Smart Media Journal / Vol.12, No.11 / ISSN:2287-1322
https://dx.doi.org/10.30693/SMJ.2023.12.11.134

Forecasting Fish Import Using Deep Learning: A Comprehensive Analysis of Two Different Fish Varieties in South Korea

Abhishek Chaudhary , Sunoh Choi

Abstract

Nowadays, Deep Learning (DL) technology is being used in several government departments. South Korea imports a lot of seafood. If the demand for fishery products is not accurately predicted, then there will be a shortage of fishery products and the price of the fishery product may rise sharply. So, South Korea's Ministry of Ocean and Fisheries is attempting to accurately predict seafood imports using deep learning. This paper introduces the solution for the fish import prediction in South Korea using the Long Short-Term Memory (LSTM) method. It was found that there was a huge gap between the sum of consumption and export against the sum of production especially in the case of two species that are Hairtail and Pollock. An import prediction is suggested in this research to fill the gap with some advanced Deep Learning methods. This research focuses on import prediction using Machine Learning (ML) and Deep Learning methods to predict the import amount more precisely. For the prediction, two Deep Learning methods were chosen which are Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM). Moreover, the Machine Learning method was also selected for the comparison between the DL and ML. Root Mean Square Error (RMSE) was selected for the error measurement which shows the difference between the predicted and actual values. The results obtained were compared with the average RMSE scores and in terms of percentage. It was found that the LSTM has the lowest RMSE score which showed the prediction with higher accuracy. Meanwhile, ML's RMSE score was higher which shows lower accuracy in prediction. Moreover, Google Trend Search data was used as a new feature to find its impact on prediction outcomes. It was found that it had a positive impact on results as the RMSE values were lowered, increasing the accuracy of the prediction.

Keywords : Machine Learning| Deep Learning| Artificial Neural Network| Long Short-Term Memory| Prediction

I. INTRODUCTION

South Korea is one of the high-trading countries with a high number of both imports and exports. The imports and exports depend upon market supplies. Seafood is also one of the major imports of South Korea. Various species of fish are imported and exported throughout the year. Since then, it has imported various fish continuously but in the past few years data shows that there has been a disbalance between consumption and production. In other words, the market was not able to supply enough number of fish products to the customers because the amount of seafood was not accurately predicted. The data shows that there was a noticeable gap between the sum of consumption and export and the sum of

* This research was supported through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NO. RS-2023-00237159)
** Member, Dept of Software Engineering, Jeonbuk National University

Manuscript : 2023.10.20
Confirmation of Publication: 2023.11.16
Corresponding Author : Sunoh Choi, e-mail : suno7@jbnu.ac.kr

production as shown in Fig 1 and Fig 2.

The erratic nature of market demand is one theory as to why there is such a discrepancy. Import selections can be strongly affected by changes in consumer tastes, seasonal variances, and unforeseeable events. Imports that are not in line with actual demand can result in shortages or overstocking, both of which disturb the market's balance. Inadequate import processes can also have a long-lasting impact on future demand, leading to a reduction in market supply.

The data show that consumption was higher than the production. This problem can be solved if the amount of seafood can be accurately predicted by using some advanced techniques like Artificial Neural Network and Long Short-Term Memory [1]. These are part of Artificial Intelligence which are the most adopted technology in most of the sectors in the current age. The use of such technology has had a great positive impact. In this research, LSTM for the fish import amount prediction is proposed as the solution.

This research goes beyond the standard by combining Google Trend [2] search data into the forecasting algorithm to produce predictions for fish import that are more correct. By using external data sources to supplement the prediction process, this novel approach creates a system that is more reliable and flexible. Data on Google's trending searches is useful for understanding the public's interest in and search patterns for a range of topics, including seafood consumption. The prediction model gets a wider view on the variables affecting fish imports by incorporating this external data. The model can now incorporate subtleties that it might have otherwise missed because of its increased feature collection, which boosts forecast accuracy.

The LSTM method is presented as the ultimate solution for the prediction as the result obtained during the research shows that the LSTM method was predicting more accurately than other methods applied which was shown by Root Mean Square Error (RMSE) as a performance metric. Moreover, the prediction was improved after using the Google trend search data. The average RMSE score of Hairtail was dropped from 179.5 to 145.5 and of Pollock it was dropped from 10494.5 to 7727. This increase in accuracy was due to the increase in the number of features which were used in the prediction.

