

## Article

# Advancing Agricultural Predictions: A Deep Learning Approach to Estimating Bulb Weight Using Neural Prophet Model

Wonseong Kim <sup>1</sup>  and Byung Min Soon <sup>2,\*</sup> 

<sup>1</sup> Department of Economics and Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea; wonseongkim@korea.ac.kr

<sup>2</sup> Department of Agricultural Economics, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea

\* Correspondence: soonbm@cnu.ac.kr; Tel.: +82-42-821-6748

**Abstract:** A deep learning methodology was utilized to predict the bulb weights of garlic and onions in the Jeolla Province of Korea. The Korea Rural Economic Institute (KREI) operates the Outlook & Agricultural Statistics Information System (OASIS) platform, which provides actual measurements of garlic and onions. We trained the Neural Prophet (NP) lagged time-series model using this data. The NP model effectively handles lagged variables and their covariates by inserting a hidden layer. Our results indicate that the NP model performed with around 5% mean absolute error in predicting bulb weights, with a gap of 3.3 g and 4.7 g with average weights of 63.7 g and 129.9 g for garlic and onions, respectively. This experimental research was based on only three years of measurement data. Hence, the gap between observed and predicted data can be reduced by accumulating more measurement data in the future.

**Keywords:** deep learning; Neural Prophet; bulb weight prediction; agricultural production; measurement data



**Citation:** Kim, W.; Soon, B.M.

Advancing Agricultural Predictions: A Deep Learning Approach to Estimating Bulb Weight Using Neural Prophet Model. *Agronomy* **2023**, *13*, 1362. <https://doi.org/10.3390/agronomy13051362>

Academic Editors: Borja Espejo-García, Spyros Fountas and Georgios Leontidis

Received: 20 March 2023

Revised: 28 April 2023

Accepted: 10 May 2023

Published: 12 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Predicting crop production provides crucial information for policymaking. When an oversupply is anticipated, the government can take preemptive actions to prevent price collapses or encourage farmers to adjust production levels or switch to alternative crops. On the other hand, if a supply shortage is expected, temporary import expansion or government purchase policies may be necessary. Garlic and onions, which are particularly sensitive to climate change and prone to price fluctuations, are essential ingredients in Korean cuisine. Their demand remains stable as they are consumed both at home and in restaurants. Proactive forecasting for these sensitive crops can contribute to supply stability, helping to inform policy decisions.

Foreign agricultural agencies conduct annual field surveys of major crops. The US National Agricultural Statistics Service (USDA-NASS) [1] investigates corn, soybeans, cotton, winter wheat, and potatoes through local field surveys and predicts their production. The NASS estimation program, which involves farm-level and objective yield surveys based on field measurements at key production sites, derives yield forecasts. Collected stage-specific measurement data is used for yield prediction through mathematical models [2]. The Belgian Statistical Agency utilizes a crop growth monitoring system to predict the production of six major crops. The model uses various information, such as weather, soil, and remote sensing, to predict yield [3]. In this way, major foreign agencies aim to contribute to supply stability by predicting crop production through field surveys of sensitive crops.

Since 2020, the Korea Rural Economic Institute (KREI) has been conducting field surveys of crops in each province. The measured data are used to predict the production

volume for the current year, and various studies have been presented for this purpose. Using growth-stage field surveys and meteorological information, Kim and Kim [4] predicted the bulb weight of onions. In particular, using a multilevel model, the predictive power of onion bulb weight was increased by considering the correlation between onion growth indicators. The researchers implemented a statistical prediction, and the results demonstrated performance with less than a 10% prediction error. Kang [5] used a spatial panel model to analyze the impact of climate change on garlic and onion yields. Meteorological factors significantly impact crops, especially with strong spatial dependencies. Moon and Jung [6] improved the prediction of yields by using onion actual measurement data and meteorological factors. These previous studies predicted yields per unit area via regression analysis using growth surveys and meteorological information.

Recently, deep learning-based prediction models have been applied in various fields. Studies have examined the impact of sentiment indices based on deep learning in the agricultural market. Soon and Kim [7] used an artificial-neural-network (ANN) language model to collect real-time African swine fever news, create a sentiment index, and analyze the impact of African swine fever on meat prices via an autoregressive lag distribution model (ARDL). Cho et al. [8] derived an onion sentiment index from articles related to onions before sowing. By incorporating the onion sentiment index into an onion supply model, they analyzed the impact of changes in the onion sentiment index on the onion market. These studies indexed existing news through sentiment analysis using neural networks. Therefore, production for the current year can be predicted by applying ANNs to measurement data. Recently, Ali et al. [9] claimed that smart farming techniques enhance agricultural productivity and decision-making while improving yields and reducing costs in sustainable crop production systems. Kaur et al. [10] focus on the importance of crop yield prediction in agriculture and its role in promoting sustainable agricultural practices. Remote sensing offers an efficient and cost-effective solution for measuring key agricultural parameters, such as soil moisture levels. However, retrieving soil moisture content from coarse-resolution datasets, particularly microwave datasets, is challenging. The development of data sourcing and management is a crucial step toward implementing sustainability and precision in agriculture. Technological advancements enable data collection at a low cost [11]. These technological advancements contribute to more efficient and sustainable agricultural practices, benefiting both farmers and policymakers.

The current study predicts the final production of garlic and onions by inputting the actual measurement data into the Neural Prophet Autoregressive Exogenous-Variable (NP-ARX) model. The NP-ARX model enhances predictive power in various sectors. Based on three years of measurement data from the KREI, the Neural Prophet's lag model was used to predict the bulb weight of harvested crops. Through this study, it is expected that the prediction accuracy of the final yield of garlic and onions using measurement data can be improved and applied to existing economic models. In addition, by utilizing our results, we aim to contribute to supply stability by preemptively predicting production, thereby increasing the welfare of both farmers and consumers.

The rest of this paper is organized as follows. The analysis model and data are described in Sections 2 and 3, respectively. The analysis is divided into explanations of the "Neural Prophet Time-Series Prediction Model" and "Application of Garlic and Onion Bulb Weight Prediction." The estimation results are evaluated in Section 4. Finally, a summary of the results and future improvements are discussed in Section 5.

