Exhibition and Seminar on Science and Creative Technology, University of Al-Azhar Indonesia (EXSACT-A 2024) Proceeding

DOI: https://doi.org/10.36722/exc.v2i1.2341

Decision Model of Interest in Buying East Asian Food Products based on Halal Awareness and Religiosity Values using the Classification Machine Learning Algorithm Method

Salsa Fairuz[1*], Budi Aribowo[2]

[1,2]Industrial Engineering, Faculty of Science and Technology, University Al-Azhar of Indonesia, Al Azhar Grand Mosque Complex, Jalan Sisingamangaraja, Kebayoran Baru, South Jakarta 12110

Corresponding author /E-mail: salsafairuzz@gmail.com

Abstract –**Indonesia has not yet developed a strong halal food industry despite its potential and remains more of a target market. With rapid digital transformation, cultural influences, and social media trends, halal awareness among Indonesians needs to be strengthened. This study aims to explore the most influential factors on buying interest in East Asian food products based on halal awareness and religiosity values, to model purchase decisions, and stimulate the model. A descriptive quantitative method was used with 302 Muslim respondents aged 15-60+ in Jabodetabek who are active social media users. Data were analyzed using five machine learning classification algorithms: Naïve Bayes, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression. Validity and reliability test were conducted to ensure the quality of the instruments. Results show that the key factors influencing buying interest are halal certification, certification institutions, and brand image. Among the models tested, the K-Nearest Neighbors algorithm delivered the performance with 97% accuracy and an AUC of 0,91. This model effectively classifies consumer interest in East Asian food products. The findings highlight the importance of halal assurance and brand trust in shaping consumen behavior in Indonesia's growing halal market.**

Keywords– *East Asian food products, Halal awareness, Machine learning classification, Purchase intention, Religious values*.

INTRODUCTION

According to the 2022 State of The Global Islamic Economy Report (GIER) released by the Dinar Standard, March 31, 2022. Indonesia is ranked fourth in the halal food and beverage category. However, even though Indonesia is a country with the highest level of consumption of halal food in the world, the development of the halal food industry in Indonesia is still not advanced.

In addition, digital transformation is currently growing rapidly due to COVID-19. According to research conducted [1] when the AFTA (Asean Free Trade Agreement) was enacted in 2015 as a form of globalization trend in this century, the flow of export-import trade between countries in the region is getting higher and without limits. Long before that, very intensive global trade had taken effect which also made the boundaries between countries in the world almost non-existent. According to Robertson. [2]. Globalization refers to the incentive narrowing of the world and the increase in our awareness of the world, namely by increasing global connections and our understanding of these connections. The emergence of culinary tourism trends in Indonesia is due to the diversity of local and international specialties. The development of the food and drink services sector along with the growth of the tourism sector has resulted in an increase in

culinary tourism activities. Culinary tourism is not only for learning about other cultures, but also for going on food tasting adventures [2].

Free markets and globalization certainly have positive and negative impacts on the development of the food market in Indonesia. One of them is East Asia. The influence of East Asia in Indonesian culinary trends can be found in various corners of the city as if East Asian food has become the stigma of typical Indonesian food. For example Korea, not only music and drama series are able to hypnotize fans from Indonesia, Korean food has also gained popularity. Before Korean specialties mushroomed, Japanese specialties were already known to the Indonesian people. There are several factors that influence East Asian food in culinary trends in Indonesia, namely the taste factor, East Asian food tends to be in demand by Indonesians because the taste is not much different from the Indonesian tongue. Vira, an Indonesian of Arab descent admitted that his taste buds are fine with Japanese food. "As long as it's halal," he said. Apart from that, Indonesians are also easily influenced by new things that seem more fun and worthy of interest. In addition to the taste factor, the mushrooming of sellers of various food and beverages also encourages people to accept them easily. Finally, there are world trend factors that influence food trends in Indonesia. In earlier times, food trends headed towards Europe. Now, people are bored and start looking at Asian food which is considered richer in taste. Even the fusion food trend is mushrooming. The mushrooming of traders of various foods and beverages also encourages people to accept them easily. Finally, there are world trend factors that influence food trends in Indonesia. In earlier times, food trends headed towards Europe. Now, people are bored and start looking at Asian food which is considered richer in taste. Even the fusion food trend is mushrooming. The mushrooming of traders of various foods and beverages also encourages people to accept them easily. Finally, there are world trend factors that influence food trends in Indonesia. In previous times, food trends were towards Europe. Now, people are bored and are starting to look at Asian food which is considered richer in taste. Even the fusion food trend is mushrooming [3].

The current halal lifestyle is no longer just a habit and consumption for Muslims, but has started to become a trend that is considered healthy and good for the health of the world's population, especially countries with a minority or relatively small total Muslim population [4].

Based on this, it can be concluded that Indonesia has not been able to develop a good halal food industry, even though there are many opportunities, Indonesia is still comfortable as a target market for halal food. Then, with the increasingly rapid digital transformation, the growing acculturation of foreign cultures entering Indonesia, trends on social media, the issue of halal and haram in food, the Indonesian people should increase their awareness of halal products to be consumed. In the phenomenon that is happening a lot today, especially related to halal-haram in food, it shows that halal products are becoming more relevant and become one of the concerns in the midst of the large number of global products entering Indonesia. which encourages local and global business actors to obtain halal certification from the Indonesian Ulama Council (MUI) for their products. This can be seen from the number of fast food outlets and imported food products that are trying to get halal certificates from the MUI. This shows that halal products are rapidly gaining worldwide recognition as a new standard for human safety and product quality assurance. This is because products with halal certification are not only more confident to be accepted by Muslim consumers, but also by consumers of other religions This shows that halal products are rapidly gaining worldwide recognition as a new standard for human safety and product quality assurance. This is because products with halal certification are not only more confident to be accepted by Muslim consumers, but also by consumers of other religions This shows that halal products are quickly gaining recognition worldwide as a new standard for human safety and product quality assurance. Because, products with halal certification are not only more confident in being accepted by Muslim consumers, but also by consumers of other religions [5].

