BUSINESS-TO-BUSINESS SALES OPPORTUNITY MANAGEMENT STRATEGY

A Dissertation

by

MUZEEB SHAIK

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Shrihari Sridhar
Co-Chair of Committee,	Vikas Mittal
Committee Members,	Eli Jones
	Srikanth Paruchuri
Head of Department,	Manjit S. Yadav

August 2022

Major Subject: Marketing

ABSTRACT

To meet sales targets with finite resources, business-to-business (B2B) sellers need to prioritize promising sales opportunities from a large pipeline of possibilities. Extant evidence in theory and practice suggests that B2B sellers use arbitrary or gut-based decision rules to prioritize sales opportunities, which leads to sub-optimal sales opportunity management.

I draw from the relationship management and organizational buying literature to build a conceptual framework relating buyer-class typology (e.g., new bid vs. modified rebid), and opportunity characteristics (e.g., opportunity size) to a seller's decision to bid, and the bid outcome. I test the framework using archival data from a major B2B on-site services provider. The data span sales outcomes (did not bid, bid and won, bid and lost) for 4,564 sales opportunities, spanning 10 years, and 23 countries. Model-free evidence indicates an ongoing tension in sales pursuit: while the best projects to bid on are ones with the best relationship typology (e.g., direct relationship) and lowest risk (e.g., small opportunity size), such opportunities are not enough to grow the sales in each region.

Accordingly, I first develop an ensemble machine learning framework to predict the focal seller's propensity to win the opportunity, as a function of complex non-linear interplay among the sales opportunity characteristics. Subsequently, I propose to build a combinatorial optimization approach that allows a company to take the right level of risk (i.e., some new bids, some large opportunities) so that each region can maximize its sales potential given finite capacity. I demonstrate that using the ensemble approach

enable the focal company to improve its predictive validity over simple models by 11.09%. Moreover, by using a combinatorial optimization approach in conjunction with the predictive ensemble model, I retrospectively demonstrate that the focal firm could have increased its sales by 22% while bidding on 38% fewer projects.

DEDICATION

This dissertation is dedicated to my parents.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my dissertation committee chair, and mentor Dr. Shrihari Sridhar for his continued guidance and encouragement during the course of my PhD education and beyond. I would also like to thank my dissertation co-chair Dr. Vikas Mittal for his mentorship and support. Working with Hari and Vikas helped me become a well-rounded individual and scholar.

I also thank my committee members Dr. Eli Jones and Dr. Srikanth Paruchuri for their time and valuable feedback in developing my research agenda. I am also thankful to all my professors for the excellent training I have received in Marketing, Economics and Statistics during the many seminars and classes.

I also wish to thank Dr. Alina Sorescu for being a persistent support system for all graduate students, including me, and for being a champion of the marketing PhD program. I would like to express my sincere gratitude to Dr. Manjit Yadav, our department head, for his leadership and support to the PhD program. Shoutout to all marketing department faculty, doctoral colleagues, and department staff for making my time in the department a memorable one.

Lastly, I would like to thank my friends and family for always being there for me. In particular, I would like to express my appreciation and regards to Parvez, Revathy, Sai Chand, Narendra, Sai Harsha, and Sandip for always being just a call away in times of need.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Shrihari Sridhar [chair of dissertation committee] of the Department of Marketing, Dr. Vikas Mittal [co-chair of dissertation committee] at the Jones Graduate School of Business of Rice University, Dr. Eli Jones of the Department of Marketing, and Dr. Srikanth Paruchuri of the Department of Management.

The data analyzed for this dissertation was provided by Dr. Shrihari Sridhar and Dr. Vikas Mittal.

All other work conducted for the dissertation was completed by the student under the supervision of the dissertation committee.

Funding Sources

Graduate study was supported by a graduate assistantship from the Mays Department of Marketing, and a PhD Enhancement Fellowship from the Mays Business School, Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

According to a recent practitioner report (Bonde et al. 2019), business-to-business (B2B) transactions account for 51% of the revenues generated in the US Economy highlighting the importance of B2B selling. B2B sales opportunities require considerable effort and time from the seller since they are characterized by heterogenous customer requirements, high technical complexity, large economic value, and large number of decision makers due to extensive buying-center involvement (Grewal et al. 2015). Close to 50% of new B2B deals taking more than 7 months to complete (CSO Insights 2019). Given that the cost of a B2B direct sales rep is approximately $1.9K per business day worked (Peter 2019), firms incur huge costs in pursuing sales opportunities.

Considering the limited resources available to sellers, and the effortful nature of sales pursuit in B2B markets, sellers need effective ways to prioritize their sales opportunities. Accordingly, sellers invest extensively in sales opportunity management, the process of prioritizing specific opportunities in the sales pipeline (Söhnchen and Albers 2010). Successful prioritization can help sellers aggressively pursue a subset of high potential opportunities from a large pipeline and provide accurate sales forecasts to the C-suite to catalyze strategy planning (Mittal and Sridhar 2021).

However, given the high complexity and transaction values of projects, and multiple agents from the buyer-side (e.g., procurement/buying center, legal compliance) and seller-side (inside/outside sales, project management) being involved in the sales process (Grewal et al. 2015; Grewal and Sridhar 2021), it is challenging for B2B salespeople to predict winning sales opportunities. Given this unpredictability, sellers use arbitrary stop/go decision heuristics to rank order which opportunities to pursue (Söhnchen and Albers 2010). Xu et al. (2021) find that companies that use such heuristics end up pursuing more medium-sized opportunities, deemphasize small and large opportunities, and overestimate conversion rates by up to 100%. It is not surprising that sales opportunity conversion rates are poor, ranging from 1%-5% (D'Haen and Van den Poel 2013), with surveys showing that 63% of sales managers believe that their firms are not doing a good job of sales opportunity management (Miller 2019).

Three disparate streams of research in business-to-business marketing can be applied to the context of sales opportunity management. The literature on salesperson judgement and decision making has documented the different biases salespeople have while prioritizing opportunities (Hall, Ahearne, and Sujan 2015; Sabnis et al. 2013; Xu et al. 2021). Similarly, literature on relationship marketing has shown that buyer's relationship strength (Dwyer, Schurr and Oh 1987) is associated with an increase in a seller's bidding success and naturally salespeople are likely to go after opportunities with high relationship strength. However, these streams of research have not been applied to predictive modeling of sales opportunity pursuit and prioritizing the pipeline of sales opportunities. A recent third stream of research in marketing and engineering uses buyer characteristics (e.g., company size, industry type) to predict the win probability of B2B sales bids (D'Haen and Van den Poel 2013). For example, Jahromi, Stakhovych and Ewing (2014) and Mortensen et al. (2019) use buyers' past purchase histories within machine learning frameworks to predict winning sales opportunities for sellers. However, this stream of research does not provide a theoretical understanding of the attributes determining the seller's bid outcome and therefore lacks normative guidance to sellers that can improve sales opportunity management.

I bring together these streams of literature to address the problem of sales opportunity management. In keeping with the first two streams of literature, I develop a theoretical framework of factors and bid attributes effecting the focal firm's decision to bid and subsequently winning the bid. Using this framework, I develop a predictive machine learning model that helps a seller predict the outcome of every B2B sales opportunity in the pipeline which can be combined with the proposed optimization framework to identify a subset of opportunities to pursue.

To empirically validate my modelling framework, I use bid-level data from a major food services and facilities management company with global operations. My data covers 4,574 B2B sales opportunities, across ten years, with each opportunity being pursued by the focal firm and its competitors. Using regression analysis, I study the effect of bid characteristics including size and relationship type on customers' decision to withdraw the bid, focal firm 'decision to bid for the opportunity and focal firm winning the bid. Results show that customers are less likely to withdraw rebid opportunities. Also, the focal firm is less likely to bid on larger opportunities and the likelihood of bidding increases as the strength of their relationship increases. There are no significant interaction effects of size of the opportunity and relationship type on the focal firm's likelihood to bid on the opportunity. Similarly, the focal firm is less likely to win larger opportunities and the likelihood of winning the bid increases as the strength of their relationship increases. I find significant non-linear interaction effects of size of the opportunity and relationship type on the likelihood to win the bid.

Subsequently, I use an ensemble machine learning framework by stacking five different machine learning models – random forest, gradient boosted machines, logistic regression, naïve bayes classifier, and support vector machine models – with another random forest as a meta-learner combining the predictions from the five individual models to predict the focal firm's likelihood to win a bid. I find that the ensemble model outperforms all the individual models used in stacking in both in-sample and out-of-sample performance on different metrics such as accuracy, sensitivity and specificity. I use partial dependence plots proposed by Friedman (2001) to understand the marginal dependence of predicted outcome of the ensemble model focal firm's likelihood to win a bid on different independent variables. The partial dependence plots show that the size of the project has an overall negative marginal effect on focal firm's likelihood to win the bid and as the strength of the relationship increases, focal firm's likelihood to win the bid increases. This shows that the focal firm's customers value relationship. Further, using the machine learning model, I develop an optimization framework that can be used by firms to prioritize and pursue a subset of opportunities within the strategic capacity constraints. The results of the optimization model document the value of having a predictive model to prioritize opportunities while maximizing different objectives within different constraints. I retrospectively demonstrate that the focal firm could have increased its sales by 22% while bidding on 38% fewer projects.

This study makes three main contributions to the marketing theory and practice. First, I add to the growing stream of literature on sales opportunity management (D'Haen and Van den Poel 2013; Söhnchen and Albers 2010) by empirically integrating well-known theoretical constructs in the B2B buyer-seller relationship literature (Andreassen and Lervik 1999; Dwyer, Schurr and Oh 1987) in a practical way to enable normative sales opportunity management. I document the biases held selling organizations have while prioritizing opportunities to pursue. Results from the regression model and the machine learning model show that firms avoid larger opportunities and are also drawn towards opportunities where they have a stronger relationship with the customers.

Second, I add to the literature on predictive sales analytics (D'Haen and van den Poel 2013; Mortensen et al. 2019; Rezazadeh 2020) by demonstrating an application of predicting model in the sales opportunity management context. Heinitz, Hable and Alavi (2021) posit that sales literature on predictive analytics has ignored questions about the adoption of predictive models by the firms and how firms should develop and implement decision making rules based on predictive models. This study retrospectively demonstrates the value of having a predictive model making it easier for firms to convince sales teams to adopt predictive tools in prioritizing opportunity pursuit. Moreover, the proposed optimization framework can be used as a template for firms to build prescriptive frame works based on predictive models.

Finally, I contribute to the literature on marketing decision support systems (Chica and Rand 2017; Silva-Risso and Ionova 2008; Zoltners and Sinha 2005) by demonstrating an empirical application of predictive and prescriptive analytics in

B2B sales. My results also indicate that the focal firm can use my approach to improve sales outcomes by prioritizing bids. More generally, my framework can be used by Chief Sales Officers and Chief Executive Officers in the context of quarterly sales forecasts and strategy planning, both of which inherently rely on sales opportunity pipelines.

CHAPTER II

BACKGROUND AND CONCEPTUAL FRAMEWORK

Sales Opportunity Management

Sales opportunity management or pipeline management, the process of prioritizing specific opportunities in the sales pipeline, is one of the most crucial activities performed by sales teams. B2B opportunities usually have larger buying cycles involving multiple decision makers and sellers need to expend time and effort to complete deals. Given the effortful nature of sales pursuit in B2B markets sellers cannot go after all the opportunities they encounter in the market. However, they also need to meet their revenue and margin targets across different locations and different product offerings. This makes it important to have effective ways to identify a subset of opportunities to pursue and win these opportunities.

Literature has shown that sales managers use arbitrary stop/go decision heuristics to identify these subset of opportunities (Söhnchen and Albers 2010). Judgment and decision-making literature in the sales context has documented the cognitive biases held by the salespeople (Brown et al. 1981; Lam and van der Borgh 2021) across different stages in the selling process and these biases are likely to impact opportunity management (Scherm 2021). For example, salespeople overestimate the probability of winning resulting in lower conversion rates than expected and this in turn leads to sale teams not achieving their targets. Similarly, salespeople are likely to pursue opportunities where success seems to be more certain, even if other opportunities have greater potential and are more profitable. The following set of opportunity attributes are known to influence the propensity to pursue the opportunity as well as likelihood of winning the opportunity.

