

## 농경지 휴경면적 예측력에 대한 모형 간 비교 연구

임청룡<sup>1</sup>, 이항아<sup>2</sup>, 최돈우<sup>2\*</sup>

<sup>1</sup>한국농어촌공사 농어촌연구원, <sup>2</sup>경상북도 농업기술원

### A Study on the Inter-Model Comparison of the Predictive Power for Idle Area of Farmland

Cheong-Ryong Lim<sup>1</sup>, Hang-Ah Lee<sup>2</sup>, Don-Woo Choi<sup>2\*</sup>

<sup>1</sup>Rural Research Institute, Korea Rural Community Corporation

<sup>2</sup>Gyeongsangbuk-do Agricultural Research & Extension Services

**요약** 이 연구에서는 지역단위 농지 휴경면적 예측모형을 개발하기 위해 토양DB, 농업총조사 원자료, 기상DB 등으로 구성된 농업빅데이터를 활용하였다. 분석모형을 구축함에 있어서 휴경면적에 영향을 미치는 변수들을 활용하여 다수의 휴경면적 예측모형을 구축 및 추정하였으며 분석과정은 총 3단계로 구분할 수 있다. 1단계는 통계분석 분야에서 광범위하게 쓰이고 있는 기초적인 기본모형에 해당하는 회귀분석, 의사결정나무, 인공신경망, 서포트벡터머신 등 기법을 활용하였고, 2단계로는 예측분야에 있어서 1단계보다 이론적으로 개선된 머신러닝기법 중 앙상블모형인 배깅, 부스팅, 랜덤 포레스트 등 모형들을 활용하였으며, 3단계에서는 자체적으로 모형 예측력을 개선할 수 있는 딥러닝기법 중 심층신경망을 활용하였다. 휴경면적에 대한 여러 예측모형 추정결과를 보다 객관적으로 비교하기 위해 분석으로부터 도출된 평균오차제곱을 기준으로 모형 간 예측력을 비교하여 최적 모형을 선택하였다. 비교 결과 2단계 모형들이 1단계 모형들에 비해 예측력이 보다 우수하고, 3단계 모형이 2단계 모형에 비해 예측력이 더 좋은 것을 확인하였다. 이러한 분석결과는 휴경면적 객관적 예측 및 휴경면적 효율적 활용에 있어서 참고자료로 활용될 것으로 여겨진다.

**Abstract** In this study, agricultural big data composed of soil databases, agricultural census raw data, and meteorological databases were utilized in constructing a prediction model for estimating regional idle farmland areas. In constructing the analysis model, several models for predicting the idle area were constructed and estimated by utilizing the variables affecting this area. The analysis process can be divided into three stages. In the first stage, basic models such as regression analysis, decision trees, artificial neural networks, and support vector machines were used. In the second stage, ensemble models such as bagging, boosting, and the random forest model were used, and in the third stage, a deep neural network was used based on deep learning techniques. The optimal model was selected by comparing the predictive power between models based on the mean square error derived using the estimation results of several models. It was confirmed that the second stage models had better predictive power than the first stage models and that the third stage model had better predictive power than the second stage models. This analysis is expected to be used as a reference in the objective prediction of idle farmland areas and the efficient use of this area.

**Keywords** : Idle Farmland, Big Data, Ensemble Model, Deep Learning, Predictive Power, MSE(Mean Square Error).

본 논문은 농촌진흥청 연구과제(과제번호: PJ015331062021)의 지원으로 수행되었음.

