Understanding and Predicting Project Payment Latency in the Covid-19 Era

Masters Student Paper Competition Submission

Erika Ergart, Rassul Yeshpayev, Sheng Yang Chou, Theo Ginting, Matthew A. Lanham
Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907
eergart@purdue.edu; ryeshpay@purdue.edu; chou101@purdue.edu; tginting@purdue.edu; lanhamm@purdue.edu

ABSTRACT

This study develops an order-to-cash process map predictive solution to better understand home remodeling project pre- and post-covid-19 payment latency. In remodeling projects, the day a project is sold (and past rescission) to the time payment is received or installment is first accepted is defined as "order-to-cash." This time window often has many sequential or overlapping tasks that must be performed before the firm receives its payment. The motivation for the current study is that while order-to-cash is often a challenge to predict and minimize before the covid-19 pandemic, it has been even more challenging for businesses to estimate since the pandemic. The risks of delayed processes and delay customer payments can hurt the company's solvency and financial stability. In collaboration with a national home remodeling company, we develop an order-to-cash process map and redesign their predictive modeling approach to show where the most uncertainty is coming from and provide empirical-based operational recommendations showing how they could reduce order-to-cash not only before the pandemic but also during. Our solution was able to improve predictive accuracy during all periods in our study. We believe practitioners and scholars alike focused on pre-and post-pandemic forecasting, particularly related to accounts receivable or queuing-based problems would find our work valuable.

Keywords: Cash Flows, Payment Latency, Prediction, CatBoost

INTRODUCTION

Despite the overall trend of economic downturn, the home remodeling industry has remained relatively robust. Due to the shift of more people working remotely from their homes, an increasing number of homeowners are partaking in a variety of house improvement projects, both in a DIY (do-it-yourself) manner or with a professional remodeling company. This uptick is the primary driver that sustained the industry, even amidst the pervasive economic downturn. However, one perennial problem plagues the industry - delays. Remodeling projects are inherently prone to variability because of the many intricately interlinked processes, components, and dependencies that in sum complete the project.

Delays earlier in a project can cast a ripple effect onto the completion of any project. Because payment can only be initiated at the completion of any project (i.e., all components of said project are done), these delays pose direct impact toward the time of payment. Subsequently, not only will this impact the bottom line of any remodeling company, but also the firm's financial solvency. Formally referred to as project payment latency, or more colloquially referred to as order-to-cash (O2C), represents the gap between the start of a remodeling project and the occurrence of the payment for the project. An example of a typical order-to-cash process is shown below in Figure 1.

