Volume XVI Number 2, 2024

Analysing Household Food Consumption in Turkey Using Machine Learning Techniques

Burak Oztornaci[1], Baris Ata[2], Serkan Kartal[2]

[1] Department of Agricultural Economics, Cukurova University, Adana, Turkey

[2] Department of Computer Engineering, Cukurova University, Adana, Turkey

Abstract

The fluctuations in food prices have highlighted the significance of analysing the factors influencing household food consumption. Recent advancements in data analysis have opened new avenues for investigating this subject. While studies have employed novel data analysis methods to examine the factors impacting household food consumption, the effect of the chosen analysis method on the research outcome remains unexplored. In this study, we aimed to investigate household food consumption in Turkey between 2012-2019 using various data analysis techniques (Linear Regression, Support Vector Machine, Random Forest, eXtreme Gradient Boosting, and Multi-Layer Perception). Our findings reveal that income emerged as the most influential factor in household food consumption across all methods. However, the impact of other factors varied depending on the method employed. This suggests that the method chosen to analyse factors other than income in studies of this nature can significantly impact the results. Researchers should exercise caution when selecting their analysis method..

Keywords

Machine learning, data analysis, food consumption, income.

Oztornaci, B., Ata, B. and Kartal, S. (2024) "Analysing Household Food Consumption in Turkey Using Machine Learning Techniques", *AGRIS on-line Papers in Economics and Informatics*, Vol. 16, No. 2, pp. 97-105. ISSN 1804-1930. DOI 10.7160/aol.2024.160207.

Introduction

Investigating household food consumption behaviours has been a long-standing area of scientific research. The pioneering studies in this field include those of David Davies and Frederick Eden in the 18th century and the work of Frederick Le Play and Edouard Ducpetiaux. Ernst Engel's contributions in the late 19th century, particularly in economics, were seminal, resulting in the formulation of Engel's law (Engel, 1857). Subsequently, numerous scientists have researched household food consumption behaviours, building upon the foundations laid by Engel.

Research on this topic has often focused on the effects of household income on food consumption. The impact of income on household food consumption has been examined by Staudigel and Schröck (2015) for Russia, Tiffin and Tiffin (1999) for England, Balli and Tiezzi (2010) for Italy, Holcomb, Park, and Capps (1995) for the USA. For worldwide, Clements and Si (2018) examined how the food composition of households is formed and what factors are affected.

The focus of these studies has generally been the impact of household incomes. However, studies in the literature examine the effects of other factors on household food consumption. For example, on issues such as food subsidies and the effects of income and price policies, Unnevehr et al. (2010); household size, Baragani et al. (2009); household location and housing tenure, Lund and Derry (1985); education level of women, Rae (1999); age, gender and etc., Göktolga et al. (2006) have been researched.

While these researches on the food consumption of households are being carried out worldwide, similar studies have been carried out in Turkey. Because it can be predicted that the food prices index, which has increased in recent years, affects the food consumption of households in Turkey. On the other hand, econometric analysis methods and multivariate statistical analysis have usually been used in these studies up to now. These studies generally used econometric analysis methods such as AIDS, the Working-Leser model, etc. However, depending on the changes in data analysis methods, it is possible to deal with the issue with new methods, such as Machine Learning (ML).

Studies using these new analysis methods, such as ML, are quite new in the literature. ML is frequently used for some estimation problems. Moreover, ML methods are also becoming widespread in food consumption research. Martini et al. (2022), Deléglise et al. (2022) are a few of the current studies using these techniques in food consumption research. The use of ML to analyze the relationship between household food consumption and other household characteristics is one of the important aspects of this study.

This study aims to investigate the factors influencing food consumption in households in Turkey using various analytical methods. By doing so, we aim to understand the impact of different analysis techniques in this field. Furthermore, our study seeks to take into account the latest global and national developments in the field and provide valuable insights and recommendations for researchers working in this area of study.

