

# Deep learning-based modeling of second-hand ship prices in South Korea

Changro Lee<sup>1</sup>, Keyho Park<sup>2</sup>

<sup>1</sup>Department of Real Estate, Kangwon National University, Chuncheon, South Korea

<sup>2</sup>Department of Geography, Seoul National University, Seoul, South Korea

## Article Info

### Article history:

Received Sep 15, 2021

Revised Mar 8, 2022

Accepted Apr 6, 2022

### Keywords:

Categorical data

Entity embedding

Neural network

Ship valuation

Structured data

## ABSTRACT

Accurate ship valuation can encourage transparency and reliability in the shipping industry. In this age driven by artificial intelligence; however, deep learning approaches have not yet taken root in ship valuation. Despite the significant achievements of deep learning algorithms in the field of unstructured data such as computer vision, the same cannot be said for the structured data-dominant areas, including the shipping industry. Neural networks (NNs), the most common algorithms for implementing deep learning, are known not to have a relative advantage in handling structured data, particularly in processing categorical data. The inefficiency of NNs for processing categorical data significantly degrades their performance when categorical data occupy a significant portion of a dataset. In this study, we employed a NN to estimate second-hand ship prices. Its architecture was specified using entity embedding layers to enhance the performance of the network when categorical variables were highly cardinal. Experimental results demonstrated that the information contained in categorical data can be efficiently extracted and fed into a NN using the entity embedding technique, thereby improving the prediction accuracy for ship valuation. The network architecture specified in this study can be applied in wider valuation areas where categorical data are prevalent.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Changro Lee

Department of Real Estate, Kangwon National University

1 Kangwondaehak-gil, Chuncheon-si, Gangwon-do, 24341, South Korea

Email: spatialstat@naver.com

## 1. INTRODUCTION

Ship prices are notorious for volatility due to peaks and valleys in the shipping industry, and there are always rent seekers attempting to gain financial returns by exploiting the unpredictability of ship prices. This situation has become more apparent during the coronavirus disease (COVID-19) pandemic. Although the ship sale and purchase market thrives on price volatility [1], it also indicates difficulties that valuation practitioners experience in estimating ship prices. Deep learning algorithms such as neural networks (NNs) have been appearing in the valuation literature since the 1990s [2], [3] to boost the predictive accuracy of asset valuation and have now started to replace the conventional models in the real estate valuation [4], [5]. However, in contrast to the active application of NNs to real estate valuation, few such studies have been conducted in the domain of ship valuation.

In this study, we attempt to apply a NN for estimating ship prices. First, we specify an appropriate NN architecture for application to tax returns of ship acquisition to estimate second-hand ship prices. We then compare the NN's prediction accuracy to that of a few baseline models. Finally, we interpret the result produced by NN training to provide useful insights for ship valuation stakeholders. Ship valuation plays a

key role in the shipping market, such as investment decisions of shipping investors, and ship loan approvals by financial institutions. In contrast to the new building price of a ship, which is essentially a forward contract for the delivery of an age-zero vessel in the future, the price of second-hand ships is remarkably unpredictable, and thus more difficult to estimate to a reasonable degree, as explained above. This study aims to estimate second-hand ship prices.

It is well known that deep learning has been disproportionately applied in unstructured-data-laden sectors such as computer vision and natural language processing industries. The use of the NN algorithm in this study is expected to significantly promote the adoption of deep learning tools in structured-data-dominant sectors including the shipping industry. Additionally, the empirical findings identified in the results of the NN may provide useful insights for ship valuation practitioners.

This paper is constructed. Section 2 describes background information on ship valuation and the entity embedding technique. Section 3 explains the data used and the NN architecture selected for estimating ship prices. Results and implications are interpreted in section 4. Lastly, conclusions are presented in section 5.

## 2. LITERATURE REVIEW

### 2.1. Ship valuation approach

The literature on ship valuation has largely focused on the analysis of the time-series properties of ship prices, such as analyzing co-integrating relationships between ship prices and the time-charter rate, or market trends in the new building prices and second-hand prices of ships [6]–[8]. Although this approach can provide an estimate of the value of a standardized generic ship based on a constructed time series of ship values, it cannot produce a price estimate for a specific ship. To provide price estimates at a vessel-specific level, valuation must be performed based on micro-scale cross-sectional data such as individual ship sales records. A few studies have been conducted based on cross-sectional data to generate ship-specific price estimates [1], [9], [10]. In this line of research, one of three valuation methods, namely the income approach, sales comparison approach, and cost approach, is typically employed. The income approach estimates ship prices using discounted cash flow analysis or Monte Carlo simulation techniques based on the freight rate or time-charter rate. Because the income approach is based on cash flow generated in the future, it is typically favored by financial institutions when approving loan applications for ships. In the sales comparison approach, actual transactions between buyers and sellers are collected and utilized for estimating ship prices. Because this method directly relies on market evidence, its estimates are generally accepted as convincing estimates by market participants. The cost approach estimates ship prices by subtracting depreciation from replacement costs and is widely adopted in tax assessment because it is relatively simple and easy to implement. However, estimates from the cost approach tend to deviate from the market value as a ship becomes obsolete. In an ideal market, price estimates from all three valuation approaches tend to converge.

