Lee J.H., Kim H. S., Lee E. and Lee S.Y. (2020). "Development of AI-based engineering Big Data integrated analysis system for decision-making support in the Engineering-Procurement-Construction (EPC) industry" In: *Proc. 37th CIB W78 Information Technology for Construction Conference* (CIB W78), São Paulo, Brazil, pp. 58-68. DOI: <a href="http://dx.doi.org/10.46421/2706-6568.37.2020.paper005">http://dx.doi.org/10.46421/2706-6568.37.2020.paper005</a>

# DEVELOPMENT OF AI-BASED ENGINEERING BIG DATA INTEGRATED ANALYSIS SYSTEM FOR DECISION-MAKING SUPPORT IN THE ENGINEERING-PROCUREMENT-CONSTRUCTION (EPC) INDUSTRY

Jung Hyun. Lee<sup>1</sup>, Hyun Soo. Kim<sup>2</sup>, Eul-bum. Lee<sup>3</sup> and Seung Yeab. Lee<sup>4</sup>

**Abstract:** Plant Engineering-Procurement-Construction (EPC) industry is one of the complex industries going through various stages from bidding to engineering, construction and operation and maintenance (O&M). A systematic management system is needed to address these complexities. However, many EPC companies in Korea are having difficulty managing their projects due to the lack of data-based systematic decision-making, and are suffering heavy losses in overseas projects.

The AI-based engineering big data integrated analysis system proposed by this study aims to minimize project losses and eventually to enhance the technical skills and competitiveness of the Korean plant industry through decision-making support, combining big data and AI in the entire EPC project life cycle. In this study, knowledge base was established to utilize various data generated during the entire EPC project life cycle in AI-based engineering big data integrated analysis systems. And a machine learning integrated platform specialized in the engineering industry was developed to support feature engineering, model learning and model operation processes. Using various algorithms from the machine learning integration platform and the knowledge base, five main decision-making applications were developed: analyzing bidding documents, predicting design costs, analyzing design errors, analyzing change order, and plant equipment prediction maintenance.

Based on the predicted information, the system could help EPC project managers identify and manage risks at each stage of the project in advance to make decisions that minimize project loss. Furthermore, the information predicted at each stage may be circulated or used as feedback for decision making at other stages.

**Keywords:** Plant Project, Engineering, Construction Lifecycle, Smart Decision-making support system, Big-Data, Artificial Intelligence.

#### 1 Introduction

The plant industry is made up of large, complex industries having various phases, ranging from bidding to engineering, construction, operation and maintenance, as well

Graduate Student, Pohang University of Science and Technology(Postech), Pohang, Korea, <a href="leejunghyun@postech.ac.kr">leejunghyun@postech.ac.kr</a>

Professor, Pohang University of Science and Technology(Postech), Pohang, Korea, dreblee@postech.ac.kr

<sup>&</sup>lt;sup>4</sup> Chief Researcher, Pohang University of Science and Technology(Postech), Pohang, Korea, sylee7@postech.ac.kr

as high-tech manufacturing technologies, and technology-intensive industries that require knowledge services such as design, production and finance.

However, many EPC companies are having difficulties in Life-cycle construction management because they do not have a data-based, systematic decision-making system such as a track record. In addition, insufficient management of unknown risks such as Country Risk and Schedule Risk often results in losses in construction.

In addition, in order for Korean EPC companies to survive the competition with major global EPC companies, it is imperative to secure competitiveness against plant value chains, including high value upstream parts such as Project Management Consultancy (PMC) and Front-End Engineering Design (FEED).

Therefore, in order to lay the foundation for securing the competitiveness of the Korean plant industry, we suggest AI-based engineering big data integrated analysis system technology that enables optimal decision making and execution in the pre-plant cycle process so that Korean EPC companies could predict and respond to risks in the bidding, execution, construction and maintenance phases of the project in advance.

To support decision-making by project stage, the research team built a knowledge base ranging from ITB (Invitation to Bid, bid document) analysis at the bidding stage to plant facility forecast maintenance of the O&M stage, select-ed features, and combined algorithms of machine learning platforms to create a machine learning solution.