\*Corresponding Author : Don-Woo Choi(Gyeongsangbuk-do Agricultural Research & Extension Services)

email: gogemaroo@korea.kr

Received July 5, 2021

Accepted August 5, 2021

Revised August 2, 2021

Published August 31, 2021



|                            |                       |         |          |         |          |
|----------------------------|-----------------------|---------|----------|---------|----------|
| Drainage grade             | Very bad              | 53.24   | 144.85   | 56.75   | 156.00   |
|                            | Bad                   | 29.83   | 119.28   | 32.14   | 125.55   |
|                            | Slightly bad          | 299.23  | 372.34   | 302.26  | 376.10   |
|                            | Slightly good         | 256.17  | 245.06   | 253.86  | 245.78   |
|                            | Good                  | 431.70  | 857.77   | 431.80  | 849.49   |
| Crop type                  | Paddy rice            | 100.7   | 78.8     | 101.0   | 78.8     |
|                            | Food                  | 230.8   | 145.4    | 231.0   | 146.5    |
|                            | Wild vegetables       | 278.3   | 167.3    | 278.4   | 169.0    |
|                            | Other field crops     | 345.1   | 209.4    | 344.9   | 211.9    |
|                            | Facility vegetables   | 807.1   | 483.3    | 806.6   | 488.6    |
|                            | Other Facilities      | 1.0     | 4.0      | 0.9     | 3.9      |
|                            | Main fruit            | 7.9     | 30.0     | 8.0     | 30.3     |
|                            | Special crop          | 17.8    | 27.9     | 17.9    | 27.7     |
| Farm type                  | 1 person              | 12.5    | 13.8     | 12.6    | 13.8     |
|                            | More than 1 person    | 14.4    | 18.3     | 14.4    | 18.1     |
| Age                        | Less then 50          | 17.9    | 19.5     | 17.9    | 19.4     |
|                            | 50-59                 | 114.0   | 116.4    | 115.1   | 116.6    |
|                            | 60-69                 | 104.8   | 72.1     | 104.0   | 72.1     |
|                            | Over 70               | 407.8   | 246.8    | 408.2   | 249.2    |
| Main industry              | Agriculture           | 433.2   | 432.7    | 387.6   | 415.9    |
|                            | Forestry              | 84.4    | 119.4    | 78.7    | 113.6    |
|                            | Fishery               | 72.8    | 130.1    | 65.8    | 118.4    |
|                            | Manufacture           | 7,287.3 | 84,823.7 | 9,211.3 | 76,765.8 |
|                            | Construction industry | 24.1    | 62.3     | 19.4    | 49.5     |
|                            | Wholesale and retail  | 9.9     | 34.9     | 8.4     | 27.1     |
|                            | Lodging and food      | 81.0    | 143.6    | 48.0    | 100.5    |
|                            | Other                 | 22.5    | 42.1     | 25.3    | 37.5     |
| Cultivated area per person |                       | 1.21    | 0.65     | 1.21    | 0.65     |
| GRDP                       |                       | 2,297   | 0.003    | 2,886   | 0.003    |
| Average temperature        |                       | 12.91   | 1.11     | 12.90   | 1.11     |
| Average wind speed         |                       | 1.93    | 0.58     | 1.93    | 0.58     |
| Average humidity           |                       | 67.35   | 3.98     | 67.36   | 4.01     |

Source: Kosis.kr, Rural Development Administration

### 3. 분석모형 및 추정방법

#### 3.1 분석모형

이 연구에서는 휴경지 예측모형을 개발하기 위해방법론적 측면에서 선행 연구를 고찰하였으며, 농가단위에서 내부적 요인과 외부적 요인으로 변수를 구성하였다. 또

한 분석에 활용한 데이터의 성격에 부합하고 논의 휴경에 영향을 미칠 수 있는 변수들을 선정하 후 다음과 같은 함수형태로 설정하였다.

$$PL_i = F(land_i, agri_i, farm_i, region_i, climate_i) + \epsilon_i \quad (1)$$

식(1)에서  $PL_i$ 는  $i$ 읍·면의 논 휴경면적이고,  $land_i$ 는 토양특성 변수들에 대한 벡터이고,  $agri_i$ 는 농업특성 변수들에 대한 벡터이며,  $farm_i$ 는 농가특성 변수들에 대한 벡터이다.  $region_i$ 은 지역변수들에 대한 벡터이고,  $climate_i$ 는 기상변수들에 대한 벡터이며, 잔차항은  $\epsilon_i$ 로 설정할 수 있다.