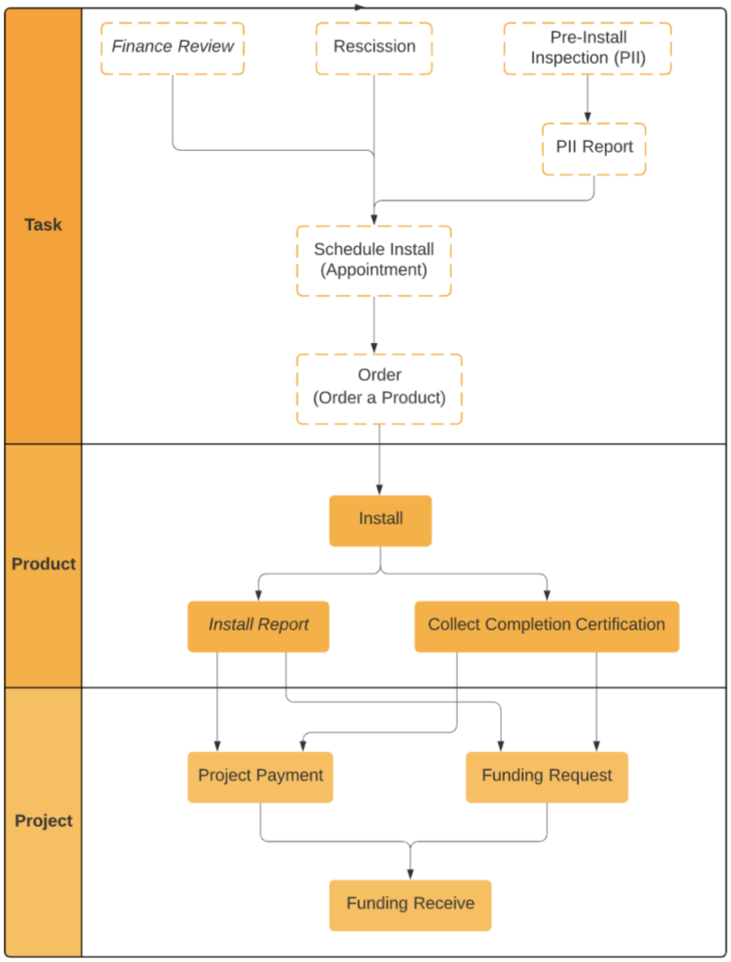

Figure 1: Order-to-Cash process map

The home remodeling industry is poised to reach $352B in project spending in the year 2021 with a sustained growth amidst the pandemic (Will 2021). Given the size of this industry and the deep rootedness of the industry towards the United States economy, it is imperative that firms can thoroughly understand, predict, and even mitigate some of the patency latency risk to increase financial stability. This has always been a relevant and timely problem, even more so in the time of the pandemic where bracing and preparing for uncertainty will be the crux of many industries.

O2C is a challenging variable to predict, yet invariably, it is the core measurable component required to understand and estimate the potential delays in the construction process, as well as strive towards a more stable financial position, eschewing insolvency, especially relevant in this Covid-19 era. Inaccurate O2C forecasts and/or project completion times could erode the valuation of any companies and the trust of stakeholders. Inability to provide accurate project completion time quotes could steer potential customers away. Bearing in mind that we can expect people to stay at their home for longer periods of time, having a project delayed past initial forecasted completion time would create unwanted intrusion toward the customers' daily routine, posing unwanted negative sentiment toward the company. All of these poor planning and transparency consequences stemming from a poor forecast could naturally impede future growth of the firm, especially in the remodeling industry where competition is high, and the industry is in its mature stage.

Another interesting aspect of our research is that the use of advanced analytics and data science is not new in the home remodeling industry. Remodeling and construction might not be the expected industries where one would think predictive modeling is being developed and implemented for on-the-ground decision-support, but we found that it is. In collaboration with a national remodeling company, we utilized a CatBoost model to improve the prediction accuracy of the Order-to-Cash (O2C) time.

Through this collaboration we answer several questions. First, what is the process map of O2C to identify likely drivers that may be required to construct a useful predictive model for payment latency? Secondly, what methodology would be ideal to develop the most accurate predictive model for time to cash (with provided prediction intervals)?

To answer the abovementioned questions, the remainder of this paper is organized in the following manner. At the initial stage, we provide a review of previous academic and professional publications that investigate similar problems. Next, we discuss the dataset provided by our industry collaborator and show where we identified interesting relationships. Then we outline our carefully designed methodology that considers the process flow of construction activities, pre- and post-covid modeling cross-validation caveats, and feature-type considerations. The models we examine were chosen based on these considerations. In the Results section, we depict how the various modeling experiments performed from a statistical perspective and discuss the business implications on O2C of using our solution on our partner's business. Lastly, we provide answers to our specified research questions, discuss implications of this study and provide some thoughts on future research endeavors for this problem.

LITERATURE REVIEW

Academic research on predictive methods deployed in the home remodeling industry remains rare. However, the problem of predicting order-to-cash has multiple parallels and similarities in other industries where forecasting duration or delays of any given task or project is a beneficial objective function. The most immediate and logical industry to broaden our scope of search leads us to studies conducted in the construction industry, but even higher than that in the field of project management, given the many parallels to task flows and task dependencies.

Hou, Liu, et al. (2009) proposed a system dynamic model to solve China's construction industry ubiquitous payment problem. This study lends a unique perspective into the pervasive delays and the subsequent payment problem. Unlike previous studies that focus on establishing the upstream causes and what methods would remedy these causes, this study focused on the consequences of these payment delays onto the construction companies involved in the projects. To solve this issue, the study developed a system dynamics model on the cash balance and profitability of construction project under multiple scenarios. Hou, Liu, et al. concluded that through simulations, generally the longer the payment is delayed the lower the cash balance becomes and profitability declines. Whilst the descriptive nature of the study does not provide us any insight into predictive methods to inspire our study, it does provide baseline motivation of why our objective is a relevant one: the longer the order-to-cash period for any project, cash balance and profitability would decline.