The following Table 1 shows the five modules of this engineering decision-making technology development.

Project Stage	Module	Contents
Bidding	Engineering design cost prediction	Analysing past engineering design man-hour and cost to predict accurate engineering design man-hour of new project
Bidding	Engineering ITB analysis	Establish bidding strategy through analysis of con-tract key issues and respond to risks through analysis of risk clause
Engineering	Engineering Design Error Analysis	Analysis of design error report (crash, missing report) information to provide proactive design risks and types of equipment and equipment with high potential for design delay and error
Engineering &Construction	Engineering Change Order Analysis	Analyse the causes of design changes and present risk impact and trend by types, and provide information that can be reflected in contract, design, purchase, and construction work.
Operation &Maintenance	Plant Equipment Predictive Maintenance	Predict maintenance items for major plant facilities

Table 1: Summary of five study parts.

In addition to introducing the above five key decision support modules in Section 3, this Paper will cover one of the decision support modules in Section 4 in more detail.

#### 2 LITERATURE REVIEW

In recent years, the construction industry has become more complex, and the need for advanced project management is increasing. As a result, there are a growing number of cases and studies that apply big data, AI, and machine learning algorithms to the construction industry. [1] James. D(2005) of AACE present-ed a correlation and prediction study between equipment and plant engineering design man-hours using regression analysis. [2] Mohamed Marzouk(2016) developed a water treatment plant construction estimating model using the Artificial Neutral Network Method in 2016. [3]Sphurti S. Arage (2017) and [4] Igor Pesko (2017) developed a Civil Construction Cost Estimation model by applying machine learning algorithms in their respective studies.

However, compared to other industries, the plant industry is relatively slow in applying machine learning and quantitative statistical analysis. In particular, researches for predicting plant engineering design man-hours are not actively con-ducted, and each EPC company or Engineering company adheres to the tradition-al method of estimating and calculating quantities. Therefore, this study intends to present engineering design man-hour prediction model after collecting historical data and performing quantitative analysis.

#### 3 FIVE DECISION SUPPORT MODULES INTRODUCTION

As mentioned above, this study consists of five modules, such as engineering design cost prediction, engineering design error analysis, engineering design change analysis, engineering ITB analysis, and plant facility forecasting maintenance, and this section provides a summary of the objectives and development methods of these five modules.

## 3.1 Engineering Design Cost Prediction

Estimating design costs is the work performed in the initial bidding phase of the project and has a great impact on the decision making of the project participants. EPC Company or Engineering Company shall estimate accurate design costs and reflect them in the schedule and expense plan before embarking on plant design work. However, since there is limited information available at the beginning of the project, qualitative factors such as the experience of the engineers in charge are often reflected, and hence the accuracy of the estimate has been often less accurate. Thus, this research team collected about 40 past EPC plant project data consisting of drawing list, equipment list and project information, etc. and con-ducted analysis from a big data perspective. As a result, a model was developed to estimate the engineering design man-hour using information from the beginning of the project, and a function was implemented to select and provide useful reference data to project participants.

# 3.2 Engineering Design Error Analysis

In plant projects, where a large amount of complex design is executed in a short period of time, design errors and omissions often occur. These design errors and omissions are one of the main risk factors that cause schedule delay, cost in-creases, and quality degradation during the subsequent construction phase, and not just during the design phase. Therefore, in this module after collecting historical plant project data, we developed an automatic classification system of design errors and an algorithm to predict the risks from the cost aspects of each type in order to minimize a project risk due to design errors by utilizing big data statistical analysis and machine learning.

# 3.3 Engineering Change Order Analysis

Due to the nature of long-running plant projects, design changes are often caused by internal causes of the project, such as scope, function and use, and by external factors such as political, economic and environmental changes. However, it is not uncommon for compensation from owners for delays and increases in costs caused by design changes to be paid improperly. And this leads to a loss to project participants. Therefore, this module developed a design change type automatic classification system and severity prediction algorithm by classifying de-sign changes by type and calculating the severity by type based on historical data so that users can proactively recognize and cope with the risks of design changes occurring.