In this study, the sales comparison approach was used for ship valuation. Its estimates are based on market evidence; thus, they are considered reliable and convincing by stakeholders, as explained earlier. Additionally, the prices of second-hand ships vary more drastically than those of new building prices depending on vessel-specific characteristics such as shipbuilding materials, engine manufacturers, and age. This individual heterogeneity cannot be effectively captured by the income approach or cost approach. In contrast, the sales comparison approach can capture heterogeneity; thus, it was adopted for estimating second-hand ship prices in this study.

### 2.2. Neural network (NN) with entity embedding

Deep learning has achieved excellent performance in terms of utilizing unstructured data such as images, audio, video, and free-form text. For example, it intensively exploits imagery data to operate self-driving cars [11], [12] or detect concrete building defects [13], [14]. In contrast to the unstructured data that are commonly used in deep learning areas, the dominant data type found in the shipping industry is spreadsheet-like structured data. Ship transaction records, ship inspection certificates, and safety construction certificates are specific examples of structured data observed in shipping businesses. These structured data were utilized thoroughly in this study.

Structured data comprise continuous variables and categorical variables. Categorical variables assume values that are names or labels such as vessel body color and engine type, which can be represented by fixed numbers. These numbers are called levels or elements and provide no meaningful information directly. Categorical variables are observed frequently in the shipping industry. For example, ships are often classified as passenger ships, fishing boats, or cargo ships. Ship valuation agents provide price estimates by considering key factors such as ship type and the presence/absence of specific features (with or without automatic radar plotting aid, and with or without lifeboats). These valuation factors are types of categorical variables. In the ship valuation literature, most studies have paid little attention to employing categorical

variables or have been reluctant to use them because efficient tools for processing such variables were unavailable. In [1] estimated ship prices using a multivariate density estimation method and the employed explanatory variables were deadweight tonnage, age, and the time-charter rate. All these are continuous variables that are convenient to use for density estimation. In [10] estimated second-hand ship prices using NNs, but their study also had limitations in that only continuous variables (age, time-charter rate, new building price, and scrap value) were utilized for price estimation. One reason for using only continuous variables is that NNs are natively well suited to processing continuous data but inefficient at handling categorical data.

Categorical variables need to be converted to numerical representations to allow a quantitative model to process them. Several methods have been developed for this purpose. The simplest method, namely the one-hot encoding approach, converts each element in a categorical variable into a new categorical column and assigns a binary value of one or zero to those columns. Although this method has been used commonly in the valuation literature [15], it has two disadvantages. First, when there are high-cardinality variables such as ship types or ZIP codes, one-hot encoding incurs excessive computational demand. Second, it treats different elements of categorical variables completely independently of each other and does not account for their informative interrelations.

If there is a clear ordering of the elements in a categorical variable, they can be translated into a set of numerical scores based on domain-specific knowledge. For example, ship agents can rank the grade of ship maintenance on an integer scale. A ship in a poor maintenance state may be scored as five, whereas an optimally maintained ship could be given a score of one. This method can be used efficiently in processing ordinal categorical variables [16], but has the drawbacks of demanding domain knowledge in the form of expert advice or consultation.

The entity embedding technique has been proposed as an alternative for extracting meaningful information from categorical variables more efficiently. Entity embedding is a technique for mapping categorical values into a multi-dimensional space with fewer dimensions than the original number of levels. In this space, values with similar function outputs are close to each other [17].

Figure 1 presents an example of categorical data embedding methods. If there is a categorical variable, day of week, each element in it is converted into a separate binary variable under the one-hot encoding approach. Figure 1(a) shows an example of this approach. The problem with it is that the informative relationship between each element is lost during conversion. In contrast, an embedding is a vector representation of a categorical variable, and the example can be represented with four numbers for each element. Figure 1(b) shows this method, and the number four is referred to as the embedding dimension.

