Hindawi
Advances in Multimedia
Volume 2022, Article ID 5034976, 12 pages
https://doi.org/10.1155/2022/5034976

Research Article

Research on Forecasting Methods of Agricultural Products Consumption Behavior Based on Unsupervised Learning

Hui Yang[1], **Tian Yang**[1], **Xiangxin Ma**[1], **and Jia Lin**[2]

[1]*College of Economics and Management, Northeast Agricultural University, Harbin 150030, China*
[2]*Department of Economic Management, Harbin Institute of Vocational and Technical Applications, Harbin 150078, China*

Correspondence should be addressed to Hui Yang; yanghui@neau.edu.cn

Received 19 November 2021; Revised 16 December 2021; Accepted 24 December 2021; Published 19 January 2022

Academic Editor: Qiangyi Li

Constructing a perfect urban fresh agricultural product supply system is the basic guarantee for the stability of urban life, and it is also the basic condition for supporting urbanization. This paper combines unsupervised learning algorithms to predict and analyze agricultural product consumption behaviors, determines whether to generate new neurons through the cumulative error value of the winning neurons, and gives a network model that can dynamically self-grow. Moreover, this paper constructs an agricultural product consumption behavior prediction model based on unsupervised learning and uses data to verify the performance of the algorithm in this paper. After confirming the performance test of the algorithm, it verifies the prediction effect of this method on the consumption behavior of agricultural products. Through statistical analysis of data, it can be known that the prediction method of agricultural consumption behavior based on unsupervised learning has a certain effect.

1. Introduction

Consumer demand refers to the consumer's demand and desire for consumer products. This need must meet two basic elements together: purchasing power and purchasing desire. Consumers need to have many characteristics such as diversity, variability, and relevance, so it is extremely complicated to study consumer needs [1]. Maslow's needs theory is what we often call the "basic needs theory". It was put forward by the aesthetics scientist Maslow. The theory divides the basic needs from low to high into five levels from the perspective of the different development stages of human needs. Safety needs and physiological needs are the most basic needs of human beings. All activities must be based on the satisfaction of basic life; otherwise, they will not operate normally. Social needs, respect needs, and self-realization needs are high-level needs, and all human material activities are a process of transition from low-level to high-level [2]. Demand theory starts from the basic needs of people to find ways to inspire people and grasps the core of the problem. Maslow said that people's needs are progressive, and they have experienced a continuous development process from low-level to high-level. They are in line with the law of development of personal needs. However, the shortcoming of this theory is that it has separated from the basic conditions of society, separated from the historical development process of human beings in different periods and social practice to simply consider human needs and demand structure. Its theoretical basis is that the nature of human beings is beyond social history, and human beings are abstracted as "natural persons" instead of "social persons", so there are certain limitations [3].

Clarifying the factors that affect consumers' online shopping of specialty agricultural products under the socialized e-commerce model can not only effectively expand the channels for smallholder farmers and enterprises to sell specialty agricultural products but also avoid vicious competition in the same industry under the same model. Therefore, companies and suppliers must fully consider the changes in consumer demand and purchasing methods before formulating marketing strategies. Consumer research is a compulsory course for companies and merchants to carry out marketing activities. Through behavioral research on consumers' online shopping process, determine which factors are affected by consumers in the process of consumer behavior, increase the driving force of positive factors, and eliminate negative factors, so as to maximize the satisfaction of consumer needs and then achieve the purpose of marketing, that is, to obtain consumer recognition, so as to obtain greater benefits. This article combines the rise of the social e-commerce model and the penetration rate of various social software applications and, on this basis, studies many e-commerce models and conducts research on consumer online shopping behaviors of characteristic agricultural products under the social e-commerce model. It is of great practical significance to explore the factors affecting consumers' online purchase of characteristic agricultural products.

This article combines the unsupervised learning algorithm to predict and analyze the consumption behavior of agricultural products and test the method proposed in this paper to improve the prediction effect of subsequent agricultural product consumption behavior, which provides a theoretical reference for subsequent agricultural product marketing.

2. Related Work

Consumer needs and changes in purchasing methods. Consumer research is a compulsory course for companies and merchants to carry out marketing activities. Through behavioral research on consumers' online shopping process, determine which factors are affected by consumers in the process of consumer behavior, increase the driving force of positive factors, and eliminate negative factors, so as to maximize the satisfaction of consumer needs and then achieve the purpose of marketing, that is, to obtain consumer recognition, so as to obtain greater benefits [4]. Combining with the rise of social e-commerce model and the penetration rate of various social software applications, on this basis, research on many e-commerce models, conduct research on consumer online shopping behaviors of characteristic agricultural products under the social e-commerce model, and discuss the influencing factors of consumers' online shopping of characteristic agricultural products are of very important practical significance [5]. Under the social e-commerce model, research on the behavior of consumers to purchase characteristic agricultural products online through social software will help small farmers, retail suppliers, and agricultural products enterprises to optimize the marketing strategies of characteristic agricultural products, improve their deficiencies, and attract more consumers. Increase the transaction volume of characteristic agricultural products in social software, improve the level of social e-commerce, and increase the economic benefits of farmers, retailers, and enterprises [6]. Determine whether there is any merit in selling characteristic agricultural products through emotional and social relationships under the social e-commerce model. Under the social e-commerce model, whether the method of selling characteristic agricultural products can exist stably between consumers and merchants [7]. Special agricultural products occupy an important position in daily life. With the acceleration of life and the popularization of online shopping, consumers can purchase special agricultural products online through various e-commerce platforms [8]. Literature [9] verifies the influence of the professionalism of the sender of Internet Word of Mouth, the trust of the receiver, and the degree of risk perceived by consumers on consumption behavior by establishing a theoretical model. Literature [10] takes the three variables of communicator, recipient, and word-of-mouth content as independent variables and consumers' perceived risk and trust as intermediate variables and establishes a model of the influence mechanism of negative Internet word-of-mouth on consumer behavior intentions, indicating the communicator that the strength of the relationship with the receiver is directly proportional to the receiver's perceived risk. Literature [11] uses fine processing models and cognitive matching–related theories to study a series of consumer behaviors in processing Internet word-of-mouth and the effect of Internet word-of-mouth on purchasing behavior. Literature [12] studied the mechanism of online word-of-mouth impact on book sales by studying online reviews, and the results showed that Internet word-of-mouth expands sales by increasing its popularity and influence, thereby convincing consumers to buy is not obvious. Literature [13] divides customer perceived value into three measurement dimensions: functional value, emotional value, and social value, and proposes related hypotheses: Internet word-of-mouth has a positive impact on the three dimensions of customer perceived value.