Having established precedence for our objective, we scoured for potential predictive methods that would be applicable and salient to our project. We once again broaden our scope of search discovering a mix method to predict late payment. Smirnov (2016) provided a mix of survival analysis and random forests techniques to model and predict late invoice payments in business-to-business sales process using sales ledgers data. This study aimed to perform experimental analysis to model payment behavior of the debtors and find solutions to provide predictive analysis that could be used in the decision making of account receivables collection. A Cox Proportional Hazard model and a novel ensemble method, Random Survival

Forests were used in their analysis. In conclusion, this study used a two prong prediction tools to predict behavior of new and returning debtors. From this study, we discovered a new scikit learn python library for ensembles tree methods and survival analysis. Given the time component of our problem, we posit survival analysis is an appropriate technique, coupled with the robustness of tree-based predictive method, could be provide beneficial.

Kwon, Lippman et al. (2010) proposed an unadjusted work-rate model to capture the unique nature of project management contracts with delayed payments. This study suggests that there is a specific condition where a delayed payment can actually be beneficial for firms involved in the contract. This study concluded that there is a two "payment-regime," delayed and non-delayed, both of which are beneficial for different type of revenue size and number of parties involved in the contract. This study helped us frame the order-to-cash problem from a different perspective. The two "payment-regime" might indicate that, while predicting the order-to-cash, which is effectively payment delay, we would want the prediction to be as accurate as possible and later be minimized through optimization, it might be beneficial to not have the singular goal of minimizing. Perhaps, it is beneficial to optimize the order-to-cash to a timeframe that will maximize the revenue.

Denis (2011) conducted a study that compares differing architectural approaches to dive deep into the topic of financial flexibility and corporate liquidity. This study summarized the determinants and consequences of corporate cash holdings by utilizing a reduced form Stein model. In conclusion, the study found that firms attain financial flexibility through the management of corporate liquidity, through capital structure policies, and through payout policy.

Shifting our scope to another industry, Krumholz, Warner et al. (2019) conducted a study that developed a model to predict payments in the healthcare industry using data from the Centers for Medicare & Medicaid Services (CMS) claims. In the healthcare industry, predicting payment for specific medical conditions is the essential foundation to create "*Hospital Value-Based Purchasing*," an incentive payment to hospitals based on the quality of care provided to patients. Furthermore, this study also aims to test if changing the predictive variables in the model CMS already deployed would improve the prediction of total payment. They conclude that by using individual codes, leveraging the POA designations, and separating index from historical codes, we could improve the performance of the models.

Lastly, we seek to understand what studies had been conducted to see the feasibility and accuracy of various predictive modelling methods to predict duration, which is a close proxy to order-to-cash. Wauters and Vanhoucke (2016) compared Decision Trees, Bagging, Random Forest, Boosting, and SVM. These machines learning predictive modeling methods outperform the best performing Planned Value, Earned Duration, Earned Schedule and Elshaer forecasting methods (the traditional forecasting method) in their study. Although promising, this study had not yet been empirically validated. Given the real-world data we use in our study, we are poised to provide some degree of empirical validation to the forecasting methods we examine. Table 1 summarizes the studies we found related and guiding our ideas and how our study relates to those.

Table 1: Summary of literature review

Studies	Construction Industry	Descriptive Analysis	Predictive Analysis	Survival Analysis	Tree-Based Predictive Methods	Duration as target
(Kwon, Lippman et al. 2010)	✓	✓				
(Hou, Liu et al. 2011)	✓	✓				
(Denis 2011)	✓					
(Wauters and Vanhoucke 2016)			✓		✓	✓

(Smirnov 2016)			✓	✓	✓	✓
Our Study	✓	✓	✓	✓	✓	✓

DATA

We obtained data from a national home remodeling company that contained multiple tables with information on our collaborator's operation. The initial dataset had variables on aspects such as product, project, task, financing information, and customer leads. After developing a deeper understanding of the data and the business with their guidance, we were able to arrive at a smaller dataset containing variables that were related to our order-to-cash prediction focus. This entailed 65 candidate predictive variables and 290 thousand observations. In general, these 65 variables were able to be grouped into seven primary categories: Project, Product, Previous Home, Parent Project, Project Track, Finance & Lender, and PII. Table 2 provides a description for each of these variable categories.

