Application Machine Learning in Construction Management

Nguyen, Phong Thanh

Ho Chi Minh City Open University

29 December 2020

Application Machine Learning in Construction Management

Phong Thanh Nguyen

Ho Chi Minh City Open University, Ho Chi Minh City, Vietnam

***Abstract* – Machine Learning is a subset and technology developed in the field of Artificial Intelligence (AI). One of the most widely used machine learning algorithms is the K-Nearest Neighbors (KNN) approach because it is a supervised learning algorithm. This paper applied the K-Nearest Neighbors (KNN) algorithm to predict the construction price index based on Vietnam's socio-economic variables. The data to build the prediction model was from the period 2016 to 2019 based on seven socio-economic variables that impact the construction price index (i.e., industrial production, construction investment capital, Vietnam's stock price index, consumer price index, foreign exchange rate, total exports, and imports). The research results showed that the construction price index prediction model based on the K-Nearest Neighbors (KNN) regression method has fewer errors than the traditional method.**

***Keywords* – Artificial Intelligence, K-Nearest Neighbors (KNN), machine learning, price index, construction management**

1. Introduction

The transition from a centrally planned economy to a market economy has helped Vietnam transform from a poor country to a low-middle-income country [1]. It is one of the most competitive and emerging countries in the East Asia Pacific region, with a fast growth rate (i.e., GDP increased by seven percent in 2019) [2]. This is due to the industrialization and urbanization of a significant part of the construction industry [3, 4]. However, due to cost overruns, many civil and infrastructure projects have been failures [5, 6]. In the construction industry, projects are characteristically enormous requiring major capital outlays. This leads to risk and uncertainty in terms of cost estimation and management. Over time, changes in the cost or price of materials, labour, and machine shifts in the market economy make it difficult for estimators to forecast a project's total budget [7, 8]. Therefore, developing useful cost prediction models to overcome is essential.

The construction price index accurately predicts project costs, prepares the budgets at the initial planning level, and manages and tracks costs during the project life cycle [9, 10]. Using data over time on socioeconomic variables, this paper introduces a new construction price index prediction model based on a machine learning approach, namely the K-Nearest Neighbors (KNN). The K-Nearest Neighbors (KNN) regression method's benefit is that its data is split into two sections: training data and validation data. The research results show that the construction price index prediction results based on the modified KNN regression have fewer errors than the conventional method.

2. Research background and methodology

Artificial Intelligence (AI) is one of the core technologies in the world for Industrial Revolution 4.0 [11]. It is a human-programmed intelligence to help computers automate intelligent behaviours such as those of the humans. In other words, using computers, especially computer systems, AI simulates human thought, and learning processes. This method involves understanding (the compilation of facts and guidelines for the use of information), logic (the use of rules to arrive at approximate conclusions), decision-making, and self-correction.

Machine learning is a subset and technology developed from the field of AI [12, 13]. It is a multidisciplinary field in which the knowledge of statistics, computer science, and optimization are combined to solve problems such as classification, clustering, and regression [14]. In other words, machine learning is a system that can learn from data. It can handle uncertainty and incomplete data, as well as to make decisions based on experience from analogous cases [15].

Although still quite limited, in recent years, several studies have been conducted on the use of machine learning in construction and project management by several researchers. To classify construction documents based on project elements, Caldas and Soibelman [16] applied the support vector machine (SVM) algorithm. It was used to enhance the quality of the existing construction information management system. Chen [17] used the KNN classification algorithm to build a model of awareness and information exchange. The proposed model can be used to assist multiple stakeholders in participating in projects to determine and mitigate the risk of litigious change disputes.

The cost of construction projects and a schedule using an artificial neural network (ANN) and support a vector machine (SVM) was projected by Wang, et al. [18]. The findings suggested that early planning is a crucial factor in a project's success. To categorize the clauses of general conditions in contracts, Salama Dareen and El-Gohary Nora [19] created a machine-learning-based text- classification algorithm. This method facilitates the automatic compliance checking of textual construction contracts.

Using neural networks and linear regression, Elfahham [20] estimated a construction cost index for concrete buildings based on historical reports of the main construction costs. Elfahham [20] analysis key contribution was to provide stakeholders with a credible method for predicting prices for future project developments.

In Nguyen and Nguyen [21] research, in-depth interviews with construction experts and correlation relationships were conducted to find key socio-economic variables that influence the construction price index in Vietnam. Table 1. summarizes seven socio-economic variables as raw data for the KNN prediction model of the construction price index.