Materials and methods

Data

The data used in the research was created from the Household Budget Survey cross-section micro data set collected for 2012-2019 by the Turkish Statistical Institute (TURKSTAT). The data set contains information on 89.135 households. Household Budget Surveys is a comprehensive data set used to determine households' monthly consumption expenditures, especially in creating the consumer price index. Households are visited 8 times a year on average by TURKSTAT officers, and data are collected. Households are selected to represent Turkey. This survey includes consumption, income, and various household characteristics. This survey details the economic status of a Household, including demographic status etc. The survey also accounts for household members' age, gender, education, and employment status. Descriptive statistics of the variables used in the study are given in Table 1.

Problem formulation

Within the scope of the research, the prices of the products consumed by the households will not be dealt with. For this reason, econometric demand models such as AIDS or QUAIDS, frequently used in this type of research, are unsuitable. Working-Leser (Working, 1943; Leser, 1963), a model that does not include price data, was used to measure the effect of household income on food demand.

$$W_F = B_0 + B_1 LN_{INCOME} + E_F \qquad (1)$$

where: W_F represents the average monthly food expenditure ratio in total monthly household income (Average Monthly Food Expenditure/Total Monthly Household Income), LN_{INCOME} represents the natural logarithm of total monthly household income, and E_F represents the error term.

When other variables that are predicted to affect the food consumption of the households are added to the model, the model is established as follows:

$$W_F = B_0 + B_1 LN_{INCOME} + B_2 A + B_3 HS + B_4 SG + B_5 HO + E_F \qquad (2)$$

where: A represents the age of the head of the household, HS represents the number of people living in the household, SG represents the educational institution where the household head last graduated, HO represents whether the family owns a home.

Methods

Linear Regression (LR)

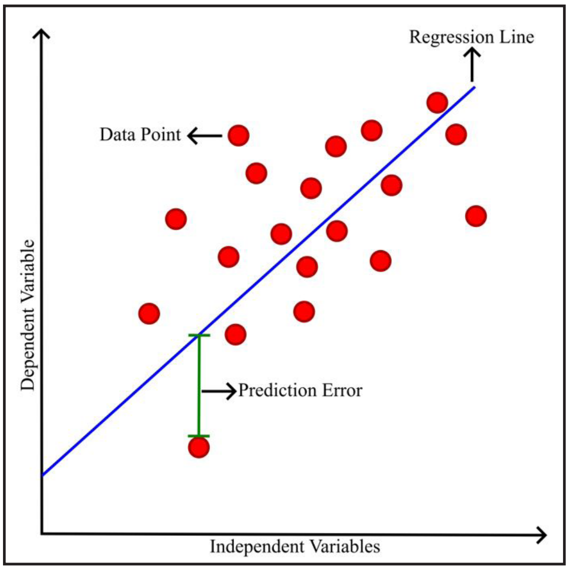
Regression is a statistical method to determine the linear and nonlinear relations between a dependent variable and one or more independent variables. The regression approach can be

Variables	Description	Mean	Std. Dev.
Household food expenditure share	Share of food expenditures in total household expenditures (%)	25.71	13.82
Income	Household total monthly income (TL)	3667.32	3575.71
Age	Age of the head of the household	50.17	14.65
School Graduation	The educational institution where the household head last graduated (1: not graduated from school; 2: primary school; 3: secondary school; 4: high school; 5: university)	2.80	1.29
Household Size	Number of people living in the household	3.53	1.83
Home Ownership	1 if the residence is property, 0 otherwise	1.24	0.42

Note: Household Budget Survey is calculated using 2012-2019 data.
Source: Author´s work

Table 1: Descriptive statistics.

named single or multiple regression according to the number of independent variables. A simple linear regression model is illustrated in Figure 1 where the lines between the regression line and data points indicate the prediction errors. The regression line here can be determined by minimizing the sum of squares of the prediction errors.



Source: Author´s work

Figure 1: Linear regression model.