Table 2: Categorization of potential model input variables

Variable Categories	Description
Project Variables	Features related to project timelines, type, current state, and project lead.
Product Information and Adjustments Variables	A project can be composed of multiple products, e.g., roofing, siding, door, insulation, or a combination of multiple products. Product Information variables contain flags indicating what type of product(s) is/are in any given project. Relating back to the variability of a remodeling project, changes in the cost of any project is important information to keep track of. Given the opportunity for further analysis on how the changes come to be and how it relates to O2C, the adjustment variables are created.
Previous Home Variables	A remodeling project has to have certain logistical and safety precautions in place before the project can go underway. This group of variable serves as a flag whether having had a previous relationship with the customer would expedite the O2C process since the required precautions would take less time to complete.
Parent Project Variables	Relating to the adjustment variables, we must have some way to keep track of how the changes relate to the initial project. Since an adjusted project creates a new project_id, we use the parent project variables to relate the new project to the initial project.
Project Track Variables	Given the many parts that form the workflow and dependencies of a project, the Project Track Variables keep track of what process has been done and how many times has it been done, and flag which suppliers are being utilized for any given project.
Finance & Lender Variables	To keep track of how a project is paid and if applicable, what financial institution provided the payment of any given projects, we use the flags under these groups.
PII Variables	The last category of variable falls under the PII -related variables. PII stands for Pre-Install Inspection. It is a process where a representative from our partner conducts an inspection to gather the necessary information required to manufacture parts and get the project underway. Given the limited availability of this resource, we posit that PII representative work rate have a significant impact on the O2C duration.

Finally, our dependent variable is the O2C gap. Defined as the gap in days between when the project was created, and cash was received (when installment started / funding received).

EDAs

The importance of some features was facilitated by EDAs. In this paper, four of the most insightful EDA's will be presented. In Figure 2, it is clearly indicated that the age of a territory plays a significant role in task duration variability. The territory where the company has served its clients for longer, on average, has more projects in progress. As a result, on average, there is a higher variability of task duration completions and, therefore, order-to-cash gap.

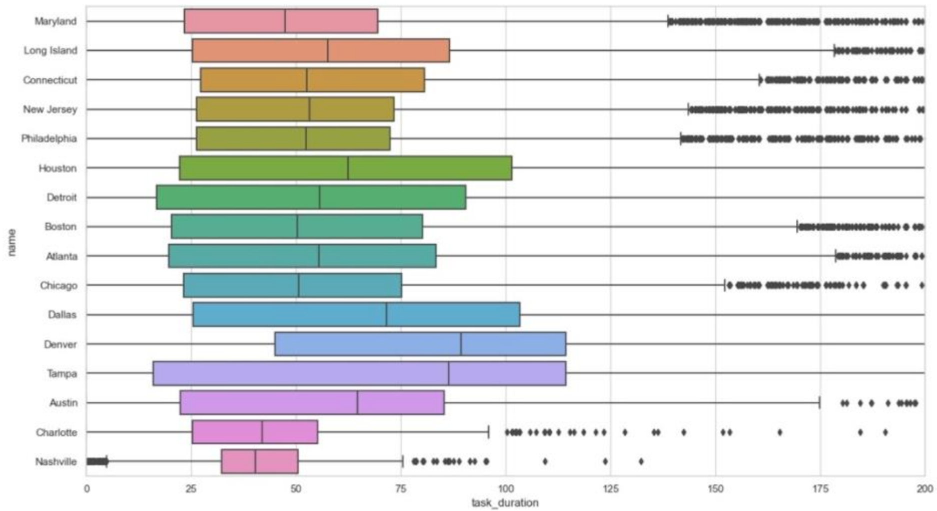

Figure 2: Completion period by regions

Figure 3 represents the average task durations for each month (histogram) and the total count of projects each month (line plot). Comparing the overall task duration averages by product, windows most closely resembled the global average. This can be explained by the fact that the majority of the products are indeed windows. The confirmation of this claim is provided in Figure 4, which confirmed that windows take up the largest section of the national remodeling company's portfolio.

