



Conference Article

Man-Hour Estimation Model Development in BIM-Based Projects with Machine Learning-Regression Algorithms

Nevsat Irem Yurttakal^{1*}

¹ ENKA Insaat, Tasarım Merkezi, Orcid ID: <https://orcid.org/0009-0000-9156-4490>,
irem.yurttakal@enka.com,

(First received September 18, 2023 and in final form December 22, 2023)

**3rd International Conference on Design, Research and Development
(RDCONF 2023)
December 13 - 15, 2023**

Reference: Yurttakal, N., İ. Man-Hour Estimation Model Development in BIM-Based Projects with Machine Learning-Regression Algorithms. *Orclever Proceedings of Research and Development*,3(1), 207-220.

Abstract

Construction management is a service that uses project management techniques to planning, design and construction of a project. Another branch of construction management is the planning made during the bidding stages before the construction starts. At this stage, the scheduling of the work, mobilization and most importantly the project design and drawing time should be analyzed and accurate estimates should be made. For this estimation, the experience of the bidding organization is the biggest reference. Construction management techniques in a BIM-based project are challengeable. The biggest reason is that knowledge about BIM is much less than the others and still a developing system. Therefore, the margin of error in the estimated project duration to be predicted in the bids to be made for such projects increases. Considering that man-hour estimations are financially very important for the company. This study is aimed to make observations with advanced technologies which is applied ML algorithms and analysis methods for man-hour estimations in BIM projects at the bidding stage. The project information to perform a man-hour analysis of completed construction projects was requested from offices that have gained experience in BIM projects. The project information of 16 projects were tabulated. The collected information was converted into a data set and transferred into regression analysis algorithms



using Python format. Data set, which was turned into a matrix, was analysed and the consistency of the parameters in data set was tested. Spent man-hour estimation was made, graphed over parameters with simple and multiple linear regression algorithms.

Keywords: Building Information Modelling (BIM), construction management, bidding, project management, Artificial Intelligence (AI), machine learning, regression analysis

1. Introduction

Construction management is a specialized service that uses project management techniques to oversee the planning, design and construction of a project from start to finish. Project management, which encompasses multifarious procedures, disciplines, and teams, has been widely adopted in the architecture, engineering, construction, and operation (AECO) industry to organize building production [1]. Another branch of construction management is the planning made during the bidding and tendering stages before the construction starts. At this stage, the scheduling of the work, mobilization, duration of the productions to be made, material supply and most importantly the project design and drawing time should be well analyzed and accurate estimates should be made. For this estimation, the experience of the bidding organization is the biggest reference. Manpower forecasting is an essential and important strategic managerial practice for government and business organizations [2].

Construction management techniques in a BIM-based project differ. With fast development of modern science and technology, informatization level of all industries is improving continuously and building industry, a kind of traditional industry applies increasingly informatization method to reinforce management and improve economic performance. BIM technique, a new carrier of project information, runs through all stages of the construction process throughout the life cycle of the project, which renders all project participants to obtain proper information at a due time as necessary, assists project understanding, reduces common mistakes and supports the project-related decision [3, p. 302]. Practical significance of BIM is to apply a kind of data to multiple purposes, and project basic data made by BIM can be used as the basis for investment assessment, bill of quantities, bidding, contract signing as well as estimation and audit of object budget. Different from the traditional 3D model, BIM model also integrated time and cost information inside which can be called 5D model [3, p. 304]. Developing traditional project management processes is a challenge at some points. The biggest reason for this is that the knowledge about BIM projects is much less than the traditional method and it



is still a developing system. Despite BIM technique has many applications, practical application is few and without good entry point when applied to the project. BIM technique is currently not mature with some difficulties in practice, and in addition, project personnel is not familiar with BIM technique, all contributing to difficult field application of BIM technique [3, p. 310].

For these reasons, there are different working techniques and these techniques are still developing, the margin of error in the estimated project duration can be misleading. Considering that man-hour and project cost estimations are financially very important for the company. This study is aimed to make observations with advanced technologies and different analysis methods for man-hour estimations in BIM projects at the bidding stage.