The contribution of the paper are as follows. First, to the best of our knowledge, this is the first paper to propose a method for predicting seafood import volume using South Korea's seafood import volume data [3]. Second, we showed that LSTM for predicting seafood import gave the highest accuracy. Third, we showed that the use of Google Trend data for predicting seafood import was effective.

The paper is further explained in this way, In Section II, we introduced the Related work, which is related with similar research fields and has used similar methods. In Section III. fish import prediction methods were introduced where fish import data was analyzed and import prediction methods were also introduced along with the proposed model. In Section IV, experimental results were presented. There are two result sections.

First, with the existing data and second, with the Google trend search data. Lastly, in Section V, we gave the Conclusion.

II. RELATED WORK

In this section, the related works are listed. In a paper by Kexian Zhang and Min Hong [1] proposed a method to forecast crude oil price and concluded with three outcomes. First, the LSTM model has strong generalization and stable applicability for forecasting in different time scales. In the first case, they also found that the fitting effect of LSTM was a little weaker than ANN model. Second, the LSTM has higher prediction accuracy in timescale prediction, but the result also showed that forecasting accuracy and forecasting stability of LSTM was slightly worse than ANN model. And third, the LSTM has lower accuracy as the time increases which requires other factors in model.

A study by Hamid Reza Niazkar and Majid Niazkar [4] gave the result that fourteen ANN-based models were able to predict the COVID-19 outbreak. They found that the implementation of the incubation period of COVID-19 in the ANN-based prediction model led to more accurate prediction. The limitation of this research was the lack of proper data where all the details were not available.

Research conducted for solar radiation validation by Zahraa E. Mohamed [5] said that the ANN-based model is an efficient method with higher precision which lower RMSE score. The first algorithm with Basic Backpropagation (Bp) was outperformed by the second algorithm with Bp with learning rate and momentum coefficient. The statistical errors RMSE, MAPE, MABE, r and R2 were used to compare the models.

Utsav Poudel [6] in an article concluded that forecasting model using LSTM adjusted in complex time series and made more accurate prediction. Monthly airline passenger data for the year 1960 was used for twelve months to make predictions. The RMSE value obtained was 30.5.

Jingyi Shen and M. Omair Shafiq [7] build a comprehensive feature engineering model (LSTM) which outperformed the often-used Machine Learning models. Chinese stock market data of two years were collected, the system was redesigned in such a way that it works in feature engineering procedure for prediction.

In research by R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande and P. Singh [8] concluded that, in a comparison between machine and deep for the prediction of heart disease, deep learning obtained 94.2% accuracy. The dataset used was not large enough so they added the dataset size can increase the accuracy with the model being optimized.

Joerg Evermann, Jana-Rebecca Rehse and Peter Fettke [9] explained the deep learning for predicting the next event in business process where the result surpass the state-of-art in prediction precision more than 80% on many problems. Deep learning and Recurrent Neural Network methods were chosen as they both are novel and explicit process models.

III. FISH IMPORT PREDICTION

METHODS

1. *Fish Data Analysis*

1) Fish Data without Google Trend Data

The supply amount depends upon the market demand, which shows the need for that product. But the data [3] shows that the market demand or consumption was much higher than productions amount. At the same time, there was export which decreased the supply in the market. Among the numbers of species of fish, Hairtail and Pollock species were selected.

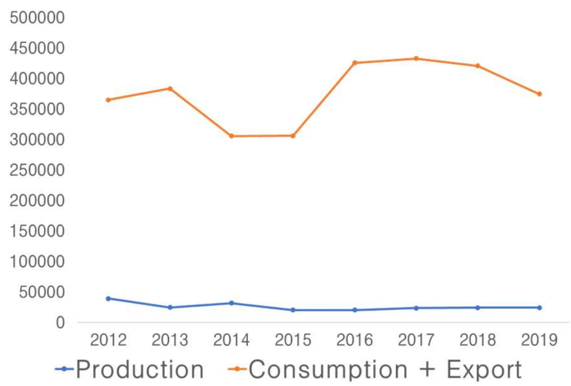


Fig 1. Production Vs Consumption + Export of Hairtail

The difference between productions and sum of consumption and export are illustrated using the following Fig 1 for Hairtail and Fig 2 for Pollock.

