

## Article

# Using Deep Learning Techniques to Forecast Environmental Consumption Level

Donghyun Lee <sup>1</sup> , Suna Kang <sup>2</sup> and Jungwoo Shin <sup>3,\*</sup>

<sup>1</sup> Department of Business Administration, Korea Polytechnic University, 237 Sangidaehak-ro, Siheung, Gyeonggi 15073, Korea; madeby2@gmail.com

<sup>2</sup> Korea Environment Institute, 370 Sicheong-daero, Sejong 30147, Korea; sakang@kei.re.kr

<sup>3</sup> Department of Industrial and Management Systems Engineering, Kyung Hee University, 1732 Deogyeong-daero, Giheung-gu, Yongin, Gyeonggi 17104, Korea

\* Correspondence: shinjung11@gmail.com; Tel.: +82-31-201-3687; Fax: +82-31-204-8114

Received: 3 August 2017; Accepted: 16 October 2017; Published: 20 October 2017

**Abstract:** Artificial intelligence is a promising futuristic concept in the field of science and technology, and is widely used in new industries. The deep-learning technology leads to performance enhancement and generalization of artificial intelligence technology. The global leader in the field of information technology has declared its intention to utilize the deep-learning technology to solve environmental problems such as climate change, but few environmental applications have so far been developed. This study uses deep-learning technologies in the environmental field to predict the status of pro-environmental consumption. We predicted the pro-environmental consumption index based on Google search query data, using a recurrent neural network (RNN) model. To verify the accuracy of the index, we compared the prediction accuracy of the RNN model with that of the ordinary least square and artificial neural network models. The RNN model predicts the pro-environmental consumption index better than any other model. We expect the RNN model to perform still better in a big data environment because the deep-learning technologies would be increasingly sophisticated as the volume of data grows. Moreover, the framework of this study could be useful in environmental forecasting to prevent damage caused by climate change.

**Keywords:** artificial intelligence; artificial neural network; consumption index; deep-learning technology; pro-environment

## 1. Introduction

Recently, the seriousness of environmental pollution—such as that of air and water—has become more evident. Environmental pollution is a global problem, rather than a national one. To solve such problems, governments need to control the actions of businesses and individuals to reduce their environmental pollution and embrace sustainable consumption. Such encouragements are typically made through policy, but they do not easily lead to pro-environmental consumption practices. For example, in the case of South Korea, over 800 government agencies spent 2.2 trillion Korea Won on eco-products in 2014 [1]; however, green products are rarely purchased outside these agencies. This phenomenon occurs because there is a gap between consumer attitudes and behavior [2]—that is, environmental attitude is a major factor in decision-making vis-à-vis the consumption of “green” goods and services [3]. Therefore, it is necessary to understand those consumer attitudes that will lead to sustainability-conducive behavior and consumption.

Similarly, both the 9th OECD Working Party on Integrating Environmental and Economic Policies [4] and Lee et al. [5] examined how to measure residents’ attitudes regarding the environment, by using online search query data; they also examined how to apply these attitudinal data to the introduction of environmental policy. Lee et al. [6] also measured a pro-environmental consumption

index by using Google search data; they verified that these index values are reliable as green consumption index values, by conducting correlation analyses with an existing green consumption index measure.

Most studies analyze the determinants of pro-environment attitudes or consumption and measure this index. To find a way to both stabilize and activate green consumption, it is essential to predict pro-environmental consumption index values that elucidate the status quo of green consumption. For this reason, we look to predict pro-environmental consumption index values by using artificial intelligence (AI)—an approach that is currently in the scholarly “spotlight”.

Artificial intelligence (AI), first mentioned in 1956 by McCarthy et al. [7], was defined as a “technology to mimic the human brain using the software technology” [8]. With the explosive growth in data volumes and increase in computing power due to the emergence of cloud computing recently, artificial intelligence and deep learning have been recognized as key futuristic technologies in the field of science.

Among the artificial intelligence technologies, deep learning is the main driver of technology generalization and improvement. The definition of deep learning differs according to the researcher. In general, the deep-learning technology is an advanced technology based on an increasing number of hidden layers in a simple neural network [9]. In addition, this technology is used as a de-noising technique to solve overfitting and to model an auto-encoder that can grasp data characteristics through unsupervised learning. With overfitting, a model is acquiring too much data—an overfitted model, for example, can include information such as random errors and noise [10]. The auto-encoder is used to reconstruct the output layer results so that they align as closely as possible with the input data [11].

The characteristics of deep-learning technologies seem to be barely different from those of neural networks, but the paradigm has greatly changed—from an increase in the number of hidden layers and breakthrough in performance to greater expressiveness, for example. Furthermore, as the auto-encoder model combines supervised and unsupervised learning, the deep-learning model automatically extracts the feature needed for analysis in contrast to the existing neural network model, which needs expert knowledge in related fields. In addition, the deep-learning model has the advantage that it is implemented as one model, since the data features are extracted and classified in one classifier [9]. While econometric models are chosen and applied to specific situations under various assumptions, the deep-learning model has the advantage of great versatility of applicable data.

Based on its excellent skills in deep learning, Google declared that it will use the deep-learning technology in the environment field—in climate change forecasting, for example [12]. However, not many environmental studies have used the deep-learning technology. In particular, studies on consumption prediction in the environment field are limited to energy consumption [13,14], and prediction research related to pro-environmental consumption is scarce.

The purpose of this study is to forecast pro-environmental consumption using deep-learning technologies. To do so, this study first proposes a pro-environmental consumption index based on big data queries in Google Trend to measure the pro-environmental consumption level of each country. In this study, we used the recurrent neural network (RNN) model, which is one of the deep-learning technologies for the prediction of the pro-environmental consumption index, and compared RNN with the OLS regression and artificial neural network (ANN) models to verify prediction accuracy. Based on the results of this study, we identify the framework scalability of this study in environment-related fields, and identify the pros and cons of deep-learning technologies.

This paper is structured as follows. Section 2 reviews the research on pro-environmental consumption levels and examines consumption predictions using big data search queries. We also examine consumption prediction in the environment field, using ANN and RNN. In addition, we investigate advanced research on the ANN and RNN development process. Section 3 derives factors affecting the pro-environmental consumption index. Section 4 compares the prediction accuracy of the regression, ANN, and RNN models. Finally, we discuss our conclusions, as well as the implications and limitations of our study, in Section 5.

## 2. Literature Review

We reviewed the literature on the pro-environmental index to assess the environment level, artificial intelligence, and deep-learning technologies in terms of the methodology. We divided the pro-environmental index into the pro-environmental policy index and the pro-environmental consumption index, and investigated advanced research on consumption predictions based on postings in consumer portals or social network service (SNS) sites. In addition, we investigated consumption predictions in the environment field. Finally, we compared the ANN and RNN methodologies.

### 2.1. Pro-Environment Index

Various pro-environmental indexes have been developed for each country as indicators of pro-environmental policy, consumption, and so on [15–17]. Among the first were the environmental performance index (EPI) and the environmental sustainability index (ESI), which were related to sustainable policy. EPI is an indicator that quantitatively assesses policy performance. This indicator was derived through calculation and aggregation using more than 20 variables that reflect national environmental data. In 2016, the EPI framework was divided into environmental health and ecosystem vitality. Those variables reflect nine issues. “Environmental health” concerns the extent of protection to human health from environmental damage. “Ecosystem vitality” measures ecosystem protection and resource management. “Environmental health” consists of health impacts, air quality, and water and sanitation. The EPI index is calculated using the “proximity to target” methodology. “Proximity to target” is a methodology to calculate how close the policies of each country are to the environmental objectives established by international organizations. This index ranges from a minimum of 0 to a maximum of 100. A score of 100 means the country’s policy wholly meets global standards [16]. South Korea, at an average level of about 70.61 scores, ranked 81st among 180 countries in 2016 [16]. Although the EPI score of South Korea slightly increased from 63.79 in 2014, its ranking declined from 43 in 2014 [15]. The ESI was investigated from 1999 to 2005 prior to the EPI. The ESI, which was developed to assess the relative sustainability of each country’s environment, evolves into the EPI.