Literature [14] concluded in the research that the purchaser who cares about safety directly determines the purchase and payment of green agricultural products. Literature [15] found that the demand for safe and nutritious agricultural products is very high. The reason is that ordinary agricultural products will be affected by environmental pollution and other aspects, leading to quality and safety problems. Under such circumstances, consumption and purchase of organic food are healthy and environmentally friendly. Literature [16] mainly analyzes the perception of consumers in the research and concludes that what consumers ultimately buy is not only the product but also the value of the product. Literature [17] believes that consumers are cognized by buyers in terms of perception. When performance and cognitive input are balanced, buyers' willingness to buy will be low at this time. Literature [18] pointed out that when consumers' willingness to consume is considered by price factors, they will make corresponding decisions while paying attention to the quality of agricultural products and finally make corresponding value evaluations. When choosing guaranteed agricultural products, cultural background, social factors, and personal psychology will all affect consumers' decision-making [19]. The demand for green agricultural products consumption is very high in western developed countries. Consumers in western developed countries are willing to buy organic agricultural products, and almost all consumers have purchased organic agricultural products [20].

3. Unsupervised Learning Prediction Method

The mapping process of the unsupervised learning network is slightly different from that of the standard learning network. It can be divided into the following three phases: the initialization phase, the self-growth phase, and the stable phase.

The self-growth stage is the core stage of the unsupervised learning network. In this stage, the algorithm will not only add new neurons at the neuron with the largest cumulative error but also use the probability distribution of the training sample as a measure of the self-growth mechanism. This overcomes the limitations caused by the fixed network structure of the standard machine learning network. In the process of self-growth, the grid keeps adding or deleting neurons on the basis of keeping square neighborhoods until the data mapping reaches the required granularity. The specific manifestation is that the input pattern is mapped to clusters with obvious boundaries, the data within the clusters have similar characteristics, and the data between the clusters have different characteristics.

In the self-growth stage, unsupervised learning will be carried out according to the following algorithm:

(1) The algorithm inputs the sample data into the initialized network model.

(2) The algorithm uses Euclidean distance to find the weight vector closest to the input vector (similar to machine learning).

This process can be summarized as finding the neuron q to make

$$\forall q \in N, \left|v - w_{q'}\right| \leq \left|v - w_q\right|. \quad (1)$$

Here, v and w are the input vector and the weight vector respectively, and q is the vector coordinate of the neuron.

(3) The adaptation of the weight vector is only applicable to the neighborhood of the winning neuron or the winning neuron itself. Even in the neighborhood, the adaptability of the weights close to the winning neuron is better than those farther away. The adaptive learning rate will decrease exponentially in the iterative process. The adaptive process of weights can be described by formula (1):

$$w_j(k+1) = \begin{cases} w_j(k), & j \notin N_{k+1} \\ w_j(k) + \mathrm{LR}(k) \times \left(x_k - w_j(k)\right), & j \in N_{k+1} \end{cases} \quad (2)$$

Here, LR(k) is the learning rate, $w_j(k)$ is the weight of the weight vector of neuron j before the structure adjustment, $w_j(k+1)$ is the weight of the weight vector of neuron j after structural adjustment, and N_{k+1} is the neighborhood of the winning neuron in the $k+1$ iteration. The deceleration rate of the learning rate LR(k) depends on the number of neurons in the unsupervised learning network at time t.

(4) The algorithm accumulates the error value of the winning neuron i (Total Error, TE, that is, the difference between the input vector and the weight vector).

(5) When TEi ≥ GT, if neuron i is a boundary neuron, a new neuron is added to its neighborhood. If neuron i is a nonboundary neuron, TE is assigned to other neurons in its neighborhood.

(6) The algorithm initializes the weight vector of the newly generated neuron.

(7) The algorithm assigns an initial value to the Learning Rate (LR).

(8) The algorithm repeats steps (2)~(7) until all input patterns have been input, and the growth rate of the neuron has been reduced to a small, fixed value.

The set of neuron weight vector w_i can be regarded as a vector quantization of the input space. Therefore, each neuron i can be used to represent a region V_i. All points in this area are closer to w_i than the weight vector of any other neuron. V_i is called the Voronoi region, that is, the weight vector divides the input space into several Voronoi regions, and each region is represented by a neuron. In the self-growth stage, when the sample data is input into the network, according to the above algorithm, a winning neuron will be generated. The difference TE between the input vector and the weight vector of the corresponding winning neuron will be added to the neuron. If neuron i contributes significantly to TE, the Voronoi area will be represented by neuron i. To better represent this area, a new neuron will be generated as a neighbor of neuron i.

Here, the error distance E needs to be defined:

$$E_i(t+1) = E_i(t) + \sqrt{\sum_{k=1}^{\mathrm{Dim}} \mathrm{Met}\left(v_k, w_{ik}\right)^2}. \quad (3)$$

Here, i and t represent the t-th iteration of neuron i, Dim is the dimension of the input data, v represents the input vector, and w represents the weight vector. Met is a function that calculates the distance between the vectors v and w. For each winning neuron, the difference between the weight vector and the input vector will be calculated as the error value. If a neuron wins multiple times, then this error value will be accumulated during the iteration process. The Euclidean function is used to calculate the distance, so formula (1) can be rewritten as follows:

$$E_i(t+1) = E_i(t) + \sqrt{\sum_{k=1}^{\mathrm{Dim}} \left(v_k - w_i\right)^2}. \quad (4)$$

For each weight update, if $E_i^{\mathrm{new}} > H_{\mathrm{Err}}$, then there is $H_{\mathrm{Err}} = E_i^{\mathrm{new}}$, otherwise the value of H_{Err} remains unchanged. Therefore, H_{Err} will always maintain the maximum error value of the neuron. The error value of each neuron can be regarded as a quantized error, and the sum of the quantized error values is as follows:

$$QE = \sum_{i=1}^{N} E_i. \quad (5)$$

Here, N is the number of neurons in the network and E is the error value of neuron i. The sum of quantization errors, QE, is used to determine when to generate new neurons. If neuron i contributes the most to the total error value QE, the Voronoi region V of the input space where it is located will be represented by neuron i, so a new neuron will be generated and used to share the responsibility of neuron i.

Neuron i should generate a new neuron:

$$\frac{\partial QE}{\partial E} > GT. \tag{6}$$

Since H_{Err} records the maximum error value of the neuron, if $H_{\mathrm{Err}} > GT$, a new neuron should be generated.

