American-Eurasian J. Agric. & Environ. Sci., 13 (10): 1314-1324, 2013
ISSN 1818-6769
© IDOSI Publications, 2013
DOI: 10.5829/idosi.aejaes.2013.13.10.1443

Profitability Prediction in Agribusiness Construction Contracts: A Machine Learning Approach

Anand N. Asthana

Centrum Católica Graduate Business School,
Pontificia Universidad Católica Del Perú

Abstract: As agriculture transforms itself from a subsistence activity to agribusiness across the world, the importance of agribusiness construction is increasing. At the same time, on account of globalisation and technological developments more and more agribusiness construction companies are becoming multinational and have started using better data processing techniques. The actual process of how contractors and their clients negotiate and agree to price is complex and not clearly articulated in the literature; but in almost all cases the profitability of a contract is the most important factor for any decision on the bid. Commercial managers employed by construction companies are asked to estimate the expected profit on a prospective contract to either decide whether to proceed with the project or to aid in financial forecasting for the company. The estimation of a prospective contract's profitability is generally done by intuition. A mathematical model to aid in predicting the profitability of a prospective contract would be of immense use to a commercial manager. It may be used as the primary source of predicting profitability whilst some commercials managers may use it as a valuable second opinion. Furthermore, it would of considerable interest to commercial managers to know the effect on predicted profitability of a contract should they change the value of an attribute of a prospective contract. the paper demonstrates that both the VSM and KRR routines are fairly simple to implement in a commercial setting. Commercial application will, however, require close interaction between scientists and business scholars.

Key words: Agribusiness · Construction · Profitability · Machine Learning

INTRODUCTION

Construction plays an important role in the development of a strong agricultural economy. This is evidenced by the need to construct efficient farm-to-market roads, irrigation channels, bridges, grain silos and facilities to produce and store agricultural goods. Agricultural construction spans a wide range of projects which can be split into primary, secondary and tertiary groups. Primary projects are those that directly affect farmers and their ability to work. These projects include the building of barns and silos, seed and grain processing, hog production and dairy production facilities. Secondary projects include essential infrastructure within a country. These construction projects involve the building of farm-to-market roads, bridges, railroads and similar projects. Tertiary projects are the peripherals related to the agricultural community, including, hotels, motels, office buildings and grocery stores [1].

One of the tasks of a commercial manager in an agribusiness construction company is to estimate the expected profit on a prospective contract. On the basis of this assessment, the company can decide whether to bid for the contract and the amount and nature of bid. Formal and analytical risk models prescribe how risk should be incorporated into construction bids. However, the actual process of how contractors and their clients negotiate and agree to price is complex and not clearly articulated in the literature [2]. In any case, the company needs to estimate the profitability before any decision on the bid can be taken.

The estimation of a prospective contract's profitability is difficult due to the range of size and types of contracts and the types of work undertaken. Furthermore, some agribusiness construction companies

Corresponding Author: Dr. Anand N. Asthana, Centrum Católica Graduate Business School, Pontificia Universidad Católica Del

specialise into a particular type of work whereas others take on many different types and sizes of work. Moreover, the profitability of a contract would certainly be influenced by the attitude of the client. While some may be extremely austere on payments made to the agribusiness construction company and often hold back payment (a process known as retention) until the very last stages of the contract, others may be less stringent due to internal factors.

Internal management of the contract heavily influences the profitability of a contract. The performance of the personnel assigned to the construction project has an influence on profitability. Other factors that influence the profitability of a contract include suppliers, productivity and availability of labour. Furthermore, most agribusiness construction companies employ subcontractors, which are other companies, on medium to large contracts usually for over half the work on the contract – and sometimes for the most of the work. The performance of the subcontractors can greatly affect the profitability of a contract if not supervised correctly.

Finally, apart from contract types and internal management, the profitability of a contract can be affected by unforeseen circumstances [3]. For example, a new government or local scheme can change the availability of labour and timely completion of a contract. If a contract requires specialist materials from a distant supplier, a sudden rise in global oil prices will increase costs for the contract and if it is not possible to pass this extra cost onto the client, the profitability will be severely affected. In agribusiness construction, uncertainties are more as most of the works are 'off road'. Consequent to globalisation of agribusiness, agribusiness construction companies are spreading their business to developing countries. This internationalisation has increased risk for companies as developing countries pose greater uncertainties to these companies [4]. In most countries agriculture as a subject is devolved to local governments. Many projects in developing countries are assigned by local governments where level of corruption is arguably higher and for the companies uncertainties on account of corruption remain even in presence of competitive bidding through open tenders [5]. Other risk factors are approvals and permits, changes in law and government policy, law enforcement, local partner's creditworthiness, political instability, higher inflation and changing interest rates and government influence on dispute resolution. The risks at country level are more severe than that at market level and the latter are more severe than that at project level [6].