## 3.4 Engineering ITB Analysis

The bidding phase of a plant project is a very important period for project participants. Early detection of potential risk conditions on bid documents and prevention of losses is a prerequisite for successful project execution within a limited period of time, and is generally required by a large number of experts in each field. In this module, the NLP(Natural Language Process) and text-mining technology, and machine learning are used to prepare systematic procedures for the extraction of risk clauses on bid documents, and to automate tasks to enable more objective and quick review of bid documents.

#### 3.5 Plant Equipment Predictive Maintenance

Ensuring the stability of plant facility system operation is of paramount importance to prevent major accidents. Although preventive maintenance and parts replacement are carried out regularly to prevent the failure of the facility system, such existing methods are limited in preventing the sudden failure. Therefore, to minimize the loss of plant operators, this study developed a technique for predicting equipment failure by using anomaly detection and machine learning techniques. By using the developed prediction algorithm, accurate pre-diagnosis can reduce unnecessary maintenance costs, increase the stability and reliability of the system, and prevent system failures and accidents.

#### 4 Engineering Design cost prediction

This section dwells on the design cost prediction module, one of the five modules introduced above. The purpose of this study is to identify the correlation between information (Basic Drawing list, Equipment list, Line list, etc.) that can be used in the initial bidding stage of a project and actual design man-hour, and to create a prediction model.

The collected drawing list was reclassified into 27 design activities using SMEs, and Unit man-hour was calculated for each activity. Then, the design man-hour trend was analysed according to the project country, duration, and scale. User can calculate the general design man-hour by inputting the required quantity based on the 27 design activity lists. The developed correction factor can be used to increase the accuracy of the estimate according to the characteristics of the user project. In addition, a similar project recommendation function using the CBR technique was implemented. The step-by-step flow is summarized in Figure 1 below, and each of these steps is described in detail in the following sections.

Development of AI-based engineering Big Data integrated analysis system for decision-making support in the Engineering-Procurement-Construction (EPC) industry

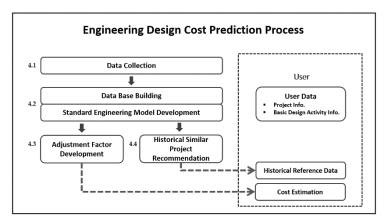


Fig. 1: Engineering design man-hour prediction algorithm development flow

#### 4.1 Data Collection

Many historical plant project data have been collected.

This section introduces historical plant project data collected from project participants. In particular, this study focused on collecting and analysing data related to piping design, which takes up 40-50% man-hour of the plant's overall design work volume and 25-34% of the total construction cost of the plant. Data from 40 projects comprising drawing list, equipment list, and project information were collected. (Figure

	A Project												
	WBS		Drawing	lory	was was	Completion Percentage of Stage(Plan)							
공종	WBSCODE	DWG No.	DWG Title	g g	Manhour	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6		
						START (0%)	STUDY (5%)	DRAFT (10%)	IFR (30%)	IFC (55%)			
PR	A0EA0PR0020501	U00-K-0021-RP	WASTE DISPOSAL REPORT	Α	71	2012-11-01	2012-11-06	2012-12-24	2012-12-31	2013-02-28	2013-04-16		
						START (0%)	STUDY (5%)	DRAFT (10%)	IFR (10%)	IFD (15%)	AFD (10%)		
PR	A0EA0PR0020601	U00-K-0026-SA	OPERATING WINDOWS AND SAFE OPERATING ENVELOPS	Α	72	2012-11-30	2012-12-05	2013-10-04	2013-10-22	2013-12-20	2014-01-31		
						START (0%)	STUDY (5%)	DRAFT (10%)	IFR (30%)	IFC (55%)			
PR	A0EA0PR0020801	A0EA0PR0020801	OPTIMIZATION STUDIES & EVALUATION REPORT	R	70	2013-08-01	2013-08-15	2013-10-31	2013-11-15	2014-01-16			
						START (0%)	IFI (100%)						
PR	A0EA0PR0020901	U00-K-0025-PR	VENDOR INFORMATION ON PACKAGE UNITS	I	67	2013-07-01	2013-09-30						
						START (0%)	STUDY (5%)	DRAFT (10%)	IFR (30%)	IFC (55%)			
PR	A0EA0PR0021001	U00-K-0023-PD	ALARM & TRIP VALUES	R	374	2013-07-01	2013-07-09	2013-09-25	2013-10-07	2013-12-31			
						START (0%)	STUDY (5%)	DRAFT (10%)	IFR (30%)	IFC (55%)			
	A0EA0PR0021101	U00-K-0019-PR	CATALYST AND CHEMICALS HANDLING AND LOADING PROCEDURES	R	70	2013-09-02	2013-09-11	2013-12-18	2013-12-31	2014-02-20			

