MAP: A Versatile Platform for Predictive Algorithms in Construction and Beyond

Ramyani Sengupta[1], Emad Elwakil[1] and Yi Jiang[1]

[1]School of Construction Management Technology, Purdue University, USA
sengup11@purdue.edu, eelwakil@purdue.edu, jiang2@purdue.edu

Abstract– The construction industry in North America is a dynamic and rapidly evolving sector characterized by its fast-paced operations and constantly shifting demands. Despite its significance, the industry has been slow to adopt emerging technologies compared to other fields. This lag extends to the integration of artificial intelligence (AI), where vast opportunities for innovation remain underutilized. While large, well-funded firms have begun leveraging AI to develop proprietary applications capable of analyzing historical data to predict patterns and trends, smaller firms often encounter significant barriers, including limited resources and technical expertise, which hinder their ability to adopt such advanced tools.

This study introduces a robust, user-friendly platform that integrates commonly employed machine learning algorithms. The platform empowers smaller firms by enabling them to train, validate, and test their data effectively without requiring extensive technical expertise. Beyond addressing the immediate needs of the construction sector, the platform is intentionally designed for versatility and adaptability, making it applicable across various industries. It supports tasks such as pattern recognition and data visualization, providing a reliable solution for non-experts to evaluate and select the most suitable models for their unique datasets and variables.

Keywords – Modeling, Algorithms, Machine learning; Simple Neural Network, Multi-Layer Perceptron, Support Vector Machine, Random Forest, Regression, Construction

1 Introduction

The construction industry in North America has become a dynamic and rapidly growing sector, with its profits soaring to an impressive $1.8 trillion in 2023 [1] and contributing to 4.4% of the GDP in the United States in 2023 [1]. This remarkable growth highlights the industry's critical role in driving regional economic development and infrastructure advancements. However, in an environment of increasing competition and evolving market demands, every organization within the industry must strive to carve out a distinctive edge that sets it apart from its closest competitors. This differentiation often hinges on how companies embrace and integrate innovative technologies into their operations. Technology adoption has emerged as a key factor in determining which firms can stay ahead of the curve, enhancing overall performance [2]. In an industry where traditional methods and opinions can still hold considerable sway in many areas, the construction industry has often fallen behind in this area [3], and today, the ability to leverage cutting-edge tools and systems can be the most defining characteristic of successful organizations.

Alongside automation, robotics, and drone technology, to name a few, artificial intelligence or 'AI' algorithms has become a quickly growing technology across all industries. Adopting it within the construction industry can help improve efficiency, safety, and overall project outcomes by streamlining project planning, resource management, and real-time decision-making. Currently, available AI tools can improve performance, make firms more competitive, and thus give them an 'edge' [4]. As more and more US construction companies invest in in-house AI algorithms for their historical data to predict project timelines and future issues, this paper aims to streamline the process by creating a platform with different ML algorithms. As it can be challenging to determine which algorithm works best for different types of data, the 'Modelling and Prediction Platform ('MAP' platform) application can help determine the best-fit model for the data. Users can run their data through MAP and easily choose among six different machine learning algorithms (a simple neural network, multi-layer perceptron, support vector machine with five different kernel options, random forest, ridge regression, and lasso regression) to determine which model – with the least amount of error – best fits the data and predicts a pattern between the variables. Certain algorithms have additional parameters that can be used to customize the users' choices. This application can be used for datasets beyond the construction industry, as its visualization, modeling, and predictive techniques are not limited to construction datasets or project management.