In Fig 1, the enormous difference between the production and sum of consumption and export can be seen for the Hairtail fish. Such huge differences can create a negative impact in the market regarding the product. The fluctuation can be seen in Fig 1 as there was a huge drop in production affecting consumption and export. Even though production was increased in 2017, it could not meet the requirement of the market supply.

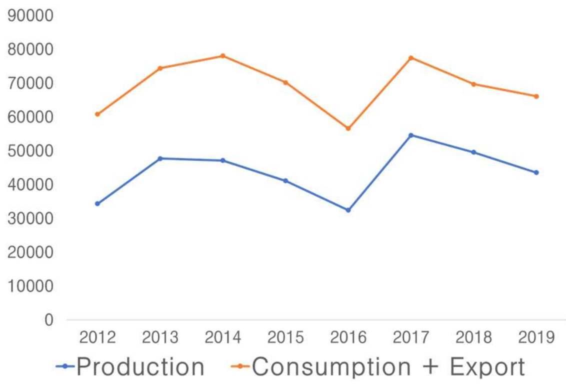


Fig 2. Production Vs Consumption +Export of Pollock

In Fig 2, the production was much lower than the sum of consumption and export in terms of Pollock fish, it means the supply for the Pollock was not meeting the market requirement but there was much higher demand. Such a dramatic difference may create an increase in the price of the product due to the low amount of supply.

The solution to decrease the gap and maintain the balance between the supply and demand can be import. Importing cannot be done randomly but it needs to be planned. Fig 1 and 2 illustrate the gap and give the numbers to be imported which can be done with the help of some advanced technique which will help to predict more accurately to keep the balance between the consumption, export and production including the import.

2) Fish Data with Google Trend Data

Fig 3 shows the google search of Hairtail yearly data from 2012 to 2019 in South Korea which shows the search increase and reach highest in 2014 but then it shows decrease in search number. This search can be considered as the interest and the popularity of the product in the market. This data can be considered as a key factor in defining the import amount.

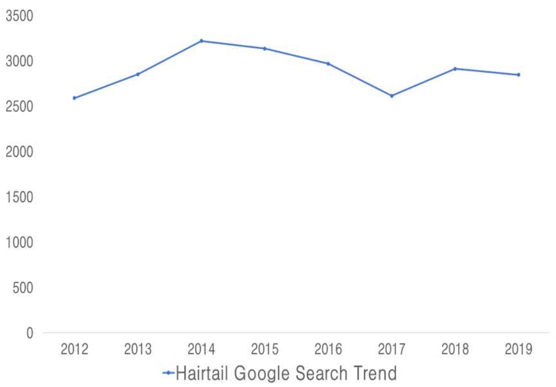


Fig 3. Hairtail Google Search Trend

Similarly, in Fig 4, the search data of Pollock fish also shows the increase in number. Even though it has some fluctuation, it keeps growing. Imports without such knowledge may affect the market and price of the product on the market.

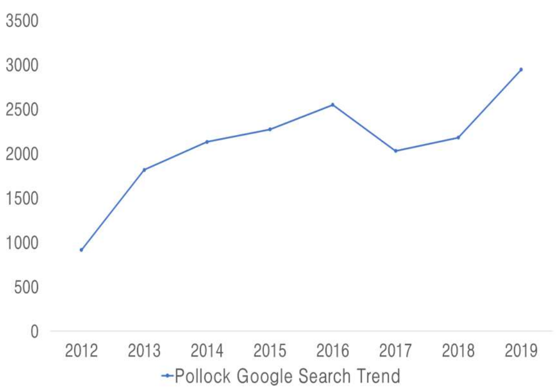


Fig 4. Pollock Google Trend Search

2. *Fish Import Prediction Methods Using Machine Learning and Deep Learning*

1) Machine Learning (ML)

It is a branch of artificial intelligence that deals with building models and algorithms that let computers learn from data. Machines employ data patterns rather than being expressly designed to get better over time. By examining and learning from examples, ML systems can be trained to make predictions, classify items, or solve challenging problems [10].