2,3 below).

Fig. 2: Drawing List Sample

#### **INFORMATION & VENDOR PRINT STATUS**

					i				D	ATASHEET	
UNIT ITEM I	ITEM NO. 1		ITEM NO.	2	DESCRIPTION	EQUIPMENT TYPE	QTY	REV.	RECEIVED	DIMENSION	FOR NOZZLE
								District Co.	200000000000000000000000000000000000000	YES	NO
Unit A	C-74101	С	74101		RECYCLE COMPRESSOR	COMPRESSOR	1	В0	2008.01.30		0
Unit A	C-74102A/B	С	74102	A/B	MAKE-UP COMPRESSOR	COMPRESSOR	2	B1	2009.06.01		0
Unit A	C-74201	С	74201		RECYCLE COMPRESSOR	COMPRESSOR	1	B0	2008.01.30		0
Unit A	C-74202	С	74202		MAKE-UP COMPRESSOR	COMPRESSOR	1	B1	2009.06.01		0
Unit A	C-74401	С	74401		RECIRCULATION COMPRESSOR	COMPRESSOR	1	B1	2009.06.01		0
Unit A	DS-74301	DS	74301		MP STEAM DESUPERHEATER	MISCELLANEOUS	1	B1	2010.01.22		0
Unit A	DS-74302	DS	74302		STRIPPER DESUPREHEATER	MISCELLANEOUS	1	B1	2010.01.22		0
Unit A	DS-74303	DS	74303		LP STEAM DESUPERHEATER	MISCELLANEOUS	1	B1	2010.01.22		0

Fig. 3: Equipment List Sample

# 4.2 Standard Master Drawing Register for piping development and data standardization

#### 4.2.1 Standard Master Drawing Register (SMDR)

<u>Standard Master Drawing Register (SMDR) was developed and the collected data was standardized.</u>

Drawing lists of the collected projects were reorganized to the same standards in collaboration with plant engineers with more than 20 years of experience. And database was built (Table 2 below). Based on the unit man-hours of drawing list of each project, the average value for each of the 27 standardized design activities was calculated as the standard unit man-hour. Table 2 below shows 27 design activity lists and unit man-hours.

Table 2: SMDR (Standard Master Drawing Register)

No.	Design Activity List	Unit Man-hour(hours)
1	Line list & Tie in List	123
2	Piping Information drawing	215
3	Vendor Print Review & Data sheet	499
4	UFD & PNID Review	91
5	Plot Plan & Equipment Arrangement drawing	91
6	Utility Station & Safety Shower Location Plan	18
7	Steam Tracing Diagram	28
8	Tracing Circuit drawing	4
9	Stress Geometry drawing	4
10	Pulsation Study Diagram	45
11	B/M Sketch drawing & MTO with key Punch	2
12	Key plan drawing.	17
13	Standard & Special Support Detail drawing	4
14	Stress Geometry drawing	4
15	Stress analysis_Specialty item data sheet for stress	5
16	Stress analysis_Piping Flexibility analysis_Critical	239
17	Stress analysis_Piping Flexibility analysis_Non Critical	162
18	Stress analysis_AIV&FIV analysis & report	97
19	Conceptual Piping Design_Process Area	76
20	Conceptual Piping Design_P/R & Off site area	60
21	Conceptual Piping Design_U/G area	23
22	PDMS Modeling(Equipment)	55
23	PDMS Modeling(Layout)	207