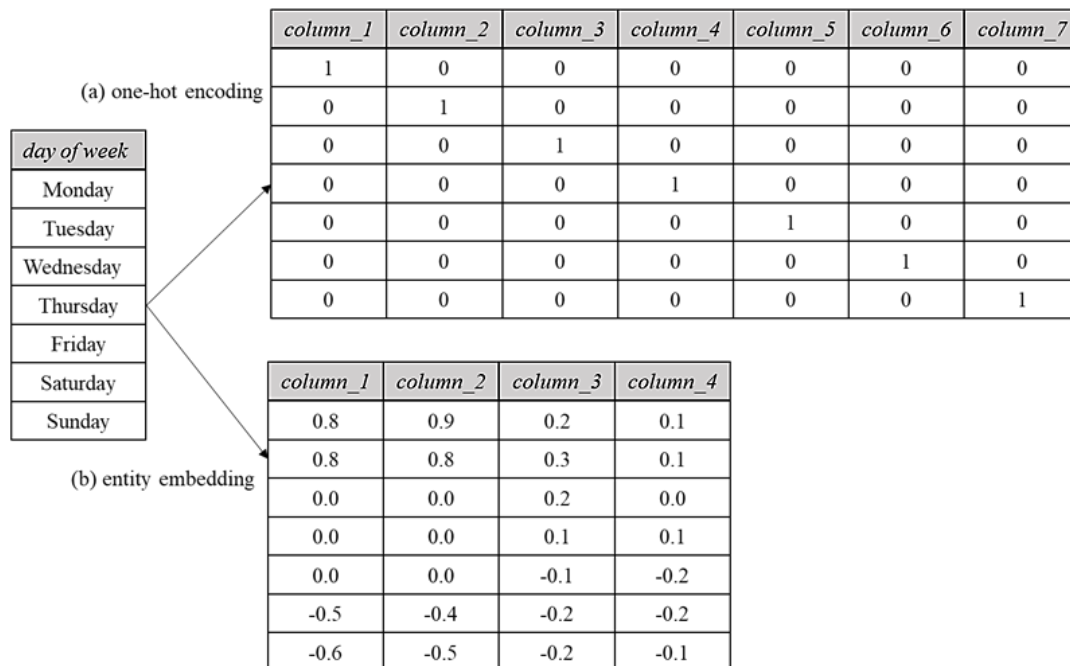


Figure 1. Example of categorical data encoding methods (a) one-hot encoding and (b) entity embedding

In Figure 1(b), Monday and Tuesday, Wednesday through Friday, and Saturday and Sunday are similar to each other, respectively, indicating that the entity embedding matrix reasonably captures interrelations between elements within the variable, day of week. Each value in the resultant embedding matrix represents the weights connecting the input layer to the embedding layer in an NN. Therefore, the embedding matrix contains the weights of the embedding layer and can be learned in the same manner as the parameters of other NN layers [17]. Through this representation, rich information among elements in a categorical variable can be captured efficiently, and can subsequently be utilized in an NN training. Lee [18] applied this approach to real estate valuation.

The problems discussed next can be alleviated by using the entity embedding technique. First, excessive computational resource consumption caused by the one-hot encoding of high-cardinality variables can be avoided. Second, different levels of categorical variables can be handled in a meaningful manner instead of these being treated completely independently of each other. Third, domain knowledge is not required once an NN is trained to learn the relationships between values of the same categorical variable efficiently. Finally, learned embeddings can be visualized using a dimensionality reduction technique, which can provide useful insights for stakeholders.

Several studies have estimated ship prices [9], [19]–[21], and recently, NNs have begun to be utilized to enhance the predictive accuracy of real estate valuation [22]–[24] and ship valuation [10]. However, most studies have not employed categorical variables for valuation, and even the few studies that have used these variables have not explicitly exploited the advantages of the aforementioned entity embedding technique. In this study, we attempted to fill this research gap by employing categorical variables actively in deep learning-based valuation.

The extracted features such as the embedding vectors can be reused in any subsequent models including a support vector machine, a random forest, and an NN. However, the main focus of this study is to create embedding vectors for categorical variables by using an NN and demonstrate their benefits in the context of asset valuation. Thus, an NN is adopted as the main model in this study.

### 3. METHOD

#### 3.1. Dataset

The dataset used in this study contained tax returns for property acquisitions of ships traded in 2018. The ship attributes available in the tax returns include the deadweight tonnage, age, and acquisition price, as reported by taxpayers. Table 1 presents descriptive statistics for the 3,475 ships used in our analysis.

The initial dataset included over 4,000 ships and the following records were removed during data preprocessing: records with missing values, redundant records, and implausible records such as an acquisition price reported as less than 10,000,000 KRW (approximately 9,000 USD). The median acquisition price for all ships in the dataset is 38,000,000 KRW (approximately 34,000 USD). The median deadweight tonnage is 5 tons and 99.4% of the ships (3,453 of 3,475) are less than 50,000 tons. Overall, this table indicates that small- and medium-sized ships are typically traded and reported for the purposes of property acquisition taxes. Ships with deadweight tonnage values of up to 50,000 tons are generally referred to as “handy-size” vessels and are mainly constructed in shipyards in South Korea, China, Japan, and Vietnam. Province refers to a local government having jurisdiction over the port of ship acquisition; 35.1% of the ships (1,218 of 3,475) were registered in Jeolla Province in 2018. The dominant ship type is a fishing boat (49.6%, 1,725 of 3,475).