2. Materials and Methods

In this study, it is planned to use applied machine learning algorithms to analyze the man hour estimations of BIM projects at the bidding stage.

Machine learning (ML) is the subset of artificial intelligence (AI) that focuses on building systems that learn—or improve performance—based on the data they consume. AI is a broad term that refers to systems or machines that mimic human intelligence [4].

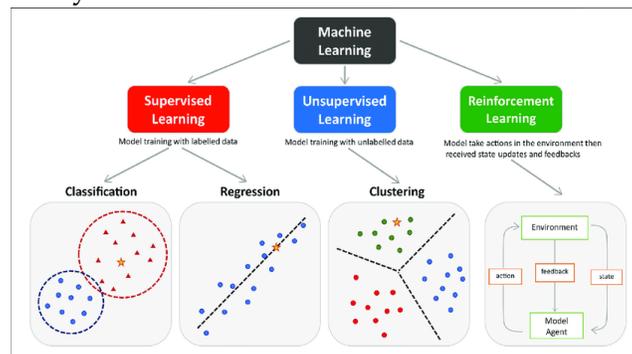


Figure 1 Types of machine learning

ML, a subset of AI, is also classified within itself and the method used in this study will be regression analysis in the Supervised Learning category [Fig.1].

In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome' or 'response' variable, or a 'label' in machine learning parlance) and one or more independent variables (often called 'predictors', 'covariates', 'explanatory variables' or 'features'). The most common form of regression analysis is linear regression, in which one finds the line (or a



more complex linear combination) that most closely fits the data according to a specific mathematical criterion. The types of regression analysis can shown as Table 1.

Table 1: Types of Regression Analysis Techniques

Types of Regression Analysis Techniques
Lineer Regression
Logistic Regression
Polynomial Regression
Quantile Regression
Ridge Regression
Lasso Regression
Elastic Net Regression
Principal Components Regression (PCR)
Partial Least Squares (PLS) Regression
Support Vector Regression
Ordinal Regression
Poisson Regression
Negative Binomial Regression
Quasi Poisson Regression
Cox Regression
Tobit Regression

Not all types of analysis will be included in this study. To summarize the study, in the first section, the information to perform man-hour analysis with completed construction projects was requested from offices and individuals who have gained experience in BIM projects. As a result of a research, the project information of 16 projects were tabulated. In Section 2, the collected information was converted into a data set and transferred into regression analysis algorithms using Jupyter notebook(Python format). The data set,



which was turned into a matrix within the algorithm, was analysed within itself and the consistency of the parameters in data set was tested. In section 3 of the study, spent man hour estimation was made and graphed over various parameters with the linear regression algorithm.

2.1. Collecting Project Information

In the first section of the study, in order to perform the targeted analyses, some information that should be used in the algorithm had to be collected by the target audience. The target audience was people who had previously worked in BIM projects, offices with expertise and experience in BIM project applications. In order to collect this information from the target group, an Excel spreadsheet was sent by e-mail to the people identified and they were asked to fill it in. The information requested from the experts is as follows;

- **Building Type:** This parameter represents the difficulty level of the project. The difficulty level of the project directly affects the number of man hours spent. There are different challenges that each type of project may pose, and it is relatively unclear to compare this. For this reason, experts were asked to rank the construction project types according to their difficulty levels, while the information of the project type was requested. The type of the project was not written directly into the data matrix, but the difficulty level was entered numerically. In this digitization action, the ranking method was performed.
- **Finish Year of Construction:** The end date of the project was requested from the experts in order to be able to deal with the foreign currency values of this period if a cost analysis was desired.
- **Project Cost:** This information was requested from the experts in order to answer the question of how much the project cost is consistent with the man-hours spent.
- **Total Construction Area:** The size of the project in square meters will be directly proportional to the man-hours to be spent, because considering that each element of the project is modelled in BIM projects, the more square meters, the more work to be done.
- **Number of Architectural Drawing Sheets:** As number of sheets increases in the project, the work to be drawn also increases, so man-hour will directly affect it.
- **Spent Man-Hour:** This parameter is the main parameter of the analysis.
- **Project Delivery Methods:** The project delivery method affects the processes in the project. It was aimed to observe the effect of this factor on the man-hour analysis.
- **Architectural Project Type:** Project type affects how detailed the project will be drawn.