Development of AI-based engineering Big Data integrated analysis system for decision-making support in the Engineering-Procurement-Construction (EPC) industry

24	PDMS Model Review	80
25	ISO drawing & BM Generation	2
26	3D Piping plan drawing Generation	13
27	Piping Engineering follow up(INCL. AS Built)	450

#### 4.2.2 Data Standardization

The collected data was standardized in the same form.

For the convenience of comparison and analysis, all project data were converted as follows. The period was calculated on a monthly basis and the project areas were classified into Korea, Middle East Asia, and Southeast Asia. Project scales (project cost) were converted to Dollar in 2019 using the CPI(Consumer Price Index). In the case of the equipment list, information on the quantity of equipment considered to affect the design time was calculated. The Fig 4 below show the standardized data set form.

	Plant Engine	erin	g F	Proj	ect	Da	ıtab	ase	)		
Project Number			1			2			3		
	Project Name	l	Project /	A		Project I	В		Project	c	
	•		•			•			-		
	Onshore / Off shore		Onshore	•		Onshore	•		Onshore	2	
	Plant Type		refinery			refinery			Chemica	al	
	Contractor	DI	EPC Com	pany	D	EPC Com	pany	DI	EPC Com	pany	
	Owner	s o	wner Con	npany	FO	wner Con	nnany	F O	wner Con	nnany	
	Country		Korea		Middl	e East Asi	a (Iran)	Middl	e East As	a (Iran)	
	Project Duration (Months)	l	27			38			16		
	Project Scale (Dollar)		2076812	10		1517565	6		590191	9	
	Equipment Quantity		163			63			21		
	Piping Line Quantity		11110			666		473			
No.	Standard Drawing List for Piping	Unit MH	Q'ty	Total MH	Unit MH	Q'ty	Total MH	Unit MH	Q'ty	Total Mi	
1	Line list & Tie in List	168	1	168	55	1	55	23	1	23	
2	Piping Information drawing	337	1	337	109	1	109	45	1	45	
3	Vendor Print Review & Comments	1,010	1	1,010	328	1	328	136	1	136	
4	UFD & PNID Review	100	1	100	100	1	100	100	1	100	
5	Plot Plan & Equipment Arrg't drawing	95	7	665	95	2	190	95	2	190	
6	Utility Station & Safety Shower Loaction Plan	20	3	50	19	2	38	19	1	19	
7	Steam Tracing Diagram	30	2	50	29	1	29	29	1	29	
8	Tracing Curcuit Drawing	2	90	180	2	91	181	2	220	524	
9	Stress Geometry drawing	4	280	1,067	4	420	1,596	4	220	836	
10	Pulsation Study Diagram	40	23	900	48	20	950	48	20	950	
11	B/M Sketch dwg. & MTO with key Punch	1	4,500	5,130	1	2,000	2,000	1	225	257	
12	Key plan dwg.(2D New & Exist)	40	1	40	10	1	10	10	1	10	
13	Standard & Special Support Detail drawing	4	30	120	4	30	120	4	30	120	
14	Stress Geometry drawing	4	300	1,200	4	300	1,200	4	300	1,200	
15	Stress analysis _ Specialty item data sheet for stress	4	35	140	5	30	143	5	30	143	
16	Stress analysis_ Piping Flexbility analysis_Critical	23	210	4,828	20	225	4,500	20	225	4,500	
17	Stress analysis_ Piping Flexbility analysis_Non Critical	19	185	3,480	15	200	3,000	15	200	3,000	
18	Strss analysis_AIV&FIV anaysis & report	100	1	100	100	1	100	100	1	100	
19	Conceptual Piping Design_Preocess Area Conceptual Piping Design P/R & Off site area	81	35	2,826	81	6	485	81	2	162	
20		52 24	25	1,306	52	3	157	52	1	52	
21			1	24	24	1	24	24	1	24	
22	PDMS Modeling(Equipment)	42	19	820	36	1	36	35	1	35	
23	23 PDMS Modeling(Layout) 24 PDMS Model Review			6.347	214	20	0.105	245		1 200	
25	ISO Dwg. & BM Generation	25 150	257	150	150	29	6,195 150	215 150	1	1,290	
26	3D Piping plan drawing Geneation	250	1	250	250	1	250	250	1	250	
26	Piping Engineering follow up(INCL, AS Built)	400	1	400	400	1	400	400	1	400	
21	Total	3.025	10	31,688	2,154	10	22,345	1.866	8	14,543	

