Journal of Namibian Studies, 35 S1 (2023): 3055-3070 ISSN: 2197-5523 (online)

Predicting The Consumer Purchase Behavior Of Organic Food Using Decision Tree Algorithm

Aswanth C[1] , Dr. Rampilla Mahesh[2] , Alex Thomas[3]

[1]Department of Commerce and Management
Amrita School of Arts and Sciences, Mysuru
Amrita Vishwa Vidyapeetham, India
aswanthc422@gmail.com
[2]Department of Commerce and Management
Amrita School of Arts and Sciences, Mysuru
Amrita Vishwa Vidyapeetham, India
rammahesh333@gmail.com
[3]Department of Commerce and Management
Amrita School of Arts and Sciences, Mysuru
Amrita Vishwa Vidyapeetham, India
Alexthomas2706@gmail.com

Abstract:

Consumers have become more conscious of health and diet. Consequently, they are interested to take better care of their health by having healthier food. Thus, the consumer purchase behavior of organic food is an important study area for businesses and researchers nowadays. This study presents an approach to predicting consumer purchase behavior of organic food using the decision tree algorithm. Decision trees are a popular machine-learning technique that can provide insights into consumer decision factors. The present study focused on demographic characteristics, purchase behavior, product satisfaction, product knowledge from media, health consciousness, and food safety of university students in tier 2 cities in Karnataka. A decision tree model can be built and evaluated by collecting relevant data on consumer demographics, income levels, health consciousness, environmental awareness, and previous organic food purchase history. The insights into the decision tree structure provide valuable information to businesses seeking to recognize the features motivating organic food purchases and prepare their strategies.

keywords: Decision tree, organic food, Consumer Purchasing behavior, Product satisfaction, Product advertisement, Perceived barriers.

I. INTRODUCTION

In recent years, the demand for organic food has grown significantly, and it is expected that this trend will continue for the coming years [1][2]. Thus, the rising customer demand for healthy and organic food products is responsible for this expansion. Organic food products are defined as those grown or processed without synthetic pesticides, fertilizers, and other harmful chemicals [3][4]. Especially, Young people's concerns about their health and safety are increasing their interest in organic food products [5][6]. The organic farming industry eliminates consumers' concerns about conventional food production. It makes use of certain organic products that are higher in vitamins, minerals, antioxidants, and nutrients than regular products, helping to build a stronger immune system and lowering the risk of cancer, heart disease, and high blood pressure [7]. Consequently, organic products ensure good for human health. On the other hand, organic farming practices ensure environmental sustainability [8]. As a result of these, **c**onsumers have become more conscious of health and diet. Consequently, they are interested to take better care of their health by having healthier food [9].

At the same time, consumers' lifestyles have been changing. Therefore, consumers have preferred to consume nutritious foods due to the changes in Lifestyle and awareness [10][11] Nutritional and wellness food has significant health advantages over other foods [12] On the other hand, Consumers are interested in consuming ecological products because they like to preserve the environment for the next generations [13]. Consequently, consumers are ready to stop purchasing products that lead to pollution and switch to organic products [14]. However, organic food products are costlier than other products. So, it will only be affordable to some consumers. Thus, increasing organic production is necessary with decreased trading expenses, which may lower sales prices and boost additive free goods consumption [15]. On the other hand, organic food products are less available in the market where consumers demand them [16]. Furthermore, the perceived quality of traceability information influences the consumers' intentions to buy organic food, but individual differences modify this causal effect. Therefore, it is necessary to assess characteristics like past behavior, knowledge, gender, and wealth [17].