Due to the number of variables and a large number of attribute values of the variables, it is not possible to use traditional *if-then-else* type of deterministic programming to make predictions about the profitability of a prospective contract. In such situations, application of Machine Learning is gaining wide acceptance as a useful tool in business research [7]. While popular business applications of machine learning are in the field of finance and marketing, newer applications are other businesses in many sectors in the service industry, like healthcare [8]. The objective of this paper is to create a Contract Profitability Prediction System using a Machine Learning algorithm that would predict the expected profitability of contracts at their starting point as well as to identify contract attributes which most influence profitability. Unfortunately, no prior Contract Profitability Prediction System exists which could have served as a template to improve upon. This paper describes the system developed and the data analysis undertaken and attempts to apply existing mathematical techniques and algorithms as a solution to a commercial problem.

Managing Contracts in Agribusiness Construction: In the construction business we find mainly two types of contracts. Fixed-price contracts provide strong cost-minimization incentives for the construction company, but raise the spectre of hold-up when the contract must be renegotiated to accommodate modifications to the project. In contrast, cost-plus contracts provide flexibility, since the principal continues to direct work on the project, but create essentially no incentive for cost-minimisation since the construction company is fully reimbursed for its costs [9]. In agribusiness construction, fixed price contracts are more common.

Estimation of the value of construction works of a contract undertaken by an agribusiness construction company is done by a Quantity Surveyors (QS). The QS keeps control of the costs and revenues of the contract as well as dealing with unforeseen circumstances and delays which may affect the profitability of the contract [10]. The QS generally submits a Cost Value Reconciliation (CVR) either monthly or quarterly which informs the management about the state of the contract. Commercial managers in agribusiness construction companies are usually senior or former QSs, who assist the management in bidding for prospective contracts and assist in the management of ongoing contracts. The QS keeps control of the costs and revenues of the contract and deals with unforeseen circumstances and

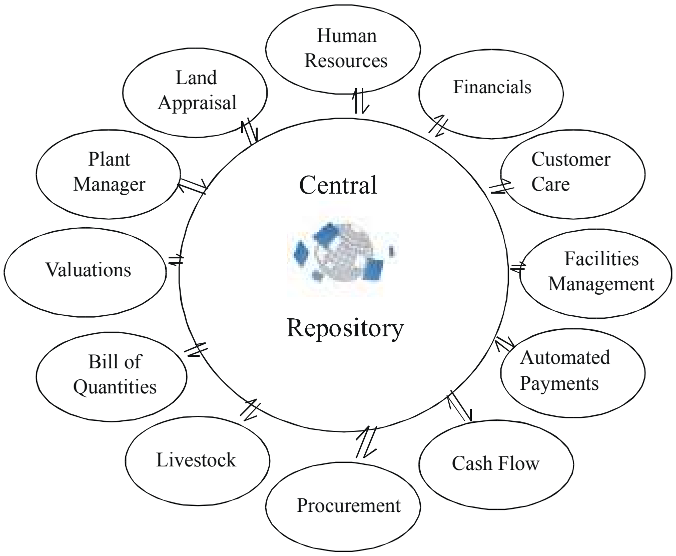

Fig. 1: ERP in a typical Agribusiness construction company

delays which may affect the profitability of the contract. The QS submits a Cost Value Reconciliation (CVR) either monthly or quarterly which informs the management about the state of the contract.

One of the most pervasive organisational change activities that occurred in the last decade of the twentieth century is the implementation of Enterprise Resource Planning (ERP) systems [11; 12]. An ERP system is a packaged business software system that enables a company to manage the efficient and effective use of resources (materials, human resources, finance, etc.) by providing an integrated solution for the organization's information processing needs [13]. The architecture of the software facilitates transparent integration of modules providing flow of information between all functions within the construction company in a consistently visible manner. Corporate computing with ERP system allows construction companies to implement a single integrated system by replacing or re-engineering their mostly incompatible legacy information systems [14]. Figure 1 shows a typical ERP system in a typical agribusiness construction company.

A typical ERP implementation in a large agribusiness construction firm takes between one and three years to complete and costs tens to hundreds of thousands of dollars. Several practitioners are of the view that ERP implementations yield more failures than successes in large construction firms [15]. ERP casts a big shadow on the employees, changing the nature of tasks and workflows and often the jobs themselves [16]. The agribusiness construction industry is characterized by activities that are discontinuous, dispersed, diverse and distinct in nature. Construction work is a demanding and stressful and construction teams often work day and night under incessant pressure to meet deadlines. The main concern of the project personnel is 'to get the work done' as early as possible to reduce project time. Under such circumstances it is extremely difficult for the people to provide a creative response to proposed changes. A major change is bound to cause problems [17]. The success or failure of an ERP system implementation is rarely tied to the features of the technology itself; more often it is linked to the job and processes reengineering that typically accompany such systems [18].

Notwithstanding these problems, more and more agribusinesses construction companies are switching over to ERP, not as an end in itself but for realisation of organisational goals [19]. Popular commercial ERP systems include SAP Business Suite, JD Edwards EnterpriseOne, Oracle E-Business Suite and PeopleSoft (by Oracle) and Microsoft Dynamics. GNU Enterprise (GNUe) is a popular open-source free-to-use ERP system.