For the generation of a new neuron, the following restrictions are required:.

(1) Determination of the location of newly generated neurons: newly generated neurons must be generated from boundary neurons, as shown in Figure 1.

(2) Initialization of the weights of newly generated neurons: newly grown neurons will be given initialization weights. For initializing new nodes, the following four situations need to be considered, as shown in Figure 2.

W_1 and W_2 are the weight of the growing source neuron, and W_{new} is the weight of the newly growing neuron. At the beginning of the unsupervised learning network, there are four initialization neurons. New neurons are always generated as neighbors of an existing node. In Figure x, (a) and (b) are the most frequently occurring situations. The initialization formula for the weight of the newly grown neuron is given below:

In Figure 2(a), the newly generated neuron is on one side of two consecutive neurons. If $w_2 > w_1$, then $w_{\mathrm{new}} = w_1 - (w_2 - w_1)$, otherwise $w_{\mathrm{new}} = w_1 + (w_1 - w_2)$.

In Figure 2(b), the newly generated neuron is between two neurons, and $w_{\mathrm{new}} = (w_1 + w_2)/2$;

In Figure 2(c), the newly generated neuron is directly connected to only one neuron, and the other neuron connected to this neuron is on the side opposite to the newly generated neuron, and the three neurons are in an "L" shape. If $w_2 > w_1$, then $w_{\mathrm{new}} = w_1 - (w_2 - w_1)$, otherwise $w_{\mathrm{new}} = w_1 + (w_1 - w_2)$. This situation is the same as (a), except that the location of neighboring neurons is slightly different.

In Figure 2(d), the newly generated neuron has only one neuron as a neighbor. This happens at the beginning of self-growth or when a neuron becomes an isolated neuron due to neuron deletion. $w_{\mathrm{new}} = (r_1 + r_2)/2$, where r_1 and r_2 are the minimum and maximum values of the weight vector distribution interval.

In cases (a) ~ (c), if $r_1 \leq w_{new} \leq r_2$, the new neuron will be given a weight. In case (d) or w_{new} exceeds the range of $[r_1, r_2]$, w_{new} will be assigned to $(r_1 + r_2)/2$, where the value is generally 0.5. The self-organizing network will quickly adjust this value to adapt to the neighborhood. It can be said that in case (a(c), the weight of the new neuron will be close to the weight of the neighboring neuron. In other cases, the initial value of the weight of the new neuron will be assigned an intermediate value. Then, the self-organizing network adjusts the value.

The adaptive learning rate is defined as follows:

$$\mathrm{LR}(t+1) = \mathrm{LR}(t) \times \alpha. \tag{7}$$

α is the reduction coefficient of the learning rate, and $0 < \alpha < 1$. LR(t) is the learning rate of the t-th iteration. When $t \longrightarrow \infty$, LR(t)$\longrightarrow$0. LR, like the machine learning network, is generally assigned a large value (close to 1) during initialization. However, since there are only a small number of neurons (usually 4) when unsupervised learning is initialized, a situation that may arise at this time is that even with the same neuron set weight vector, the network may grow in a completely different direction.

To better solve this problem, a new learning coefficient rule is introduced here, and the number of neurons in the current neural network is added to the learning coefficient. Therefore, the adaptive learning rate can be defined as follows:

$$\mathrm{LR}(t+1) = \alpha \times \psi(n) \times \mathrm{LR}(t). \tag{8}$$

Here, n is the number of neurons in the network and $\psi(n)$ can control the value of the learning rate according to the value of n. For example, the $\psi(n)$ function can be set to $(1 - R/n(t))$, and when the number of initialized neurons is $n = 4$, $R = 3.8$ is generally adopted. Thus, there are the following:

(1) In the initialization phase, when the number of neurons is small, the LR value will be effectively reduced by $\psi(n)$, which will reduce the weight fluctuation of the network.

(2) When the network topology continues to increase and the number of neurons continues to increase, $\psi(n)$ will also gradually increase and approach 1, and LR will not fluctuate by a large margin as in the case (1).

The unsupervised learning network skips the value setting stage, and the weight of the new neuron is initialized to the weight of the surrounding neighborhood. In the self-growth stage, the unsupervised learning network will initialize the learning rate and neighborhood size Ng for each new input. The weight is continuously adjusted adaptively when the range of the neighborhood, and the learning rate are gradually reduced until the number of neurons in the neighborhood is 1, and then the new input is initialized again. The weight adaptive adjustment can be expressed by formula (8).

$$\frac{\mathrm{d}w_i}{\mathrm{d}t} = \mathrm{LR} \times (v - w_i),\ i \in N_k. \tag{9}$$

It is stipulated here that if a neuron is not a boundary neuron of the network, it will not generate a new neuron, and use formula (9) to transfer the cumulative error value of the neuron to its neighboring neurons:

$$E_{t+1}^{w} = \frac{\mathrm{GT}}{2}. \tag{10}$$

Here, E_t^{w} is the error value of the winning neuron. The error value of the neighborhood neuron of the winning neuron will increase according to formula (10):

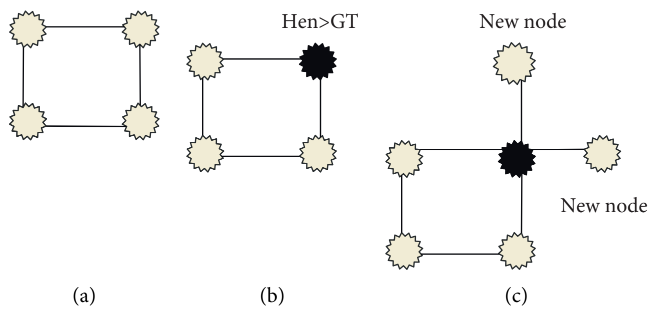

Figure 1: New neurons generated from boundary neurons.

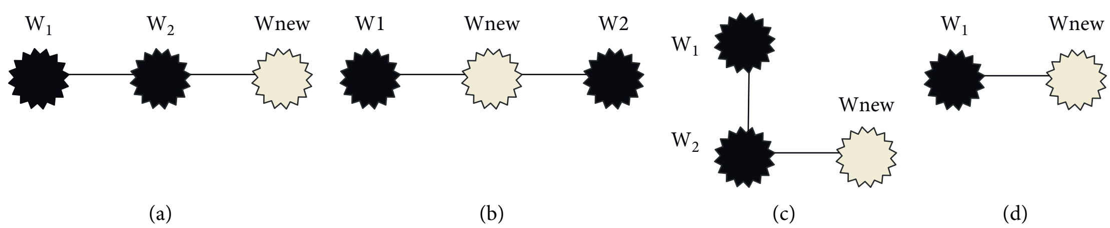

Figure 2: Initialization of the weights of newly generated neurons.

$$E_{t+1}^{n_i} = E_t^{n_i} + \gamma E_t^{n_i}, \tag{11}$$

where $E_{t+1}^{n_i}\,(i = 1, 2, 3, 4)$ is the error value of the i-th neighbor of the winning neuron (its error value is $E_t^{n_i}$). y is a constant value, which is called a distribution factor (FD), and is used to control the accumulation of error values. The value of FD is generally between 0 and 1.