#### 3.2 추정방법

이 연구에서 읍·면별 논 휴경면적 예측모형을 추정함에 있어서 3단계로 거쳐서 분석을 진행하였다. 1단계는 기본모형인 회귀분석, 의사결정나무, 인공신경망, 서포트 벡터머신 등을 활용하고, 2단계로는 머신러닝기법 중 앙상블모형인 배깅, 부스팅, 랜덤포레스트 등 모형들을 활용하며, 3단계에서는 딥러닝 기법 중 심층신경망을 활용하였다.

##### 3.2.1 의사결정나무(Decision Tree)

의사결정나무는 주어진 입력 값에 대하여 출력 값을 예측하는 모형으로 분류나무와 회귀나무모형이 있다. 의사결정나무는 지도학습기법으로 각 변수의 영역을 반복적으로 분할함으로써 전체 영역에서의 규칙을 생성하는 것이다[7].

##### 3.2.2 신경망(Neural Networks)

신경망은 생물학적 신경망 구조로부터 착안된 학습 알고리즘이다. 특히 입력 값과 출력 값 간의 매우 복잡한 형태의 비선형 모형화 방법이다. 신경망은 인간의 두뇌 구조를 모방한 지도학습법이며, 신경망의 기본적인 작동 원리는 여러 개의 뉴런들을 상호 연결하여 입력 값에 대한 최적의 출력 값을 예측하는 것이다. 다층신경망 구조에서는 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)으로 구성되며, 단층신경망은 입력층과 출력층만으로 구성된다[8,9].

##### 3.2.3 서포트벡터머신(Support Vector Machine)

서포트 벡터 머신(support vector machine, SVM)

은 자료 분석을 위한 지도 학습 모델이며, 주로 분류와 회귀 분석을 위해 사용한다. 서포트 벡터 머신은 분류 또는 회귀 분석에 사용 가능한 초평면(hyperplane) 또는 초평면들의 집합으로 구성되어 있다. 초평면이 가장 가까운 학습 데이터 점과 큰 차이를 가지고 있으면 분류 오차가 작기 때문에 좋은 분류를 위해서는 어떤 분류된 점에 대해서 가장 가까운 학습 데이터와 가장 먼 거리를 가지는 초평면을 찾아야 한다. 데이터가 선형 구분이 되지 않는 문제를 해결하기 위해 초기 문제의 유한 차원에서 더 높은 차원으로 대응시켜 분리를 쉽게 하는 방법이 제안되었다[10,11].

### 3.2.4 배깅(Bagging)

배깅은 불안정한 예측모형에서 불안정성을 제거함으로써 예측력을 향상시키기 위하여 개발되었으며, 학습 자료의 작은 변화에 의해 예측모형이 크게 변하는 경우 그 학습방법은 불안정하다고 할 수 있다. 불안정성은 통계학의 로버스트(robust)성질과 유사하며 배깅은 주어진 자료에 대하여 여러 개의 붓스트랩(bootstrap) 자료를 생성하고, 각 붓스트랩 자료에 예측모형을 만든 후 결합하여 최종 예측모형을 만드는 방법이다[12].

### 3.2.5 부스팅(Boosting)

부스팅의 기본 아이디어는 예측력이 약한 모형(weak learner)들을 결합하여 강한 예측모형을 만드는 것이며, 여기서 약한 예측모형이란 랜덤하게 예측하는 것보다 약간 좋은 예측력을 지닌 모형을 말하고 강한 예측모형이란 예측력이 최적에 가까운 예측모형을 가리킨다. 이 연구에서는 샘플의 가중치 수정 대신 예측기가 만든 잔여 오차에 새로운 예측기를 학습시키는 방식인 그레이디언트 부스팅(Gradient Boosting)을 활용하였다[13,14].

### 3.2.6 랜덤 포레스트(Random Forest)

랜덤 포레스트는 의사결정나무의 특징인 분산이 크다는 점을 고려한 것으로서 배깅과 부스팅보다 더 많은 무작위성을 주어 약한 학습기들을 생성한 후 이를 선형 결합하여 최종 학습기를 만든다. 랜덤 포레스트는 무작위성을 최대로 주기 위해 붓스트랩과 더불어 입력변수들에 대한 무작위 추출을 결합한다. 따라서 서로 연관성이 약한 학습기를 여러 개 만들어 내는 기법이다[15].