2 AI and Construction

Although challenging, multiple attempts have been made to define AI and its function. "*A being concerned with methods of achieving goals in situations in which the information available has a certain complex character*" was used by McCarthy in a 1988 textbook as an attempt to define artificial intelligence. Since then, its definition has constantly evolved, and its functions have surpassed expectations. Although it can still be termed a comparatively modern form of technology [5], the number of companies using AI technologies has increased by 270% between 2014 and 2018, almost a three-fold increase [6]. This begs the obvious question today: *if everyone is using it, how does it lend a competitive edge*? Mikalef and others, in their 2021 study, argue that in-house tools, i.e., AI capabilities developed by firms themselves, can enhance performance and give companies an edge [4]. Such engineered solutions can allow project managers to track project progress, resolve conflicts and issues, and make informed decisions from any location. A good example of such 'in-house' tools is using historical data for predictions. Some of the heavyweight construction firms in the country, such as Turner Construction, Suffolk Construction, DPR Construction, and AECOM, among others, routinely train AI algorithms nowadays on their proprietary company data from previous decades to predict future scheduling timelines, delay causes, project performance, etc. The idea is that such algorithms can quickly be trained to analyze historical data to detect current issues a*nd predict potential issues* in construction projects. Many of the top 10 construction companies in the US (big names in the industry such as Turner, AECOM, Suffolk, etc.) have investing 'wings' within their firm that heavily research these topics and then infuse capital into startups focusing on specific tools that can improve their projects. This can help enable proactive planning and resource allocation within the company *before* such issues arise. Due to budget constraints and lack of manpower, smaller names often fall behind, and the tool described in this study may help them compete.

2.1 Background on AI Adoption

Case studies can significantly help strengthen the argument that incorporating AI techniques and tools improve company performance and can reiterate the importance of more compact firms having access to such technology. A 2020 study including 150 firms of various sizes investigated the use of eleven types of AI capabilities (generic AI such as task and decision automation, machine learning algorithms, deep learning algorithms, cognitive, cognitive cyber security, natural language processing, robotic personal assistant, pattern recognition, chatbots, neural networks, and virtual companions). The study found that AI use within firms had an incredibly positive result with increased organizational performance at the company level, and process-related variables also improved the company [7]. Professional readiness and AI technologies correlate as well [8]. Another 2022 study investigated 299 Scandinavian countries and found that AI platforms strongly impact organizational performance, whether in the social aspect or financial performance [2]. A 2021 study among 530 employees across Indian telecommunication companies (Vodafone, Idea Telecom, Bharti Airtel, Reliance Jio, etc.) reported that using AI applications and algorithms helps managerial employees get real-time data on employee performance quite effectively and that automated processes and other AI tools can influence performance. Companies can "overachieve", "deliver sustainable results", "optimize resources better", and "do the right thing" far more often if they use AI tools to drive performance [9]. Another empirical study from 2022 among 380 employees of Dubai-based companies found that AI use had positive relationships with organizational structure, process innovation, and employee expertise/performance.

When combined with onsite tools/field work, AI technologies can use technologies like tracking and computer vision algorithms for remote inspection while also maintaining quality assurance/quality control (QA/QC) [10]. Construction robotics and automation also transform how resources are used on construction sites [11]. Robotic systems can perform repetitive or dangerous tasks, such as bricklaying, concrete pouring, or welding, with greater precision and speed than human workers [10, 12, 13]. This increases productivity and reduces the risk of human error and accidents. Today, a combination of sensors and algorithms can be used for cranes (e.g., versatile.ai), which connect to a local IoT system for easy tracking by all stakeholders in a project. Currently, versatile.ai is almost unanimously favored by the top construction companies in the US for site work and remote supervision.

AI technologies, thus, show promise when used for a myriad of onsite and offsite work in construction. It is no surprise that large firms across North America are spending millions of dollars on AI tools for all kinds of construction-related variables – from office productivity to onsite safety and PPI use. As discussed earlier, not all firms are large enough to fund decades of AI research, and the platform discussed in this paper can help smaller firms overcome the labor of building complex tools for their historical project data – a platform like MAP can help them determine which algorithms can help predict and highlight any existing patterns in their datasets.