## 2. Model

### 2.1. Main Features in Neural Prophet

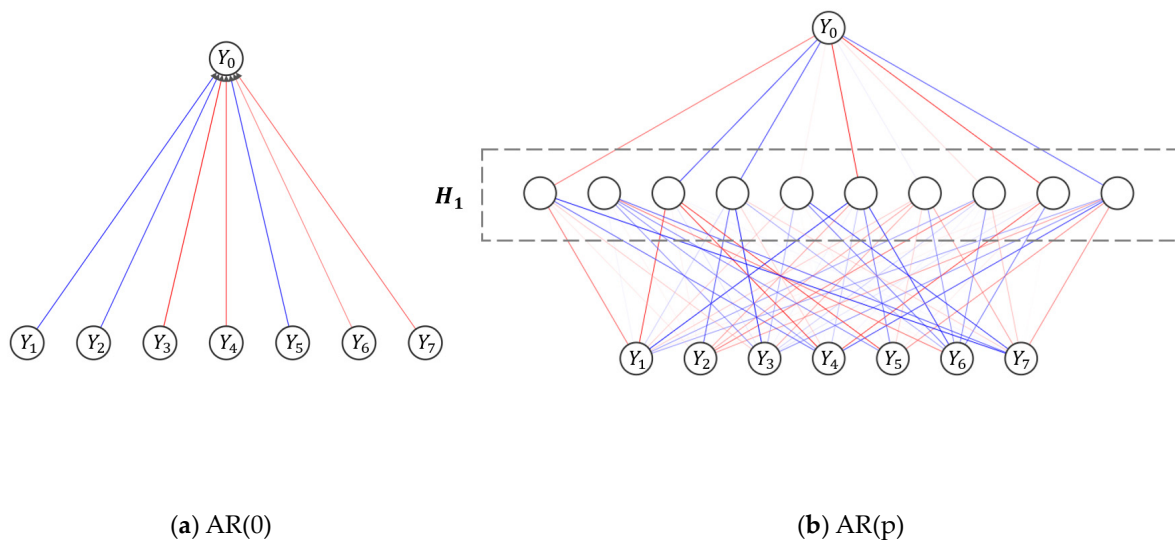
Neural Prophet is an extension of the Prophet package, a time-series forecasting package released by Facebook in 2017 [12], which not only applies a deep learning structure but also enhances performance using an autoregressive network (AR-Net) that combines feedforward ANNs with traditional autoregressive time-series models [13]. In Equation (1), the first three independent variables (trend (T), seasonality (S), and events (E)) are included

in the basic Prophet model, while the Neural Prophet adds the latter three components (regressive (F), autoregressive (A), and lag (L)). The AR-Net is applied to interpret the lag effects of independent variables, which is considered suitable for effectively handling crop measurement data [13]. The six components in the Neural Prophet can be excluded depending on the research design, and trend and seasonal components are assigned by default. Considering the characteristics of crop growth, the occurrence of unforeseen events or the impact of holidays on crop growth was insignificant. Otherwise, Fourier analysis in the Neural Prophet, which decomposes the frequency of these components and helps improve pure forecasting performance, was used to internally analyze trend, seasonality, and event components [12].

$$\hat{Y} = T(t) + S(t) + E(t) + F(t) + A(t) + L(t) \quad (1)$$

In the given equation,  $T(t)$  represents trend effects,  $S(t)$  represents seasonal effects,  $E(t)$  represents event and holiday effects,  $F(t)$  represents regression effects,  $A(t)$  represents autoregression effects for the future, and  $L(t)$  represents lag effects.

Generally, as the AR process becomes higher-dimensional, the model's predictive power improves, but the interpretability decreases. Moreover, lag variables are often correlated with themselves and other independent variables. The Neural Prophet in Figure 1b eases those issues by the AR-Net inserting hidden layers between the target variable at the present time ( $p = 0$ ) and the lagged dependent variables ( $p = 1 \sim 7$ ) [14]. In particular, the weights determined with the hidden layers decide whether to strengthen or weaken certain connections, and the optimal weight values are derived through iterations of the training process.



**Figure 1.** Artificial-neural-network autoregressive structure. Source: Created by author. (a) AR(0): Structure without hidden layers. (b) AR(p): Structure with hidden layers (H).

The AR-Net structure renders the lag factors of independent variables ( $L(t)$ ) in the prediction. The AR-Net comprehensively considers the set lag (lag  $p$ ) and generates additional independent variables for each  $p$ , enabling predictions for  $h$  steps ahead. For the five independent variables, seven lag variables were created, and weights ( $w_p$ ) of their lag factors were multiplied. The lag-reflected independent variables are likely to be correlated, and the Neural Prophet is designed to separately construct the covariates (covariates,  $x$ ) as lag factors ( $L(t)$ ) and influence the prediction. As demonstrated in Equation (2), the lag factor becomes the sum of the influences given by the covariates of each lag up to  $p$  [13].

Indeed, this advantage is not exclusive to the NP model; many deep learning architectures share this benefit.

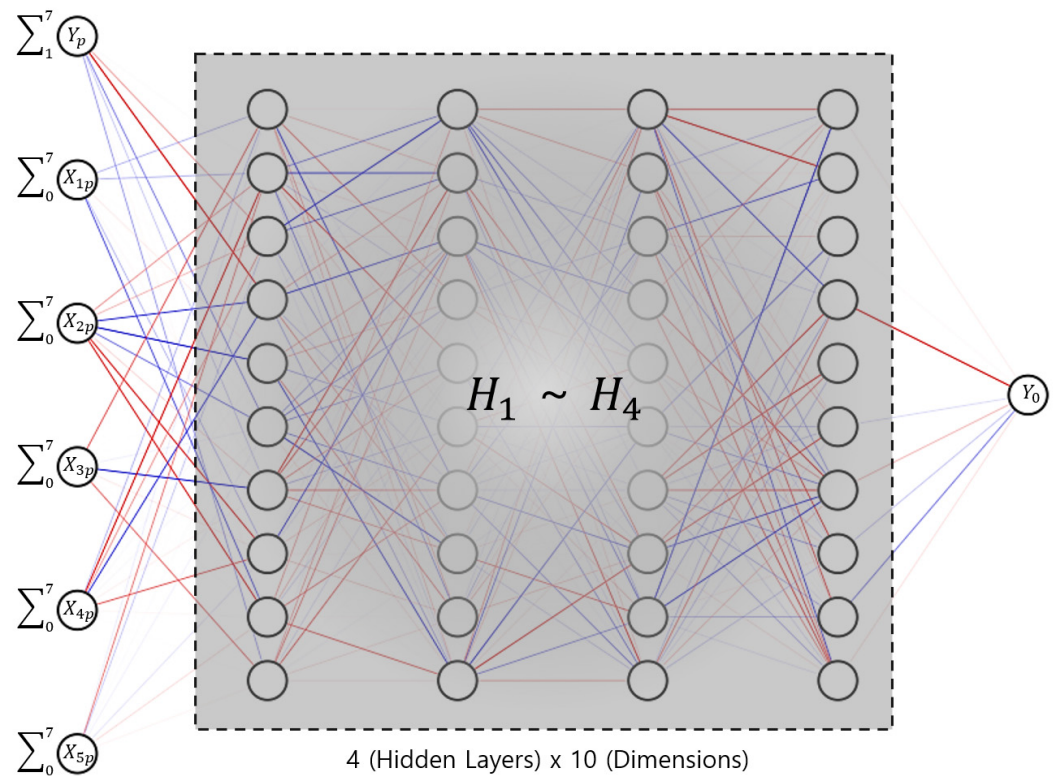
$$L(t) = \sum_{x \in X} L_x(x_{t-1}, x_{t-2}, \dots, x_{t-p}), X \in \mathbb{R}^{T \times n_l} \quad (2)$$