This research aims to find out and explore further what factors have the highest level of importance in buying interest in East Asian food products based on halal awareness and religiosity values which are contained in 15 research indicators, namely halal awareness, halal certification, halal certification institutions , halal labels, product naming, religiosity, food ingredients, environment, social media, influencers, family, subjective norms, marketing/promotion, price, brand image,

and purchase intention as a reference for purchasing interest decisions.

METHOD

This research is classified as a quantitative descriptive research that combines descriptive and quantitative analysis approaches. According to [6] quantitative descriptive research is describing, researching and explaining something that is studied as it is, and drawing conclusions from phenomena that can be observed using numbers. According to [7] quantitative descriptive research is research that only describes the contents of a variable in the research, not intended to test a particular hypothesis. Therefore, based on its definition, quantitative descriptive is data obtained from a sample of a population, analyzed using statistical methods and described with a sentence that describes the results of the analysis. In this research, a quantitative approach was used for psychographic segmentation data, and a descriptive approach was used for demographic and geographic segmentation data of respondents. The purpose of this research are:

1. Testing research instruments using validity and reliability tests.
2. Analyzing consumer behavior towards Muslim consumers' buying interest in East Asian food products based on halal awareness and religious values
3. Knowing the most underlying variables for consumers to buy East Asian food products.
4. Identifying the level of purchase interest decision model that produces the best performance in classifying research instruments.
5. Carry out a simulation of the modeling that has been created.

In this study, researchers used market segmentation methods (demographic, geographical, and psychographic), descriptive analysis, data preprocessing, model analysis on classification Machine Learning algorithms (Naïve Bayes, Decision Tree, SVM, KNNeighbors, Logistic Regression), as well as analysis of decision models .

Data collection was carried out in this research using the method of distributing questionnaires to respondents in accordance with the research criteria. The distribution of this questionnaire is related to the consumer market segmentation of East Asian food products which includes demographics, geography and psychographics. The stages in distributing the questionnaire are explained as follows:

a. Preparation of a market segmentation questionnaire for consumers of East Asian food products.
b. Analysis of each variable based on halal awareness and religious values.
c. Trial the consumer's understanding response to the questions asked.
d. Distribution of questionnaires via social media, such as WhatsApp groups, line groups, telegram, Instagram, and word of mouth by researchers for a duration of two weeks.
e. The number of respondents who filled out the questionnaire according to the criteria was 302 respondents.
f. All data from the questionnaire proceed to the data processing stage.

Processing data using descriptive statistics. Descriptive statistics are carried out by means of descriptive analysis. Descriptive analysis was carried out to describe respondents' demographic and geographic segmentation data, while a combination of descriptive analysis, data preprocessing and classification machine learning algorithm modeling was used to process respondents' psychographic data. Next, the results of the decision model analysis using a classification machine learning algorithm with the highest AUC value are used as a reference for the decision model regarding interest in purchasing East Asian food products.