Size of the opportunity

Recent literature on salesperson judgment and decision making has documented the effect of opportunity size on the propensity to pursue. Using a scenario experiment, Xu et al. (2021) show that the size of the project has an inverted U relationship with salespeople's willingness to pursue different opportunities. Meaning, as the opportunity size increases from low to medium, the willingness of the salespeople to pursue the opportunity increases and as the opportunity size further increases from medium to high, the willingness to pursue decreases. They also find that this effect is mediated by the resource slack available to the salesperson validating their argument that salespeople do an expected cost-benefit analysis at the portfolio level and try to conserve resources while making sure their targets are met. They document a similar (inverted U) relationship between the size of the project and winning the projects.

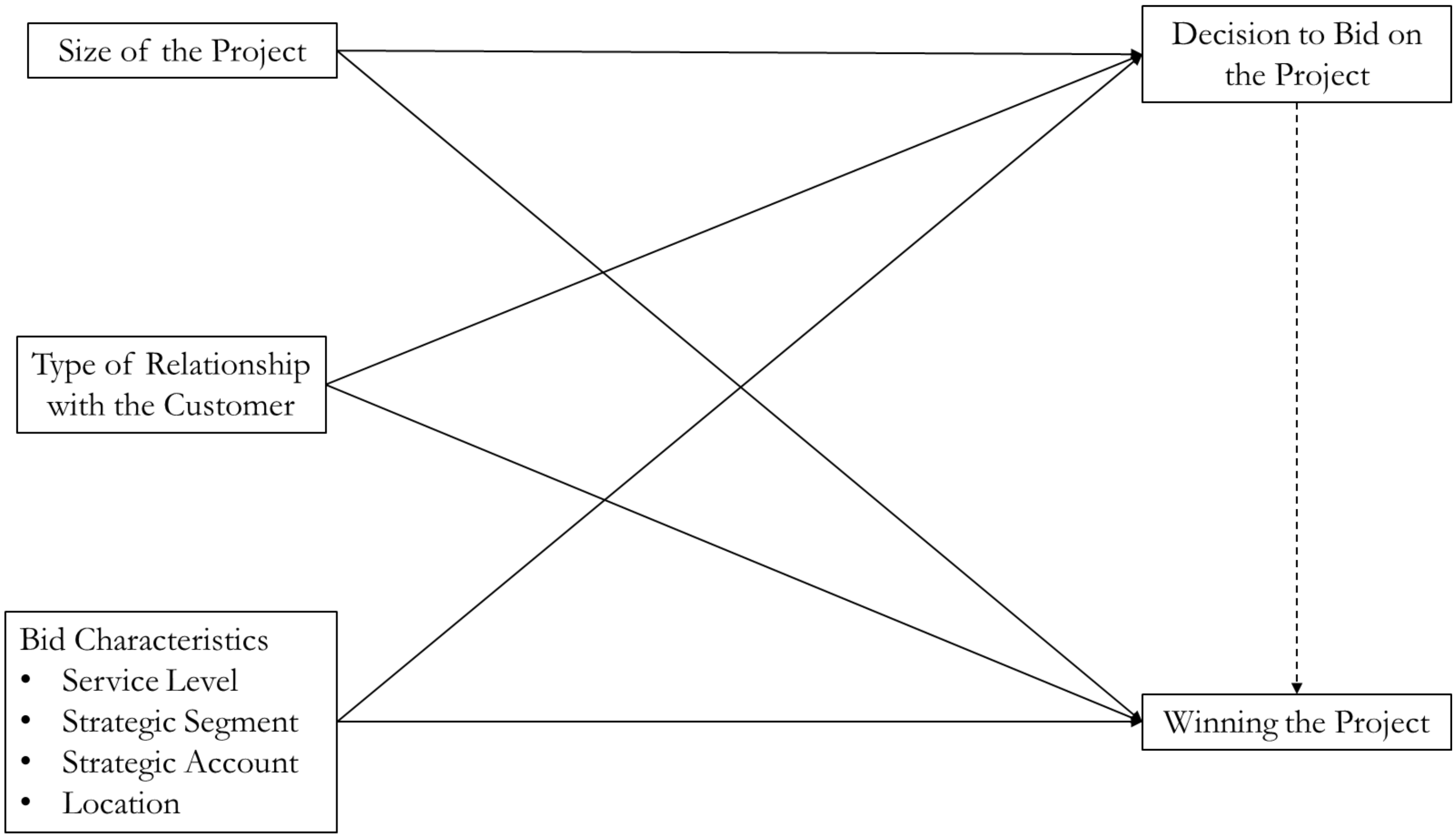
Relationship with the customer

Literature on relationship marketing has robustly shown that value of having relationships with customers. Dwyer, Schurr, and Oh (1987), in their seminal article, discuss the need to adopt a relational view of buyer-seller exchanges and proposed a framework for developing buyer-seller relationships. Having stronger relationships with the customers will increase their commitment towards the seller which is known to have a positive effect on the customer repurchase intentions (Bosukonda et al. 2020). Given the robust effect of relationships on the buyers' willingness to purchase, sellers are ought

Table 1: Relationship Typology

Type of Relationship	Description
Direct Relationship	Focal-firm has direct relationship with the customer at the focal site
Indirect Relationship and New Location	Focal-firm has a relationship with the customer parent company at other locations and the focal site is either a green field/self-serviced location
Indirect Relationship	Focal-firm has a relationship with the customer parent company at other locations and the focal site is either serviced by competitors
No Relationship and New Location	Focal-firm has no relationship with the customer parent company at other locations and the focal site is either a green field/self-serviced location
No Relationship	Focal-firm has no relationship with the customer parent company at other locations and the focal site is either serviced by competitors

Figure 1: A Theoretical Framework of Factors Effecting Opportunity Pursuit and Winning the Opportunity



to be pursuing opportunities where they have strong relationships with the customers. However, literature has also shown that salespeople are likely to misjudge their customers' commitment levels. Relationship-related factors such as frequency of interactions and the length (time) of the relationship influence the extent to which salespeople misjudge the commitment levels (Homburg, Bornemann, and Kretzer 2013). Overall, the relationship with the customers will likely have an effect on the choice of opportunities to pursue and subsequent winning of the bid. I my context, leveraging the rich data capturing the relationship with the parent companies, I develop a relationship typology shown in Table 1 that captures the focal firm's direct and indirect relationship with the customer. This typology can complement the BuyGrid framework (Robinsion, Faris and Wind 1967) used to understand B2B buying by capturing a relationship perspective of the seller.

In addition to the above mentioned factors, other bid characteristics are also likely to impact the sellers decision to pursue an opportunity and a buyer's decision to award the contract to the focal sellers. Figure 1 shows a theoretical frame of association between size of the project, relatonship type and bid charactristics and the focal firm's decision to bid and focal firm winning the project.

CHAPTER III

EMPERICAL ANALYSIS

Institutional Context and Data

I use data from a leading business-to-business on-site services provider (I refer to this as “ServiceCo” hereinafter) with global operations. ServiceCo competes primarily with four major players globally and several local service providers. All the sales opportunities encountered by ServiceCo can be either a re-bid opportunity where ServiceCo is already engaged with the customer in the focal site or new-bid opportunity. Within the new-bid opportunities, some opportunities are serviced by one of ServiceCo's competitors, whereas some opportunities are self-serviced by the customers, and a few are new site openings. Most of ServiceCo’s clients have multiple sites in operation which opens the possibility of ServiceCo having a relationship with the parent company of a specific customer at a different location. I use this information to develop the relationship typology.

The customer relationship management system data provides a static snapshot of 4,574 sales opportunities along with the customer account histories from fiscal year 2010 to 2021. Of the 4,574 opportunities, 550 opportunities (12.04%) are withdrawn by the customers, for 1,203 opportunities (26.03%) ServiceCo decided not to bid, and they submitted bids for 2,821 opportunities (61.67%). This indicates that across all the opportunities, ServiceCo has a walk away rate of 30%. Figure 2 provides a distribution of old provide and new provider (who won the bid) across different opportunities

Figure 2: Distribution of Old Service Provider and New Service Provider across all Opportunities