A prospective contract is entered in the Contract Status Ledger. If it is decided that the company should proceed with the contract and all the legal agreements have been concluded with the client, the Bill of Quantities (BOQ) for the contract would be imported into the *Valuations* module. The BOQ contains all the items of work required to be completed. As the work commences on the contract, the QS in charge of the contract, would update the BOQ items in terms of percentage complete.

Using this information, the QS would bill the client using Contract Sales Ledger certificates. The client themselves will employ a QS, known as a Principal Quantity Surveyor (PQS) who will inspect the claims from the QS to determine the payment made to the agribusiness construction company. The amount claimed for and amount received will be stored on the certificate in Contract Sales Ledger. This will update the revenues of the contract.

As the work on the contract progresses, *Procurement* would be used to place orders from the selected suppliers, which would automatically update the costs of the contract. *HR & Payroll* will be used to pay the workers on the contract and these modules will also update the costs for the contract. For work that is done via subcontractors, orders will be placed via *Subcontract Ledger*. The subcontractor will follow a similar system for the work obtained. The subcontractor will then bill the agribusiness construction company for the work completed via subcontract certificates in the Subcontract Ledger. The subcontract certificate will contain both the applied-for amount by the subcontractor and the actual amount paid to the subcontractor. This module will also update the costs for the contract. At monthly or quarterly intervals, the QS will complete a Cost Value Reconciliation (CVR), which amongst other things contains the QS's forecasts for future costs and revenue for the contract. These values are loaded into the Contract Status Ledger for forecasting. The *Financials* module will retain a summary of all the costs and revenues for the contract. The reporting can be done at sub-contract, contract, group, or company level.

MATERIALS AND METHODS

Data Set: The data set was extracted from the live financial data and restricted to completed contracts which are upwards of $100,000 in costs incurred. The total number of contracts available in the data set is 934. Figure 2 displays the range of profit percentages. The distribution is skewed to the right, indicating that the number of contracts that were profitable is greater than the number of contracts which were loss-making. Approximately 40% of the contracts are in the 5% to 14% profit range, which is an encouraging news for this sector.

Extraction and Setup: Contracts below and above the -20% to 20% profit were rejected as outliers. The contract data is extracted from the live system by performing a database dump of table jc_job into a database dump-file.

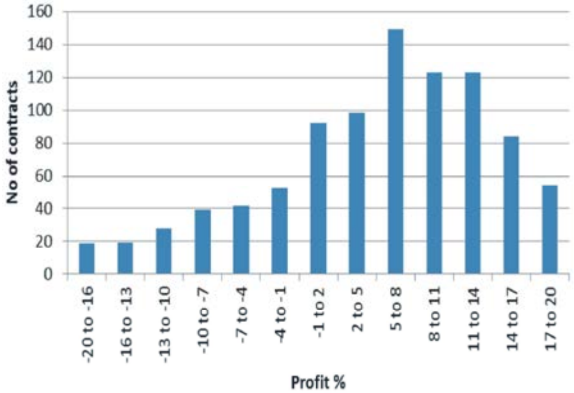

Fig. 2: Profits in agribusiness construction contracts

The dump-file is then used to create table jc_job of the same structure as the live system in a locally accessible database. Financials and Contract Status Ledger reports are run to extract the final cost incurred and revenue received for all the completed contracts. In the local database, fields prj_cost, prj_rev, prj_profperc are created on table jc_job. An index is created on table jc_job containing the following fields (in ascending order): job_complete, prj_cost, job_num. The cost and revenue data extracted from the reports are loaded into the new jc_job fields and profit percentage is calculated from cost and revenue. Since all the data is now in one database table, we can run a simple Progress queries on the contracts of interest, as follows:

```
for each jc_job no-lock where
jc_job.kco = 1 and
jc_job.job_complete and
jc_job.prj_cost >= dMinCost and
jc_job.prj_rev > 0 and
(jc_job.prj_profperc >= dMinProfitPerc and
jc_job.prj_profperc <= dMaxProfitPerc):
/* code */
end.
```

By specifying job_complete and prj_cost in the query, the new index created in step 5 above is automatically invoked and as a consequence, even though the database table jc_job contains a very large number of contracts, the completed contracts of over certain cost incurred, which are of interest to us, are retrieved very efficiently.

