4th Construction Specialty Conference
4e Conférence spécialisée sur la construction

Montréal, Québec
May 29 to June 1, 2013 / *29 mai au 1 juin 2013*

Automation of Markup Decisions in Construction Projects through Machine Learning (ML)

Tarek Mahfouz
Department of Technology, Ball State University, Indiana, USA
James Jones
Department of Technology, Ball State University, Indiana, USA

Abstract: The markup value included in a bid is usually the outcome of a complicated analysis that requires experience. It entails taking into account different factors that could be internal and/or external. However, the precedence of one factor to the other and its effect on a decision to bid varies due to the dynamic nature of the industry and the complexity of the projects. Consequently, accurate estimation of markup values through the assessment of the different factors and their associations is essential. This paper proposes an automated decision support tool for markup values through Machine Learning. To that end, the adopted methodology (1) utilizes data from a set of completed projects; (2) defines a list of factors based on a comprehensive literature review and previous experiences; (3) proposes an automated methodology through Support Vector Machines (SVM), Naïve Bayes (NB), and Decision Tree (DT) for the assessment of markup values; and (4) compares the outcomes of the developed models to select the best suitable for the current research task. The models retrieve the closest case to a newly encountered one, and report the estimated markup value to be adopted. The outcomes of the current research illustrate that ML modeling has high potential in this field.

1 Introduction

Construction is a complex and highly dynamic industry in which decision are based on multifaceted analysis of interrelated and integrated factors that are continuously changing. Due to this dynamic nature, mere consideration of factors must be augmented by previous experiences. Such phenomena make Construction a knowledge intensive industry where lessons learned provide a summary of previous experiences needed for future decision making. A decision about the markup value to be included in a bid estimate is no different in this regards. It entails the analysis of complex set of factors that could be internal and/or external. The former is defined as consideration related to the performance and ability of a firm to perform the required work under a newly introduced contract like number of current projects undertaken by the firm, the capacity, ability, and experience of the in house team, nature of the project and the scope of work, experience with similar projects, financial capacity of the firm, etc. On the other hand, the latter includes associations to the construction market at which the work is performed including, market stability, local/ foreign, market risk, owner's financial capacity, level of competition, etc. A problem manifests when a number of these factors coexist, for the precedence of one factor to the other and its effect on a decision to bid varies due to the dynamic nature of the industry and the complexity of the projects. Consequently, accurate estimation of markup values through the assessment of the different factors and their associations through lessons learned is essential.

In an attempt to alleviate risks associated with this problem and to provide decision makers with better understanding of their decision making processes, researchers utilized Artificial Intelligence and statistical methodologies to analyse different components of construction projects costing and budgeting. They covered a wide range of spectrum ranging between statistical models (Russell and Zhai 1996, Hwang 2009), Case-Based Reasoning (CBR) (Dogan et al. 2008), Machine Learning (ML) (An et al. 2007, and Mahfouz 2011), MATLAB (Lee et al. 2012), and hybrid models (Son et al. 2012). Although these researches have improved knowledge in this field, none of them provided a prediction tool for markup value in construction bids through the use of factors proven to be statistically significant through lessons learned.

Consequently, this paper presents a decision support methodology for predicting markup value through Machine Learning (ML). To that end, the adopted research methodology (1) utilizes a set of completed projects; (2) uses a set of statistically significant factors; and (3) develops and compared the outcomes of three ML algorithms namely, Support Vector Machines (SVM), Naïve Bayes (NB), and Decision Tree (DT) for the assessment of markup values. The outcomes of the current research illustrate that ML modeling performs accurately in assigning markup values, making it a powerful tool for decision making. The rest of the body of this paper describes the followings.

- Background;
- Research Methodology;
- Results and discussion; and
- Conclusion.