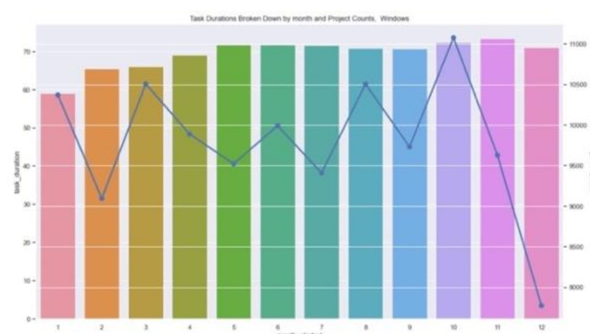

Figure 3: Completion period by month & number of projects each month

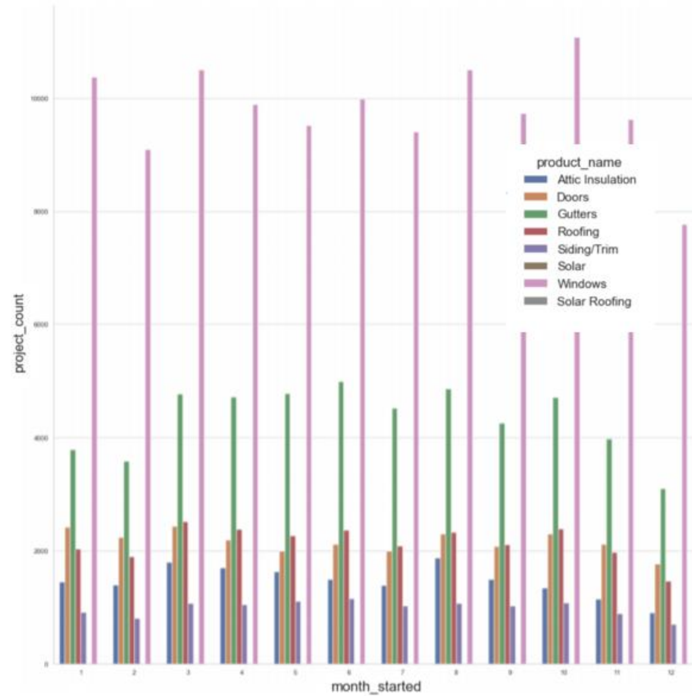

Figure 4: Breakdown of type of project by month

Even though in the current study the goal is to improve the already existing model, which is based on the cumulative duration of all tasks, there are further steps required to enhance it. Average task duration, in days, is provided below for the mentor-specified tasks. However, due to task dependency, each of the successive task duration is cumulative. For example, task 23 depends on tasks 5, 1, 7, and 8 being completed. To shorten the length of the order-to-cash-gap, identification of top 2-3 most prolonged tasks is required.

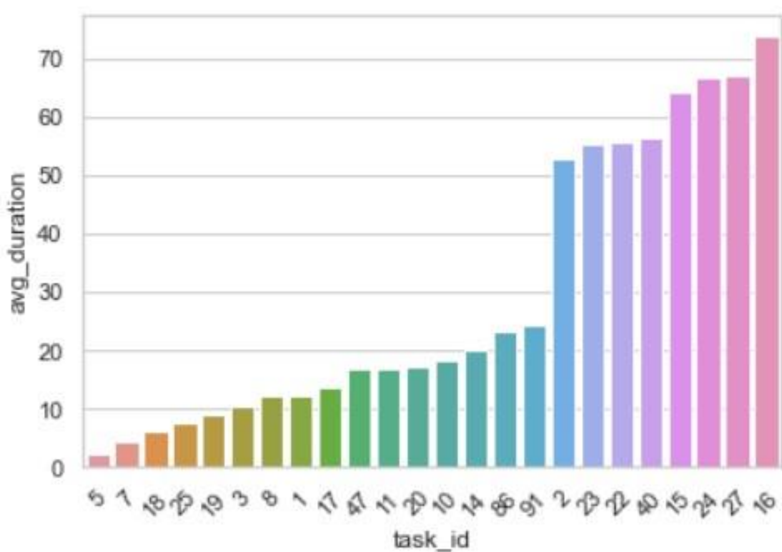

Figure 5: Task duration in days

METHODOLOGY AND MODELS

An attractive aspect of our data preparation process that other firms will likely be tasked with considering is that we account for pre-, during-, and post-covid in our training and evaluation process. Because a traditional cross-validation method would lead missing values in data which time-series cannot deal with, we applied a rolling forecast origin technique to partition the data. Since the pandemic poses a unique challenge given the sudden difference in data behavior, we decided to partition this data in a fashion suitable to handle a behaviorally time-series data.