Establishment and implementation of national environmental policies in line with international standards contribute to sustainable development, so the present policy condition needs to be identified through indexes such as the EPI and ESI. In addition, since individual pro-environmental consumption such as energy consumption and recycling has a great influence on sustainable development, it is also important to identify a pro-environmental consumption index that comprehensively represents consumers’ sustainable behaviors. Actually, Kao and Tu [18] found, through an analysis of correlation between pro-environmental consumption index and sustainable attitude and behavior, that pro-environmental consumption is closely associated with pro-environmental behavior. That is, they validated a relationship between pro-environmental consumption index and behavior. Therefore, our research focuses on predicting the pro-environmental consumption index, which indicates sustainable behavior.

We examined the existing pro-environmental consumption indexes such as the consumer environment index (CEI) and the Greendex. CEI indicates consumption behavior change and the impact of consumption on the environment. This index tracks the impact of production, use, and recycling of products purchased by Washington consumers each year, as well as the environmental emission trends. CEI calculates the environmental impact and waste caused by the production, use, and disposal of products and services purchased by consumers based on their economic life cycle combined with consumption patterns. CEI mainly deals with contents related to public health harm from climate change, microfossils, and carcinogenic substances and ecosystem harm from toxic materials and measures the potential effect of our behaviors on the environment, climate change, and health. This index footprints track the environmental impact on product purchasing. Therefore, government organizations and corporations utilize this index as a guide to stimulate sustainable behavior and reduce environmental impacts [19].

In addition, Greendex [20] monitors consumer behavior to measure sustainable consumption in National Geographic and GlobeScan. Greendex studied 18,000 consumers in 18 countries during a specific period. The questionnaire consists of 32 items, including energy use and conservation, transportation, food, relative use of eco products versus traditional products, sustainable attitudes, and knowledge of environmental issues. Greendex measures lifestyle and consumption behavior and is divided into four categories: housing, transportation, food, and goods. This index reflects the environmental impact related to consumption patterns of respondents. In addition, Greendex scores are indicated for 1000 consumers by country on a scale of 0–99. The indicators are derived with weights assigned to each category. Housing and transportation are weighted 30%, and food and goods 20%. Greendex also estimates changes in consumer behavior over time using the Market Basket index. Greendex measures energy-saving consumer behavior, whereas the Market Basket index indicates whether a country's energy consumption increases or decreases.

## 2.2. Consumption Prediction

Recently, many researchers have studied pro-environmental consumption and household indexes as well as suicide rate predictions using messages posted by Internet users on Google Trend, Tweets, and so on. Lee et al. [5] analyzed the impact of the pro-environmental consumption attitudes of residents on the adoption of green policies. They estimated pro-environmental attitudes using search query data provided by Google Trend and confirmed, through OLS regression analysis, that a pro-environment attitude has a positive correlation with the pro-environmental attitude index. They also explained that environment-friendly attitudes of residents play an important role in policy making. In the past, most household consumption indexes were calculated through surveys, but big data have recently gained research attention. Vosen and Schmidt [21] calculated the household consumption index using data related to Internet search behavior provided by Google Trend. They explained that search query data can be a new indicator of household consumption because it is more closely related to individual consumption decisions than the market propensity index provided in the survey. Meanwhile, Lee et al. [22] measured past orientation using search queries on Google Trend. They also investigated the relationship between unemployment, the Gini coefficient, the population growth rate, gross state product (GSP), past orientation, and the suicide rate. Using least-squares regression as an analysis methodology, they found a positive correlation between past orientation, the unemployment rate, the Gini coefficient, the population increase rate, and the suicide rate. However, the suicide rate has a negative correlation with GSP.

In addition, some researchers studied consumers' consumption status using tweets. Korpusik et al. [23] argued that consumption can be identified through tweet data analysis, and applied his argument using a neural network model and a regression model. Monitoring consumer interest in the product based on vocabulary data, they analyzed whether product recommendations would be accepted by the consumers. The research targets are cameras and mobile devices, and tweet users are categorized according to whether or not they have purchased the product in the past and want or do not want to purchase it based on specific vocabulary. The research was conducted on 2403 mobile device users and 1252 camera users. This study also compared prediction accuracy using logistic regression, feed-forward (FF) network analysis, RNN, and Long-Short Term Memory (LSTM) analysis. Korpusik et al. [23] constructed a model by including the input layer, the relevance sub-network layer, the buy sub-network layer, and the output layer in order to predict whether the tweet users would purchase a product or not. The accuracy of the FF network analysis was 81.2% for mobile phones and 80.4% for cameras, and the performance of this methodology was the highest among all methodologies. In addition, RNN and LSTM indicated almost similar accuracy. The accuracy of RNN analysis was 80.1% for mobile phones and 79.2% for cameras, and the accuracy of LSTM analysis was 80.2% for mobile phones and 77.0% for cameras.

Consumption prediction studies have also been conducted in the field of environment. Most studies predicted energy and power consumption. Kalogirou and Bojic [13] forecasted the energy

consumption of a solar building using ANN. They classified buildings into insulated buildings and partially insulated buildings, and constructed a dynamic adiabatic building based on time and volume to evaluate thermal patterns. Their results showed a discriminant coefficient ( $R^2$ ) value of 0.9985 from a training data set and a discriminant coefficient ( $R^2$ ) value of 0.9991 from a test data set. Marvuglia and Messineo [14] used RNN to predict power consumption after one hour. They used weather and power consumption data for 79 weeks. Their prediction error was 1.5%, and the maximum error was 4.6%. Although energy and power consumption in the environment field have been predicted, few studies have predicted environmental consumption or the pro-environmental consumption index. Therefore, we attempt to predict the pro-environmental consumption index by using deep-learning technology, specifically the RNN, and compare the prediction accuracy of the model with that of other models such as regression and ANN.

### 2.3. Artificial Intelligence

ANN, developed by McCulloch and Pitts [24], is a non-linear algorithm that generalizes the human brain to a mathematical model. They proposed a computing element model called the McCulloch–Pitts neuron. The significance of ANN is that it forms algorithms by mathematically modeling the activation of neural networks, consisting of synapses that interconnect neurons, but it has a limitation in that its learning performance declines because the weight is fixed. Then, Rosenblatt [25] developed perceptron in 1957. Perceptron has the advantage of repeated learning in which weights are adjusted to improve learning performance. However, this methodology has limitations that apply only to linear separable analyses such as AND and OR operations [25]. Minsky and Papert [26] developed a neural network model that can perform a XOR operation to solve the limitations of perceptron. However, the neural network model has received little research interest due to its large computational cost and functional limitations. However, neural networks have been actively researched since the introduction of multi-layer perceptron in 1986 [27]. The multi-layer perceptron solved the non-linear separation problem, and became the basic structure of the current neural network model. The multi-layer perceptron model performs the XOR operation by stacking perceptron and determines the weight required to minimize error through repeated learning of input and output data. This methodology consists of an input layer, an output layer, and a hidden layer, and is divided into an FF neural network, a backpropagation model, and RNN according to the learning method. The FF neural network is a model in which vector values are transferred from the input layer to the hidden layer and from the hidden layer to the output layer, but no circulating path exists [28]. The backpropagation model updates the weights by calculating the error between the predicted and actual output values [29].