3 The 'MAP' Platform

Six algorithms were chosen for this platform due to their already well-established status in data analysis – they are well-documented in literature and can work for varying data types. The application, named 'Modelling & Prediction Platform' or 'MAP,' displays an error histogram and a graph between the actual data and predicted data alongside the Mean Squared Error and R-squared coefficient values. This can be used to predict future technology adoption within the industry. The findings can help offer a new appreciation for the challenges and opportunities that may be analogous to adopting AI in construction. The application is currently in its testing phase, and a case study conducted with MAP and real construction data is ongoing and thus is beyond the scope of the current discussion.

3.1 Algorithms in MAP

Keeping in mind the dynamic needs of the construction industry, the following six algorithms have since been encoded into the MAP UI (written in Python ver3.12.6, with sci-kit-learn and Matplotlib – discussed later in this section):

- Simple Neural Network
- Multi-layer Perceptron
- Support Vector Machine (with five kernel functions the user can choose from – linear, polynomial, sigmoid, precomputed, and radial basis function or rbf)
- Random Forest
- Ridge Regression
- Lasso Regression.

A Simple Neural Network is great for learning patterns in data that traditional algorithms might miss. Its main advantage is flexibility—it can adapt to many types of data, like images, text, or numerical data [14] (*in the current version of the MAP UI, however, only numerical data can be used as input and/or training and/or test variables*). It is also quite robust to noise, meaning it can handle data that is not perfectly clean (*again, it is advised to use cleaned data for the MAP platform for good visualizations*). With the right training, it can generalize well to predict new, unseen data. The Multi-Layer Perceptron can be considered a powerful extension of simple neural networks. Its biggest benefit is its ability to model complex relationships by using multiple layers of interconnected nodes. MLPs can solve problems that are not linearly separable, making them ideal for tasks like image classification or natural language processing [15]. They also perform well with large and diverse datasets, which can be helpful for construction companies.

Support Vector Machines, or SVMs, are highly effective for classification and regression tasks. SVMs are particularly good with small to medium-sized datasets and are less prone to overfitting. One of their strengths is the flexibility provided by kernel functions, which lets them work well with both linear and non-linear data [16, 17]. Kernel functions help to convert input data into the required hyperplane/ multidimensional space, and the five kernels that are included in the current version of the MAP model (that can be chosen with ease by the user, depending on their dataset, variable types and/or any modeling requirements) are:

- The linear kernel is an ideal option for simple, linearly separable data,
- The Polynomial kernel excels at capturing complex relationships,
- The sigmoid kernel mimics the activation function in neural networks,
- The precomputed kernel allows customization based on prior knowledge and
- The Radial Basis Function (RBF) is versatile for non-linear patterns.

Random Forest is a versatile and reliable algorithm with extensive uses in classification and regression. Its primary benefit is robustness – it combines the predictions of many decision trees to improve the overall accuracy of the output and reduce overfitting. It also does well at handling missing data and can work with numerical and categorical variables [17]. Additionally, Random Forest provides feature importance scores, helping identify which features matter most [17, 18]. Ridge Regression improves upon linear regression by adding a penalty to large coefficients, and its key benefit is reducing overfitting, which makes it ideal for datasets with many correlated features [19]. This regularization ensures that the model generalizes new data better. Ridge Regression can be particularly useful when all input features contribute to the outcome but need their influence 'moderated' to a certain extent [20]. The final algorithm, called 'Lasso Regression,' is similar to Ridge Regression but has an additional benefit: it can select features by shrinking some coefficients to zero. This makes it a great choice for simplifying models and improving interpretability, especially when dealing with datasets that have many irrelevant features [21]. Thus, the model can be computationally efficient while excelling in identifying the most important variables. All machine learning models in MAP were encoded into the platform using 'sci-kit-learn' (also referred to as 'sklearn'), a freely available and easy-to-use open-source machine learning library kit for the Python programming language, and all visualization techniques were incorporated into the platform by using 'Matplotlib' (a blend of 'MATLAB' – the popular programming environment, 'plot', and 'library') which is another commonly used free, open-source plotting library for the Python programming language. The user interface was built entirely with Python (the code editor was Visual Studio Code). The user interface (or 'UI') was built to import a comma-separated values (csv) file, read and identify each variable from each column, and then automatically display them in a window for the user to choose input and output variables. The following section discusses the UI in detail.