Table 1. Explanatory and dependent variables

Variable names in the KNN model	**Description**
	Explanatory Variables
X1	Industrial production
X2	Construction investment capital
X3	Consumer price index
X4	Foreign exchange rate
X5	Total export
X6	Total import
X7	Vietnam's stock price index
	Dependent variable
Y	Construction price index

K-Nearest Neighbors (KNN) is a supervised learning algorithm [22, 23]. The benefit of the KNN regression method is that its dataset is split into two sections: training data and validation data. In a prediction problem using KNN regression, a data point's output will be determined based on the output of neighbouring points and their distance. The KNN algorithm suggested that similar data would exist close together in a space to find the K points nearest to the input data to be examined. The algorithm makes decisions based on these nearest points [24, 25]. Thus, the crux of the algorithm is the calculated distance between these points. In general, each data point is represented by one feature vector, so the distance between two points is the distance between two vectors.

There are many ways to calculate the distance between the nearest points in the K-Nearest Neighbors as shown in formulars from (1) to (3) (i.e., Euclidean, Manhattan, and Minkowski distance metrics) [26]. Commonly, the Euclidean formula is used as follows (i.e., the distance between 2 points is the length of the line connecting them):

$$d_i = \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2} \quad (1)$$

$$d_i = \sum_{i=1}^{k} |x_i - y_i| \quad (2)$$

$$d_i = \left(\sum_{i=1}^{k} \left(|x_i - y_i| \right)^q \right)^{1/q} \quad (3)$$

3. Research Results

Using Python language and scikit-learn library, the construction price index prediction model based on K-Nearest Neighbors was developed using raw data of seven socio-economic variables in Ho Chi Minh City, Vietnam as training data in Table 2.:

After it was pre-processed, the entire data set composed of input variables were divided into two data sets (i.e., training dataset and validation dataset). The last twelve (12) data points are used as a test set to check the prediction model's accuracy. Next, the deviation and percentage error between the results from the construction department in Ho Chi Minh City predicted by the K-Nearest Neighbors model were compared with the state agencies' calculation results. Based on the research result summarized in Table 3. and Figure 1., we can see that the percentage error was relatively low, so it can be concluded that the construction price index prediction model is relatively accurate and feasible.

Table 3. Construction price index prediction results and comparison

t	Actual	Forecast	Deviation	Percentage Error
37	100.74	99.36	1.38	1.37%
38	100.72	98.75	1.97	1.96%
39	101.02	98.93	2.09	2.06%
40	101.24	98.94	2.30	2.27%
41	101.27	100.23	1.04	1.03%
42	101.40	99.42	1.98	1.95%
43	101.40	99.63	1.77	1.74%
44	101.08	100.05	1.03	1.02%
45	101.08	100.23	0.85	0.84%
46	100.79	99.67	1.12	1.11%
47	100.57	100.19	0.38	0.38%
48	100.74	100.01	0.73	0.73%

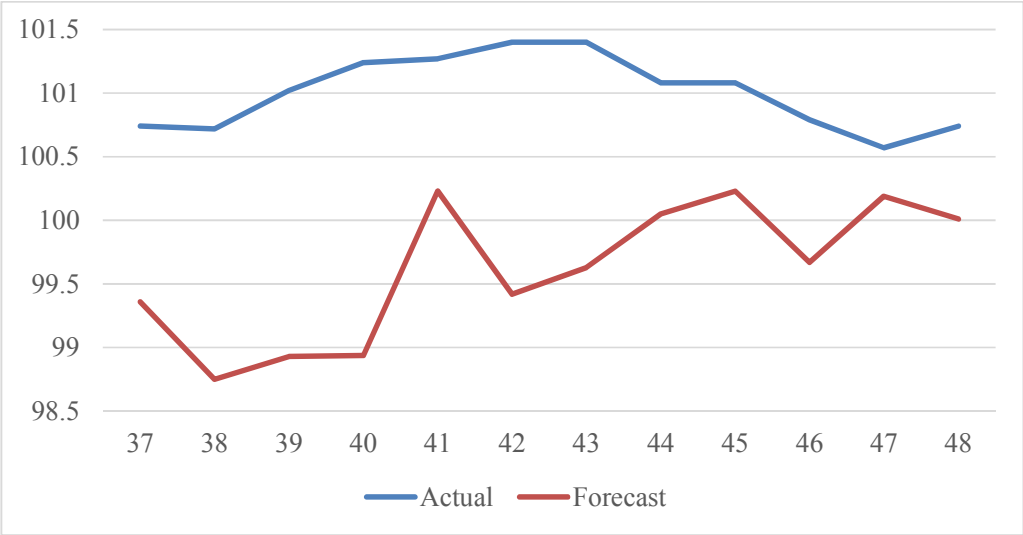

Figure 1. Construction price index comparison

Table 2. Monthly data of variables in the construction price index prediction model