## 2.2. Neural Prophet Autoregressive Exogenous-Variable (NP-ARX) Model

The model used to predict the bulb weights of garlic and onions is presented in Equation (3). To predict the bulb weight ( $\hat{Y}$ ) two months later, six independent variables are set: autoregressive term ( $Y$ ), plant height ( $X_1$ ), leaf count ( $X_2$ ), leaf sheath length ( $X_3$ ), stem diameter ( $X_4$ ), and leaf tip dying length ( $X_5$ ). Each independent variable reflects a lag of up to seven periods (7-day window). Accordingly, each of the five independent variables generates seven additional independent variables, resulting in a total of 47 independent variables. For instance, the independent variables on April 15th ( $X$ ) are used to predict the bulb weight ( $\hat{Y}$ ) on June 15th, two months later.

$$\hat{Y} = \sum_{t=1}^7 Y(t) + \sum_{t=0}^7 X_1(t) + \sum_{t=0}^7 X_2(t) + \sum_{t=0}^7 X_3(t) + \sum_{t=0}^7 X_4(t) + \sum_{t=0}^7 X_5(t) \quad (3)$$

The ANN structure applied to predict the bulb weight ( $\hat{Y}$ ) is illustrated in Figure 2. The measured data were used after SOFT1 normalization, and the loss function was set to minimize the mean absolute error (MAE). The “SOFT1” normalization method adjusts the minimum value to 0.1 and the 95th percentile to 1.0, which is effective for data with outliers. This method was applied because the observed bulb weight of onions significantly exceeded the average. The loss function refers to the criterion used to calculate the difference between the actual and predicted values. It represents the degree of error for the model’s training data, and the goal is to identify values that minimize the loss function as much as possible. The optimizer used was AdamW [15], which is the default setting of the Neural Prophet, and a total of four hidden layers were constructed, each with 10 dimensions, to predict the bulb weight. The point where the loss function is minimized and the method of finding this optimal point are called the optimizer. The optimizer is an algorithm that adjusts the model’s parameters iteratively to minimize the loss function and improve the model’s performance. Adam, a type of optimizer, is a combination of the first letters of Adaptive GD and Momentum, two different optimizers. AdamW optimizer refers to the addition of a weight decay method to the original Adam optimizer. In the context of gradient descent, weight decay reduces the size of weights by a certain ratio when updating them, which helps prevent overfitting. Additionally, the training was conducted with the following settings: Batch size = 64, Epochs = 100, Learning rate = 0.1, Activation Function = ReLU, and random seed = 456. One epoch refers to the entire training dataset passing through the ANN once. Batch size is a unit that indicates how many parts the dataset is divided into for one epoch. In this study, 1 epoch is composed of 64 batches, and the training was conducted 100 times. Learning rate, also known as the learning step, refers to the amount by which the learning progresses to reach the optimal point. Activation functions are responsible for transforming input values non-linearly, yielding output values. The input layer contains six independent variables, and the additional seven lagged variables generated are not presented in Figure 2. The independent variables input is calculated in the hidden layers, which is often referred to as the “black box” area, where interpretation is not easy. Recently, research has been actively conducted to address the black box problem using eXplainable AI (XAI) to improve model interpretability [16].



**Figure 2.** Neural Prophet lag model structure. Note: (1) In the input layer, the lag variables for each independent variable are not shown in the schematic diagram, and they are denoted by the shorthand notation sigma ( $\sum X_{ip}$ ). (2)  $Y_0 = \hat{Y}$  refers to the predicted value 2 months later. Source: Created by the author.

### 3. Data

We used the actual measurement data in the Agricultural Economic Research Institute's OASIS system (<https://oasis.krei.re.kr/>; Accessed on 09 December 2022). We collected growth survey data for garlic and onions in Jeollado from 2020 to 2022. The garlic cultivar used for the analysis was a warm-region species (Namdo, Dae-seo, Republic of Korea), and the onion cultivar was a middle-aged species. The garlic varieties are classified into warm-region and cold-region, and this study analyzes the main variety, the warm-region species. The onion varieties are classified into early-maturing and middle-aged, and this study analyzes the main variety, the middle-aged species. The application of methodology on each variety will be executed in further research. This study's area was limited to Jeollado (Jeollanam-do and Jeollabuk-do, Republic of Korea). The Jeolla Province is located in the southeastern part of South Korea and is the main agricultural production region of garlic and onions. Although the original data are differentiated up to the city and county levels, we calculated the average values to create average growth data for Jeollado. A visual explanation of the measured part of crops is presented in Figure 3, and the descriptive statistics of the actual measurement data are reported in Table 1. Examination of the daily actual measurement data revealed garlic growth had a maximum bulb weight of 120 g, plant height of 89 cm, leaf count of 9, leaf sheath length of 42 cm, stem diameter of 22 mm, and leaf tip dying length of 21 cm. For onions, the growth reached a maximum bulb weight of 357 g, plant height of 91 cm, leaf count of 9, leaf sheath length of 22 cm, stem diameter of 26 mm, and leaf tip dying length of 13 cm.