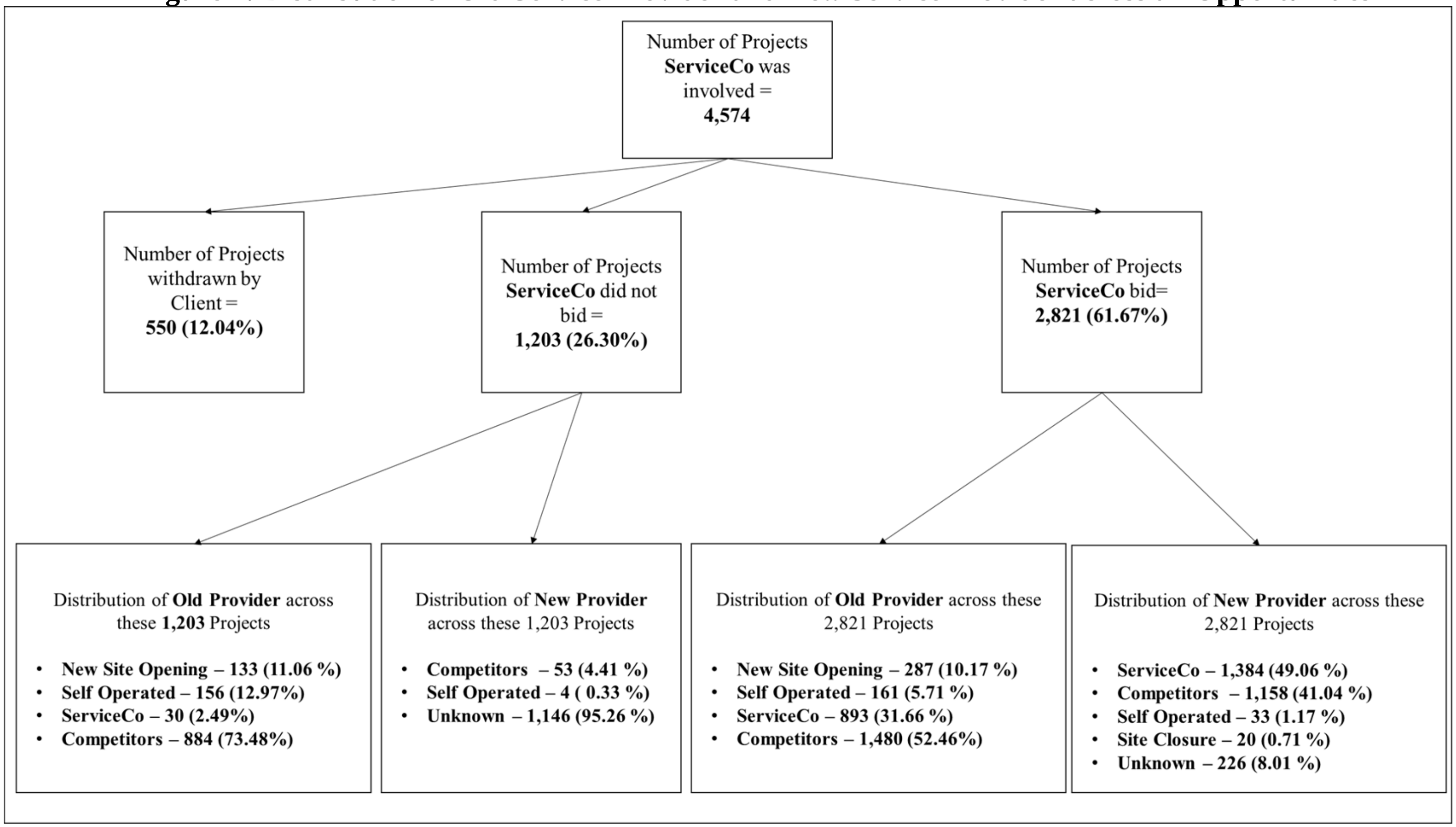


Table 2: Summary Statistics

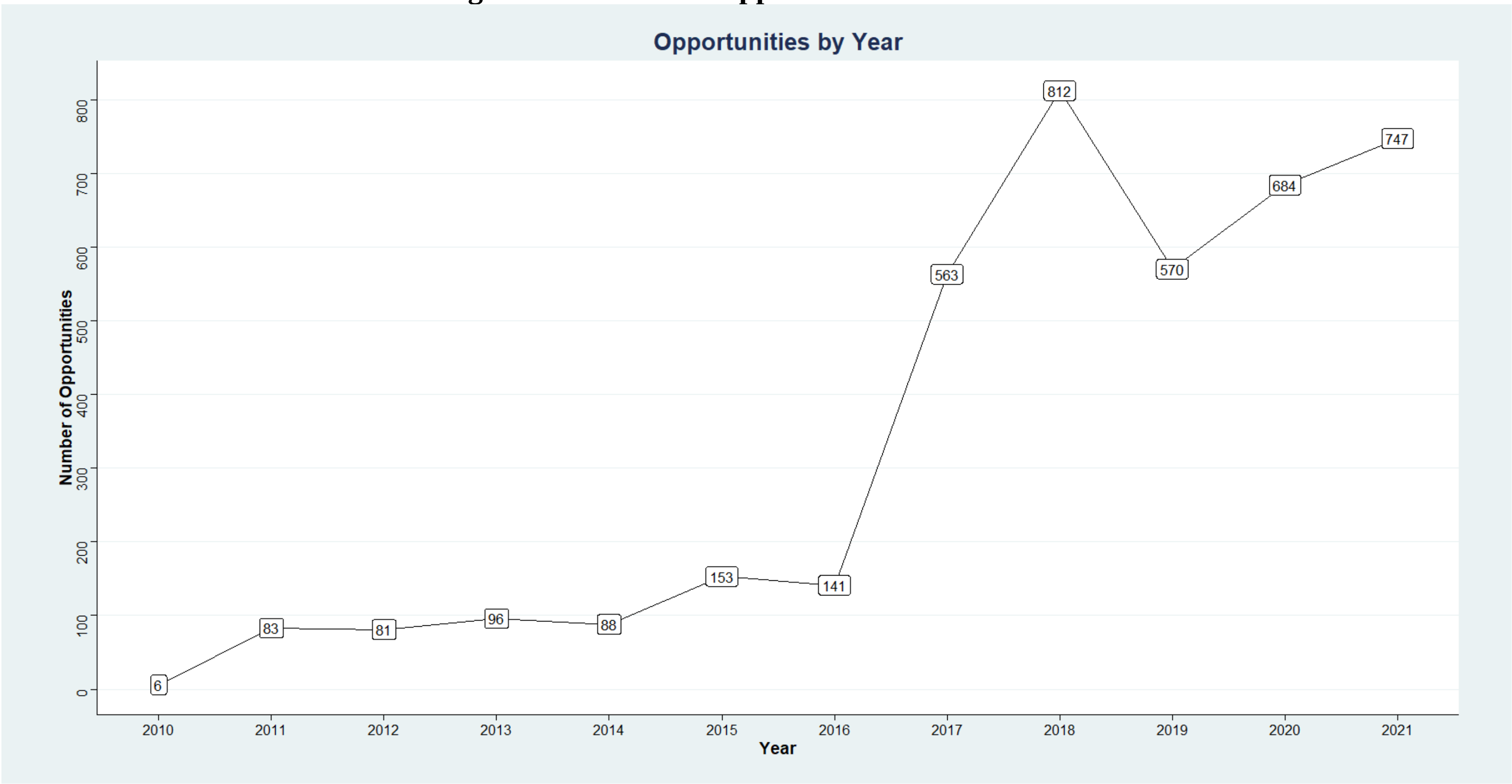
Variable	Definition	Mean (SD)	Min	Max
Dependent Variable				
Customer Withdraws the Opportunity	Yes = 1, otherwise = 0	12.02%*	0	1
ServiceCo Bids for the Opportunity	Yes = 1, otherwise = 0	70.10%	0	1
ServiceCo Wins the Opportunity	Yes = 1, otherwise = 0	48.71%**	0	1
Independent Variables				
Size of the Project				
Revenue	Expected revenue from the opportunity in millions of dollars	1.57 (3.73)	0	66.05
Relationship Type				
Direct Relationship with Buyer on Site	Yes = 1, otherwise = 0	22.94%	0	1
Indirect Relationship at Old Location	Yes = 1, otherwise = 0	23.14%	0	1
Indirect Relationship at New Location	Yes = 1, otherwise = 0	6.21%	0	1
No Relationship at New Location	Yes = 1, otherwise = 0	12.10%	0	1
No Relationship at Old Location	Yes = 1, otherwise = 0	35.61%	0	1
Strategic Segment				
Mining	Mining = 1, otherwise = 0	26.59%	0	1
Corporate Services	Corporate Services = 1, otherwise = 0	0.45%	0	1
Energy	Energy = 1, otherwise = 0	72.56%	0	1
Government Services	Government Services = 1, otherwise = 0	0.25%	0	1
Healthcare	Healthcare = 1, otherwise = 0	0.05%	0	1
No Segment	No Segment = 1, otherwise = 0	0.07%	0	1
Schools	Schools = 1, otherwise = 0	0.02%	0	1
Service Level				
1	Service Level 1 = 1, otherwise = 0	9.82%	0	1

Table 2 Continued: Summary Statistics

Variable	Definition	Mean (SD)	Min	Max
2	Service Level 2 = 1, otherwise = 0	25.72%	0	1
3	Service Level 3 = 1, otherwise = 0	36.98%	0	1
4 (Integrated Services)	Service Level 4 = 1, otherwise = 0	27.49%	0	1
Strategic Account				
Yes	Customer is a strategic account = 1, otherwise = 0	16.55%	0	1
Country				
Australia	Yes = 1, otherwise = 0	9.82%	0	1
Belgium	Yes = 1, otherwise = 0	0.67%	0	1
Brazil	Yes = 1, otherwise = 0	20.38%	0	1
Canada	Yes = 1, otherwise = 0	12.33%	0	1
Chile	Yes = 1, otherwise = 0	5.07%	0	1
Colombia	Yes = 1, otherwise = 0	2.93%	0	1
Denmark	Yes = 1, otherwise = 0	0.02%	0	1
Gabon	Yes = 1, otherwise = 0	0.12%	0	1
Global Segment	Yes = 1, otherwise = 0	0.92%	0	1
India	Yes = 1, otherwise = 0	7.93%	0	1
Indonesia	Yes = 1, otherwise = 0	0.20%	0	1
Laos	Yes = 1, otherwise = 0	0.05%	0	1
Malaysia	Yes = 1, otherwise = 0	0.22%	0	1
Mexico	Yes = 1, otherwise = 0	0.75%	0	1
Myanmar	Yes = 1, otherwise = 0	0.40%	0	1
Netherlands	Yes = 1, otherwise = 0	3.01%	0	1
Norway	Yes = 1, otherwise = 0	1.64%	0	1
Peru	Yes = 1, otherwise = 0	6.44%	0	1
Thailand	Yes = 1, otherwise = 0	0.32%	0	1
USA	Yes = 1, otherwise = 0	16.80%	0	1
United Arab Emirates	Yes = 1, otherwise = 0	0.07%	0	1
United Kingdom	Yes = 1, otherwise = 0	9.87%	0	1
Vietnam	Yes = 1, otherwise = 0	0.05%	0	1

* N = 4,574; ** N = 2,821

Figure 3: Number of Opportunities in Each Year



ServiceCo was involved in. Re-bid opportunities are identified with ServiceCo as the old provider.

ServiceCo's opportunities are grouped into different *Strategic Segments* with majority of opportunities coming from mining and energy segments. ServiceCo offers services in three primary domains (referred to as *Service Levels*) and some customers award contracts that combine services across different *Service Levels* (referred to as *Integrated Services*). Most of ServiceCo's customers have opportunities at different locations and each of these locations operate as an individual entity. In my data, I get to observe the parent account associated with each opportunity and I use this information to develop a variable which captures the relationship typology discussed earlier. Also, ServiceCo treats some of these customers as strategic accounts and I use this information in my modeling framework. For the 4,024 opportunities not withdrawn by the customers, Figure 3 provides a snapshot of number of opportunities in each year and Table 2 provides a description of each variable, and descriptive statistics.

Model Free Evidence

To obtain model-free evidence, I compare walk away rate, absolute win rate, win rate conditional on bidding across different levels of the variable to be included in the analysis. Walk away rate is calculated as the percentage of opportunities ServiceCo did not bid among the opportunities not withdrawn by the customers. Absolute win rate is calculated as the percentage of opportunities ServiceCo won among the opportunities not withdrawn by the customers. Win rate conditional on bidding is calculated as the percentage of opportunities ServiceCo won among the opportunities ServiceCo bid on.

Figure 4: Model Free Evidence

To compare walk away rate, absolute win rate, win rate conditional on bidding across opportunities of sizes, I grouped the opportunities into low revenue (1st quartile), medium revenue (2nd and 3rd quartiles), and high revenue (4th quartile). Panel A in Figure 4 shows the walk away rate, absolute win rate, win rate conditional on bidding across opportunities of different sizes. The difference in walk away rate between low revenue opportunities and high revenue opportunities is statistically significant ($p < 0.01$). Similarly, the difference in walk away rate between medium revenue opportunities and high revenue opportunities is also statistically significant ($p < 0.05$). The difference in absolute win rate between low revenue opportunities and high revenue opportunities is statistically significant ($p < 0.01$). Similarly, the differences in absolute win rate between medium revenue opportunities and high revenue opportunities is also statistically significant ($p < 0.01$). The difference in win rate conditional on bidding between medium revenue opportunities and high revenue opportunities is statistically significant ($p < 0.01$).

Panel B in Figure 4 shows the walk away rate, absolute win rate, win rate conditional on bidding among opportunities with different relationship types. The differences in walk away rates between different groups of opportunities based on relationship type are statistically significant ($p < .05$). Overall, the walk rate decreases as the strength of relationship becomes stronger. The differences in absolute win rate rates between different groups of opportunities based on relationship type are statistically significant ($p < .01$) expect for the difference between indirect relation group and no relationship at new location group. Similarly, the differences in win rate conditional on bidding between different groups of opportunities based on relationship type are statistically significant ($p < .01$) expect for the difference between direct relation group and no relationship at new location group. Appendix A shows the walk away rate, absolute win rate, win rate conditional on bidding among opportunities with different relationship types within low revenue, medium revenue, and high revenue groups.

Panel C in Figure 4 shows the walk away rate, absolute win rate, win rate conditional on bidding among opportunities with different service levels. The differences in walk away rates between different groups of opportunities based on service levels are statistically significant ($p < .1$) expect for the difference between service levels 1 and 3, and service levels 3 and 4. The difference in absolute win rate rates between service levels 3 and 4 is statistically significant ($p < .01$). Similarly, the differences in win rate conditional on bidding between service levels 2 and 4, and service levels 3 and 4 are statistically significant ($p < .01$).

Panel D in Figure 4 shows the walk away rate, absolute win rate, win rate conditional on bidding among opportunities of different strategic segments. There are no statistically significant differences in walk away rate, absolute win rate, win rate conditional on bidding among opportunities of these different strategic segments.

Panel E in Figure 4 shows the walk away rate, absolute win rate, win rate conditional on bidding among opportunities from strategic-account customers and non-strategic-account customers. There is a statistically significant difference in walk away rate ($p < .01$) and absolute win rate ($p < .05$) between opportunities from strategic-account customers and non-strategic-account customers. However, there is no significant difference on win rate conditional on bidding between the two groups.

Regression Analysis

I use moderated regression analysis to examine the effect of bid characteristics including size and relationship type on customers' decision to withdraw the bid, ServiceCo's decision to bid for the opportunity and ServiceCo winning the bid. I model these decisions in three different stages and accordingly I estimate three separate equations.