2 Background

A decision about the value of markup to be included in a new project is an essential one that in some cases could be a decisive edge in winning or losing bids for a contractor. In most cases, it is a paramount decision that affects all project stakeholders. However, the evaluation process to make such a decision cannot be made only based on project characteristics. It involves a complex analysis of factors that are interrelated. In most cases, the importance and weight of one factor in comparison to another is based on previous experiences. Consequently, the required knowledge to make a decision about markup values for the construction industry is present in experiences gained through previously performed projects. However, it is not available in an explicit form; rather it exists in a latent form through decisions made in similar projects after a comprehensive analysis of all associated factors. Some of these experiences have led to success, while others had far less achievement. However, all of these experiences are essential for future markup value analysis so that more informed decisions are made and errors are eliminated. As a result facilitating the evaluation processes through lessons learned has become a necessity to provide decision makers with better understanding of the outcomes of their judgements and to focus their efforts on more value adding tasks. A number of research studies have addressed this issue. They adopted research methodologies that implemented statistical models, Case-Based Reasoning (CBR), Machine Learning (ML), MATLAB, and Hybrid modeling.

In 2012, Tamer, Yoon, and Hastak created a decision support methodology through the development of a protocol for the analyses of construction projects profitability. The protocol provided decision makers with the ability to "(1) recognize the changes in profit margins of projects, (2) improve their overall profitability by eliminating problems from the relationships for the prospective projects, and (3) be selective of more profitable projects in the bidding phase." (Tamer et al. 2012). Hyojoo Son, Changmin Kim, and Changwan Kim (2012) developed a hybrid model for the prediction of cost performances of commercial projects. The model integrated principal component analysis and support vector regression for its prediction. It utilized a total of 64 factors that related to pre-project planning stages. Furthermore, Lee, Lim, and Arditi developed an automated tool that utilizes MATLAB for the analysis of project activities for the development of best fit cashflow overdrafts and profit (Lee et al 2012). Mahfouz (2011) and An et al. (2007) utilized Latent Semantic Analysis and Support Vector Machines, respectively, for the evaluation and assessment of conceptual cost estimate through project related factors. Hwang (2009) and Dogan et al. (2008) developed assessment models for the prediction of construction cost. The former utilized dynamic regression modeling for the forecasting of construction cost indexes in attempt to achieve more informed and accurate cost estimates. The former utilized data related to design parameters and structural components' costs for the model development and assessment. The adopted research methodology established factors weights through decision trees that were further implemented through case based reasoning system. Ammar et al. (2003) studied the relation between profitability and firms' sizes for electrical contractors who are members of the Federated Electrical Contractors. Russell and Zhai (1996) predicted contractors' failure through stochastic analysis of economic and financial variables. Although these researches have contributed tremendously to the advancement and knowledge in this area, they did not consider the associations and interrelations between decision governing factors.

To fill this research gap, this paper proposes an automated decision support methodology for markup value through ML techniques, namely Support Vector Machines (SVM), Naïve Bayes (NB), and Decision Trees, while incorporating internal and external factors that are proven to be statistically significant for the problem analyzed.

3 Research Methodology

To achieve the aforementioned goal, the adopted research methodology for the current research task, (1) utilized data about internal and external factors from a set of completed construction projects; (2) identified a set of factors that govern decision making related to markup value and proven to be statistically significant for current research task; (3) developed four (4) ML models (2 SVM, 1 NB, and 1 DT) to predict markup values for future projects through lessons learned in previous projects and the identified factors; and (4) compared the outcomes of the developed ML models to adopt the most suitable for the current research task based on performance measure (see Figure 1). The rest of this section details the stages of development and implementation of the adopted research methodology, namely (1) Data Preparation, (2) Factors Identification, (3) Model Development and Implementation, and (4) Model Testing and Validation.

3.1 Data Preparation

The adopted research methodology utilized a set of 27 completed projects. Projects were selected based on the following criteria.

(1) Projects are completed;
(2) Full set of documents is available; and
(3) Final markup percentages after completion of all works are known.

In an attempt to have a comprehensive model for the construction industry, analyzed data included completed contracts as well as change orders. Consequently, if a change order existed within project documents, it was separated and dealt with as a separate project. To that end, the compiled set of data constituted of 15 completed projects and 12 change orders (Figure 2).