Fig. 4: Standardized data set example

# 4.3 Adjustment Factor Development

Adjustment factors were developed to reflect trends in design man-hour according to project characteristic information. After setting the project characteristic classification criteria (Table 3) through the SME(Subject Matter Expert)s survey technique, the factor was calculated by comparing relative increments and reductions.

Currently, the country factor, period factor, and scale factor have been developed, and detailed development methods are as follow. User can apply by multiplying each factor according to the user's project condition by the standard de-sign man hour value calculated using the SMDR of 4.2section

Table 3: Project Characteristic Classification Criteria

Adjustment Factors	Classification				
Country	Korea (Standard)	Korea			
	Middle East Asia	Malaysia			
	Southeast Asia	Saudi Arabia Iran			
Duration	0 - 12 mo	nths			
	13 - 24 months (Standard)				
	25 - 36 mc	onths			
	37 - 48 mc	onths			
Scale	Less than \$1	million			
	More than \$1 million and less than \$4 million (Standard)				
	More than \$4 million and less than \$6 million				
	More than \$6 million				

#### 4.3.1 Country Factor

The country factor was developed to reflect the design man-hour change by country condition. An engineering man-hour prediction regression model was developed for 20 projects conducted in Korea. After that, the characteristic information variables of other country project were inputted into the Korean project regression prediction model, and the increase/decrease rate was regarded as a difference according to the country of the project. The regression model for Korea Project results and developed factors are shown in Table 4 below.

Table 4: Summary of Korea Project Regression Model and Country Factor

Regressio	on Model for Kor	ea Project		Country Factor	
MAPE(%)	R-squared	P-value	Korea	Middle East Asia	Southeast Asia
11.41	0.94	$1.042 \times 10^{-8}$	1.0	1.20	0.83

#### 4.3.2 Duration Factor

Similar to the above method, we created a regression model that predicts man-hours for projects ranging from 13 to 24 months. And design man-hour was predicted by

inputting variables of different period groups. Like the Country Factor, the difference between the predicted and actual values was considered to be due to the duration of the project. The regression result and developed factors are summarized in the table5 below.

Table 5: Summary of Normal Duration Project Regression Model and Duration Factor

Regression	Model Norm Project	al Duration		Duration	n Factor	
MAPE(%)	R-squared	P-value	0-12 months	13-24 months	25-36 months	37-48 months
19.40	0.87	$2.867 \times 10^{-5}$	0.93	1.0	1.13	1.35

#### 4.3.3 Scale Factor

Scale factors have also been developed similar to the above method. After generating a man-hour prediction regression model with a project of the general scale, the design man-hour was estimated using small-scale and large-scale data as in-put variables. The difference between the design man-hour value and the actual value of the two project scale groups predicted by the model was considered to be due to the tendency to follow the group. The regression model result and developed factors are shown in Table6 below.

Table 6: Summary of Normal Scale Project Regression Model and Scale Factor

Regression	Model for No Project	ormal Scale		Scale 1	Factor	
MAPE(%)	R-squared	P-value	Less than \$ 1 million	More than \$ 1 million and less than \$ 4 million	More than \$ 4 million and less than \$ 6 million	More than \$ 6 million
12.87	0.90	$7.189 \times 10^{-6}$	0.83	1.20	1.17	1.37

## 4.4 Historical Similar Project Recommendation

In addition, similar project suggestion function using CBR method has been developed. CBR method is a technique that finds the most similar cases among past cases and uses that for problem solving. In this study, the priority of each factor in judging project similarity was investigated through SME questionnaire. These factors refer to the standardized project data set introduced in section 4.2. The survey results of SME were used to establish similar project selection criteria and the criteria in-formation is shown in Table 7 below.