- Construction System: The difficulty of the construction system of the project means more work and time consumption. In this parameter, Concrete-Steel and similar types given by the experts were turned into a numerical data by the ranking method according to their difficulty levels.
- BIM Level: As the BIM level increases in the BIM project, the layers of the project and the work to be modelled increase.
- LOD Level: As the LOD level increases, in some project's elements modeled from scratch can cause a huge man-hour spent.
- Interdisciplinary Coordination: Interdisciplinary coordination in BIM requires different working principles compared to the traditional project drawing process, it can negatively affect the man-hours spent.

As a result of the interviews with experts, information on 16 projects was collected as shown in Figure 2.

GENEL PROJE BİLGİLERİ											BEP (BIM EXECUTION PLAN) BİLGİLERİ		
PROJE ADI	KULLANIM AMACI (PROJE TİPİ)	İNŞA YILLARI	PROJE MALİYETİ - RANKIN (T/US D/EUR)*	TOPLAM İNŞAAT ALANI (m2)	MİMARİ PAFTA SAYISI	HARCANA N ADAM SAAT***	PROJE TESLİM YÖNTEMİ (DELIVERY METHODS)	MİMARİ PROJE TÜRÜ (AVAN PROJE, UYGULAMA PROJESİ VS.)	YAPIM SİSTEMİ (BETONARME, ÇELİK, PREFABRİK, KARMA)	(A)BİM LEVELS /BİM DİMENSİYONLARI	(B)LOD SEVİYESİ	(C)DISİPLİNLER ARASI KOORDİNASYON	
1 A PROJECT	Alışveriş merkezi, Aquapark, Spa, Otel	2021-2022	-	60000	390	9720	Design Bid Build	(C&S)	Betonarme	LEVEL 2	LOD 350	EVET	
2 B PROJECT	Havallımanı	2011-2015	-	168275	400	300000	Design Bid Build	Uygulama projesi (C&S)	Karma	LEVEL 2	LOD 300	EVET	
3 C PROJECT	Yönetim Binası	2013	10	2750	126	10000	Design Bid Build	Uygulama projesi	Betonarme	LEVEL 1	LOD 300	HAYIR	
4 D PROJECT	Yönetim Binası	2013	12	1500	123	16000	Design Bid Build	Uygulama projesi	Betonarme	LEVEL 1	LOD 300	HAYIR	
5 E PROJECT	AVM/Ofis (karma)	2015	1	15000	909	180000	Design Bid Build	Uygulama projesi	Karma	LEVEL 2	LOD 300	EVET	
6 F PROJECT	AVM	2015-2018	4	190000	579	150000	Design Bid Build	Uygulama projesi	Karma	LEVEL 2	LOD 300	EVET	
7 G PROJECT	Ofis	2006-2009	5	109000	476	41000	Design Bid Build	Uygulama projesi	Karma	LEVEL 2	LOD 300	EVET	
8 H PROJECT	Ofis	2008-2010	5	190000	314	35000	Design Bid Build	Uygulama projesi	Karma	LEVEL 2	LOD 300	EVET	
9 I PROJECT	AVM	2016	-	130000	400	280000	Design Bid Build	Uygulama projesi	Karma	LEVEL 2	LOD 300	EVET	
10 J PROJECT	AVM	2014-2016	2	159000	515	90000	Design Bid Build	Uygulama projesi	Karma	LEVEL 2	LOD 300	EVET	
11 K PROJECT	AVM	2016	3	200000	410	33000	Design Bid Build	Uygulama projesi	Karma	LEVEL 2	LOD 300	EVET	
12 L PROJECT	Konut	2017	0	100000	295	20000	Design Bid Build	Uygulama projesi	Betonarme	LEVEL 2	LOD 300	EVET	
13 M PROJECT	Kreş	2018	11	3000	110	5800	Design Bid Build	Uygulama projesi	Betonarme	LEVEL 2	LOD 200	EVET	
14 N PROJECT	Stad	2019	5	140000	-	70000	Design Bid Build	Uygulama projesi	Karma	LEVEL 2	LOD 300	EVET	
15 O PROJECT	Antreman Kompleksi	2018	9	16000	-	35000	Design Bid Build	Uygulama projesi	Karma	LEVEL 2	LOD 300	EVET	
16 P PROJECT	Otel	2017	8	17000	-	40000	Design Bid Build	Uygulama projesi	Betonarme	LEVEL 2	LOD 300	EVET	