To do this, we initially had the idea to first partitioned multiple folds with the first observation in each fold going forward in time. This is to achieve a representative model that is sensitive towards the "abnormal" observations caused by the pandemic. Using this approach allowed us to account for the pandemic's abnormal behavior and to achieve a higher accuracy and to avoid overfitting on the train set. Then, we specified the horizon to be bigger than one to have more than one observation as part of the validation set. We experimented with 80/20, 70/30, 60/40, and 50/50 partitioning methods and determined that the optimal method during the model development was 80/20. After tuning the data, we merged all the data together and grouped by territories which provided useful information and business insight for us to predict the payment latency in various geographical locations. It is necessary to sort and clean the data before training the machine learning algorithm. After data preparation, we created a pipeline for the machine learning model.

Ensemble Method

Utilizing the increasingly popular H2O machine learning library, specifically the H2O AutoML wrapper functionality, we attempt to develop an ensemble method predictive solution. Given the robustness, array of algorithms within the wrapper, and the generally accurate prediction (LeDell, Erin, and S. Poirier.) we attempted to use this rather "black-box" approach to machine learning with a two-folds goal: First, to see if this approach may be a suitable solution for predicting Order-to-Cash. Second, to see if we can peel the layers and gain a better understanding of how to best tune some of the parameters especially to fit out problems.

Recurrent Neural Networks

Time-series forecasting has always been a very important area of research in many domains because many different types of data are temporal in nature, as with our problem. Developing Recurrent Neural Networks (RNNs) which each connection between neurons has a corresponding trainable weight and organizes into successive layers. We solved some problems using traditional machine learning models for time-series forecasting. For instance, RNNs can not only find the complex patterns in the input time-series but also give good results in forecasting more than few-steps. However, the training of a RNNs is hard to parallelize and is also computationally expensive (Bengio, et al. 2015).

Random Survival Forest

Other methods in applying statistical machine learning are survival analysis. Random Survival Forest (RSF) is an active area of research in biostatistics, which focuses on a time-to-event outcome that is typically censored. Compared with regression models, RSF not only represents a suitable tool for exploratory analysis but also limits univariate regression approaches such as overfitting, unreliable estimation of regression coefficients. However, RSF does not give precise continuous nature prediction, which cannot predict beyond the range in the training data. Thus, overfitting data sets that are particularly noisy.

Light Gradient Boosting Machines
With the gradient boosting framework that uses a tree-based learning algorithm, Light Gradient Boosting Machines (GBM) grows tree leaf-wise while other algorithms grow level-wise (Fei, et al. 2018) When growing the same leaf, the leaf-wise algorithm can reduce more loss than a level-wise algorithm. The data we possess are extremely large which Light GBM can play important role in dealing with overfitting problem. Light GBM can not only handle the large size of data and takes lower memory to run, but also focus on accuracy of results. However, Light GBM is sensitive to overfitting and can easily overfit small data.

CatBoost
Due to diverse data types, we investigated the CatBoost algorithm on our forecasting problem, which has a very good vector representation of categorical data and takes concepts of ordered boosting. CatBoost can be explained by Category and Boosting, which not only yields state-of-the-art results without extensive data, but also provides powerful out-of-the-box support for the more descriptive data formats that accompany many business problems (Dorogush et al. 2018). However, CatBoost does not support sparse matrices and takes time to train compared to Light GBM.

We choose to use mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE) statistical evaluation metrics to assess the performance of the predictive models in our study because we found them to be the most popular metrics to compare time-series models. The methodology can be seen in Figure 6.