2) Artificial Neural Network (ANN)

ANN are computational models that are based on the composition and operation of biological neural networks in the human brain. ANNs are made up of interconnected "neurons," or nodes, arranged in layers. Every neuron process input and sends its results to the layer below [11].

3) Long Short-Term Memory (LSTM)

It is a recurrent neural network (RNN) architecture type that was created to deal with data sequences. The inability of conventional RNNs to recognize distant dependencies in sequences is addressed by LSTMs. For tasks involving sequential data, such as time series prediction, speech recognition, and natural language processing, LSTMs are successful [12].

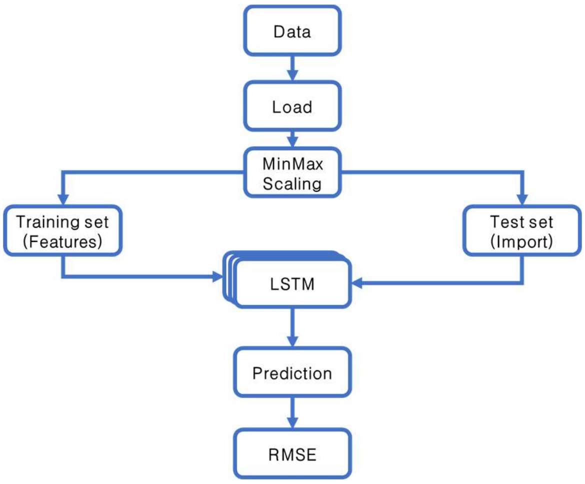


Fig 5: Proposed Model

As shown in Fig 5, the first step was to load the data which further goes through the MinMax scaling. After that, the data was divided into training and test set. All the features from the dataset except import were set as training set and the import was only used as test set. After that, LSTM was applied to predict the result. The output was compared with actual value along with accuracy and RMSE value to compare the efficiency of the method.

IV. EXPERIMENTAL RESULTS

1. *Setup*

The machine used for the setup was Lenovo i7 with NVIDIA GeForce RTX 3060 Laptop GPU. A virtual environment was created with TensorFlow 2.10.0, sk-learn 1.2.2, python 3.9.17 and Keras 2.10.0.

2. *Performance Metric*

1) Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) is used to measure the accuracy of predictive models. It calculates the average size of the discrepancies between actual observed values and those that were projected. RMSE is a crucial measure for examining the predictive accuracy across many domains since it can be used to compare the performance of various models or decide how a model has improved over time [13]. The equation for RMSE is shown below:

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

where, n is the number of data points, $\hat{y_i}$ is the i-th measurement and y_i is the corresponding prediction.

2) Accuracy

The percentage is one of the most used metrics to measure the success rate. It also helps to understand the error rate and the efficiency and quality of the outcome. It is simple and easy to understand [14].

Here is the formula used for the calculation:

$$Accuracy = \frac{Predicted}{Actual} \times 100\ \%$$

Where Predicted is the predicted value, Actual is the actual value.

3. *Results without Google Trend Search Data*

Google Trend Search data was considered as a new feature which was used for the prediction to find its impact on the results.

Since there are two different fish data from 2012 to 2019, the training and test set were divided as follows:

Table 1. Test and Training set

Steps	Years	Training	Testing
Step 1	2012-2016	2012-2015	2016
Step 2	2013-2017	2013-2016	2017
Step 3	2014-2018	2014-2017	2018
Step 4	2016-2019	2015-2018	2019

The training and testing were done in four steps for each fish individually applying the three selected methods to find which method will work effectively. Further processes are explained below:

1) Hairtail

The results obtained after applying the methods are illustrated in the given Table 2. Further, the accuracy was also calculated in individual prediction to find which method predicted more accurately. And lastly, average accuracy was calculated to find which method has predicted better on average.

The results obtained are compared in terms of accuracy. It was found that the average accuracy of LSTM was 99.37% while ML predicted 116.11% which is 16.11% higher than actual prediction. Similarly, on average ANN's accuracy was 111.92% which is lower than ML but still 11.92% higher than actual values to be predicted.