Table 1. Descriptive statistics for 3,475 ships (transactions in 2018)

Variable	Min.	Mean	Median	Max.
Acquisition price (million KRW)	10	437	38	67,886
Deadweight tonnage	0.5	984	5	172,146
Age	1	16	15	74
Province (12 levels)	Jeolla: 1,218, Gyeongsang: 709, Busan: 488, Jeju: 327, Gyeonggi: 213, Incheon: 155, Chungcheong: 127, Gangwon: 105, Ulsan: 59, Daegu: 47, Gwangju: 18, Daejeon: 9			
Ship type (11 levels)	Fishing boat: 1,725, Small motor boat: 884, Combination carrier: 249, Barge: 158, General cargo ship: 147, Oil tanker: 104, Tugboat: 102, Passenger ship: 51, Yacht without motor: 43, Yacht with motor: 6, Special cargo ship: 6			

#### 3.2. Architecture of a neural network (NN)

Ship prices are affected by various factors, and selecting relevant variables involves a trade-off between valuation theory and data availability. Nine inputs were employed to estimate ship prices, as shown in Table 2. All variables excluding deadweight tonnage, age, and assessed value are categorical data. The

continuous variables were scaled to have a mean of zero and a standard deviation of one. The target variable was the acquisition price, which was log-transformed to alleviate a right-skewed distribution. The 3,475 ships were split into training data (80%, 2,780 ships) and test data (20%, 695 ships) in a random manner, and the training data were further divided into two sets for ordinary training (2,224 ships) and validation purposes such as hyperparameter tuning (556 ships). The test data were reserved for the final evaluation of NN performance.

Table 2. Input variables

Data type	Variable	Number of levels	Remarks
Categorical	Province	12	Embedding dimension: 4
	Ship type	11	Embedding dimension: 4
	Shipbuilding material	5*	One-hot encoding
	Engine make	2 (domestic, foreign)	One-hot encoding
	Transaction type	4**	One-hot encoding
	Buyer type	2 (individual, corp.)	One-hot encoding
Continuous	Deadweight tonnage	-	Scaling
	Age	-	Scaling
	Assessed value	-	Scaling

\*: wood, light-steel, steel, fiber-reinforced plastic, and others, \*\*: sales, auction, import, and others

This study adopted a fully connected layer NN or dense NN. Input layers were created, and embedding layers for the two high-cardinality categorical variables (province and ship type) were additionally created and joined to the architecture. A proper number of dimensions had to be defined for each embedding layer, and the prediction performance under various dimension sizes was investigated using the usual validation process. The number of dimensions assigned to each categorical variable based on this validation was four, as shown in the last column of Table 2. Three hidden layers were joined to the architecture to include more parameters to capture minor information contained in data. The final architecture of the NN is schematized in Figure 2.

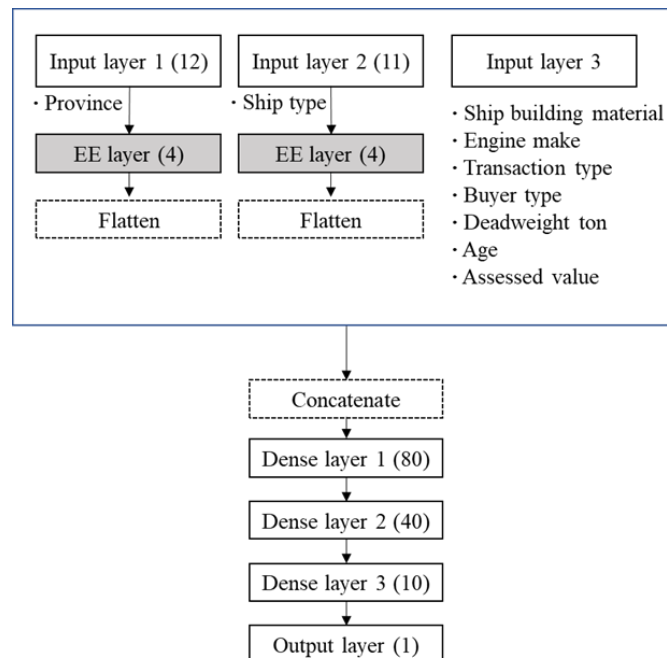


Figure 2. Final NN architecture

In Figure 2, EE stands for entity embedding. The number of levels in input layer 1 (Province) is 12, and its embedding dimension is four. The number of levels in input layer 2 (ship type) is 11, and its embedding dimension is four. The numbers in parentheses in dense layers 1, 2, and 3 and the output layer

indicate the numbers of units (neurons) in each layer. That is, the three dense layers following concatenation contain 80, 40, and 10 neurons in order. The output layer is a linear layer with one neuron representing the ship price.

The specific implementation details are: a gradient descent optimizer with momentum and Glorot initialization with a uniform distribution were used. A constant learning rate of 0.001 was adopted because the result was only trivially changed by alteration in the learning rate or learning schedule (e.g., exponential scheduling and power scheduling). A rectified linear unit (ReLU) activation function was used for all layers, except that a linear activation function was used for the output layer. The NN was trained for 30 epochs with a batch size of 64, and mean squared error was adopted as a loss function.