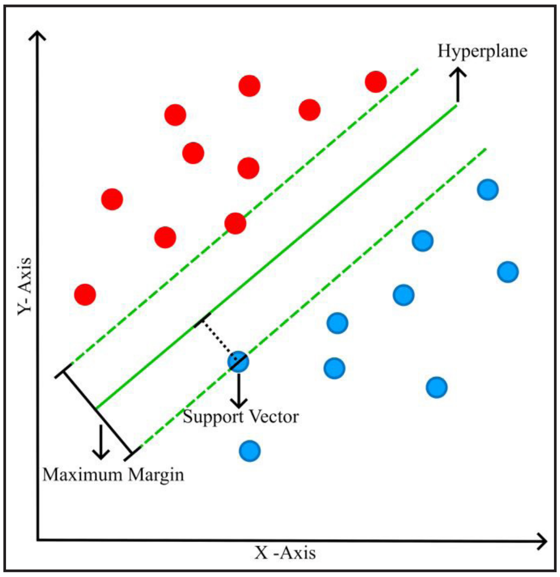
Multiple linear regression extends simple linear regression to include more than one independent variable. The basic multiple linear regression model is defined as follows

$$Y = B_x + c \quad (3)$$

where: Y is the dependent variable, $X = (1, x_1, x_2, \ldots, x_n)$ represents the independent variable, $B = (B_0, B_1, \ldots, B_n)$ represents the predicted coefficients of the independent variable, and c represents the constant. The coefficients of the independent variables can be estimated by the least square method (Grégoire, 2014).

Support Vector Machine (SVM)

The support vector machine (SVM) algorithm originated by (Boser et al., 1992) is a supervised machine learning method that is based on the statistical learning theory, and it can be applied to both linear and non-linear datasets for classification and regression purposes (Trafalis and Gilbert, 2007; Xu et al., 2009). Assuming that the number of features in the dataset is n, the SVM algorithm maps each item in the dataset into an n-dimensional feature space. Then it selects a hyperplane to divide items in the dataset into two separate classes. This hyperplane is selected to maximize the marginal distance, which is the distance between the selected hyperplane and the nearest data item of a class (Vapnik, 2000; Xu et al., 2009). A simplified version of an SVM classifier is illustrated in Figure 2 where the hyperplane is defined to differentiate two classes by the maximum marginal distance. By defining the hyperplane that results in possible prediction errors, the SVM algorithm can be used for regression tasks to make accurate predictions.

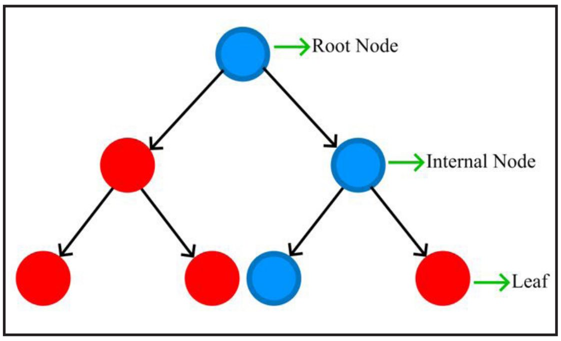


Source: Author´s work

Figure 2: A simplified illustration of the support vector machine model.

Random Forest (RF)

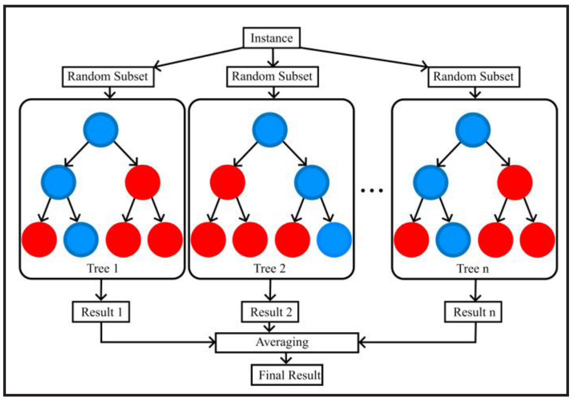
The decision tree method, which models the data items in a dataset into a tree-like structure for classifying them, can be considered one of the earliest machine learning approaches (Breiman, 1998). In this approach, the nodes on the tree-like structure have three major categories. The top-level node is called the root node; every node with at least one child node is called an internal node, and nodes with no child nodes are called a leaf node or a terminal node. All internal nodes in a decision tree represent tests on attributes of input, based on the results of these tests, the algorithm moves to the appropriate child node, and this process repeats until reaching a terminal node which represents a decision as shown in Figure 3.



