To bid, or not to bid?

Designing a machine learning model to support bid/no-bid decision-making for large Dutch construction projects

Complex Systems Engineering and Management

T.H. van der Sluijs

Delft University of Technology

TU Delft

To bid, or not to bid?

Designing a machine learning model to support bid/no-bid decision-making for large Dutch construction projects

by

T.H. van der Sluijs

To obtain the degree of
Master of Science
in Complex Systems Engineering and Management
at the Faculty of Technology, Policy, and Management of Delft University of Technology

Graduation committee

First supervisor and chair:	Dr. N. Mouter
Second supervisor:	Dr. A.A. Ralcheva
Third supervisor (Count & Cooper):	Drs. D. van Uffelen

Cover: Van Brienenoordbrug by Rijkswaterstaat

TU Delft COUNT & COOPER

Preface

In front of you lies my master's thesis, *"To bid, or not to bid: Designing a machine learning model to support bid/no-bid decision-making for large Dutch construction projects"*. This thesis has been at the centre of my universe for the past few months. Now that it is over, I can look beyond it again. That may sound as if I did not enjoy the process, but I genuinely did. I know I work well when I can fully focus on one thing, and this thesis gave me exactly that. For many, a thesis marks the end of their student days. For me, it was no different. A lot has changed over the course of this thesis, and perhaps by the end of it, I was not entirely a student anymore: going to the C&C office five days a week, treating my thesis like a 9-5, living together with my girlfriend, drinking tea in the evening, going to bed early?! And honestly, I am loving it!

What I also enjoyed about this thesis is that it concludes my academic journey at TU Delft. During my time at TPM, I found it difficult to choose between the different tracks, because all of them were interesting. This thesis' subject is a combination of the final choices I made: the Data & Digitalisation track inspired the machine learning approach, and the tendering and infrastructure context came from the Transport, Infrastructure, and Logistics track. In addition, doing this thesis at C&C made the experience even better. Five days a week, at my personal flex desk, this thesis grew from research ideas to proposal, from dataset to final results, and from draft to this final version. And what really made this journey memorable was the support, interest, coffee breaks, lunches, walks, and Friday drinks here at Count & Cooper. So in that regard, I would like to thank my company supervisor Dirk van Uffelen for providing me with the opportunity to do a graduation internship at C&C, and for his guidance, advice, one-liners, and stories; it was a pleasure working with you.

Sincere gratitude also goes to my academic supervisors: Niek Mouter and Aleksandrina Ralcheva. The combination of sharp and critical questions, advice, and discussions made this thesis a lot more coherent, structured, and academically grounded. Your feedback and input have made a substantial impact on the quality and clarity of this thesis.

Finally, I would like to thank my family, friends, and old and new roommate(s) for their constant support, encouragement, enthusiasm, listening ear, and home-cooked meals. Your patience and absorbance of my monologues meant more to me than you know.

It has been an amazing journey.

T.H. van der Sluijs
Delft, March 2026

Summary

Large Dutch construction tenders require contractors to invest time, expertise, and resources to prepare final submissions. In recent years, project risks and tendering effort have increased. For contractors, the bid/no-bid decision is therefore a critical early decision point: it determines whether capacity is allocated to preparing a full tender submission, at a stage when important information is still uncertain. A wrong decision can have serious consequences: a wrong bid can lead to excessive tender preparation costs or an unprofitable project, while a wrong no-bid may mean missing opportunity for profit. Despite its importance, the bid/no-bid decision in practice is still largely driven by experience and intuition, and contractors typically lack systematic and predictive decision support for this high-stakes choice.

This thesis analyses to what extent an explainable machine learning approach can *predict* and *explain* bid/no-bid decisions for large Dutch construction projects using characteristics that are available early in the tendering process. The study is conducted as a case study at Count & Cooper (C&C), a Dutch project-management and tender advisory consultancy in the construction sector. C&C specialises in large, complex construction projects and is active across all fronts of a project, including tendering, project management, planning, stakeholder coordination, and risk management. This setting provides rare access to real-world tender decision data and enables an assessment of how predictive modelling could contribute to bid/no-bid discussions in practice. The central research question is: *To what extent can an explainable machine learning model predict and explain the bid/no-bid decision for a Dutch contractor?*

To answer this research question, a dataset was constructed from historic tender information from C&C. Each row ($N = 101$) represents one project, and each column represents an operationalised tender characteristic, including project size and duration, tender duration, tender-document quality, project type, contract form and conditions, procurement procedure, and the price-quality ratio in the award criteria. A codebook with explicit definitions and coding rules was used to translate tender document information into structured variables in a consistent way. Importantly, the dataset focuses on characteristics that are typically available early, before a contractor has invested in producing a full tender submission. During dataset construction, two limitations emerged that affected the modelling approach. First, several potentially influential organisational and strategic variables (e.g., strategic potential, current workload availability, and client capability) could not be reconstructed reliably from the available C&C data and were therefore excluded. Second, the dataset contains survivorship bias due to an August 2024 system migration: historic projects with a "bid" outcome were imported into the new tender management tool more consistently than "no-bid" projects. As a result, the bid/no-bid balance shifts over time, which makes a simple "train on older projects, test on newer projects" evaluation unreliable. The model was therefore evaluated using 5-fold cross-validation: the dataset is split into five parts, the model is trained on four parts, and tested on the remaining part. This process is repeated five times so that each part serves once as the test set.

Several machine learning models were developed and compared, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machines (SVM), and XGBoost. Models were evaluated using common classification metrics, including Accuracy, False Positives (incorrectly predicting a bid), and False Negatives (incorrectly predicting a no-bid). Besides predictive performance, this thesis also examines why models make certain predictions. Linear models are interpreted using their coefficients. For the tree-based models, feature-importance methods are used to show which variables most strongly influence the predictions. These explanations are used to understand model behaviour, not to claim a causal relation between the variables and the bid/no-bid decision.

The results show that machine learning can predict the bid/no-bid decisions with an accuracy of about 71-74%. Random Forest achieved the strongest overall predictive performance (Accuracy = 0.741, or 74%), closely followed by Logistic Regression (73%) and SVM (71%). Across models, variables such as tender duration, contract duration, project value, project type, and contract form repeatedly

emerge as important drivers. Therefore, the bid/no-bid decision can be predicted and explained to a moderate extent, which makes the model more useful as transparent screening support than as an automated decision-maker. In addition, better registration of internal decision factors and continued dataset expansion should further enhance the predictability of the model.

The thesis concludes that explainable machine learning can provide meaningful predictive and interpretable patterns for bid/no-bid decision-making within the case study context, but that the current results should be treated as a proof-of-concept rather than a deployable decision tool. The main limitations are data-related: the analysis is based on a single-company case study; the dataset is relatively small; several important internal decision factors are missing (e.g., strategic potential, current workload, and client capability); and survivorship bias limits the validity of time-based analysis.

For practice, the most appropriate role of the model is as an additional, transparent input alongside expert judgement, rather than a full replacement of the current decision makers. The thesis therefore recommends improving future data capture (particularly the currently missing internal drivers) and documentation of decision reasoning, and continuing to expand the dataset over time to enable time-based evaluation. In practice, this means capturing internal context at the decision moment (e.g., capacity constraints, strategic fit, and key reasons for bid/no-bid), ideally in a consistent format.

Finally, future research should explicitly consider the costs of different types of mistakes. Predicting *bid* when the correct outcome is *no-bid* (False Positive) can lead to excessive tender-preparation costs and potentially unsuitable projects, while predicting *no-bid* when the correct outcome is *bid* (False Negative) can mean missing an opportunity for profit or strategic positioning. Studying the cost trade-off between false positives and false negatives could help optimise decision thresholds and class weights so that models are evaluated in terms of decision quality rather than accuracy alone.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
(F)AHP	(Fuzzy) Analytic Hierarchy Process
(A)NN	(Artificial) Neural Network
AI	Artificial Intelligence
ARW	Works Procurement Regulations
AUC	Area Under the Curve
CA	Contracting Authority
C&C	Count & Cooper
CD	Competitive Dialogue procedure
CoSEM	Complex Systems Engineering and Management
CPV	Common Procurement Vocabulary
CPwN	Competitive Procedure with Negotiation
CV	Cross-Validation
D&C	Design and Construct
DBFM(O)	Design, Build, Finance, Maintain, (and Operate)
DT	Decision Tree
EMVI	Economisch Meest Voordelige Inschrijving (MEAT/EMAT)
ESPD	European Single Procurement Document
EU	European Union
F1	F1-score (harmonic mean of precision and recall)
FN	False Negative
FP	False Positive
KNN	K-nearest neighbours
MEAT/EMAT	Most Economically Advantageous Tender
ML	Machine Learning
NOS	Dutch Broadcasting Foundation
PIANOo	Dutch Public Procurement Expertise Centre
RBF	Radial Basis Function
RF	Random Forest
RVO	Rijksdienst voor Ondernemend Nederland
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine
SQ	Sub-question
TN	True Negative
TP	True Positive
UAV-GC	Uniform Administrative Conditions for Integrated Contracts
UAV/RAW	Uniform Administrative Conditions
XGBoost	eXtreme Gradient Boosting

1

Introduction

1.1. Problem context

Tendering ensures transparency, competition, and efficiency in public spending, but it also introduces high stakes for contractors (PIANOo, n.d.-a). Preparing a competitive tender submission is a costly and time-intensive process, often requiring several months of dedicated work from multidisciplinary teams (Brainial, 2025). These preparation costs are not always compensated, making this bid/no-bid decision to participate in a tender far from trivial.

In addition to the financial burden, tender participation exposes contractors to strategic and operational risks. Large-scale construction projects often require advanced engineering solutions, tight scheduling, and careful coordination with ongoing traffic flows (Brainial, 2025). Contractors must not only account for these technical challenges, but also manage long-term contractual commitments, potential penalties for underperformance, and reputational risks in the event of delays or cost overruns (Lesniak *et al.*, 2018; Sonmez & Sözgen, 2017). As a result, participation in such tenders represents a high-stakes strategic gamble, particularly for complex projects. If the tender is won with an optimistic bid, the risk of cost overruns is significant. This *winner's curse* makes contractors more hesitant to tender (Thaler, 1991).

This bid/no-bid decision is becoming more important as project complexity increases. Infrastructure projects must increasingly comply with stricter environmental requirements, higher traffic volumes, rapid developments in digitalisation and cybersecurity, the need for climate resilience, and ambitions related to standardisation and sustainability (Knoope *et al.*, 2022; National Coordinator for Counterterrorism and Security, 2024). For contractors, such developments translate into more demanding tender requirements, higher upfront effort, and greater uncertainty at the early moment the bid/no-bid choice must be made.

These effects can be seen in the selection of projects contractors are willing to pursue. For example, BAM (one of the Netherlands' largest construction companies) has announced that it will no longer bid on projects larger than 150 million euro (Trouw, 2024). Similarly, the first procurement round for the renovation of the Van Brienenoord Bridge attracted only a single proposal, which was ultimately rejected because the contractor requested an "excessive" level of financial compensation (NOS, 2025). Consequently, the renovation has been delayed, with only minor maintenance activities being carried out in the meantime (Rijkswaterstaat, 2025). Together, these cases illustrate that the procurement of large projects is increasingly shaped by a contractor's decision to bid or not to bid.

For this reason, the bid/no-bid decision has become one of the most critical strategic choices for contractors. It serves as a gatekeeping mechanism that determines whether resources should be allocated to a tender process. Currently, this decision is based on expertise and experience of management (Ketaren & Sianturi, 2017; Lowe & Parvar, 2004).

Public construction projects are typically procured through a competitive tendering process in which the client publishes a call for tender, contractors decide whether to participate, submit a proposal if they

do, and the client evaluates bids and awards the contract (Rijksdienst voor Ondernemend Nederland (RVO), 2024). For contractors, the bid/no-bid decision is the internal go/no-go to commit to preparing a full tender submission. A visualisation of the simplified tendering process and the early in time bid/no-bid decision is shown in Figure 1.1.

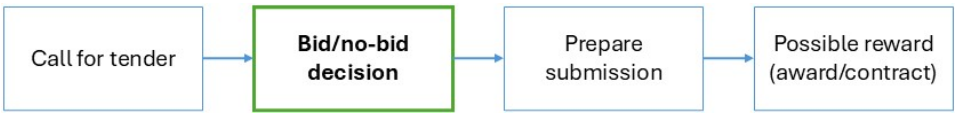


Figure 1.1: Simplified tendering process.

1.2. Previous studies

Given the high effort and uncertainty surrounding tender participation, research has increasingly examined how contractors make bid/no-bid decisions and whether this choice can be supported by formal decision frameworks and predictive models.