The RNN model includes a connection node that can cycle the previous output values into the input values. That is, RNN is an extension of the multilayer perceptron. The multilayer perceptron simply connects to the output vector from the input vector, but the RNN can remember the previous output values and use it as the input value again [30]. RNN has the advantage that it can use values calculated in the past through context units. Table 1 presents a comparison between the ANN and RNN models. This study verifies the prediction accuracy of the pro-environmental consumption index based on RNN by comparing the model with OLS linear regression and ANN.

In this paper, we tried to analyze data using OLS regression, ANN, and RNN. A Neural Network (NN) is a mathematical modeling method that mimics the process by which the brain transmits information through the synapses between neurons. NN is a nonlinear model that learns to optimize data [31]. Therefore, it is suitable for analyses that comprise a variety of causal factors. An OLS regression is a model used to explain the linear relationship between independent and dependent variables [32]. If the linear relationship between the independent and dependent variables is strong, OLS regression offers a good analytical result. However, social phenomena are generally composed of non-linear relationships and multiple causalities. Pro-environmental consumption index prediction is also a social phenomenon and a complex decision with diverse causality. Therefore, we have tried to implement an optimized algorithm using ANN and RNN.



**Table 1.** Comparison of ANN and RNN [30].

Category		ANN	RNN
Model		Input layer, hidden layer, output layer	Input layer, hidden layer, output layer, recurrent weight
Learning method		In Feed-Forward Neural network, values are transferred from the input layer to the hidden layer and from the hidden layer to the output layer, but no circulating path exists. Backpropagation is a model that updates the weights by calculating the error between the output value and the actual value learned in the same way as the Feed-Forward Neural network	BPTT (Back-Propagation Through Time) * allows time to unfold through recurrent weight

Note: \* Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., & Khudanpur, S. (2010).

### 3. Methodology

#### 3.1. Model

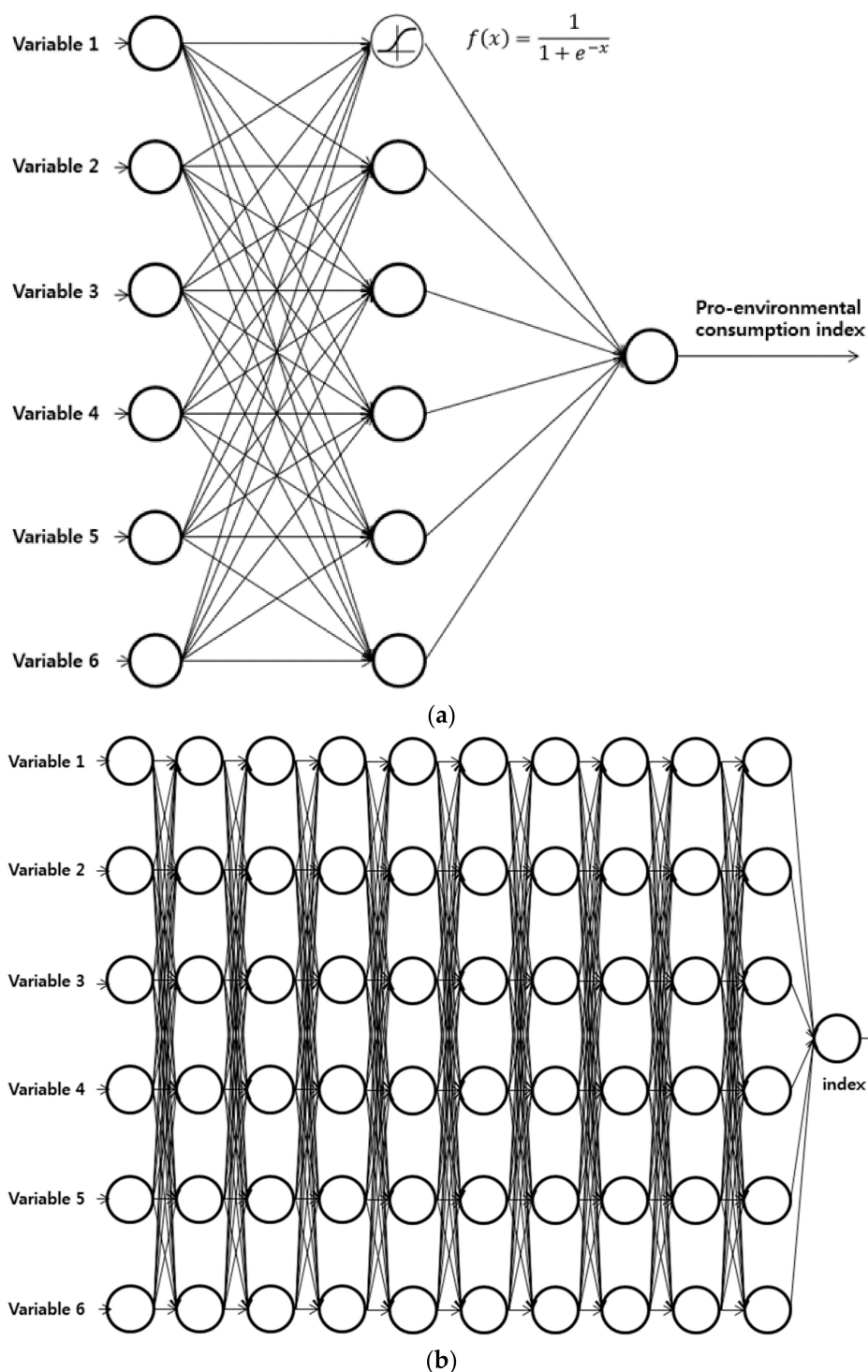
In order to compare the prediction performance for the pro-environmental consumption index and investigate the possibility of applying artificial intelligence to environment studies, this study considers the traditional regression analysis model, ANN, and RNN. We selected the OLS regression model from among the traditional regression analysis models based on constructed data and used variables satisfying the OLS regression model assumption. We analyzed the prediction performance by comparing the root mean square error (RMSE) between the predicted and actual values. The regression model is defined in Equation (1), where  $i$  and  $t$  represent the country and the year, respectively.

$$\begin{aligned}
 \text{Pro-environmental consumption index}_{it} &= \alpha_i + \beta_1(\text{Health expenditure}_{it}) + (\text{Age 65 and above}_{it}) \\
 &+ \beta_3(\text{Preprimary education}_{it}) + \beta_4(\text{Low GDP country} \times \text{GDP}_{it}) \\
 &+ \beta_5(\text{High GDP country} \times \text{GDP}_{it}) + \beta_6(\text{Past orientation}_{it} + \epsilon_i)
 \end{aligned} \quad (1)$$

As for the ANN model, this study analyzed the prediction values after constructing three ANN models with 1, 9, and 100 hidden layers, respectively. The hidden layer was designed differently to capture the effect of the number of hidden layers on the prediction performance. For the implementation of ANN, we used Frauke Günther and Stefan Fritsch's neuralnet algorithm [33]. Specifically, the commonly used sigmoid (or logistic) function was utilized as the perceptron activation function of input  $x$  and output  $y$ , as shown in Equation (2).

$$\begin{aligned}
 \text{net value} &= \sum_{i=0}^n w_i x_i \\
 y = f(\text{net value}) &= \frac{1}{1 + e^{-\sum_{i=0}^n w_i x_i}}
 \end{aligned} \quad (2)$$

Furthermore, we use a resilient backpropagation with a weight-backtracking algorithm to find the weight ( $w_i$ ) that minimizes the error. Of the three models, the ANN model has the same structure regardless of whether it has 1 hidden layer (or indeed none at all) or 100 hidden layers, except that the number of layers increases (see Figure 1).



