi Halal dan Implikasinya Bagi Bisnis Produk Halal di Indonesia. *Al Maal: Journal of Islamic Economics and Banking*, **2**(1): 98-112. https://doi.org/10.31000/almaal.v2i1.2803.

Yunus, R., Arif, O., Afzal, H., Amjad, M. F., Abbas, H., Bokhari, H. N., Haider, S. T., Zafar, N., and Nawaz, R. (2019). A Framework to Estimate the Nutritional Value of Food in Real Time Using Deep Learning Techniques. *IEEE Access*, **7:** 2643–2652. https://doi.org/10.1109/ACCESS.2018.2879117**.**