Across the bid/no-bid literature, studies have used surveys, questionnaires, and interviews to identify and rank factors that influence the bid/no-bid decision (Aje *et al.*, 2017; Egemen & Mohamed, 2007; Mahamid, 2022; Shokri-Ghasabeh & Chileshe, 2016; Wanous *et al.*, 1998). Some studies go further by trying to predict this bid/no-bid decision using quantitative methods, including machine learning (Cheng *et al.*, 2011; Gunduz & Lutfi, 2021; Ketaren & Sianturi, 2017; Lowe & Parvar, 2004; Sonmez & Sözgen, 2017). In this thesis, machine learning (ML) refers to training supervised learning models on historical tender cases to estimate the probability of a bid/no-bid outcome using information available at the earliest stage of a tender. The aim is not only prediction, but also explanation: interpreting which characteristics drive model outputs so the results can inform decision-making.

However, a recurring limitation across many bid/no-bid models is a form of circular reasoning. In numerous cases, the variables used to train models are derived from surveys of managers' subjective opinions (Jang *et al.*, 2015; Ketaren & Sianturi, 2017; Lesniak *et al.*, 2018; Lowe & Parvar, 2004; Sonmez & Sözgen, 2017). These opinions are then processed through factor analysis and statistical modelling, producing models that replicate managerial choices with accuracies of up to 90%. In addition, the ranking of the projects' features were also decided on by the same respondents. Such results only show that managerial judgement can be quantified and maybe even predicted, but they do not establish whether those judgements are correct or lead to better project outcomes.

In addition to these empirical contributions, systematic literature reviews have examined the use of machine learning in the construction sector (Prasetyo *et al.*, 2025; Rampini & Cecconi, 2022; Taboada *et al.*, 2023). Within these reviews, studies on ML applications for bid/no-bid decisions, bidding, or procurement were also considered. In one of these reviews, only a single ML study related to bid/no-bid decisions was identified, and this study was categorised under "training and skill development" rather than procurement (Prasetyo *et al.*, 2025). This suggests that the bid/no-bid decision has not yet been treated as its own application area. In addition, these reviews show that the number of relevant contributions to bid/no-bid is small, especially compared to risk assessment, and schedule, budget and deliverable management.

To conclude, there is still limited evidence on whether an explainable ML model trained on a contractor's actual tender history and its project, tender, contract, and client/strategic characteristics can both predict and explain bid/no-bid decisions in practice.

1.3. Problem definition and relevance

The bid/no-bid decision is a crucial step in the procurement process, as it determines whether contractors allocate resources to pursue a tender. In practice, this decision is still largely based on experience and intuition (Ketaren & Sianturi, 2017; Lowe & Parvar, 2004), while existing studies remain descriptive, highly context-specific (Aje *et al.*, 2017; Egemen & Mohamed, 2007; Mahamid, 2022; Shokri-Ghasabeh & Chileshe, 2016; Wanous *et al.*, 1998), or primarily validate managerial opinion (Jang *et al.*, 2015; Ketaren & Sianturi, 2017; Lesniak *et al.*, 2018; Lowe & Parvar, 2004; Sonmez & Sözgen, 2017). As

a result, contractors lack systematic and predictive decision support for this high-stakes choice, particularly in large construction projects where risks and sunk tendering costs are substantial. Building on the gaps identified in Section 1.2, this study therefore examines whether an explainable machine learning model, trained on a contractor's own tender history, can help to predict and explain bid/no-bid decisions.

In this thesis, machine learning is understood as a set of data-driven methods that learn patterns from historical observations in order to make predictions on new cases (El Naqa & Murphy, 2015). It is relevant here because the bid/no-bid decision depends on multiple project, contract, tender, and strategic characteristics, while the relationships between these characteristics and the final decision are not known beforehand (Lowe & Parvar, 2004; Sonmez & Sözgen, 2017). The remainder of this thesis therefore uses machine learning as an analytical tool to examine whether such patterns can be identified in historical tender data, to assess how well bid/no-bid outcomes can be predicted, and to explore which factors appear to be most influential in those predictions.

This study contributes to the scientific domain in two ways: (1) by moving beyond survey-based factor analysis and stated preferences (Aje *et al.*, 2017; Egemen & Mohamed, 2007; Mahamid, 2022; Shokri-Ghasabeh & Chileshe, 2016; Wanous *et al.*, 1998) to incorporate observed project and organisational characteristics into predictive models, and (2) by exploring model interpretability to identify which factors are most influential (Carvalho *et al.*, 2019; Louhichi *et al.*, 2023). In doing so, it also responds to a recurring limitation in the bid/no-bid modelling literature: models often reproduce subjective inputs from the same decision-makers they aim to predict, which risks circular reasoning and limits decision-support validity (Jang *et al.*, 2015; Ketaren & Sianturi, 2017; Lesniak *et al.*, 2018; Lowe & Parvar, 2004; Sonmez & Sözgen, 2017).

This study also addresses a contractor-side decision problem in the Dutch construction sector. Dutch contractors are becoming increasingly hesitant to participate in complex tenders, partly because unsuccessful bids involve substantial sunk tendering cost (NOS, 2025; Trouw, 2024). In this context, a systematic and explainable machine learning modelling approach may help contractors make more well-founded, consistent, and transparent early screening decisions.

By developing and validating an explainable machine learning approach based on real project and organisational data from a single Dutch contractor, the study is both scientifically and societally relevant. This connects directly to the MSc CoSEM framework, which equips students to address complex socio-technical systems where technical, institutional, and behavioural factors intersect. The bid/no-bid decision reflects such a system: a decision where technical project characteristics, formal institutional procedures, and informal dynamics between clients, contractors, and other stakeholders all play a role. Taken together, this positions the study at the intersection of scientific and societal relevance, and provides the basis for the research objectives and questions in the next section.

1.4. Research objective and questions

The aim of this study is to explore how explainable machine learning (ML) can be employed to predict and explain a Dutch contractor's bid/no-bid decisions in large construction projects. The research objective is to develop and test explainable supervised learning models on the contractor's observed tender history, using information available at the earliest stage of the tendering process to provide systematic, data-driven and interpretable insights for decision-making. Machine learning is appropriate because bid/no-bid decisions are influenced by multiple project and tender characteristics whose relationship with the outcome is not known beforehand (Lowe & Parvar, 2004; Sonmez & Sözgen, 2017); by comparing logistic regression as an interpretable baseline with more flexible models, this study examines both whether greater model complexity improves predictive performance and whether the main decision drivers remain consistent across models. To address this objective, the study is guided by the following main research question:

> ***To what extent can an explainable machine learning model predict and explain the bid/no-bid decision for a Dutch contractor?***

To answer the main research question, the study proceeds in four steps. First, it examines how the tendering process for large Dutch construction projects unfolds and which decision points are most

critical for contractors. Second, it identifies the factors that influence contractors' bid/no-bid decisions and reviews how these have been addressed in literature and practice. Third, it develops and compares multiple machine learning models to assess how well bid/no-bid decisions can be predicted from early-available tender characteristics. Fourth, it interprets the resulting models to determine which variables are most influential in shaping their predictions. These four steps are reflected in the following sub-questions:

SQ1: How does the tendering process for large Dutch construction projects unfold, and which decision points are most critical for contractors?

SQ2: Which factors influence the bid/no-bid decision of contractors, and how have these been addressed in existing literature and practice?

SQ3: How does the predictive performance of different machine learning algorithms (such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machine, and XGBoost) compare when modelling the contractor's bid/no-bid decisions?

SQ4: Which variables emerge as most influential in the predictive models, and how can these be explained and interpreted?

Count & Cooper (C&C) is used as the empirical case for this thesis. C&C is a Dutch project-management and tender advisory consultancy in the construction sector. The firm specialises in large, complex construction projects and is active across all fronts: tendering, project management, planning, stakeholder coordination, and risk management.

The study draws on C&C's internal historic tender data to obtain the bid/no-bid outcomes of their selected projects. This case setting provides rare access to real-world decision data. In addition, the C&C experts provide additional support, knowledge, practical perspective, and validation.

This thesis aims to deliver both predictive and explanatory insight into a contractor's bid/no-bid decisions. Rather than automating the decision itself, the goal is to provide transparent decision support that can help contractors make more systematic, and better grounded screening choices. By combining predictive performance with model interpretation, the study evaluates not only whether bid/no-bid behaviour can be predicted, but also whether the main decision drivers can be explained.

1.5. Thesis outline

The remainder of this thesis consists of five chapters, followed by the references and appendices. Chapter 2 first establishes the context in which the bid/no-bid decision is made and positions this decision within that broader context. Appendix A provides additional supporting information on the tendering context and the contractor's decision within it. Chapter 3 then identifies the factors that may influence the bid/no-bid decision. By reviewing prior literature and incorporating expert validation, it derives and defines the set of factors that form the basis for the modelling approach.

Building on these first two chapters, Chapter 4 contains the methodology. It explains which machine learning models are selected, how the variables are operationalised, how the modelling dataset is constructed, the model validation strategy, and how the performance and explainability are evaluated. Appendix B provides additional background on the selected model families, Appendix C presents the full operationalisation codebook, and Appendix D reports additional analyses related to model validation.

Chapter 5 presents the empirical results of the modelling approach. It first compares the predictive performance of the selected models and then interprets which variables emerge as most influential in their predictions. In this way, the chapter addresses both the predictive and explanatory parts of the research objective. Appendix F contains supplementary results from the comparative analysis.

Finally, Chapter 6 synthesises the main findings and reflects on their meaning for both research and practice. It answers the main research question, discusses the study's main limitations, provides recommendations for Count & Cooper, and discusses future research possibilities. Appendix G contains an additional analysis related to the practical application of the model.

2

Context exploration

In this chapter the first sub-question is discussed: *How does the tendering process for large Dutch construction projects unfold, and which decision points are most critical for contractors?* SQ1 provides the context in which the bid/no-bid decision is made, which is necessary to understand in order to identify and define the factors influencing this decision (as will be discussed in Chapter 3). To answer this question, first the formal procurement context is outlined (Section 2.1), and afterwards the contractor's decision points within that context are located (Section 2.2).

2.1. Formal structure of the procurement process

This section sets the formal structure for construction procurement in the Netherlands. It begins with the legal basis from EU directives into Dutch practice (§2.1.1), then introduces the four procedures and their flows (§2.1.2), explains the selection and award criteria (§2.1.3), and closes with the most common contract forms and their associated risk allocation (§2.1.4).

2.1.1. From EU to NL legal framework

The legal basis for the selection of public procurement procedures in the countries of the European Union (EU) is Directive 2014/24/EU of 26 February 2014 on public procurement (Directive 2014/24/EU, 2014). Public procurement as defined in the Directive 2014/24/EU is 'the acquisition by means of a public contract of works, supplies or services by one or more contracting authorities from economic operators chosen by those contracting authorities'. Public procurement should furthermore safeguard the principles of equal treatment, non-discrimination, proportionality and transparency. In the Netherlands, Directive 2004/18/EC (2004) (the predecessor of 2014/24/EU) was implemented in the Aanbestedingswet (*Stb*. 2012, 542) (this translates to Dutch Public Procurement Act), and was later updated to implement Directive 2014/24/EU. The Aanbestedingswet operationalises four core principles (PIANOo, n.d.-a):

- **Non-discriminatory:** no distinctions may be made on the basis of nationality.
- **Equal treatment:** all economic operators are treated equally and receive the same information.
- **Transparency:** the process is open and comprehensible; expectations are clear in advance and decisions are carefully explained.
- **Proportionality:** the requirements set are proportionate to the nature and scale of the contract.

According to Directive 2014/24/EU, when awarding public contracts above a certain value (called a threshold), the EU procurement obligation applies (Marinelli & Antoniou, 2019; Plebankiewicz, 2024). The threshold for 2026-2027 is EUR 5,404,000 (PIANOo, n.d.-c). In addition to the EU procurement obligations, the Netherlands has instruments that help with the implementation and approach of procurement, and with the legal framework. These are the Aanbestedingsreglement Werken 2016 (ARW 2016), Gids Proportionaliteit, and the UEA/ESPD (Uniform Europees Aanbestedingsdocument/European Single Procurement Document). For a compact explanation of these instruments, see Appendix A.1.

2.1.2. Four procedures and their flows

According to EU regulations, a public owner can use one of several procedures. Various procedures have different options for assessing the contractor and evaluating the tender, and each has both advantages and disadvantages. Knowing them allows one to choose the most effective procedure, tailored to the conditions of a given order (Directive 2004/18/EC, 2004; Directive 2014/24/EU, 2014; Plebankiewicz, 2024). There is no procedure strategy that is a perfect fit for all public work procurement (Schrijfgroep Gids Proportionaliteit, 2022).

Below, the four most used standard procedures for above-threshold public works are introduced: (1) Open, (2) Restricted, (3) Competitive Procedure with Negotiation (CPwN), and (4) Competitive Dialogue (CD). Each follows the same basic stages but structures selection and award differently, and offers a different level of interaction between client and bidders. A short description is provided for each procedure; for a more detailed description and a visualised comparison of each procedure, please see Appendix A.2.