3.2 Factors Identification

The set factors utilized for the development of ML models were identified in two folds. First, a comprehensive literature review of previous researches performed in this area was performed for initial factor identification. These factors were further augmented with parameters believed to be important based on the research team experiences. Table 1 illustrates the full set of the 32 initially identified factors. Since these factors are interrelated, it was crucial to analyze the individual effect of each one on decision making processes related to markup value and to identify a combination that has the most prominent impact on consideration of such a decision. To achieve this goal, individual effects and optimum combination analysis was performed through statistical modeling, namely Ordered Probability Modeling, under a separate research task. The aforementioned analysis identified a set of 14 factors (Table 2) to be statistically significant for markup value decisions. For full details of the analysis, please refer to Mahfouz et al. 2013.

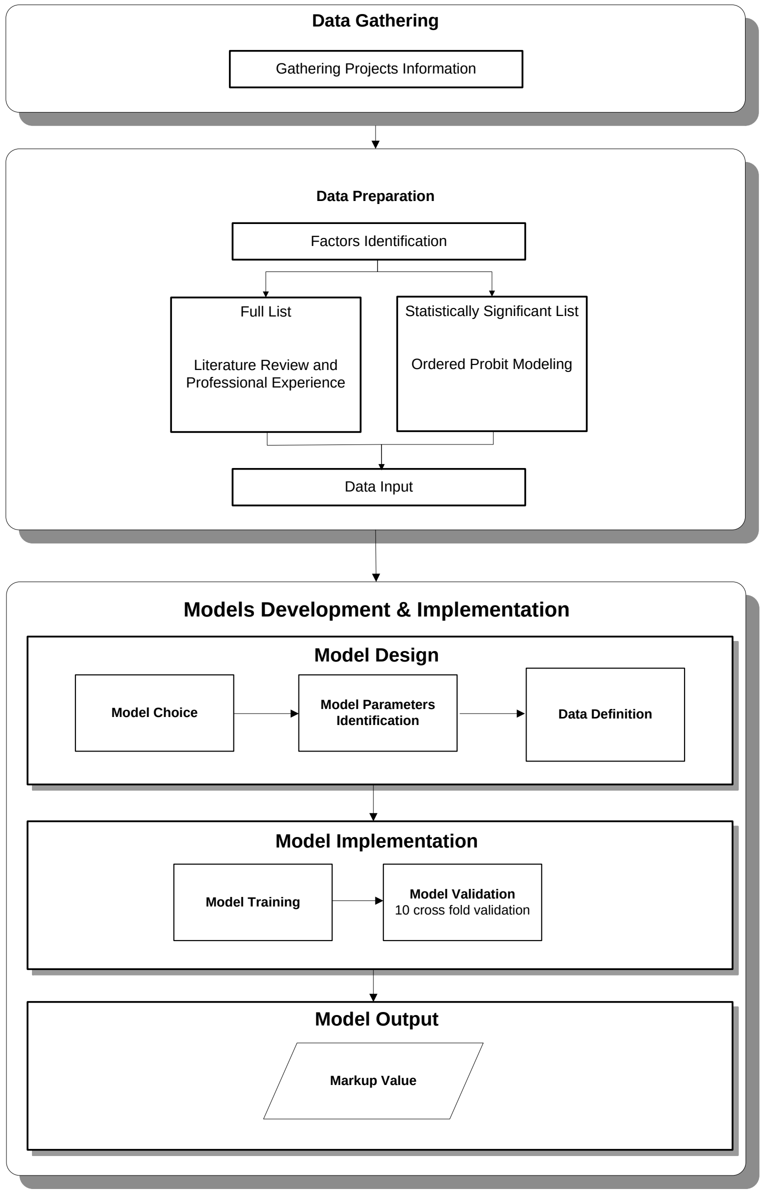

Figure 1. Research methodology

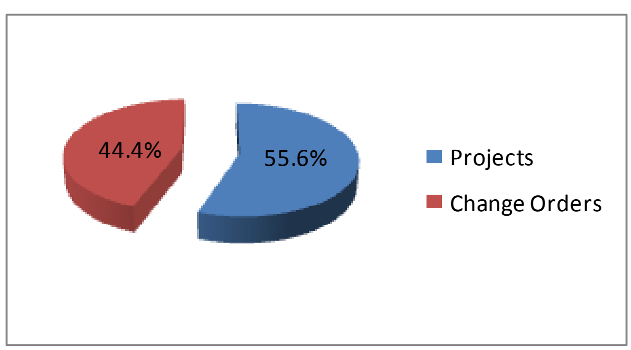

Figure 2. Data percentage distribution

Table 1: Initially identified set of factors

Item	Factor Definition	Type of Factor
1	Intensity of the site visit	Ordinal 5:high–0:none
2	Site clearness of obstacles during site visit	Ordinal 5:high–1:low
3	Possibility of differing site conditions	Ordinal 5:high–0:none
4	Level of site survey	Ordinal 5:high–1:low
5	Experience with similar projects	Ordinal 5:high–1:low
6	Details of existing data	Ordinal 5:high–0:none
7	Level of details in project definition	Ordinal 5:high–1:low
8	Level of details in project scope statement	Ordinal 5:high–1:low
9	Level of details of the project drawings	Ordinal 5:high–1:low
10	Level of details of the project technical specifications	Ordinal 5:high–1:low
11	Level of details of the project general conditions	Ordinal 5:high–1:low
12	Level of details of the project supplementary conditions	Ordinal 5:high–1:low
13	Level of commitment of the company to the project	Ordinal 5:high–1:low
14	Financial capacity of the company	Ordinal 5:high–1:low