Figure 2: Excel file of the collected data before normalizations

Data Set

In the second stage, normalization processes were carried out in order to make the information obtained processable. The answers of the parameters with verbal answers were converted into numerical values with the ranking method and as a result, the matrix in Fig. 3 was created. The created dataset was introduced to the algorithm to be analyzed as a cvs. file.



BUILDING OF TYPE	FINISH YEAR	PROJECT COST-RANKING	TOTAL CONSTRUCTION AREA	NUMBER OF ARCHITECTURAL DRAWING SHEETS	SPENT MAN-HOUR	PROJECT DELIVERY METHODS	ARCHITECTURAL PROJECT TYPE	CONSTRUCTION SYSTEM	BIM LEVEL	LOD LEVEL	INTERDISCIPLINARY COORDINATION
7	2022	1	60000	390	9720	1	1	1	2	3	2
9	2015	16	168275	400	300000	1	1	2	2	2	2
3	2013	6	2750	126	10000	1	2	1	1	2	1
3	2013	4	1500	123	16000	1	2	1	1	1	1
7	2015	15	260000	909	160000	2	2	2	2	2	2
7	2018	12	190000	579	150000	1	2	2	2	2	2
2	2009	11	109000	476	41000	1	2	2	2	2	2
2	2010	11	190000	314	35000	1	2	2	2	2	2
7	2016	2	130000	400	260000	2	2	2	2	1	2
7	2016	14	195000	515	90000	1	2	2	2	2	2
7	2016	13	200000	410	33000	1	2	2	2	2	2
4	2017	3	100000	295	20000	1	2	1	2	2	2
5	2018	5	3000	110	5800	1	2	1	2	1	2
8	2019	11	140000	400	70000	1	2	2	2	2	2
1	2018	7	16000	200	35000	2	2	2	2	2	2
6	2017	8	17000	200	40000	2	2	1	2	2	2

Figure 3 Matrix of Dataset

For the first stage of the analysis, it was aimed to analyze the consistency of the parameters determined in the study. As a first step, a python script was created on the Jupyter notebook (Fig. 4) and the created dataset was introduced to the script.

```
In [1]: from warnings import filterwarnings
filterwarnings('ignore')
```

```
In [2]: import pandas as pd
ad = pd.read_csv("Man-hour Analysis-data.csv")
df = ad.copy()
df.head()
```

```
Out[2]:
```

	BUILDING TYPE	FINISH YEAR OF CONSTRUCTION	PROJECT COST-RANKING	TOTAL CONSTRUCTION AREA	NUMBER OF ARCHITECTURAL DRAWING SHEETS	SPENT MAN-HOUR	PROJECT DELIVERY METHODS	ARCHITECTURAL PROJECT TYPE	CONSTRUCTION SYSTEM	BIM LEVEL	LOD LEVEL	INTERDISCIPLINARY COORDINATION
0	7	2022	1	60000	390	9720	1	1	1	2	3	
1	9	2015	16	168275	400	300000	1	1	2	2	2	
2	3	2013	6	2750	126	10000	1	2	1	1	2	
3	3	2013	4	1500	123	16000	1	2	1	1	1	
4	7	2015	15	260000	909	160000	2	2	2	2	2	