Open procedure: Used when the works are well specified and of low to moderate complexity, and the authority wants broad competition without pre-shortlisting. It is a single-stage route: the contracting authority publishes a notice and makes the full tender documents available to all interested suppliers; bidders submit both selection (qualification) information and their tenders in one go. Selection checks and award evaluation are applied to identify the most suitable contractor (more on this in §2.1.3) (Marinelli & Antoniou, 2019; Plebankiewicz, 2024; SIGMA, 2016).

Restricted procedure: Chosen when capability should be vetted before pricing and the client wants to manage the number of full bids. It runs in two stages: after the notice, economic operators submit requests to participate (UEA/ESPD) for the selection stage; the authority may then draw up a shortlist and invite only those bidders to submit tenders with the full invitation-to-tender documents. This limits the number of tenders and reduces wasted effort (Marinelli & Antoniou, 2019; Plebankiewicz, 2024; SIGMA, 2016).

Competitive Procedure with Negotiation (CPwN): Suitable where project specifications are not yet set in stone, and/or the type of work is complex. After selection (requests to participate and shortlisting), the authority invites shortlisted bidders and issues initial tender documents; bidders submit initial tenders and then enter one or more negotiation rounds. When negotiations close, an invitation to submit final tenders is issued and final tenders are evaluated. This provides structured interaction, but requires careful governance and equal-treatment records (Plebankiewicz, 2024; SIGMA, 2016).

Competitive Dialogue (CD): Used for high-complexity or innovative projects where the solution cannot be fully defined upfront and outcomes must be shaped with market input. After selection and shortlisting, the authority issues descriptive/dialogue documents and conducts one or more dialogue rounds to develop the solution. Once the dialogue is formally closed and the requirements are fixed, the authority issues invitations for the final tenders (Lenferink *et al.*, 2013; Marinelli & Antoniou, 2019; Plebankiewicz, 2024; SIGMA, 2016).

In sum, the four procedures follow the same backbone but differ in how they handle entry to the competition, the degree of interaction, and when final tenders are requested. These design choices influence what and when bidders must submit information, and how authorities review tenders. The next subsection explains the selection and award logic that contracting authorities use.

2.1.3. Selection and award criteria

Contracting authorities use selection and award criteria to identify the most suitable contractor and to choose the winning tender in a way that is transparent, fair, proportionate and non-discriminate (Koppenjan *et al.*, 2024). These criteria are standardised within the EU/Dutch framework and apply across the four procedures introduced in §2.1.2. In short, selection concerns the bidder's capability to participate, whereas award concerns the comparative evaluation of compliant tenders (Schrijfgroep Gids Proportionaliteit, 2022). Below a short description is provided for each type of criteria; for a more in-depth description, please see Appendix A.3.

Selection criteria

Selection criteria consist of two main elements: exclusion grounds and suitability requirements. Exclusion grounds address issues such as legal or tax non-compliance and serious professional misconduct, and lead to a binary decision on admissibility, while suitability requirements concern legal standing, financial and economic capacity, and technical or professional capacity (PIANOo, n.d.-a; SIGMA, 2016). These requirements must be proportionate to the contract's size, risk profile, and allocation of responsibilities; for example, they may relate to turnover thresholds, references, certifications, or key personnel qualifications (Schrijfgroep Gids Proportionaliteit, 2022). In the open procedure, selection checks are performed after submission, whereas in the other three procedures they occur at the request-to-participate stage and may be followed by shortlisting, which limits the number of invited bidders (SIGMA, 2016).

Award criteria

Award criteria determine which compliant tender is the most economically advantageous. This is usually done under the MEAT framework (Most Economically Advantageous Tender, also called EMAT, and in Dutch *EMVI*), which combines qualitative criteria with price or cost using predefined weights and a stated calculation method (Fuentes-Bargues *et al.*, 2017; PIANOo, n.d.-a; Plebankiewicz, 2024; SIGMA, 2016). Typical qualitative elements include the plan of approach, risk management, planning/phasing, stakeholder and environmental management, and sustainability. For simple and well-specified works, lowest price may be used instead, although this can encourage artificially depressed prices and thereby lead to unforeseen costs (Bochenek, 2014; SIGMA, 2016).

2.1.4. Contract forms and risk allocation

Contract forms are instruments to allocate risk and responsibility across project phases. They are chosen to fit the level of uncertainty and the client's capacity to define, manage and assure the work (van Ham & Koppenjan, 2002; Verweij & van Meerkerk, 2021). In practice, the choice determines who is accountable for shaping and checking the design, coordinating interfaces between parts of the work, and meeting performance in use (e.g., requirements completeness, integration and commissioning, long-term reliability) (PIANOo, n.d.-b). Figure 2.1 shows responsibility shifting from client to contractor as one moves from traditional contract forms towards more integrated models (Source: adjusted from Count & Cooper). The darker shading indicates a larger share of responsibility for the contractor across earlier phases of design and, at the far end, into maintenance. The contracts listed in the figure are Cost-Plus, RAW/Bestek, Bouwteam, Design & Build (D&B), Turnkey, Design Build Finance Maintain (and Operate) (DBFM(O)). For a description of each contract, please see Appendix A.4. Note: these are not all the possible contract forms, but cover the most used in the Netherlands.

As can be seen in Figure 2.1: the more integrated the contract form (rows), the more responsibility transfers to the contractor (columns). Under traditional contracting models, the tender primarily demonstrates how the client's design will be executed in combination with an accurate view of quantities and productivity. Under D&B and Turnkey models, the tender must also show how the contractor will shape and validate their own design. Under long-term forms, the tender must show how the solution will remain reliable in use and how it will be maintained.

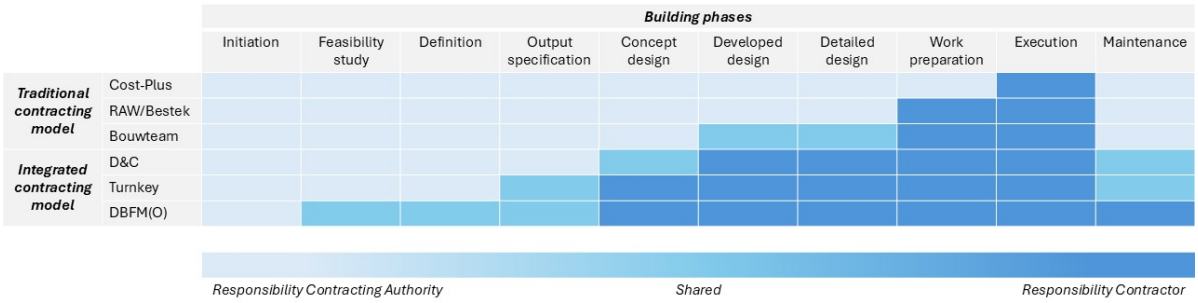


Figure 2.1: Distribution of responsibility per contract form (Source: adjusted from Count & Cooper).

In addition to these contract form, Dutch public works contracts typically refer to a standard set of con-

tract conditions that shapes the legal relationship between client and contractor (PIANOo, n.d.-f). This is commonly the UAV 2012 (Uniform Administrative Conditions), which is written for contracts where the client provides the design basis and the contractor is primarily responsible for work preparation and execution. For integrated contracts, clients can also use UAV-GC. UAV-GC 2025 is made for projects with a design, construct, and/or maintain component (CROW kennisplatform, n.d.). As a result, two projects with the same “contract form” label may still differ in risk allocation depending on whether UAV or UAV-GC applies. For this reason, this thesis treats contract form and contract conditions as separate constructs.

2.2. The bid/no-bid decision

With the procurement context established, this section turns to the bid/no-bid decision from the contractor perspective. In this thesis, the bid/no-bid decision is defined as the contractor’s internal go/no-go decision to commit time, money, and organisational capacity to the preparation of a full tender submission. This decision is critical because tender participation requires considerable investment, while important uncertainties about the project, client, risks, and competitive position may still remain (Gunduz & Lutfi, 2021; Lowe & Parvar, 2004). An incorrect decision can result either in sunk tendering costs and exposure to an unsuitable project, or in a missed opportunity for profit and strategic positioning. Its importance increases further for larger and more complex projects, where the expected bid effort and associated risks are higher.

To structure the contractor perspective in a transparent way, this thesis distinguishes six decision points in the tendering process. These decision points are developed through discussions with experts at Count & Cooper. While presented in a sequential order for clarity, the experts emphasised that decision points (2)-(4) are typically intertwined and may overlap in practice.

There are six identified decision points for contractors during the entire procurement process. These are listed below in chronological order. During the *pre-market phase*, a contractor can decide whether or not to participate in the market consultation; in this research this is called the (1) **early scan** decision. During the *selection phase*, the four main decisions must be made: the (2) **pre-qualification decision**, the (3) **teaming/consortium strategy**, the (4) **solution strategy**, and the (5) **bid economics** decision. Bid economics may also happen during the *interaction phase*, depending on the chosen procurement process. Lastly, a final decision must be made whether to submit the final tender or not: the (6) **final submit/no-submit** decision. Appendix A.5 explains these six decision points in more detail.

In this thesis, the bid/no-bid decision is positioned during pre-qualification (2) and teaming/consortium strategy (3). At this point, the contractor decides whether to proceed with a bid or to withdraw from the opportunity.

Early available information

The timing of the bid/no-bid decision has an important implication for this thesis: this decision is made under incomplete information. At the moment a contractor decides whether to proceed, not all relevant project details are known yet, and many strategic considerations have not yet been fully worked out. This timing is important for the remainder of the thesis. Since the objective is to analyse and model the bid/no-bid decision using historical tender data, the focus is placed on characteristics that are typically available at this early decision moment and that can be reconstructed consistently from documentary sources.

3

Related work

Building on the tender context in Chapter 2, this chapter synthesises bid/no-bid literature, desk research, and expert validation to address the second sub-question: *Which factors influence the bid/no-bid decision of contractors, and how have these been addressed in existing literature and practice?* SQ2 matters because before predicting (and explaining) the bid/no-bid decision, one must first identify the factors that influence such a decision, so that those factors can be defined and turned into measurable inputs for the model. Now that the context in which the bid/no-bid decision takes place is clear, the next step is to determine what influence it. First, a literature review is conducted to see how prior work identifies and uses factors (Section 3.1). Second, the factors are given a definition that will be used throughout the thesis (Section 3.2).

3.1. Bid/no-bid factors and modelling approaches in literature

This section reviews how earlier studies have treated the bid/no-bid problem, what factors they considered, which ML models they used, and their findings. For clarity purposes, the papers are grouped into two groups (identifying/ranking vs. modelling/frameworks). The goal is to identify what factors influence the bid/no-bid decision.

The review is conducted using the databases Scopus, Google Scholar, and the TU Delft repository, focussing on peer-reviewed literature in both English and Dutch. The search and selection process concentrates on three themes: (i) bid/no-bid decision-making in construction, (ii) procurement and contracting processes, and (iii) applications of machine learning in construction and procurement. Together, these domains cover both the institutional and decision-making context as well as the potential of data-driven models. Searches were performed using predefined keywords related to the these three topics and combinations of those keywords (Table 3.1). A total of eleven papers that research the bid/no-bid decision were included in this literature review. Since the research on this subject is limited, there was no geographical filter and older relevant sources were allowed. The papers range from the year 1998 to 2022. The synthesis of the literature review is illustrated in Table 3.2.

Table 3.1: Research topics and example keywords used for the literature search.

Research topic	Example keywords
(i) Bid/no-bid decision	*bid/no-bid decision*, *tender decision*, *contractor bidding*, *tender selection*
(ii) Procurement procedure	*public procurement*, *construction procurement*, *tendering process*, *procurement procedure*
(iii) ML application	*machine learning*, *classification*, *SHAP*, *feature importance*

A factor would be included if three criteria could be applied: (1) it was mentioned in at least three reviewed studies or explicitly emphasised by the case experts, (2) it is observable at the bid/no-bid

moment, and (3) it can be logically operationalised into a usable feature for machine learning. The second criteria is based on the projects' selection or award guidelines provided by clients. The third criteria is based on the expertise of the researcher. This synthesis resulted in the 14 factors listed in Table 3.2, which are defined in Section 3.2, and the operationalisation into measurable variables is described in Chapter 4.

Note that across both groups, the reviewed studies use different factor labels and levels of detail, meaning that reported features do not always map one-to-one onto the 14-factor set in Table 3.2. To enable synthesis, this thesis applies a closest-match interpretation: reported variables are linked to the most similar factor, or treated as a proxy when only partial overlap exists. For example, terms such as "scope fit", "type of works", or "experience with similar projects" are mapped to *Project type*, while features such as "competition intensity" or "market share" are mapped to *Potential for new projects*.