Stage 0: Customer's decision to withdraw the opportunity

The following equation represents the model used in my analysis for this stage. In the below model, β_t represents time-period dummies and β_j opportunity country dummies.

$$
\begin{aligned}
(1) \quad & \mathit{Customer\ Withdraws\ the\ Opportunity}_{\mathrm{ijt}} \\
&= \beta_0 + \beta_j + \beta_t + \beta_1\left(\mathit{Size\ of\ the\ Project}_{\mathrm{ijt}}\right) \\
&+ \beta_2 \left(\mathit{Relationship\ type}_{\mathrm{ijt}}\right) \\
&+ \beta_3 \left(\mathit{Size\ of\ the\ Project}_{\mathrm{ijt}}\right) x \left(\mathit{Relationship\ type}_{\mathrm{ijt}}\right) \\
&+ \beta_4 \left(\mathrm{Strategic\ Segment}_{\mathrm{ijt}}\right) + \beta_5 \left(\mathrm{Service\ Level}_{\mathrm{ijt}}\right) \\
&+ \beta_6 \left(\mathrm{Strategic\ Account}_{\mathrm{ijt}}\right) + \varepsilon_{\mathrm{ijt}}
\end{aligned}
$$

Table 3: Results of Stage 0 Regression Analysis

VARIABLES	Stage 0 (1)		(2)	
	Customer Withdraws the Opportunity (1 or 0)			
	Est	**SE**	**Est**	**SE**
Size of the Project				
Revenue in Million Euros	-0.001	(0.002)	-0.002	(0.002)
Relationship Type				
Direct Relationship with Buyer on Site	-0.071***	(0.025)	-0.069***	(0.023)
Indirect Relationship at Old Location	-0.020	(0.026)	-0.029	(0.025)
Indirect Relationship at New Location	0.007	(0.041)	0.004	(0.040)
No Relationship at New Location	0.020	(0.031)	-0.004	(0.027)
No Relationship at Old Location				
Relationship Type x Size of the Project				
Direct Relationship x Revenue	-0.004	(0.004)	-0.002	(0.004)
Indirect Relationship at Old Location x Revenue	0.002	(0.004)	0.001	(0.005)
Indirect Relationship at New Location x Revenue	-0.000	(0.003)	-0.001	(0.003)
No Relationship at New Location x Revenue	-0.002	(0.003)	0.001	(0.003)
No Relationship at old Location x Revenue				
Strategic Segment				
Mining				
Corporate Services	0.121	(0.105)	0.171*	(0.095)
Energy	-0.010	(0.028)	0.015	(0.027)
Government Services	0.110	(0.140)	0.147	(0.118)
Healthcare	0.207	(0.247)	0.275	(0.257)
No Segment	-0.223***	(0.083)	-0.184*	(0.105)
Schools	-0.089***	(0.029)	-0.016	(0.024)
Service Level				
1				
2	-0.025	(0.024)	-0.048**	(0.023)
3	-0.037	(0.030)	-0.045	(0.030)
4 (Integrated Services)	-0.028	(0.025)	-0.032	(0.025)

Table 3 Continued: Results of Stage 0 Regression Analysis				
	Stage 0			
VARIABLES	(1)		(2)	
	Customer Withdraws the Opportunity (1 or 0)			
	Est	SE	Est	SE
Strategic Account				
No				
Yes	0.006	(0.030)	0.006	(0.027)
Opportunity Country FE	Yes		Yes	
Year FE	Yes		Yes	
Year FE * Opportunity County FE	No		Yes	
Constant	0.342	(0.239)	0.510*	(0.261)
Observations	4,574		4,574	
R^2	0.123		0.228	

Robust standard errors in parentheses, clustered at Parent ID level
*** p<0.01, ** p<0.05, * p<0.1

Table 3 shows the results from the moderated regression analysis for the model specified in equation (1). Given the large number of fixed effects, I use a linear probability model following Goldfarb and Tucker (2011). The standard errors in the regressions are robust and are clustered at the parent company of the opportunity level. In Table 3, Model (1) shows the regression results for model with opportunity country and opportunity year fixed effects whereas Model (2) includes opportunity country – opportunity year fixed effects along with the opportunity country and opportunity year fixed effects.

In Model (2), the coefficient of *Size of the Project* ($\beta = -.002$, $p > .1$) is not statistically significant. This indicates that the size of the project has no effect on the customer's likelihood to withdraw an opportunity and they are equally like to withdraw a smaller project as they are to withdraw a larger project. For *Relationship type*, the significant coefficient on direct relationship type ($\beta = -.069$, $p < .01$) indicates that compared to opportunities where ServiceCo has no relationship with customer, opportunities where ServiceCo has a direct relationship with customer at the focal site are significantly less likely to be withdrawn. The non-significant coefficients ($p > .1$) on other relationship types indicate that the likelihood of withdrawal for opportunities with these relationship types is no different from likelihood of withdrawal for opportunities where ServiceCo has no relationship with the customer at an old location.

For *Strategic Segment*, compared to opportunities in mining segment, opportunities in corporate services segment are more likely to be withdrawn ($\beta = .171$, $p < .1$), opportunities with no segment are less likely to be withdrawn ($\beta = -.184$, $p < .1$), and for opportunities in other segments, there is no difference in likelihood of withdrawal ($p > .1$). For *Service Level*, compared to opportunities in service level 1, opportunities in service level 2 are less likely to be withdrawn ($\beta = - .184$, $p < .1$), and for opportunities in service levels, there is no difference in likelihood of withdrawal ($p > .1$). For *Strategic Account*, there is no difference in likelihood of withdrawal for opportunities from strategic accounts compared to opportunities from non-strategic accounts ($\beta = .006$, $p > .1$).

$$(2) \quad ServiceCo\ Bids\ for\ the\ Opportunity_{\mathrm{ijt}}$$

$$= \gamma_0 + \gamma_{\mathrm{j}} + \gamma_{\mathrm{t}} + \gamma_1 (Size\ of\ the\ Project_{\mathrm{ijt}})$$

$$+ \gamma_2\ (Relationship\ type_{\mathrm{ijt}})$$

$$+\ \gamma_3\ (Size\ of\ the\ Project_{\mathrm{ijt}}) x (Relationship\ type_{\mathrm{ijt}})$$

$$+ \gamma_4\ (\text{Strategic Segment}_{\mathrm{ijt}}) + \gamma_5\ (\text{Service Level}_{\mathrm{ijt}})$$

$$+ \gamma_6 (\text{ Strategic Account}_{\mathrm{ijt}}) + \mu_{\mathrm{ijt}}$$

Stage 1: ServiceCo's decision to bid for the opportunity

The following equation represents the model used in my analysis for this stage. In the below model, γ_t represents time-period dummies and γ_j opportunity country dummies. Table 4 shows the results from the moderated regression analysis for the model specified in equation (2). I include opportunities that are not withdrawn by the customers in this analysis (N=4,024). Similar to Stage 0, I use a linear probability model for my analysis, the standard errors in the regressions are robust and are clustered at the parent company of the opportunity level. In Table 4, Model (1) shows the regression results for model with opportunity country and opportunity year fixed effects whereas Model (2) includes opportunity country – opportunity year fixed effects along with the opportunity country and opportunity year fixed effects.

In Model (2), the coefficient of *Size of the Project* ($\beta = -.006$, $p < .05$) is statistically significant. This indicates that as size of the project increases, ServiceCo is

Table 4: Results of Stage 1 Regression Analysis

VARIABLES	Stage 1 (1)		(2)	
	ServiceCo Bids for the Opportunity (1 or 0)			
	Est	**SE**	**Est**	**SE**
Size of the Project				
Revenue in Million Euros	-0.007**	(0.003)	-0.006**	(0.003)
Relationship Type				
Direct Relationship with Buyer on Site	0.425***	(0.036)	0.423***	(0.036)
Indirect Relationship at Old Location	0.225***	(0.048)	0.198***	(0.040)
Indirect Relationship at New Location	0.346***	(0.052)	0.321***	(0.052)
No Relationship at New Location	-0.026	(0.078)	-0.020	(0.068)
No Relationship at Old Location				
Relationship Type x Size of the Project				
Direct Relationship x Revenue	0.003	(0.006)	0.001	(0.006)
Indirect Relationship at Old Location x Revenue	0.002	(0.006)	0.001	(0.005)
Indirect Relationship at New Location x Revenue	-0.001	(0.007)	0.000	(0.006)
No Relationship at New Location x Revenue	-0.002	(0.007)	-0.004	(0.006)
No Relationship at old Location x Revenue				
Strategic Segment				
Mining				
Corporate Services	-0.116	(0.226)	-0.008	(0.160)
Energy	-0.035	(0.040)	-0.036	(0.037)
Government Services	0.137*	(0.077)	0.136**	(0.056)
Healthcare	0.112*	(0.058)	0.125	(0.097)
No Segment	0.697***	(0.116)	0.670***	(0.115)
Schools	0.357***	(0.071)	0.196**	(0.078)
Service Level				
1				
2	0.098**	(0.039)	0.085***	(0.032)
3	0.049	(0.045)	0.030	(0.036)
4 (Integrated Services)	0.073*	(0.040)	0.065*	(0.034)

Table 4 Continued: Results of Stage 1 Regression Analysis				
	Stage 1			
VARIABLES	**(1)**		**(2)**	
	ServiceCo Bids for the Opportunity (1 or 0)			
	Est	**SE**	**Est**	**SE**
Strategic Account				
No				
Yes	0.018	(0.041)	-0.004	(0.040)
Opportunity Country FE	Yes		Yes	
Year FE	Yes		Yes	
Year FE * Opportunity County FE	No		Yes	
Constant	0.433***	(0.115)	0.759***	(0.143)
Observations	4,024		4,024	
R^2	0.216		0.306	

Robust standard errors in parentheses, clustered at Parent ID level
*** p<0.01, ** p<0.05, * p<0.1

less likely bid on the opportunity. For *Relationship type*, the significant coefficient on direct relationship type (β =.423, $p < .01$) indicates ServiceCo is more likely to bid on opportunities of that relationship type compared to opportunities where ServiceCo has no relationship at an old location. Similarly, the significant coefficient on indirect relationship at old location type (β =. 198, $p < .01$) and indirect relationship at new location type (β = .321, $p < .01$) indicate ServiceCo is more likely to bid on opportunities of those relationship types compared to opportunities where ServiceCo has no relationship at an old location. The non-significant coefficient on no relationship at new location (β = -.020, $p > .1$) indicates that there is no difference in ServiceCo's likelihood to bid on opportunities of that relationship type compared to opportunities where ServiceCo has no relationship at an old location. The non-significant coefficients ($p > .1$) on interaction terms between *Relationship type* and *Size of the Project* indicate that none of the above-mentioned effects of *Relationship type* are moderated by the *Size of the Project*.

For *Strategic Segment*, compared to opportunities in the mining segment, ServiceCo is more likely to bid on opportunities in government services segment ($\beta = .136$, $p < .1$), opportunities with no segment ($\beta = .670$, $p < .1$), and opportunities in schools' segment ($\beta = .196$, $p < .1$). For opportunities in other segments, there is no difference in ServiceCo's likelihood to bid ($p > .1$) on them compared to opportunities in the mining segment. For *Service Level*, compared to opportunities in service level 1, ServiceCo is more likely to bid on opportunities in service level 2 ($\beta = .085$, $p < .01$), and service level 4 ($\beta = .065$, $p < .1$). However, there is no difference in ServiceCo's likelihood to bid on opportunities in service level 1 and service level 3 ($\beta = .030$, $p > .1$). For *Strategic Account*, there is no difference in ServiceCo's likelihood to bid on opportunities from strategic accounts compared to opportunities from non-strategic accounts ($\beta = -.004$, $p > .1$).

Stage 2: ServiceCo wins the opportunity

The following equation represents the model used in my analysis for this stage. In the below model, δ_t represents time-period dummies and δ_j opportunity country dummies.

$$\text{(3)} \quad \textit{ServiceCo Wins the Opportunity}_{\mathrm{ijt}}$$

$$= \delta_0 + \delta_{\mathrm{j}} + \delta_{\mathrm{t}} + \delta_1(\textit{Size of the Project}_{\mathrm{ijt}})$$

$$+ \delta_2 (\textit{Relationship type}_{\mathrm{ijt}})$$

$$+ \delta_3 (\textit{Size of the Project}_{\mathrm{ijt}}) x (\textit{Relationship type}_{\mathrm{ijt}})$$

$$+ \delta_4 (\text{Strategic Segment}_{\mathrm{ijt}}) + \delta_5 (\text{Service Level}_{\mathrm{ijt}})$$

$$+ \delta_6 (\text{Strategic Account}_{\mathrm{ijt}}) + \sigma_{\mathrm{ijt}}$$

Table 5 shows the results from the moderated regression analysis for the model specified in equation (3). I include opportunities that ServiceCo bid on in this analysis (N=2,821). Similar to Stage 0 and Stage 1, I use a linear probability model for my analysis, the standard errors in the regressions are robust and are clustered at the parent company of the opportunity level. In Table 5, Model (1) shows the regression results for model with opportunity country and opportunity year fixed effects whereas Model (2) includes opportunity country – opportunity year fixed effects along with the opportunity country and opportunity year fixed effects. These results reflect ServiceCo's likelihood of winning a bid conditional on bidding for the opportunity. Results of a model with all opportunities that are not withdrawn by the customers (N=4,024) documenting ServiceCo's likelihood of winning a bid conditional on customers not withdrawing the opportunity are reported in Appendix B.