Figure 4 Python Script File

With the Describe() method, the statistical values of the data we have are shown and the number, standard deviation, minimum value, max value of the parameters are analysed. As can be seen in Fig. 3, when the parameters of 16 projects are analysed, the oldest project in the project string belongs to 2009 and the newest project belongs to 2022. The standard deviation of the project delivery method and the next parameters is low, so it is difficult to make a prediction or comment on these parameters because the values of all of them are the same or very close to each other in many projects and there is very little diversity in the data.



```
In [4]: df.describe().T
```

```
Out[4]:
```

	count	mean	std	min	25%	50%	75%	max
BUILDING TYPE	16.0	5.3125	2.468974	1.0	3.00	6.5	7.00	9.0
FINISH YEAR OF CONSTRUCTION	16.0	2015.7500	3.296463	2009.0	2014.50	2016.0	2018.00	2022.0
PROJECT COST-RANKING	16.0	8.6875	4.826576	1.0	4.75	9.5	12.25	16.0
TOTAL CONSTRUCTION AREA	16.0	111407.8125	85673.440905	1500.0	16750.00	119500.0	190000.00	260000.0
NUMBER OF ARCHITECTURAL DRAWING SHEETS	16.0	365.4375	204.211645	110.0	200.00	395.0	426.50	909.0
SPENT MAN-HOUR	16.0	79720.0000	91184.671226	5800.0	19000.00	37500.0	105000.00	300000.0
PROJECT DELIVERY METHODS	16.0	1.2500	0.447214	1.0	1.00	1.0	1.25	2.0
ARCHITECTURAL PROJECT TYPE	16.0	1.8750	0.341565	1.0	2.00	2.0	2.00	2.0
CONSTRUCTION SYSTEM	16.0	1.6250	0.500000	1.0	1.00	2.0	2.00	2.0
BIM LEVEL	16.0	1.8750	0.341565	1.0	2.00	2.0	2.00	2.0
LOD LEVEL	16.0	1.8750	0.500000	1.0	2.00	2.0	2.00	3.0
INTERDISCIPLINARY COORDINATION	16.0	1.8750	0.341565	1.0	2.00	2.0	2.00	2.0

Figure 5

When the distribution of variables and the correlation between them are analyzed with the Corr() method, the results in Figure 6 are obtained.

```
In [6]: df.corr()
```

```
Out[6]:
```

	BUILDING TYPE	FINISH YEAR OF CONSTRUCTION	PROJECT COST-RANKING	TOTAL CONSTRUCTION AREA	NUMBER OF ARCHITECTURAL DRAWING SHEETS	SPENT MAN-HOUR	PROJECT DELIVERY METHODS	ARCHITECTURAL PROJECT TYPE	CONSTRUCTION SYSTEM	
BUILDING TYPE	1.000000	0.518089	0.327622	0.489637	0.478099	0.597230	-0.015094	-0.424910	0.209264	0.3
FINISH YEAR OF CONSTRUCTION	0.518089	1.000000	-0.298543	-0.137713	0.000272	-0.001375	0.135665	-0.325649	-0.222461	0.3
PROJECT COST-RANKING	0.327622	-0.298543	1.000000	0.717632	0.587785	0.381971	-0.084935	0.015164	0.694074	0.2
TOTAL CONSTRUCTION AREA	0.489637	-0.137713	0.717632	1.000000	0.855795	0.522614	-0.039378	-0.012437	0.753554	0.4
NUMBER OF ARCHITECTURAL DRAWING SHEETS	0.478099	0.000272	0.587785	0.855795	1.000000	0.490880	0.180489	-0.056510	0.619374	0.4
SPENT MAN-HOUR	0.597230	-0.001375	0.381971	0.522614	0.490880	1.000000	0.287926	-0.321673	0.550970	0.2
PROJECT DELIVERY METHODS	-0.015094	0.135665	-0.084935	-0.039378	0.180489	0.287926	1.000000	0.218218	0.149071	0.2
ARCHITECTURAL PROJECT TYPE	-0.424910	-0.325649	0.015164	-0.012437	-0.056510	-0.321673	0.218218	1.000000	0.097590	-0.1
CONSTRUCTION SYSTEM	0.209264	-0.222461	0.694074	0.753554	0.619374	0.550970	0.149071	0.097590	1.000000	0.4
BIM LEVEL	0.365621	0.325649	0.298235	0.497933	0.460563	0.285627	0.218218	-0.142857	0.487950	1.0
LOD LEVEL	0.141759	0.262908	0.203733	0.230823	0.318542	-0.164706	-0.149071	-0.487950	0.066667	0.2
INTERDISCIPLINARY COORDINATION	0.365621	0.325649	0.298235	0.497933	0.460563	0.285627	0.218218	-0.142857	0.487950	1.0