Table 3.2: Synthesis of the literature review on the bid/no-bid decision research.

	Group one					*Group two*					
Factor	**(A)**	**(B)**	**(C)**	**(D)**	**(E)**	**(F)**	**(G)**	**(H)**	**(I)**	**(J)**	**(K)**
Project size	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*
Project nature/type	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*
Project planning	*x*		*x*		*x*	*x*				*x*	*x*
tender duration	*x*	*x*		*x*	*x*	*x*	*x*			*x*	
Tender documents quality	*x*	*x*	*x*	*x*	*x*			*x*			*x*
Contract type/collaboration form		*x*	*x*			*x*	*x*				
Contract conditions		*x*			*x*	*x*	*x*			*x*	*x*
Contract payment terms	*x*	*x*	*x*		*x*	*x*	*x*				
Procurement procedure			*x*			*x*					*x*
Selection criteria	*x*	*x*		*x*	*x*		*x*				*x*
Award criteria										*x*	*x*
Potential for new projects		*x*						*x*	*x*		
Current workload	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*
Client capability	*x*	*x*	*x*		*x*	*x*	*x*			*x*	*x*

(A): Wanous *et al.* (1998), (B): Egemen and Mohamed (2007), (C): Shokri-Ghasabeh and Chileshe (2016),
(D): Aje *et al.* (2017), (E): Mahamid (2022), (F): Lowe and Parvar (2004), (G): Cheng *et al.* (2011),
(H): Ketaren and Sianturi (2017), (I): Sonmez and Sözgen (2017), (J): Lesniak *et al.* (2018), (K): Gunduz and Lutfi (2021).

The reviewed papers can be categorised into two groups: (1) the identifying and ranking research, and (2) the decision frameworks or modelling approaches group. The first group (A, B, C, D, and E) all used surveys to capture the factors influencing the bid/no-bid. The second group (F, G, H, I, J, and K) used multiple different approaches such as logistic regression, (F)AHP, or machine learning models.

In the first group, the studies consistently treat the bid/no-bid decision as a multi-criteria decision and aim to prioritise the underlying considerations. The studies in Group one were conducted in different contexts and economically different countries, but they report recurring characteristics that have significant impact on the bid/no-bid decision. Project-related characteristics (e.g., project size, project type), contract-related characteristics (e.g., payment terms), tender-related characteristics (e.g., selection criteria), and client-related characteristics (e.g., financial stability, promptness) were among the highest scoring factors.

In the second group, studies move from factor ranking towards decision-support and predictive modelling. Across these studies, the bid/no-bid decision is modelled using approaches such as logistic regression, SVM/NN classifiers, fuzzy multi-criteria methods, and tree-based models. The factors reported as influential can be clustered into familiar characteristics: project (e.g., scope fit, price level), contract (e.g., contract conditions, payment terms), and client/strategic (e.g., personal relationship with client, political position).

Table 3.2 summarises which factors are actually in play across the reviewed work. Project size, project type, and current workload are discussed in every paper. Client capability/reputation and contract/-

payment conditions are widely referenced, and tender documents quality and tender duration occur in multiple sources. In contrast, procurement procedure, potential for new projects, and award criteria show low coverage. However, expert validation stated that these are definitely taken into account. In the next section, the factors will be given their definition that will be used in the rest of the thesis.

3.1.1. Expert validation of factor relevance

Expert input from Count & Cooper is used as a validity check on whether the factor set derived from the literature and document analysis matches how bid/no-bid decisions are discussed in practice.

Within C&C's way of work, experts indicated that the factors included in this thesis align well with the elements considered in their internal process. Experts stated that factors (such as tender documents quality, separating contract form from contract conditions, payment terms, potential for follow-on projects, current workload, and client capability) are also discussed during meetings, but not explicitly documented. At the same time, the experts emphasised that their decision includes an important qualitative component related to the overall "attractiveness" of a project. As one interviewee stated, "C&C does not do boring projects". While this project attractiveness is clearly relevant in practice, it is not easily observable or definable from tender documents. As a result, it cannot be operationalised reliably within the this thesis and is therefore treated as out-of-scope.

3.2. Defining the factors

This section defines the factors identified in the previous section that will be used throughout the remainder of this thesis. To provide structure, the factors are grouped into four clusters: (1) project characteristics, (2) contract characteristics, (3) tender characteristics, and (4) client/strategic characteristics. These definitions clarify the conceptual meaning of each factor in the context of the contractor's bid/no-bid decision for large and complex construction projects, and if they are available/reconstructable at the decision moment. Their operationalisation into measurable variables, including coding schemes and value buckets, follows in Chapter 4.

Project characteristics

The first cluster consists of *Project size*, *Project nature/type*, and *Project planning*. Project size captures the overall scale of the works relevant to capacity, area, volume, and number of sites; and is defined primarily as the estimated contract value of the package to be bid. This factor is recognised across all reviewed studies and was identified by Egemen and Mohamed (2007) as the most significant in the decision process. In practice and literature, project size is often expressed in value buckets or categories such as small, medium, and large. For this thesis, it is important to note that the case study contractor Count & Cooper is specialised in large and complex projects, including megaprojects above €500 million. Because only 3% of procured projects in the Netherlands exceeds €10 million (Hardeman, 2012), the eventual value ranges used in this study must be shifted to better match this market segment.

Project nature, or project type, represents the sector and type of works to be delivered. In this thesis, it is defined as the sector classification of the project, such as infrastructure, buildings, rail, energy, or marine works. This factor is also mentioned in all reviewed studies and scores high in stated-preference research (Lesniak *et al.*, 2018; Lowe & Parvar, 2004). Closely related to this is project planning, which captures the execution timeline and planning constraints of the project. It is defined by the required start and finish window. Egemen and Mohamed (2007) and Aje *et al.* (2017) similarly refer to whether the allowed project duration is sufficient or the proposed timescale is realistic.

At the bid/no-bid moment, information on project size, project type, and the broad project planning is typically already observable in the tender notice, tender documents, or estimated by experts. This makes these factors both practically relevant and operationalisable for retrospective analysis.

Contract characteristics

The second cluster consists of *Contract type/collaboration form*, *Contract conditions*, and *Contract payment terms*. Contract type/collaboration form describes how scope and responsibilities are allocated between parties (see §2.1.4) and at what stage the contractor becomes involved. The emphasis of this

factor is on who does what and when across the project lifecycle, and on the degree of collaboration implied by the chosen form.

Contract conditions capture the legal and technical framework governing the project. In Dutch and international construction practice, these are often based on standardised conditions such as UAV 2005 or UAV-GC 2025, usually supplemented with project-specific clauses. This factor intentionally excludes the pricing and payment mechanism, which is treated separately. Contract conditions are particularly relevant because they shape the allocation of design, execution, and maintenance responsibilities, as well as broader project risks. Two of the reviewed studies identified contract conditions as among the most influential factors in the bid/no-bid decision (Lesniak *et al.*, 2018; Shokri-Ghasabeh & Chileshe, 2016).

Contract payment terms capture how the contractor is compensated and how cost risk is shared between the contracting parties. In this thesis, payment terms refer specifically to the nature and timing of payments and the compensation basis during project execution, rather than to broader legal clauses or delivery structures. Egemen and Mohamed (2007) and Mahamid (2022) both conclude that payment arrangements play a significant role in the decision whether or not to prepare a tender.

These contract characteristics are not equally observable at the bid/no-bid moment. Contract type/collaboration form and contract conditions are typically already stated in the tender notice, tender guideline, or draft contract; which makes them suitable for operationalisation. Contract payment terms are more difficult to capture. As experts mention: for large and complex projects it is often not yet fixed at the bid/no-bid stage and may remain subject to negotiation after contract award. As a result, payment terms are expected to be less observable and reconstructable than the other contract characteristics.

Tender characteristics

The third cluster consists of *tender duration*, *Tender documents quality*, *Procurement procedure*, and *Selection* and *award criteria*. tender duration is the time available to prepare and submit a bid. After deciding to bid, the contractor must prepare planning documents, pricing, method statements, and other submission materials to convince the client of its capability to execute the project successfully. As noted in Chapter 1, tendering is a costly and time-intensive process. The available tender period therefore directly affects the feasibility and attractiveness of participation, and may vary depending on the procurement procedure, contract form, and award set-up.

Tender documents quality refers to the quality, completeness, clarity, and accuracy of the tender documents issued by the contracting authority. Across the literature, this factor appears under different labels, including completeness of bid documents, rigidity of specifications, clarity of work, clarity of documents, and accuracy of contract documents (Aje *et al.*, 2017; Egemen & Mohamed, 2007; Shokri-Ghasabeh & Chileshe, 2016). In this thesis, these dimensions are combined under the broader label of tender documents quality. This factor matters because better-quality documents enable more accurate estimates of cost, time, and resource requirements, thereby reducing bid risk. Aje *et al.* (2017) even found this factor to be statistically significant in the bid/no-bid decision.

Procurement procedure, in this thesis, refers to the four EU-standard procedures for above-threshold public works: Open, Restricted, Competitive Procedure with Negotiation (CPwN), and Competitive Dialogue (CD). Selection criteria and award criteria are also part of this tender-related cluster. Selection criteria are used to determine whether firms are admissible and suitable to participate, for example on the basis of legal compliance, financial capacity, or technical experience. Wanous *et al.* (1998) found selection criteria to rank among the highest in stated-preference research. Award criteria determine which tender is considered the most economically advantageous, either through lowest price or through a MEAT/EMVI approach in which qualitative criteria are combined with price or cost using published weights and a stated scoring method (PIANOo, n.d.-a; SIGMA, 2016).

These tender characteristics are also suitable for inclusion in this thesis, because they are generally observable at the bid/no-bid moment. The tender notice and documents almost always specify the tender duration, procurement procedure, selection and award criteria.

Client/strategic characteristics

The fourth and final cluster consists of *Potential for new projects*, *Current workload*, and *Client capability*. Potential for new projects captures the strategic value of participating in a tender beyond the immediate

contract. In this thesis, it is defined as the anticipated potential to create a strong reference project, strengthen credentials in a specific sector, build a relationship with a client, or improve the chance of future invitations and wins. Interviewees indicated that firms may sometimes decide to tender not only because of the project itself, but also because it helps them enter a new sector or strengthen their market position.

Current workload captures the contractor's existing commitments and capacity utilisation at the time of tender and across the expected execution window. It reflects the degree to which resources are already allocated to ongoing projects, and whether sufficient capacity remains to prepare a competitive bid and deliver the project if awarded. This factor is widely recognised in the literature as an important determinant of the bid/no-bid decision, especially because tendering itself requires substantial time and effort.

Client capability captures the reliability and competence of the client to fund, manage, and govern the contract. It can be understood in terms of financial stability, payment behaviour, decision-making quality, and experience with similar projects or delivery models. This factor is important because it influences payment risk, change risk, and coordination risk during project execution. Mahamid (2022) likewise found the client's financial stability to be one of the highest-ranked factors in their survey results.

These client and strategic characteristics are also relevant because they are typically observable at the bid/no-bid moment. Potential for new projects, current workload, and client capability are usually assessed by the contractor based on internal knowledge, existing client relationships, and prior project experience.

This chapter has synthesised the literature to identify and define the factors that influence the bid/no-bid decision. It has grouped prior work into identifying/ranking versus modelling algorithms. The result is a set of factors that can be further operationalised so that they become suitable for analysis (SQ2). In Chapter 4 these factors are operationalised into measurable variables (level of measurement, coding schemes, value buckets).

4

Methodology

This chapter explains how the identified bid/no-bid factors are translated into a structured case study dataset and how the machine learning analysis is designed. To predict and explain the bid/no-bid decision, these factors must be operationalised into measurable variables, combined into a modelling dataset, and analysed using suitable machine-learning models. The chapter therefore discusses the case study setting and data sources, the selection of suitable machine-learning models, the operationalisation of variables and dataset construction, the validation strategy, and the performance evaluation and explainability methods.

This thesis adopts a quantitative research approach. The objective is to assess to what extent project characteristics can predict and explain the bid/no-bid outcome. The quantitative approach is appropriate when the research relies primarily on numerical data and on the evaluation of patterns and model performance (Creswell, 2009).

The design consists of four components. First, the case study setting and data sources are introduced, and the model requirements and selection are discussed (Section 4.1). Second, the factors identified in the previous chapters are operationalised into measurable variables and combined into a modelling dataset (Section 4.2). Third, the selected models are developed and validated using a consistent modelling strategy (Section 4.3). Finally, predictive performance is evaluated using classification metrics appropriate for the bid/no-bid context, and explainability methods are applied to interpret the contribution of individual variables and to assess the extent to which the resulting models can support decision-making in practice (Section 4.4).