**Figure 1.** ANN model. (a) ANN model with one hidden layer; (b) ANN model with nine hidden layers.

Finally, we forecasted the pro-environmental consumption index using the RNN technology, which is one of the deep-learning technologies developed through the ANN model. In the analysis of time series data, the traditional ANN does not consider time, so the effects of previous relationships are difficult to reflect. On the other hand, RNN can consider previous relationships. That is, it incorporated data of a prior period ( $h_{t-1}$ ) recursively as input data and calculated the output  $y_t$  for time  $t$ .

$$y_t = f(h_{t-1} \times W^h + x_t \times W^x) \quad (3)$$

We implemented the RNN algorithm of Bastiaan Quast using the R package [34]. The RNN model was designed with Gompertz of activation function, 100 repetitions of Epoch, 10 hidden layers, 20 batch sizes, and a learning rate of 0.02. Data training and prediction proceeded as follows: 85% of data were used as model estimation and training values in all models, including regression analysis. We also used 15% of the data as verification data to compare prediction values. To improve the verification reliability, we undertook repeated experiments with 84 different datasets. We have used an i7-4790k quad-core CPU, a GTX970 GPU, and a 32GB RAM for our analysis.

### 3.2. Data

We use the RNN, ANN, and OLS methodologies to predict pro-environmental consumption index values. In this chapter, we select the variables that affect pro-environmental index values, by undertaking a literature review to improve the prediction accuracy of the three models (RNN, ANN, and OLS). Data are collected from 13 countries worldwide—namely, Argentina, Australia, Brazil, Canada, China, Hungary, Japan, Mexico, South Africa, South Korea, Spain, the United Kingdom, and the United States—collected through the World Bank [35] and Google Trends.

#### 3.2.1. Dependent Variable: Pro-Environmental Consumption Index

Lee et al. [6] propose a pro-environmental consumption index that leverages the search words “recyclables” and “disposables”; they define the index as the ratio of the number of search queries for “recyclables” to the number of search queries for “disposables.” Lee et al. [6] selected the nouns “recyclables” and “disposables” as keywords, reasoning that these nouns are used to assess the index value by determining product consumption, rather than people’s attitudes. Additionally, compared to verbs and adjectives, nouns and noun phrases could better reflect whether or not people actually bought a product. The pro-environmental consumption index is calculated as follows.

$$\text{Pro – environmental consumption index} = \frac{(\text{Number of search queries for “recyclables”})}{(\text{Number of search queries for “disposables”})} \quad (4)$$

When the number of search queries for “recyclables” is identical to that for “disposables,” the index value is 1; hence, the index value grows when the number of search queries for “recyclables” exceeds that for “disposables” [6]. To demonstrate the usefulness of this index, Lee et al. [6] conducted correlation analysis with the Greendex consumption index. (The Greendex index is a global index developed by National Geographic and GlobeScan research.) Lee et al. [6] found a positive correlation between the pro-environmental index and the Greendex index, with a correlation coefficient of 0.3716 at the 99.9% significance level. Having thoroughly reviewed this study, we decided to use as a dependent variable the index suggested by Lee et al. [6], to predict pro-environmental consumption index values. We constructed index data by leveraging search queries performed in the main language of each of the 13 countries.

#### 3.2.2. Independent Variable

##### (1) Health expenditure

Several studies have demonstrated that the more people are interested in health, the more they consume organic products, one of the sustainability product types. Grankvist [36] and Magnusson et al. [37] assert that health concerns were the strongest predictor of organic product consumption. Magnusson et al. [37] proposed that organic product consumption reflects on the pro-environmental behavior index. That is, organic food consumption and health concerns are related to purchase frequency. In addition, the intention to consume organic food is closely linked to environmental interest [38,39].

Lee et al. [6] investigated the relationship between health expenditure variables and a pro-environmental consumption index, and found it to be negative. They found that if per-capita health expenditures were to be increased by USD 1000, the pro-environmental consumption index value would decrease by 2.1 points.



This study uses health expenditure as a variable of health concerns based on previous research findings that higher health concerns lead to more pro-environmental consumption and behavior. The variable values obtained from per capita medical spending data for each country provided by the World Bank.

## (2) Age 65 and above

Many studies have investigated the relationship between age and pro-environmental behavior. Poortinga et al. [40] discussed the determinants of environmental behavior in the field of household energy use. They classified pro-environmental behavior as home energy use and transport energy use. Age was significantly negatively correlated with transport energy use. In addition, Zimmer et al. [3] argued that consumer perception about green issues is related to marketing. They found a statistically significant negative relationship between age and green attitude. Moreover, Liere and Dunlap [41] emphasized that young people were more interested in the environmental quality than older people. However, some studies argue a positive relationship between age and pro-environmental behavior. Vining and Ebreo [42] examined the differences in recycling by voluntary users using variables such as demographic characteristics, knowledge, and motivation. The study found that older residents were more likely to recycle than young residents.

Overall, many researchers have considered the relationship between age and pro-environmental behavior, but express conflicting opinions. This study predicts the pro-environmental index using the age variable. The variable value is obtained from World Bank data.

## (3) Preprimary education

Preprimary education means an earliest step of regular education that provides a similar school environment to help young children adapt before entering into the education system [35]. Many researchers have studied the relationship between the education level and pro-environmental behavior. Poortinga et al. [40] demonstrated that a higher education level leads to saving of transport energy. In addition, Buttel and Flinn [43], Roberts [44], and Tilikidou [45] showed that the relationship between educational level and pro-environmental behavior is positive. According to Whitmarsh and O'Neill [46], eco-shopping and eating had a positive relationship with a high level of education. However, regular water/energy conservation had a positive relationship with a low level of education. In analyzing the disturbance factor of pro-environmental behavior, Kollmuss and Agyeman [47] found that concerns regarding environmental issues relate to having attained higher levels of education; however, this relationship is not statistically significant. This study also drew conflicting conclusions regarding the relationship between education level and pro-environmental attitude and behavior.

Meanwhile, some researchers have proposed a positive relationship between childhood experience and pro-environmental behavior. Chawla [48] interviewed 30 environmentalists to determine the factors behind their environmental dedication. This study investigated the life experience (childhood, university years, adulthood) of the environmentalists. The analysis results revealed the childhood factors associated with an environmental commitment: experience of natural areas, organizations, education and family. That is, pro-environmental attitudes are related to childhood experiences. Lee et al. [6] discovered that there is a positive relationship between pre-primary education and pro-environmental consumption index values, at the 0.1% significance level. With one additional year of pre-primary education, the pro-environmental consumption index value was found to increase by 4.3 points. Based on a review of relevant research, we used a preprimary education (childhood education) variable to predict the pro-environmental consumption index.

## (4) Low/High GDP country $\times$ GDP

Many researchers have analyzed the relationship between income and environment-friendly behavior, whereas studies on the relationship between GDP and pro-environmental attitudes are limited. Berger [49] analyzed the influence of the demographics variable on pro-environmental behavior for 43,000 households through a survey. They used indicators such as income and

education. They found that income was a significant determinant of pro-environmental behavior. Gatersleben et al. [50] studied pro-environmental behavior in a psychological setting, and compared it to the usual pro-environmental behavior in the social science sphere. They argued that organic food and recycling are strongly related to age and income. In addition, Clark et al. [51] concluded that the income variable has a statically significant positive correlation with the green electricity participation decision.

Franzen and Meyer [52] investigated the determinants of pro-environment interests based on the wealth and income variables. They used per capita GDP as the purchasing power variable. The analysis showed that developed countries, with higher per capita GDP, are more concerned about the environment than developing countries.