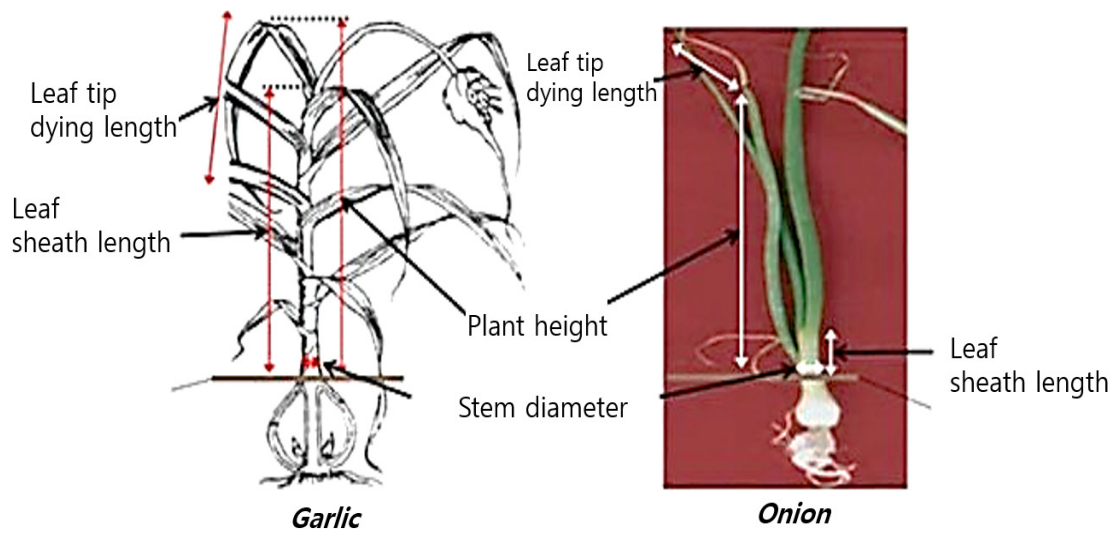


Figure 3. Visual representation of measurement data. Source: Farmers News Korea (<https://www.nongupin.co.kr/news/articleView.html?idxno=90052>; Accessed on 25 December 2022).

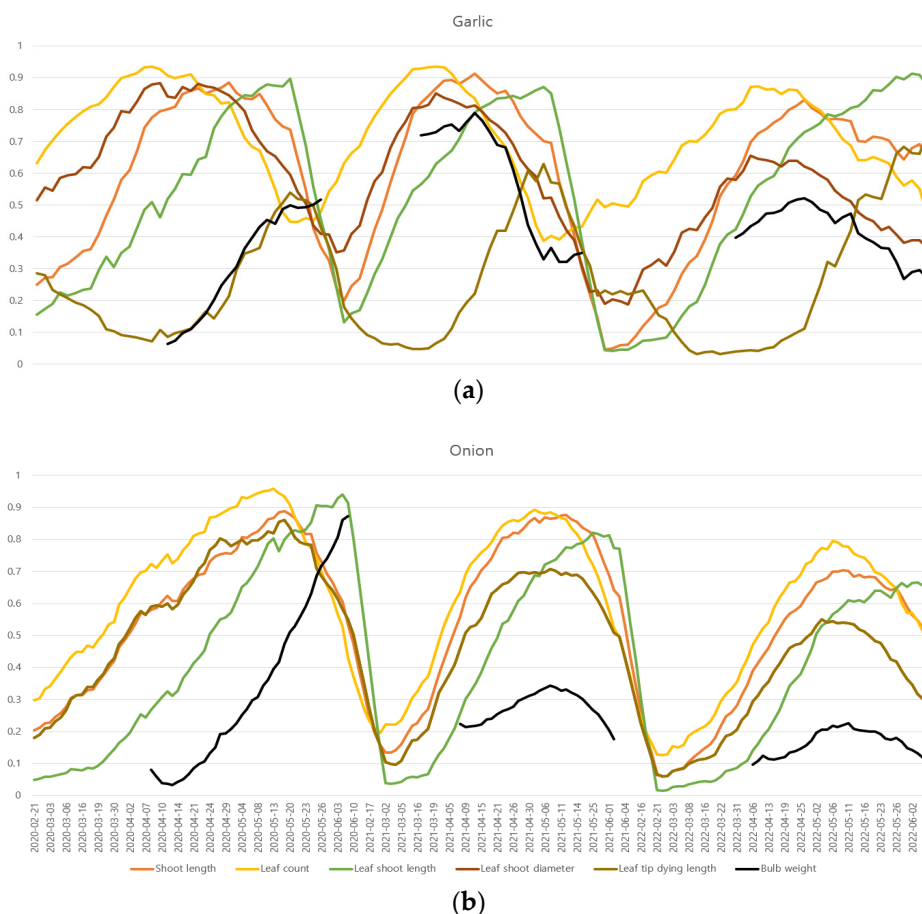
Table 1. Descriptive statistics of garlic and onion actual measurement data (2020–2022).

Stat.	Plant Height (cm)	Leaf Count (leaves)	Leaf Sheath Length (cm)	Stem Diameter (mm)	Leaf Tip Dying Length (cm)	Bulb Weight (g)	
Garlic	Min.	20.62	2.64	2.46	7.29	18.73	
	Max.	88.75	8.90	41.81	21.66	120.22	
	Mean	59.42	7.07	22.46	15.67	63.69	
	Median	66.58	7.26	24.61	15.73	64.07	
	SD	19.85	1.18	12.67	3.11	22.25	
	Variance	393.88	1.40	160.51	9.70	495.04	
	Skewness	−0.46	−0.79	−0.15	−0.27	1.21	0.18
	Kurtosis	−1.21	0.72	−1.52	−0.26	0.79	−0.34
Onion	Min.	9.20	2.76	0.99	4.03	10.10	
	Max.	91.43	9.15	21.59	25.56	356.94	
	Mean	54.85	6.65	9.87	14.89	129.89	
	Median	58.84	6.99	9.85	16.09	110.70	
	SD	22.24	1.67	6.78	5.37	83.30	
	Variance	494.80	2.80	46.01	28.87	6939.22	
	Skewness	−0.39	−0.40	0.14	−0.40	2.28	1.10
	Kurtosis	−1.12	−1.03	−1.51	−0.92	4.55	0.54

Source: Prepared by the author (originally sourced by the Korea Rural Economic Institute).