Artificial intelligence plays a predominant role in this dynamic market. Customer analytics can employ consumer related data to estimate consumers' purchasing patterns using big data sources that produce information patterns showing significant insights about consumer behavior [18][19]. Therefore, firms can anticipate sales, market optimization, inventory planning, fraud detection, and more [20]. Consequently, Big data driven consumer analytics are increasingly used in strategies and tools that enable businesses to predict the purchasing habits of consumers [21], which can favorably accelerate operational and strategic value. Thus, Machine learning predicts consumer behavior through information patterns. So, it is high time for marketers and other institutions to be aware of the significance of the customers' attitudes toward organic food products [22][23]. Overall, this research paper seeks to contribute to the existing literature on organic food products and consumer behavior and to inform the development of marketing strategies that target young consumers. So, the present study aims to find the consumer purchase intention of tier 2 cities using the decision tree algorithm.

II. RELATED WORK

Consumer health consciousness drastically raised. Therefore, they prefer organic food for consumption; consequently, organic food demand is rising eventually. Due to the demand increase, sellers are interested in consumer behavior. However, the consumer mind is like a black box, making predicting difficult. Even though, Numerous techniques are available to predict consumer behavior. Researchers are especially employing machine learning techniques. Below is a list of recent contributions in this field.

The present organic vegetable industry's supply chain is expanded. [13] stated that health-related factors, past purchasing habits, knowledge, cost, and trust in the label's organic certification are the main variables affecting the purchase of organic goods. Thus, consumers believe regularly purchasing organic food worthwhile and enjoyable. Consumers' awareness of organic food, information about it, worries about food safety, and the value of organic food all influence them to buy [24]. Similarly, the association between perceived uncertainty and purchase intention is positively moderated by product knowledge [17][15]. Furthermore, psychological variables influence the decision-making process a consumer uses to purchase food goods. Therefore, consumers purchase a specific food item for their health and wellness. More specifically, [5] Middle-aged, middle-class, university educates prefer to consume organic food. So, [8] the researchers found that relationship between organic attitudes, subjective norms, environmental concerns, and organic knowledge. While familiarity was the only component significantly related to organic purchasing behavior, other variables influenced purchase intentions, including familiarity, quality, subjective norms, and health consciousness. On the other hand, availability, health consciousness, and education positively affect customer attitudes toward organic food. Customers are generally happier with organic food than inorganic food. Consumers tend to be conscious of environmental and public health issues [11].

Additionally, safety, innovation, and quality are the primary characteristics influencing customers' perceptions. These elements have been attained by using a qualitative word root analysis technique [25]. The machine Learning Model predicts consumer behavior in a particular product based on demographic information [26]. They also train our data using ten classification techniques and assess the model's performance. Random Forest, Decision Tree Classifier, and Stochastic Gradient Descent are the most accurate methods. [27] Decision trees provided more accurate results than logistic regression and the study is important because it identifies problems that might occur when machine learning is used to make marketing decisions online and offers solutions. prior studies on Machine learning for Predictive Reasons claim that a variety of machine learning approaches may enhance marketing decision-making [28][21].

The decision tree, which is highly intuitive and simple to comprehend, starting at the decision node and ending at the terminal node and this is how the decision tree was trained for variable selection. Finally, [15] Producers can develop more effective marketing plans for product features, market attributes, and psychographics.

III. PROPOSED METHODOLOGY

This research aims to develop a model that can forecast consumer behavior. We must go through several steps, including data gathering, preprocessing, model implementation, and so forth, to achieve our aim. The workflow for the study is shown in Figure. 1.

Figure. 1. Workflow diagram - Process of Decision tree Algorithm

A. Training points and Data set:

Predicting consumer purchase behavior for organic food products has increased prominence due to increased demand from the younger generation. Therefore, Data was collected from university students in tier 2 cities in Karnataka using the convenient sampling method. A structured questionnaire is used to collect information on consumer demographics, past purchases, and preferences for organic food to predict purchase behavior. From the sample, 346 responses in total were gathered. However, 228 responses were considered for the analysis due to the missing values. Table 1. Shows the independent variable, dependent variable, training dataset, and testing dataset.