SME	Result #1	Result #2	Result #3	Result #4	Result #5	Weight Factor for each characteristic (Average Value of distributed points)
On shore/ Off shore	35	25	30	30	30	30
Project Country	20	25	20	20	15	20
Plant Type	20	20	25	15	20	20
Project Duration	15	10	10	5	10	10
Scale Total	5	10	10	15	10	10
Equipment Ouantity	5	10	5	15	15	10

Table 7: Similar Project Selection Criteria and Survey Result

When the user enters a project information variable, the similarity score is calculated by the designed similar project selection criteria. The three projects with the highest scores are selected and provided with basic information such as plant type, duration and scale, and details such as plumbing drawings and man-hour information.

#### 5 CONCLUSION

In this study, a man-hour prediction algorithm and tool for plant piping design was developed. The development process is as follows.

- Step 1. Data collection: About 40 historical plant project data were collected.
- Step 2. Develop a standard pipe design model and build a database: A database was built by standardizing the collected plant project data in a consistent format. The user can estimate the typical design man-hour by simply entering the required quantity.
- Step 3. Development of correction factors: Country, Duration, and Scale Factors were developed using a machine learning algorithm.

By applying the coefficient to the general design man-hour value calculated in Step 2, we can obtain the design man-hour value that reflects the difference according to the project characteristics.

Step 4. Similar project recommendation function: select similar projects with the user's project based on project characteristic information and present as a reference.

This study is meaningful as it collects historical data and applies statistical analysis and machine learning algorithms in quantitative terms. Using the developed algorithm, project performers, such as EPC Company or Engineering Company, can relatively easily calculate design man-hours using limited information at the beginning of the bid.

#### 6 LIMITATION AND FUTURE WORKS

As mentioned above, the scope of this study is basically to predict only the design man-hours corresponding to the piping discipline, and the other disciplines are replaced Development of AI-based engineering Big Data integrated analysis system for decision-making support in the Engineering-Procurement-Construction (EPC) industry

by the method using the ratio of design time between the discipline of commercial data. However, the correlation between project factors (period, etc.) and design time may vary depending on the type of work. Therefore, data on other types of work will be collected in the future, and studies should be conducted to find out the correlation between project types and overall design man-hours as well as each type of work. If data can be collected on a project basis, design man-hour trends by plant type and a customer can also be investigated. Existing regression and correction factors can also be improved.

#### 7 FUNDING ACKNOWLEDGEMENT

The authors acknowledge that this research was sponsored by the Korea Ministry of Trade Industry and Energy (MOTIE/KEIT) through the Technology Innovation Program funding by KEIT for (1) Artificial Intelligence Big-data (AI-BD) Platform for Engineering Decision-support Systems (grant number = 20002806) and (2) Intelligent Project Management Information Systems (i-PMIS) for Engineering Projects (grant number=10077606)...

#### 8 REFERENCE

- James, D. F. (2005) AACE International Transactions : A Methodology for Estimating Engineering Details.
- Marzouk, M., Elkadi. M.(2016) Estimating water treatment plants cost using factor analysis and artificial neural networks. Journal of Cleaner Production 112(2016). http://doi.org/10.1016/j.jclepro.2015.09.015
- Sphurti, S., Nagarai, V. (2017) Cost estimation of civil construction projects using machine learning paradigm. International Conference on I-SMAC 2016, pp. 1–5. IEEE (2017). http://doi.org/10.1109/I-SMAC.2017.8058249
- Pesko, I., Muscenski. V.(2017) Estimation of Costs and Durations of Construction of Urban Roads Using ANN and SMV. pp.13 (2017). http://doi.org/10.1155/2017/2450370