The distribution of the data used in this study is represented in Figure 4. Bulb weight reaches its maximum around May, and the trend changes similarly to that of the five independent variables. During the growth period, a total of 10 surveys were conducted, and missing values between the first and tenth surveys were resolved using linear interpolation to organize the dataset by date. For example, if the first and second surveys are on 14th and 17th of April, respectively, the data for the 15th and 16th are connected linearly between the 14th and 17th to input data for the dates without measured data. When we used the measured data for model training, owing to differences in measurement units between variables, the independent variables were inputted as normalized values through the SOFT1 method. However, considering the difficulty in updating the measured data of onion bulb weight in June 2020, which recorded 357 g, significantly exceeding the overall average of 130 g, outliers were not handled separately and included the data in the training. As a result, it is judged that the error of the estimated value ( $\hat{Y}$ ) of the daily production

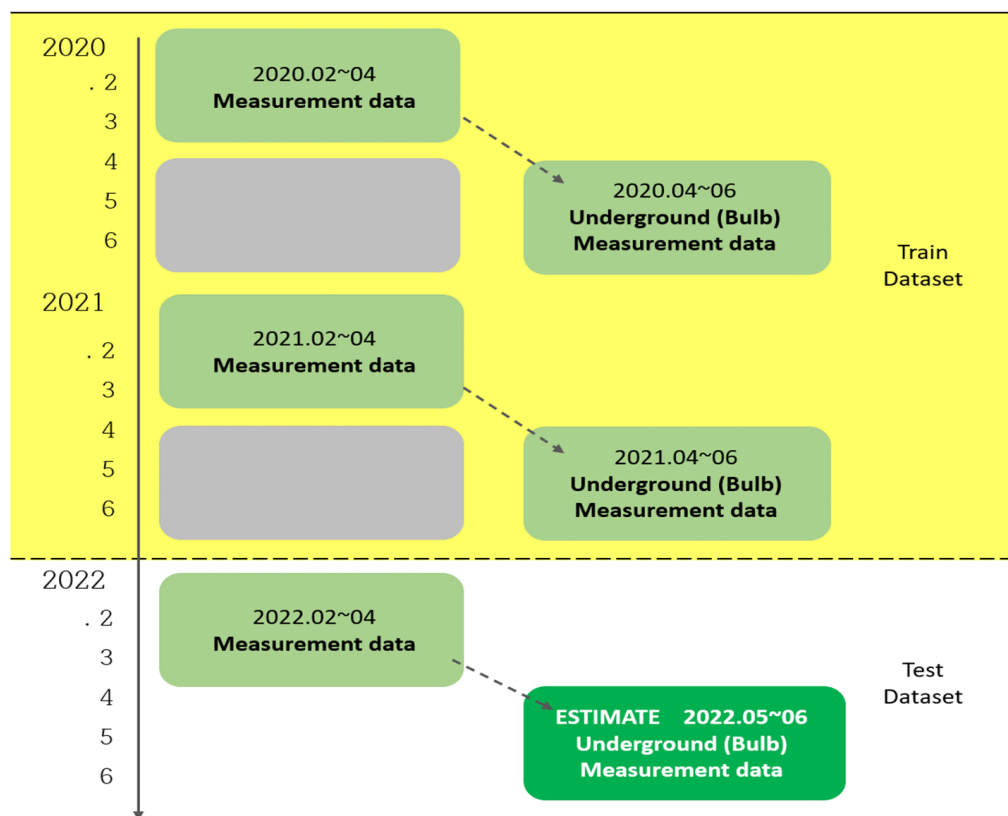
prediction results displays a negative value, as the standard deviation of the onion is higher than that of the garlic, and the normal measured information is underestimated.



**Figure 4.** Daily production in average (2020~2022). Note: (1) Raw data were normalized using the min.–max. method. The measurement data show between 0 and 1 (y-axis). (2) Normalized data are applied to the moving-average (7-day window) method to show the intuitive relationship between dependent (Bulb weight) and independent variables. Source: Created by the author. (a) Garlic. (b) Onion.

This study predicts the bulb weight between April and June using the five independent variables measured between February and April. The data learning structure is presented in Figure 4, where the aboveground measured data from February to April for each year are used to predict the underground (bulb) measured data from April to June. For example, the aboveground measured data recorded on February 14th are used to estimate the underground measured data on April 14th. As a result, the completed lag model can predict the bulb weight two months later.

Currently, the agricultural observation statistical system shares measurement data only for three years. This study was designed to train the Neural Prophet lag model with the data from 2020 and 2021 (training data and then predict the bulb weight from April to June in 2022 (test dataset) using the aboveground observation data from February to April 2022. During the training, the learning rate was set to 10% of the training dataset, which means that the model was trained on 10% of the dataset while using the remaining 90% to evaluate its performance. Here, the aboveground observation data also existed in May and June but were removed (gray box in Figure 5) to align with the goal of pre-estimating the bulb weight before proceeding with the training. Therefore, the model predicts the bulb weight from April to June using only the aboveground measured data from February to April, without knowing the aboveground measured data from May to June.



**Figure 5.** Training and testing data structure. Source: Created by the author.

#### 4. Performance Evaluation

The performance of the deep learning-based regression model was evaluated using the MAE and root-mean-squared-error (RMSE) metrics. The MAE represents the average of the absolute differences between the actual and predicted values, while the RMSE is the square root of the mean squared error (MSE) and is less sensitive to outliers compared to the MSE. (Refer to Equation (4) for details.) Both metrics are considered to have higher predictive accuracy when their values are closer to 0. The MAE is robust to outliers among the three metrics (MAE, RMSE, and MSE) and assigns equal weights to errors. The model was designed to minimize the MAE, and since garlic and onions exhibit annual periodicity, the model takes the yearly seasonality into account. Additionally, the SmoothL1Loss performance metric was considered, which combines the advantages of both L1 and L2 functions. SmoothL1Loss is a loss function that represents the error between the actual and predicted values in absolute terms. It aims to address the limitations of the traditional L1 loss and L2 loss functions, which are either non-differentiable or sensitive to outliers. SmoothL1Loss primarily uses the L1 loss function but transitions to the L2 loss function as the error value decreases. By combining the benefits of both L1 and L2 loss functions, SmoothL1Loss offers a more robust and smooth approach to minimizing errors during the model training process. This allows the model to achieve better performance in handling small errors and outliers, ultimately leading to more accurate predictions. When the error value is large, the gradient decreases steadily, and when it is less than 1, the gradient decreases smoothly, making it easier to determine the optimal value.