Source: Author´s work

Figure 3: Decision tree.

The random forest approach can be considered as a collection of many decision trees, and it can be used for both classification and regression tasks (Breiman, 2001). Decision trees are quite sensitive in their training; this can make them error-prone to test data and lead to overfitting on training data. To prevent these drawbacks, different decision trees can be trained using random parts of the training data on the random forest approach, as shown in Figure 4. Each decision tree dives into a decision, and the RF algorithm takes the mean of these individual predictions to solve regression tasks.

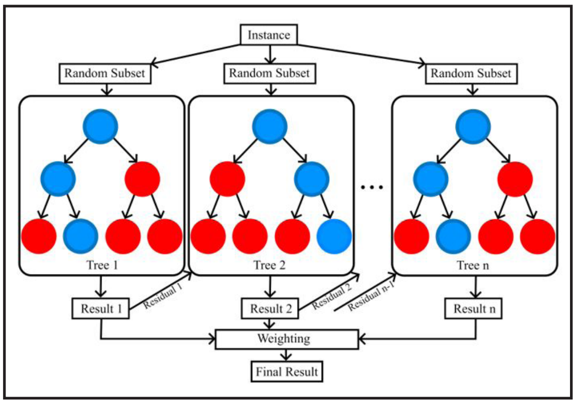


Source: Author´s work

Figure 4: Simplified structure of Random Forest.

eXtreme Gradient Boosting (XGBoost)

XGBoost stands for extreme gradient boosting, which is a very efficient ML method based on a gradient boosting algorithm proposed by (Chen and Guestrin, 2016). Basically, the XGBoost approach employs multiple DTs as in the RF algorithm. However, unlike the RF algorithm, the boosting algorithms' trees are connected sequentially. XGBoost is an iterative DT algorithm where it adds a new tree each time by learning a new function to fit the residual of the last prediction, as shown in Figure 5. When a predefined number of trees are obtained after training, XGBoost combines the predictions of each individual tree by weighting them according to their relative importance as the predicted value instead of voting.

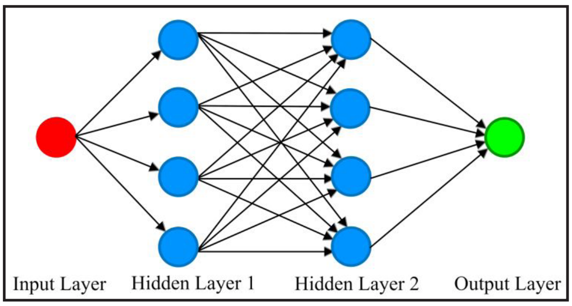


Source: Author´s work

Figure 5: Simplified structure of XGBoost.

Multi-Layer Perception (MLP)

Artificial Neural Network (ANN) models are inspired by the ability of human neurons to generate new logical results based on previous experiences (Rumelhart et al., 1986; Hill et al., 1994). Neurons in the human brain are connected to each other through axons. Similar to that graph-like architecture, the ANNs can be modelled as interconnected groups of different groups of neural networks. These types of ANNs are called as Multi-Layer Perceptron (MLP) neural networks. MLP architecture includes at least one hidden layer apart from an input layer, one or more hidden layers, and an output layer. Every layer consists of one or more nodes, and the output of one node goes as input of another node. Edges have weights, so input values are multiplied by weights while transmitting to the next layer. In the end, the output layer produces the final output by the values computed in the hidden layers. The MLP neural network updates randomly selected weight values based on the training data and can make a prediction for the test data. The general structure of an MLP neural network is illustrated in Figure 6.



Source: Author´s work

Figure 4: Simplified structure of Random Forest.