Table 1. Training points and Data set

Independent variable	**Dependent Variable**	**Training dataset**	**Testing dataset**
Gender, Age, Education, Family type, Family income, Satisfaction, Media exposure, Health and wellness, Stores availability, Price	Purchasing history	182	46

B. Variable Selection:

Identify the most relevant attributes by a literature review that influence consumer purchase behavior for organic food. The following are variables.

Consumer behavior organic food products: Consumer behavior regarding organic food products refers to consumers' attitudes, preferences, and purchasing patterns. It is driven by health and nutrition concerns, environmental sustainability, ethical considerations, and the desire for natural and chemical-free products. Price sensitivity and value perception are important considerations when choosing organic food products. Consumer behavior regarding organic food products refers to consumers' attitudes, preferences, and purchasing patterns. It is driven by health and nutrition concerns, environmental sustainability, ethical considerations, and the desire for natural and chemical-free products. Price sensitivity and value perception are important considerations when choosing organic food products.

Media exposure: Advertising organic food products through media channels aims to promote the unique qualities and benefits of organic foods to consumers. Effective advertising requires conveying a unique value proposition, connecting with target consumers' values, and differentiating the brand from competitors [3].

Health consciousness: Organic food consumers prioritize their health by opting for organic products that offer higher nutritional value and are free from synthetic pesticides and herbicides. They are concerned about reducing their exposure to harmful chemicals and pesticide residues in conventional foods [15].

Food safety: Organic food products are safe and undergo rigorous testing and certification processes to meet regulatory requirements. These processes involve regular inspections and compliance checks to ensure adherence to strict standards [24].

Perceived barriers: Perceived barriers to consuming organic food products can vary among individuals, but some common factors include thefollowing:

Price: One of the primary barriers is the perception that foods produced organically are more expensive than those made traditionally. The higher cost is often attributed to organic farming practices, certification fees, and lower economies of scale.

Limited Availability: Some consumers may perceive a lack of accessibility and limited availability of organic food products in their local area. It can be due to fewer organic food retailers or limited organic options in conventional grocery stores [3].

Model Selection:

Numerous algorithms are available in machine learning for building the prediction model. In this case, a decision tree algorithm is selected due to its ability to handle categorical variables and interpretability.

Decision Tree Algorithm

This method, which is a member of the Supervised Learning family, is unique in that it may be used broadly for issues involving regression and classification. The primary goal is to utilize the dataset to build a training model, which will then be used to predict the assigned class from the dataset and produce the output for that prediction [29]. Businesses can gain insights into customers' preferences, identify patterns, and make informed predictions about food purchasing behaviors using the decision tree algorithm. This information can be valuable for marketing strategies, inventory management, and personalized recommendations to enhance customer satisfaction and optimize business operations.

Decision trees differ from other machine learning models in several ways. Unlike linear models, decision trees can capture nonlinear relationships and handle complex interactions among features. They also provide interpretable models, allowing users to understand the decision making process and trace predictions through the tree structure. In contrast to neural networks, decision trees are more transparent, requiring less computational power for training and prediction. Decision trees can handle numerical and categorical data without extensive preprocessing, making them more flexible.

Additionally, decision trees offer feature importance measures, enabling the identification of critical factors influencing predictions. Compared to instance based models like k-nearest neighbors, decision trees provide a compact representation of the decision boundaries, making them more efficient for large datasets. These unique characteristics make decision trees a valuable tool in machine learning, suitable for various prediction tasks.

Formula: Precision = True Positives / (True Positives + False Positives)

Recall = True Positives / (True Positives + False Negatives)

Here's a breakdown of the terms used in the formulas:

- True Positives (TP): The number of correctly predicted positive instances (e.g., correctly predicted instances where food was purchased).
- False Positives (FP): The number of instances incorrectly predicted as positive (e.g., instances predicted as food purchased when it was not).
- False Negatives (FN): The number of instances incorrectly predicted as negative (e.g., instances predicted as not purchasing food when they did).
- True Negatives (TN): The number of correctly predicted

negative instances (e.g., correctly predicted instances where food was not purchased).