Figure 6

Figure 6 shows that there is a positive relationship between some variables and a negative relationship between others. Within these 16 projects, it can be interpreted that as the end date of the project increases, the square meters of the projects decreases. In the same way, we can also examine the correlation graphs of the two variables with the help of the seaborn library via Python. We can see the relationship between Building Type and Spent Man-Hour variables with the graph in Figure 7.



```
In [14]: sns.jointplot(x='BUILDING TYPE',y='SPENT MAN-HOUR',data=df,kind='reg')
Out[14]: <seaborn.axisgrid.JointGrid at 0x1d4f9fdb160>
```

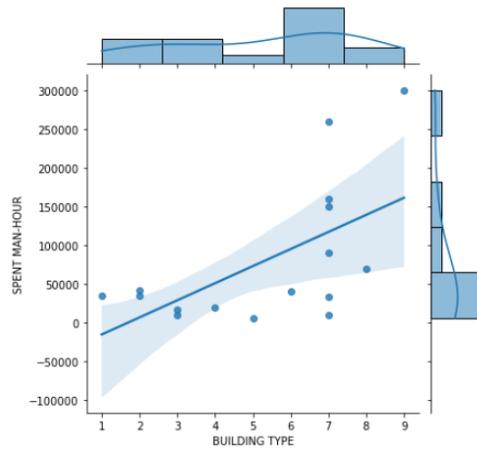


Figure 7

Since our main objective in the study is to analyze the man-hours spent, in Section 3 the analysis is continued on two variables: total construction area and spent man-hours.

Analysis

When we look at the correlation graph(Fig. 8) between Spent Man-hour and total construction area variables, we see that as the total square meters increases, the man-hour spent also increases.

```
In [10]: sns.jointplot(x='TOTAL CONSTRUCTION AREA',y='SPENT MAN-HOUR',data=df,kind='reg')
Out[10]: <seaborn.axisgrid.JointGrid at 0x26d64fc8d60>
```

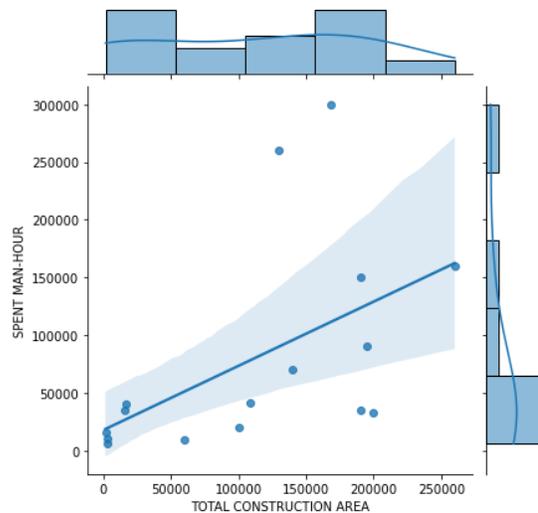


Figure 8



Simple Linear Regression Analysis

When the linear regression model was created with the `statmodels` library in the regression algorithm, the results were as shown in Figure 9. As seen in the model, it is seen that the R-square value is 20%. This value indicates that 20% of the change in man hour value can be explained by the change in total construction area.

Out[40]: OLS Regression Results

Dep. Variable:	SPENT MAN-HOUR	R-squared:	0.273				
Model:	OLS	Adj. R-squared:	0.221				
Method:	Least Squares	F-statistic:	5.261				
Date:	Mon, 11 Jul 2022	Prob (F-statistic):	0.0378				
Time:	16:19:57	Log-Likelihood:	-202.36				
No. Observations:	16	AIC:	408.7				
Df Residuals:	14	BIC:	410.3				
Df Model:	1						
Covariance Type:	nonrobust						
	coef	std err	t	P> t	[0.025	0.975]	
	const	1.775e+04	3.37e+04	0.527	0.606	-5.45e+04	9e+04
	TOTAL CONSTRUCTION AREA	0.5562	0.243	2.294	0.038	0.036	1.076
Omnibus:	10.225	Durbin-Watson:	2.396				
Prob(Omnibus):	0.006	Jarque-Bera (JB):	6.729				
Skew:	1.425	Prob(JB):	0.0346				
Kurtosis:	4.404	Cond. No.	2.33e+05				