4.1. Case study setting, data sources, and model selection

This section explains the case study setting, the data sources, and how the machine-learning models were selected for the comparative analysis. Because the aim of this study is not only to predict the bid/no-bid outcome, but also to explain which project characteristics contribute to that prediction, the selected models had to satisfy both predictive and interpretability-related requirements. In addition, they needed to be suitable for a relatively small, structured dataset consisting of numerical, ordinal, and categorical variables.

This thesis compares Logistic Regression as an interpretable baseline with several more machine learning models. This comparison is relevant because the bid/no-bid decision may depend on multiple interacting project characteristics, while the exact relationships between these characteristics and the final decision are not known *a priori* (Lowe & Parvar, 2004; Sonmez & Sözgen, 2017). More flexible models may therefore capture patterns that a simpler baseline cannot, while explainability methods can help assess which variables drive the resulting predictions (El Naqa & Murphy, 2015). In this way, the study examines whether machine learning provides added value for predicting and explaining the bid/no-bid decision in practice.

Case study setting and data sources

This study is conducted with a case study at Count & Cooper (C&C), a Dutch project-management consultancy in the construction sector. The firm specialises in large, complex construction projects and is active across multiple project phases, including tendering, project management, planning, stakeholder coordination, and risk management. This case study provides access to real-world historic tender data and expert knowledge on how bid/no-bid decisions are made in practice.

The main data source is C&C's internal historic tender administration, exported from Brainial, a third-party tender-management system used within the company. These records were used to reconstruct the bid/no-bid outcome and to identify core project, tender, and contract characteristics for each observation. To supplement and verify this information, tender documents such as contract notices, guidelines, draft contracts, and award-related documents were consulted where available. In addition, informal discussions and validation sessions with C&C experts were used to clarify ambiguous cases, confirm coding choices, and assess whether the reconstructed variables were consistent with practice.

Model requirements

Four requirements guide the model selection. First, the models have to be capable of binary classification, since the target variable is whether a contractor decides to bid or not to bid. Second, the models have to offer at least some degree of explainability, because the study also aims to identify which variables contribute most strongly to the predictions. Third, since the relation between the factors and the bid/no-bid decision is not known, nonlinearity and interaction have to be explored. Lastly, since the list of features is long, relative to the number of assumed datapoints, overfitting risk is also taken into account.

Selected models

Based on these requirements, five supervised classification models were selected: Logistic Regression, Decision Tree, Random Forest, XGBoost, and Support Vector Machine. Logistic Regression was included as a baseline, because it is widely used for binary classification and provides a transparent reference model through its coefficients (Park, 2013; Singh *et al.*, 2016). Support Vector Machine was included because prior bid/no-bid research reports good performance for SVM-based approaches, and because SVMs can be effective when the number of features is relatively high compared to the number of observations (Han & Jiang, 2014; Singh *et al.*, 2016; Sonmez & Sözgen, 2017). SVMs behave differently depending on their chosen kernels; in this thesis, a linear and a non-linear kernel have been chosen (linear and rbf, respectively) (Navia-Vázquez & Parrado-Hernández, 2006). Finally, tree-based models were included because they represent a different family of learning mechanisms: a single Decision Tree provides an interpretable nonlinear benchmark, while Random Forest and XGBoost allow the comparison to include more flexible ensemble-based tree methods that may better capture interactions and more complex patterns in structured tender data (Carvalho *et al.*, 2019; Li, 2022; Murel & Kavlakoglu, n.d.). Together, these five models provide a balanced comparative set for an exploratory analysis of the bid/no-bid decision. A full explanation of the models principles, and their main differences across the requirements can be found in Appendix B.1 and Appendix B.2, respectively.

4.2. Variable operationalisation and dataset construction

This section explains how the factors identified in Chapters 2 and 3 are translated into a structured modelling dataset. The objective is to operationalise those characteristics that could be observed consistently and that are typically available early in the tendering process. To support this process, a codebook is developed in which each factor is assigned a definition, coding rule, and data source. This codebook serves as the basis for translating tender information into a quantitative dataset suitable for machine-learning analysis. The full codebook, including the operationalisation of each variable, is provided in Appendix C.1.

Researcher judgement and expert validation

Although the codebook provides a structured basis for operationalisation, some factors requires subjective judgement because the relevant information was qualitative, only partially documented, or dependent on the specific tender context. In addition, certain variables that are relevant in practice could not be reconstructed reliably from the available historic records, because they depend on internal and

time-specific company information. For these reasons, the operationalisation process required both researcher judgement and expert validation.

In particular, ordinal variables such as project size, tender-document quality, and selection criteria had to be translated into practical categories that were meaningful in the Count & Cooper context. Expert validation was therefore used to assess whether these encodings reflected realistic distinctions between projects, whether the required information is typically available at the bid/no-bid decision moment, and whether observed patterns of missingness were plausible in practice. This is especially relevant for variables that literature identifies as important, but that are often not explicitly documented in historic tender files. The detailed operationalisation rationale and expert validation outcomes are described in Appendix C.1 and Appendix C.1.1, respectively.

Initial dataset and diagnostic assessment

After the codebook has been applied, an initial dataset was constructed in which each observation represented one tender opportunity for which the bid/no-bid outcome was known. The initial dataset contained $N = 101$ observations and 16 variables. The target variable was relatively balanced, with 48 bid cases and 53 no-bid cases. A diagnostic assessment was then conducted to obtain an overview of the dataset composition and to identify potential data-quality issues before modelling. The full dataset overview and diagnostic results are provided in Appendix C.2.1. The assessment of the data quality can be found in Appendix C.2.2. Lastly, the crosstabs between each variable and the bid/no-bid outcome is provided in Appendix C.2.3.

Data cleaning and preprocessing

The next step is to assess the quality of the initial dataset in more detail. This assessment focuses on missing values, “unknown” categories, and skewed distributions. Based on this assessment, the dataset is cleaned and preprocessed to make it suitable for modelling. Variables that could not be reconstructed reliably *post hoc* or that contained excessive missingness, were excluded from the modelling dataset. In addition, some variables were imputed to reduce missingness, and the categorical variables were dummy encoded into a model-ready format. All the preprocessing and cleaning steps are documented in Appendix C.2.4.

Final modelling dataset

Finally, the cleaned variables were converted into the format required for model development. Nominal categorical variables were transformed using dummy encoding, whereas ordinal variables were treated as ratio variables. After cleaning, reduction, and encoding, the final modelling dataset consisted of 101 observations, and 19 model features. The list of features and variables after dummy encoding are provided in Appendix C.2.5

4.3. Model validation strategy

A common validation approach is a single 80/20 train/test split (Ying, 2019). However, with a limited number of observations, relying on one split can lead to unstable estimates that depend strongly on which observations end up in the test set (Molinaro *et al.*, 2005). Therefore, this study uses a resampling-based validation strategy.

Survivorship bias

An important limitation of the dataset concerns its time coverage. The observations were exported from Brainial, a third-party tender management system used by Count & Cooper. During the data collection and validation process, it became clear that the dataset is not a complete chronological record of all tender opportunities. In particular, tenders dated before August 2024 are disproportionately represented by *bid* decisions, because during a system migration Count & Cooper primarily imported their historic tenders for which they had actively participated. As a result, historical *no-bid* decisions are underrepresented in the earlier part of the dataset: out of the 36 projects pre August 2024, 34 were ”bids”.

After August 2024, the distribution shifts in the opposite direction: bids become rare in the more recent observations (12 out of 65). When the dataset is sorted by the tender notice date, the bid/no-bid ratio therefore varies systematically over time. This pattern is a form of survivorship bias, indicating that the

observed sample reflects what was retained and migrated rather than the full population of bid/no-bid decisions (Carpenter & Lynch, 1999).

This skew has direct implications for model validation. A strictly time-based train/test split would produce training and test sets with different class distributions (many more bids in the earlier period than in the later period), making performance estimates unstable and difficult to interpret. This survivorship bias has impact on the choice of validation strategy.

Resampling options

Given the limited sample size, resampling is used to obtain more stable performance estimates by reusing the available data. Two standard options are cross-validation and bootstrapping, which involve a bias/variance trade-off (Burzykowski *et al.*, 2023; Kohavi, 1995; Molinaro *et al.*, 2005; Pargent *et al.*, 2023). Bootstrapping can have lower variance, but may suffer from larger bias in some settings (Kohavi, 1995; Molinaro *et al.*, 2005). Since this dataset already contains structural bias (as described above), bootstrap is less attractive for the present case. While cross-validation does not remove bias, repeated cross-validation can reduce the variance of the estimates compared to a single k-fold run (Kohavi, 1995; Molinaro *et al.*, 2005; Van Winckelen & Blockeel, 2012).

Time dependence and validation implications

The ultimate goal of the model is to predict and explain the bid/no-bid decisions for future tenders, which raises the question whether validation should be time-based. Standard k-fold cross-validation can be problematic for time-series forecasting, because shuffling can violate temporal causality and lead to optimistic estimates (Bergmeir *et al.*, 2018). In this case, however, strict time-based splits are not well-suited due to the small dataset and the Brainial-migration shift in class balance. A time split would place a relatively large share of ”bids” in the training set and relatively few in the test set, making performance estimates highly sensitive to the chosen cut-off. Prior work shows that, under stated assumptions and with careful overfitting control, cross-validation can still be used in time-indexed data and may perform well in practice (Bergmeir & Benítez, 2012; Bergmeir *et al.*, 2018; Pinto & Marçal, 2019). An analysis using a time-ordered split has been performed to validate this claim, and can be found in Appendix D.

Final validation approach

The final validation approach for model comparison is the *repeated* 5*-fold cross-validation*: the data is split into five folds, and each fold is evaluated once, and this procedure is repeated 10 times with different random seeds. This yields 50 train/test evaluations per model, and performance is reported as mean $\pm$ standard deviation over these repeats. This approach aims to (i) use the dataset efficiently, (ii) reduce variance in the estimates, and (iii) keep model comparison fair across algorithms (Kohavi, 1995; Nti *et al.*, 2021; Van Winckelen & Blockeel, 2012). During optimisation, conservative settings are used to prevent overfitting.

Finally, the k-fold cross-validation choice is also consistent with prior work in the construction tendering literature, where k-fold cross-validation is used to evaluate predictive models on relatively small datasets. For example, Sonmez and Sözgen (2017) apply 10-fold cross-validation for a bid/no-bid classifier, while related tender studies (bidding price and duration estimation) likewise report 10-fold cross-validation as their primary validation approach (Petruseva *et al.*, 2016; Tirataci & Yaman, 2023).

The repeated 5-fold cross-validation results are used to compare models relative to each other and to assess stability across resamples. Due to the dataset limitations (survivorship bias and time-varying class balance), the reported scores should not be interpreted as unbiased estimates of future real-world performance. Instead, they provide a best-effort and transparent comparison under the constraints of the available data. This limitation will be elaborated on in Chapter 6.

4.4. Performance evaluation and explainability methods

To compare the models on equal terms, five standard classification metrics were used: Accuracy, Precision, Recall, F1-score, and the Area Under the Curve (AUC). Accuracy was included as a general measure of overall correctness, while Precision and Recall were used to distinguish between false-positive and false-negative behaviour. The F1-score was included because it balances Precision and Recall into a single threshold-based measure, and AUC was used to assess the models' ranking ability

across possible decision thresholds. The definitions of the metrics and their formulas are provided in Appendix E.1.

In addition to mean test performance, model stability and overfitting risk were assessed by comparing training and test performance across the repeated resamples. Given the limited sample size, the aim was not to maximise predictive performance through extensive tuning, but to use conservative model settings that kept the comparison fair and reduced the risk of overly optimistic results. Class weights were not adjusted and the default decision threshold of 0.5 was kept, meaning that false positives and false negatives were treated equally in the model comparison. This was considered appropriate for the present exploratory study, since no reliable cost model was available to assign different practical weights to the two error types. Class weights, threshold, and error types will be further discussed in Section 6.4.2. The detailed hyperparameter settings are reported in Appendix E.2.

To examine explainability, model-specific interpretation methods were used. For the linear models, Logistic Regression and the linear Support Vector Machine, explainability was based on model coefficients, which indicate the direction and relative strength of the association between each predictor and the predicted class (Park, 2013). For the tree-based models, interpretability was assessed using feature-importance measures, with SHAP values used as the primary explanation method for the Random Forest model because they provide both local and aggregated views of feature contributions (Li, 2022; Louhichi *et al.*, 2023). For XGBoost, SHAP-based importance was likewise used as a complementary interpretation method (Li, 2022). For the nonlinear SVM with the rbf kernel, permutation-based feature importance was used to assess how strongly predictive performance changed when individual variables were added/removed (Liu *et al.*, 2011). Together, these methods made it possible to compare which variables emerged as influential across the different models, while recognising that the resulting importance measures are not numerically identical and should therefore be interpreted mainly in terms of direction, relative ranking, and cross-model consistency.