Contract Attributes: A contract entered in Contract Status Ledger, has several attributes which will serve as our predictor variables. 10 attributes were chosen some of

Table 1: Contract attributes

Number	Name	Description	Unique values
1	jcl_loc	The location of the contract	29
2	jgr_grp	The group within the company undertaking the contract	8
3	job_anl[1]	Attributes are used by agribusiness companies to enter information of their choosing. This could be for accounting or reporting purposes, or could be information like Group/Regional Manager	53
4	job_anl[2]		79
5	job_anl[3]		72
6	job_anl[4]		46
7	job_arc	The architect used for the contract	36
8	job_qsr	The QS in charge of the contract	92
9	jty_typ	The contract type. This could be revenue type, e.g. cost-plus or Pain/Gain, or could be another way of classifying contracts	31
10	rcm_num	The client for the contract	265

which may be extremely relevant toward contract profitability, whereas others may be completely irrelevant. Though we may have some prior knowledge or an intuition about which attributes will be relevant, we will not encode this information into the system; instead we will test the predictions of the system against our prior knowledge. All the attributes are nominal multinomial, i.e. the values are alpha-numeric codes which cannot be ranked. The breakdown of these attributes is presented in Table 1. The attributes extracted from the contracts are set when the prospective contract is input and are not changed once the contract has commenced. While the suppliers and subcontractors used while the contract is underway may contribute to the contract profitability, since we are making the contract profitability prediction of a prospective contract these operatives do not figure in our calculations.

Attribute Combinations: Apart from the main goal of making predictions on contract profitability, we also need to identify the attributes which contribute towards contract profitability. Possible combinations of the 10 attributes are:

$$_{10}C_1 + {}_{10}C_2 + {}_{10}C_3 + {}_{10}C_4 + {}_{10}C_5 + {}_{10}C_6 + {}_{10}C_7 + {}_{10}C_8 + {}_{10}C_9 + {}_{10}C_{10}$$
$$= 10 + 45 + 120 + 210 + 252 + 210 + 120 + 45 + 10 + 1 = 1023.$$

Cross-Validation: The experiments were done using 10-fold cross-validation which is commonly used [20]. The data is partitioned into 10 subsamples. Of the 10, each one in turn is used as test set and the other 9 as the training set. Leave-one-out cross-validation is not used due to the fact that it would prove to be computationally extremely expensive. However, we cannot divide the contracts into 10 subsamples as extracted from the database table. This is due to the fact that contract name is in the table index, which implies that contracts will appear in ascending alpha-numeric order. This could be a potential problem if similar contracts have similar names. In this case, we may end up with the scenario that contracts within each subsample may be very similar to each other but very different to contracts in another subsample. To overcome this difficulty, we pick contracts at random into the subsamples with the following algorithm:

```
define temp-table ttJob with fields i and name indexed by i
define temp-table ttFold with fields iFold and name indexed by iFold and name.
set total = 0 & folds = 10.
loop through all contracts filtered by cost and profit percentage increment total.
    create an entry in ttJob with ttJob.i = total & ttJob.name = contract name.
end loop
set foldsize = floor(total / folds).
loop variable i from 1 to (folds – 1)
    set j = 0.
    repeat until j < foldsize
        set x = random integer between 1 and total
        find ttJob where ttJob.i = x.
        if found ttJob
            create an entry in ttFold with ttFold.iFold = i & ttFold.name = ttJob.name.
            delete record from ttJob.
            increment j.
        end if
    end loop
    set total = total – foldsize.
    set j = 0.
    loop through all ttJob
        increment j.
        set ttjob.i = j.
```

```
end loop
end Loop
loop through all ttJob
create an entry in ttFold with ttFold.iFold = folds and ttFold.name = ttJob.name
end loop
export ttFold to text file for future use.
```

Vector Space Model (VSM): To make predictions about the profitability of a prospective contract, we can start by making an assumption that similar contracts will have similar profitability. For example, a contract to demolish an unused office building and to clear the area in a given location, managed by quantity surveyor QS1 and Regional Manager RM1 should be similar in profitability of another contract of the same type of work and managed by the same people which is undertaken a few months later, since the type of work, location of work and the personnel involved are the same.

To find similar contracts to a prospective contract, we use the VSM which is used to rank or classify textual documents in Information Retrieval. VSM is based on linear algebra and converts documents into vectors of index terms. One of the measures used to identify similarity is cosine similarity, which measures the angle between two vectors of *n* dimensions [21]. Given two vectors A and B, the cosine similarity is given by their dot product and magnitude:

Cos(θ) = $A•B$ / □□A□□ □□B□□

In information retrieval the document vectors would be represented by TF-IDF (Term Frequency – Inverse Document Frequency) which is one of the most commonly used statistical weighting schemes in today's information retrieval systems to evaluate how important a word is to a document or a corpus [22]. However, in our case this is not required or applicable since each contract attribute can take only one value and hence each contract can be represented as a vector containing attribute values, whose maximum length can be only 10. (While performing cosine similarity, normalizing by magnitude is required, as there exists a possibility that a particular attribute may not be set – i.e. blank/unknown value - on the contract).

Finally, in Information Retrieval, the "top-n" documents are retrieved for a given query. The value of *n* can have an effect on precision and recall. However, this is also not required or applicable in our case and we can set the value to a reasonable number, say 10 or 15. The system calculates the top most similar contracts for the one we're trying to predict and takes the average profitability of all the calculated similar contracts as the prediction. All the 1,023 attribute combinations are processed and predictions made for every contract using 10-fold cross validation. The best three and the worst three predictions are listed in columns 2 to 4 of Table 2.