IV. RESULTS AND DISCUSSION

Figure 2 shows the correlation matrix. It allows us to identify features and make informed decisions about feature selection. It is often beneficial to choose less correlated features to maintain model simplicity and reduce the risk of overfitting. The below matrix shows the correlation between the features.

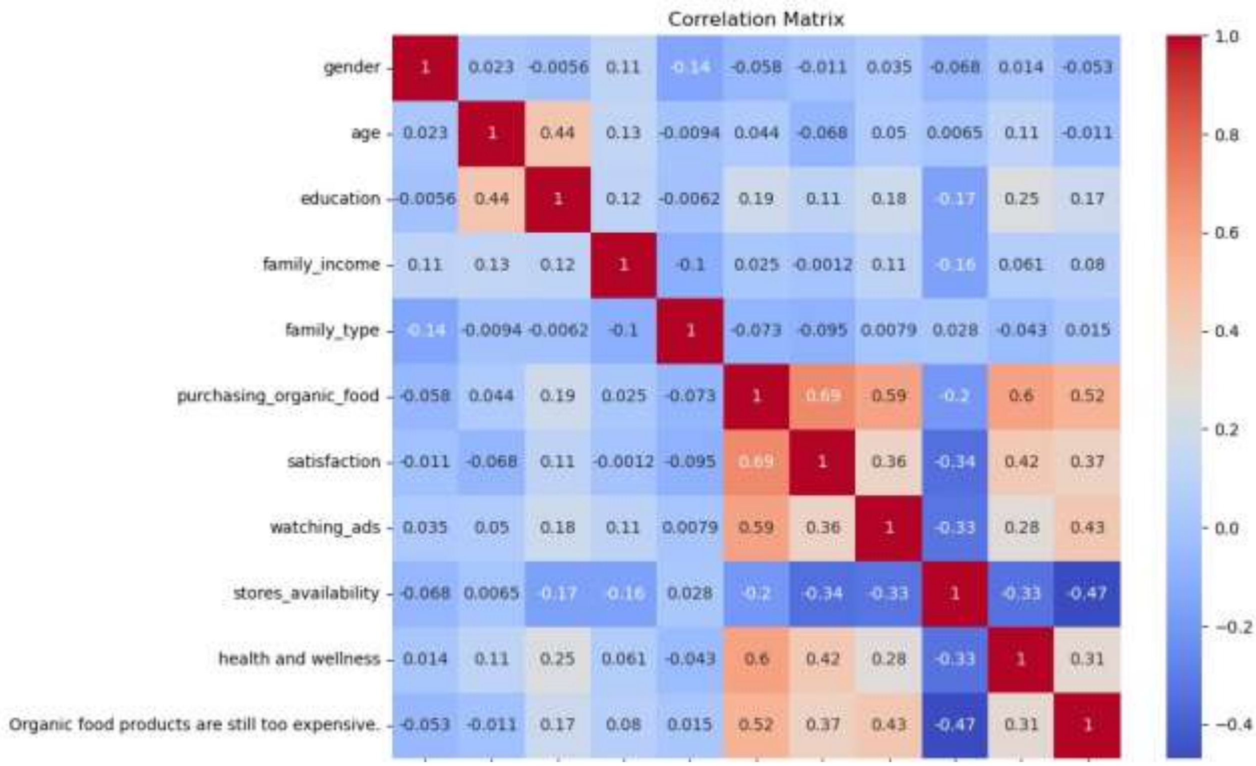

Figure 2. Correlation matrix

Table 2. VIF values

Sl.no	Feature	VIF
1	Gender	1.664041
2	Age	3.856648
3	Education	1.795824
4	Family income	2.613801
5	Family type	1.230366
6	Satisfaction	3.330684
7	Watching ads	2.893471
8	Stores availability	1.715052

9	Health and wellness	2.566613
10	Organic food is expensive	3.311265

In table 2 refers the Variance Factor (VIF) measures multicollinearity. A high VIF value indicates high multicollinearity, suggesting that the predictor variables are highly correlated. From the below table, we can find VIF values. The average VIF value across all the features is 2.4978. This indicates that, on average, the model's predictor variables are moderately correlated. VIF values below 5 are generally acceptable, suggesting that multicollinearity is not a significant issue in this case. Therefore, there is no severe multicollinearity present among the predictor variables.

