

Article

Not peer-reviewed version

Machine Learning for Profitability Prediction in Agribusiness Construction: A Novel Approach Using Vector Space Model and Kernel Ridge Regression

[Mary Smith](#) *

Posted Date: 13 September 2024

doi: 10.20944/preprints202409.1043.v1

Keywords: Agribusiness Construction; Contract Profitability Prediction; Machine Learning; Vector Space Model; Kernel Ridge Regression (KRR); Driven Decision Making



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Machine Learning for Profitability Prediction in Agribusiness Construction: A Novel Approach Using Vector Space Model and Kernel Ridge Regression

Mary C Smith

* Correspondence: marycsmithuk@gmail.com

Abstract: In the dynamic realm of agribusiness construction, firms are striving to bolster their operational efficiency and profitability amidst the global shift from traditional farming to commercial agriculture. This transition has intensified the demand for sophisticated infrastructure development, presenting new challenges for commercial managers. A critical hurdle is the accurate estimation of profitability for prospective contracts, a task often reliant on intuition rather than data-driven methods. To address this, we propose the development of a mathematical model utilising machine learning techniques to predict contract profitability and identify influential factors. This model aims to aid in bid decision-making, financial forecasting, and enhancing competitiveness in the marketplace. Furthermore, it would provide valuable insights into how altering specific contract attributes might affect predicted profitability. The implementation of such a system necessitates close collaboration between IT professionals and construction executives. This study delineates the development of this predictive model, outlining the data analysis process and the application of machine learning algorithms to tackle this complex commercial challenge. The ultimate goal is to bridge the gap between intuitive estimation and data-driven prediction, thereby enhancing the financial performance and strategic decision-making capabilities of firms in the agribusiness construction sector.

Keywords: agribusiness construction; contract profitability prediction; machine learning; vector space model; Kernel Ridge Regression (KRR); driven decision making

Introduction

In the global agricultural landscape, construction plays a pivotal role in developing robust agricultural economies, particularly in emerging markets. This is evident in the need for effective infrastructure for producing and storing agricultural products. Agricultural construction encompasses a broad spectrum of projects, ranging from major initiatives that have immediate effect on cultivators' work capabilities, such as barns, silos, seed and grain processing facilities, and livestock production units, to secondary projects involving essential infrastructure like large warehouses and transport networks. As the agricultural sector evolves, construction firms specialising in agribusiness face the challenge of estimating expected profits on prospective contracts in an increasingly competitive market. This crucial task informs bidding decisions, including whether to pursue an agreement and the nature and amount of the offer. Accurately forecasting profitability remains essential for informed decision-making prior to submitting any bid. The transformation of the agricultural sector in developing economies has ushered in a new era where profit is no longer considered taboo, necessitating more sophisticated approaches to contract management and profitability prediction in agribusiness construction. Because contracts come in a variety of sizes, shapes, and forms, as well as from different types of employment, it can be challenging to estimate how profitable a potential contract will be. Moreover, while some agribusiness construction businesses focus exclusively on a single kind of work, others accept a wide range of projects of all sizes. Furthermore, the client's attitude would undoubtedly affect a contract's profitability. Some may be less strict because of internal issues.

The profitability of a contract is significantly impacted by internal management. Profitability is impacted by how well the employees allocated to the construction project perform. The profitability of a contract is also influenced by personnel availability, productivity, and suppliers. Moreover, on medium-to large-scale contracts, the majority of agriculture construction companies use subcontractors, or other businesses, for more than half of the work, and occasionally for the majority of the job. If subcontractors' performance is not properly controlled, it can have a significant impact on a contract's profitability.

Finally, in addition to contract types and internal management, unanticipated events may have an impact on a contract's profitability. For instance, the timely completion of a contract and the availability of labour may alter due to new municipal or government initiatives. A sudden increase in oil prices will drive up the cost of a contract if it requires specialised components from a far-off source. If the extra cost cannot be passed on to the client, the profitability will be negatively impacted. Since the majority of the work is done "off road," there are more unknowns in agricultural building. Agribusiness construction businesses are expanding their operations into developing nations as a result of the globalisation of the agribusiness industry.

Businesses now face more risk as a result of this internationalisation since developing nations present more uncertainty. Approvals and permits, modifications to laws and policies, law enforcement, the creditworthiness of local partners, political unpredictability, increased inflation, fluctuating interest rates, and government influence over dispute settlement are additional risk concerns.