Most past studies have identified the relationship between income and pro-environmental behavior. In addition, researchers discovered that the GDP variable has a positive relationship with environmental concern. Therefore, we predict the pro-environmental consumption index using GDP variables, divided into high-GDP (USD 30 k and higher) and low-GDP (up to USD 10 k) dummy variables.

#### (5) Past orientation

Studies have investigated the relationship between time orientation and pro-environmental behaviors. Corral-Verdugo et al. [53] investigated the relationship between the time perspective and sustainable behavior (water conservation). In this study, time orientation is divided into the past, present, and future. They claimed that the relationship between future-oriented individuals and water conservation had a higher correlation, but a past orientation did not affect sustainable behavior. Milfont et al. [54] suggested a high correlation between the time perspective and pro-environmental behavior. Future orientation had a higher association with sustainable behavior than past orientation. In other words, future orientation played an important role as a determinant of pro-environmental behavior.

Meanwhile, Preis et al. [55] developed a past and future orientation index using search queries on Google Trend. They measured this index as the ratio of number of Google search queries for past years to number of Google search queries for future years. Lee et al. [22] also used a past orientation index using search queries on Google Trend. Therefore, we use search queries on Google Trend to measure past orientation, as suggested by Lee et al. [22] and Preis et al. [55], in order to predict the pro-environmental consumption index.

We assume that the independent variables affect the dependent variables one year later, so we used each independent variable with a lag. The data used for the analysis are shown in Tables 2 and 3.

**Table 2.** Data description.

Variable	Description	Unit	Scale	Source
Pro-environmental consumption index	Lag value of pro-environmental consumption index	0–100 (Index)	/10 <sup>2</sup>	Google Trends ( <a href="http://www.google.com/trends/">http://www.google.com/trends/</a> )
Health expenditure	Lag value of health expenditure		/10 <sup>3</sup>	World Bank
Age 65 and above	Lag value of population aged 65 and above (% of total)	%	/10 <sup>2</sup>	World Bank
Preprimary education	Lag value of preprimary education	Years		World Bank
Low-GDP country × GDP	Interaction term: (country with GDP of USD 10,000 or less) × Lag value of GDP per capita	USD	/10 <sup>4</sup>	World Bank
High-GDP country × GDP	Interaction term: (country with GDP of USD 30,000 or more) × Lag value of GDP per capita	USD	/10 <sup>4</sup>	World Bank
Past orientation	Lag value of past orientation	0–100 (Index)	/10 <sup>2</sup>	Google Trends

**Table 3.** Summary statistics.

Variable	Obs.*	Mean	s.e.	Min	Max
Pro-environmental consumption index	130	0.049	0.063	0.000	0.280
Health expenditure	130	2.450	2.284	0.071	8.988
Age 65 and above	130	0.121	0.050	0.045	0.250
Preprimary education	130	2.308	0.879	1.000	4.000
Low-GDP country $\times$ GDP	130	0.214	0.358	0.000	1.304
High-GDP country $\times$ GDP	130	1.934	2.186	0.000	6.765
Past orientation	130	0.012	0.004	0.006	0.027

Notes: s.e.: standard error, \* 130 Obs. = 13 countries  $\times$  10 years (2005–2014).

#### 4. Results and Discussion

In this study, we used the RNN model to predict the pro-environmental consumption index. In addition, this study investigated the pros and cons of the RNN model in terms of prediction power by comparing the prediction accuracy between the existing regression analysis model and the ANN model. We randomly selected 85% of the data as training data for model estimation and learning. The remaining 15% of the data were used to test the predictive power. We repeated this process 84 times and averaged. The determinants of the pro-environmental consumption index based on the regression analysis model are shown in Table 4.

**Table 4.** Regression results.

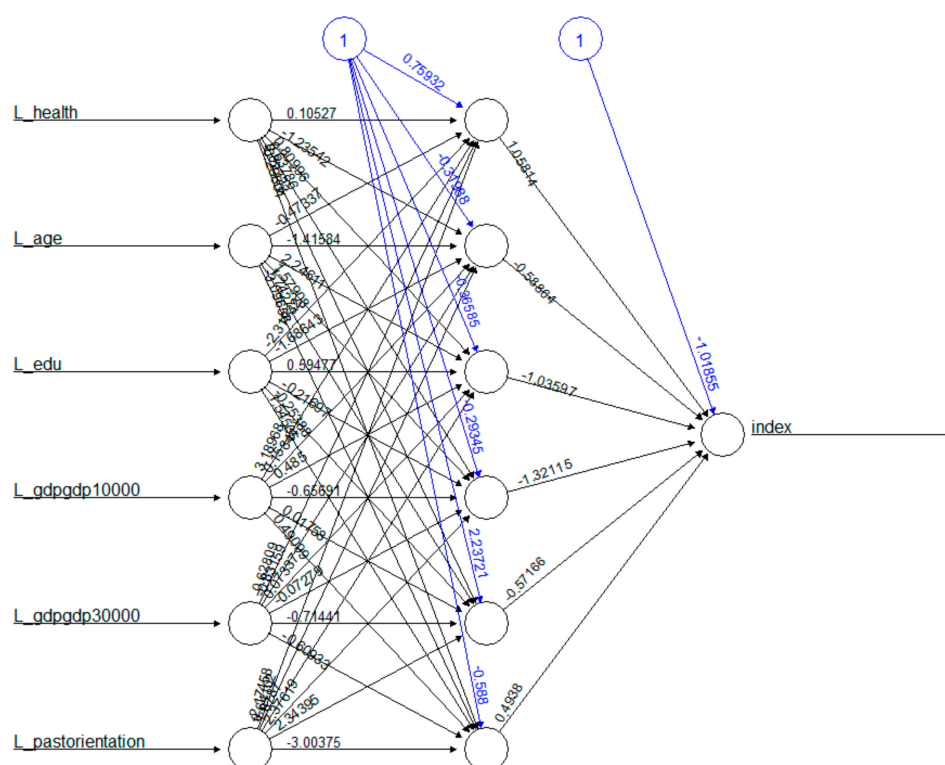
	Coef.	s.e.	P > T
Health expenditure	−0.018 ***	0.006	0.003
Age 65 and above	−0.772 ***	0.213	0.000
Preprimary education	0.028 ***	0.009	0.003
Low-GDP country $\times$ GDP	0.012	0.018	0.502
High-GDP country $\times$ GDP	0.009	0.008	0.251
Past orientation	−4.982 ***	1.243	0.000
_cons	0.168 ***	0.031	0.000
P > F	0.000		
R <sup>2</sup>	0.471		
Adjusted R <sup>2</sup>	0.440		
Number of observations	110		

Notes: \*\*\* Significance Level  $p < 0.01$ .

The one-year lag value of health expenditure, age 65 and above, and past orientation show a negative correlation with Google queries related to pro-environmental consumption at the significance level of 1%. However, preprimary education shows a statistically significant positive correlation with Google queries. This study predicted the pro-environmental consumption index based on the regression result to compare the prediction power. From the comparison, the RMSE of the regression model was 0.0765.