Flowchart Study

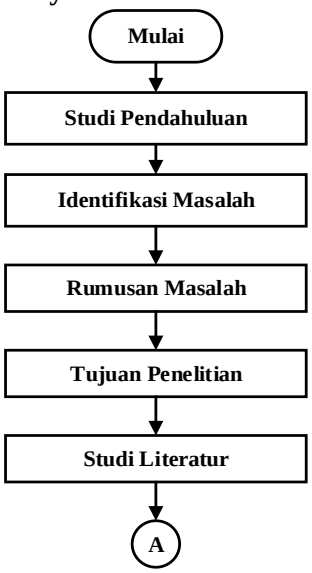

Figure 1. Research Flowchart

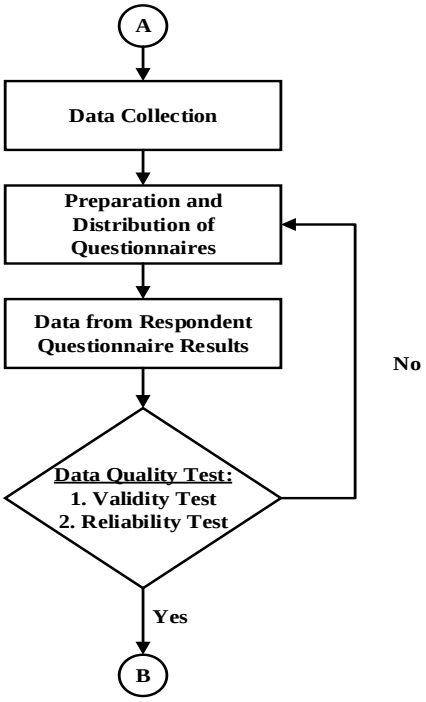

Figure 2. Research Flowchart

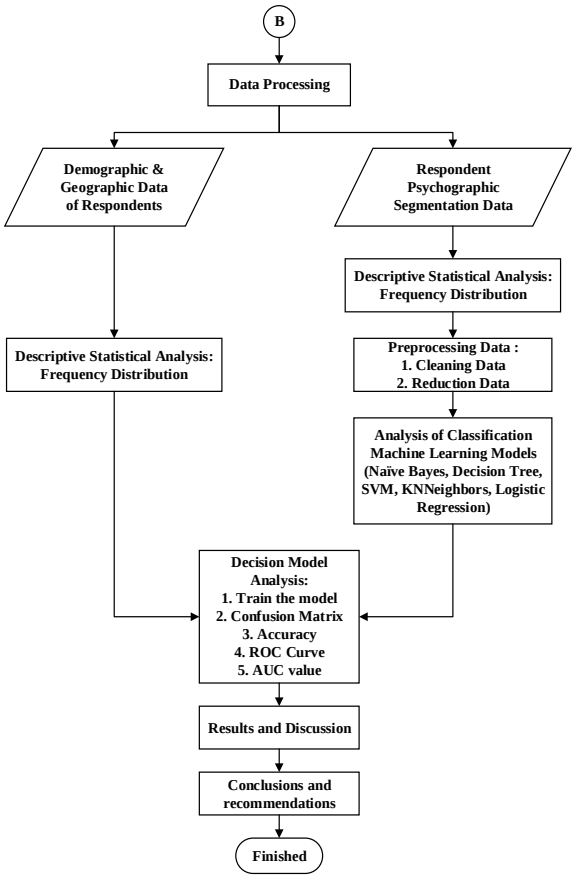

Figure 3. Research Flowchart (Continued)

Place and time of research

This research was conducted online in collecting data through questionnaires distributed in Jabodetabek. This research was conducted for approximately 3 months (April – June 2023) with a 3 week questionnaire distribution period in June.

Data collection

Data collection was carried out in this research using the method of distributing questionnaires to respondents in accordance with the research criteria. The distribution of this questionnaire is related to the consumer market segmentation of East Asian food products which includes demographics, geography and psychographics. The stages in distributing the questionnaire are explained as follows:

a. Preparation of a market segmentation questionnaire for consumers of East Asian food products.
b. Analysis of each variable based on halal awareness and religious values.
c. Trial the consumer's understanding response to the questions asked.
d. Distribution of questionnaires via social media, such as WhatsApp groups, line groups, telegram, Instagram, and word of mouth by researchers for a duration of two weeks.
e. The number of respondents who filled out the questionnaire according to the criteria was 302 respondents.
f. All data from the questionnaire proceed to the data processing stage.

Measurement Scale

According to [8] a scale is a tool used in the measurement process to be able to obtain information such as direction, level, and intensity of the attachment between indicators owned by a variable. In this study using interval level measurements, namely measurements that can show the existence of categories or differences, levels or levels, and the distance of a variable. The interval scale used is the Likert scale. The Likert scale is a psychometric scale that has four or more questions that are combined to form a score that represents individual characteristics, such as attitudes, knowledge, and behavior [9].The following is the Likert scale score used in this study.

Table 1. Likert Scale

Information	Score
Strongly Disagree (STS)	1
Disagree (TS)	2
Neutral (N)	3
Agree (S)	4
Strongly Agree (SS)	5

Based on the information in Table 1, the author uses a five-point Likert scale regarding the respondent's agreement with the statement or statements put forward prior to the answer options provided. This is supported by [10] Regarding the advantages of the 5-point Likert scale, namely, the questionnaire is able to accommodate answers that are neutral or uncertain. Additionally, According to, [11] a Likert scale of 5 points is more widely used because a Likert scale of 7 points or more will make it more difficult for respondents to distinguish between scale points and cause respondents to have difficulty processing information. The data used in this study were obtained from the results of a Google form questionnaire that had been distributed via social media and word of mouth by the researchers. Researchers provided 15 statements related to the problems in the research. Each answer from the respondent becomes a reference in research which has the meaning to test which independent variables have the most influence on the dependent variable.

Data processing

After data collection, data processing is then carried out using descriptive statistics. Descriptive statistics are carried out by means of descriptive analysis. Descriptive analysis was carried out to describe respondents' demographic and geographic segmentation data, while a combination of descriptive analysis, data preprocessing and classification machine learning algorithm modeling was used to process respondents' psychographic data. Next, the results of the decision model analysis using a classification machine learning algorithm with the highest AUC value are used as a reference for the decision model regarding interest in purchasing East Asian food products.

RESULTS AND DISCUSSION

Validity test

The significance test was carried out by comparing the value of rcount with rtable for the degree of freedom (df) = N-2. N is the number of samples. The questionnaire consists of 15 variables with 1 statement each, so the number of statements or statement items is 15 items with a total of 302 respondents. The number of samples (N) is = 302 and the size of df is df = N-2 = 302 – 2 = 300. In this case the sample = 300 and the significance level is 0.05 with a two-tailed test, so rtable = 0.113. In testing the validity of all statement items that produce an rcount value greater than 0.113, it is declared valid. Table 2 shows the results of the validity test calculations performed.

Table 2. 1 Results of Research Instrument Validity Test

Variable	Indicator	rcount	rtable	Decision
Halal Awareness	I know the halal criteria for an ingredient/East Asian food product that I will consume	0.444	0.113	Valid
Halal Certification	I will only buy East Asian food that has been officially certified by a halal certification agency	0.516	0.113	Valid
Halal Certification Institute	I believe that restaurants/food vendors in East Asia that are halal certified by any halal certification agency will guarantee halal status for all of their menus.	0.466	0.113	Valid
Halal label	The halal label is a consideration for me when buying East Asian food	0.596	0.113	Valid
Product Naming	Product names are my consideration when buying East Asian food	0.565	0.113	Valid
Religiosity	I believe that by consuming halal ingredients/products, life will be a blessing and more focused	0.441	0.113	Valid
Food material	Before buying East Asian food or coming to an East Asian food	0.542	0.113	Valid