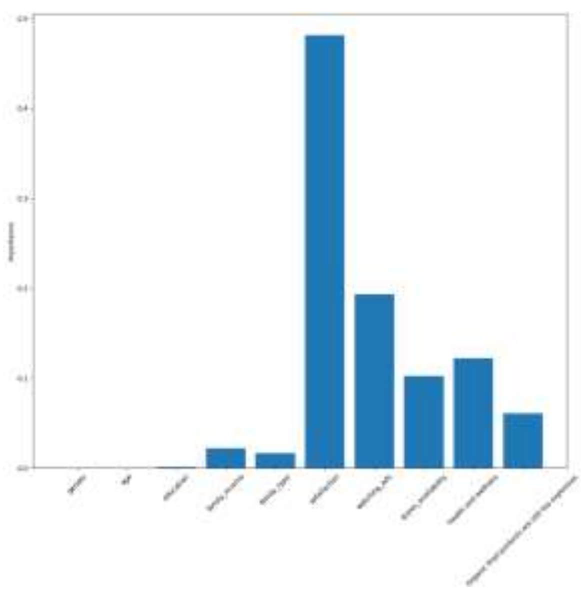

Figure 3. Decision Tree Feature Importance

In figure. 3 shows the importance value of each factor considered for prediction. Here, it is clear that the most important factors are satisfaction, watching ads, store availability, health-wellness and organic food is expensive. The remaining variables, however, are of limited significance in prediction.

Confusion matrix

In figure. 4 The accuracy score is 0.983, indicating that the model correctly predicted 98.3 of the instances in the dataset. The confusion matrix summarizes the model’s predictions compared to the actual labels. True Negative (TN): In the upper-left cell of the confusion matrix, 107 instances were correctly predicted as the negative class (Class 0). False Positive (FP): In the upper right cell, 0 instances were incorrectly predicted as the positive class (Class 1). True Positive (TP): In the lower right cell, 119 instances were correctly predicted as the positive class (Class 1).

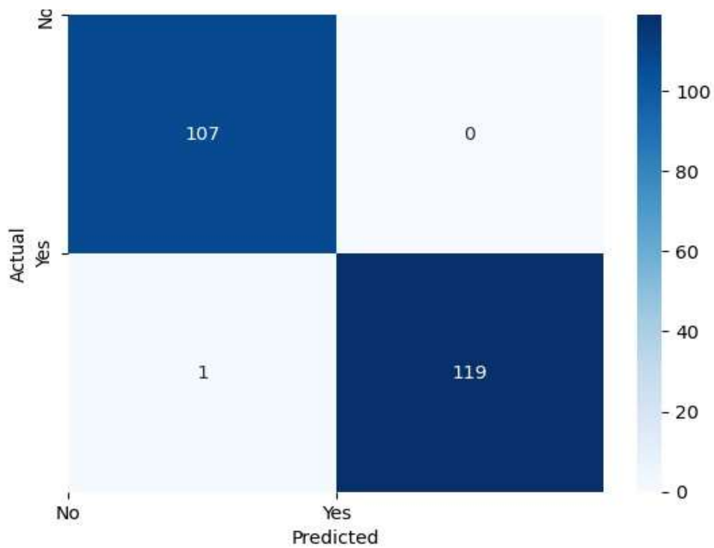

Figure 4 Confusion matrix

Table 3. Decision Tree Classification Report

	Precision	Recall	F1-score	Support
0	0.84	1.00	0.91	16
1	1.00	0.90	0.95	30
Accuracy			0.93	46
Macro avg	0.92	0.95	0.93	46
Weighted avg	0.95	0.93	0.94	46

In table. 3 refers to the comprehensive evaluation of the model's performance in each class. For Class 0, Precision: The precision for Class 0 is 0.84. Recall: The recall for Class 0 is 1.00. F1-score: The F1-score for Class 0 is 0.91. For Class 1, Precision: The precision for Class 1 is 1.00. Recall: The recall for Class 1 is 0.90. F1-score: The F1-score for Class 1 is 0.95.