When the cumulative error value exceeds the growth threshold GT, the neuron will grow by itself, and the total error value of neuron i can be calculated as

$$TE_i = \sum_{H_i} \sum_{j=1}^{D} \left(x_{ij} - w_j\right)^2. \tag{12}$$

Here, H_i is the number of wins of neuron i, D is the dimension of the data set, and x_{ij} and w_j are the input and weight vector of neuron i, respectively. For a boundary neuron, if it wants to grow a new neuron, it requires

$$TE_i \geq GT. \tag{13}$$

The value of GT is derived experimentally based on the requirements of network growth. It can be seen from the formula (12) that the dimensionality of the data set has a very important influence on the cumulative error value TE.

Because of $0 < x_{ij}, w_j < 1$, the maximum contribution of an input attribute to the error value is 1, that is, $\max|x_{ij} - w_j| = 1$. Therefore, $TE_{\max} = D \times H_{\max}$, where $TE_{\max}$ is the maximum accumulated error, and $H_{\max}$ is the maximum number of times that this neuron becomes a winning neuron. If H(t) is considered as a function of time (number of iterations) t, then there is $0 \leq GT \leq D \times H(t)$.

The definition of GT needs to consider the expansion factor of the network, which depends on the dimensionality of the data set and the number of wins of the neuron. For different data sets, there will be different GT values. Therefore, the GTs of different data sets are not comparable. To solve this problem, the spreading factor SF (Spread-Out Factor) is introduced here to control and calculate the GT value of the unsupervised learning network.

The GT value can be defined as follows:

$$GT = D \times f(SF). \tag{14}$$

Here, $SF \in R, 0 \leq SF \leq 1$, and f(SF) are functions of SF.

The range of the total error value TEi of neuron i is as follows:

$$0 \leq TE_i \leq TE_{\max}. \tag{15}$$

Here, TEmax is the maximum cumulative error value, and then formula (14) can be obtained:

$$0 \leq \sum_{H_i} \sum_{j=1}^{D} \left(x_{ij} - w_j\right)^2 \leq \sum_{H_{\max}} \sum_{j=1}^{D} \left(x_{ij} - w_j\right)^2. \tag{16}$$

Since the purpose of GT is to allow the network to grow new neurons by providing a threshold for the cumulative error value, there is also formula (16):

$$0 \leq GT \leq \sum_{H_{\max}} \sum_{j=1}^{D} \left(x_{ij} - w_j\right)^2. \tag{17}$$

Since $H_{\max}$ can be infinite in theory, $0 \leq GT \leq \infty$. According to the definition of the spreading factor, a function f(SF) needs to be defined such that $0 \leq \mathrm{SF} \leq 1$ and $0 \leq D \times f(SF) \leq \infty$. In other words, the value range of the function $f(x)$ is from zero to infinity, and the value range of the parameter x is between 0 and 1.

We give $y = -a \times \ln(1 - x)$, which can meet the above requirements. If we set $\eta = 1 - SF$ and $GT = -D \times \ln(1 - \eta)$, then $GT = -D \times \ln(SF)$. Therefore, the expansion factor SF can now be used to replace the growth threshold GT, thereby eliminating the trouble of setting different GT values for different data sets. Through the expansion factor, different network mappings can be compared. Figure 3 is a graph showing changes in GT values of data sets of different dimensions with a given expansion factor SF.

Now, the GT value of the unsupervised learning network can be calculated by a given expansion factor. The range of the expansion factor SF is between 0 and 1. If the data set has no prior knowledge, the recommended value range of the initial SF is 0 to 0.3. This will show the most important clusters in the unsupervised learning network. Based on these most basic clusters, you can decide whether to conduct a deeper data analysis. We can select some areas that have been clustered and use a larger SF value for further data analysis, as shown in Figures 4 and 5.

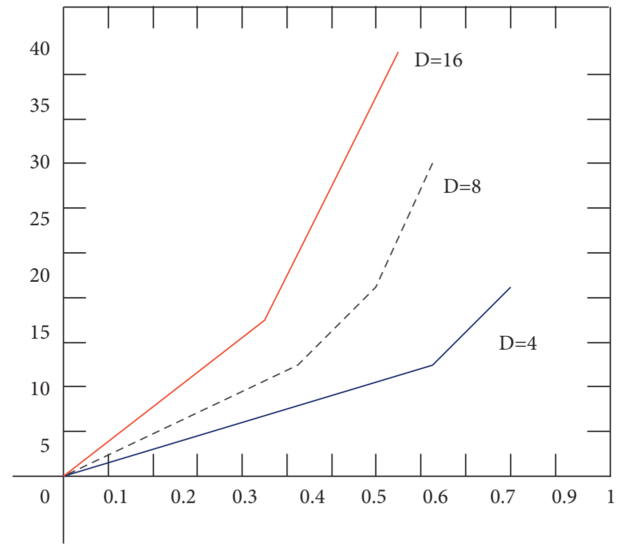

Figure 3: Expansion factor-GT value curve.

4. Prediction of Agricultural Product Consumption Behavior Based on Unsupervised Learning

This paper uses unsupervised learning method to predict the consumption behavior of agricultural products. This article takes fresh agricultural products as an example to study. Figure 6 shows the relationship between "eating, clothing, housing, transportation, and play" in the traditional sense and the hierarchical model of consumption needs. It is also easy to understand the importance and basis of "eating," and the consumption of fresh agricultural products is also the embodiment of "eating."

Based on the existing consumer behavior selection process model, this paper combines the observation and analysis of consumption behavior of fresh agricultural products to divide the consumption behavior of fresh agricultural products into four stages: purchase demand formation, purchase decision making, purchase decision implementation, and post-purchase evaluation feedback. Moreover, this article believes that the final performance of consumer behavior is produced under the dynamic and continuous influence of internal and external factors, that is, consumer behavior is a dynamic and continuous process, and the specific behavior of fresh agricultural products is shown in Figure 7.