Variable	Indicator	rcount	rtable	Decision
	restaurant, I first ask the seller or restaurant about the halal critical points regarding the ingredients used directly or through communication media.			
Environment	Before buying East Asian food or coming to an East Asian food restaurant, I first pay attention to the environment around which it is sold	0.533	0.113	Valid
Social media	East Asian food that goes viral on social media often attracts my attention to immediately buy it	0.492	0.113	Valid
influencers	*influencers*which I follow often plays an important role in purchasing decisions (especially for East Asian food)	0.561	0.113	Valid
Family	My family/relatives influenced me to eat East Asian food which is clearly halal	0.596	0.113	Valid
Subjective Norms	The religious leaders I follow influenced me to consume East Asian food which is clearly halal	0.626	0.113	Valid
Marketing/Promotion	The promotions offered by East Asian restaurants/food sellers made me interested in buying them	0.526	0.113	Valid
Price	Price influences me to buy East Asian food	0.415	0.113	Valid
Brand Image	*Brand image*or the brand image of an East Asian food restaurant influences me to buy the products offered	0.564	0.113	Valid

Reliability Test

The reliability test in this research was carried out to test the consistency or reliability of the research instrument so that it can be trusted in the measurement process. Reliable and reliable calculation results show that the research is able to provide identical or consistent results when repeated measurements are carried out. In this case, the researcher uses the Cronbach's Alpha coefficient formula to test the reliability of research indicators provided that the coefficient exceeds or is equal to the alpha value of 0.60, then the indicator is declared reliable [12]. Table 2.2 is the result of the research instrument test measurements that have been carried out.

Table 2. 2 Results of Research Instrument Reliability Test

Reliability Coefficient	*N of items*	Decision
0.812	15	Very Reliable

Data processing

After data collection, data processing is then carried out using descriptive statistics. Descriptive statistics are carried out by means of descriptive analysis. Descriptive analysis was carried out to describe respondents' demographic and geographic segmentation data, while a combination of descriptive analysis, data preprocessing and classification machine learning algorithm modeling was used to process respondents' psychographic data. Next, the results of the decision model analysis using a classification machine learning algorithm with the highest AUC value are used as a reference for the decision model regarding interest in purchasing East Asian food products.

Data source

The data source in this research comes from data collection through questionnaires and analyzed using a machine learning classification algorithm consisting of the Naïve Bayes classification model, Decision Tree, KNeighbors, Support Vector

Machine (SVM) and Logistic Regression. The data source comes from distributing questionnaires via Google Form covering the population of Muslim/Muslim women consumers who live in Jabodetabek.

Preprocessing Data

Preprocessing aims to convert data into a more appropriate and structured format to facilitate and optimize the classification process.

Data Cleaning

Before proceeding to the modelling stage, data cleaning must be performed to ensure the data is suitable for processing. The purpose of data cleaning is to reduce noise, ensure data quality, and prepare the data for the next steps so that the classification results are more accurate and the risk of errors is minimized. In this study, the data cleaning process was carried out to detect missing values from the research instruments. The following are the stages of the data cleaning process as well as the data cleaning classification program code. Missing values are detected by empty or infinite entries or NaNs [13].

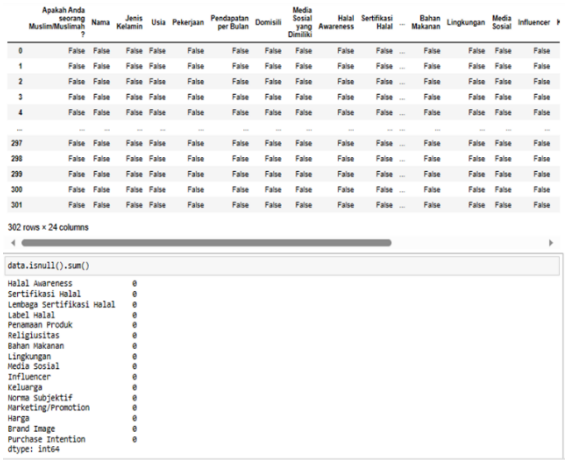

Figure 4. Output data cleaning

Based on Figure 4, it can be seen that the number of columns is 24 and the number of rows is 302. This means that the columns are all the variables and research subjects, while the rows are data from the research subjects, in this case the respondents' answers to the research instrument. In the figure it can be seen that after data cleaning was carried out to detect missing values, the result was that no missing values were found in the data, so no additional data was needed because there was no empty data. Furthermore, the data can be processed using data reduction using feature selection.

Data Reduction

Data on the decision to buy East Asian food products taken by distributing questionnaires are then entered at the data selection stage. This data selection aims to select attributes that are considered as attributes that influence the classification of interest in buying East Asian food products.

Table 3. Output feature selection

No	*Variables*	*Score*	*Values*
2	Halal Certification	3.08608099e+00	0.07998496
3	Halal Certification Institute	2.62854865e+00	0.10600812
15	*Brand Image*	2.01404236e+00	0.15688772

In Table 3. It can be concluded that the variables taken based on the results of data selection, namely the variable halal certification, halal certification bodies, and brand image for these variables are processed at a later stage until the evaluation of the appropriate modeling.

Data Modeling

At this stage the process of data modeling is carried out. Researchers used five classification analysis methods to find out the most appropriate modeling for decisions regarding purchasing interest in East Asian food products. The five analysis methods in question consist of Naïve Bayes classification analysis, Decision Tree, KNeighbors, Support Vector Machine, and Logistic Regression. Model testing produces accuracy values and AUC values which are used as references for each classification analysis model.

The following are the process stages for determining the level of accuracy in machine learning modeling:

1) Data collection

   Before the data is processed, the data needs to be imported using the pandas module into Jupyter Notebook, in this case the researcher uses excel data.