Results show that the model achieved high precision and recall for both classes. Class 0 has a slightly lower recall than Class 1, indicating that a few instances of Class 0 might have been misclassified. However, the overall F1 scores for both classes indicate strong performance.

Table 4. Cross-Validation

Cross-Validation Scores:
[0.95652174 0.91304348 0.95555556 0.93333333 0.95555556]
Mean Accuracy: 0.94

In table 4 The cross-validation scores indicate the model's accuracy across multiple folds (5 in this case) of the dataset. The cross-validation scores range from 0.913 to 0.956, with a mean accuracy of 0.94. These scores reflect the model's performance on different subsets of the data and provide an estimateof its generalization ability. Overall, the decision tree classifierdemonstrates strong performance, with high precision, recall, and F1 scores for both classes. The accuracy scores fromcross-validation further support the model's effectiveness in generalizing to unseen data.

Receiver Operating Characteristic Curve

In Figure. 3 AUC of 0.93 implies that the model has demonstrated strong performance in accurately classifying instances and making predictions, with a high likelihood of correctly identifying positive instances while minimizing false positives.

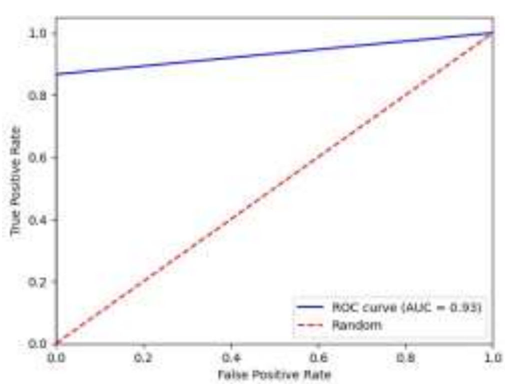

Figure 5. Receiver Operating Characteristic Curve

V. CONCLUSION

Overall, predicting the consumer behavior of organic food products gained more emphasis in the present market. Thus, Marketing researchers make many substantial contributions to consumer purchase behavior and inform the sellers to develop marketing strategies targeting young consumers. The present study aimed to find the consumer purchase intention of tier 2 cities using the decision tree algorithm. This proposed algorithm examines several factors, such as age, gender, education, family income, family type, satisfaction, watching ads, store availability, health wellness, and the cost of organic food. The average VIF value across all the features is 2.4978, indicating that the model's predictor variables are moderately correlated. The decision tree classifier demonstrates strong performance, with high precision, recall, and F1 scores for both classes. The cross-validation scores range from 0.913 to 0.956, with a mean accuracy of 0.94. An AUC of 0.93 implies that the model has demonstrated strong performance in accurately classifying instances and making predictions. Finally, the decision tree showed the important features as satisfaction, watching ads, store availability, health wellness, and organic food being the most critical factors. Consequently, we can use these important features to predict the consumer behavior of organic food products and policy makers can prepare strategies. The main challenges of decision trees are susceptibility to bias towards features with high cardinality or those that appear earlier in the tree, potentially affecting the fairness and accuracy of the predictions. Addressing these challenges involves techniques such as pruning, regularization, ensemble methods, and careful feature engineering to improve the robustness and reliability of decision tree analysis. However, several variables that need to be tested include self-identity, subjective norms, food taste, and purchase attitude. It is one of the limitations of our study.

Reference

[1] G. N. Lodorfos and J. Dennis, "Consumers' intent: In the organic food market," Journal of Food Products Marketing, vol. 14, no. 2, pp. 17–38, May 2008, doi: 10.1080/10454440801918218.