Table 5: Results of Stage 2 Regression Analysis

VARIABLES	Stage 2 (1)		(2)	
	ServiceCo Wins the Opportunity (1 or 0)			
	Est	**SE**	**Est**	**SE**
Size of the Project				
Revenue in Million Euros	-0.009	(0.007)	-0.011*	(0.006)
Relationship Type				
Direct Relationship with Buyer on Site	0.292***	(0.058)	0.336***	(0.054)
Indirect Relationship at Old Location	0.091	(0.059)	0.108*	(0.055)
Indirect Relationship at New Location	0.494***	(0.060)	0.509***	(0.055)
No Relationship at New Location	0.286***	(0.067)	0.266***	(0.063)
No Relationship at Old Location				
Relationship Type x Size of the Project				
Direct Relationship x Revenue	0.033***	(0.011)	0.024**	(0.010)
Indirect Relationship at Old Location x Revenue	-0.006	(0.009)	-0.011	(0.010)
Indirect Relationship at New Location x Revenue	0.010	(0.009)	0.009	(0.009)
No Relationship at New Location x Revenue	-0.029***	(0.010)	-0.021**	(0.010)
No Relationship at old Location x Revenue				
Strategic Segment				
Mining				
Corporate Services	0.255**	(0.107)	0.230*	(0.134)
Energy	-0.060	(0.048)	-0.039	(0.048)
Government Services	0.127	(0.150)	0.124	(0.158)
Healthcare	-0.109	(0.229)	-0.003	(0.299)
No Segment	-0.691***	(0.137)	-0.744***	(0.120)
Schools	0.735***	(0.104)	0.765***	(0.178)
Service Level				
1				
2	-0.053	(0.060)	-0.024	(0.057)
3	-0.136*	(0.076)	-0.151**	(0.070)
4 (Integrated Services)	0.002	(0.055)	0.018	(0.050)

Table 5 Continued: Results of Stage 2 Regression Analysis				
	Stage 2			
VARIABLES	**(1)**		**(2)**	
	ServiceCo Wins the Opportunity (1 or 0)			
	Est	**SE**	**Est**	**SE**
Strategic Account				
No				
Yes	-0.020	(0.050)	-0.006	(0.047)
Opportunity Country FE	Yes		Yes	
Year FE	Yes		Yes	
Year FE * Opportunity County FE	No		Yes	
Constant	0.711***	(0.112)	1.012***	(0.146)
Observations	2,821		2,821	
R^2	0.254		0.369	

Robust standard errors in parentheses, clustered at Parent ID level
*** p<0.01, ** p<0.05, * p<0.1

In Model (2), the coefficient of *Size of the Project* (β = -.011, p < .1) is statistically significant. This indicates that, conditional on ServiceCo biding for the opportunity, as size of the project increases, ServiceCo is less likely to win. For *Relationship type*, the significant coefficient on direct relationship type (β =.336 p < .01) indicates that, conditional on bidding, ServiceCo is more likely to win opportunities of that relationship type compared to opportunities where ServiceCo has no relationship at an old location. This positive effect is significantly strengthened as the size of the project increases as indicated by the coefficient on the interaction term between direct relationship type and revenue (β = .024 p < .05). Similarly, the significant coefficient on indirect relationship at old location type (β =. 108, p < .1), indirect relationship at new location type (β = .509, p < .01), and no relationship at new location type (β = .266, p < .01) indicate that, conditional on bidding, ServiceCo is more likely to win opportunities of those relationship types compared to opportunities where ServiceCo has no relationship at an old location. The positive effect of no relationship at new location relative to no relationship at old location is weakened as the size of the project increases as indicated by the coefficient on the interaction term between no relationship at new location relationship type and revenue ($\beta = -.021$, $p < .05$).

For *Strategic Segment*, compared to opportunities in the mining segment, conditional on bidding, ServiceCo is more likely to win opportunities in corporate services segment ($\beta = .230$, $p < .1$), and opportunities in schools' segment ($\beta = .735$, $p < .01$). However, compared to opportunities in the mining segment, conditional on bidding, ServiceCo is less likely to win opportunities in with no segment ($\beta = -.744$, $p < .01$). For opportunities in other segments, conditional on bidding, there is no difference in ServiceCo's likelihood to win ($p > .1$) compared to opportunities in the mining segment. For *Service Level*, compared to opportunities in service level 1, conditional on bidding, ServiceCo is less likely to win opportunities in service level 3 ($\beta = -.151$, $p < .05$. However, conditional on bidding, there is no difference in ServiceCo's likelihood to win opportunities in service level 1 and service level 2 ($\beta = -.024$, $p > .1$), and service level 1 and service level 4 ($\beta = .018$, $p > .1$). For *Strategic Account*, conditional on bidding, there is no difference in ServiceCo's likelihood to win opportunities from strategic accounts compared to opportunities from non-strategic accounts ($\beta = -.006$, $p > .1$).

Machine Learning Model to Predict ServiceCo's Likelihood of Winning the Bid

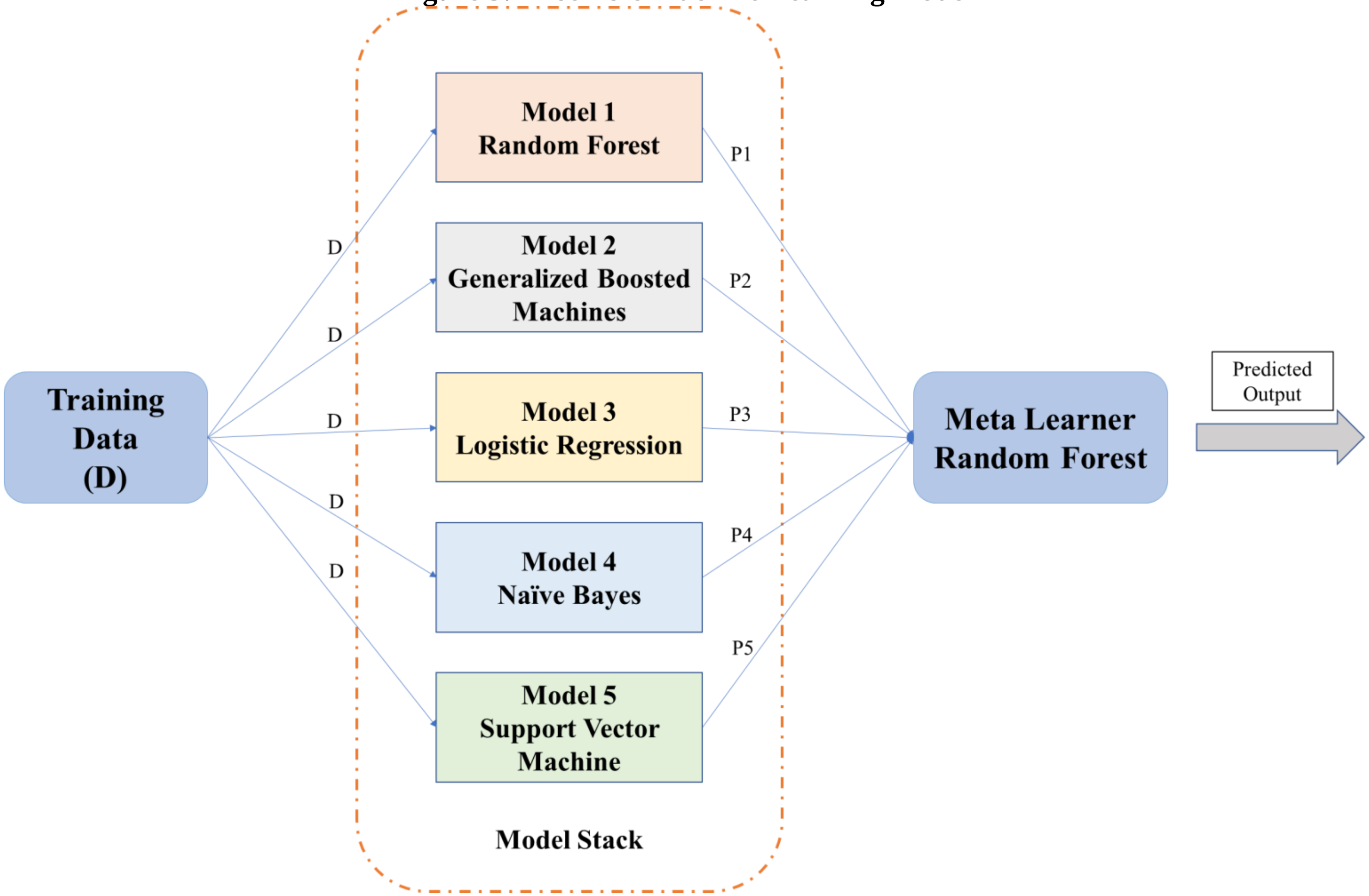
In this section, I explain the ensemble learning approach used to develop a supervised machine learning model to predict ServiceCo's likelihood of winning the bid.

Ensemble approaches are used to create a single optimal prediction model by combining multiple machine learning algorithms or models. More specifically, I use an ensemble stacking approach in which multiple machine learning models are trained in parallel and then a meta-learner is trained to predict the outcome using predictions of the earlier learned models. In my context, I use predictions from the below mentioned machine learning models as input to the meta-leaner. Figure 5 presents a visualization of the ensemble stacking approach used in my context. Appendix C provides a review of studies using predictive models in the context of B2B opportunity management.

Random Forest (RF): A random forest algorithm, developed by Breiman (2001), builds a function capturing the relationship between independent variables and the dependent variable by building a large set of decision trees. Each decision tree is created from a bootstrap sample of training examples and a limited number of randomly selected attributes. During the classification process, each tree predicts a class for each testing instance and a majority vote is used to arrive at the overall prediction for the testing instance. I use Kuhn et al.'s (2020) implementation of the random forest algorithm in R to develop my RF model.

Generalized Boosted Machines (GBM): A GBM algorithm builds a function capturing the relationship between independent variables and the dependent variable in three steps: first, it fits a decision tree (tree 1) to the data; second, it fits another decision tree (tree 2) to the residuals of the first tree; third, it fits a boosting ensemble (trees 1 and 2) to the data; and finally, it iteratively fits a boosting ensemble (trees 1 and 2) to the data. Iterative refinement of models based on prior modeling representation errors leads to an

Figure 5: Ensemble Machine Learning Model



accurate final prediction. I use Kuhn et al.'s (2020) implementation of the GBM algorithm in R to develop my GBM model.

Logistic Regression: A logistic regression models the log-odds of a data instance belonging to a specific class as a linear function of the independent variables. The estimated coefficients are used to predict the probability of the data instance belonging to a specific class.

Naïve Bayes: Naïve Bayes classifier is a probabilistic classifier that applies Bayes' theorem with a strong (naïve) assumption of independence between the independent variables in the model. The assumption is that the presence of a particular level of independent variable in a dependent variable class is unrelated to the presence of any other independent variable. For example, if a fruit is red, spherical, and roughly 3 inches in diameter, it is classified as an apple. Even if these characteristics are dependent on one another or on the presence of other characteristics, Naïve Bayes classifier assumes that each of these characteristics contributes on its own to the likelihood that the fruit is an apple. I use Meyer et al.'s (2015) implementation of the Naïve Bayes algorithm in R to develop my Naïve Bayes model.

Support Vector Machine (SVM): SVM is a semiparametric approach developed by Cortes and Vapnik (1995) to separate data instances of different classes (Win/Loss) by estimating a hyperplane in the multidimensional space that maximizes the distance between the nearest data points (support vectors) of the two classes. I use Karatzoglou et al.'s (2004) implementation of the SVM algorithm in R to develop my SVM model.

Random Forest Meta Learner: The main idea behind the meta-learner is to combine the predictions from different individual models used in the ensemble model and come up with the overall prediction. A meta-learner used the predictions from individual models as independent variables to predict the dependent variable. In my context, I use *Random Forest* as a meta-learner taking the predicted likelihood of winning from five different models as independent variables to predict ServiceCo's overall likelihood of winning.