## 4. RESULTS AND DISCUSSION

### 4.1. Results

The mean absolute percentage error (MAPE) was used to evaluate model performance, as expressed by the following equation:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y - \hat{y}}{y} \right| \quad (1)$$

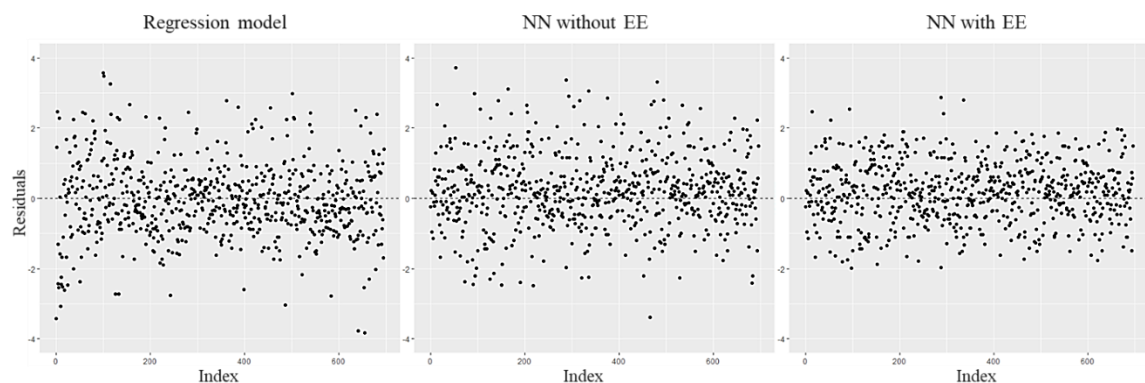
where  $y$  indicates the observed price, and  $\hat{y}$  denotes the estimated price from the NN. MAPE is a measure that is frequently used by property valuation agencies [25]. MAPE indicates the prediction error as a percentage and is convenient for comparisons across different valuation models. According to [26], [27], MAPE values greater than 10.0 are typically regarded as inappropriate for loan collateral programs and a valuation model with MAPE values in excess of 20.0 should not be used in applications involving risk.

The dataset of 3,475 records was randomly split into training (80%) and test (20%) datasets. Table 3 shows the MAPE results based on the test dataset. The performances of a few baseline models are presented for comparative purposes. Specifically, we consider a regression model and an NN without entity embedding layers. One-hot encoding was accepted for the categorical variables in both the regression model and NN without entity embedding. As listed in the table, MAPE is generally less than 10.0, indicating that the fitted models do not present serious drawbacks. The MAPE of the NN utilizing entity embedding layers is the lowest by a meaningful margin.

Table 3. Comparisons of model performance

	Regression model	NN without entity embedding	NN with entity embedding
MAPE	4.6	3.9	3.0

Figure 3 presents the distributions of residuals from the three models in Table 3. There are no specifically notable patterns in any of the three models. One clear result is that the range of residuals was reduced remarkably in the case of the NN with entity embedding layers, indicating its excellent prediction performance.



Note: EE stands for entity embedding.

Figure 3. Residual distributions for each model on the test dataset

#### 4.2. Interpreting learned embedding space

The primary aim of entity embedding is to map similar elements in a categorical variable close to each other in an embedding space. Then, a relevant question arises: how does the distribution of elements in a categorical variable appear in the embedding space? To present high-dimensional embeddings visually, t-SNE was employed to map the embeddings to a 2D space. t-SNE is a dimensionality reduction technique that produces outstanding visualizations by reducing the tendency to crowd points together in a map [28]. Figures 4 and 5 present the results of the province and ship type embeddings represented in a 2D space.

The province variable had 12 levels and was transformed to an embedding layer with four dimensions. Figure 4 presents the learned province embedding, which is visualized in a 2D space. Three distinct clusters can be observed. One group consisting of Incheon, Chungcheong, and Gyeonggi shares the same location, that is, all three provinces face the west coast of the Korean Peninsula. Another group consisting of Ulsan, Jeolla, Jeju, and Busan also shares the same location, that is, all four provinces in the group are located in the south of the Korean Peninsula facing the south coast. The third group consisting of the remaining five provinces seems to defy a relevant interpretation. However, it is promising that some of the geographically adjacent provinces are clustered together, despite the algorithm's ignorance of the geography in South Korea.