The basis for the ultimate realization of the purchase of fresh agricultural products is the formation of purchase motivation. That is to say, only when consumers feel or realize that they need to buy fresh agricultural products can the final purchase behavior occur. The so-called no cause means no result. Consumer purchase motivation is formed or disappeared under the combined effect of demand factors and constraints. The content and mechanism of demand factors and constraints formed by purchase motivation are shown in Figure 8.

The purchase decision itself is a complicated psychological process, and the measurement of "profit" and "strength" is a huge project. On the other hand, the focus of this article is not to measure the utility of the sales terminal but to explore the development direction of the urban fresh agricultural product supply system through the performance of consumer behavior. Therefore, in the research process, this article ignores the measurement standards of the "strength" of each fresh agricultural product sales terminal at the factor level and the quantitative research process of the consumer's "profit" at the factor level. To facilitate research, a comprehensive system for purchasing decision research is established as shown in Figure 9.

On the basis of the above analysis, we conduct research on the prediction methods of agricultural product consumption behavior and study the distribution of consumer household income under different attitudes. The answer can be clearly obtained from the statistics and analysis in Figure 10. When we analyze the distribution of consumer household income under different attitudes, it can be seen that the lower the average household income, the higher consumers' expectations of "lower prices". Conversely, the higher the average household income, the weaker the consumer's expectation of "lower prices". This shows that although the "lower price" is the "profit" side, it shows different "profits" according to the level of consumer income.

In order to show the influence of the factors of each sales terminal more clearly, draw the radar distribution diagram of the influence intensity, as shown in Figure 11. Through the comparison of the main influencing factors of various sales terminals, it can be found that the same attributes of different sales terminals have very different influences for consumers, but different terminal advantages for the operation of sales terminals. Based on the specific scope of this survey (Chengdu), the advantages of supermarkets in operating fresh agricultural products over other sales terminals are mainly reflected in safety (quality and safety), selectivity (multiple product types), demand relevance (buying other products, like shopping), comfort (good shopping environment), and credibility (high business reputation). The advantages of the farmer's market in operating fresh agricultural products are reflected in economy (low price level),

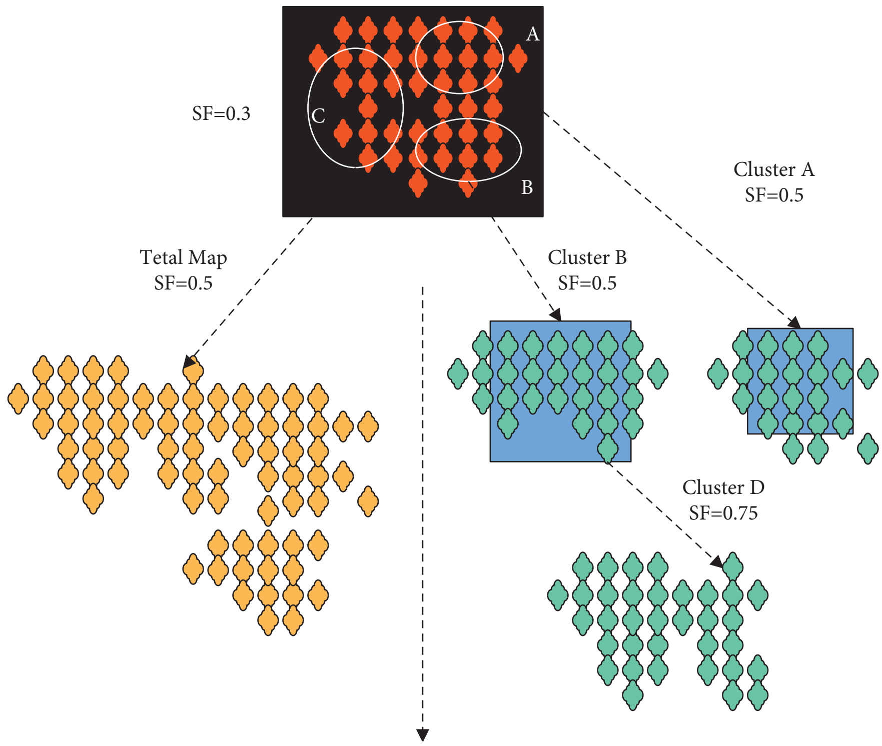

Figure 4: Data analysis based on the value of SF..

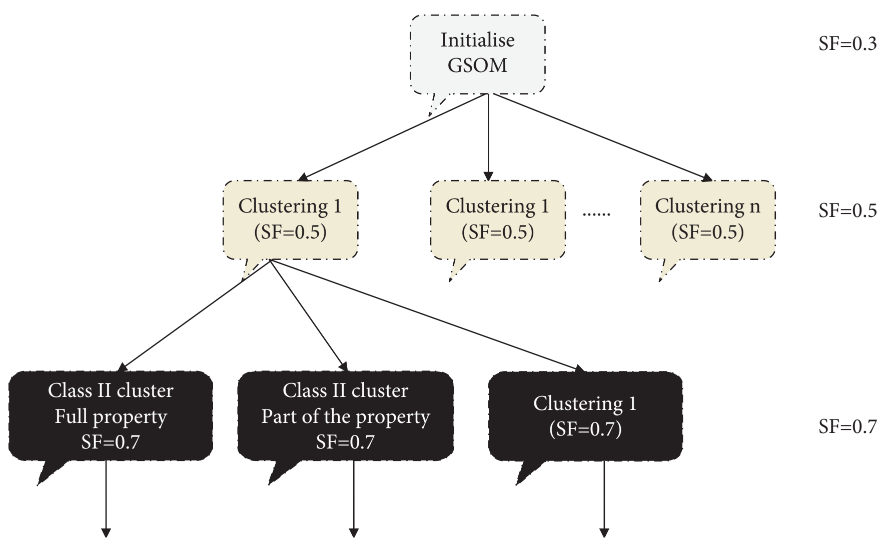

Figure 5: Hierarchical clustering based on the value of SF.

efficiency (less queues to save time), convenience (short travel distance), and product appearance (product freshness). The advantages of roadside vendors in operating fresh agricultural products are reflected in convenience (short travel distance) and economy (low price level).

The above-mentioned research not only analyzes the prediction of agricultural product consumption behavior but also verifies that the prediction method of agricultural product consumption behavior based on unsupervised learning has certain effects.