t	X1	X2	X3	X4	X5	X6	X7	Y
1	90.60	103.10	99.52	100.58	85.50	88.60	91.30	97.94
2	73.00	101.20	100.02	99.79	80.00	74.00	103.60	98.08
3	130.70	102.50	100.10	98.86	129.00	145.00	71.70	97.13
4	93.30	106.20	100.47	99.82	101.00	103.00	186.20	97.28
5	100.50	101.30	100.82	99.98	111.50	105.00	75.50	97.23
6	99.20	109.10	100.80	100.26	105.10	102.50	94.20	97.26
7	96.40	101.50	100.19	99.80	102.10	100.90	111.30	96.30
8	100.81	101.80	99.80	99.96	98.10	100.70	102.20	96.33
9	103.46	109.94	100.43	100.01	98.00	104.00	94.80	96.35
10	105.03	102.10	100.62	100.14	102.00	105.60	96.74	96.35
11	103.18	105.70	100.55	100.33	96.30	101.20	102.88	96.43
12	103.86	109.26	100.52	102.77	113.00	100.50	96.27	96.55
13	86.94	96.40	100.85	99.75	87.00	90.00	111.31	97.67
14	92.85	101.30	100.50	99.46	85.00	89.50	58.59	97.71
15	108.20	111.10	99.91	99.95	120.00	111.00	194.08	98.07
16	98.98	102.40	99.82	99.60	94.00	96.00	136.91	98.27
17	103.09	102.50	100.04	100.03	112.00	106.00	77.67	98.45
18	102.72	106.10	100.00	99.82	106.00	95.00	137.38	98.44
19	99.64	106.60	99.84	100.30	102.00	96.00	96.65	99.70
20	102.99	106.20	100.50	99.89	101.00	101.50	96.62	100.08
21	100.49	110.20	100.90	99.99	96.00	98.00	106.26	100.16
22	104.100	105.60	100.63	99.99	103.50	104.40	66.41	99.70
23	102.88	112.20	100.17	99.96	97.00	105.00	115.67	99.79
24	106.37	118.40	100.05	100.00	107.00	103.40	139.58	99.79
25	91.29	104.20	100.19	99.98	106.00	103.00	87.54	99.71
26	74.15	104.50	100.34	99.98	67.00	68.00	366.92	99.69
27	142.12	104.50	99.70	100.21	159.00	145.00	55.90	100.35
28	102.95	106.20	100.12	100.17	110.00	103.00	149.28	100.54
29	103.85	106.80	100.43	99.91	108.50	108.00	81.02	100.20
30	104.81	111.70	100.55	100.18	105.00	102.00	91.81	100.41
31	101.71	109.60	99.91	100.84	105.30	106.00	91.79	100.98
32	106.15	113.80	100.48	101.18	107.30	106.20	107.48	101.10
33	98.31	114.10	100.81	100.09	87.00	95.00	111.55	101.38
34	103.49	122.30	100.64	100.21	105.00	102.00	91.05	101.53
35	102.16	118.50	99.75	99.96	105.30	106.00	117.01	101.52
36	102.24	120.30	100.75	99.95	112.00	110.00	73.76	101.42

4. Conclusion

Vietnam's economy currently has one of the best growth rates in the East Asia region. This was due to the growth of industries including the construction industry. Construction is one of the primary industries in the national economy and plays a crucial role in projects serving the infrastructure of the society. For these projects to be successful, estimators need to apply new methods of the new achievements of the Fourth Industrial Revolution in the construction of accurate cost forecasting models. One of the achievements has been the machine learning methods of artificial intelligence. It is a multidisciplinary field that combines the knowledge of statistics, computer science, and optimization to solve many problems such as classification, clustering, regression, prediction, etc. One of the most widely used machine learning algorithms is the K-Nearest Neighbors (KNN) approach because it is a supervised learning algorithm. In a prediction problem by KNN regression, the output of a data point will be determined based on the output of neighbouring points, and their distance. After examining the main socio-economic factors that affect the construction price index, this study applied the KNN to model it. The KNN method has been proved by researchers to be fair with more accurate results than traditional methods (i.e., the time series method). Also, not merely building a prediction model based on time series or using direct factors such as materials, labour, machine shifts, this KNN model also considers socioeconomic factors that affect the construction price index. This model reflects a more general and complete picture of the construction price index forecast.

Acknowledgements

This research is funded by the Ministry of Education and Training, Vietnam, and Ho Chi Minh City Open University, Vietnam under the grant number B2019-MBS-05.