15	Financial capacity of the client	Ordinal 5:high–1:low
16	Time to estimate	Numerical days
17	Difficulty of the estimating procedures	Ordinal 5:high–1:low
18	Estimator's career experience	Numerical years
19	Estimator's field work experience	Numerical years
20	Estimator's experience with similar projects	Ordinal 5:high–0:none
21	Estimator's experience with field work in similar projects	Ordinal 5:high–0:none
22	Capacity of the estimating team	Ordinal 5:high–1:low
23	Number of others projects under estimation	Numerical integer
24	Capacity of the architectural team	Ordinal 5:high–1:low
25	Capacity of the procurement team	Ordinal 5:high–1:low
26	Capacity of the technical office team	Ordinal 5:high–1:low
27	Capacity of the quality control team	Ordinal 5:high–1:low
28	Capacity of the quality control team	Ordinal 5:high–1:low
29	Capacity of client	Ordinal 5:high–1:low
30	Level of construction difficulty	Ordinal 5:high–1:low
31	Level of competition	Ordinal 5:high–1:low
32	Type of Project	Integer 1:Original Project –2: Change Order

Table 2: Statistically significant factors

Item	Factor Definition	Type of Factor
1	Intensity of the site visit	Ordinal 5:high–0:none
3	Possibility of differing site conditions	Ordinal 5:high–0:none
5	Experience with similar projects	Ordinal 5:high–1:low
9	Level of details of the project drawings	Ordinal 5:high–1:low
10	Level of details of the project technical specifications	Ordinal 5:high–1:low
11	Level of details of the project general conditions	Ordinal 5:high–1:low
13	Level of commitment of the company to the project	Ordinal 5:high–1:low
14	Financial capacity of the company	Ordinal 5:high–1:low
16	Time to estimate	Numerical days
20	Estimator's experience with similar projects	Ordinal 5:high–0:none
23	Number of others projects under estimation	Numerical integer
30	Level of construction difficulty	Ordinal 5:high–1:low
31	Level of competition	Ordinal 5:high–1:low
32	Type of Project	Integer 1:Original Project –2: Change Order

3.3 Model Development and Implementation

To that end, the adopted research methodology utilized three techniques of ML, namely SVM, NB, and DT. The following sections provide insight into these models.