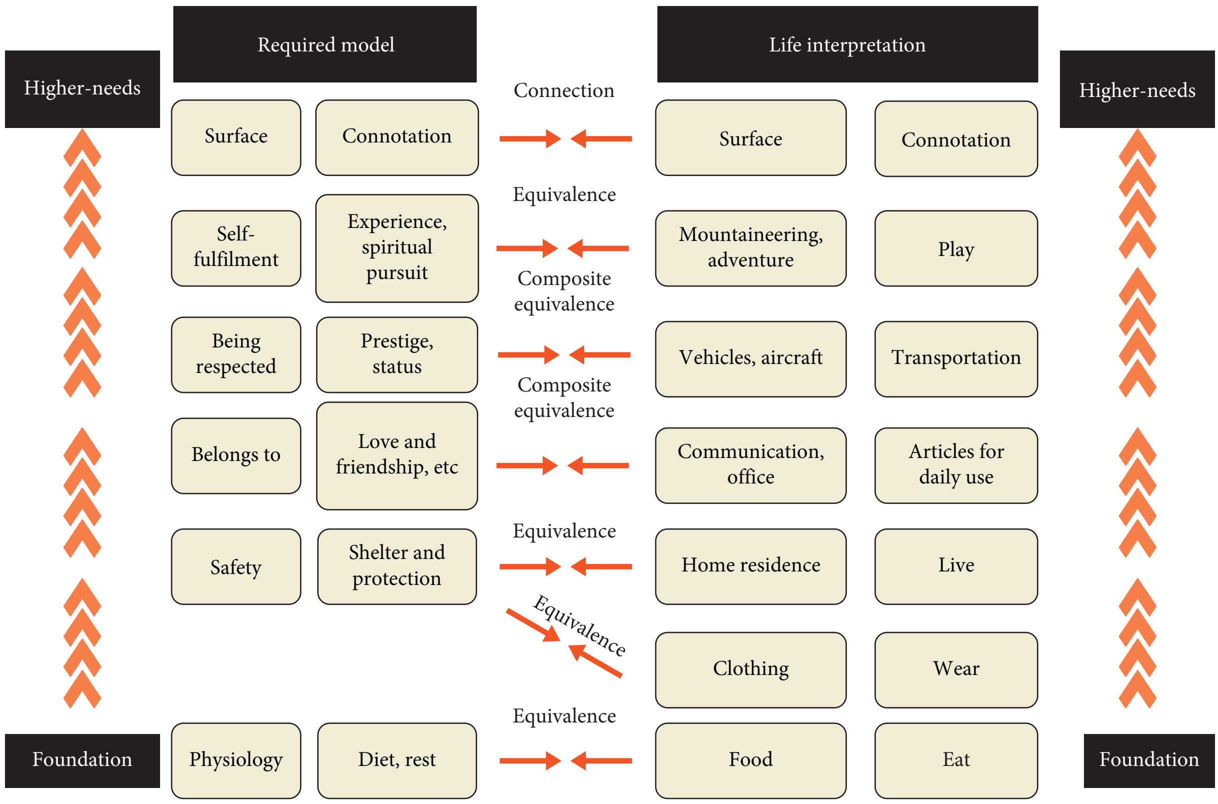

Figure 6: The relationship between life needs and needs models in the traditional sense.

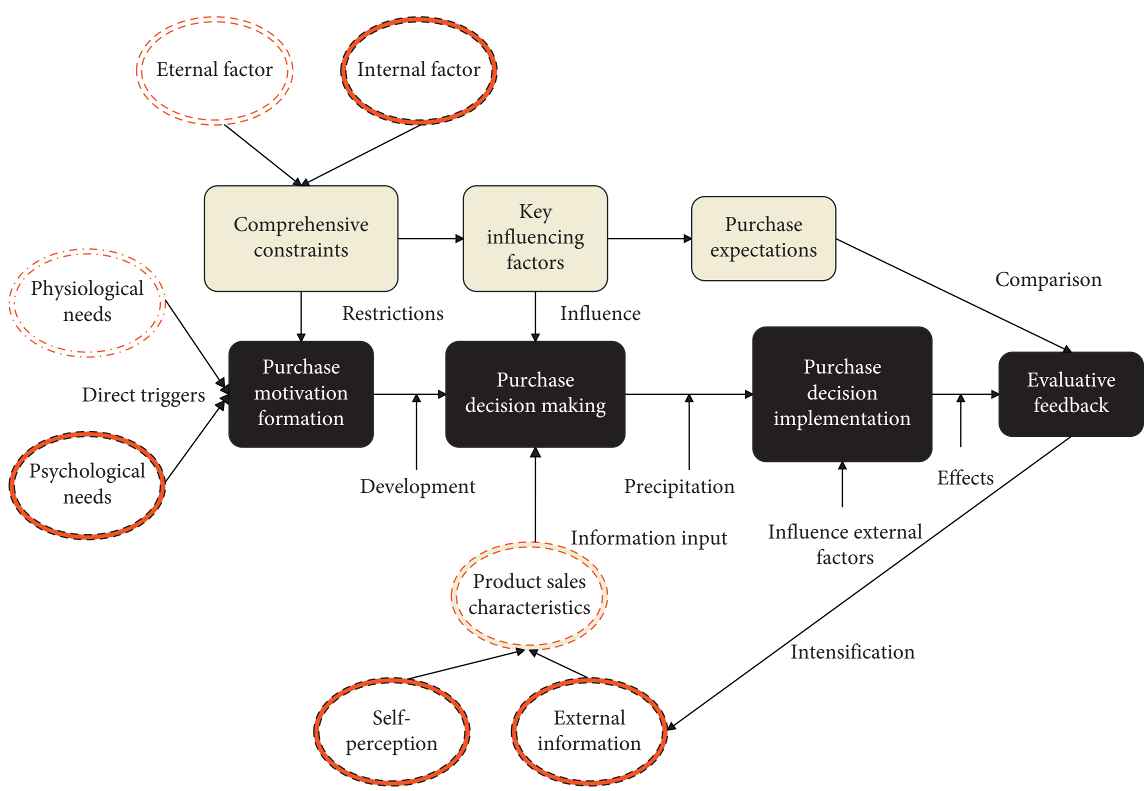

Figure 7: Structure diagram of consumption behavior pattern of fresh agricultural products.