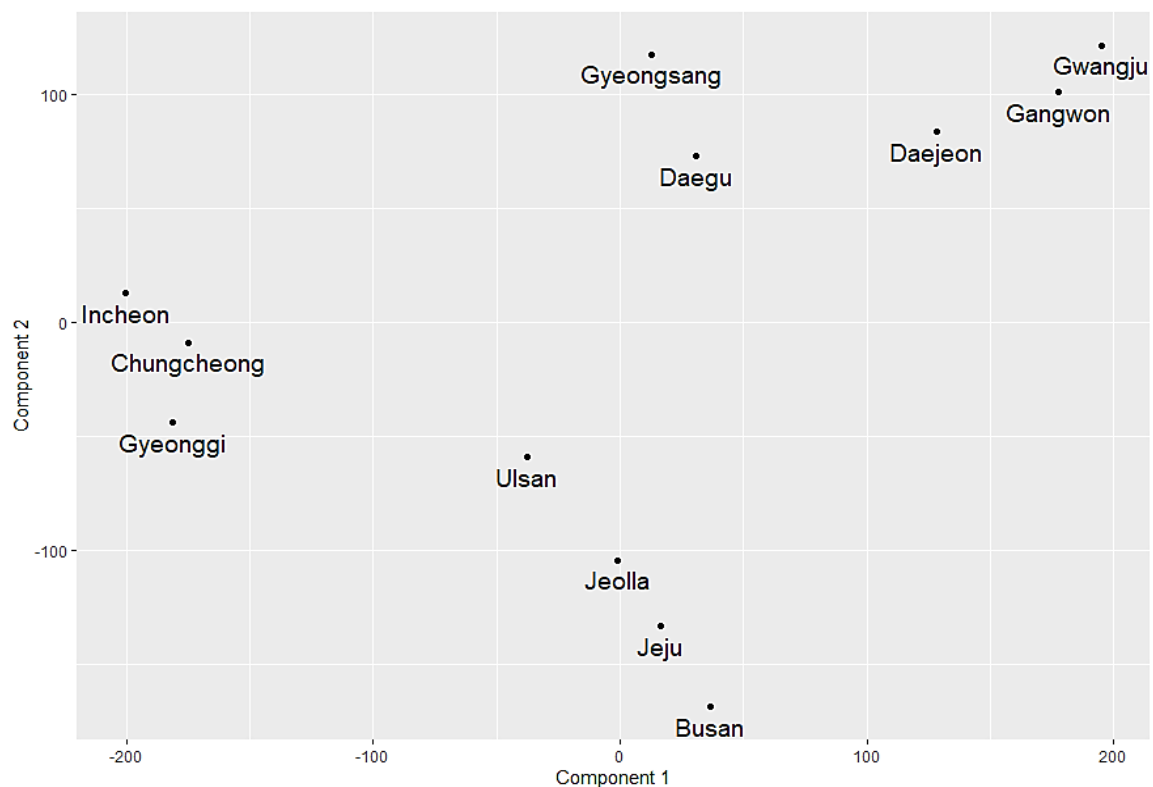


Figure 4. Province embedding mapped to a 2D space using t-SNE

The ship type variable had 11 levels and was converted into an embedding layer with four dimensions. Figure 5 presents the learned ship type embedding in a 2D space. Two main groups are identifiable in the figure. Passenger ship, general cargo ship, oil tanker, and special cargo ship are close to each other. The ships in this group are relatively expensive vessels with large sizes. The other ships are clustered in an oppositional position in the figure, and the ships in the second group can be interpreted as relatively cheap and small boats. These clustering results are notable because the NN algorithm is unaware of the characteristics of each ship type.

Although interpreting resultant embedding patterns is difficult and necessarily involves subjective judgment, it is for high-cardinality variables such as the province and ship type that the entity embedding technique exhibits noteworthy performance compared to the one-hot encoding approach.