The error distribution of the best attribute combination is shown in Figure 3a.

Outlier Elimination: Outliers are observations that are numerically distant from the rest of the data [23]. There are many approaches to dealing with outliers. The most common one which we follow is to 'throw the rascals out'. Detection of outliers is more problematic as the classic estimates of the mean and covariance matrix using all the data are extremely sensitive to the presence of outliers [24]. Mahalanobis distances provide the standard test for outliers in multivariate data in case of normal distribution. However, the performance of the test depends crucially on the subset of observations used to estimate the parameters of the distribution [25]. To identify outliers we use Random Sample Consensus (RANSAC) which is an iterative method of eliminating outliers by iteratively selecting a random subset of the given data as hypothetical inliers to calculate the true outliers of the data. The system calculates the top most similar contracts for the one we're trying to predict, performs outlier elimination and takes the average of the remaining inliers as the predicted profitability. The predictions are made by the system for every contract using 10-fold cross validation. Prediction error is shown in Figure 3b. The mean absolute error is 5.41 and the median absolute error is 4.28.

Weighted Nearest Neighbour: Performing outlier elimination on the results of Vector Space Model improves both the mean and the median absolute error. We know that the profitability of the majority of contracts lies in the 5% to 8% range (Figure 2). We use this knowledge by weighting the contracts which fall in this range higher than other contracts. Instead of taking the mean of the remaining inliers, we take the weighted mean:

$\Sigma w_i x_i / \Sigma w_i$

The system calculates the top most similar contracts for the one we're trying to predict, performs outlier elimination and takes the weighted mean of the remaining inliers as the predicted profitability. The predictions are made by the system for every contract using 10-fold cross

Table 2: VSM and KRR results

Rank	VSM			KRR			
	Fields	Mean Absolute Error	Median Absolute Error	Fields	Mean Absolute Error	Median Absolute Error	λ
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Best results							
1	jcl_loc, jgr_grp, job_anl[4], jty_typ, rcm_num	5.68	4.76	jcl_loc, jgr_grp, job_anl[4], jty_typ, rcm_num	5.01	4.13	0.1
2	jcl_loc, jgr_grp, job_anl[4], job_qsr, jty_typ, rcm_num	5.73	4.79	jcl_loc, jgr_grp, job_anl[4], job_qsr, jty_typ, rcm_num	5.07	4.17	1
3	jgr_grp, job_anl[4], job_qsr, jty_typ, rcm_num	5.80	4.85	jgr_grp, job_anl[4], job_qsr, jty_typ, rcm_num	5.09	4.19	0.1
Worst results							
1021	job_anl[1], job_anl[3] job_anl[2],	7.13	5.90	jgr_grp, job_anl[1], job_anl[3], job_arc, job_qsr, rcm_num	7.81	5.79	0.01
1022	jgr_grp, job_anl[1], job_anl[2], job_anl[3], job_arc	7.19	5.92	jcl_loc, jgr_grp, job_anl[1], job_anl[2], job_anl[3], job_anl[4], job_arc, job_qsr, rcm_num	7.97	5.85	0.01
1023	job_anl[2]	7.31	5.95	job_anl[1], job_anl[2], job_anl[3], job_anl[4], job_arc, job_qsr, rcm_num	7.97	5.87	0.01

validation and weighted nearest neighbour (WNN) are shown in Figure 3c. The mean absolute error is 5.09 and the median absolute error is 4.17.

Kernel Ridge Regression (KRR): A system with weights trained by regression can then be used to make predictions. Linear regression attempts to find a linear relationship:

$$Xw = Y$$

while the optimal value of weight w can be found using Ordinary Least Squares:

$$w = (X^{T}X)^{-1}\, X^{T}y$$

Ridge regression is useful when $(X^{T}X)^{-1}$ does not exist or inversion is numerically unstable. A problem that often arises in regression is overfitting when the model describes noise instead of the underlying relationship. One of the common techniques to combat this issue is to