[2] A. R. Da Liang, "Enthusiastically consuming organic food:

An analysis of the online organic food purchasing behaviors of consumers with different food-related lifestyles," Internet Research, vol. 24, no. 5, pp. 587–607, Sep. 2014, doi: 10.1108/IntR-03-2013-0050.

[3] T. H. Pham, T. N. Nguyen, T. T. H. Phan, and N. T. Nguyen, "Evaluating the purchase behaviour of organic food by young consumers in an emerging market economy," Journal of Strategic Marketing, vol. 27, no. 6, pp. 540–556, Aug. 2019, doi: 10.1080/0965254X.2018.1447984.

[4] R. Saleki, F. Quoquab, and J. Mohammad, "Business and Management Research," 2020.

[5] R. Rodríguez-Bermúdez, M. Miranda, I. Orjales, M. J. Ginzo-Villamayor, W. Al-Soufi, and M. López-Alonso, "Consumers' perception of and attitudes towards organic food in Galicia (Northern Spain)," Int J Consum Stud, vol. 44, no. 3, pp. 206–219, May 2020, doi: 10.1111/ijcs.12557.

[6] L. Zepeda and D. Deal, "Organic and local food consumer behaviour: Alphabet theory," Int J Consum Stud, vol. 33, no. 6, pp. 697–705, 2009, doi: 10.1111/j.1470-6431.2009.00814.x.

[7] L. H. Essoussi and M. Zahaf, "Decision making process of community organic food consumers: An exploratory study," Journal of Consumer Marketing, vol. 25, no. 2, pp. 95–104, 2008, doi: 10.1108/07363760810858837.

[8] S. Smith and A. Paladino, "Eating clean and green? Investigating consumer motivations towards the purchase of organic food," Australasian Marketing Journal, vol. 18, no. 2, pp. 93–104, 2010, doi: 10.1016/j.ausmj.2010.01.001.

[9] H. Quang Doan, "Factors Affecting Customers' Organic Vegetable Buying Intentions: A Case Study in the Northern Provinces of Vietnam." [Online]. Available: https://www.researchgate.net/publication/365126683

[10] A. R. Da Liang, "Enthusiastically consuming organic food: An analysis of the online organic food purchasing behaviors of consumers with different food-related lifestyles," Internet Research, vol. 24, no. 5, pp. 587–607, Sep. 2014, doi: 10.1108/IntR-03-2013-0050.

[11] J. Paul and J. Rana, "Consumer behavior and purchase intention for organic food," Journal of Consumer Marketing, vol. 29, no. 6, pp. 412–422, Sep. 2012, doi: 10.1108/07363761211259223.

[12] E. Tsakiridou, C. Boutsouki, Y. Zotos, and K. Mattas, "Attitudes and behaviour towards organic products: An exploratory study," International Journal of Retail and Distribution Management, vol. 36, no. 2, pp. 158–175, 2008, doi: 10.1108/09590550810853093.

[13] N. Dangi, S. A. Narula, and S. K. Gupta, "Influences on

purchase intentions of organic food consumers in an emerging economy," Journal of Asia Business Studies, vol. 14, no. 5, pp. 599–620, Dec. 2020, doi: 10.1108/JABS-12-2019-0364.

[14] A. Bodini, T. Richter, and R. Felder, "Give to AgEcon Search Quality related communication approaches for organic food Quality related communication approaches for organic food." [Online]. Available: http://ageconsearch.umn.edu

[15] T. Ali, A. Alam, and J. Ali, "Factors Affecting Consumers' Purchase Behaviour for Health and Wellness Food Products in an Emerging Market," Global Business Review, vol. 22, no. 1, pp. 151–168, Feb. 2021, doi: 10.1177/0972150918795368.

[16] L. Nafees, E. M. Hyatt, L. L. Garber, and N. Das, "Exploration of the organic food-related consumer behaviour in emerging and developed economies: the case of India and the US," 2020.