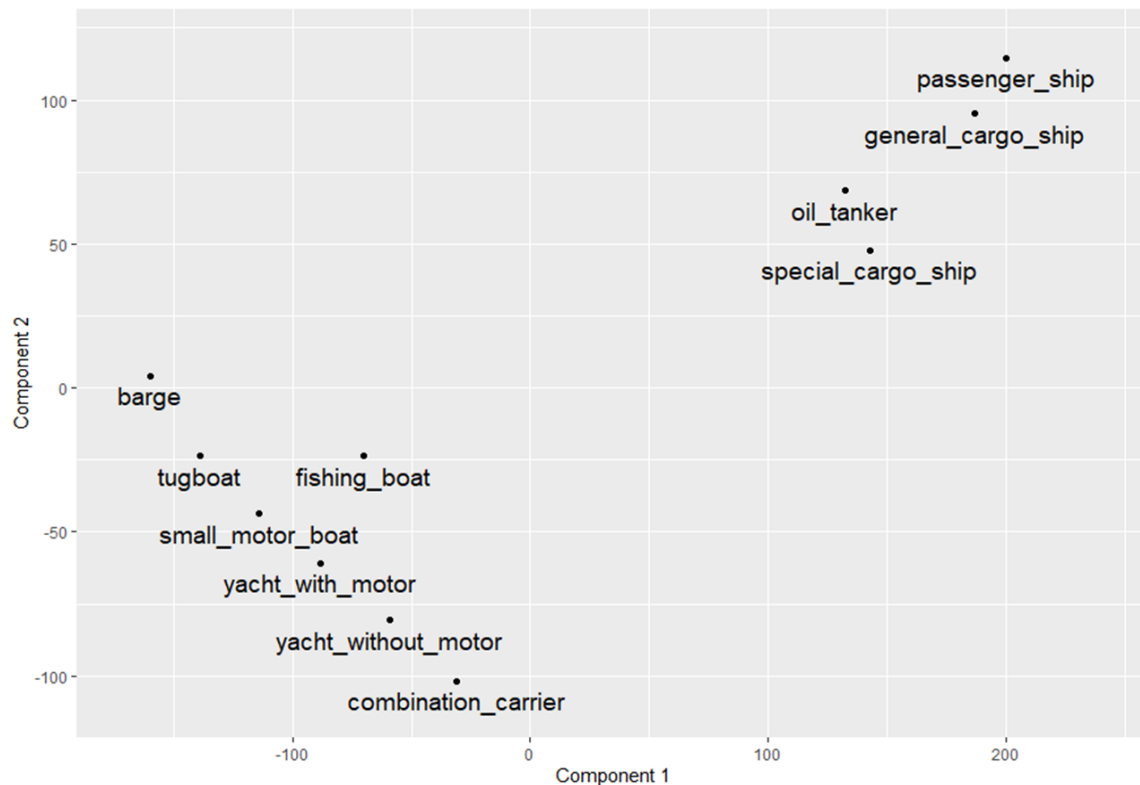


Figure 5. Ship type embedding mapped to a 2D space using t-SNE

## 5. CONCLUSION

Categorical variables are abundant in ship valuation, and the most common approach to handle these variables is to use one-hot encoding; thus, the advantages of entity embedding have not yet been explicitly exploited. We employed the entity embedding technique in this study to improve the performance of an NN by specifying and fitting an NN with entity embedding layers to tax returns for ship acquisitions. The results demonstrated that the NN with entity embedding layers outperformed the baseline models (regression model and NN without entity embedding layers). This improvement can be attributed to the capabilities of the entity embedding layers to capture informative relationships between the elements in each categorical variable. A study limitation is that we only discussed the clustering patterns in a compressed 2D embedding space and did not interpret the values in the embedding matrices. For example, the embedding matrix for the provinces had 12 rows and 4 columns. The element values in this matrix were learned during NN training. Subsequent studies should interpret the values in embedding matrices to provide more explainable deep learning models for stakeholders. Another limitation is the small number of variables used in the valuation model. The variables frequently reported in the literature include the time-charter rate, scrap value, oil price, and London Interbank Offered Rate. These variables are primarily responsible for reflecting changes in the shipping market and need to be utilized in future studies, particularly when the dataset consists of time-series data.

## REFERENCES




- [1] R. Adland and S. Koekebakker, "Ship valuation using cross-sectional sales data: a multivariate non-parametric approach," *Maritime Economics and Logistics*, vol. 9, no. 2, pp. 105–118, Jun. 2007, doi: 10.1057/palgrave.mel.9100174.
- [2] M. M. Lenk, E. M. Worzala, and A. Silva, "High-tech valuation: should artificial neural networks bypass the human valuer?," *Journal of Property Valuation and Investment*, vol. 15, no. 1, pp. 8–26, Mar. 1997, doi: 10.1108/14635789710163775.
- [3] S. McGreal, A. Adair, D. McBurney, and D. Patterson, "Neural networks: the prediction of residential values," *Journal of Property Valuation and Investment*, vol. 16, no. 1, pp. 57–70, Mar. 1998, doi: 10.1108/14635789810205128.
- [4] O. Poursaeed, T. Matera, and S. Belongie, "Vision-based real estate price estimation," *Machine Vision and Applications*, vol. 29, no. 4, pp. 667–676, May 2018, doi: 10.1007/s00138-018-0922-2.
- [5] S. Law, B. Paige, and C. Russell, "Take a look around," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 5, pp. 1–19, Sep. 2019, doi: 10.1145/3342240.
- [6] S. D. Tsolakis, C. Cridland, and H. E. Haralambides, "Econometric modelling of second-hand ship prices," *Maritime Economics and Logistics*, vol. 5, no. 4, pp. 347–377, Dec. 2003, doi: 10.1057/palgrave.mel.9100086.
- [7] R. Adland, H. Jia, and S. Strandenes, "Asset bubbles in shipping? an analysis of recent history in the drybulk market," *Maritime Economics and Logistics*, vol. 8, no. 3, pp. 223–233, Sep. 2006, doi: 10.1057/palgrave.mel.9100162.