This chapter explained how the study was designed to analyse the bid/no-bid decision in a structured and reproducible way. First, the case study setting and data sources were introduced, followed by the model requirements and the selection of five supervised machine-learning models for comparative analysis. Next, the factors identified in the previous chapters were operationalised into a structured modelling dataset, including the use of a codebook, expert validation, diagnostic assessment, and preprocessing steps. The final dataset consisted of 101 observations and 19 model features.

Second, the model validation strategy was defined. Given the limited sample size and the presence of survivorship bias in the historic tender records, repeated 5-fold cross-validation was selected as the main validation approach. Finally, the chapter described how predictive performance and explainability would be assessed, using a combination of classification metrics, overfitting checks, and model-specific interpretation methods.

Together, these methodological choices provide a transparent basis for comparing the predictive and explanatory value of different machine-learning approaches for the bid/no-bid decision. The next chapter presents the results of this comparative analysis.

5

Results

This chapter contains the results of the predictive modelling approach and answers research sub-questions SQ3 and SQ4: (SQ3) *How does the predictive performance of different machine learning algorithms (such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and XGBoost) compare when modelling the contractor's bid/no-bid decisions?* and (SQ4) *Which variables emerge as most influential in the predictive models, and how can these be explained and interpreted?*. Together, SQ3 and SQ4 provide the empirical basis to answer the main research question by (i) establishing how well the outcome can be predicted from the available characteristics and (ii) explaining which characteristics most consistently contribute to those predictions.

This chapter is structured as follows. Section 5.1 presents the predictive performance comparison (SQ3), including an overall metrics table, an error analysis using confusion matrices, and diagnostics for stability and potential overfitting. Section 5.2 then addresses variable importance and interpretability (SQ4), first for the linear models and then for the tree-based models, before concluding with a cross-model agreement discussion. Appendix F contains supplementary results for the predictive performance and the explainability analysis.

5.1. Predictive performance comparison

This section answers SQ3 by comparing all candidate models in terms of predictive performance under the repeated cross-validation procedure described in Section 4.3. Performance is evaluated using Accuracy, F1, AUC, Precision, and Recall. This section has three components: the overall predictive comparison between the models (§5.1.1), their error behaviour (§5.1.2), and their stability and overfitting signals (§5.1.3).

5.1.1. Overall comparison table

Table 5.1 summarises the mean test performance and standard deviations over 10 repetitions of 5-fold cross-validation. Across the evaluated models, the Random Forest achieves the strongest overall performance, with the highest Accuracy (0.741), F1 (0.710), AUC (0.798), and Precision (0.758). This indicates that, on average, it provides the best balance between overall correctness (Accuracy), performance on the positive class (F1), ranking ability (AUC), and the reliability of positive predictions (Precision). Logistic Regression and the linear SVM perform almost the same, but with slightly lower Accuracy and F1 than the Random Forest. XGBoost follows with intermediate performance across all metrics (Accuracy $= 0.703$, AUC $= 0.750$), suggesting that it generalises reasonably well but does not match the Random Forest under the current hyperparameter settings.

The Decision Tree shows the weakest overall performance (Accuracy $= 0.639$, AUC $= 0.698$) but attains the highest Recall (0.694), implying that it captures a larger share of positive cases while producing less precise predictions (Precision $= 0.606$). Finally, the nonlinear SVM (rbf) exhibits relatively modest threshold-based performance (Accuracy $= 0.666$, F1 $= 0.645$) despite maintaining a comparatively strong AUC (0.763).

Overall, although the Random Forest achieves the highest average predictive performance, its advantage over Logistic Regression is limited: the difference in accuracy is only around one percentage point (0.741 vs. 0.733). Given that this thesis aims not only to predict the bid/no-bid outcome, but to also provide the explainability, Logistic Regression can therefore be regarded as the most suitable model for subsequent analysis. More on explainability in Section 5.2.

Table 5.1: Mean (standard deviation) test performance over 10 repetitions of 5-fold cross-validation.

Model	Accuracy	F1	AUC	Precision	Recall
Logistic Regression	0.733 (0.023)	0.699 (0.035)	0.785 (0.012)	0.752 (0.038)	0.670 (0.037)
Decision Tree	0.639 (0.025)	0.630 (0.049)	0.698 (0.022)	0.606 (0.029)	**0.694** (0.086)
Random Forest	**0.741** (0.026)	**0.710** (0.035)	**0.798** (0.014)	**0.758** (0.039)	0.688 (0.022)
XGBoost	0.703 (0.016)	0.679 (0.020)	0.750 (0.023)	0.709 (0.018)	0.668 (0.025)
SVM (linear)	0.712 (0.036)	0.690 (0.047)	0.777 (0.012)	0.711 (0.046)	0.689 (0.054)
SVM (rbf)	0.666 (0.020)	0.645 (0.027)	0.763 (0.011)	0.657 (0.030)	0.650 (0.029)

5.1.2. Error behaviour

It is important to understand the kinds of errors the models make. There are four possible outcomes for predicting a single project: correctly predicting a project that was indeed a bid (True Positive (TP)), correctly predicting a no-bid (True Negative (TN)), incorrectly predicting a bid (False Positive (FP)), and incorrectly predicting a no-bid (False Negative (FN)). Figure 5.1 reports the confusion matrices for Logistic Regression and Random Forest, the two overall strongest performers. In the bid/no-bid context, false negatives and false positives can have different practical implications: a false negative corresponds to rejecting a case that would belong to the positive class, which could result in missed opportunity costs, while a false positive corresponds to accepting a case that would belong to the negative class, which could result in excessive business costs. Both outcomes, therefore, result in extra costs; however, in this thesis FPs and FNs will be treated as equally bad (more on this in Section 6.4).

Figure 5.1 visualises the error behaviour of the Logistic Regression (LR) and Random Forest (RF) models using confusion matrices. For LR, on average 41.9 negative cases are classified correctly and 11.1 are misclassified as positive, while 32.1 positive cases are classified correctly and 15.9 are missed. RF predicts the same number of true negatives and false positives as LR, but a small shift within the positive class: fewer false negatives (15.1 vs. 15.9) and more true positives (32.9 vs. 32.1). This pattern is consistent with the slightly higher Recall and F1-score observed for the Random Forest in Table 5.1, as RF correctly identifies a marginally larger share of positive cases while maintaining a similar false-positive rate. The confusion matrices for the remaining models are provided in Appendix F.1.

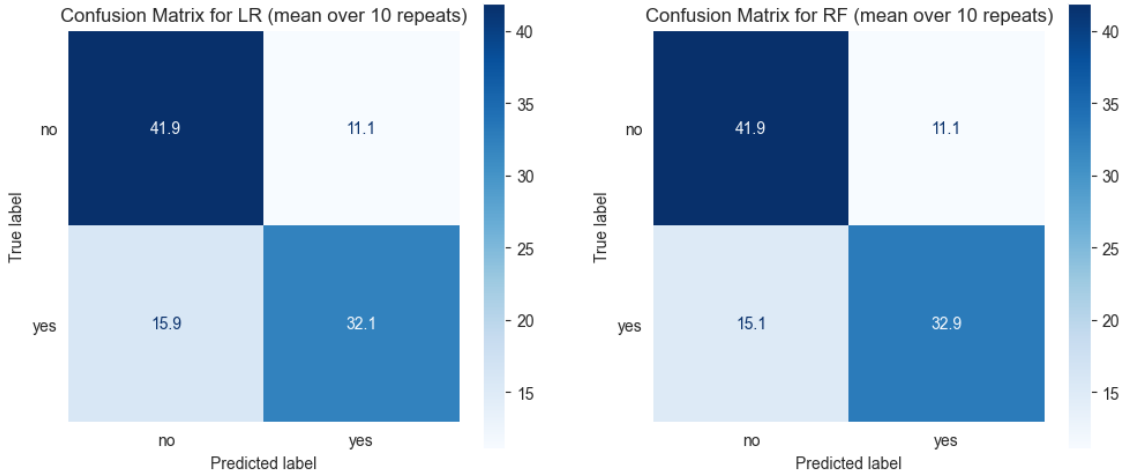


Figure 5.1: Confusion matrix for Logistic Regression (left) and Random Forest (right).

5.1.3. Stability and overfitting diagnostics

While repeated k-fold cross-validation provides a robust estimate of generalisation performance, model stability remains a concern in relatively small and heterogeneous datasets (Kohavi, 1995; Ying, 2019). Table 5.2 therefore reports mean train and test performance, along with train/test gaps for Accuracy and AUC. Larger gaps indicate a higher risk of overfitting, meaning the fitting of the training folds do not generalise to the test folds (Ying, 2019). None of the models show alarming signs of overfitting (ΔAcc < 10), and it is not uncommon for models to have a slightly higher train accuracy compared to their test accuracy (Ying, 2019). Among the evaluated models, the shallow Decision Tree and the linear SVM show the largest gaps. In contrast, XGBoost and the rbf SVM show smaller gaps, suggesting better controlled model complexity under the chosen hyperparameters, although their absolute test performance is lower than the Logistic Regression. In addition, LR has a low accuracy difference as well, while having one of the strongest test performance. This suggests that Logistic Regression is not only predictive, but also stable; making it a strong candidate for the explainability analysis.

Table 5.2: Train vs test performance and train-test gaps (mean values) over 10 repetitions of 5-fold cross-validation.

Model	Train Acc	Test Acc	$\triangle$ Acc	Train AUC	Test AUC	$\triangle$ AUC
Logistic Regression	0.778	0.733	0.045	0.844	0.785	0.059
Decision Tree	0.724	0.639	0.085	0.799	0.698	0.101
Random Forest	0.806	0.741	0.065	0.871	0.798	0.073
XGBoost	0.742	0.703	0.039	0.803	0.750	0.053
SVM (linear)	0.798	0.712	0.086	0.851	0.777	0.075
SVM (rbf)	0.702	0.666	0.037	0.813	0.763	0.051

5.2. Variable importance and interpretability

This section addresses SQ4 by examining which variables contribute most to the model predictions and whether the different models have the same thought-process. Interpretability between models cannot be done side-by-side, it is treated per model: linear models (Logistic Regression and linear SVM) can be compared through coefficients, tree-based (Decision Tree, Random Forest, and XGBoost) are analysed using feature importance and SHAP analyses, and finally for the rbf SVM the feature importance is evaluated through adding and removing a feature and record the change in predictive accuracy; this is called Recursive Feature Elimination (RFE) (Liu *et al.*, 2011). In this section, only the linear models (§5.2.1) and the Random Forest SHAP analysis (§5.2.2) are shown. Afterwards, the results of these three interpretability analyses will be compared (§5.2.3). The remaining feature importance results can be found in Appendix F.2.

5.2.1. Linear models

Table 5.3 reports coefficients for the Logistic Regression and the linear SVM models. For Logistic Regression, coefficients are expressed in log-odds units: holding other features constant, positive coefficients are associated with a higher predicted probability of the positive class, while negative coefficients are associated with a lower predicted probability (Park, 2013). For the linear SVM, coefficients represent decision-function weights; their signs can be interpreted directionally in the same way, but their magnitudes are not directly comparable to Logistic Regression due to the different objective function and scaling (Liu *et al.*, 2011; Navia-Vázquez & Parrado-Hernández, 2006). The table is divided per variable, and because of the dummy encoding, the reference feature is shown for clarity purposes.

Both models show multiple variables with the same directional effects. While the weight of the coefficients cannot be compared directly, their relative weight can be observed. In particular, *project type infrastructure*, *contract form d&c*, *contract conditions uav-gc*, and *procurement procedure other* in both models have the highest positive impact. In addition, *project type modern cities*, *contract form uav/raw*, and *procurement procedure open* all have the highest negative coefficient in both models. The signs and relative of the ordinal and ratio variables are also mostly aligned. This alignment of signs and relative significance of both models suggests that the linear approach is consistent across the two models.

Table 5.3: Coefficients of the two linear models. (Coefficients are not directly comparable across models).