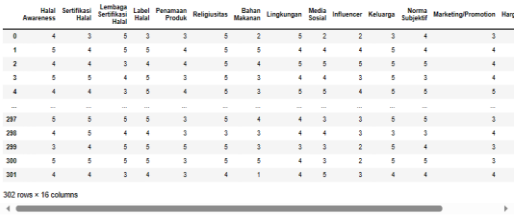

Figure 5. Output import data

2) Defining the independent variable (X) and the dependent variable (Y)
   At this stage, machine learning needs to define the independent variables and dependent variables that will be processed.

X

	Sertifikasi Halal	Lembaga Sertifikasi Halal	Brand Image
0	3	5	3
1	4	5	4
2	4	3	4
3	5	4	4
4	4	3	5
...	...	...	...
297	5	5	4
298	5	4	5
299	4	5	4
300	5	5	5
301	4	3	5

302 rows × 3 columns

Figure 6. output for defining variable X

3) Divide training data and testing data
   In this research, the author divides each data set into 80% or 0.8 training data and 20% or 0.2 testing data. Training data is used to train the classification model, while testing data is used to test the model's performance with data that is assumed to have never been 'seen' or 'digested' by the model. Separating training data and testing data can keep the testing process pure from data manipulation. The process of dividing training data and testing data or called train-test-split in this case uses the scikit-learn library.

Table 4. Train-split-test results

Information	Training Data	Data Testing	Total
Amount	3,624	906	4,530
Percentage	80%	20%	100%

In Table 4, the total data used in this research is 4,530 which was obtained from the results of respondents' answers to the research instrument, one respondent answered 15 statements with a five-point scale, then the 15 statements were multiplied by the number of respondents, namely 302 respondents. The composition of 80% of trainingdata is 3,624 data, and the remaining 906 data is for data testing.

4) Testing machine learning models using data

X_train

	Sertifikasi Halal	Lembaga Sertifikasi Halal	Brand Image
163	5	5	5
25	5	5	3
49	4	4	4
97	5	4	5
269	5	5	5
...	...	...	...
110	5	4	5
66	5	5	5
95	5	4	4
264	3	4	4
88	4	5	4

241 rows × 3 columns

Figure 7. Output variable X after train-test-split

The next stage after dividing the training data and testing data, namely classification analysis using five classification methods. The classification methods tested in this study are Naïve Bayes, Decision Tree, KNeighbors, Support Machine Vector (SVM), and Logistic Regression classifications. The modeling process uses the skicit-learn library.

5) Testing the accuracy level of each machine learning model uses testing data and predictive data
   Furthermore, each model is trained using training data that has gone through the process of data training and data testing before. Model testing produces an f1 score that researchers use to compare the performance of each model using the skicit learn library.

Naïve Bayes Classification Analysis

Naïve Bayes is a simple probabilistic classification method based on Bayes' Theorem which is carried out through training sets of a number of data efficiently. Naive Bayes assumes that the value of an input attribute in a given class does not depend on the values of other attributes [14].

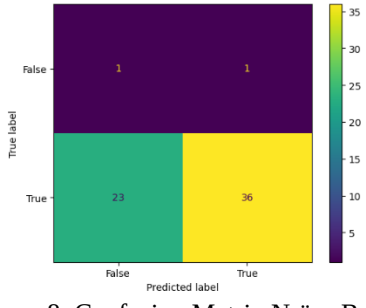

Figure 8. Confusion Matrix Naïve Bayes

Based on Figure 8, there are 61 data included in the testing data used to compile the confusion matrix. Based on the confusion matrix, it can be seen that the naïve Bayes model predicts the purchase interest decision "Yes" 59 times, and predicts the purchase interest decision "No" 2 times.

	precision	recall	f1-score	support
0	0.04	0.50	0.08	2
1	0.97	0.61	0.75	59
accuracy			0.61	61
macro avg	0.51	0.56	0.41	61
weighted avg	0.94	0.61	0.73	61

Figure 9. Naïve Bayes accuracy results

The results of the accuracy calculation in Figure 9. show that the classification performance using the naïve Bayes model algorithm has a score of 0.6065573770491803 or 0.61 or the equivalent of 61%, which is able to classify 59 samples out of a total of 61 samples of testing data.

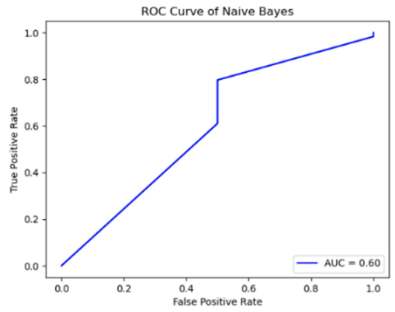

Figure 10. Naïve Bayes ROC curve

After creating the ROC curve, then to calculate and ensure the accuracy of the measurements in the naïve Bayes model, the AUC calculation is used.The naïve Bayes model produces an AUC value of 0.5974576271186441 or 0.60. This AUC value is used to determine which classifier model is superior in predicting purchasing interest decisions.

Decision Tree Classification Analysis

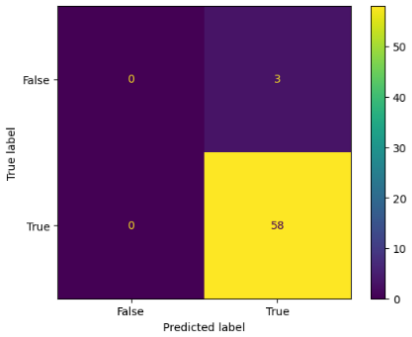

Figure 11. Confusion matrix decision tree

Based on 11, there are 61 data included in the testing data used to compile the confusion matrix. Based on the confusion matrix, it can be seen that the Decision Tree model predicts a decision of buying interest "Yes" 58 times, and predicts a decision of buying interest "No" 3 times.

	precision	recall	f1-score	support
0	0.00	0.00	0.00	3
1	0.95	1.00	0.97	58
accuracy			0.95	61
macro avg	0.48	0.50	0.49	61
weighted avg	0.90	0.95	0.93	61

Figure 12. Naïve Bayes accuracy results

The results of the accuracy calculation in Figure 12 show that the classification performance using the decision tree model algorithm has a score of 0.9508196721311475 or 0.95 or the equivalent of 95%, which is able to classify 58 samples out of a total of 61 samples of testing data.