Table 2. Hairtail Prediction and accuracy results

Year	Actual value	ML	ANN	LSTM
2016	28731	28643 **(99.69%)**	29737 (103.5%)	28015 (97.51%)
2017	23432	26037 (111.12%)	24374 (104.02%)	23431 **(99.99%)**
2018	17722	23346 (131.73%)	22131 (124.88%)	17722 **(100%)**
2019	15659	19091 (121.92%)	18051 (111.92%)	15658 **(99.99%)**
Average		116.11%	111.92%	**99.37%**

The RMSE values of the methods applied are listed in Table 3.

Table 3. Hairtail RMSE of ML, ANN, and LSTM

Years	ML	ANN	LSTM
2016	88	1006	716
2017	2605	942	1
2018	5624	4409	0
2019	3432	2392	1
Average	2937.25	2187.25	**179.5**

From Table 3, we can see that the average RMSE value of LSTM is comparatively lower than that of ML and ANN. It is almost sixteen times lower than ML and almost twelve times lower than ANN.

2) Pollock

The result obtained after the experiment showed that the LSTM was outperformed by the ANN method. The average accuracy obtained by the ANN is 101.90% while LSTM obtained 97.22%. The average accuracy of ML is 104.82%. The predicted output and the average accuracies are listed in Table 4.

Table 4. Pollock Prediction and accuracy results

Year	Actual value	ML	ANN	LSTM
2016	398343	409314 (102.75%)	382785 (96.09%)	386614 (**97.06 %)**
2017	405355	399282 (98.5%)	406167 (**100.2%**)	394903 (97.42%)
018	401403	393657 (98.07%)	394532 (**98.29%)**	391201 (97.46%)
2019	314001	376681 (119.96%)	354905 (113.03%)	304406 (**96.94** %)
Avera ge		104.82%	**101.90%**	97.22%

The average RMSE score from Table 5 shows that the LSTM has outperformed other methods. The RMSE score of ML and ANN are two times and one and half times higher than LSTM, respectively. The average RMSE values of ML, ANN and LSTM are 21867.5, 16036.25 and 10494.5, respectively.

Table 5. Pollock Prediction RMSE

Years	ML	ANN	LSTM
2016	10971	15558	11729
2017	6073	812	10452
2018	7746	6871	10202
2019	62680	40904	9595
Average	21867.5	16036.25	10494.5

4. *Results with Google Trend Search Data*

The accuracy of the prediction results has been increased as the number of features has increased. Google Trend search data was added as a new feature for the prediction. It showed the positive impact on the outcome by increasing the accuracy and decreasing the RMSE values.

1) Hairtail

The outcome of the LSTM for Hairtail with and without the Google Trend Search are listed in Table 6.

Table 6. Hairtail LSTM Prediction and accuracy with and without Google data

Year	Actual Value	LSTM (Accuracy)	LSTM_GTrend (Accuracy)
2016	28731	28015 (97.51%)	28163 (98.02%)
2017	23432	23431 (99.99%)	23431 (99.99%)
2018	17722	17722 (100%)	17722 (100%)
2019	15659	15658 (99.99%)	15658 (99.99%)
Average		99.37%	99.50%

From Table 6, we can see that the predicted value for the year 2016 was improved while other values still are unchanged. Even though, there was only one improvement which was still enough to increase the accuracy and showed the effect of the increase of features. The accuracy increased from 99.37% to 99.50%.

In Table 7, the RMSE value and accuracy are listed after the calculation. It showed that the RMSE value was dropped from 179.5 to 142.5.

Table 7. LSTM RMSE with and without Google data

Years	LSTM	LSTM_GTrend
2016	716	568
2017	1	1
2018	0	0
2019	1	1
Average	179.5	**142.5**

2) Pollock

The prediction results of Pollock obtained after adding the Google Trend Search data as a new feature are listed in Table 8.