Traditional programming methods using if-then-else logic are inadequate for predicting contract profitability due to the complexity and multitude of variables involved. Machine Learning has emerged as a valuable tool in business research to address such challenges. While commonly used in finance and marketing, machine learning is now finding applications in sectors like healthcare. The system's goals are twofold: to forecast the expected profitability of contracts at their inception and to identify the key contract attributes that significantly impact profitability. The absence of existing models in this specific area necessitated the creation of an original system. The paper outlines the developed system and the data analysis process, demonstrating how modern methods can be applied to solve a real-world problem.

Management of contracts

Construction contracts typically fall into two categories: fixed-price and cost-plus. Fixed-price contracts incentivize cost reduction but may lead to renegotiation issues when project modifications are needed. Cost-plus contracts offer flexibility as the client maintains project control, but they provide little motivation for cost-saving since the construction company recovers all expenses.

In agribusiness construction, Quantity Surveyors (QS) evaluate contract values and oversee costs, revenues, and unexpected issues that might impact profitability. They regularly submit Cost Value Reconciliation (CVR) reports to keep management informed about contract progress. Commercial managers, often experienced QSs, aid in bidding for new contracts and managing ongoing ones. The QS's role involves monitoring financial aspects, addressing unforeseen circumstances, and providing periodic updates to management through CVR reports.

Using Enterprise Resource Planning (ERP) systems became one of the most widespread organisational reform initiatives of the final decade of the 20th century. ERP allows a business to manage the effective and efficient use of resources (materials, human resources, money, etc.). The software's architecture makes it easier for modules to integrate transparently and provide information flow that is consistently visible across all processes inside the construction organisation. Construction companies can replace or re-engineer their generally incompatible legacy information systems with a single integrated system by utilising corporate computing in conjunction with ERP systems.

Implementing an ERP system in a large agribusiness construction firm is a significant undertaking, typically spanning one to three years and costing substantial amounts. Many industry experts believe that ERP implementations in large construction firms often result in more failures

than successes. ERP systems profoundly impact employees, altering task nature, workflows, and sometimes entire job roles. Work is tough and frequently requires working around the clock to meet tight deadlines. Project personnel primarily focus on completing work quickly to decrease duration. In this environment, it is challenging for individuals to respond creatively to proposed changes. Major changes inevitably lead to complications. Despite these challenges, an increasing number of agribusiness construction companies are adopting ERP systems, not as an end goal, but as a means to achieve organizational objectives.

The process of managing a contract within an ERP system begins with entering a prospective contract into the Contract Status Ledger. Once the company decides to proceed and legal agreements are finalised, BOQ is introduced which lists items of work. Quantity Surveyor is responsible for updates of the completion percentages. Principal Quantity Surveyor (PQS) of the client reviews these claims. Both amounts are recorded updating the revenues. As work continues, a Procurement module is used to place orders with selected suppliers, automatically updating contract costs. HR & Payroll modules are used for worker payments, also updating contract costs. For subcontracted work, orders are placed via the Subcontract Ledger. Subcontractors follow a similar system.

Methods and Measures

Data: Contracts that were completed with expenses over \$100,000 in US dollars were used for data. There are 934 contracts in all available in the data collection. The data suggests that there were more profitable contracts than loss-making ones, as the distribution is tilted to the right. The good news for this industry is that almost 40% of the contracts have profits between 5% and 14%.

Data Extraction and Preparation: Contracts with profit margins outside the -15% to 15% range were excluded. The data was obtained by exporting the “jc_job” table. This locally recreates the “jc_job” table with an identical structure. “Financial and Contract Status Ledger Reports” get generated for extraction of the costs and revenues. New fields – “prj_cost, prj_rev” and “prj_profperc” - get added to the “jc_job” table. The profitability can be computed based on these figures. With consolidation, straightforward Progress queries are executed on the relevant contracts. For each “jc_job no-lock”:

```
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 */.
```

Contract Characteristics: When a contract arrives into the Ledger, it is associated with various attributes that will be “predictor variables”. Ten highly relevant attributes can be selected. All the chosen attributes are “nominal multinomial”, meaning they consist of alphanumeric codes that cannot be ranked. These attributes remain unchanged. Other operational factors are not included in our calculations since we're predicting profitability for prospective contracts. Our main goal is to anticipate contract profitability, but we also want to know which attributes make a contract profitable. The following are possible pairings of the ten attributes:

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

Cross-Validation Approach: Our approach involves dividing the data into 10 subsets. Nine subsets are utilised as training sets, and each subset is used once as a test set. However, we cannot simply split the data into ten subsamples as they appear in the table. With a view to avoid subsamples which are internally very similar but significantly different from other subsamples. To address this challenge, we randomly assign contracts to subsamples using:

“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".