I split the 2,824 opportunities ServiceCo bid on into a training set that includes a random sample of 80% of the opportunities and test set that includes the remaining 20% of the opportunities. To avoid overfitting, I used a 5-fold cross-validation procedure for hyperparameter tuning.

Model Evaluation

In this section, I compare the predictive performance of different machine learning models stacked in the ensemble model with the ensemble model in both train and test samples. Any binary classification problem usually has two classes – a positive class (ServiceCo won the bid) and a negative class (ServiceCo lost the bid) – and the model predictions can be classified into true positives (*tp*) where the model predicts a win and ServiceCo actually won the bid, false positives (*fp*) where the model predicts a win but ServiceCo actually lost the bid, true negatives (*tn*) where the model predicts a loss and ServiceCo actually lost the bid, and false negatives (*fn*) where the model predicts a loss but ServiceCo actually won the bid. I use these four values to calculate the following metrics to compare the performance of different models.

Accuracy is the proportion of correct predictions across all the predictions of a model.

$$Accuracy = \frac{(tp + tn)}{(tp + fp + tn + fn)} * 100$$

Sensitivity, also referred to as recall, is a metric used to capture a machine learning model's ability to predict true positive values. It is defined as the proportion of actual positive predictions (*tp*) across all the positive predictions of a model (*tp*+*fn*). This measure is important in applications where Type 2 errors or false negatives are costly.

$$Sensitivity = \frac{(tp)}{(tp + fn)} * 100$$

Specificity is a metric used to capture a machine learning model's ability to predict true negative values. It is defined as the proportion of actual negative predictions (*tn*) across all the negative predictions of a model (*tn*+*fp*). This measure is important in applications where Type 1 errors or false positives are costly.

$$Specificity = \frac{(tn)}{(tn + fp)} * 100$$

I use 80% stratified sample to train the models and the remaining 20% of the data as test sample. The test sample performance highlights a model's predict ability on data unseen by the model and in my context, on new opportunities encountered by ServiceCo. Table 6 shows the performance of the five different models used in the ensemble model and the ensemble model. The ensemble model outperforms all the individual models in *Accuracy*, *Sensitivity*, and *Specificity* for the training sample. The next best model is the random forest model. For the test sample, the ensemble model outperforms all the individual models in *Accuracy*, and *Specificity* followed by the random forest model. On

Sensitivity in the test sample, the ensemble model and the random forest model perform equally well.

Partial Dependence Plots

In this section, I plot and analyze partial dependence plots (PDPs) based on Friedman (2001) to understand the marginal dependence of predicted outcome of the ensemble model (ServiceCo's propensity to win the bid) on different independent variables. A detailed approach to developing PDPs is explained in Appendix D. The PDPs obtained for the ensemble model are shown in Figure 6. Each plot represents the change in ServiceCo's propensity to win a bid as a function of the different independent variables included in the ensemble model. I use Tukey's honest significance test to compare the differences is marginal effects of various levels of categorical independent variables included in the ensemble model.

For *Size of the Project*, first plot in Figure 6 captures the non-linearities in the relationship with ServiceCo's propensity to win. The marginal win propensity fluctuates between the 50% threshold between $0 and $8 million USD, and subsequently the win propensity follows a downward trend between $8 million and $43 million USD with slight fluctuations. The overall slope of the curve is -.156 and this slope is significantly different from 0 ($p < .01$) indicating that the size of the project has an overall negative marginal effect on ServiceCo's propensity to win the bid.

For *Relationship type*, second plot in Figure 6 captures ServiceCo's marginal win propensity for different relationship types. The average win propensity for bids where

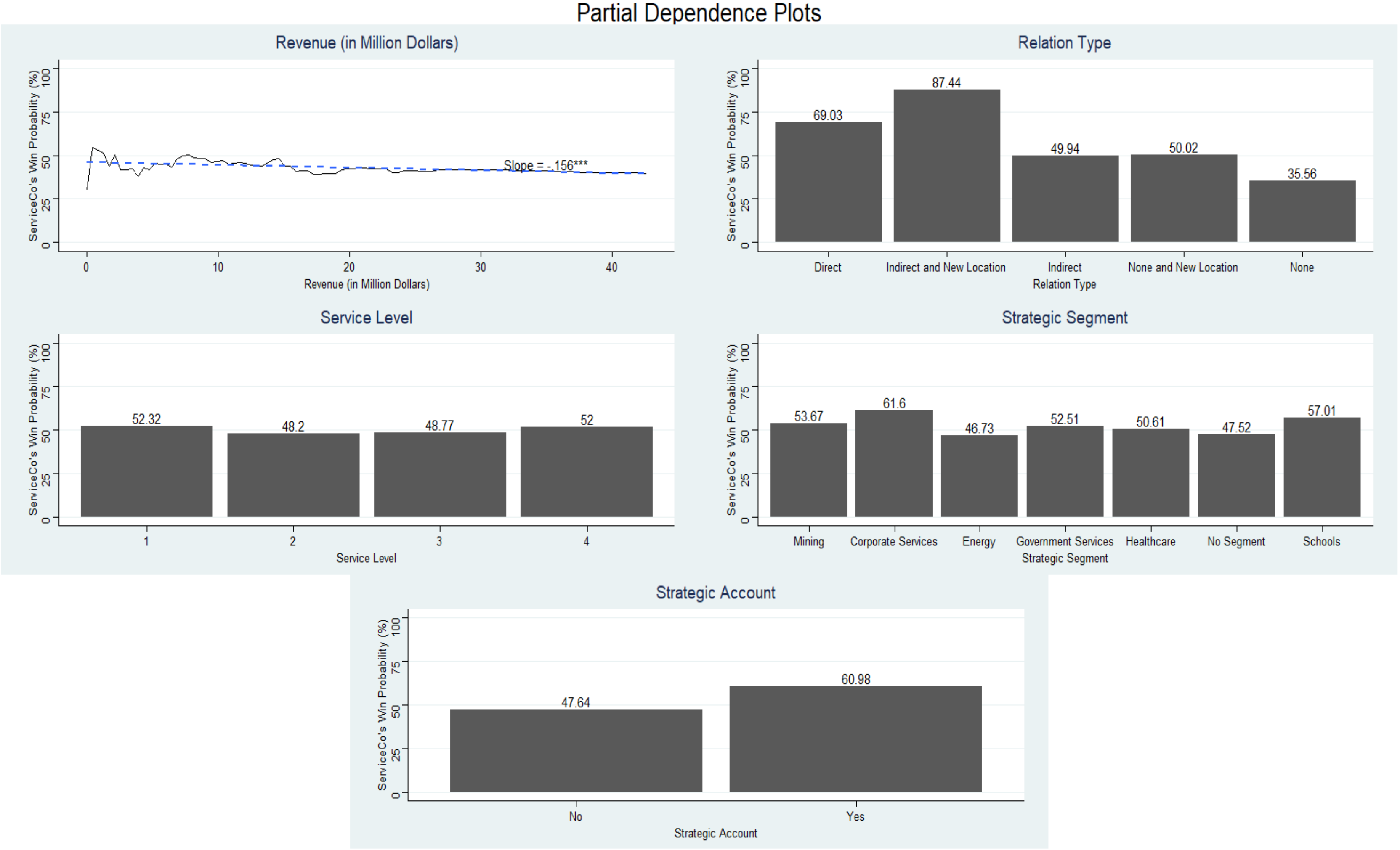
Table 6: Performance of Different Machine Learning Models

Machine Learning Model	Training sample (80 % Data)			Test Sample (20% Data)		
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
Random Forest	94.72	92.97	96.37	76.42	75.18	77.59
Generalized Boosted Machines	76.03	74.64	77.36	72.16	67.88	76.21
Logistic Regression	69.95	62.5	77.01	68.79	59.85	77.24
Naïve Bayes	64.67	72.45	57.3	65.25	72.99	57.93
Support Vector Machine	74.61	71.9	77.18	72.16	68.25	75.86
Ensemble Model	**99.25**	**99.18**	**99.31**	**77.48**	**75.18**	**79.66**

ServiceCo has a direct relationship with the customer is 69.03 % and this value is significantly different from other types of relationships ($p < .01$). The average win propensity for bids where ServiceCo has an indirect relationship with the customer and a new project location is 87.44 % and this value is significantly different from other types of relationships ($p < .01$). The average win propensity for bids where ServiceCo has an indirect relationship with the customer and an old project location is 49.94 % and this value is significantly different from all other types of relationships ($p < .01$) except no relationship and new project location ($p > .1$). The average win propensity for bids where ServiceCo has no relationship with the customer and a new project location is 50.02 % and this value is significantly different from all other types of relationships ($p < .01$) except indirect relationship and old project location ($p > .1$). The average win propensity for bids where ServiceCo has no relationship with the customer and an old project location is 35.56 % and this value is significantly different from other types of relationships ($p < .01$). Overall, this shows that relationships are valued by ServiceCo's customers and as the strength of the relationship increases, ServiceCo's marginal win propensity increases.

For *Strategic Segment*, fourth plot in Figure 6 captures ServiceCo's marginal win propensity for different types of strategic segments. The average win propensity for bids across mining segment is 53.67%, corporate services is 61.60%, energy is 46.73%, government services is 52.51%, healthcare is 50.61%, bids with no segment is 47.52%,

Figure 6: Partial Dependence Plots following Freedman (2001)



and schools is 57.01%. There are significant differences ($p < .1$) between the different types of strategic segments except for mining – government services, mining – health care, energy – no segment, and government services – health care.

For *Service Level*, third plot in Figure 6 captures ServiceCo's marginal win propensity for different service levels. The average win propensity for bids across service level 1 is 52.32% and this value is significantly different from bids across service level 2 ($p < .01$) and service level 3 ($p < .05$) and not different from bids across service level 4 ($p > .1$). The average win propensity for bids across service level 2 is 48.20% and this value is significantly different from bids across service level 1 ($p < .01$) and service level 4 ($p < .01$) and not different from bids across service level 3 ($p > .1$). The average win propensity for bids across service level 3 is 48.77% and this value is significantly different from bids across service level 1 ($p < .05$) and service level 4 ($p < .01$) and not different from bids across service level 3 ($p > 0.1$). The average win propensity for bids across service level 4 is 52.00% and this value is significantly different from bids across service level 2 ($p < .01$) and service level 3 ($p < .05$) and not different from bids across service level 1 ($p > .1$).

For *Strategic Account*, fifth plot in Figure 6 captures ServiceCo's marginal win propensity for bids from strategic-account customers and non-strategic-account customers. The average win propensity for bids from non-strategic-account customers is 47.64 % and this value is statistically different ($p < .01$) from marginal win propensity for bids from strategic-account customers of 60.98 %.

CHAPTER IV

OPTIMIZING OPPORTUNITY PURSUIT PORTFOLIO

The primary purpose of the machine learning model is to capture the functional relationship between different attributes of an opportunity and the likelihood of ServiceCo's propensity to win. However, sales managers need a practical tool that recommends a subset of opportunities to pursue from a large pool of opportunities while considering different strategic capacity constraints. Accordingly, in this section I develop a normative optimization model wherein I view the sales manager's objective as determining the specific set of opportunities to go after under different constraints of capacity. For example, some firms might want to maximize revenue, some firms would like to maximize profits, and some firms might want to maximize sales while maintaining a specific percentage level of profits. Firms might even want to maintain a specific level of business in different sub segments (service levels or strategic segments or different relationship types). Irrespective of the firms' objectives, this general framework will help firms make use of the machine learning framework and identify a subset of opportunities to pursue while not expending effort on opportunities where they have a lower likelihood of winning.