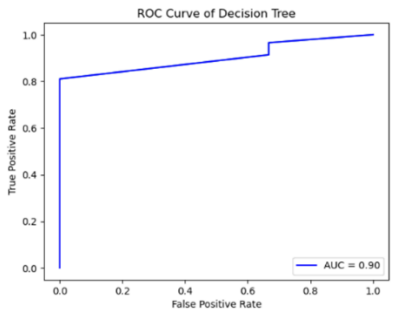

Figure 13. ROC decision tree curve

After creating the ROC curve, then to calculate and ensure the accuracy of measurements in the Decision Tree model, the AUC calculation is used. The Decision Tree model produces an AUC value of 0.9022988505747127 or 0.90. This AUC value is used to determine which classifier model is superior in predicting purchasing interest decisions.

Support Vector Machine (SVM) Classification Analysis

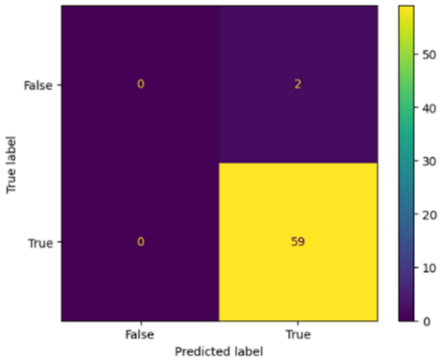

Figure 14. SVM confusion matrix

Based on Figure 14, there are 61 data included in the testing data used to compile the confusion matrix. Based on the confusion matrix, it can be seen that the SVM model predicts a purchase intention decision of "Yes" 59 times, and predicts a purchase interest decision "No" 2 times.

	precision	recall	f1-score	support
0	0.00	0.00	0.00	2
1	0.97	1.00	0.98	59
accuracy			0.97	61
macro avg	0.48	0.50	0.49	61
weighted avg	0.94	0.97	0.95	61

Figure 15. SVM accuracy results

The results of the accuracy calculation in the figure above show that the classification performance using the SVM model algorithm has a score of 0.9672131147540983 or 0.97 or the equivalent of 97%, which is able to classify 59 samples out of a total of 61 samples of testing data.

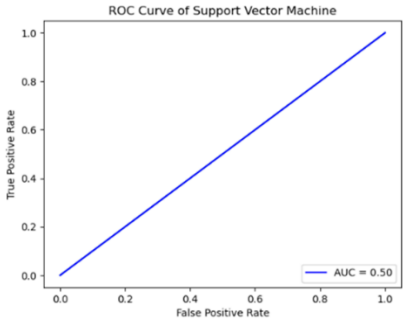

Figure 16. ROC SVM curve

After creating the ROC curve, next to calculate and ensure the accuracy of the measurements in the SVM model, the AUC calculation is used.The SVM model produces an AUC value of 0.5. This AUC value is used to determine which classifier model is superior for predicting purchasing interest decisions.

K-Nearest Neighbors (KNN) Classification Analysis

Apart from using the Naïve Bayes, Decision Tree and SVM algorithms, this research also uses the K-Nearest Neighbor (KNN) algorithm which is a type of supervised machine learning whose classification system takes into account the distance between observation data. Just as SVM has parameters for classification, k-NN also has parameters symbolized by k which will then be optimized to determine the best parameters with the best classification performance results [15].

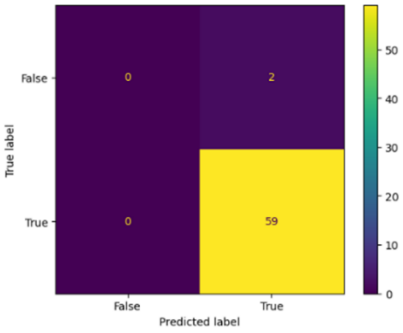

Figure 17. KNN confusion matrix

Based on Figure 17, there are 61 data included in the testing data used to compile the confusion matrix. Based on the confusion matrix, it can be seen that the KNN model predicts buying interest decisions "Yes" as many as 59 times, and predicts buying interest decisions "No" as many as 2 times.

	precision	recall	f1-score	support
0	0.00	0.00	0.00	2
1	0.97	1.00	0.98	59
accuracy			0.97	61
macro avg	0.48	0.50	0.49	61
weighted avg	0.94	0.97	0.95	61

Figure 18. KNN accuracy results

The results of the accuracy calculation in the figure above show that the classification performance using the KNN model algorithm has a score of 0.9672131147540983 or 0.97 or the equivalent of 97%, which is able to classify 59 samples out of a total of 61 samples of testing data.

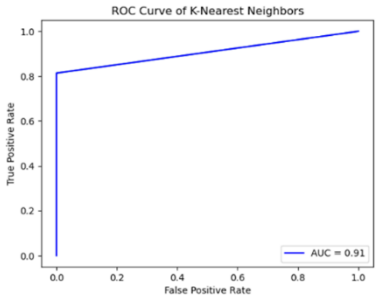

Figure 19. ROC KNN curve

After creating the ROC curve, then to calculate and ensure the accuracy of measurements in the KNN model, the AUC calculation is used. The KNN model produces an AUC value of 0.9067796610169492 or 0.91. This AUC value is used to determine which classifier model is superior in predicting purchasing interest decisions.