VSM: To predict a new contract's profitability, we operate under the premise that contracts with similar characteristics are likely to have comparable profitability. For instance, a project involving the demolition of an unused office building and site clearance at a specific place, overseen properly, is likely to have profitability comparable to another project of similar nature managed by the same team a few months later. This assumption is based on the similarities in work type, location, and personnel involved. To identify contracts similar to a prospective one, we employ the Vector Space Model (VSM), a technique commonly used in Information Retrieval for ranking or classifying textual documents. VSM, rooted in linear algebra, transforms "documents" into "vectors of index terms". A method employed to determine similarity is "cosine similarity", as represented by the formula:

$$\text{Cos}(\theta) = A \bullet B / \|A\| \|B\|$$

In the realm of information retrieval, document vectors are typically represented using TF-IDF (Term Frequency – Inverse Document Frequency), a widely adopted statistical weighting scheme used in contemporary information retrieval systems to assess a word's importance within a document or corpus. However, this approach is neither necessary nor applicable in our contract analysis scenario. This is because each contract attribute can only assume a single value, allowing us to represent each contract as a vector of attribute values with a maximum length of ten. Normalisation by magnitude is necessary when calculating cosine similarity since it is possible that an attribute for a given contract may not be provided, meaning its value may be blank or unknown. By computing their cosine similarity, our system finds the contracts that are closest to the one we are trying to anticipate. The forecast for the new contract is then based on the mean profitability of these comparable contracts. This process is applied to all 1,023 possible attribute combinations, with predictions generated for every contract using the 10-fold cross-validation method.

Elimination of Outliers: Addressing outliers is a crucial step in data analysis, with various approaches available. Detecting outliers is particularly challenging. For multivariate data with normal distribution, Mahalanobis distances are typically used as the standard test for outliers. The efficacy of this test is heavily contingent upon which subset is employed for estimation of the parameters of the distribution. In our approach, we employ Random Sample Consensus (RANSAC) to identify outliers. RANSAC is an iterative method that repeatedly selects random subsets of the data to determine the outliers. Our system identifies the most similar contracts to the one being predicted, applies outlier elimination using RANSAC, and then calculates the mean profitability.

Weighted Nearest Neighbour: The mean and median absolute error of the Vector Space Model's output are both improved by doing outlier removal. We are aware that most contracts have a profitability of between 5% and 8%. We make use of this information by giving contracts that fall inside the range more weightage. We use the weighted mean rather than the mean of the remaining inliers.

$$\frac{\sum w_i x_i}{\sum w_i}$$

The method finds the contracts that are most similar to the one we are attempting to predict, eliminates outliers, and uses the weighted mean of the inliers that remain to determine the expected profitability. Using 10-fold cross validation, the system generates forecasts for each contract. 5.09 is the mean absolute error, and 4.17 is the median absolute error.

KRR: Predictions can then be made using a system with weights that has been trained using regression.

$$X_w = Y$$

whereas OLS can be used to determine the ideal weight w value:

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

When $(X^T X)^{-1}$ is nonexistent or the inversion is numerically unstable, ridge regression can be helpful. Overfitting occurs if there is noise rather than the primary association, which is a common regression problem. A popular method to address this problem is to include a regularizer (λ). This promotes weight values to decay towards zero, much like in a sequential learning system, unless they are backed by data. The ideal “weight vector” value is determined using L training samples as follows:

$$w = (X^T X + \lambda I_n)^{-1} X^T y$$

$$w = \lambda^{-1} X^T (y - X w) \quad X^T y = X^T \alpha \quad w = \sum \alpha_i x_i$$

$$\alpha = (X^T X + \lambda I_n)^{-1} y$$

and the following can provide the “prediction function:

$$\langle w, x \rangle = \sum \alpha_i \langle x_i, x \rangle$$

Kernel Functions and Indicator Variables: In order to perform regression, we must convert all of our nominal multinomial predictor variables into binary indicator variables. For every property, a different file is created by the process. The files analysed are horizontally concatenated when the system encounters a specific combination of attributes:

$$\phi : D \rightarrow F, K(d_i, d_j) = \langle \phi(d_i), \phi(d_j) \rangle$$

We will attempt to duplicate the Vector Space Kernel, which is created by multiplying the term-document matrix (D) by its transpose in order to create the kernel:

$$K = D D^T$$

The term frequencies are contained in the term-document matrix; and “indicator variable matrix” multiplied by its “transpose” will be Kernel matrix in this instance. Every combination of attributes is handled using different values of λ . Every contract is predicted using 10-fold cross validation.