References

[1] Mishra, D. (2011). *Vietnam development report 2012: Market economy for a middle-income Vietnam* (No. 65980, pp. 1-90). The World Bank.

[2] Nguyen, H. M., & Bui, N. H. (2020). Revisiting the relationship between energy consumption and economic growth nexus in Vietnam: new evidence by asymmetric ARDL cointegration. *Applied Economics Letters*, 1-7.

[3] Van, D. T. B., & Bao, H. H. G. (2018). The role of globalization on CO2 emission in Vietnam incorporating industrialization, urbanization, GDP per capita and energy use. *International Journal of Energy Economics and Policy, 8*(6), 275.

[4] Nguyen, T. P., & Chileshe, N. (2015). Revisiting the construction project failure factors in Vietnam. *Built Environment Project and Asset Management*.

[5] Nguyen, P. T., Likhitruangsilp, V., & Onishi, M. (2020). Success factors for public-private partnership infrastructure projects in Vietnam. *International Journal on Advanced Science, Engineering and Information Technology, 10*(2), 858-865.

[6] Le-Hoai, L., Lee, Y., & Cho, J. W. (2009). Construction of time-cost model for building projects in Vietnam. *Korean Journal of Construction Engineering and Management, 10*(3), 130-138.

[7] Xie, S., & Fang, J. (2018). Prediction of construction cost index based on multi variable grey neural network model. *International Journal of Information Systems and Change Management, 10*(3), 209-226.

[8] Cao, Y., & Ashuri, B. (2020). Predicting the Volatility of Highway Construction Cost Index Using Long Short-Term Memory. *Journal of Management in Engineering, 36*(4).
[9] Baldassari, P., & Roux, J. D. (2017). Industry 4.0: preparing for the future of work. *People & Strategy, 40*(3), 20-24.
[10] Rudin, C., & Wagstaff, K. L. (2014). Machine learning for science and society, Springer.
[11] Canhoto, A. I., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons, 63*(2), 183-193.
[12] Rebala, G., Ravi, A., & Churiwala, S. (2019). Machine Learning Definition and Basics. In *An Introduction to Machine Learning* (pp. 1-17): Springer.
[13] Ziora, L. (2020). Machine learning solutions in the management of a contemporary business organisation. *Journal of Decision Systems*, 1-8.
[14] Caldas, C. H., & Soibelman, L. (2003). Automating hierarchical document classification for construction management information systems. *Automation in Construction, 12*(4), 395-406.
[15] Chen, J.-H. (2008). KNN based knowledge sharing model for severe change order disputes in construction. *Automation in Construction, 17*(6), 773-779. https://doi.org/10.1016/j.autcon.2008.02.005
[16] Xie, S., & Fang, J. (2018). Prediction of construction cost index based on multi variable grey neural network model. *International Journal of Information Systems and Change Management, 10*(3), 209-226.
[17] Salama Dareen, M., & El-Gohary Nora, M. (2016). Semantic Text Classification for Supporting Automated Compliance Checking in Construction. *Journal of Computing in Civil Engineering, 30*(1), 04014106.
[18] Elfahham, Y. (2019). Estimation and prediction of construction cost index using neural networks, time series, and regression. *Alexandria Engineering Journal, 58*(2), 499-506.
[19] Nguyen, P. T., & Nguyen, Q. L. H. T. T. (2020). Critical Factors Affecting Construction Price Index: An Integrated Fuzzy Logic and Analytical Hierarchy Process. *The Journal of Asian Finance, Economics and Business, 7*(8), 197-204.
[20] Lin, W.-C., Ke, S.-W., & Tsai, C.-F. (2017). Top 10 data mining techniques in business applications: a brief survey. *Kybernetes*.
[21] Kohli, S., Godwin, G. T., & Urolagin, S. (2020). Sales Prediction Using Linear and KNN Regression. In *Advances in Machine Learning and Computational Intelligence* (pp. 321-329): Springer.
[22] Aswani, R., Ghrera, S., Kar, A. K., & Chandra, S. (2017). Identifying buzz in social media: a hybrid approach using artificial bee colony and k-nearest neighbors for outlier detection. *Social Network Analysis and Mining, 7*(1), 38.
[23] Ashuri, J. W. A. B. (2016). Predicting ENR Construction Cost Index Using the Modified K Nearest Neighbors (KNN) Algorithm. In *Construction Research Congress 2016* (pp. 2502-2509).
[24] Singh, A., Yadav, A., & Rana, A. (2013). K-means with Three different Distance Metrics. *International Journal of Computer Applications, 67*(10).