Experimental Setup

In this study, a series of experiments were conducted using LR, SVM, and MLP models, along with RF and XGBOOST models. The LR model was trained using default parameters from the scikit-learn library (Pedregosa et al., 2011), while a linear kernel was employed for the SVM model. The MLP model was configured with two hidden layers, each containing 50 neurons. Additionally, an early stopping criterion enabled to prevent overfitting.

On the other hand, RF and XGBOOST models fine-tuned using the Fast Library for Automated Machine Learning & Tuning (FLAML) developed by Microsoft (Wang et al.,2021). FLAML is a Python library that automates machine learning model selection and hyperparameter tuning. Instead of employing a grid search, FLAML takes the available computing time as a parameter and attempt to identify optimal hyperparameters within the allotted time.

For this study, the RF and XGBOOST models were optimized for 1 hour on a Windows workstation featuring an Intel Core i7 3.5 GHz CPU and NVIDIA GeForce GTX 1070 (8 GB) GPU. This methodology facilitated the identification of the most suitable hyperparameters such as number of trees and number of tree levels, to achieve optimal performance of the RF and XGBOOST models.

Evaluation Metrics

To assess the overall performance of the models, three widely used evaluation metrics are used, namely, Root Squared (R2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). Equations of the metrics are given below.

$$R2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (5)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i|^2 \quad (6)$$

where: y_i and $\hat{y}_i$, are the ith actual value and the ith predicted value among n samples, and $\bar{y}$ is the mean of the y values. The smallest RMSE, MAE values indicate the highest prediction accuracy.

Results and discussions

Experiments have been performed with the five ML models. Since the correlation between food consumption and household income can be directly evaluated through the coefficients acquired by the LR model, in this study, the LR model was considered as the baseline. According to these coefficients, it is possible to calculate the ratio of change in food consumption depending on the change in each household income parameter. The coefficients acquired by the LR algorithm are given in Table 2.

To evaluate the performance of the models, R2, RMSE, and MAE scores of the prediction processes carried out with both the LR algorithm and other ML algorithms are given in Table 3. The models were evaluated by performing a five-fold cross-validation. Results show that the SVM model gives almost the same results as LR, while the RF, XGBoost, and MLP models give better results.

Although it is possible to make better predictions with RF, XGBoost, and MLP algorithms, these algorithms do not provide information about how much each household income parameter directly affects the ratio of food consumption. On the other hand, it is possible to find out how much each feature affects the ML models' prediction performance. Accordingly, the magnitude of the LR coefficients

BIAS	A	SG	HO	HS	LN_{INCOME}
63.73	0.20	-0.99	-3.04	1.52	-6.75

Note: *A* - age of the head of the household; HS - the number of people living in the household; SG - the educational institution where the household head last graduated; HO - whether the family owns a home.
Source: Author´s work

Table 2: Acquired coefficients of the LR models on the food consumption dataset.

	LR	SVM	RF	XGBOOST	MLP
R2	0.25	0.24	0.28	0.29	0.29
RMSE	11.89	11.99	11.69	11.63	11.62
MAE	9.04	8.92	8.89	8.83	8.80

Source: Author´s work

Table 3: R2, RMSE, and MAE scores of the LR, SVM, RF, XGBOOST, and MLP models.