### 3.2.7 심층신경망(Deep Neural Network)

심층 신경망(Deep Neural Network, DNN)은 입력층(input layer)과 출력층(output layer) 사이에 여러 개의 은닉층(hidden layer)들로 이뤄진 인공신경망(Artificial Neural Network, ANN)이다. 심층 신경망은 일반적인 인공신경망과 마찬가지로 복잡한 비선형 관계(non-linear relationship)들을 모델링할 수 있다 [16,17].

## 4. 휴경면적 예측모형 분석결과 비교

### 4.1 적정 반복회수 설정

앙상블 모형 중 배깅, 부스팅 및 랜덤포레스트 등을 추정하기 위해 트리모형에 대한 반복 회수 설정이 필요하며, 이 연구에서는 충분한 크기의 반복회수를 설정한 후 평균오차제곱(Mean Square Error)이 안정화되는 상황을 보고 결정하였다. 부스팅일 경우 최대반복회수 1,000회에 가까워짐에 따라 MSE가 안정됨을 알 수 있다 (Fig. 2).

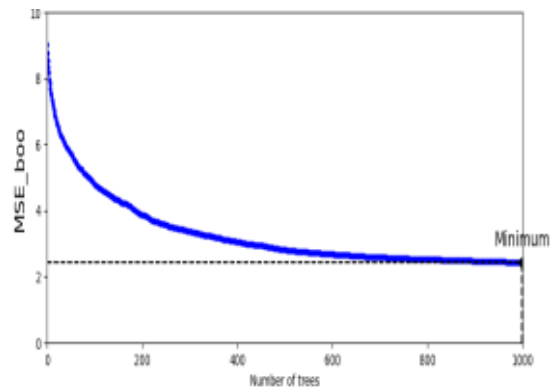


Fig. 2. MSE Change According to the Number of Trees Applied with Boosting

심층신경망 구조 설정은 입력층과 출력층을 제외한 은닉층을 4개로 구성하였으며, 은닉층 1-3까지는 200개의 뉴런을 가지고, 은닉층 4에는 100개의 뉴런으로 설정하였다(Table 2).

Table 2. Deep Neural Network Structure

|                | Number of Node | Number of Parameters |
|----------------|----------------|----------------------|
| Input layer    | 37             | 0                    |
| Hidden layer 1 | 200            | 7,600                |
| Hidden layer 2 | 200            | 40,200               |
| Hidden layer 3 | 200            | 40,200               |
| Hidden layer 4 | 100            | 20,100               |
| Output layer   | 1              | 101                  |

Total Parameters: 108,201

심층신경망 모형의 충분한 학습을 위해 적정 에포크 수가 필요하며 이 연구에서는 1,000회로 설정하였으며, 2010년 자료를 활용하여 학습하고, 2015년 자료를 활용하여 검정하였다. 분석결과 에포크 증가에 따라 학습데이터 MSE(loss)는 지속적으로 감소하지만, 검정데이터 MSE(val\_loss)는 일정 수준에 수렴하는 것을 알 수 있었다(Fig. 3).

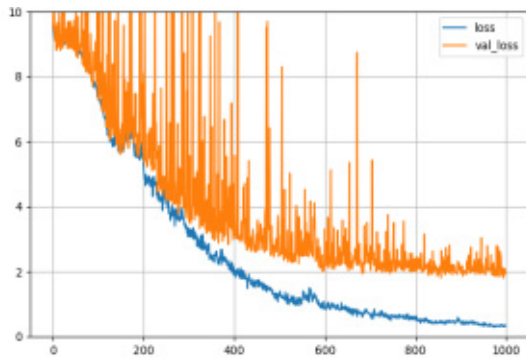


Fig. 3. MSE Change According to the Number of Epochs

#### 4.2 모형 간 예측력 비교

모든 모형은 2010년 자료를 활용하여 분석한 후, 2015년 자료를 활용하여 검정을 하였다. 실측치와 예측값에 대한 산점도를 통해 시각화하였으며, 우선 기본모형들에 대한 산점도를 살펴보면 대각선 상으로 수렴하지 않아서 예측력이 좋지 않음을 알 수 있다(Fig. 4).