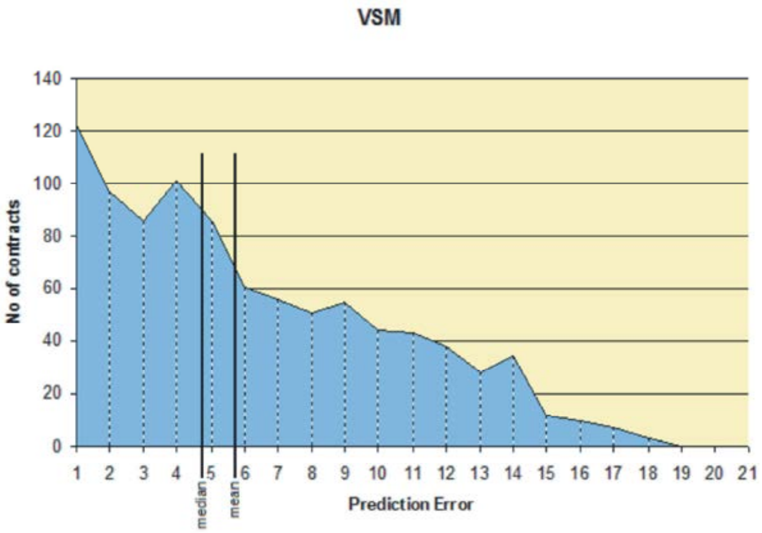

Fig. 3a: Using VSM

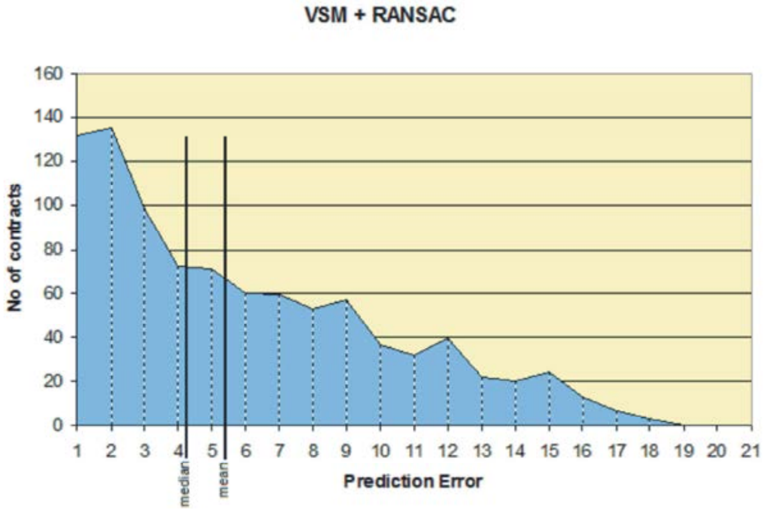

Fig. 3b: Using VSM and RANSAC

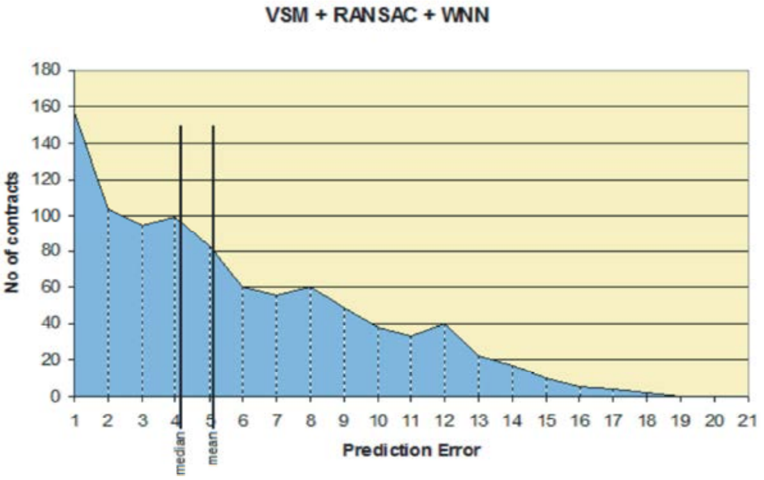

Fig. 3c: Using VSM, RANSAC and WNN

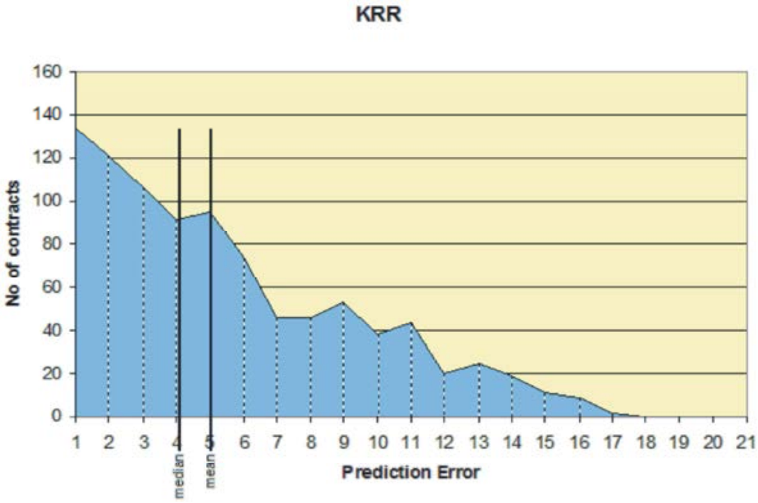

Fig. 3d: Using KRR

Fig. 3: Error distribution using various methods

introduce a regulariser (λ). This acts as weight decay, as in a sequential learning algorithm, it encourages weight values to decay towards zero, unless supported by data. With L training examples, the optimal value of weight vector with dimension n of the feature space can then be found as:

$$w = (X^TX + \lambda I_n)^{-1}X^Ty$$
$$w = \lambda^{-1}X^T\ (y\text{-}Xw)\ X^T\ y = X^T\ \alpha$$
$$w = \Sigma\alpha_i x_i$$
$$\alpha = (X^TX + \lambda I_L)^{-1}y$$