KNN has parameters symbolized by k which will later be validated to determine the best parameters with the best classification performance results.

Table 5. KNN parameter validation

Parameter	accuracy
K = 3	0.97
K = 5	0.97
K = 7	0.97
K = 9	0.97
K = 11	0.97

Table 5 shows the results of the k parameter validation using the KNN algorithm. These results were obtained using training data, namely 80 percent of the dataset. Based on the validation results, it was found that the accuracy value for each parameter was the same, namely 0.97 or 97% percent, meaning that each k parameter tested, namely k=3, k=5, k=7, k=9, and k=11 had the same performance.

Logistic Regression Classification Analysis

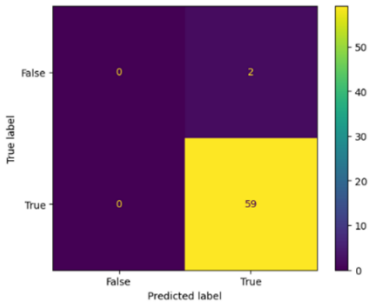

Figure 20. Logistic regression confusion matrix

Based on Figure 20, there are 61 data included in the testing data used to compile the confusion matrix. Based on the confusion matrix, it can be seen that the Logistic Regression model predicts the purchase interest decision "Yes" 59 times, and predicts the purchase interest decision "No" 2 times.

	precision	recall	f1-score	support
Tidak	0.00	0.00	0.00	2
Ya	0.97	1.00	0.98	59
accuracy			0.97	61
macro avg	0.48	0.50	0.49	61
weighted avg	0.94	0.97	0.95	61

Figure 21. Logistic regression accuracy results

The accuracy calculation results in Figure 21 show that the classification performance using the Logistic Regression model algorithm has a score of 0.9672131147540983 or 0.97 or the equivalent of 97%, namely being able to classify 59 samples from a total of 61 testing data samples.

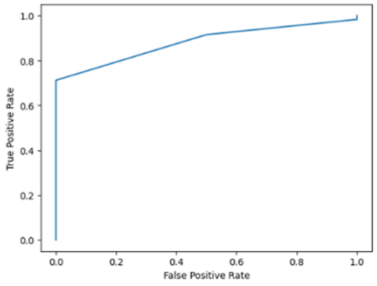

Figure 22. Logistic regression ROC curve

After creating the ROC curve, next to calculate and ensure the accuracy of the measurements in the logistic regression model, the AUC calculation is used. The Logistic Regression Model produces an AUC value of 0.8559322033898304 or 0.85. This AUC value is used to determine which classifier model is superior in predicting purchasing interest decisions.

Comparison Results of Classification Machine Learning Models

One of the goals of this research is to find out the classification machine learning algorithm model that produces the best performance in classifying research instruments, namely Naïve Bayes, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression. Therefore, the next stage is to compare the five algorithms based on their classification performance.

Table 6. Model comparison results

Algorithm	accuracy	Train-Test	AUC value	rank
KNNeighbors	0.97	0.2	0.91	1
SVM	0.97	0.2	0.5	5
Logistic Regression	0.97	0.2	0.85	3
Decision Trees	0.95	0.2	0.90	2
Naïve Bayes	0.61	0.2	0.6	4

Based on Table 6, it can be concluded that the application of the KNNeighbors algorithm model has better performance for creating purchasing interest decision models.

Model Simulation

Next, after a series of modeling stages have been carried out, the final stage is simulating the algorithm mode that has the best performance for purchasing interest decisions based on the selected

variables from the previous feature selection. Based on the comparison results of each algorithm model described above, it is known that the KNNeighbors algorithm model has the best performance among other algorithm models as seen from its AUC value. Therefore, model simulation will be applied to the KNNeighbors algorithm model. The input values used to simulate the model come from respondents' answers in the form of a five-point Likert scale on selected independent variables, namely Halal Certification, Halal Certification Institute, and Brand Image.

```
Sertifikasi Halal :5
Lembaga Sertifikasi Halal :2
Brand Image :1
[1]
```

Figure 23. Example of model simulation on KNN

Based on Figure 23, it can be seen that when the input value is entered into the modeling, it will produce a value of 1, meaning that with this input value, the predicted output is the answer "Yes" to the purchasing decision.

CONCLUSION

1. The research instrument used as a reference in this study needs to pass the validity and reliability test stages, so that it can be known whether the research instrument is valid and reliable or not. The results of the instrument validity test in this study show that each indicator in the research instrument can be used to measure research variables because the answers from the respondents who are used as samples can be trusted. The results of the reliability test of this research instrument show that it is reliable and suitable to be used as a reference in research.
2. Consumer behavior regarding interest in purchasing East Asian food products based on halal awareness and religious values can be determined by the distribution of research instrument data which shows that the majority of respondents answered agree and strongly agree with the research indicator statements. This means that from 15 research indicator variables based on halal awareness and religiosity values, researchers concluded that the majority of respondents were quite aware of halal awareness for their decision to purchase interest in East Asian food products.
3. Data from research instruments from Muslim/Muslim women respondents, domiciled in Jabodetabek, who have one or more active social media on interest in buying East Asian food products based on halal awareness and religiosity values which are processed using a machine learning classification algorithm through feature selection, it is known that the indicators The most underlying consumer decisions regarding interest in purchasing East Asian food products are indicators of halal certification, halal certification institutions, and brand image or product brand image.
4. Modeling the decision to buy East Asian food products using a classification machine learning algorithm produces a decision model that has the highest level of accuracy, namely the KNNeighbors model with an accuracy result of 0.97 and an AUC value of 0.91, then the Decision Tree model with an accuracy result of 0, 95 and the AUC value is 0.90. Furthermore, the Logistic Regression model with an accuracy of 0.97 and an AUC value of 0.85. then the Naïve Bayes model with an accuracy value of 0.61 and an AUC value of 0.6. finally, the Support Vector Machine (SVM) model with an accuracy of 0.97 and an AUC value of 0.5. So it can be concluded that the decision model that has the best performance for the classification of the decision to buy East Asian food products is the K-Nearest Neighbors model.
5. Model simulation is carried out last after the calculation stage of the accuracy value and AUC value of each model is completed. Based on the results of the comparison of each algorithm model described above, it is known that the KNNeighbors algorithm model has the best performance among other algorithm models as seen from its AUC value. Therefore, model simulation will be applied to the KNNeighbors algorithm model. However, the author also presents model simulations on other algorithm models, namely the Naïve Bayes algorithm.