The results of the deep learning-based regression model for predicting the bulb weights of garlic and onions indicate promising performance, as indicated by the low values of the MAE and RMSE metrics. This demonstrates that the model can make accurate predic-



tions using the aboveground observation data, which can help farmers and agricultural professionals make better-informed decisions regarding crop management and planning.

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum |Y - \hat{Y}| \\ \text{RMSE} &= \sqrt{\frac{\sum (Y - \hat{Y})^2}{n}} \\ \text{smooth L1Loss} &= \begin{cases} 0.5(Y - \hat{Y})^2 & \text{if } |Y - \hat{Y}| < 1 \\ |Y - \hat{Y}| - 0.5 & \text{otherwise,} \end{cases} \end{aligned} \quad (4)$$

The performance metrics for garlic and onion are presented in Table 2. The NP-ARX model, which incorporates the aboveground observation data for May and June, attained an MAE of 3.3 g and 4.7 g and RMSE of 5.4 g and 8.7 g for the garlic lag and the onion models, respectively. The learning performance of the former was higher than that of the latter, and both models exhibited an average daily error of less than 5 g (3–6% error rate) based on the MAE.

**Table 2.** Estimation performance.

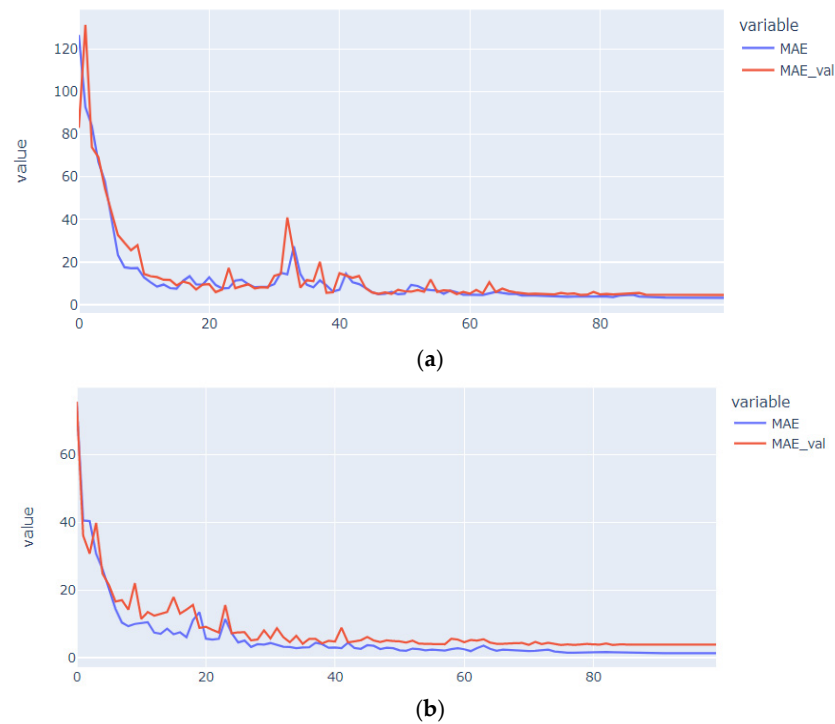
Garlic		Onion	
MAE (Train)	1.41 g	MAE (Train)	3.30 g
MAE (Test)	3.34 g	MAE (Test)	4.66 g
Error Rate to Average	5.24%	Error Rate to Average	3.59%
RMSE (Train)	3.60 g	RMSE (Train)	10.16 g
RMSE (Test)	5.36 g	RMSE (Test)	8.68 g
Error Rate to Average	8.42%	Error Rate to Average	6.68%
SmoothL1Loss	0.001	SmoothL1Loss	0.003

Note: (1) The training data error results are the prediction results of the 2020 and 2021 training data, and test data error results are the prediction results for 2022 based on the training data. (2) The average bulb weights of garlic and onions are 63.69 g and 129.89 g, respectively. Source: Prepared by the author.

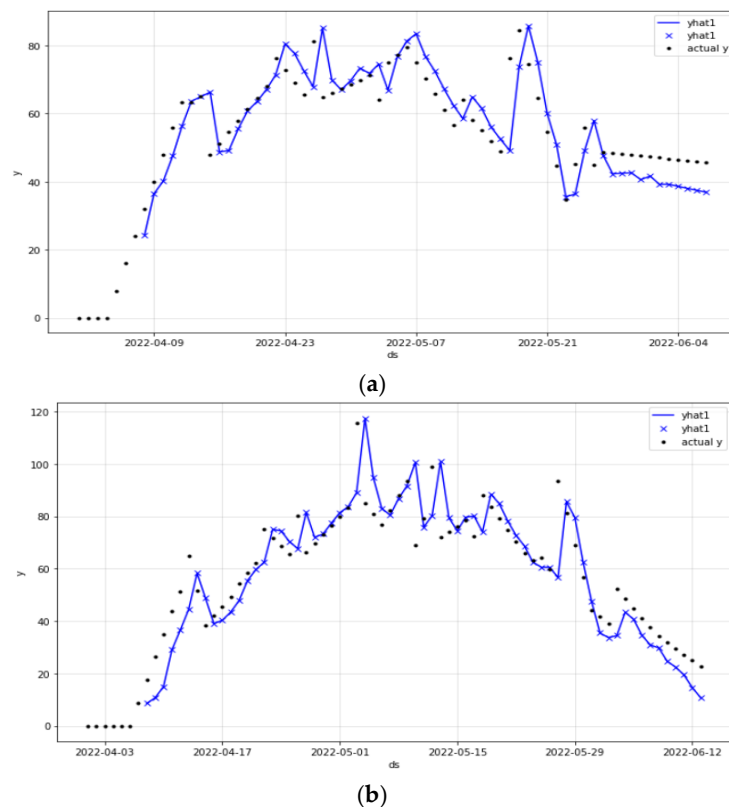
This study aimed to predict daily bulb weight, and the models are considered to have high applicability due to the highly directed nature of the measurement data used for prediction. Furthermore, they exhibited an error rate within 6% of the average bulb weight, further supporting their potential practical use in the field. However, the difference between the training and test data errors can be attributed to the relatively short observation period used for training the dataset (only two years: 2020 and 2021). This limitation can be improved in the future as more observation data is accumulated.