3.3.1 Support Vector Machines (SVM)

Hypothesis space of linear functions in high dimensional space is used by SVMs learning systems from optimization theory, trained with a learning algorithm that employs learning biased driven from statistical learning theory (Shawe-Taylor and Cristianini 2000). Within the current research SVM Classification targets to divide the training projects into positive and negative classes based on the 14 identified factors. In its linear form, binary categorization is executed by means of a real-valued hypothesis function, equation 1, where input x (project) is allocated to the positive class (Specific Markup Value) if $f(x) \geq 0$; if not, it is allotted to the negative class (Mahfouz and Kandil 2012, Mahfouz 2012).

$$[1]\ y = < wx > b$$

Since multiple hyperplanes can exist, in binary linear separation problem a hyperplane which is allocated to be $f(x) = 0$, the separation (γ) is maximized where the vector ***w*** (weight vector) and ***b*** (functional bias) (Figure 3).

If the data are not linearly separable, during the formation of the proposed SVM models, a problem manifests. Complex real life problems are rarely linearly separable which is stated by Shawe-Taylor and Cristianini (2000). In such cases, a transformation of the input factors into a higher dimension space is mandated to shift the problem from linearly inseparable to linearly separable (Mangasarian and Musicant 1999). As noted in the literature, Kernel transformation function Ø (Equation 2) presents a solution by transforming the projects vector into a higher dimension space that is linearly separable (Equation 3) (Mahfouz et. al. 2010).

[2] $\mathbf{F} = \{Ø(\mathbf{x})\ |\mathbf{x} \in X\}$ where Ø is the kernel transformation function

[3] $\mathbf{x}=(x_1,\ldots,x_n) \rightarrow Ø(\mathbf{x})=\left(Ø_1(\mathbf{x}_1),\ldots,Ø_n(\mathbf{x_n})\right)$

According to the present categorization problem, this paper investigates a variety of kernel transformation which is an effort to derive the best prediction model.

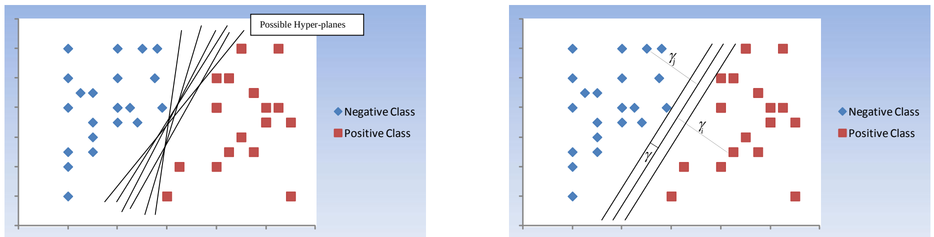

Figure 1a. Possible Hyperplanes in SVM

Figure 1b. Hyperplane Representation in SVM

Where γ_i and γ_j defines the Euclidean distance of two points from the decision boundary in the input space; and γ is the maximum geometric margin over all possible hyperplanes

Figure 3. Support Vector Machine (SVM) hyperplane (adopted from Mahfouz 2012)

3.3.2 Naïve Bayes (NB)

This type of classification is based on conditional probabilities. In other words, the model learns the classification from a set of projects and predicts the outcomes of a newly encountered one as the highest probability of a specific markup value while having prior knowledge of the values of the 14 identified factors. However, the calculations are extremely simplified by assuming that the factors are mutually exclusive, thus the Naïve part of the name (Alqady 2012, Mahfouz and Kandil 2012, and Mitchell 2010). Consequently, a newly analysed project can only belong to one class (markup value). Accordingly, “the classification is performed by choosing the class with the maximum posterior probability”(Alqady 2012) in accordance with equation 4.