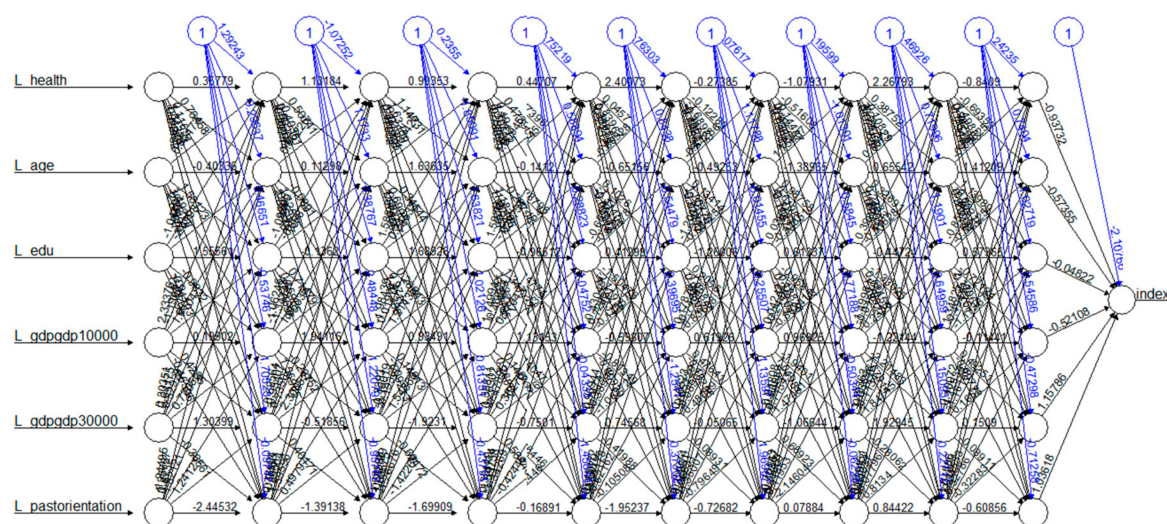
Next, we predicted the pro-environmental consumption index using the ANN model. Among the ANN models considered in this research, the ANN model with one hidden layer was the one first analyzed, from which we derived model shown in Figure 2.

Since the ANN model has a hidden layer, a simple interpretation of the weights is difficult. Running the data through the model, we obtained an RMSE value of 0.0712 by comparing the predicted and actual values. This was slightly lower than the RMSE value obtained through OLS.



**Figure 2.** Results of ANN model with one hidden layer.

We also analyzed a deep neural network with nine hidden layers, the results of which are shown in Figure 3.

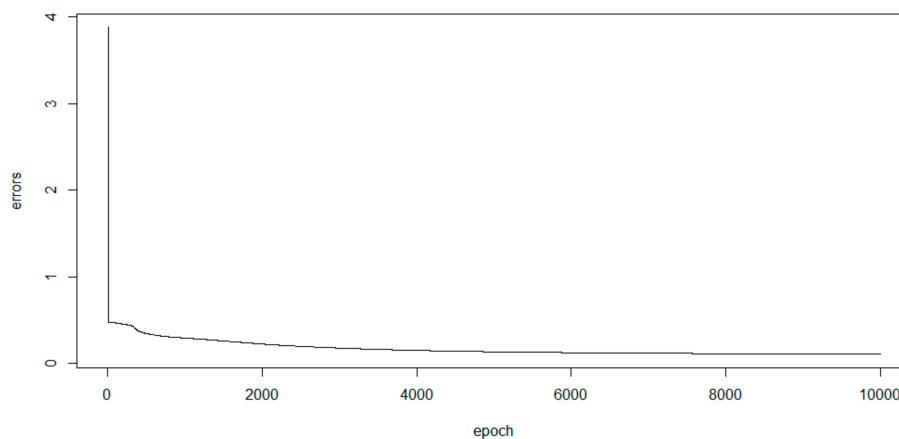


**Figure 3.** Results of ANN model with nine hidden layers.

The model is difficult to interpret as in the previous ANN. Since the coefficients were calculated by the interaction of several hidden layers through the backpropagation process, the effect of each hidden node needs to be analyzed by classifier rather than according to the unilateral influence of the specific variable. However, we can see the influence of the variables on the output by summing the weights. The ANN model with nine hidden layers yielded an RMSE of 0.0677, lower than the RMSE based on the ANN model with one hidden layer.

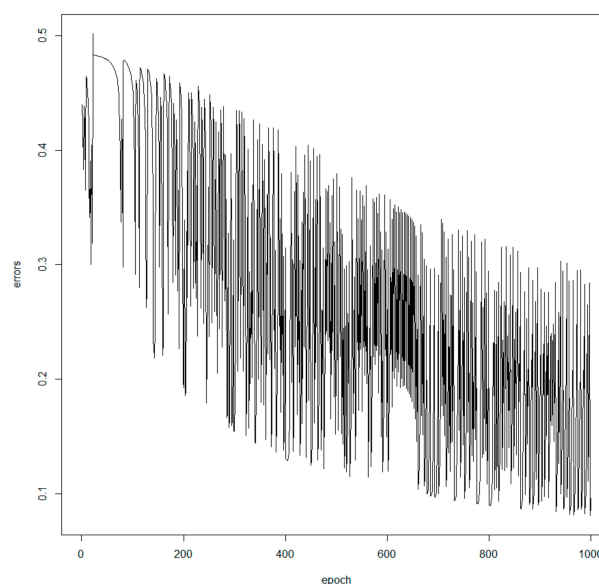
A model with 100 hidden layers provided an RMSE of 0.0675. This result indicates that the RMSE of the model did not significantly improve, in line with the rapid increase from 9 to 100 hidden layers.

Finally, we designed an RNN model, which is one of the deep-learning technologies, and trained it with data. We set parameters such as epoch and learning rate for the RNN model. We discovered from this study that as the epoch increased, the error tended to decrease, first rapidly and then at a low rate. Figure 4 shows the error change in line with the change in epoch.



**Figure 4.** Change in error value according to epoch in RNN.

In addition, we supposed that the model might be subject to a robustness problem due to large error fluctuations when the learning rate is too high. Figure 5 shows that setting an appropriate learning rate is important for RNN analysis.



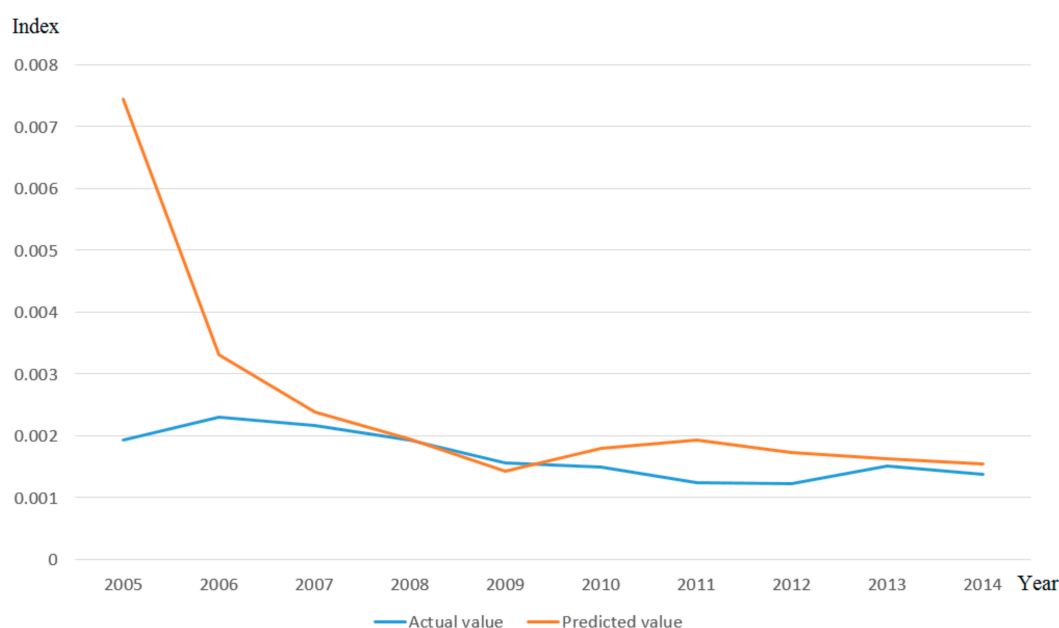
**Figure 5.** Learning rate.