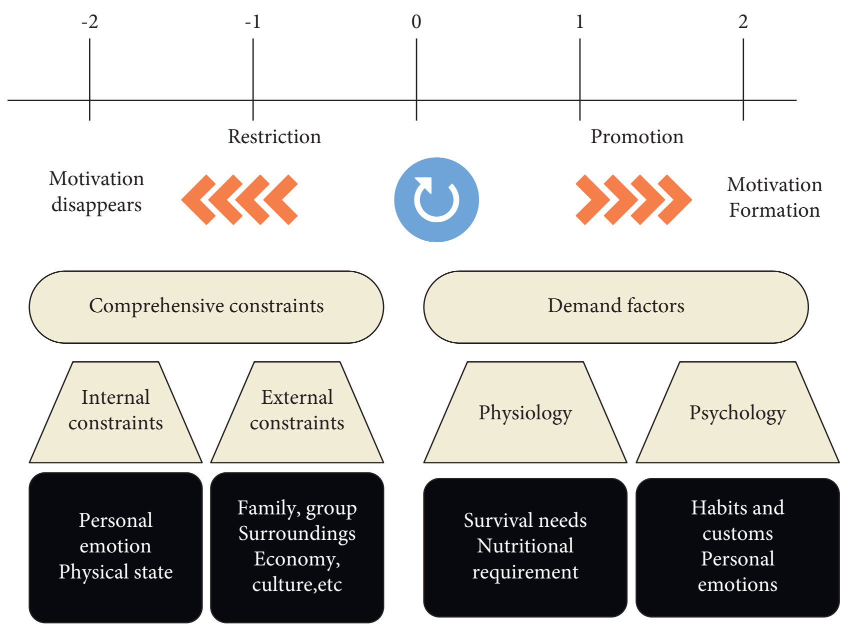

Figure 8: The influencing factors and mechanism of the formation of buying motives.

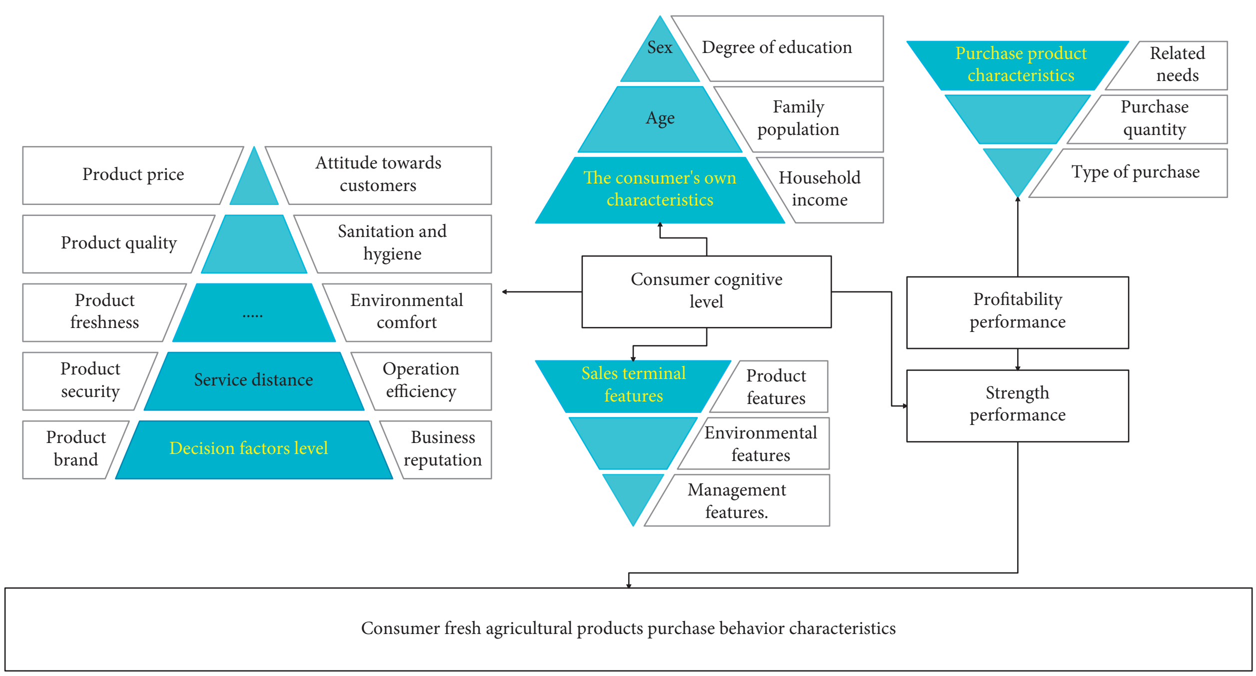

Figure 9: The comprehensive system for purchasing decision-making research on fresh agricultural products.

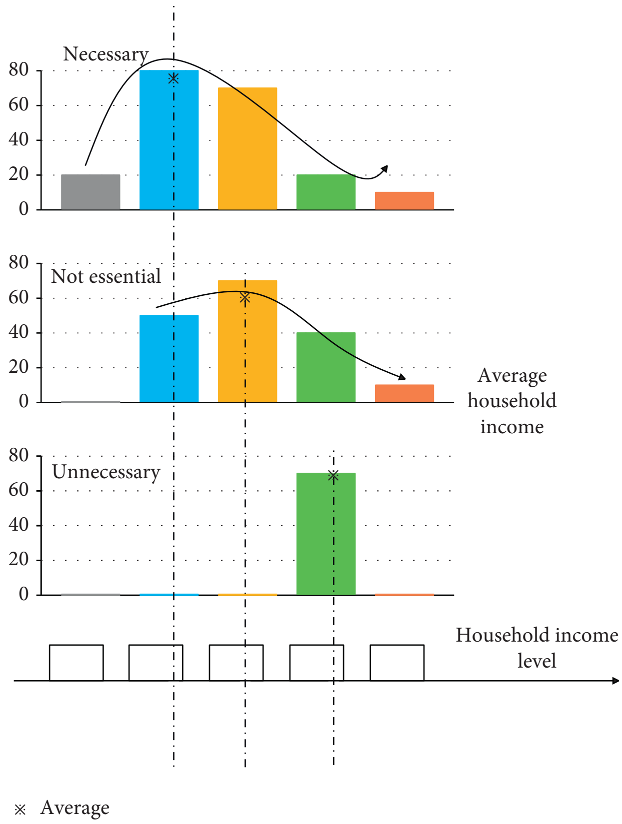

FIGURE 10: Average level of household income under different price attitudes.