Figure 9

When we estimate with the `Print` function, we can formulate (Equation 1) the relationship between spent man-hour and total construction area as follows. The output of the function is as follows.(Equation 2)

```
“ print("SPENT MAN-HOUR = " + str("%.2f" % model.params[0]) + " + TOTAL
CONSTRUCTION AREA" + "*" + str("%.2f" % model.params[1]))“
```

(Equation 1)

$$\text{SPENT MAN-HOUR} = 17751.33 + \text{TOTAL CONSTRUCTION AREA} * 0.56$$

Equation 2

When we graph this equation, it is as in Figure 10.

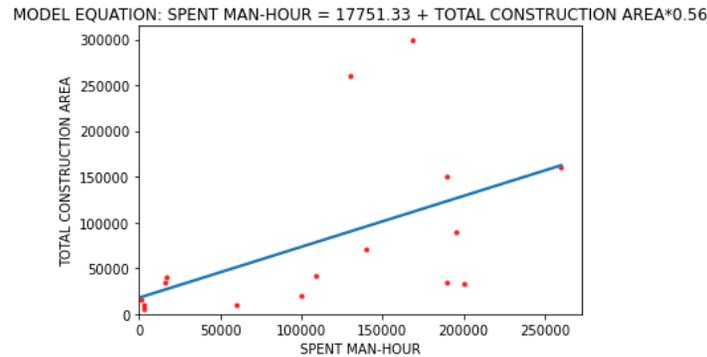


Figure 10

Multiple Linear Regression

In this section, it is assumed that 20% of the data is tested with machine learning logic and a prediction is made after the necessary change analysis is done. When we proceed through an example prediction, let's assume that the model equation takes the following scenario A as a reference.

Company A will bid for the tender of a shopping center project, which is planned to be completed in 2025, where architectural modeling and interdisciplinary coordination is required to be done at Level 2 in BIM. It is stated in the BEP that the average LOD level in the project is 300. According to the ranking of the project group in the cost dataset allocated for the project, it corresponds to the 7th place. The total square meter area of the project is 200,000 m² and the application project of the plan project will be drawn in a mixed structure system. There will be a total of 450 sheets targeted to be delivered. Contractual requirements should be taken into consideration for the design-build project. The following estimation model has been created for the estimation of the project in this scenario.



OLS Regression Results

Dep. Variable:	SPENT MAN-HOUR	R-squared:	0.963
Model:	OLS	Adj. R-squared:	0.796
Method:	Least Squares	F-statistic:	5.757
Date:	Sun, 17 Jul 2022	Prob (F-statistic):	0.157
Time:	17:34:43	Log-Likelihood:	-131.44
No. Observations:	12	AIC:	282.9
Df Residuals:	2	BIC:	287.7
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
BUILDING TYPE	1.011e+04	3.86e+04	0.262	0.818	-1.56e+05	1.76e+05
FINISH YEAR OF CONSTRUCTION	-2128.7788	2.59e+04	-0.082	0.942	-1.14e+05	1.09e+05
PROJECT COST-RANKING	-1.087e+04	6107.448	-1.780	0.217	-3.72e+04	1.54e+04
TOTAL CONSTRUCTION AREA	0.0271	0.347	0.078	0.945	-1.468	1.522
NUMBER OF ARCHITECTURAL DRAWING SHEETS	113.4444	117.042	0.969	0.435	-390.147	617.036
PROJECT DELIVERY METHODS	6.732e+04	5.75e+04	1.171	0.362	-1.8e+05	3.15e+05
ARCHITECTURAL PROJECT TYPE	2.097e+06	2.61e+07	0.080	0.943	-1.1e+08	1.14e+08
CONSTRUCTION SYSTEM	9.4e+04	1.02e+05	0.920	0.455	-3.46e+05	5.34e+05
BIM LEVEL	-1.43e+04	3.02e+04	-0.473	0.683	-1.44e+05	1.16e+05
LOD LEVEL	-1.242e+04	3.37e+04	-0.369	0.748	-1.57e+05	1.32e+05
INTERDISCIPLINARY COORDINATION	-1.43e+04	3.02e+04	-0.473	0.683	-1.44e+05	1.16e+05