The MAE graph, which is the learning evaluation metric for the training and validation data, is presented in Figure 6. The index represents the epoch, and the MAE value decreases as the model goes through 100 iterative learnings. Both garlic and onion present stable MAE values after epoch = 80, and the ‘MAE\_val’ (validation value) is somewhat higher but still stable. This graph is visualized to check for overfitting; when overfitting occurs, the validation value displays an upward or unstable trend, meaning the model is overtrained on the training dataset and does not work properly on the validation dataset. In the current study, there is no overfitting, and it also demonstrates fitting results with the validation data.

The predicted average bulb weights of garlic and onions for April to June 2022, presenting both actual values (actual  $y$ ) and estimated values ( $\hat{y}$ ), are visualized in Figure 7. The actual values reveal a pattern of starting the first bulb weight observations in early April, reaching a peak in May, and then decreasing in June. The predicted and actual values exhibit a similar trend, and the daily predictions can derive the bulb weight within a certain margin of error (Table 3). The predicted values for garlic tend to be higher than the actual values, while those for onions tend to be lower.



**Figure 6.** Learning rate of weight estimation. Source: Prepared by the author. (a) Garlic. (b) Onion.



**Figure 7.** Visualization of estimated results. Note: (1) Although there are actual bulb weight measurements of garlic and onions in February in the raw data, the model’s training is focused on the main harvest season from April to June, so the data from February are excluded. (2) The actual values are presented as continuous points in some cases because daily measurement data do not always exist. For data for missing days, linear interpolation is applied to insert the expected values. Source: Created by the author. (a) Garlic. (b) Onion.

**Table 3.** Daily average estimation of bulb weight (2022.05).

Observation Data	Garlic			Observation Data	Onion		
	Act. (y)	Approx. (yhat)	Gap (g)		Act. (y)	Approx. (yhat)	Gap (g)
Mean	60.5	58.8	1.7	Mean	80.2	79.0	−1.2
Std Error	14.1	13.3	9.7	Std Error	13.8	16.3	17.7
2 May 2022	71.2	71.8	−0.6	2 May 2022	83.3	83.7	−0.4
3 May 2022	64.1	74.6	−10.5	3 May 2022	115.6	89.1	26.6
4 May 2022	75.1	66.9	8.2	4 May 2022	85.1	117.5	−32.4
6 May 2022	79.6	81.3	−1.8	6 May 2022	76.8	83.0	−6.2
11 May 2022	56.6	62.4	−5.8	9 May 2022	93.6	91.6	1.9
12 May 2022	64.1	58.6	5.5	10 May 2022	69.0	100.6	−31.6
13 May 2022	58.1	64.9	−6.8	11 May 2022	79.3	75.9	3.4
16 May 2022	49.0	52.5	−3.6	12 May 2022	99.0	80.3	18.7
17 May 2022	76.3	49.2	27.1	13 May 2022	72.1	101.0	−28.9
18 May 2022	84.4	73.7	10.7	16 May 2022	78.4	79.7	−1.3
23 May 2022	34.7	35.6	−0.9	17 May 2022	72.3	80.1	−7.9
24 May 2022	45.1	36.4	8.8	18 May 2022	87.9	74.0	13.9
25 May 2022	56.0	49.1	6.9	23 May 2022	66.0	68.7	−2.7
26 May 2022	44.8	58.0	−13.1	24 May 2022	63.1	62.5	0.6
27 May 2022	48.7	47.7	1.1	25 May 2022	64.3	60.6	3.7
				26 May 2022	59.7	60.5	−0.8
				27 May 2022	93.7	56.7	37.0
				31 May 2022	44.3	47.5	−3.2

Source: Created by the author.

The underestimation of onion predictions is related to the existence of outliers that greatly exceed the average actual value. This may also be associated with the underrepresentation of the data during the normalization process following the min.–max. method. This aspect is expected to be adjusted in the future as raw data is updated.

The daily bulb weight prediction results for garlic and onions in May, which is the main production period and the month when the bulb weight reaches its maximum, are reported in Table 3. On average, prediction errors of 1.7 g and 1.2 g were observed for garlic and onions, respectively. The standard deviation of the predicted values for garlic is 13.3 g, which is lower than that of the actual values, which is 14.1 g. This confirms that the model was applied more stably for garlic than for onions, which have a higher standard deviation (16.3 g). The Neural Prophet lag model effectively used autoregressive (AR) variables and lagged independent variables through high-dimensional deep learning structures. Despite using relatively high-frequency data for prediction, the model quickly adapted to the shocks of the latest actual information and made predictions accordingly. For example, on May 17th, 2022, the actual datum of 76.3 g for garlic was collected, and the model generated a prediction of 49.2 g, an error of 27.1 g. Subsequently, the model quickly adjusted to an error of 10.7 g on May 18th and 0.9 g on May 23rd when the next actual information was collected. The significance of these prediction results is that they were achieved using only growth measurement data, even though no other growth information (such as weather data or growth indices) was included as independent variables. Measurement data is a direct indicator for predicting crop yield, and its inclusion in forecasting models could potentially enhance their accuracy by incorporating previously unconsidered data. The results suggest that improving the precision and frequency of updates to the forecasting model could lead to better predictions of crop yield.

## 5. Summary and Conclusions

Garlic and onions are key vegetables used for seasoning in Korea, consumed stably throughout the year. It is necessary to establish supply policies that stabilize the market through production forecasts. This study predicts the bulb weight of harvested crops using garlic and onion growth-stage measurement data and contributes to minimizing fiscal

expenditures for related policy authorities while increasing the welfare of farmers and consumers through proactive preparation of supply.

Since 2020, the Korea Rural Economic Institute (KREI) has been conducting field surveys of crops in each province. The measurement data of garlic and onions from the Jeolla Province were used in the current study. The observed bulb weight of garlic was stably measured with an average of 19–120 g over three years, while that of onions exhibited large fluctuations with an average of 10–357 g over three years. In particular, the bulb weight of onions in 2020 exceeded the average of 130 g by more than three times, requiring a meticulous review of the observed values. Additionally, the measurement data was randomly conducted per field, which is considered suitable for predicting the average total production.