[4] $C_{max} = argmax_{c_{j \in C}} p(c_j) \prod_i^m p(f_i|c_j)$

where c_j is a specific class among the set of classes C, f_i is factor i among m number of factor in the test set, $p(c_j)$ is the prior probability of class c_j; and $p(f_i|c_j)$ is the probability of a factor f_i given class c_j (adopted from Alqady 2012).

3.3.3 Decision Tree (DT)

In comparison to the aforementioned models, Decision Trees (DT) classification algorithms predicts an outcome (markup value) of newly introduced project based on rules that were derived from a training set of previous examples (completed projects) (Mahfouz and Kandil 2012, and Bramer 2007). The prediction rules learned from previous projects are set in the form of a tree expanding downward, where each split (leaf) is governed by a specific induced rule, hence the name Decision Tree (DT). For more illustration, for a given training data set, decision rules are generated in accordance with a process known as splitting on the value of attributes. In this case, each project characteristic within the set of completed projects is tested for all of its possible values. Within the current data, for a discrete characteristic like the type of project, a split leaf is created for each possibility. However, continuous variables like time to estimate are branched at values like “less than or equal to a value”, “greater than or equal to a value”, “less than a value”, “greater than a value” … etc. The splitting mechanism is continued until all attributes are tested and each complete branch is associated with one classification (Mahfouz and Kandil 2012).

3.3.4 Model Training

To that end, the developed models were trained using the data set of 27 completed projects, where the 14 statistically significant factors represented the attribute of each project and the final markup value is the prediction parameter. The models were trained in a 10 fold cross validation process. For more clarification, the full set of projects is divided into two subsets. The first includes 10% of the projects and the second is compromised of the other 90%. The model is trained using the subset of 90% and tested on the 10% one. This process is repeated in an iterative manner until the developed model is trained and tested over all projects in the initial set.

3.4 Model Testing and Validation

The model testing and validation is based on average performance measures achieved through the 10 fold cross validation plan, namely Prediction Accuracy (PA), Precision (P), Recall (R), and F-Measure (F) (Mahfouz and Kandil 2012). Within the scope of this research, Prediction Accuracy is defined as the ratio of projects correctly predicted by the model to the total number of projects tested. Precision (P), Equation 5, is defined as the ratio of projects with a markup category correctly predicted to the number of projects predicted as belonging to that markup category. In comparison, Recall (R), Equation 6, is defined as the proportion of projects belonging to a specific category of markup prediction that the model correctly assigned. Due to the fact that there is always a trade-off between P and R, where a model can attain 100% P at 0% R and vise versa, an overall performance measure combining P and R can be calculated through F-Measure (F), Equation 7 (Mahfouz and Kandil 2012 and 2010, and Mahfouz et al. 2010). The aforementioned measures of performance are reported for each fold of cross validation and their averages are reported as the final measure of each developed model.

[5] Precision (P) = $t_p / (t_p+f_p)$

[6] Recall (R) = $t_p / (t_p+f_n)$

[7] F- Measure (F) = 2PR / (R+P)

Where t_p is the true positive prediction of the model, f_p is the false positive prediction of the model, and f_n is the false negative prediction of the model.

4 Results and Discussion

The results of the implementation of the adopted research methodology discussed in section 3 are illustrated in Table 3 and Figures 4 and 5. It is worth mentioning that for the sake of the comparison, the developed models were tested using the full set of project factors (32) and the statistically significant ones (14). By comparing the final performance measures achieved by the 4 developed models, it is evidently clear that:

- The performance of all developed models increased by using the statistically significant factors except for the Naïve Bayes (NB) model (refer to Table 3 and Figure 4). This fact is attributed to the nature of the model. Since NB prediction is based on conditional probability, as the number of factors in the training set increases, the knowledge about the prior projects increase, thus enhancing the prediction capability.
- The highest increase in prediction accuracy (16.34%) was attained using third polynomial degree kernel SVM (refer to Figure 4). This is attributed to the active learning characteristic of SVM algorithms. As the ratio of attributes to the total number of instances decreases, the transformation of a data set into a higher dimension space enhances the prediction accuracy of the model.
- The lowest increase in prediction accuracy (3.33%) was achieved using Decision Trees (DT) (refer to Figure 4). The low effect due to factors reduction is related to the mechanism by which the DT algorithm operates. As discussed earlier, splits are generated after testing all attributes (factors). Consequently, the smaller set of factors was originally tested within the full set and rules were defined. However, the reduction only resulted in better rule generation that had minor enhancement on the model's prediction ability.
- The highest average prediction accuracy, precision, recall, and F-measure, while using the 14 factors, were achieved by the Linear SVM model. It attained an increase of 6.3% over 3rd polynomial Kernel SVM and 17.67% over NB and DT. This fact is related to the active learning characteristic of SVM discussed previously.
- The highest average prediction accuracy, precision, recall, and F-measure, while using the 32 factors, were achieved by the NB model. It attained an increase of 1.67%, 13.34%, and 11.67% over Linear SVM, 3rd polynomial Kernel SVM and DT models respectively. This fact is related to the nature of NB illustrated above.

Table 3: Model Comparison

	SVM Linear		SVM 3rd Degree		NB		DT	
	32	14	32	14	32	14	32	14
P A	47.00%	58.00%	35.33%	51.67%	48.67%	40.33%	37.00%	40.33%
P	48.90%	48.90%	42.00%	51.67%	54.10%	40.84%	53.11%	53.67%
R	93.00%	93.00%	72.00%	67.67%	69.92%	54.34%	54.50%	65.75%
F-Measure	64.10%	64.10%	53.05%	58.60%	61.00%	46.63%	53.80%	59.10%

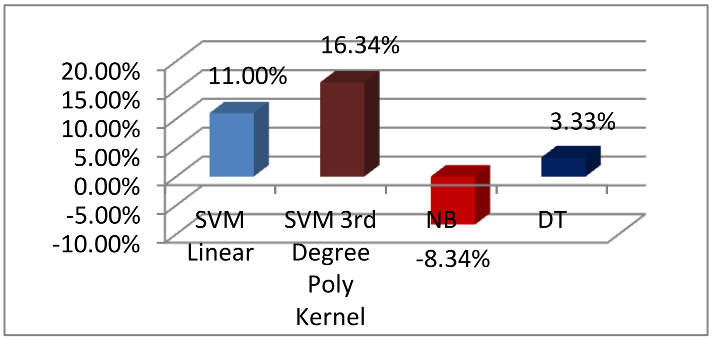

Figure 4. Prediction accuracy enhancement due to factors reduction

5 Conclusion, Limitations, and Future Works

This paper proposed an automated methodology for assessing markup values for the construction industry. To that end, the adopted research methodology utilized a set of 27 completed projects, identified a set of factors that affect decision making related to markup values, developed 4 ML models, namely Linear SVM, 3rd degree polynomial kernel SVM, Naive Bays, and Decision Tree. The implementation of the aforementioned research methodology yielded that the best model to be used for the current research task is Linear SVM. It achieved a prediction accuracy of 58% while using a smaller set of factors proven to be statistically significant.

The adopted research methodology has some limitations that should be discussed. First, the number of projects used for training, testing, and validating the models should be increased. It is instigated that as the number of instances increase, the performance of all models will be enhanced. Second, although the use of smaller set of factors enhanced the model performance, it is clear that these factors should be examined in more depth. The set of identified factors were based on literature review and the research team experiences. However, include experiences of other professional practitioners in the construction industry is mandated. Furthermore, all factors utilized for the model developed are equally weighed. This might not be true for all construction projects. Consequently, the future work of the research team will address the above-mentioned limitations through (1) gathering a more comprehensive set of projects, (2) integrating other professionals' insight through personal interviews and survey instruments, and (3) developing weighing mechanism that captures precedence of one factor to another.

6 References

Alqady, M. A. (2012). "Leveraging text content for management of construction project documents." Ph.D. thesis, College of Engineering, Division of Construction Engineering and Management, Purdue Univ. West Lafayette, IN.