[17] X. Wu, J. Xiong, J. Yan, and Y. Wang, "Perceived quality of traceability information and its effect on purchase intention towards organic food," Journal of Marketing Management, vol. 37, no. 13–14, pp. 1267–1286, 2021, doi: 10.1080/0267257X.2021.1910328.

[18] F. Taghikhah, A. Voinov, N. Shukla, and T. Filatova, "Shifts in consumer behavior towards organic products: theory-driven data analytics." [Online]. Available: http://creativecommons.org/licenses/by-nc-nd/4.0/Thedefinitivepublisherversionisavailableonlineat[https://www.sciencedirect.com/science/article/pii/S0969698921000825?via%3Dihub]

[19] J. Li, S. Pan, L. Huang, and X. Zhu, "A machine learning based method for customer behavior prediction," Tehnicki Vjesnik, vol. 26, no. 6, pp. 1670–1676, Nov. 2019, doi: 10.17559/TV-20190603165825.

[20] P. S. Aithal, "Quantitative ABCD Analysis of Organic Food Product and its Impact on Purchase Intention." [Online]. Available: https://ssrn.com/abstract=4136661

[21] A. Kundu, Institute of Electrical and Electronics Engineers, Institute of Electrical and Electronics Engineers. Kolkata Section. PES Chapter, and Institute of Electrical and Electronics Engineers. Kolkata Section, 2020 IEEE International Conference for Convergence in Engineering : proceedings : 5-6 September, Kolkata, India.

[22] E. Nica, "Buying organic food as sustainable consumer decision-making behavior: Cognitive and affective attitudes as drivers of purchase intentions toward environmentally friendly products," SHS Web of Conferences, vol. 74, p. 04018, 2020, doi: 10.1051/shsconf/20207404018.

[23] G. Hustvedt and M. A. Dickson, "Consumer likelihood of

purchasing organic cotton apparel: Influence of attitudes and self-identity," Journal of Fashion Marketing and Management, vol. 13, no. 1, pp. 49–65, 2009, doi: 10.1108/13612020910939879.

[24] T. Le-Anh and T. Nguyen-To, "Consumer purchasing behaviour of organic food in an emerging market," Int J Consum Stud, vol. 44, no. 6, pp. 563–573, Nov. 2020, doi: 10.1111/ijcs.12588.

[25] B. Prasad and I. Ghosal, "Forecasting Buying Intention Through Artificial Neural Network: An Algorithmic Solution on Direct-to-Consumer Brands," FIIB Business Review, vol. 11, no. 4, pp. 405–421, Dec. 2022, doi: 10.1177/23197145211046126.

[26] M. A. Rubi, M. Hasan Imam Bijoy, S. Chowdhury, and M. K. Islam, "Machine Learning Prediction of Consumer Travel Insurance Purchase Behavior," in 2022 13th International Conference on Computing Communication and Networking Technologies, ICCCNT 2022, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICCCNT54827.2022.9984470.

[27] D. L. Olson, D. Delen, and Y. Meng, "Comparative analysis of data mining methods for bankruptcy prediction," Decis Support Syst, vol. 52, no. 2, pp. 464–473, Jan. 2012, doi: 10.1016/j.dss.2011.10.007.

[28] J. Lee, O. Jung, Y. Lee, O. Kim, and C. Park, "A comparison and interpretation of machine learning algorithm for the prediction of online purchase conversion," Journal of Theoretical and Applied Electronic Commerce Research, vol. 16, no. 5, Aug. 2021, doi: 10.3390/jtaer16050083.

[29] N. Manohar and A. U. Archana, "Cloud-Based Flood Prediction Using IoT Devices and Machine Learning Algorithms," in Proceedings of the 2nd International Conference on Electronics and Sustainable Communication Systems, ICESC 2021, 2021, pp. 754–762. doi: 10.1109/ICESC51422.2021.9532916.