3.2 Methodology of MAP

The following sections outline the processes behind the user interface design, each option in the application window, the algorithms and their testing alongside the function of each output window.

3.2.1 The UI

The MAP platform is designed to take in a .csv file as an entry (toggle *1. Select .csv file*), parsing the variable list and automatically identifying the variables. Once the file is chosen, the window beneath this option will display a complete list of all the variables. The user can then choose which variables to choose as 'input' options and which elements are to be displayed as 'output' options, and they appear in the next two terminals (underneath the buttons '*2. Choose input variables*' and '*3. Choose output variables*'). A list of algorithms is provided on the right (*4. Choose a model*) from which the user can select any option to train and validate data from the .csv file of their choosing. The figure below shows what the interface of the modeling platform looks like:

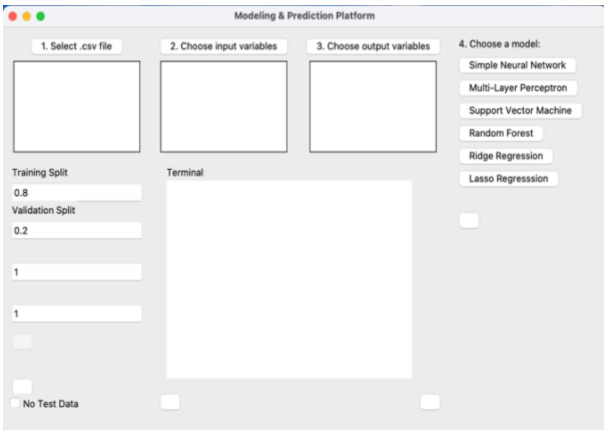

Figure 1. The MAP Platform

As discussed earlier, certain algorithms require the user to make further choices. Additional parameters appear in windows like the one below (Figure 2 below shows an example if the user chooses the Support Vector Machine option):

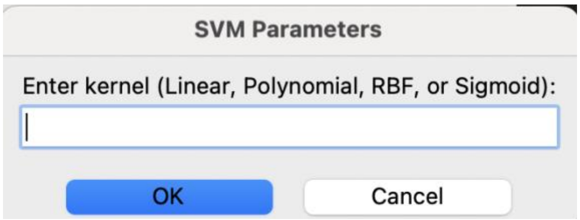

Figure 2. Supplementary parameters for some algorithms

The platform also provides a terminal window (Figure 1) that displays the mean squared error and R-squared values to the user. This can be particularly useful to quickly check if the chosen model suits the user's data or if the parameters need to be customized to fit their needs. Since values for both the training and validation datasets are displayed, it can help the user understand if a different modeling algorithm should be tested.

3.2.2 Model testing and validation

As seen in Figure 1, the platform allows each dataset to be split into two sets for training and validation. Depending on the user and the type of model they choose, one can split the data into any ratios (*they must add up to 1.0 to utilize the entire dataset – in the example figure (Figure 1), the ratios are 0.8 and 0.2*). For example, historical project data can be used for training, while current data can be used for the output.

3.2.3 Output windows

The final output is divided into two graphs, one for the error histogram of the output variable and another for the actual data vs predicted data for the chosen output variable. In the actual output, the error histogram is displayed toward the left of the final output window instead of the top, and the graph of actual vs. predicted values is displayed toward the right side of the output window. Output images from a test training and validation set (using a synthetic dataset for preliminary work) are displayed below in Figures 3 and 4 for a fictional variable from a test dataset:

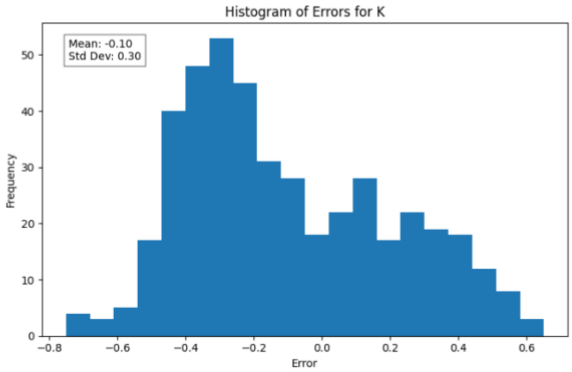

Figure 3. Test output showing an error histogram for a fictional variable '*K*' from a test dataset

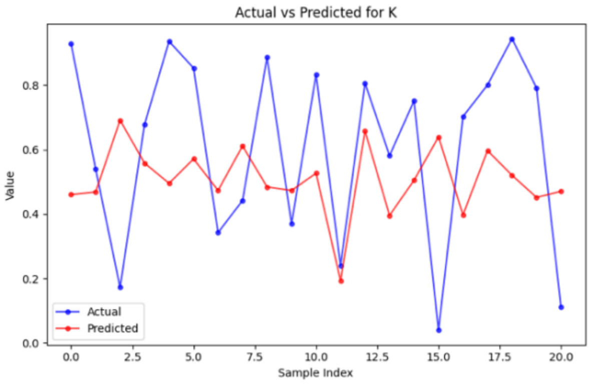

Figure 4. Test output showing the actual vs. predicted data plot for a fictional variable '*K*' from a test dataset

The test dataset shows a weak resemblance between the actual and predicted values, as the dummy dataset does not have identifiable relationships between the variables and were used purely for testing each algorithm. Data from the error histogram was used as one of many identifiers of whether the algorithm is a good choice for the dataset. The values of standard deviation and mean squared and R-squared error (pictured in the top left corner of the error histogram – Figure 3 alongside the terminal window – in Figure 1) also contribute to the user being able to determine of the model is a good fit for their chosen dataset.

4 Conclusion

In stark contrast to other industries, despite such clear positives, stakeholders and investors in this industry are still quite unsure of the potential benefits and overall financial worth of the construction and AI team-up [22]. As the industry changes and adjusts, investigating and investing in AI technologies will be crucial to maintaining competitiveness and meeting the increasing demand for smarter and greener projects. The future of construction lies in the successful adoption of AI, making this an essential area of exploration for the industry. Accurate information from real employees in the industry is crucial for researchers to gain more perspective on the usefulness of AI in construction.

While addressing the current needs of the construction industry, the platform's design is also simple, versatile, customizable, and adaptable, allowing it to be used in a wide range of other industries for tasks such as pattern recognition and data visualization. The platform's customization options also enable users to tailor its functionalities to suit their specific datasets and variables. This flexibility makes it particularly appealing for users who are not overly familiar with ML algorithms, as it provides an accessible way to experiment with different models and determine which one performs best for their unique business needs. The simplicity of the interface also ensures that using it requires little training and can be mostly intuitive. By bridging the gap between AI capabilities and practical usability, this platform offers smaller firms an opportunity to compete on a level playing field with their larger counterparts while encouraging broader AI adoption across various industries. Such algorithms hold immense potential to revolutionize small construction companies, as they can provide firms with evidence-based, data-driven insights that were previously only accessible to larger firms. MAP can help more compact companies optimize their project timelines by using their historical data, similar to the larger firms, and then predict any probable delays or other issues. It can also help to improve estimation by identifying patterns (if any) in past projects, helping employees build more accurate and competitive bids. Most importantly, platforms like MAP need not be limited to any singular industry; the application can be used to predict patterns in other firms looking for a simple application to test out predictive algorithms for their data. Future and ongoing work on MAP includes testing and validating it on real-world construction data from construction companies across the continent.