양상블 모형들과 심층신경망 모형에 대한 실측값과 예측값에 대한 산점도를 보면 기본모형보다 대각선에 잘 분포되어 있으며, 그 중 심층신경망이 가장 집중적으로 나타남을 알 수 있다(Fig. 5).

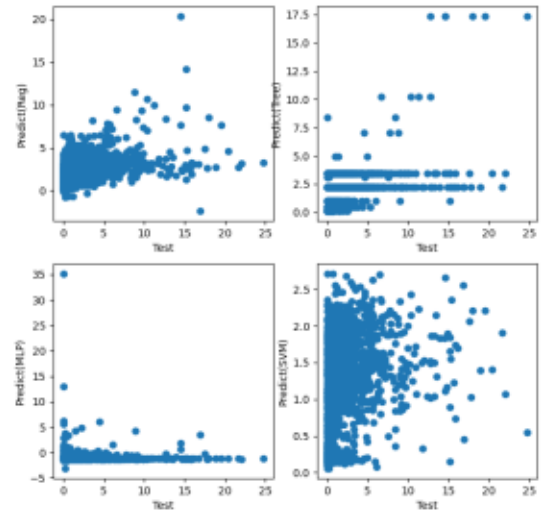


Fig. 4. Comparison of Estimation Results for Basic Models

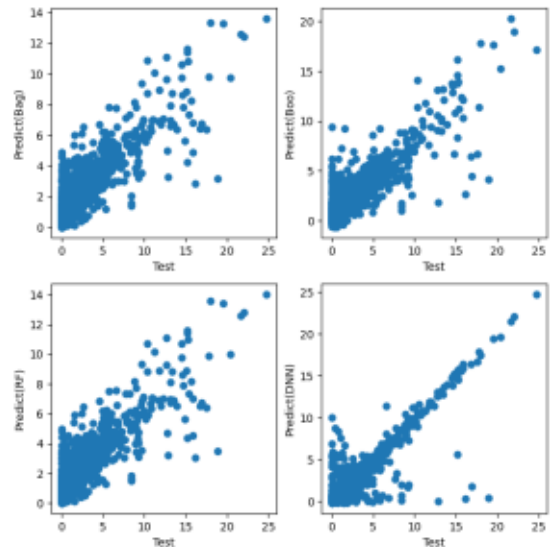


Fig. 5. Comparison of Ensemble Model and Deep Learning Estimation Results

모형들의 예측력을 통계량 기준으로 살펴보면, 전체모형 중 심층신경망 모형의 MSE가 가장 작게 나타나 예측력이 가장 우수한 것을 알 수 있다. 또한 예측값과 실측값 평균 비율이나 변이계수비율을 살펴보아도 심층신경망이 가장 우수한 것을 알 수 있다(Table 3).

Table 3. Comparison of Prediction Results for Estimation Models

|                |               | MSE   | Mean  | C.V.  | Mean ratio | C.V. ratio |
|----------------|---------------|-------|-------|-------|------------|------------|
| Idle farmland  |               | -     | 2.10  | 0.69  | -          | -          |
| Basic model    | Regression    | 7.63  | 2.28  | 1.54  | 108.4      | 48.51      |
|                | Decision Tree | 7.33  | 2.07  | 1.40  | 98.60      | 48.43      |
|                | ANN           | 20.52 | -1.06 | -0.96 | -50.3      | 36.20      |
|                | SVM           | 9.51  | 1.15  | 1.76  | 54.63      | 21.36      |
| Ensemble Model | Bagging       | 3.40  | 2.21  | 1.24  | 105.1      | 58.48      |
|                | boosting      | 2.44  | 2.17  | 0.91  | 103.1      | 77.93      |
|                | Random Forest | 3.35  | 2.20  | 1.23  | 104.9      | 58.83      |
| Deep Learning  | DNN           | 1.81  | 2.05  | 0.70  | 97.73      | 95.83      |

2015년 자료를 활용한 읍면단위 예측값과 실측값 중 랜덤으로 4개 추출하여 살펴볼 경우 심층신경망이 예측력이 가장 우수함을 알 수 있다(Table 4).