Feature	LR coeff.	SVM (linear) coeff.
project type infrastructure	0.325	0.460
project type marine	0.119	0.058
project type modern cities	-0.292	-0.272
project type energy (ref.)		
contract form d&c	0.297	0.500
contract form dbfm(o)	0.062	0.143
contract form framework agreement	0.019	0.055
contract form other	-0.181	-0.342
contract form uav/raw	-0.278	-0.563
contract form bouwteam (ref.)		
contract conditions uav-gc	0.185	0.237
contract conditions uav	-0.068	0.005
contract conditions project-specific (ref.)		
procurement procedure other	0.194	0.255
procurement procedure restricted	0.100	0.133
procurement procedure open	-0.198	-0.188
procurement procedure cd (ref.)		
fictitious discount yes	-0.004	-0.109
fictitious discount no (ref.)		
project value bucket	0.147	0.109
contract duration	0.195	0.234
tender duration	0.006	0.005
tender docs quality	0.026	0.083
%price quality	0.114	0.208
Intercept	**-2.153**	**-2.255**

5.2.2. Tree-based models

The tree-based models (Decision Tree, Random Forest, and XGBoost) can capture nonlinear relationships and interaction effects that are not represented in the linear models (Singh *et al.*, 2016). To answer SQ4, interpretability is therefore assessed using SHAP (SHapley Additive exPlanations) for the Random Forest model (Figure 5.2). They allow you to understand the specific contribution of each feature to the prediction, and to quantify the impact of that feature of the prediction (Louhichi *et al.*, 2023). Positive SHAP values indicate that a feature pushes the model output towards the positive class, while negative SHAP values push it towards the negative class; the magnitude reflects the strength of that push for a given observation. Aggregating these local contributions across the dataset yields a global importance measure, commonly reported as the mean absolute SHAP value, which summarises how strongly each feature influences predictions on average (Carvalho *et al.*, 2019).

Figure 5.2 shows that the Random Forest's explanatory power comes from a small number of variables (the numerical values can be found in Appendix F.2). The most influential feature is *tender duration* (mean$|$SHAP$| = 0.0555$), which dominates all other features. A second tier of features includes *contract duration* (0.0285), *contract form uav/raw* (0.0268), and *project type infrastructure* (0.0253), indicating that contract length, a specific contract form, and infrastructure project type are recurrent determinants of the predicted class. *Project type modern cities* (0.0170), *project value* (0.0168), *contract conditions uav-gc* (0.0133) and *contract conditions uav* (0.0132) have smaller but still non-negligible average impact. Other variables, such as *tender documents quality* (0.0088), *%price quality* (0.0086), and *procurement procedure open* (0.0097), are almost negligible. Finally, several dummy variables have near-zero mean$|$SHAP$|$ values, implying that they rarely change predictions. The complete numerical SHAP importance values for the Random Forest, together with additional interpretability outputs for the Decision Tree and XGBoost models, are provided in Appendix F.2. For completeness and comparability across

the models, permutation-based importance for the rbf SVM is also included in this appendix.

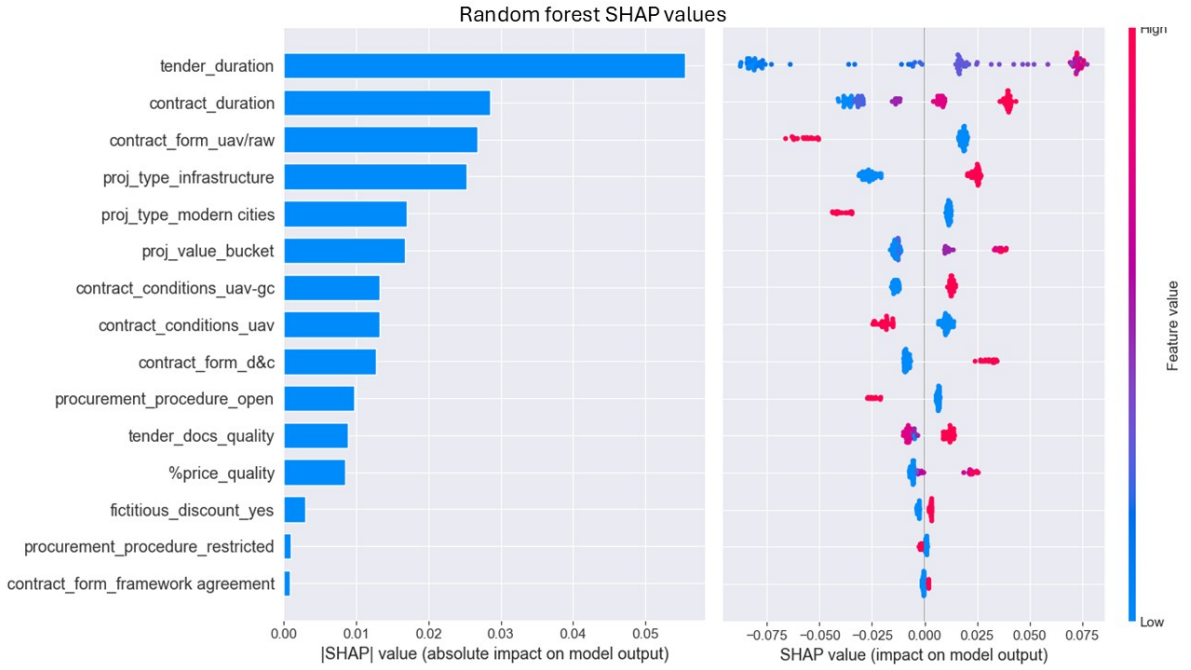


Figure 5.2: Random forest SHAP values.

5.2.3. Cross-model agreement

A key objective of this thesis is to explain each model and identify similar explanatory drivers across the different models (SQ4). This is done through analysing the coefficients of the two linear models and the SHAP values of the Random Forest model. Because the results cannot be compared directly, there will only be looked at the sign and ranking of a feature, rather than the exact numerical value.

In the Random Forest, *tender duration* is by far the dominant driver (highest mean|SHAP|), and *contract duration* also ranks among the most important features. In the linear models, *tender duration* and *contract duration* are included as an explanatory variable as well. In the LR model, *tender duration* has a coefficient of 0.006, this seems low but considering tender duration is operationalised per day and its average is 132.23, it has high impact ($0.006 \times 123.32 = 0.793$). The same goes for *contract duration*, the coefficient is 0.195 and its weighted average is 3.198 ($0.195 \times 3.198 = 0.624$). Second, project type has similar consistency: *project type infrastructure* appears among the top contributors in the Random Forest and carries relatively high positive coefficients in both linear models, indicating that this category is associated with a higher predicted likelihood of the positive class. Third, *contract conditions uav-gc* is a notable variable in the Random Forest and has a positive coefficient in both linear models, while contract form such as *uav/raw* and *d&c* appear as important in the Random Forest and carry comparatively large coefficients (in absolute value) in the linear specifications.

Taken together, these repeated appearances suggest that the models converge on a similar set of factors: time-related variables, project type, and contract forms and conditions. This cross-model agreement provides the answer to SQ4, indicating that the identified importance patterns are similar across the models. Although the Random Forest achieves the strongest overall predictive performance among the models (Table 5.1), Logistic Regression is used as the primary basis to answer SQ4 because its explainability through its coefficients is more valuable than the small predictive improvement of the Random Forest model.

6

Conclusion, discussion, and future work

This chapter synthesises the findings of the thesis and provides a concluding reflection on what can (and cannot) be claimed based on the conducted analyses. Building on this synthesis, the chapter discusses implications for practice, key limitations, and directions for future work. Finally, the chapter answers the main research question and closes this thesis with final remarks.

6.1. Synthesis of findings

This thesis examined to what extent an explainable machine learning model can *predict* and *explain* the bid/no-bid decision for large Dutch construction tenders, using a case-study dataset from Count & Cooper (C&C). The synthesis below summarises the main findings of the four sub-questions. First, SQ1-SQ2 are synthesised to describe the tendering context and which factors influence the bid/no-bid decision. Second, SQ3-SQ4 are synthesised to summarise the model comparison results and to interpret which variables drive the predictions.

6.1.1. SQ1-SQ2: Context and factors

SQ1: How does the tendering process for large Dutch construction projects unfold, and which decision points are most critical for contractors?

SQ1 (Chapter 2) is about understanding the context in which contractors must make the bid/no-bid decision. In this thesis, the bid/no-bid decision is defined as an internal ”green light” to invest substantial time and resources into developing a full proposal. Based on document analysis and expert input, the contractor’s tender effort during the procurement procedure can be structured into six decision points: (1) early scan, (2) pre-qualification decision, (3) teaming/consortium strategy, (4) solution strategy, (5) bid economics, and (6) final submit/no-submit. Importantly, these decision points are not black-and-white steps: in practice, decision points 2-4 are often intertwined and iterated as new information becomes available and strategic choices are revisited.

Within this structure, the bid/no-bid decision analysed in this thesis is positioned between (2) the pre-qualification and (3) the teaming/consortium strategy. This timing matters for modelling: at the moment of the bid/no-bid decision, there is still a lot of uncertainty, meaning that the decision is made under incomplete information. Accordingly, the factors identified in SQ2 should be limited to characteristics that are observable at the time the bid/no-bid decision is made.

SQ2: Which factors influence the bid/no-bid decision of contractors, and how have these been addressed in existing literature and practice?

The main output of SQ2 (Chapter 3) is a set of factors that influence bid/no-bid decisions, developed by synthesising prior literature and validating the factor list with C&C experts. The factors can be grouped into four clusters.

First, *project characteristics* include project size, project type, and project (contract) duration. Second, *contract characteristics* include contract form, contract conditions, and contract payment terms. Third, *tender characteristics* include tender duration, tender documents quality, procurement procedure, and selection and award criteria. Finally, *client/strategic characteristics* were repeatedly identified as influential in literature and by C&C expert. However, the last category of factors could not be reconstructed from historic tender records and were therefore treated as out-of-scope for the predictive modelling in this thesis.

6.1.2. SQ3-SQ4: Model comparison, results, and interpretation

SQ3: How does the predictive performance of different machine learning algorithms (such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and XGBoost) compare when modelling the contractor's bid/no-bid decisions?
In SQ3 (Chapter 5), multiple supervised learning models were compared: Logistic Regression, Decision Tree, Random Forest, XGBoost, and Support Vector Machines (linear and rbf kernels). Given the small sample size and the data limitations, model performance was evaluated using a repeated 5-fold cross-validation setup to obtain more stable comparative estimates across splits. The main result is that the Random Forest achieved the strongest overall test performance (74% accuracy), with Logistic Regression (73%) and linear SVM (71%) as a close second and third. Logistic Regression was selected as the primary model for further interpretation in SQ4, with the Random Forest and linear SVM models as support.

SQ4: Which variables emerge as most influential in the predictive models, and how can these be explained and interpreted?
The core output of SQ4 (Chapter 5) is an explanation of which variables matter most for the bid/no-bid predictions and how consistent these drivers are across modelling approaches. The linear models (LR and linear SVM) were interpreted using their coefficients (Table 5.3), while the Random Forest was interpreted using SHAP values to quantify feature contributions at the prediction level (Figure 5.2).

Tender duration, contract duration, project value, and infrastructure type projects repeatedly emerge as important drivers. Moreover, different models point to the same drivers, which increases confidence in the overarching explanatory story. At the same time, agreement is primarily assessed in terms of direction and ranking, because coefficient magnitudes and SHAP values are not directly comparable in scale across methods.

6.2. Recommendations

The results of this thesis suggest that explainable machine learning can support learning about bid/no-bid decisions, but that its practical value depends primarily on the quality and consistency of the underlying tender data. The main recommendation is therefore not to change how decisions are made, but to improve how decisions and their context are documented. Better documentation increases transparency in decision-making, reduces ambiguity in historic records, and creates a stronger foundation for future modelling and evaluation.

6.2.1. Practical implications for Count & Cooper

For Count & Cooper, a key improvement is to register the decision context more consistently at the moment the bid/no-bid decision is made. This includes recording where the tender is in the funnel, who was involved in the decision meeting, and the decision rationale in a short structured way (e.g., 2-3 main reasons with brief notes). The one-pagers that are currently used provide a good starting point for structuring all the released tender notices. In addition, the codebook (Appendix C.1) can be seen as an ”expanded” one-pager: a fully structured data point can always be used to generate a one-pager afterwards, while the reverse is not possible. Finally, to strengthen future modelling, C&C should start collecting the internal factors that were missing in the historic dataset (such as current workload, client capability, and strategic fit) and grow the dataset consistently over time. Until forward-looking performance can be validated in a time-based setup, model outputs should be used primarily as a conversation starter and as a transparent screening tool, rather than as an automated decision tool.

Shadow predictions of the latest bid/no-bid decisions

As an additional sense-check, a small "shadow prediction" was performed on the five most recent bid/no-bid decisions: the model predictions were compared with the outcomes stated by C&C business leads and discussed in a reflection session. This exercise is not meant as a statistical validation given the very small sample ($n = 5$), but as a qualitative check of whether the model outputs behave plausibly on recent, realistic decision situations and whether disagreements can be explained by information that is not captured in the current dataset.