The RMSE of the RNN model was 0.0576. The RMSE of the regression model, at 0.0765. This indirectly indicates that RNN and deep learning can be effectively used to make economic forecasts in the environmental field through time-series data.

Based on the RNN model, we have predicted the United Kingdom's pro-environmental consumption index in Figure 6. In the training process of this RNN model, data from the United Kingdom were not utilized. If we compare the predicted value with the actual value through the RNN,



the forecast is relatively accurate for all the years, except for 2005. If the training data is increased and the number of learning epochs is increased, the prediction rate will be higher.



**Figure 6.** The United Kingdom's pro-environmental consumption index and the predicted value of the RNN model.

Finally, RMSE Comparison between OLS, ANN, and RNN are shown in Table 5. To summarize, ANN and OLS indicated relatively similar RMSE coefficients. ANN with one hidden layer showed higher prediction power than OLS. ANN models with 9 and 100 hidden layers have somewhat higher prediction power than ANN with one hidden layer. The RMSE coefficients differed considerably between the ANN model with one hidden layer and that with nine hidden layers. However, there was no significant difference between the RMSE coefficients of the 9-layered and 100-layered models, although the model design allowed for a rapid increase in the number of hidden layers to 100. Finally, the RNN model shows the highest prediction performance because the RMSE was the lowest. Prediction accuracy improved considerably for RNN compared with other models because the previously considered RNN model could not be solved with any number of ANN layers.

**Table 5.** RMSE Comparison between OLS, ANN, and RNN.

Rank	Model	RMSE
1	RNN	0.0576
2	ANN (100 hidden layers)	0.0675
3	ANN (9 hidden layers)	0.0677
4	ANN (1 hidden layer)	0.0712
5	OLS	0.0765

The deep-learning model will increase in sophistication with the growth of data and therefore can be expected to greatly improve its performance in real big data environments. The deep-learning model is also useful for studies on climate change and in environmental fields that require elaborate forecasting.

## 5. Conclusions

In this study, we proposed a pro-environmental consumption index using big data queries to measure the environmental consumption level for each country, and predicted the proposed

index using deep-learning technology in the context of its application to environmental studies. This study used the RNN model, which is one of the deep-learning technologies, and compared its prediction accuracy with that of the OLS regression model and the ANN model to verify the prediction power of deep-learning technology. Therefore, this study derived implications for the extensibility of environmental research, as well as the advantages and disadvantages of the deep-learning technology.

The implications of this study are as follows. When forecasting the pro-environmental consumption index using the OLS, ANN, and RNN models, the predicted value of the RNN model, which is one of the deep-learning technologies, was closest to the actual value. Therefore, the RNN model and deep-learning technologies, which can be considered for time series input, are expected to be useful in the environmental field. In particular, it seems to apply to prediction research in the environmental field to prevent damage from climate change. Despite the sharp increase in the number of hidden layers in the ANN from 9 to 100, the RMSEs of the two models were almost equal, at 0.0677 and 0.0675, respectively. It is likely that the efficiency drops when the gradient tends to vanish as it passes through successive layers of the neural network.

However, the implication of each coefficient and variable in artificial neural networks such as RNN and ANN is not easy to infer, unlike in regression analysis. Although the impact of these variables can be determined by assigning weights, for example, it is difficult to clearly identify the influence, as in regression analysis. In summary, rather than increase predictability, deep-learning techniques can be seen as comprising a “black box” model that has weaker explanatory power. Therefore, it appears that in addition to the existing regression analysis, parallel research is needed to pinpoint policy implications.

However, deep-learning and artificial intelligence technologies that provide policy implications can be expected to emerge in the near future, considering that artificial intelligence, which allows data features to be automatically extracted, has recently been used to conduct unsupervised learning.

Finally, Google trends data may have some limitations. For example, Google measures data for Internet users only. In particular, using Google trends data for early 2005, when the Internet was not significantly developed, may cause some biases. To compensate for these limitations, we used pre-verified indicators that use Google trends through correlation analysis with existing indexes.

We showed that deep-learning technology can be fully utilized in environmental studies. In particular, deep-learning techniques such as RNN can be usefully applied to climate change, which could be predicted mainly from time-series data. For policy implications, however, neural networks, including deep learning, need to be analyzed in combination with regression analysis because no neural network can clearly indicate what each model coefficient means.

**Acknowledgments:** This work was supported by the Academic Promotion System of Korea Polytechnic University and this work was also supported by a grant from Kyung Hee University in 2016(KHU-20161373).

**Author Contributions:** Donghyun Lee and Jungwoo Shin conceived and designed the research; Donghyun Lee collected data and analyzed the data; Suna Kang and Jungwoo Shin contributed to progress of research idea; Donghyun Lee, Suna Kang and Jungwoo Shin wrote the paper. All authors have read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Korea Ministry of Environment. *Public Organizations Spend 2.2 Trillion Korean Won to Purchase Green Products in 2014*; Ministry of Environment: Sejong, Korea, 2015.
2. Young, W.; Hwang, K.; McDonald, S.; Oates, C.J. Sustainable consumption: Green consumer behaviour when purchasing products. *Sustain. Dev.* **2010**, *18*, 20–31. [[CrossRef](#)]
3. Zimmer, M.R.; Stafford, T.F.; Stafford, M.R. Green issues: Dimensions of environmental concern. *J. Bus. Res.* **1994**, *30*, 63–74. [[CrossRef](#)]
4. Organisation for Economic Co-operation and Development (OECD). *Society at a Glance: Asia/Pacific 2014*; OECD: Paris, France, 2014.