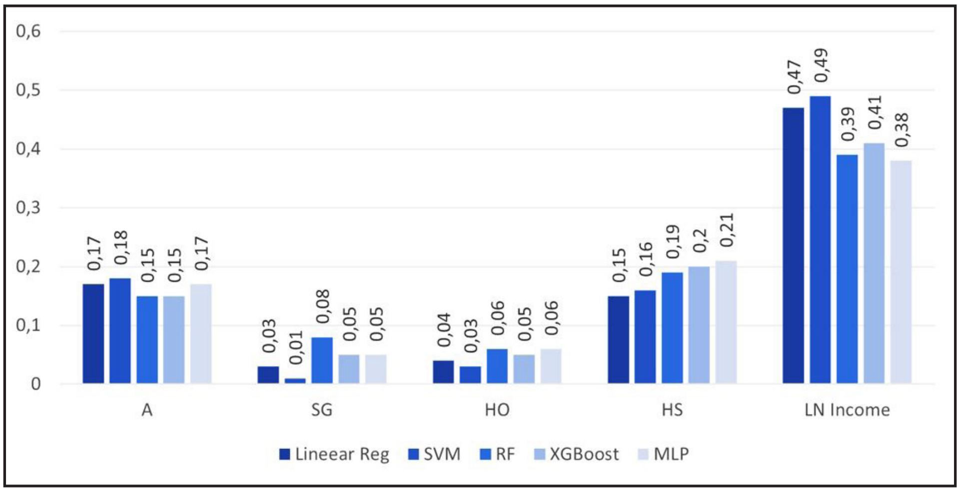
(the correlation between food consumption and household income) can be discussed depending on the feature importance scores of the LR and other ML models. In this study, the permutation importance (PI) algorithm was used to calculate the effects (importance) of the features on the prediction performance of the models. The acquired feature importance values are given in Figure 7.

With the given figure, it is seen that there is a consistency in the feature importance values; that is, all models give similar importance values to similar features. On the other hand, if the table features are examined in detail, it is clearly seen that the LNINCOME feature contributes the most to the prediction process. However, LN_{INCOME} is less important in ML models with a relatively high success rate, such as LR. Considering these values, it can be stated that LN_{INCOME} should affect the food consumption ratio less than the value given in Table 3. Furthermore, the second most important feature, *HS*, is more important in better models. Therefore, it can be concluded that *HS* affects food consumption more than the value given in Table 3. Similarly, according to the feature importance values, it can be concluded that *HO* and the *SG* features should affect food consumption more than the coefficient values given in Table 3. On the other hand, it is seen that the LR and SVM models with a low success rate and the MLP model with a high success rate give almost the same importance value to the *A* feature. Therefore, it is impossible to say anything definite about feature *A*, only considering its importance scores.

Previous studies on the factors influencing household food consumption have primarily focused on product-specific calculations, particularly those related to calorie content, making it difficult to compare their findings. However, our findings on the values of factors affecting food expenditure in Turkey are similar to those found in previous studies. Akbay et al. (2007) identified income as the most significant factor affecting household food expenditure. However, they reached varying conclusions about the impact of other factors. Our study demonstrates that the importance of factors varies depending on the analysis method used, with income being the most crucial factor. Overall, the literature suggests that income is the primary factor influencing food expenditure, but the effectiveness of other factors may vary depending on the analysis method employed.

Conclusion

In this study, the factors affecting food consumption were examined using the consumption data obtained from approximately 89 thousand households in Turkey between the years 2012-19. Consistent with Engel's Law, it was found that as the income of the households increased, their share of food expenditure decreased. It has been observed that this result is in parallel with other studies in the literature. However, the level of importance of the effects of factors other than income varies according to the analysis method. In this research, five different models, including Linear Regression (LR), Support Vector Machine (SVM), Random Forest (RF), XGBoost, and Multi-Layer Perceptron (MLP).



Source: Author´s work

Figure 7: Feature importance scores of the food consumption features according to the PI algorithm on the ML models

When these different analysis methods developed in the literature in recent years are compared, it is seen that XGBOOST and MLP models have higher R2 values and lower RMSE and MAE values compared to other methods. The LN_{INCOME} variable was calculated to be the variable with the highest significance level, consistent with previous studies. However, it was determined that the order of importance and values of other variables changed according to the analysis method. According to the LR, SVM, and MLP methods, the least effective variable on the food expenditure ratios of the households was *SG*; According to the RF method, it is the *HO* variable. According to the XGBoost method, *SG* and *HO* variables are equally important and least effective.

In conclusion, our research demonstrates that the choice of analysis method significantly impacts the results when examining household food consumption rates in Turkey. Future studies should consider utilizing new analysis methods to investigate the distribution of food consumption and the factors influencing it. Additionally, evaluating the effects of economic and food crises in the 2000s on food consumption through these novel methods could yield valuable insights for policymakers when formulating social policies. Furthermore, determining the impact of various household characteristics on food consumption using different analysis methods can assist in identifying key policy focus groups.