and the prediction function can be given by:

$$<w, x> = \Sigma\alpha_i<x_i, x>$$

Indicator Variables and Kernel Functions: Since all our predictor variables are nominal multinomial, we need to transform them into binary indicator variables for regression. The procedure creates a separate file for each attribute. When the regression system that is processing the data comes across a particular combination of attributes, it horizontally concatenates the files corresponding to the attributes in the combination being processed:

$$\square: D \rightarrow F,\ K(d_i, d_j) = <\square(d_i), \square(d_j)>$$

To construct the Kernel, we will try to replicate Vector Space Kernel, where the Kernel is term-document matrix (D) multiplied with its transpose:

$$K = DD^T$$

The term-document matrix contains the term frequencies. In our case, the Kernel matrix will be the indicator variable matrix multiplied by its transpose.

Regression: All the attribute combinations are processed with varying values of λ. A prediction is made for every contract using 10-fold cross validation. The results for the best three and the worst three predictions are listed in the last four columns of Table 2. The error distribution of the best attribute combination is shown in Figure 3d.

RESULTS

Table 3 shows results of the experiments performed by Vector Space Model and Kernel Ridge Regression the results are broadly similar. The results for KRR are slightly better than VSM when enhanced with outlier elimination and weighted nearest neighbour, but not significantly so.

Figure 4 shows the error distribution of KRR plotted against error distribution of VSM and enhanced VSM.

The most encouraging result from implementing both VSM and KRR is the fact that they both give their best result on the same attribute combination and their top 3 attribute combinations have the same attributes as evident in column 2 and 5 of tables 2. The fact that they perform badly on different attribute combinations, is of no relevance. We can thus make a decision on which

Table 3: Performance comparison

Method	Mean Absolute Error	Median Absolute Error
Vector Space Model	5.68	4.76
Vector Space Model & outlier elimination	5.41	4.28
Vector Space Model, outlier elimination, & weighted nearest neighbour	5.09	4.17
Kernel Ridge Regression	5.01	4.13

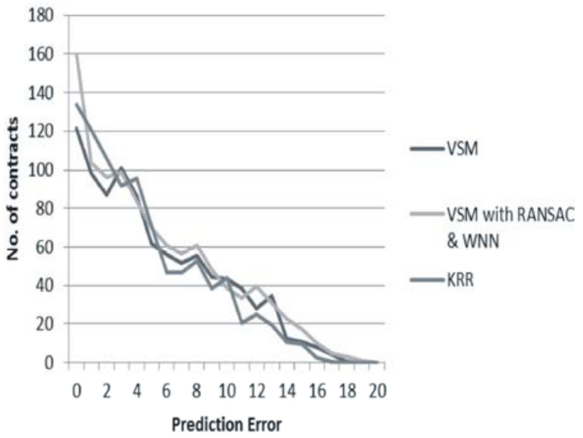

Fig. 4: Error distribution of VSM enhanced VSM and KRR

attributes contribute towards profitability and which have no effect. The attributes that influence contract profitability are: location, group, manager, QS, contract type and client.

CONCLUSION AND FURTHER WORK

As agriculture turns to agribusiness around the world, the role of agribusiness construction is increasing. The paper presents a Machine Learning approach to prediction of profitability in agribusiness construction contracts. The estimation of a prospective contract's profitability is often done by intuition by a commercial manager. A mathematical model to aid in predicting the profitability of a prospective contract would be of immense use to a commercial manager. Some commercials managers may use it as the primary source of predicting profitability whilst others may use it as a valuable second opinion. Furthermore, it would of considerable interest to commercial managers to know the effect on predicted profitability of a contract should they change the value of an attribute of a prospective contract. Both the VSM and KRR routines are fairly simple to implement in a commercial setting. Commercial application will require close interaction between scholars in the field of agricultural sciences, computer science and business. For quite some time research in business has been becoming increasingly quantitative and business scholars in the domains of Economics, Finance, Human Resource, Marketing and Industrial Organisation, among others, are aspiring to achieve the same standards of academic excellence that hard disciplines demand [26]. At the same time in the machine learning community there is greater realisation that research needs to be more relevant to business [27]. While the model measures accounting profits, one possible extension of the model could be to measure economic profits. Often the companies shift costs and revenues from one project to another or from one year to another for the purpose of tax accounting. It may be possible to extend the model to take this activity into account.

On the technical side, the system could be tested against another data source with a smaller -10% to 10% or -8% to 8% profitability range. Such a system should be able to predict with a mean/median absolute error of 2 or 3. The system could be extended to provide monthly predictions - effectively acting as a source of cost and revenue forecasting. Finally, when contract profitability prediction system is implemented, the Vector Space Models could be used to find the most similar contracts to the prospective contract which we are predicting. A routine can then examine the top similar contracts and can make recommendations about suppliers and subcontractors to use for the prospective contract to maximise profitability. We have moved on from descriptive analytics of profitability in agribusiness construction projects to predictive analytics, i.e., "what would be profitability like if...". It would require a lot more further research in algorithmic game theory and other converging branches of business and computer science to reach the level of prescriptive analytics in this field.