Results

When outlier removal and weighted nearest neighbour are added, KRR performs marginally—but not substantially—better than VSM. Their top three attribute combinations all have the same traits, and the most encouraging attribute combination is the same combination. It makes little difference if they perform poorly on various attribute combinations. Thus, we may decide which characteristics have an impact on profitability and which don't. Location, group, management, QS, contract type, and client are the factors that affect contract profitability.

Conclusion and Future Directions

The global shift from traditional agriculture to agribusiness has led to an increasing significance of agribusiness construction. This paper introduces a novel Machine Learning approach for predicting profitability in agribusiness construction contracts. The research demonstrates that estimating a prospective contract's profitability can be based on data-driven methods rather than relying solely on intuition or political considerations. The proposed mathematical model serves as a valuable tool for enterprises, offering an objective means to evaluate contract profitability and potentially mitigate political pressures. This approach provides a robust framework for decision-making, allowing organizations to make more informed choices based on quantitative analysis rather than subjective factors.

Moreover, the model offers significant benefits to commercial managers by enabling them to assess the impact of attribute changes on a prospective contract's predicted profitability. This feature allows for more nuanced contract negotiations and strategic planning, as managers can simulate various scenarios and understand their potential outcomes before committing to specific terms. The simplicity of implementing both the Vector Space Model (VSM) and Kernel Ridge Regression (KRR) routines in commercial settings makes this approach particularly attractive for practical application.

These methods can be integrated into existing systems with relative ease, providing a powerful analytical tool without requiring extensive overhauls of current processes.

However, the successful application of this model in enterprises will necessitate collaborative efforts across multiple disciplines. It calls for close cooperation between experts in agricultural sciences, computer science, and business management. This interdisciplinary approach ensures that the model not only leverages advanced computational techniques but also incorporates domain-specific knowledge and business acumen. Future work in this area could explore several avenues:

1. Refinement of the model: Incorporating more sophisticated machine learning algorithms or ensemble methods to improve prediction accuracy.

2. Dynamic attribute weighting: Developing methods to automatically adjust the importance of different contract attributes based on changing market conditions or organizational priorities.

3. Integration with other systems: Exploring ways to connect this predictive model with other enterprise systems for more comprehensive decision support.

4. Expanded data sources: Investigating the potential of including external data sources, such as economic indicators or weather patterns, to enhance prediction capabilities.

5. Adaptation to other sectors: Examining how this model could be modified for use in other areas of agribusiness or different industries altogether.

6. Long-term performance tracking: Implementing systems to monitor the model's predictions against actual outcomes over time, allowing for continuous improvement and validation.

7. Ethical considerations: Addressing potential biases in the model and ensuring its use aligns with ethical business practices and business responsibilities.

By pursuing these directions, researchers and practitioners can further enhance the utility and impact of this machine learning approach in agribusiness construction and potentially beyond, contributing to more efficient and effective decision-making in enterprises.