Corresponding author:
Serkan Kartal, Associate Professor
Department of Computer Engineering, Cukurova University, 01330, Adana, Turkey
Email: skartal@cu.edu.tr

References

[1] Akbay, C., Boz, I. and Chern, W. S. (2007) "Household food consumption in Turkey", *European Review of Agricultural Economics*, Vol. 34, No. 2, pp. 209-231. ISSN 0165-1587. DOI 10.1093/erae/jbm011.

[2] Bagarani M., Forleo M. B. and Zampino S. (2009) "Households food expenditures behaviours and socioeconomic welfare in Italy: A microeconometric analysis", Paper prepared for presentation at the *113th EAAE Seminar "A resilient European food industry and food chain in a challenging world*", Chania, Crete, Greece, date as in: September 3 - 6, 2009.

[3] Balli F. and Tiezzi S. (2010) "Equivalence scales, the cost of children and household consumption patterns in Italy", *Review of Economics of the Household*, Vol. 8, No. 4., pp. 527-549. ISSN 1569-5239. DOI 10.1007/s11150-009-9068-3.

[4] Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992) "A training algorithm for optimal margin classifiers", *COLT '92: Proceedings of the fifth annual workshop on Computational learning theory,* July 1992, pp. 144-152. DOI 10.1145/130385.130401.

[5] Breiman L., Friedman, J. H., Olshen, R. A. and Stone, Ch. J. (2017) "*Classification and regression trees*", Routledge. ISBN 9781315139470. DOI 10.1201/9781315139470.

[6] Breiman L. (2001) "Random Forests", *Machine Learning,* Vol. 45, No. 1, pp. 5-32. ISSN 0885-6125. DOI 10.1023/A:1010933404324.

[7] Chen, T. and Guestrin, C. (2016) "Xgboost: A scalable tree boosting system", *KDD '16: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794. arXiv:1603.02754. DOI 10.1145/2939672.2939785.

[8] Clements, K. W. and Si, J. (2018) "Engel's law, diet diversity, and the quality of food consumption", *American Journal of Agricultural Economics*, Vol. 100, No. 1, pp. 1-22. ISSN 1467-8276. DOI 10.1093/ajae/aax053.

[9] Deléglise, H., Interdonato, R., Bégué, A., d'Hôtel, E. M., Teisseire, M. and Roche, M. (2022) "Food security prediction from heterogeneous data combining machine and deep learning methods", *Expert Systems with Applications*, Vol. 190, p. 116189. ISSN 0957-4174. DOI 10.1016/j.eswa.2021.116189.

[10] Engel, E. (1857) "Die Productions-Und Consumtionsver-Haltnise Des Konigreichs Sachsen", *Zeitschrift Des Statistischen Bureaus Des Königlich Sächsischen Ministeriums Des Innern*, Vol. 8, pp. 1-54.

[11] Goktolga, Z. G., Bal, S. G. and Karkacier, O. (2006) "Factors effecting primary choice of consumers in food purchasing: The Turkey case", *Food Control*, Vol. 17, No. 11, pp. 884-889. ISSN 0956-7135. DOI 10.1016/j.foodcont.2005.06.006.

[12] Grégoire, G. (2014) "Multiple linear regression", *European Astronomical Society Publications Series*, Vol. 66, pp. 45-72. E-ISSN 1638-1963, ISSN 1633-4760. DOI 10.1051/eas/1466005.

[13] Hill, T., Marquez, L., O'Connor, M. and Remus, W. (1994) "Artificial neural network models for forecasting and decision making", *International Journal of Forecasting*, Vol. 10, No. 1, pp. 5-15. ISSN 0169-2070. DOI 10.1016/0169-2070(94)90045-0.

[14] Holcomb, R. B., Park, J. L. and Capps, Jr. O. (1995) "Revisiting Engel's law: Examining expenditure patterns for food at home and away from home", *Journal of Food Distribution Research,* Vol. 26, No. 2, pp. 1-8. ISSN 2643-3354. DOI 10.22004/ag.econ.27224.