Table 4. Prediction Result of Idle Farmland in Paddy Fields

|                |               | Region 1 | Region 2 | Region 3 | Region 4 |
|----------------|---------------|----------|----------|----------|----------|
| Idle farmland  |               | 0.00     | 0.76     | 2.70     | 2.43     |
| Basic model    | Regression    | 2.66     | 2.63     | 2.26     | 2.77     |
|                | Decision Tree | 2.19     | 2.19     | 2.19     | 2.19     |
|                | ANN           | -1.01    | -1.18    | -1.18    | -1.18    |
|                | SVM           | 1.05     | 1.60     | 1.30     | 0.19     |
| Ensemble Model | Bagging       | 1.13     | 1.37     | 2.31     | 2.21     |
|                | boosting      | 0.75     | 1.39     | 2.83     | 1.79     |
|                | Random Forest | 1.13     | 1.39     | 2.28     | 2.09     |
| Deep Learning  | DNN           | 0.31     | 0.95     | 3.59     | 2.54     |

## 5. 요약 및 결론

이 연구에서는 토양 DB, 농업총조사 원자료 등으로 구성된 농업빅데이터를 활용하여 회귀분석, 의사결정나무, 인공신경망, 배깅, 부스팅, 랜덤포레스트, 심층신경망 등의 모형을 활용하여 읍·면단위 휴경면적 예측모형을 선정하였으며, 분석결과를 요약하면 다음과 같다.

첫째, 딥러닝모형의 예측력이 앙상블모형보다 우수하였으며, 앙상블모형이 기본모형들보다 예측력이 우수한 것을 알 수 있었다

둘째, 실측값과 예측값에 대한 평균과 변이계수 비교를 통해 심층신경망 추정결과가 실측값에 가장 근접할 뿐만 아니라 변동성도 작아서 안정적인 것을 알 수 있다.

셋째, 앙상블모형은 부스팅의 예측력이 가장 우수한 것으로 나타났다.

이러한 분석결과로부터 다음과 같은 결론 및 시사점을 제시하고자 한다.

첫째, 종속변수에 대한 변수들의 영향력이 아닌 종속변수 예측결과가 분석목표일 경우 딥러닝 모형이 훨씬 효과적임을 알 수 있다.

둘째, 지역 내 휴경지의 정확한 예측으로 휴경면적을 과학적으로 예측하여 지역단위 농지이용계획을 효율적으로 수립하는데 기여할 수 있다.

셋째, 예측된 휴경지를 효율적으로 활용한다면 농지이용률이 제고되고, 지역경제 활성화에도 기여할 수 있을 것이다.

이 연구에서 심층신경망분석이 전통적인 모형들에 비해 예측력이 가장 우수함을 확인하였으나, 기타 딥러닝모형 간 비교분석은 이루어지지 않아 추후 딥러닝모형 간 비교 연구가 필요할 것이다.

다만 이 연구는 최근 자료인 2020년 농업총조사 원자료를 활용하지 못한 한계점을 있다.