Figure 11

As can be seen in Figure 11, the coefficients of influence of the other variables for the man hours spent variable are observed and the consistency of the factors on the constant variable is 96.3%. In order to perform the machine learning estimation, a formulation in Figure 12 was performed and as a result of this formula, the recommended duration for scenario 1 was 174.966 hours.

```
In [26]: yeni_veri = [ [7], [2025], [7], [200000], [450], [2], [2], [2], [2], [2], [2] ]
        yeni_veri = pd.DataFrame(yeni_veri).T

In [27]: model.predict(yeni_veri)

Out[27]: array([174966.96160028])
```

Figure 12

3. Result

RMSE (Root Mean Square Error) is measured to find the distance between the predicted values and the actual values of a machine learning model. RMSE is the standard deviation of the prediction errors (residuals). That is, the residuals are a measure of how far the regression line is from the data points; the RMSE is a measure of how far these residuals are spread. In other words, it tells us how concentrated that data is around the line that



best fits the data. An RMSE value of zero means that the model makes no errors. The plus-minus values for the prediction in the scenario in this study can be considered as follows. 13832 hours is the margin of error of the 80% project taught to the system. 137978 is the margin of error for 20% of the projects tested. Since the number of tested projects is very small, it is normal to have such a high deviation. (Figure 13)

```
In [38]: rmse = np.sqrt(mean_squared_error(y_train, model.predict(X_train)))
```

```
In [39]: rmse
```

```
Out[39]: 13832.157957461912
```

```
In [40]: rmse = np.sqrt(mean_squared_error(y_test, model.predict(X_test)))
```

```
In [41]: rmse
```

```
Out[41]: 137978.46448875545
```

Figure 13

4. Discussion and Conclusion

It is clearly seen that the main focus of academic studies on project management in the construction industry for years has been the project management that has been taught for years. With the introduction of BIM in our lives in the recent past and the transformation it has created worldwide, project management styles have started to change as well as working systems have changed. Although BIM aims to include the project management process in the system from level 2, the management phase of the projects continues to be carried out with the familiar methods. In project management, the development of a project in the proposal phase and the work done for the proposal phase are also of great importance. One of them is to make accurate cost calculations for the project and to make a successful approximate man-hour calculation.

Literature review shows that many studies, analyses and frameworks on manpower analysis have been developed in the academic field since many years ago, but there are very few studies that include man-hour analysis for BIM projects. In addition, it is seen that today's achievements such as artificial intelligence applications and machine learning are only recently reflected in such studies. Artificial intelligence applications will



save a lot of time for these estimations and with a successful application, the estimation errors made in the proposals will be significantly reduced thanks to machine learning. The machine learning applications shown in this report have been analysed with simple linear regression and Multiple Linear Regression within the scope of regression. Data was of great importance in the analysis. However, during the data collection process, sufficient information could not be obtained due to the confidentiality principles of the companies. These analyses, which were conducted with a total of 16 projects, could have yielded more efficient and more accurate testable results if they had been conducted with many more projects. However, realistic results were obtained with the existing projects. In the light of the results, it is seen that integrating artificial intelligence applications into this type of analysis is a great benefit for business efficiency.

References

- [1] X. D. Ma, F. Xiong, T. O. Olawumi and N. Dong, "Conceptual Framework and Roadmap Approach for Integrating BIM into Lifecycle Project Management," *Journal of Management in Engineering*, 2018.
- [2] D. Barnes, "Manpower Planning," 1975.
- [3] P. Yan, X. Xie and Y. Meng, "Application of the BIM Technique in Modern Project Management," ASCE, 2014.
- [4] Oracle, [Online]. Available: <https://www.oracle.com/data-science/machine-learning/what-is-machine-learning/>.