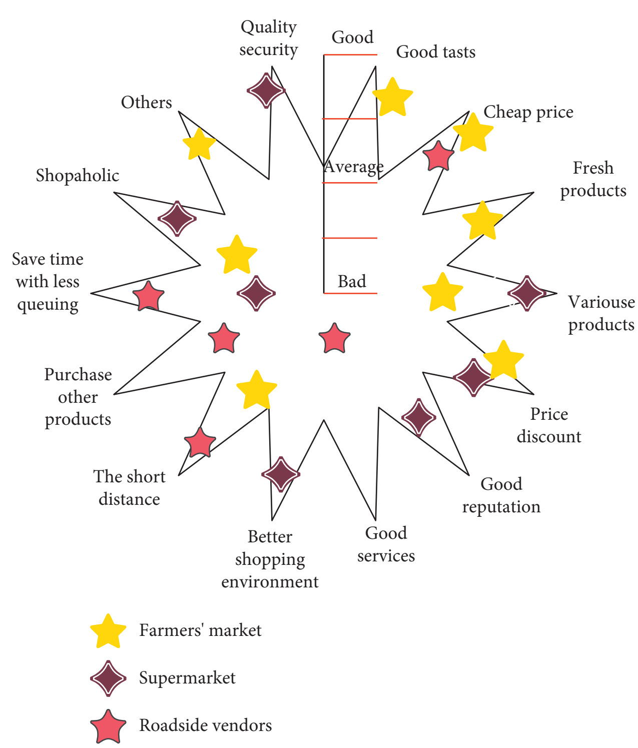

FIGURE 11: Radar distribution map of the influence intensity of various factors in different sales terminals.

5. Conclusion

This article studies the urban agricultural product supply system based on consumer behavior, that is, from the fundamental purpose of agricultural product logistics, fully considers the characteristics of consumer demand for agricultural products, and understands the key factors that affect the consumption of agricultural products. Moreover, this article theoretically enriches the ideas and methods of urban agricultural product supply chain optimization and construction and better explores the mechanism of urban agricultural product supply system construction and the key factors that determine its construction. In addition, the study of the supply system of agricultural products from the perspective of consumer behavior can have more practical significance and provide feasible decision-making reference and development ideas for the planning, management, transformation, and operation of urban agricultural product supply systems. This article combines the unsupervised learning algorithm to predict and analyze the consumption behavior of agricultural products and test the method proposed in this paper to improve the prediction effect of subsequent agricultural product consumption behavior, which provides a theoretical reference for subsequent agricultural product marketing. Finally, this paper verifies that the agricultural product consumption behavior prediction method based on unsupervised learning has certain effects through experiments.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The research was supported by 2021 Heilongjiang Philosophy and Social Sciences Research Planning Project "Research on the Effective Path of Heilongjiang Brand Strengthening Agriculture Construction under the Strategy of Rural Revitalization," (No. 21YA005) and Postdoctoral Fund of Heilongjiang Province "Research on the Effect of Milk Culture on the Development of Dairy Enterprises in Heilongjiang" (No. LBH-Z17028).

References

[1] S. R. Paramati, N. Apergis, and M. Ummalla, "Dynamics of renewable energy consumption and economic activities across the agriculture, industry, and service sectors: evidence in the perspective of sustainable development," *Environmental Science and Pollution Research*, vol. 25, no. 2, pp. 1375–1387, 2018.

[2] B. Aydoğan and G. Vardar, "Evaluating the role of renewable energy, economic growth and agriculture on CO2 emission in E7 countries," *International Journal of Sustainable Energy*, vol. 39, no. 4, pp. 335–348, 2020.

[3] I. Mocanu, I. Grigorescu, B. Mitrică, and M. Bianca, "Regional disparities related to socio-economic determinants of agriculture in the Romanian Plain," *Journal of Urban and Regional Analysis*, vol. 10, no. 1, pp. 79–99, 2018.

[4] A. Ruane, J. Antle, J. Elliott et al., "Biophysical and economic implications for agriculture of +1.5° and +2.0°C global warming using AgMIP Coordinated Global and Regional Assessments," *Climate Research*, vol. 76, no. 1, pp. 17–39, 2018.

[5] W. Khan and S. A. Ansari, "Does agriculture matter for economic growth of Uttar Pradesh (India)?" *Economy of Region*, vol. 14, no. 3, pp. 1029–1037, 2018.

[6] M. Paul, "Community-supported agriculture in the United States: social, ecological, and economic benefits to farming," *Journal of Agrarian Change*, vol. 19, no. 1, pp. 162–180, 2019.

[7] I. Pattnaik, K. Lahiri-Dutt, S. Lockie, and B. Pritchard, "The feminization of agriculture or the feminization of agrarian distress? Tracking the trajectory of women in agriculture in India," *Journal of the Asia Pacific Economy*, vol. 23, no. 1, pp. 138–155, 2018.

[8] G. Salmoral, D. Rey, A. Rudd, P. Margon, and I. Holman, "A probabilistic risk assessment of the national economic impacts of regulatory drought management on irrigated agriculture," *Earth's Future*, vol. 7, no. 2, pp. 178–196, 2019.

[9] B. Lanz, S. Dietz, and T. Swanson, "Global economic growth and agricultural land conversion under uncertain productivity improvements in agriculture," *American Journal of Agricultural Economics*, vol. 100, no. 2, pp. 545–569, 2018.

[10] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: a new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.

[11] J. N. Goetz, A. Brenning, H. Petschko, and P. Leopold, "Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling," *Computers & Geosciences*, vol. 81, pp. 1–11, 2015.

[12] H. Darabi, B. Choubin, O. Rahmati, A. Torabi Haghighi, B. Pradhan, and B. Kløve, "Urban flood risk mapping using the GARP and QUEST models: a comparative study of machine learning techniques," *Journal of Hydrology*, vol. 569, pp. 142–154, 2019.

[13] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019.

[14] Y. Xin, L. Kong, Z. Liu et al., "Machine learning and deep learning methods for cybersecurity," *IEEE Access*, vol. 6, pp. 35365–35381, 2018.

[15] B. Meredig, A. Agrawal, S. Kirklin, and S. James, "Combinatorial screening for new materials in unconstrained composition space with machine learning," *Physical Review B*, vol. 89, no. 9, Article ID 094104, 2014.

[16] L. Ward, A. Agrawal, A. Choudhary, and W. Christopher, "A general-purpose machine learning framework for predicting properties of inorganic materials," *Npj Computational Materials*, vol. 2, no. 1, pp. 1–7, 2016.

[17] P. Feng, B. Wang, D. L. Liu, C. Waters, and Q. Yu, "Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia," *Agricultural and Forest Meteorology*, vol. 275, pp. 100–113, 2019.

[18] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8–17, 2015.

[19] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, “Power to the people: the role of humans in interactive machine learning,” *AI Magazine*, vol. 35, no. 4, pp. 105–120, 2014.

[20] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas, “Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines,” *Ore Geology Reviews*, vol. 71, pp. 804–818, 2015.