ACKNOWLEDGEMENTS

The author is thankful to J. Drèze, S. Khorana, T. Sargent, A. Sen and three anonymous referees for helpful suggestions, R. Sharma and M. Dubey for programming in Progress and running the experiments and P. Salvatierra and J. Smith for research assistance.

REFERENCES

1. Jaselskis, E.J., R.L. Wilson and A. Ladson, 1997. Developing an International Agribusiness Construction Information System. Midwest Agribusiness Trade Research and Information Center. Iowa State University.
2. Laryea, S. and W. Hughes, 2011. Risk and Price in the Bidding Process of Contractors. Journal of Construction Engineering and Management, 137(4): 248-258.
3. Cooke, B. and P. Willaims, 2009. Construction Planning, Programming and Control 3rd Edition. Wiley-Blackwell.
4. Jaselskis, E. and A. Talukhaba, 1998. Bidding Considerations in Developing Countries. Journal of Construction Engineering and Management, 124(3): 185-193.
5. Asthana, A.N., 2012. Decentralisation and corruption revisited: evidence from a natural experiment. Public Administration and Development 32(1): 27-37.
6. Wang, S.Q., M.F. Dulaimi and M.Y. Aguria, 2004. Risk management framework for construction projects in developing countries. Construction Management and Economics, 22(3): 237-252.
7. Asthana, A.N. and S. Khorana, 2013. Unlearning Machine Learning: The Challenge of Integrating Research in Business Applications. Middle-East Journal of Scientific Research, 15(2): 266-271.
8. Oztekin, A., 2012. An Analytical Approach to Predict the Performance of Thoracic Transplantations. Journal of CENTRUM Cathedra, 5(2): 185-206.
9. Corts, K.S., 2012. The interaction of implicit and explicit contracts in construction and procurement contracting. Journal of Law, Economics and Organization, 28(3): 550-568.
10. Harris, F. and R. McCaffer, 2013. Modern Construction Management 7th Edition. Wiley-Blackwell.
11. Jarvenpaa, S.L. and D.B. Stoddard, 1998. "Business Process Redesign: Radical and Evolutionary Change," Journal of Business Research, (41:1): 15-27.
12. Davenport, T.H., 2000. Mission Critical, Boston: Harvard Business School Press.
13. Fui-Hoon Nah, F., J. Lee-Shang Lau and J. Kuang, 2001. Critical factors for successful implementation of enterprise systems. Business Process Management Journal, 7(3): 285-96.
14. Chan, E., 2009. Knowledge management using enterprise resource planning (ERP) system, doctoral thesis, RMIT University, Melbourne.
15. Voordijk, H., A. Van Leuven and A. Laan, 2003. Enterprise Resource Planning in a large construction firm: implementation analysis. Construction Management and Economics, 21(5): 511-521.
16. Davenport, T.H., S.L. Jarvenpaa and M.C. Beers, 1996. Improving Knowledge Work Processes. Sloan Management Review, 37(4): 53-56.
17. Johns, G., 2006. The Essential Impact of Context on Organizational Behavior. Academy of Management Review, 31(2): 386-408.
18. Peppard, J. and J. Ward, 2005. Unlocking Sustained Business Value from IT Investments. California Management Review, 48(1): 52-70.
19. Martin, T. and Z. Huq, 2007. Realigning Top Management's Strategic Change Actions for ERP Implementation: How Specializing on Just Cultural and Environmental Contextual Factors Could Improve Success. Journal of Change Management, 7(2): 121-142.
20. Bengio, Y. and Y. Grandvalet, 2004. No unbiased estimator of the variance of k-fold cross-validation. The Journal of Machine Learning Research, 5: 1089-1105.
21. Singhal, A., 2001. Modern Information Retrieval: A Brief Overview. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 24: 35-43.
22. Aizawa, A., 2003. An information-theoretic perspective of tf–idf measures. Information Processing and Management, 39(1): 45-65.
23. Barnett, V. and T. Lewis, 1994. Outliers in Statistical Data 3rd edition. John Wiley.
24. Todorov, V., M. Templ and P. Filzmoser, 2011. Detection of multivariate outliers in business survey data with incomplete information. Advances in Data Analysis and Classification, 5(1): 37-56.
25. Riani, M., A. C. Atkinson and A. Cerioli, 2009. Finding an unknown number of multivariate outliers. Journal of the Royal Statistical Society: series B (statistical methodology), 71(2): 447-466.
26. Asthana, A.N., S. Mohan and S. Khorana, 2012. Physics Envy and Natural Experiments in Business and Economics. World Applied Sciences Journal, 20(3): 464-469.
27. Wagstaff, K.L, 2012. Machine Learning that Matters. Proceedings of the 29th International Conference on Machine Learning, Edinburgh.