Optimization Model

The input data for the optimization model is derived from the predictive machine learning model developed earlier, which serves as a sales response function. The optimization model computed the optimal revenue maximizing portfolio of projects to pursue from a large pool of opportunities available for the firm to pursue. Model M1

Table 7: Optimization Results with Capacity Constraint on Total Number of Projects ServiceCo can Bid in a Year

Year	Actual Data			Projected Values Based on Predictive Model and Solver			Percentage Change in Revenue	Percentage Change in Number of Bids
	Number of Projects bid	Total Revenue Bid in Million Euros	Total Revenue Won in Million Euros	Number of Projects - Should bid	Total Revenue Should Bid in Million Euros	Total Conditional Revenue Won in Million Euros		
2010	6	1.95	1.95	1	0.57	0.57	-71.00%	-83.33%
2011	61	78.57	24.27	30	40.71	40.71	67.71%	-50.82%
2012	53	71.77	41.07	29	33.62	33.62	-18.12%	-45.28%
2013	50	111.10	61.26	49	57.11	57.11	-6.78%	-2.00%
2014	61	144.41	16.86	17	21.00	21.00	24.56%	-72.13%
2015	127	59.66	22.82	15	17.10	17.10	-25.04%	-88.19%
2016	121	176.91	77.05	48	74.70	74.70	-3.05%	-60.33%
2017	451	553.94	175.97	235	217.26	217.26	23.46%	-47.89%
2018	562	704.75	229.37	309	335.98	335.98	46.48%	-45.02%
2019	413	496.50	219.08	240	220.59	220.59	0.69%	-41.89%
2020	461	763.24	420.92	347	471.06	471.06	11.91%	-24.73%
2021	455	641.33	357.69	436	522.27	522.27	46.01%	-4.18%
Total	2821	3804.15	1648.33	1756	2011.97	2011.97	22.06%	-37.75%

describes the optimization problem to be solved to arrive at a revenue maximizing portfolio for time period t. The objective function maximizes the prospect revenue in a specific time period t and the capacity constraint restricts the capacity of the total number of projects bid on in the time period t.

$$(M1) \qquad maximize \sum R_{i,t} * X_{i,t} * \widehat{W_{\iota,t}}$$

$$subject\ to: \ \sum X_{i,t} \leq C_t$$

where $R_{i,t}$ is the revenue from project i, $X_{i,t}$ indicates the decision variable if the focal firm bid on project i or not ($X_{i,t} \in \{0,1\}$), C_t is the capacity constraint in time t, and $\widehat{W_{\iota,t}}$ is the predicted win or loss of project i from the ensemble machine learning model developed earlier.

I apply this optimization approach to the ServiceCo's 4,024 opportunities to identify the subset of opportunities ServiceCo should have pursued each year. Table 7 shows a comparison the portfolio with the actual set of projects ServiceCo bid on every year with the opportunities identified by the optimization model. The results of the optimization model indicate that ServiceCo should have bid on a total of 1,756 project with a total revenue of 2,011.97 million dollars between 2010-2021. Whereas ServiceCo bid on 2,821 projects with total value of 3,804.15 million dollars and had a revenue of 1,648.33 million dollars. This indicates that ServiceCo could have had a 22.06% revenue increase while bidding on 37.75% lesser number of projects. The results document that, by using the optimization model, ServiceCo would have gained revenue in some years whereas it lost revenue in certain years. For example, in the year 2010, ServiceCo bid on

6 projects accounting for a total of 1.95 million dollars. Whereas the optimization model recommends bidding on one project with an expected revenue of 0.57 million dollars resulting in a 71% decrease in revenue compared to the reality. Similarly, in the year 2011, ServiceCo bid on 53 projects accounting for a total revenue of 41.07 million dollars. Whereas the optimization model recommends bidding on 29 projects with an expected revenue of 33.62 million dollars resulting in a 18.12% decrease in revenue compared to the reality.

In contrast, in the year 2010, ServiceCo bid on 61 projects accounting for a total revenue of 24.27 million dollars. Whereas the optimization model recommends bidding on 30 projects with an expected revenue of 40.71 million dollars resulting in a 67.71% increase in revenue compared to the reality. Similarly, in the year 2021, ServiceCo bid on 455 projects accounting for a total revenue of 357.69 million dollars. Whereas the optimization model recommends bidding on 436 projects with an expected revenue of 522.27 million dollars resulting in a 46.01% increase in revenue compared to the reality.

Analysis of Type I and Type II Errors by ServiceCo Relative to the Optimization Model

In this section, I identify the opportunities did not bid by ServiceCo, but the optimization model identified as opportunities that ServiceCo should bid (Type I errors) and the opportunities bid by ServiceCo but the optimization model identified as opportunities that ServiceCo should not bid (Type II errors). I subsequently analyze the distribution of different opportunity characteristics of these identifies opportunities.

Table 8: Distribution of Bid Characteristics of Type I and Type II Errors by ServiceCo Relative to the Optimization Model

Variable	ServiceCo didn't bid when model suggested to bid	ServiceCo bid when model suggested not to bid
N	384	1,445
Size of the Project		
Revenue- Mean (SD)	1.04 (2.07)	1.51 (3.1)
Relationship Type		
Direct Relationship with Buyer on Site	4.30%	24.40%
Indirect Relationship at Old Location	22.70%	26.30%
Indirect Relationship at New Location	6.80%	1.90%
No Relationship at New Location	33.10%	7.80%
No Relationship at Old Location	33.10%	39.70%
Strategic Segment		
Mining	20.60%	26.20%
Corporate Services		0.10%
Energy	79.40%	73.40%
Government Services		0.10%
Healthcare		0.10%
No Segment		0.02%
Schools		
Service Level		
1	10.20%	8.40%
2	15.60%	30.20%
3	29.90%	37.10%
4 (Integrated Services)	44.30%	24.30%
Strategic Account		
Yes	13.80%	18.10%

The summary statistics of each of these groups of opportunities are shown in Table 8. There were 384 opportunities ServiceCo did not bid for which the optimization model predicted as a winning opportunity. The average revenue of these opportunities is 1.04 million dollars. These opportunities are more likely to be opportunities where ServiceCo does not have a direct or indirect relationship with the customer (66.2%). These opportunities belong to energy (79.4%) and mining service segments (20.6%). 10.2% of the opportunities belong to service level 1, 15.6% belong to service level 2, 29.9% belong to service level 3 and 44.3% belong to integrated services. 13.8% opportunities belong to strategic accounts.

Similarly, of the 2.821 opportunities ServiceCo bid on, 1,445 (51.2%) opportunities were identified by the optimization model as opportunities not to pursue. The average revenue of these opportunities is 1.514 million dollars. These opportunities are more likely to be opportunities where ServiceCo has a direct or indirect relationship with the customer (51.6%). These opportunities primarily belong to energy (73.4%) and mining service segments (26.2%). 8.4% of the opportunities belong to service level 1, 30.2% belong to service level 2, 37.1% belong to service level 3 and 24.3% belong to integrated services. 18.1% opportunities belong to strategic accounts.

CHAPTER V

DISCUSSION AND CONCLUSIONS

In this dissertation, I seek to understand the relationship between different project attributes such as size of the projects and existing relationship with the customer on a firm's propensity to bid on the project and subsequently a firm's propensity to win the bid. The results of the empirical analysis document the biases held by the sales teams in opportunity pursuit. To alleviate these inherent biases, I propose and develop a data driven decision support framework informed by extant marketing theory that can help firms prioritize the opportunities to pursue from a larger opportunity pool. These results have important theoretical and practical implications.

Theoretical Implications

My analysis offers several key insights and extend extant research. First, firms are more likely to avoid larger opportunities. The size of the project is a proxy for the complexity of the project and by extension, firms are avoiding larger more complex projects. These results are in line with the emerging literature on salesperson judgement and decision-making documenting the biases held by the salespeople in prioritizing opportunity pursuit. Also, firms are more likely to pursue opportunities where they a strong existing relationship with the customers and they are likely to leave out opportunities here they do not have any prior relationship with the customers. Moreover, there is no significant interaction effect of opportunity size and relationship with customers there by indication that this bias of large opportunity avoidance persists across opportunities where firms have an existing relationship with the customers as well as opportunities where they do not have any relationship with the customers.

Second, the results of the prescriptive optimization model document the fact that firms are leaving money on the table by not bidding on the opportunities where they have higher chance of winning. This highlights the need to adopt data driven decision support tools in the context of sales management to identify a subset of opportunities to pursue from a larger pool of opportunities available for firms to go after.

Managerial Implications

These results offer several important implications for practice, particularly senior B2B executives such as Chief Sales Officers and Chief Executive Officers. First, these results show that sales managers need to be wary of the biases they have while prioritizing sales opportunities. Moreover, the prescriptive model negative consequences of these biases in terms of the lost sales potential. In the context of our study, the focal firm could have improved their revenue by 22% by using a data driven approach for opportunity pursuit.

Second, executives can make use of the optimization framework developed in the study to identify opportunities that fit their strategic objectives. For example, if the focal firm wanted to place a higher emphasis on a specific service level, they can add those objectives as the constraints in the optimization framework and the updated model will provide a list of opportunities to pursue that fits the strategic objectives. Finally, executives can make use of this framework for effective sales forecast that will help them plan downstream production capacities.

Limitations and Future Research

First, this study uses CRM data from a single firm on opportunities the focal firm is involved. This data does not represent all the opportunities available in the market. Future research can make use of richer contexts such as B2G contracting where researchers can observe all the available opportunities through publicly available data sources and develop models for opportunity pursuit. Second, the predictive model used in this study provides a static snapshot of the predicted win probability for each opportunity. Future research can develop models which can provide a dynamic change in win probability as the opportunity advances through the sales process. This will help sales managers evaluate projects at different stages in the sales process and decide future course of action. Third, the current study makes use of structured CRM data in developing the predictive models. Future research can make use of unstructured data such as text from email exchanges with the customers, audio and video data from the sales negotiations to develop models to predict success in sales pursuit

REFERENCES

Andreassen, T. W., & Lervik, L. (1999). Perceived Relative Attractiveness Today and Tomorrow as Predictors of Future Repurchase Intention. *Journal of Service Research*, *2*(2), 164-172.

Bonde, A, Bruno, J, Wu, Ruhl, Charlie, & Birrell, Rachel (2019). U.S. B2B eCommerce Will Hit $1.8 Trillion by 2023. *Forrester Brief*, https://www.forrester.com/report/USþB2BþeCommerceþWillþHitþ18þTrillionþByþ2023/-/ERES136173.

Bosukonda, N., Shaik, M., Mittal, V., & Sridhar, S. (2020). The Asymmetric Association of Customer Commitment and Behavioral Intentions in Business-to-Business Markets. *Available at SSRN 3532546*.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Chica, M., & Rand, W. (2017). Building Agent-based Decision Support Systems for Word-of-Mouth Programs: A Freemium Application. *Journal of Marketing Research*, *54*(5), 752-767.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine learning*, *20*(3), 273-297.

CSO Insights. (2019). *Selling in the Age of Ceaseless Change: 2018-2019 Sales Performance Study*. https://suncoast-partners.com/wp-content/uploads/2020/01/2018-2019-Sales-Performance-Report-Summary.pdf.

D’Haen, J., & Van den Poel, D. (2013). Model-Supported Business-to-Business Prospect Prediction Based on an Iterative Customer Acquisition Framework. *Industrial Marketing Management*, *42*(4), 544-551.

Dwyer, F. R., Schurr, P. H., & Oh, S. (1987). Developing Buyer-Seller Relationships. *Journal of Marketing*, *51*(2), 11-27.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 1189-1232.

Grewal, R., Lilien, G. L., Bharadwaj, S., Jindal, P., Kayande, U., Lusch, R. F., ... & Sridhar, S. (2015). Business-to-Business Buying: Challenges and Opportunities. *Customer Needs and Solutions*, *2*(3), 193-208.

Grewal, R., & Sridhar, S. (2021). Commentary: Toward Formalizing Social Influence Structures in Business-to-Business Customer Journeys. *Journal of Marketing*, *85*(1), 98-102.

Heinitz, N., Habel, J., & Alavi, S. (2021). Predictive Sales Analytics: State of the Literature and a Theory of Adoption. *Available at SSRN 3994561*.

Jahromi, A. T., Stakhovych, S., & Ewing, M. (2014). Managing B2B Customer Churn, Retention and Profitability. *Industrial Marketing Management*, *43*(7), 1258-1268.

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). Kernlab-an S4 Package for Kernel Methods in R. *Journal of Statistical Software*, *11*(9), 1-20.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... & Team, R. C. (2020). Package 'CARET'. *The R Journal*, *223*, 7.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2015). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, 2015. *R Package Version*.