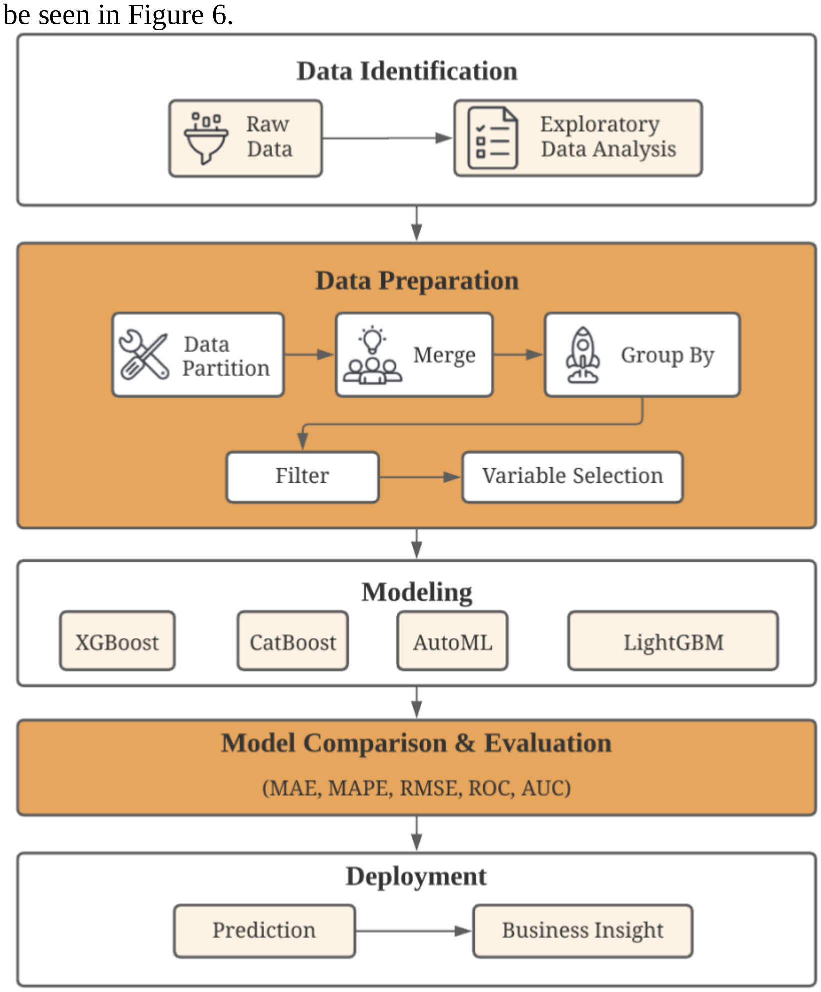

Figure 6: Methodology

RESULTS

Of the four models we attempted, we managed to achieve a training results that performed better than those of the base model provided by our partner. More specifically, the baseline model of our industry partner has a training Root Mean Squared Error (RMSE) of 15.44. We achieve a training RMSE of 13.97 for the CatBoost Model, 13.93 for the XGBoost Model, 16.67 for the LightGBM model, and 14.1 for the Stacked Ensemble from the h2o AutoML. All models were candidate models (i.e., not overfit) in our experiments. Figure 3 provides the reader an idea of how accurate our forecast was over time from 2014 thru 2021 (pre- and post-covid).

Figure 7: Best prediction versus actual

We explored various training and validation designs with this time-series data to see which would best capture the signal of O2C considering pre- and post- covid as mentioned earlier in the methodology part. One cross-validation design that was promising but did not lead to the results we had hypothesized was splitting the timeframe over a few time windows to train and evaluate the models. We found that giving differing weights during the training of the models, using the most recent data to give more signal by assigning more weights to these data and providing smaller weights to older data allowed our models to learn the changes in the data quickly and best predict the business over all time periods. This is depicted in Figure 4 below.