[15] Leser, C. E. V. (1963) "Forms of Engel functions", *Econometrica"*, Vol. 31, No. 4, pp. 694-703. ISSN 0012-9682. DOI 10.2307/1909167.

[16] Lund, P. J. and Derry, B. J. (1985) "Household food consumption: the influence of household characteristics", *Journal of Agricultural Economics*, Vol. 36, No. 1, pp. 41-58. ISSN 0021-857X. DOI 10.1111/j.1477-9552.1985.tb00155.x.

[17] Martini, G., Bracci, A., Riches, L., Jaiswal, S., Corea, M., Rivers, J., Husain, A. and Omodei, E. (2022) "Machine learning can guide food security efforts when primary data are not available", *Nature Food*, Vol. 3, No. 9, pp. 716-728. ISSN 2662-1355. DOI 10.1038/s43016-022-00587-8.

[18] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011) "Scikit-learn: Machine learning in Python", *The Journal of Machine Learning Research*, Vol. 12, pp. 2825-2830. ISSN 1533-7928.

[19] Rae, A. N. (1999) "Food consumption patterns and nutrition in urban Java households: the discriminatory power of some socioeconomic variables", *Australian Journal of Agricultural and Resource Economics*, Vol. 43, No. 3, pp. 359-383. E-ISSN 1467-8489, ISSN 1364-985X. DOI 10.1111/1467-8489.00084.

[20] Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986) "Learning representations by back-propagating errors", *Nature*, Vol. 323, pp. 533-536. ISSN 1476-4687. DOI 10.1038/323533a0.

[21] Staudigel, M. and Schröck, R. (2015) "Food demand in Russia: heterogeneous consumer segments over time", *Journal of Agricultural Economics*, Vol. 66, No. 3, pp. 615-639. ISSN 0118-6566, E-ISSN 1477-9552. DOI 10.1111/1477-9552.12102.

[22] Tiffin, A. and Tiffin, R. (1999) "Estimates of food demand elasticities for Great Britain: 1972–1994", *Journal of Agricultural Economics*, Vol. 50, No. 1, pp. 140-147. ISSN 0118-6566, E-ISSN 1477-9552. DOI 10.1111/j.1477-9552.1999.tb00800.x.

[23] Trafalis, T. B. and Gilbert, R. C. (2007) "Robust support vector machines for classification and computational issues", *Optimisation Methods and Software*, Vol. 22, No. 1, pp. 187-198. E-ISSN 1029-4937. DOI 10.1080/10556780600883791.

[24] Unnevehr, L., Eales, J., Jensen, H., Lusk, J., McCluskey, J. and Kinsey, J. (2010) "Food and consumer economics", *American Journal of Agricultural Economics*, Vol. 92, No. 2, pp. 506-521. ISSN 0002-9092. DOI 10.1093/ajae/aaq007.

[25] Vapnik, V. N. (2000) "The Nature of Statistical Learning Theory", New York, NY: Springer. 314 p. E-ISBN 978-1-4757-3264-1. DOI 10.1007/978-1-4757-3264-1.

[26] Working, H. (1943) "Statistical Laws of Family Expenditure", *Journal of the American Statistical Association,* Vol. 38, No. 221, pp. 43-56. ISSN 0162-1459. DOI 10.2307/2279311.

[27] Wang, C., Wu, Q., Weimer, M. and Zhu, E. (2021) "FLAML: A Fast and Lightweight AutoML Library", *Proceedings of Machine Learning and Systems*, Vol. 3, pp. 434-447. arXiv:1911.04706. ISSN 2640-3498. DOI 10.48550/arXiv.1911.04706.

[28] Xu, H., Caramanis, C. and Mannor, S. (2009) "Robustness and Regularization of Support Vector Machines", *Journal of Machine Learning Research*, Vol. 10, No. 7., pp. 1485-1510. E-ISSN 1533-7928.