Ammar, A., Hanna, A., Nordheim, E., and Russell, J. (2003). "Indicator Variables Model of Firm's Size-Profitability Relationship of Electrical Contractors Using Financial and Economic Data." J. Constr. Eng. Manage., 129(2), 192–197.

An, S., Park, U., Kang, K., Cho, M., and Cho, H. (2007). "Application of Support Vector Machines in Assessing Conceptual Cost Estimates." J. Comput. Civ. Eng., 21(4), 259–264.

Bramer, M. (2007). Principles of daw mining, Springer Verlag, London.

Doğan, S., Arditi, D., and Murat Günaydin, H. (2008). ”Using Decision Trees for Determining Attribute Weights in a Case-Based Model of Early Cost Prediction.” J. Constr. Eng. Manage., 134(2), 146–152.

Hwang, S. (2009). ”Dynamic Regression Models for Prediction of Construction Costs.” J. Constr. Eng. Manage., 135(5), 360–367.

Lee, D., Lim, T., and Arditi, D. (2012). ”Stochastic Project Financing Analysis System for Construction.” J. Constr. Eng. Manage., 138(3), 376–389.

Mahfouz, T (2011). “Application of Latent Semantic Analysis for Conceptual Cost Estimates Assessment in the Construction Industry” Proceedings of 2011 American Society of Civil Engineers (ASCE) Workshop of Computing in Civil Engineering, Miami, FL, USA. P. 752-759.

Mahfouz, T. (2012). “A Productivity Decision Support System For Construction Projects Through Machine Learning (ML)”. Proceedings of the 2012 CIB W078 29th International Conference On Applications Of IT In The AEC Industry; OCTOBER 17-19, 2012; Beirut, Lebanon.

Mahfouz, T., Jones, J., and Kandil, A. (2010). “A Machine Learning Approach for Automated Document Classification: A Comparison between SVM and LSA Performances.” Int. J. of Eng. Res. and Innov. (IJERI), 2(2): 53 - 62.

Mahfouz, T., Jones, J., and Kollwitz, S. (2013). “An Ordered Probability Framework to Assess Markup Values for the Construction Industry” Int. J. of Eng. Res. and Innov. (IJERI). Accepted for publication on May 15, 2013.

Mahfouz, T. and Kandil, A. (2012). ”Litigation Outcome Prediction of Differing Site Condition Disputes through Machine Learning Models.” J. Comput. Civ. Eng., 26(3), 298–308.

Mahfouz, T., and Kandil, A. (2010). "Automated outcome prediction model for differing site conditions (DSC) through support vector machines (SVM)." *Proc., Int. Con.f. on Compwing in Civil and Building Engineering (ICCCBE 2010),* W. Tizani, eel., Nollingham University Press, Nollingham, UK. 289-294.

Mangasarian, L., and Musicant, D. R. (1999). "Massive support vector regression." Technical Rep. Data Mining Institute TR-99-02, Univ. of Wisconsin, Madison, WI.

Mitchell, T. M., (2010) “Generative and Discriminative Classifiers: Naïve Bayes and Logistic Regression” Unpublished Book Chapter.

Russell and Zhai (1996). ”Predicting Contractor Failure Using Stochastic Dynamics of Economic and Financial Variables.” J. Constr. Eng. Manage., 122(2), 183–191.

Shawe-Taylor, J., and Cristianini, N. (2000). Support veer or machines and other kernel-based learning methods, Cambridge University Press, Cambridge, UK.

Son, H., Kim, C., Kim, C. (2012). “Hybrid principal component analysis and support vector machine model for predicting the cost performance of commercial building projects using pre-project planning variables.” Automation in Construction, Volume 27, November 2012, 60-66.

Tamer, Z., Yoon, Y., and Hastak, M. (2012). ”Protocol for Profitability Analysis Using Internal Entities in Organizational Structure of Construction Companies.” J. Constr. Eng. Manage., 138(12), 1394–1402.