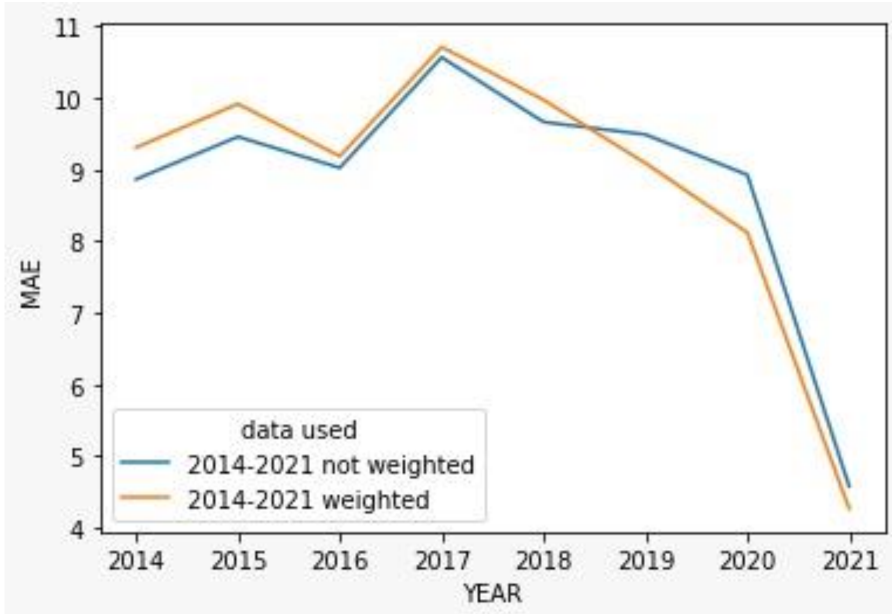

Figure 8: Weighted versus non-weighted data

Business Perspective and Impact

Given the performance of our best model, the CatBoost model, we can expect some qualitative and quantitative improvements to various aspects of the business of our partner. Firstly, the increase of prediction accuracy would allow a more efficient human resource management. Our partner would be able to allocate staff and workers more effectively if they have a better prediction to inform their decision. Customer satisfaction would also be improved. A more accurate prediction would allow the business to provide a more complete estimate of the project time. The closer the actual completion time to the estimate, the higher the total utility the customer would get out of the project.

Internally, the increase of prediction along with the increase of understanding of the drivers of the O2C duration would also allow businesses to understand the "cost of queue", essentially given the potential revenue from a project how much is not realized due to an inaccurate O2C period. As such, this also brings the possibility of using the prediction as supporting tools to make the operational decision of what task to "crash" along the chain of tasks in a given project to shorten O2C period. Quantitatively, the two days improvement in prediction accuracy can be expected to translate into 3% reduction in operating costs, primarily from labor and overhead costs. In the context of our partner's operation, this could translate into approximately $800,000 in savings through reduction in Operational Expense.

CONCLUSIONS

In conclusion, after numerous experiments, our team was able to build a model for predicting Order-to-Cash (O2C) time. We were able to achieve a significant improvement in accuracy in predicting Order-to-Cash (O2C) compared to the model that our client is currently using. Furthermore, in order to help our client, identify the areas for opportunity to improve the operations we managed to identify features that drive the increase in project length.

From a modeling perspective, to address recent events we made sure that our model is good at responding to rapid changes in trends. We achieved that by ensuring that the most recent observations have more significance compared to the old ones. We believe that sensitivity of our model towards recent events is the most significant improvement we made, since the industry was greatly affected by the pandemic and lockdowns.

Using our model, our industry partner and other business alike who relies on accurate prediction of duration can improve their customer satisfaction, better manage personnel resourcing, and potentially save cost as a result of better budgeting practice.

REFERENCES

Denis, D. J. (2011). "Financial flexibility and corporate liquidity." Journal of corporate finance **17**(3): 667-674.

Hou, W., et al. (2011). "Payment problems, cash flow and profitability of construction project: a system dynamics model." International Journal of Economics and Management Engineering **5**(10): 1266-1272.

Krumholz, H. M., et al. (2019). "Development and Testing of Improved Models to Predict Payment Using Centers for Medicare & Medicaid Services Claims Data." JAMA network open **2**(8): e198406-e198406.

Kwon, H. D., et al. (2010). "Project management contracts with delayed payments." Manufacturing & Service Operations Management **12**(4): 692-707.

Smirnov, J. (2016). Modelling late invoice payment times using survival analysis and random forests techniques, Ph. D. thesis.

Wauters, M. and M. Vanhoucke (2016). "A comparative study of Artificial Intelligence methods for project duration forecasting." Expert Systems with Applications **46**: 249-261.

Will, A. (2021). "Benchmark Update Modestly Boosts Remodeling Market Size Projections." Retrieved 2/1/2021, 2021, from https://www.jchs.harvard.edu/benchmark-update-modestly-boosts-remodeling-market-size-projections.

Dorogush, Anna Veronika, Vasily Ershov, and Andrey Gulin. "CatBoost: gradient boosting with categorical features support." *arXiv preprint arXiv:1810.11363* (2018).

Li, Fei, et al. "A light gradient boosting machine for remainning useful life estimation of aircraft engines." *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018.

Bengio, Samy, et al. "Scheduled sampling for sequence prediction with recurrent neural networks." *arXiv preprint arXiv:1506.03099* (2015).

LeDell, Erin, and S. Poirier. "H2o automl: Scalable automatic machine learning." *7th ICML workshop on automated machine learning*. 2020.