- [8] L. Fan and J. Yin, "Analysis of structural changes in container shipping," *Maritime Economics and Logistics*, vol. 18, no. 2, pp. 174–191, Jun. 2016, doi: 10.1057/mel.2014.38.
- [9] J.-S. Choi, K.-H. Lee, and J.-S. Nam, "A ship valuation model based on Monte Carlo simulation," *Journal of Korea Port Economic Association*, vol. 31, no. 3, pp. 1–14, 2015.
- [10] D. Kim and J.-S. Choi, "Development of ship valuation model by neural network," *Journal of the Korean Society of Marine Environment and Safety*, vol. 27, no. 1, pp. 13–21, Feb. 2021, doi: 10.7837/kosomes.2021.27.1.013.
- [11] M.-T. Duong, T.-D. Do, and M.-H. Le, "Navigating self-driving vehicles using convolutional neural network," in *2018 4th International Conference on Green Technology and Sustainable Development (GTSD)*, Nov. 2018, pp. 607–610, doi: 10.1109/GTSD.2018.8595533.
- [12] H. Song, "The application of computer vision in responding to the emergencies of autonomous driving," in *2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, Jul. 2020, pp. 1–5, doi: 10.1109/CVIDL51233.2020.00008.
- [13] M. Alipour, D. K. Harris, and G. R. Miller, "Robust pixel-level crack detection using deep fully convolutional neural networks," *Journal of Computing in Civil Engineering*, vol. 33, no. 6, Art. no. 04019040, Nov. 2019, doi: 10.1061/(ASCE)CP.1943-5487.0000854.
- [14] H. Perez, J. H. M. Tah, and A. Mosavi, "Deep learning for detecting building defects using convolutional neural networks," *Sensors*, vol. 19, no. 16, Art. no. 3556, Aug. 2019, doi: 10.3390/s19163556.
- [15] R. Gloudemans and R. Almy, *Fundamentals of mass appraisal*. International Association of Assessing Officers, 2011.
- [16] A. Agresti, *Analysis of ordinal categorical data*. Hoboken, NJ, USA: John Wiley and Sons, Inc., 2010.
- [17] C. Guo and F. Berkhahn, "Entity embeddings of categorical variables," Apr. 2016. Available: <http://arxiv.org/abs/1604.06737>.
- [18] C. Lee, "Enhancing the performance of a neural network with entity embeddings: an application to real estate valuation," *Journal of Housing and the Built Environment*, Aug. 2021, doi: 10.1007/s10901-021-09885-2.
- [19] A. Miroyanis, "Estimation of ship construction costs," Massachusetts Institute of Technology, 2006.
- [20] T. Syriopoulos and E. Roumpis, "Price and volume dynamics in second-hand dry bulk and tanker shipping markets," *Maritime Policy and Management*, vol. 33, no. 5, pp. 497–518, Dec. 2006, doi: 10.1080/03088830601020729.
- [21] L. Dai, H. Hu, F. Chen, and J. Zheng, "The dynamics between newbuilding ship price volatility and freight volatility in dry bulk shipping market," *International Journal of Shipping and Transport Logistics*, vol. 7, no. 4, pp. 393–406, 2015, doi: 10.1504/IJSTL.2015.069666.
- [22] O. Al-Gburi and S. Kurnaz, "Real estate price range prediction using artificial neural network and grey wolf optimizer," in *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, Oct. 2020, pp. 1–5, doi: 10.1109/ISMSIT50672.2020.9254972.
- [23] H. Seya and D. Shiroy, "A comparison of residential apartment rent price predictions using a large data set: kriging versus deep neural network," *Geographical Analysis*, Mar. 2021, doi: 10.1111/gean.12283.
- [24] L. N. Yasnitsky, V. L. Yasnitsky, and A. O. Alekseev, "The complex neural network model for mass appraisal and scenario forecasting of the urban real estate market value that adapts itself to space and time," *Complexity*, vol. 2021, pp. 1–17, Mar. 2021, doi: 10.1155/2021/5392170.
- [25] M. Ecker, H. Isakson, and L. Kennedy, "An exposition of AVM performance metrics," *Journal of Real Estate Practice and Education*, vol. 22, no. 1, pp. 22–39, Jan. 2020, doi: 10.1080/15214842.2020.1757352.
- [26] P. Rossini and P. J. Kershaw, "Automated valuation model accuracy: Some empirical testing," 2008.
- [27] H. R. Isakson, M. D. Ecker, and L. Kennedy, "Principles for calculating AVM performance metrics," *The Valuation Journal, National Association of Romanian Valuers*, vol. 16, no. 2, pp. 38–69, 2020.
- [28] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

## BIOGRAPHIES OF AUTHORS



**Changro Lee**    is a professor at the Department of Real Estate, Kangwon National University (KNU), South Korea. Before joining KNU, Lee was a researcher at Korea Institute of Local Finance (KILF). He has worked in the fields of real estate management, machine learning, and local financing of real estate. He can be contacted at email: [spatialstat@naver.com](mailto:spatialstat@naver.com).



**Keyho Park**    holds a Bachelor of Engineering (B.Eng.) in Architecture, Master of Urban Planning (MUP), and Ph.D in Computer Science. He has been a professor since 1995 at Department of Geography, Seoul National University (SNU), South Korea. His teaching and research areas of interest include geographic information science, spatio-temporal statistics and machine learning. He can be contacted at email: [khp@snu.ac.kr](mailto:khp@snu.ac.kr).