References

1. Laryea, S., & Hughes, W. (2011). Risk and Price in the Bidding Process of Contractors. *Journal of Construction Engineering and Management*, 137(4), 248–258.
2. Cooke, B., & Willaims, P. (2009). *Construction Planning, Programming and Control* 3rd Edition. Wiley-Blackwell.
3. Asthana, A. N. (2014). Thirty years after the cataclysm: toxic risk management in the chemical industry. *Journal of Toxicological Sciences*, 6(1), 01-08.
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. (2024). The Mechanism of Stress-Reduction Benefits of Yoga for Business Students. *The Seybold Report*, 19, 198-208.
6. Wang, S. Q., Dulaimi, M. F., & Aguria, M. Y. (2004). Risk management framework for construction projects in developing countries. *Construction Management and Economics*, 22(3), 237-252.
7. 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.
8. Asthana, A. N. (2023). Wastewater Management through Circular Economy: A Pathway Towards Sustainable Business and Environmental Protection. *Advances in Water Science*, 34(3), 87-98.
9. Harris, F., & R. McCaffer, R. (2013). *Modern Construction Management* 7th Edition. Wiley-Blackwell.
10. Davenport, T. H. (2000). *Mission Critical*, Boston: Harvard Business School Press.
11. Jarvenpaa, S. L., & Stoddard, D. B. (1998). Business Process Redesign: Radical and Evolutionary Change. *Journal of Business Research* 41(1), 15-27.
12. Asthana, A. N. (2022). Impact of mindfulness on irrigation water consumption. *Frontiers in Water*, 4. <https://doi.org/10.3389/frwa.2022.1062835>
13. Nah, F., Lau, J. L-S, & Kuang, J. (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, Melbourne: RMIT University..
15. Voordijk, H., A., Van Leuven, A. & Laan, A. (2003). Enterprise Resource Planning in a large construction firm: implementation analysis. *Construction Management and Economics*, 21(5), 511-521.
16. Davenport, T. H., Jarvenpaa, S. L., & Beers, M. C. (1996). Improving Knowledge Work Processes. *Sloan Management Review*, 37(4), 53-56.

17. Asthana, A. N. (2023) Determinants of Cultural Intelligence of Operations Management Educators. *The Seybold Report*, 18(6), 789-800.
18. Johns, G. (2006). The Essential Impact of Context on Organizational Behavior. *Academy of Management Review*, 31(2), 386-408.
19. Aizawa, A., (2003). An information-theoretic perspective of tf-idf measures. *Information Processing & Management*, 39(1), 45-65.
20. Asthana, A. N. (2023). Prosocial behavior of MBA students: The role of yoga and mindfulness. *Journal of Education for Business*, 98(7), 378-386.
21. Martin, T., & Huq, Z. (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.
22. Asthana, A. N., & Tavželj, D. (2022). International Business Education Through an Intergovernmental Organisation. *Journal of International Business Education*, 17, 247-266.
23. Peppard, J. & Ward, J. (2005). Unlocking Sustained Business Value from IT Investments. *California Management Review*, 48(1), 52-70.
24. Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *The Journal of Machine Learning Research*, 5, 1089-1105.
25. Asthana, A., & Asthana, A. N. (2012). Yogic science for human Resource management in business. *World Applied Sciences Journal*, 19(1), 120-130. <https://doi.org/10.5829/idosi.wasj.2012.19.01.6191>
26. Singhal, A. (2001). Modern Information Retrieval: A Brief Overview. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 24, 35-43.
27. Barnett, V., & Lewis, T. 1994. *Outliers in Statistical Data 3rd edition*. John Wiley.
28. Asthana, A. N. (1999). [Rezension] Financing decentralized expenditures, Ehtisham Ahmad(ed.): Cheltenham, UK[u. a.], Elgar, 1997. *Kyklos*, 52(1), 103-104.
29. Asthana, A. (1998). Intergovernmental fiscal relations, Ronald C. Fisher(ed.): Boston[u. a.], Kluwer, 1997. *Kyklos*, 51(4), 595-596.
30. Riani, M., Atkinson, A. C. & Cerioli, A. (2009). Finding an unknown number of multivariate outliers. *Journal of the Royal Statistical Society*, 71(2): 447-466.
31. Asthana, A. N. (2014). Profitability Prediction in Cattle Ranches in Latin America: A Machine Learning Approach. *Global Veterinaria*, 13(4), 473-495.
32. Todorov, V., Templ, M., & Filzmoser, P. (2011). Detection of multivariate outliers in business survey data with incomplete information. *Advances in Data Analysis and Classification*, 5(1), 37-56.
33. Asthana, A. N. (2015). Sustainable Fisheries Business in Latin America: Linking in to Global Value Chain. *World Journal of Fish and Marine Sciences*, 7(3), 175-184.
34. Nunnally, S. W., & Nunnally, S. W. (2001). *Construction methods and management* (Vol. 3). Prentice Hall.
35. Forcael, E., Ferrari, I., Opazo-Vega, A., & Pulido-Arcas, J. A. (2020). Construction 4.0: A literature review. *Sustainability*, 12(22), 9755.
36. Mohan, S. (2021). Non-tariff measures a trade barrier for developing countries' agricultural processed products exports. *Public Enterprise*, 25(1-2), 1-17. <https://doi.org/10.21571/pehyj.2021.2512.01>
37. Tomlinson, M. J., & Boorman, R. (2001). *Foundation design and construction*. Pearson education.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.