Miller, Andrew (2019). 5 Ways to Improve Your Sales Pipeline Management and Boost Revenue. *Prialto* (February 2019), https://www.prialto.com/blog/sales-pipeline-

management.
Mittal, V., & Sridhar, S. (2021). Strategy Planning in the Real World. In *Focus* (pp. 1-10). Palgrave Macmillan, Cham.

Mortensen, S., Christison, M., Li, B., Zhu, A., & Venkatesan, R. (2019, April). Predicting and Defining B2B Sales Success with Machine Learning. In *2019 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 1-5). IEEE.

Peter. (2019, October 28). *Go to market strategy – the cost of B2B direct selling*. brandknewmag. Retrieved May 6, 2022, from https://www.brandknewmag.com/go-to-market-strategy-the-cost-of-b2b-direct-selling/

Silva-Risso, J., & Ionova, I. (2008). Practice Prize Winner—A Nested Logit Model of Product and Transaction-Type Choice for Planning Automakers' Pricing and Promotions. *Marketing Science*, *27*(4), 545-566.

Rezazadeh, A. (2020). A Generalized Flow for B2B Sales Predictive Modeling: An Azure Machine-Learning Approach. *Forecasting*, *2*(3), 267-283.

Robinson, P. J., Faris, C. W., & Wind, Y. (1967). *Industrial Buying and Creative Marketing*. Allyn & Bacon.

Scherm, M.J. (2021). Pipeline Management: Overcoming Our Reptile Brain. In: *Scrum for Sales. Future of Business and Finance*. Springer, Cham. https://doi.org/10.1007/978-3-030-82978-0_5

Söhnchen, F., & Albers, S. (2010). Pipeline Management for the Acquisition of Industrial Projects. *Industrial Marketing Management*, *39*(8), 1356-1364.

Xu, J., Van der Borgh, M., Nijssen, E. J., & Lam, S. K. (2021). Why Salespeople Avoid Big-Whale Sales Opportunities. *Journal of Marketing*, 00222429211037336.

Zoltners, A. A., & Sinha, P. (2005). The 2004 ISMS Practice Prize Winner—Sales Territory Design: Thirty years of Modeling and Implementation. *Marketing Science*, *24*(3), 313-331.

APPENDIX A

MODEL FREE EVIDENCE – REVENUE VS. RELATIONSHIP TYPE

APPENDIX B

RESULTS OF STAGE 2 (ABSOLUTE WIN) REGRESSION ANALYSIS

Table B1: Results of Stage 2 (Absolute Win) Regression Analysis

VARIABLES	**Stage 2 - Absolute Win (1)**		**(2)**	
	ServiceCo Wins the Opportunity (1 or 0)			
	Est	**SE**	**Est**	**SE**
Size of the Project				
Revenue in Million Euros	-0.005**	(0.002)	-0.006***	(0.002)
Relationship Type				
Direct Relationship with Buyer on Site	0.422***	(0.050)	0.462***	(0.047)
Indirect Relationship at Old Location	0.164***	(0.045)	0.149***	(0.042)
Indirect Relationship at New Location	0.540***	(0.054)	0.517***	(0.052)
No Relationship at New Location	0.117*	(0.061)	0.102*	(0.057)
No Relationship at Old Location				
Relationship Type x Size of the Project				
Direct Relationship x Revenue	0.023***	(0.009)	0.011	(0.008)
Indirect Relationship at Old Location x Revenue	-0.010*	(0.005)	-0.014**	(0.006)
Indirect Relationship at New Location x Revenue	-0.004	(0.006)	-0.001	(0.005)
No Relationship at New Location x Revenue	-0.008	(0.005)	-0.007	(0.005)
No Relationship at old Location x Revenue				
Strategic Segment				
Mining				
Corporate Services	0.113	(0.194)	0.093	(0.176)
Energy	-0.055	(0.036)	-0.049	(0.035)
Government Services	0.231	(0.180)	0.182	(0.168)
Healthcare	-0.027	(0.248)	0.032	(0.289)
No Segment	-0.283***	(0.086)	-0.301***	(0.081)
Schools	0.832***	(0.080)	0.809***	(0.161)

Table B1 Continued : Results of Stage 2 (Absolute Win) Regression Analysis				
	Stage 2 - Absolute Win			
VARIABLES	**(1)**		**(2)**	
	ServiceCo Wins the Opportunity (1 or 0)			
	Est	**SE**	**Est**	**SE**
Service Level				
1				
2	0.006	(0.046)	0.028	(0.039)
3	-0.057	(0.057)	-0.076	(0.049)
4 (Integrated Services)	0.036	(0.044)	0.036	(0.038)
Strategic Account				
No				
Yes	-0.005	(0.038)	0.008	(0.036)
Opportunity Country FE	Yes		Yes	
Year FE	Yes		Yes	
Year FE * Opportunity County FE	No		Yes	
Constant	0.567***	(0.105)	0.813***	(0.118)
Observations	4,024		4,024	
R^2	0.248		0.344	

Robust standard errors in parentheses, clustered at Parent ID level
*** p<0.01, ** p<0.05, * p<0.1

APPENDIX C

RELATED LITERATURE ON PREDICTIVE MODELS IN B2B OPPORTUNITY MANAGEMENT

Reference	**Focal Dependent Variable**	**Focal Independent Variables**	**Predictive Models Used**	**Context /Data**	**Sample Size**
Bohanec et al. 2017	Win/Loss	22 attributes relevant to the specific bid context	Random Forest Naïve Bayes Support Vector Machine Decision Tree Artificial Neural Networks	B2B Software solutions Company	448 unique leads
D'Haen and Van den Poel 2013	Whether or not a Company is a client	Company Characteristics	Neural Networks	B2B company in telecommunication services industry	107 companies for training
Elite and Buxmann 2019	Conversion from 1. lead-to-opportunity 2. opportunity-to-sales 3. lead-to-sales	Customer features	CatBoost Random Forest Support Vector Machine XGBoost Decision Tree	Fortune 500 B2B software company	36,929 unique leads
Mortensen et al 2019	Win/Loss	15 opportunity level variables	Logistic Regression Decision Tree Random Forest XGBoost	Fortune 500 paper and packaging company	40,464 unique leads

Reference	Focal Dependent Variable	Focal Independent Variables	Predictive Models Used	Context /Data	Sample Size
Rezazadeh 2020	Win/Loss	20 attributes relevant to the specific bid context	35 iterations of XGBoost and Light GBM classifiers with various parameterizations are combined to generate a soft-voting ensemble classifier	Global multi-business B2B consulting firm	25,578 unique leads
Rohaan, Topan and Goothuis-Oudshoorn 2022	Win/Loss	28 Attributes relevant to the specific sales lead	Logistic Regression Gradient Boosting Classifier Random Forest	Spare part RFQ's of an after-sales service provider	NA
Yan et al 2015a	Win/Loss	Attributes relevant to the specific sales lead	Logistic Regression	500 multinational B2B technology company (IBM)	NA
Yan et al 2015b	Win/Loss over lifetime of a lead	Attributes relevant to the specific sales lead	Profile-specific two-dimensional Hawkes processes model	500 multinational B2B technology company (IBM)	NA

Reference	Focal Dependent Variable	Focal Independent Variables	Predictive Models Used	Context /Data	Sample Size
Zahid et al. 2021	Win/Loss/No Bid/Customer Did Not Pursue (Multi Class Classification)	93 features relevant to the specific sales lead	Tree-based Light GBM Classifier	One of the world's largest IT service providers (IBM)	6,002,014 unique leads

REFERENCES

Bohanec, M., Borštnar, M. K., & Robnik-Šikonja, M. (2017). Explaining Machine Learning Models in Sales Predictions. *Expert Systems with Applications*, 71, 416-428.

D'Haen, J., & Van den Poel, D. (2013). Model-Supported Business-to-Business Prospect Prediction Based on an Iterative Customer Acquisition Framework. *Industrial Marketing Management*, *42*(4), 544-551.

Eitle, V., & Buxmann, P. (2019, January). Business Analytics for Sales Pipeline Management in the Software Industry: A Machine Learning Perspective. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Rezazadeh, A. (2020). A Generalized Flow for B2B Sales Predictive Modeling: An Azure Machine-Learning Approach. *Forecasting*, *2*(3), 267-283.

Rohaan, D., Topan, E., & Groothuis-Oudshoorn, C. G. M. (2022). Using Supervised Machine Learning for B2B Sales Forecasting: A Case Study of Spare Parts Sales Forecasting at an After-Sales Service Provider. *Expert Systems with Applications*, *188*, 115925.

Yan, J., Gong, M., Sun, C., Huang, J., & Chu, S. M. (2015a). Sales Pipeline Win Propensity Prediction: A Regression Approach. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)* (pp. 854-857). IEEE.

Yan, J., Zhang, C., Zha, H., Gong, M., Sun, C., Huang, J., ... & Yang, X. (2015b). On Machine Learning Towards Predictive Sales Pipeline Analytics. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 29, No. 1).

Zahid, E., Ong, Y. J., Megahed, A., & Nakamura, T. (2021, December). Predicting Loss Risks for B2B Tendering Processes. In *2021 IEEE International Conference on Big Data (Big Data)* (pp. 2076-2083). IEEE.

APPENDIX D

PROCEDURE TO DEVELOP PARTIAL DEPENDENCE PLOTS

Partial dependence plots (PDPs), first developed by Friedman (2001), are used to understand the average behavior of a black-box machine learning model. PDPs depict the relationship between one or two focal independent variable(s) and the predicted dependent variable marginalizing over the values of all other independent variables included in the model. I use the following four-step approach to develop one-way PDPs depicting the dependence of model predictions on the values of a single independent variable. The same approach can be extended to build a two-way PDP depicting the dependence of interaction between two focal independent variables on the predicted outcome variable.

Step 0: Estimating a link function ($\hat{\boldsymbol{F}}$): Relationship between Dependent and Independent Variables

I use the ensemble approach explained in paper to build a link function **($\hat{\boldsymbol{F}}$)** that represents the relationship between my dependent variable (ServiceCo's propensity to win the bid) and independent variables. The link function **($\hat{\boldsymbol{F}}$)** built using Ensemble Algorithm is a complex object that includes five different models and a random forest meta learner to combine the predictions from the five models. To predict the outcome **($\hat{\boldsymbol{y}}$)** corresponding to a given value of independent variables, the data is passed through each of the five models and the predictions from the five models are used as input variables into the random forest meta learner.

Step 1: Simulate Data ($\boldsymbol{X_S}$) to plot partial dependence on a specific variable

a. *Categorical variables:* I simulate data ($\mathbf{X_S}$) based on different levels of the focal categorical independent variable. I force all data instances in my original data to have the same level of the categorical variable and append data corresponding to different levels. In my context, I have data for 4,024 bids. To plot PDPs for the *Relationship Class* variable which has five levels, the simulated data ($\mathbf{X_S}$) will have 20,120 (5 x 4,024) data instances.

b. *Continuous variables:* I follow a similar approach as in the case of the categorical variable, but I identify the different levels of the variable based on the minimum value, maximum value, and the granularity of the PDP I want to plot. In my context, to plot PDPs for the *Revenue (in million dollars)* variable, I simulate 100 (granularity of the graph) values between x (minimum value) and y (maximum value) and use them to create the simulated data ($\mathbf{X_S}$) that will have 402,400 (100 x 4,024) data instances.

Step 2: Predict $\widehat{Y_S}$ for simulated data (X_S) using the link function ($\hat{F}$)

Using the link function $(\boldsymbol{\hat{F}})$ developed in Step 0, I predict the dependent variable - "ServiceCo's propensity to win the bid"- for each data instance in the simulated data ($\mathbf{X_S}$).

$$\boldsymbol{\widehat{Y_S}} = \boldsymbol{\hat{F}}(\boldsymbol{X_S})$$

Where $\mathbf{X_S}$ is the data simulated in Step 1 and $\boldsymbol{\hat{F}}$ is the link function developed in Step 0 using the ensemble algorithm.

Step 3: Plot $\widehat{Y_S}$ to depict partial dependence between the focal independent variable and the dependent variable

a. *Categorical variables:* To understand the partial dependence of the focal independent variable, I plot the mean of the predicted outcome variable ($\widehat{Y_S}$) across different levels of focal independent variable.

b. *Continuous variables:* To understand the partial dependence of the focal independent variable, I plot the mean of predicted outcome variable ($\widehat{Y_S}$) corresponding to the different values of the focal independent variable in the simulated data (**X**s).