SUGGESTION

1. Future research can use other variables or add new variables that are more complete and more detailed regarding consumer buying behavior based on halal awareness so that the results are more precise.

2. To be able to produce more interesting results, researchers suggest changing and or adding research objects using non-Muslim consumers.
3. Future research can use other modeling trials in data processing. This is useful for contributing new findings in the field of science.

ACKNOWLEDGMENT

Completion of this Final Assignment would not have been completed well without help, support and prayers from various parties.On this occasion, with sincerity and deep gratitude, the author would like to express his deepest gratitude and appreciation to all parties who have helped researchers so that this research can be completed properly.

REFERENCES

[1] T. P. L. Bulan and K. Fazrin, "Pengaruh Label Halal dan Bonus dalam Kemasan terhadap Keputusan Pembelian Produk Kinder Joy pada Masyarakat Kota Langsa," *J. Manaj. dan Keuang.*, vol. 6 (2), 2017.

[2] K. Ayudya Wardani and I. Nurfaizah, "Tren Makanan Sebagai Dampak Globalisasi Terhadap Pertukaran Budaya," *Pros. Konf. Ilm. Pendidik.*, vol. 3, pp. 2963–3222, 2022, [Online]. Available: https://www.google.com/amp/s/docplayer.info/amp/72666219-Kualitatif-bogdan-dan-taylor-.

[3] E. Sukmasari, "Pengaruh Asia Timur dalam Tren Kuliner di Indonesia," 2019. https://www.cultura.id/pengaruh-asia-timur-dalam-tren-kuliner-di-indonesia (accessed Sep. 01, 2023).

[4] U. Triana, "Pengaruh Sertifikasi Halal, Kesadaran Halal, Bahan Makanan, dan Citra Merek Produk terhadap Konsumsi Produk Makanan Halal," *J. Ilm. Mhs. FEB Univ. Brawijaya*, vol. 9 (2), 2021.

[5] A. Amini, M. I. Fasa, and S. Suharto, "Urgensi Halal Food Dalam Tinjauan Konsumsi Islami," *LIKUID J. Ekon. Ind. Halal*, vol. 2, no. 2, pp. 1–14, 2022, doi: 10.15575/likuid.v2i2.16031.

[6] N. M. Listiani, "Pengaruh Kreativitas Dan Motivasi Terhadap Hasil Belajar Mata Pelajaran Produktif Pemasaran Pada Siswa Kelas XI SMK Negeri 2 Tuban," *J. Ekon. Pendidik. Dan Kewirausahaan*, vol. 2(2), p. 263, 2017.

[7] E. Marlina, "Pengembangan Model Pembelajaran Blended Learning Berbantuan Aplikasi Sevima Edlink," *J. Padegogik*, vol. 3(2), 2020.

[8] W. L. Neuman, *Basic of Social Research : Qualitative & Quantitative Approaches Third Edition*. Pearson, 2014.

[9] Maryuliana, I. M. I. Subroto, and S. F. C. Haviana, "Sistem Informasi Angket Pengukuran Skala Kebutuhan Materi Pembelajaran Tambahan Sebagai Pendukung Pengambilan Keputusan Di Sekolah Menengah Atas Menggunakan Skala Likert," *J. Transistor Elektro dan Inform.*, vol. 1, no. 2, pp. 1–12, 2016.

[10] E. Hertanto, "Perbedaan Skala Likert Lima Skala dengan Modifikasi Skala Likert Empat Skala," *J. Metodol. Penelit.*, 2017.

[11] J. F. et al Hair, . *Multivariate Data Analysis 7th Edition*. England, 2010.

[12] I. Ghozali, *Structural Equation Modeling SEM Metode Alternatif menggunakan PLS.* Badan Penerbit Universitas Diponegoro, 2006.

[13] E. Arifina, B. Mursito, and D. Hermawati Wahyuningsih, "Consumer Loyalty Reviewed From Brand Image, Social Media, Lifestyle and Product Innovation (Study on iPhone Users in Solo)," *Int. J. Econ. Account. Res. Peer Rev. J.*, vol. 7, no. 1, pp. 1–7, 2023, [Online]. Available: https://jurnal.stie-aas.ac.id/index.php/IJEBAR.

[14] A. Fauzi, F. F. Wati, I. Sulistyowati, M. F. Akbar, E. Rahmawati, and R. K. Sari, "Penerapan Metode Machine Learning Dalam Memprediksi Keberhasilan Panggilan Telemarketing Menjual Produk Bank," *Indones. J. Softw. Eng.*, vol. 6, no. 2, pp. 213–222, 2020, doi: 10.31294/ijse.v6i2.8977.

[15] H. Al Aziez and G. Anuraga, "Klasifikasi Daerah Tertinggal di Indonesia Menggunakan Algoritma SVM dan k-NN," *J. Ilmu Dasar*, vol. 22(1), pp. 31–38, 2021.