5. Lee, D.; Kim, M.; Lee, J. Adoption of green electricity policies: Investigating the role of environmental attitudes via big data-driven search-queries. *Energy Policy* **2016**, *90*, 187–201. [CrossRef]
6. Lee, D.; Kang, S.; Shin, J. Determinants of Pro-Environmental Consumption: Multicountry Comparison Based upon Big Data Search. *Sustainability* **2017**, *9*, 183. [CrossRef]
7. McCarthy, J.; Minsky, M.L.; Rochester, N.; Shannon, C.E. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Mag.* **2006**, *27*, 12.
8. Konar, A. *Artificial Intelligence and Soft Computing: Behavioral and Cognitive Modeling of the Human Brain*; CRC Press: Boca Raton, FL, USA, 1999.
9. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
10. Tetko, I.V.; Livingstone, D.J.; Luik, A.I. Neural network studies. 1. Comparison of overfitting and overtraining. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 826–833. [CrossRef]
11. Bengio, Y. Learning deep architectures for AI. *Found. Trends<sup>®</sup> Mach. Learn.* **2009**, *2*, 21–127. [CrossRef]
12. White, G. Google Chairman Wants AI Robots to “Solve Problems” of Overpopulation, Climate Change and Education. Available online: [www.newstarget.com](http://www.newstarget.com) (accessed on 24 November 2016).
13. Kalogirou, S.A.; Bojic, M. Artificial neural networks for the prediction of the energy consumption of a passive solar building. *Energy* **2000**, *25*, 479–491. [CrossRef]
14. Marvuglia, A.; Messineo, A. Using recurrent artificial neural networks to forecast household electricity consumption. *Energy Procedia* **2012**, *14*, 45–55. [CrossRef]
15. Yale University Center for Environmental Law & Policy; Center for International Earth Science Information Network. *Environmental Performance Index. The 2014 Environmental Performance Index Full Report and Analysis*; Yale University Center for Environmental Law & Policy: New Haven, CT, USA; Center for International Earth Science Information Network: Manhattan, NY, USA, 2014.
16. Hsu, A.; Alisa, Z. Environmental performance index. In *Wiley StatsRef: Statistics Reference Online*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2016.
17. SEDAC Homepage. Available online: <http://sedac.ciesin.columbia.edu/data/collection/esi/> (accessed on 24 November 2016).
18. Kao, T.; Tu, Y. Effect of green consumption values on behavior: The influence of consumption attitude. *Int. J. Arts Sci.* **2015**, *8*, 119–130.
19. Morris, J.; Matthews, H.S. Consumer Environmental Index (CEI). Background Information. Available online: [http://www.ecy.wa.gov/beyondwaste/pdf/CEI\\_Background\\_4-23-12.pdf](http://www.ecy.wa.gov/beyondwaste/pdf/CEI_Background_4-23-12.pdf) (accessed on 25 November 2016).
20. Greendex. *Consumer Choice and the Environment—A Worldwide Tracking Survey*; Greendex: Washington, DC, USA, 2014.
21. Vosen, S.; Schmidt, T. Forecasting private consumption: Survey-based indicators vs. Google Trends. *J. Forecast.* **2011**, *30*, 565–578. [CrossRef]
22. Lee, D.; Lee, H.; Choi, M. Examining the relationship between past orientation and US suicide rates: An analysis using big data-driven Google search queries. *J. Med. Internet Res.* **2016**, *18*, 1–12. [CrossRef] [PubMed]
23. Korpusik, M.; Sakaki, S.; Chen, F.; Chen, Y. Recurrent Neural Networks for Customer Purchase Prediction on Twitter. In Proceedings of the CBRRecSys 2016 3rd Workshop on New Trends in Content-based Recommender Systems, Boston, MA, USA, 16 September 2016; pp. 47–50.
24. McCulloch, W.S.; Pitts, W.H. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [CrossRef]
25. Rosenblatt, F. *The Perceptron: A Perceiving and Recognizing Automaton (Project Para)*; Cornell Aeronautical Laboratory: Buffalo, NY, USA, 1957.
26. Minsky, M.; Papert, S. *Perceptrons: An Introduction to Computational Geometry*; Expanded Edition; MIT Press: Oxford, UK, 1988; pp. 157–169.
27. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagation errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
28. Svozil, D.; Kvasnicka, V.; Pospichal, J. Introduction to multi-layer feed-forward neural networks. *Chem. Intell. Lab. Syst.* **1997**, *39*, 43–62. [CrossRef]
29. Hecht-Nielsen, R. Theory of the backpropagation neural network. In Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), Washington, DC, USA, 18–22 June 1989; pp. 593–605.

30. Mikolov, T.; Karafiat, M.; Burget, L.; Cernocky, J.; Khudanpur, S. Recurrent neural network based language model. *Interspeech* **2010**, *2*, 3.
31. Wang, H.; Raj, B.; Xing, E.P. On the Origin of Deep Learning. *arXiv* **2017**, arXiv:1702.07800.
32. Neter, J.; Kutner, M.H.; Nachtsheim, C.J.; Wasserman, W. *Applied Linear Statistical Models*; Irwin: Martinsville, OH, USA, 1996; Volume 4, p. 318.
33. Günther, F.; Fritsch, S. Neuralnet: Training of neural networks. *R J.* **2010**, *2*, 30–38.
34. Quast, B.A.; Fichou, D. RNN: A Recurrent Neural Network in R. Available online: <https://github.com/bquast/rnn> (accessed on 24 November 2016).
35. World Bank. Available online: [www.worldbank.org](http://www.worldbank.org) (accessed on 6 September 2016).
36. Grankvist, G. Purchase Criteria, Beliefs and Habits as Determinants of Choice of Eco-Labeled Products. Ph.D. Dissertation, Göteborg University, Göteborg, Sweden, 2001.
37. Magnusson, M.K.; Arvola, A.; Hursti, U.K.K.; Aberg, L.; Sjoden, P.O. Choice of organic foods is related to perceived consequences for human health and to environmentally friendly behaviour. *Appetite* **2003**, *40*, 109–117. [CrossRef]
38. Tregear, A.; Dent, J.B.; McGregor, M.J. The demand for organically grown produce. *Br. Food J.* **1994**, *96*, 21–25. [CrossRef]
39. Wandel, M.; Bugge, A. Environmental concern in consumer evaluation of food quality. *Food Qual. Preference* **1997**, *8*, 19–26. [CrossRef]
40. Poortinga, W.; Steg, L.; Vlek, C. Values, environmental concern, and environmental behavior a study into household energy use. *Environ. Behav.* **2004**, *36*, 70–93. [CrossRef]
41. Liere, K.D.V.; Dunlap, R.E. The social bases of environmental concern: A review of hypotheses, explanations and empirical evidence. *Public Opin. Q.* **1980**, *44*, 181–197. [CrossRef]
42. Vining, J.; Ebreo, A. What makes a recycler? A comparison of recyclers and nonrecyclers. *Environ. Behav.* **1990**, *22*, 55–73. [CrossRef]
43. Buttel, F.H.; Flinn, W.L. Environmental politics: The structuring of partisan and ideological cleavages in mass environmental attitudes. *Sociol. Q.* **1976**, *17*, 477–490. [CrossRef]
44. Roberts, J.A. Green consumers in the 1990s: Profile and implications for advertising. *J. Bus. Res.* **1996**, *36*, 217–231. [CrossRef]
45. Tilikidou, I. Ecologically Conscious Consumer Behaviour: A Research Project Conducted in Thessaloniki, Greece. Ph.D. Dissertation, University of Sunderland England, Sunderland, UK, 2001.
46. Whitmarsh, L.; O'Neill, S. Green identity, green living? The role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours. *J. Environ. Psychol.* **2010**, *30*, 305–314. [CrossRef]
47. Kollmuss, A.; Agyeman, J. Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environ. Educ. Res.* **2002**, *8*, 239–260. [CrossRef]
48. Chawla, L. Life paths into effective environmental action. *J. Environ. Educ.* **1999**, *31*, 15–26. [CrossRef]
49. Berger, I.E. The demographics of recycling and the structure of environmental behavior. *Environ. Behav.* **1997**, *29*, 515–531. [CrossRef]
50. Gatersleben, B.; Steg, L.; Vlek, C. Measurement and determinants of environmentally significant consumer behavior. *Environ. Behav.* **2002**, *34*, 335–362. [CrossRef]
51. Clark, C.F.; Kotchen, M.J.; Moore, M.R. Internal and external influences on pro-environmental behavior: Participation in a green electricity program. *J. Environ. Psychol.* **2003**, *23*, 237–246. [CrossRef]
52. Franzen, A.; Meyer, R. Environmental attitudes in cross-national perspective: A multilevel analysis of the ISSP 1993 and 2000. *Eur. Sociol. Rev.* **2010**, *26*, 219–234. [CrossRef]
53. Corral-Verdugo, V.; Fraijo-Sing, B.; Pinheiro, J.Q. Sustainable behavior and time perspective: Present, past, and future orientations and their relationship with water conservation behavior. *Interam. J. Psychol.* **2006**, *40*, 139–147.
54. Milfont, T.L.; Wilson, J.; Diniz, P. Time perspective and environmental engagement: A meta-analysis. *Int. J. Psychol.* **2012**, *47*, 325–334. [CrossRef] [PubMed]
55. Preis, T.; Moat, H.S.; Stanley, H.E.; Bishop, S.R. Quantifying the advantage of looking forward. *Sci. Rep.* **2012**, *2*, 350. [CrossRef] [PubMed]