Table 8. Pollock LSTM Prediction and accuracy with and without Google Trend Search data

Year	Actual Value	LSTM (Accuracy)	LTSM_GTrend (Accuracy)
2016	398343	386614 (97.06%)	**387249 (97.21%)**
2017	405355	394903 (97.42%)	**415378 (102.47%)**
2018	401403	391201 (97.46%)	**395147 (98.44%)**
2019	314001	304406 (96.94%)	**317536 (101.13%)**
Average		97.22%	**99.81%**

The bold character in Table 8 shows the improved results obtained than the earlier results. Since, we can see that the result obtained after using Google Trend Search data has shown its positive impact as all the prediction result obtained are more accurate than without Google Trend Search data.

The RMSE and accuracy obtained for the Pollock are listed in Table 9. From Table 9, we can see that the average RMSE value have been decreased from 10494.5 to 7727. Similarly, from Table 8, the accuracy has increased from 97.22% to 99.81% after adding the feature.

Table 9. LSTM RMSE with and without Google data

Years	LSTM	LSTM_GTrend
2016	11729	11094
2017	10452	10023
2018	10202	6256
2019	9595	3535
Average	10494.5	**7727**

V. CONCLUSION

Advanced Deep Learning techniques like LSTM can help in prediction in the short or long term. The Machine Learning method was outperformed by the Deep Learning approach. Among two applied Deep learning methods, LSTM performed better with higher accuracy and lower RMSE values which showed better performance and efficiency. Moreover, the increase in the number of features such as Google Trends data can help to increase accuracy and contribute to an increase in efficiency.

REFERENCES

[1] M. H. Kexian Zhang, "Forecasting crude oil price using LSTM neural networks," *AIMS Press,* vol. 2, no. 3, 2022.

[2] Google, "Google Trends," Google , [Online]. Available: https://trends.google.com/trends/.

[3] TimaxTibero Co.,Ltd, "BigdataSea," TimaxTibero Co.,Ltd, [Online]. Available: https://www.bigdata-sea.kr/.(accessed oct., 20, 2023).

[4] M. N. Hamid Reza Niazkar, "Application of artificial neural networks to predict the COVID-19 outbreak," *Global Health Research and Policy,* 2020.

[5] Z. E. Mohamed, "Using the artificial neural networks for prediction and validating solar radiation," *Journal of the Egyptian Mathematical Society,* 2019.

[6] U. Paudel, "Time and Series Forecasting with LSTM- Recurrent Neural Network," Level Up Coding, May 2023.

[7] J. S. &. M. O. Shafiq, "Short-term stock market price trend prediction using a comprehensive deep learning system," *Journal of Big Data,* 2020.

[8] A. K. M. S. G. D. S. P. P. S. Rohit Bharti, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning," Hindawi, 2021.

[9] J.-R. R. P. F. Joerg Evermann, "Predicting process behaviour using deep learning," *ScienceDirect ,* vol. 100, pp. 129-140, 2017.

[10] S. Brown, "Machine learning, explained," MIT, 2021 April 21.

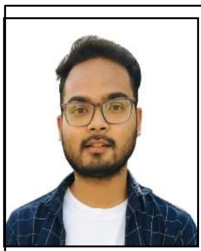
[11] S. L. Y.-S. Park, "Artificial Neural Network," *ScienceDirect,* no. 2008, pp. 237-245, 2008.

[12] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *ScienceDirect,* vol. 404, 2020.

[13] C3, "Root Mean Square Error (RMSE)," C3.ai, [Online]. Available: https://c3.ai/glossary/data-science/root-mean-square-error-rmse/.(accessed oct., 20, 2023).

[14] Analytics Vidhya, "Performance Metrics : Classification Model," Medium, [Online]. Available: https://medium.com/analytics-vidhya/performance-metrics-classification-model-69a5546b118c. (accessed oct., 20, 2023).

Authors



Abhishek Chaudhary

He received his B.S. degree in Computing from Coventry University, UK in 2022.
He is currently enrolled in Jeonbuk National University for M.S degree in Software Engineering from 2023.

Sunoh Choi

He received his B.S. and M.S. degree in Computer Science and Engineering from Korea University in 2005 and 2008. He received his Ph.D. in Electrical and Computer Engineering from Purdue University, USA in 2014. He was a senior researcher in ETRI before joining Jeonbuk National University. Since 2021, He is working as an Associate professor in Jeonbuk National University.