The lag model was used with the Neural Prophet instead of statistical analysis for 356 measurement data over 3 years. While various economic forecasts using deep learning have been applied, research using crop measurement data in agricultural economics is in its early stages. This study trained a multilayer ANN and demonstrated performance using error performance indicators between measured and predicted bulb weights. The results revealed an error of about 5% of the average bulb weight when predicting garlic and onions in 2022, suggesting the possibility of predicting underground measurements not yet observed through aboveground measurements two months earlier.

By enabling the customization of model options, the Neural Prophet facilitates the resolution of prevailing challenges in time-series modeling, namely multicollinearity, heteroscedasticity, and autocorrelation. The Neural Prophet ARX model demonstrated high accuracy in predicting the bulb weight, even with a relatively short measurement period and training data of 2 years, by controlling correlated independent variables and their lagged variables.

This study explores ANN-based time-series prediction using measurement data. To evaluate its practicality, comparisons with existing statistical methods are needed. The enhancement tasks, such as handling missing values and optimizing data processing, are unorganized. Integrating updated data and additional crop production information may refine the deep learning model. Notably, the NP-ARX is user-friendly for agricultural economists without extensive computer science expertise. While the NeuralProphet package provides an automated AR-Net structure, fine-tuning the model for better performance can be challenging. Therefore, researchers may find it advantageous to construct their own neural-network time-series models to achieve improved results.

**Author Contributions:** Theoretical framework establishment and estimation, W.K.; Conceptualization and resources, B.M.S.; Review and editing—final draft, W.K. and B.M.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2022M3J6A1084843).

**Data Availability Statement:** Raw data are available in “Outlook & Agricultural Statistics Information System (OASIS)” at <https://oasis.krei.re.kr/gis/observationResultSearch/selectGrowthList.do> (accessed on 25 December 2022), operated by the Korea Rural Economic Institute (KREI). Additionally, the Neural Prophet repository is available at <https://github.com/wonseongkim/NP-ARX.git> (accessed on 5 January 2023).

**Acknowledgments:** This study is a revised paper that the award-winning research by the Minister of Agriculture, Food and Rural Affairs and the Korea Rural Economic Institute.

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

## References

1. USDA-NASS. The Yield Forecasting Program of NASS. 2012. Available online: [https://www.nass.usda.gov/Education\\_and\\_Outreach/Understanding\\_Statistics/Yield\\_Forecasting\\_Program.pdf](https://www.nass.usda.gov/Education_and_Outreach/Understanding_Statistics/Yield_Forecasting_Program.pdf) (accessed on 5 April 2022).
2. Schnepf, R. NASS and US Crop Production Forecasts: Methods and Issues. Congressional Research Service (CRS) Report. 2017. Available online: <https://sgp.fas.org/crs/misc/R44814.pdf> (accessed on 5 April 2022).

3. Bernardi, M.; Deline, J.; Durand, W.; Zhang, N. Crop Yield Forecasting: Methodological and Institutional Aspects. Rome: FAO.33. 2016. Available online: [https://www.researchgate.net/publication/297234539\\_Crop\\_Yield\\_Forecasting\\_Methodological\\_and\\_Institutional\\_Aspects](https://www.researchgate.net/publication/297234539_Crop_Yield_Forecasting_Methodological_and_Institutional_Aspects) (accessed on 5 April 2022).
4. Kim, J.S.; Kim, S.U. Forecasting Bulb Weight of Median-Late Variety Onion Using Growth Indicators and Meteorological Factors. *J. Rural. Dev.* **2021**, *44*, 1–26. [[CrossRef](#)]
5. Kang, H.J. Analysis of the Impact of Climate Change on the Production of Garlic and Onions in Major Production Areas. 1st Agricultural Products Supply and Demand Forecast Model Contest. 2021. Available online: <https://www.mafra.go.kr/bbs/mafra/71/327878/artclView.do> (accessed on 5 April 2022).
6. Moon, J.H.; Jeong, C.E. Prediction of Onion Production Stages and Estimation of Social Welfare Effects Considering Measured Growth Information and Meteorological Information. 1st Agricultural Products Supply and Demand Forecast Model Contest. 2021. Available online: <https://www.mafra.go.kr/bbs/mafra/71/327878/artclView.do> (accessed on 5 April 2022).
7. Soon, B.M.; Kim, W.S. Impacts of the ASF Sentiment Changes on Meat Prices. *Korean J. Agric. Econ.* **2022**, *63*, 43–62. Available online: <https://kiss.kstudy.com/Detail/Ar?key=3940614> (accessed on 11 January 2023). [[CrossRef](#)]
8. Cho, S.M.; Oh, J.E.; Baek, J.H.; Soon, B.M. Modeling the Onion Market Using Sentiment Analysis with Deep Learning. *Korean J. Agric. Econ.* **2022**, *63*, 33–54. Available online: <https://kiss.kstudy.com/Detail/Ar?key=3973701> (accessed on 11 January 2023).
9. Ali, A.; Hussain, T.; Tantashutikun, N.; Hussain, N.; Cocetta, G. Application of Smart Techniques, Internet of Things and Data Mining for Resource Use Efficient and Sustainable Crop Production. *Agriculture* **2023**, *13*, 397. [[CrossRef](#)]
10. Kaur, R.; Tiwari, R.K.; Maini, R.; Singh, S. A Framework for Crop Yield Estimation and Change Detection Using Image Fusion of Microwave and Optical Satellite Dataset. *Quaternary* **2023**, *6*, 28. [[CrossRef](#)]
11. Hachimi, C.E.; Belaqziz, S.; Khabba, S.; Sebbar, B.; Dhiba, D.; Chehbouni, A. Smart Weather Data Management Based on Artificial Intelligence and Big Data Analytics for Precision Agriculture. *Agriculture* **2022**, *13*, 95. [[CrossRef](#)]
12. Taylor, S.J.; Letham, B. Forecasting at scale. *Am. Stat.* **2018**, *72*, 37–45. [[CrossRef](#)]
13. Triebe, O.; Hewamalage, H.; Pilyugina, P.; Laptev, N.; Bergmeir, C.; Rajagopal, R. Neuralprophet: Explainable forecasting at scale. *arXiv* **2021**, arXiv:2111.15397. [[CrossRef](#)]
14. Triebe, O.; Laptev, N.; Rajagopal, R. Ar-net: A simple auto-regressive neural network for time-series. *arXiv* **2019**, arXiv:1911.12436. [[CrossRef](#)]
15. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101. [[CrossRef](#)]
16. Das, A.; Rad, P. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv* **2020**, arXiv:2006.11371. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.