## References

- [1] I. H. Park, G. S. Jang, D. J. Seo, "An Analysis of the Fallow Potential in Agricultural Area by Multi-logistic Model -A Case Study of Ibang-myeon, Changnyeong-gun, Kyungsangnam-do-", Journal of Environmental Impact Assessment, vol. 15, no. 1, pp. 53-65, 2006.  
UCI : G704-000713.2006.15.1.004
- [2] H. L. Xie, L. J. Cheng, T. G. Lv, "Factors Influencing Farmer Willingness to Fallow Winter Wheat and Ecological Compensation Standards in a Groundwater Funnel Area in Hengshui", Sustainability, vol. 9, pp. 1-18, 2017.
- [3] T. E. Mauldin, A. J. Plantinga, R. J. Alig., "Determinant of Land Use in Maine with Projections to 2050", NJAF, vol. 16, no. 2, pp. 82-88, 1999.
- [4] S. E. Ahn, A. J. Plantinga, R. J. Alig., "Historical Trends and Projections of Land Use for the South-Central United States", U.S. Department of Agriculture Forest Service, Pacific Northwest Research Station, 2001.
- [5] W. S. Jang, N. S. Jung, K. W. Park, "Development of a Paddy Field Estimation Model Considering Shift-share Effects", Journal of the Korean Society of

Agricultural Engineers, vol. 50, no. 3, pp. 83-89, 2008.  
 UCI : G704-000114.2008.50.3.011

- [6] D. W. Choi, D. C. Kim, H. A. Lee, C. R. Lim, "A Study on the Efficient Utilization of Idle and Deserted Land -Focused on Oriental Melon of SeongJu and Peach of GyeongSan-", Journal of Agriculture & Life Science, vol. 53, no. 4, pp. 137-144, 2019.  
 DOI : <https://doi.org/10.14397/jals.2019.53.4.137>
- [7] L. Breiman, J. Fredman, R. Olshen, C. Stone, Classification and regression trees, Wadsworth, 1984.
- [8] T. Hastie, R. Tibshirani, J. Friedman. The element of statistical learning: data mining, inference, and prediction(2nd edition). Springer New York. 2009.
- [9] A. J. Izenman, Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. Springer New York, 2008.
- [10] C. Cortes, V. Vapnik, "Support-vector networks". Machine Learning, vol. 20, no. 3, pp 273. 1995.  
 DOI: <https://doi.org/10.1007/BF00994018>
- [11] V. Vapnik, Support Vector Machines. The nature of statistical learning theory. New York: Springer-Verlag New York, 2000.
- [12] L. Breiman, "Bagging predictors", Machine Learning, vol. 26, pp. 123-140, 1996.
- [13] L. Breiman, Arcing The Edge, Technical Report 486. Statistics Department, University of California, Berkeley, 1997.
- [14] J. H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine, Springer New York, 1999.
- [15] L. Breiman, "Random Forests". Machine Learning, vol. 45, pp. 5-32, 2001.
- [16] Y. Bengio, A. Courville, P. Vincent, "Representation Learning: A Review and New Perspectives," IEEE Trans. PAMI, special issue Learning Deep Architectures, 2013.
- [17] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview", Neural Networks, vol. 61, pp. 85-117, 2015.  
 DOI: <https://doi.org/10.1016/j.neunet.2014.09.003>

---

임 청 룡(Cheong-Ryong Lim)

[정회원]



- 2006년 2월 : 경북대학교 대학원 농업경제학과 (경제학석사)
- 2008년 2월 : 경북대학교 대학원 농업경제학과 (경제학박사)
- 2014년 9월 : 연변대학교 농림경제학과 전임강사
- 2017년 1월 : 한국농어촌공사 농어촌연구원 주임연구원

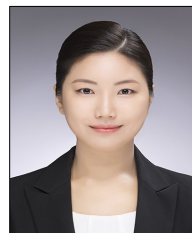
<관심분야>

농업경영, 환경경제, 농산물마케팅, 빅데이터

---

이 항 아(Hang-Ah Lee)

[정회원]



- 2017년 2월 : 경북대학교 대학원 농업경제학과(경제학석사)
- 2017년 7월 : 경상북도농업기술원

<관심분야>

농업경영, 패널분석, 소득분석

---

최 돈 우(Don-Woo Choi)

[정회원]



- 1996년 2월 : 경상북도농업기술원
- 1998년 8월 : 경북대학교 대학원 농업경제학과(경제학석사)
- 2014년 2월 : 경북대학교 대학원 농업경제학과(경제학박사)

<관심분야>

농업경영, 농업회계, 빅데이터