References

[1] Statista. Construction - statistics and facts. Available at: https://www.statista.com/topics/974/construction/#topicOverview (accessed November 19, 2024).

[2] K. Bley, S. F. B. Fredriksen, M. E. Skjærvik, and I. O. Pappas. The role of organizational culture on artificial intelligence capabilities and organizational performance. In *The Role of Digital Technologies in Shaping the Post-Pandemic World*, pages 13–24. Springer International Publishing, 2022.

[3] A. Sawhney, M. Riley, and J. Irizarry. Construction 4.0: An Innovation Platform for the Built Environment. Routledge, 1st edition, 2020.

[4] P. Mikalef and M. Gupta. Artificial intelligence capability: conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information Management*, 58(3), 103434, 2021.

[5] H. Chen, L. Li, and Y. Chen. Explore success factors that impact artificial intelligence adoption on telecom industry in China. *Journal of Management Analytics*, 8(1), 36–68, 2021.

[6] I. O. Pappas, P. Mikalef, M. N. Giannakos, J. Krogstie, and G. Lekakos. Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. *Information Systems and E-Business Management*, 16(3), 479–491, 2018.

[7] S.-L. Wamba-Taguimdje, S. F. Wamba, J. R. K. Kamdjoug, and C. E. T. Wanko. Impact of artificial intelligence on firm performance: Exploring the mediating effect of process-oriented dynamic capabilities. In *Digital Business Transformation*, volume 38, pages 3–18. Springer International Publishing, 2020.

[8] J. Choudrie and E. D. Zamani. Understanding individual user resistance and workarounds of enterprise social networks: The case of Service Ltd. *Journal of Information Technology*, 31(2, SI), 130–151, 2016.

[9] K. Bley, S. F. B. Fredriksen, M. E. Skjærvik, and I. O. Pappas. The role of organizational culture on artificial intelligence capabilities and organizational performance. In *The Role of Digital Technologies in Shaping the Post-Pandemic World*, pages 13–24. Springer International Publishing, 2022.

[10] U. Sahay and G. Kaur. Leveraging AI to drive organizational performance through effective performance. In *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC)*, pages 584–588, Palladam, India, 2021.

[11] Y. Pan and L. Zhang. A BIM-data mining integrated digital twin framework for advanced project management. *Automation in Construction*, 124, 103564, 2021.

[12] N. Emaminejad and R. Akhavian. Trustworthy AI and robotics: Implications for the AEC industry. *Automation in Construction*, 139, 104298, 2022.

[13] C. J. Liang, T. H. Le, Y. Ham, B. R. Mantha, M. H. Cheng, and J. J. Lin. Ethics of artificial intelligence and robotics in the architecture, engineering, and construction industry. *Automation in Construction*, 162, 105369, 2024.

[14] Y. Pan and L. Zhang. Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Automation in Construction*, 122, 103517, 2021.

[15] E. Brown, J. Gao, P. Holmes, R. Bogacz, M. Gilzenrat, and J. D. Cohen. Simple neural networks that optimize decisions. *International Journal of Bifurcation and Chaos*, 15(3), 803–826, 2005.

[16] M. C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7), 579–588, 2009.

[17] S. Suthaharan. Support vector machine. In Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning, pages 207–235, 2016.

[18] S. J. Rigatti. Random forest. *Journal of Insurance Medicine*, 47(1), 31–39, 2017.

[19] L. Breiman. Random forests. *Machine Learning*, 45, 5–32, 2001.

[20] G. C. McDonald. Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 93–100, 2009.

[21] W. N. van Wieringen. Lecture notes on ridge regression. *arXiv preprint arXiv:1509.09169*, 2015.

[22] J. Ranstam and J. A. Cook. LASSO regression. *Journal of British Surgery*, 105(10), 1348–1348, 2018.

[23] C. Merschbrock and B. E. Munkvold. Effective digital collaboration in the construction industry: A case study of BIM deployment in a hospital construction project. *Computers in Industry*, 73, 1–7, 2015.