Across the five projects, the discussion highlighted that several bid/no-bid decisions were driven by factors that are either not observable in historic tender documents or are highly context-dependent. Project 1 was considered a strong strategic fit and would normally be a clear "bid", but almost became a no-bid due to a capacity clash with a similar tender and concerns about intellectual property across two different external parties involved in these parallel tenders. Project 2 was described as a clear "no-bid" because it falls outside C&C's core business and does not meet the size and complexity requirements, even though it is related to the broader "Modern Cities" growth ambition. Project 3 was described as a clear "bid", but the rationale explicitly reflected a recent policy change: it fits the 2026 sector definition, whereas similar projects would not have been pursued earlier, implying that a model trained on historic decisions could not have anticipated this shift. Project 4 was a long internal debate and illustrates a consortium constraint at decision point 3: although the project itself was attractive, its geographic location and the lack of suitable local partners were decisive barriers. Finally, Project 5 was a clear "no-bid" because it was perceived as a service/maintenance framework with relatively small tasks and therefore insufficient complexity for C&C's way of work.

Overall, the filled-in shadow prediction results show a mixed but informative pattern across the five projects. Random Forest and SVM correctly matched the business-lead outcome in Projects 1, 2, and 5 (3/5), while Logistic Regression matched in Projects 2 and 5 (2/5). The project 3 and 4 mismatches are consistent across all models: Project 3 was labelled as a bid by the business leads, but predicted as a no-bid by all models, which aligns with the identified policy change in the 2026 sector definition that is not represented in the historic training data. Additionally, Project 4 was labelled as a no-bid, but predicted as a bid by all models, the consortium and location-related constraint (far away, lack of suitable partners) that is currently not encoded in the dataset. Together, these outcomes suggest that the models perform reasonably when decisions are driven by stable tender characteristics and clear scope fit (Projects 2 and 5), while they are more likely to diverge from practice in cases dominated by non-documentable constraints or shifts in organisational policy (Projects 3 and 4). Appendix G provides the explanatory results of the coefficients and SHAP for each project of each model.

Table 6.1: Shadow prediction overview for the five most recent projects.

Project	True label	LR	RF	SVM
Project 1	T	F	**T**	**T**
Project 2	F	**F**	**F**	**F**
Project 3	T	F	F	F
Project 4	F	T	T	T
Project 5	F	**F**	**F**	**F**

6.2.2. Recommendations beyond the case study

Beyond this case study, the main recommendation is to treat bid/no-bid decisions as a strategic dataset rather than as isolated outcomes. This means not only capturing what kind of project, but also the internal organisational context that drives the decision. To enable others to build on this work, the key requirements are clear ownership of the dataset and codebook, consistent definitions for labels and variables. These conditions make it possible to conduct more realistic time-based validation, and move from proof-of-concept modelling towards robust insights that generalise beyond a single case.

6.3. Limitations

This thesis provides a proof-of-concept for predicting and explaining bid/no-bid decisions using explainable machine learning. However, the conclusions that can be drawn are bounded by several limitations. This section outlines the main limitations of the study. The limitations are grouped into four categories: (i) data and sampling, (ii) measurement and label quality, (iii) validation and performance uncertainty, and (iv) interpretation and generalisability.

6.3.1. Data and sampling limitations

First, the dataset is relatively small ($N = 101$), which increases uncertainty in model performance estimates and makes results sensitive to individual observations. This limits the interpretations of the models, since they can be unstable. However, when compared to other research (Lowe and Parvar (2004) with $N = 115$ but heavily skewed towards ”bid” (99/115); and Sonmez and Sözgen (2017) with $N = 40$) the size of the dataset is arguably decent.

Second, the dataset reflects a single-company case-study context. It captures C&C’s portfolio and registration practices rather than a random or complete sample of all large Dutch tenders, which limits representativeness beyond the case context. Again, the expectation of gathering data of all Dutch tenders, or by multiple contractors is unrealistic because tendering is part of their core business and therefore confidential. In addition, Count & Cooper is specialised in large complex projects, and even megaprojects (>€500m project size). Only 3% of the procured projects in the Netherlands are larger than €10m (Hardeman, 2012), and all the projects in the dataset have a project size greater than €10m. This actually makes this thesis more relevant, because 70% of the sum of the procured project volumes is related to this 3%. However, it does result in underrepresentation of some tender types, contract forms, and procurement procedures.

Third, the dataset has a structural limitation due to the August 2024 migration, which introduced survivorship bias by overrepresenting accepted bids in earlier periods. As a result, the dataset cannot be treated as a complete chronological record of bid/no-bid opportunities. However, this migration does make the class balance stable for a k-fold validation approach (more on this in §6.3.3).

Fourth, the bid/no-bid distribution and the mix of tenders likely change over time. This potential distribution shift reduces the validity of conclusions that implicitly assume a stable decision environment across the full dataset. As can be seen in the shadow prediction in the previous section (§6.2.1), the third project was predicted as a strong ”no-bid” by all the models, however because of a recent policy change, they did decide to bid.

6.3.2. Measurement and label quality limitations

A key limitation is that several relevant drivers of bid/no-bid decisions could not be included in the modelling dataset. Internal factors such as current workload, strategic fit, and client capability were not consistently available in historic records and had many missing values. Therefore the client/strategic characteristics were removed from the final dataset, rendering them out-of-scope. Several papers did find these variables important, so removing these reduced the quality of the dataset and possibly reduced the predictive accuracy of the models (Cheng *et al.*, 2011; Ketaren & Sianturi, 2017; Sonmez & Sözgen, 2017).

The variables encoding used in the codebook (Appendix C.1 has impact on the final results. Variables such as *tender documents quality* and *selection criteria* were encoded on a scale of low to high. The values for each project were subjectively given by the researcher, introducing a researcher bias (Gao, 2020). A further encoding limitation is information loss. Buckets and ordinal scales simplify reality and may hide meaningful differences within categories, which can affect both model learning and the conclusions drawn from variable importance. Finally, for variables such as *contract conditions* some entries are straight-forward (e.g., UAV 2005/2012), because there is only one standardised version. However, UAV-GC 2005/2025 has multiple versions or applications. Meaning for one variable, the information is correctly reflected, but others may need a deeper view to better reflect reality.

6.3.3. Validation and performance uncertainty

The study does not include a realistic forward-looking evaluation on future tenders that were not yet known at the time of model development. One of the purposes of this thesis is to be able to see if ML is capable of predicting bid/no-bid decisions, including future decisions. However, with the limited datasize ($N = 101$) and the survivorship bias, a time-based split resulted in major instability and unreliable evaluation metrics (Appendix D). For this reason, a repeated 5-fold cross-validation was used to improve stability. This method was also used by other research (Sonmez & Sözgen, 2017), but this validation method mixes observations from different time periods. This complicates interpretation of the results when evaluating future prediction performance.

Other methods for stability were looked at, such as resampling. However, resampling cannot correct for structural bias in the underlying data, such as survivorship bias, missingness patterns, and researcher bias. Therefore, even stable cross-validation results may not translate directly to real-world future use. In addition, small differences in accuracy, F1, or AUC between models should not be over-interpreted as clear evidence that one model is superior to the other.

The study also does not take into account the trade-off between false positives (FP) and false negatives (FN). This matters because prediction errors result in excessive business costs (FP) or opportunity costs (FN). Knowing this trade-off could be used to adjust the decision threshold accordingly, resulting in less FP or FN, depending on adjustments. In addition, class balance also has impact on the amount of FP and FN. Because the bid/no-bid ratio in this dataset is not stable over time (due to the 2024 migration), a model trained and evaluated under this class balance may not behave the same way when the dataset better reflect the actual biddings. This interacts with threshold adjustment: the class balance affects the precision (FP-related) versus recall (FN-related) trade-off and the practical implication of FP and FN predictions, meaning that a combination of threshold and class-weighting choices has a lot of impact on the implication of the models.

6.3.4. Interpretation and generalisability

The interpretation of the results is another limitation. The explainability analysis only explains the behaviour of the model rather than causality. Coefficients and SHAP values indicate which variables the models use to get the highest accuracy, but they do not prove that those variable actually cause bid/no-bid outcomes.

Another question that remains open is if the drivers influencing the bid/no-bid decisions, are linear, nonlinear, or interaction effects. In this thesis, nonlinearity is only looked at through the SVM (with the rbf kernel) and by the assumed properties of certain models. In addition, variable interaction is not looked at, only through properties of the models. However, since the explainability of the model outcomes is limited, no final answer could be given. This further limits the interpretation of a single feature.

Generalisability is also constrained by the case study scope. The results are based on one organisation, one dataset, and a specific tendering context and time period, so transfer to other contractors, sectors, or market should be done with caution. Even if predictive performance is acceptable in-sample, practical deployment choices such as decision thresholds depend on context-specific risk preferences and the relative cost of false positives and false negatives. These preferences can differ across organisations and periods, meaning that a ”best” threshold or operating point may not transfer directly across settings.

6.4. Future work

Future work can follow two main directions. The first direction focuses on improving the current modelling approach by strengthening the data foundation, variable design, and model robustness. The second direction focuses on cost-sensitive decision support by explicitly exploring false negative versus false positive trade-offs and the implications for decision thresholds.

6.4.1. Improving data, variables, and model robustness

A first and most direct extension is to improve data capture. Future tender registrations should structurally include internal factors such as current workload, strategic fit, and client capability, as well as

short decision rationales and meeting context. This would reduce missingness in important drivers and possibly increase the predictive power; since these variables are found to be of importance in research (Cheng *et al.*, 2011; Ketaren & Sianturi, 2017; Sonmez & Sözgen, 2017). *Current workload* could be encoded as amount of FTEs available; *strategic fit* and *client capability* can both be put on an ordinal scale (1-5, for example), because these reflects the opinion of and the client's impression on the experts, which are both subjective.

A second extension is to improve variable construction and preprocessing as the dataset grows. This includes revisiting variable definitions and encodings (e.g., ordinal/buckets versus continuous measures where feasible), testing whether alternative transformations better reflect the underlying concepts, and systematically exploring feature selection. In particular, it may be valuable to evaluate whether certain variables add noise and whether the removal, imputation, or re-engineering of such features improves performance and stability.

A third direction is to strengthen robustness and generalisation through broader validation and replication. As more tenders become available, time-based evaluation can be introduced to assess forward-looking performance more realistically. In addition, applying the same operationalisation and modelling approach to other organisations would allow testing external validity and clarifying which drivers are case-specific and which patterns generalise across settings.

6.4.2. Cost-sensitive decision support (FN/FP trade-offs)

A second direction is to move from predictive performance toward decision usefulness by explicitly modelling the trade-off between false positives and false negatives. This requires defining the practical costs of both error types, for example by linking false positives to unnecessary tendering effort and false negatives to missed opportunities. An important component of this direction is the contractor's tender win percentage. Unnecessary tendering effort is only truly costly when bids are lost (or when projects that are won turn out to be unprofitable), while the cost of a missed opportunity depends on the likelihood that the contractor would have won and delivered value. In this sense, a higher win percentage increases the expected cost of false negatives and reduces the expected cost of false positives, and vice versa. By researching win percentage and approximating both cost types, future work could derive a more decision-relevant class weighting and decision threshold, leading to more practical implications if predictive performance remains reasonably high.

Based on such a cost model, future work can explore threshold selection and calibration. Rather than using the default 0.5, a more suitable threshold can be chosen to minimise decision cost under realistic class balance assumptions. This means that class balance has a high impact on the new threshold: if bid/no-bid ratios change over time, thresholds and class-weighting strategies may need to be recalibrated to preserve decision-relevant performance. Finally, cost-sensitive decision support should be tested by running the model parallel to real bid/no-bid meetings for a longer period. When time-based split results are stable and accurate enough, practical implications and actual decision support could be researched.

6.5. Final conclusion

This section answers the main research question central in this thesis: *To what extent can an explainable machine learning model predict and explain the bid/no-bid decision for a Dutch contractor?*

Machine learning can predict the bid/no-bid decisions with an accuracy of about 71-74%. Across models, variables such as tender duration, contract duration, project value, project type, and contract form repeatedly emerge as important drivers. Therefore, the bid/no-bid decision can be predicted and explained to a moderate extent, which makes the model more useful as transparent screening support than as an automated decision-maker. In addition, better registration of internal decision factors and continued dataset expansion should further enhance the predictability of the model.

To go more in depth regarding *prediction*, the results indicate that bid/no-bid outcomes can be predicted to a moderate extent from the available project, tender, and contract characteristics. In the repeated 5-fold cross-validation comparison, the Random Forest achieved the strongest overall performance (74% accuracy), with Logistic Regression (73%) and linear SVM (71%) closely behind. These results demonstrate that the constructed dataset contains measurable patterns that support prediction, but the

results should not be interpreted as proven forward-looking forecasting accuracy, given the dataset's limitations and bias. Also, because of the dataset size ($N = 101$), model stability is an issue, and therefore it cannot be concluded which model is the best fit for predicting the bid/no-bid decision.

Regarding *explanation*, the interpretability analyses show that the models provide insights into which variables affect predictions in this dataset. Linear models (Logistic Regression and linear SVM) were interpreted using their coefficients, while the Random Forest was interpreted using SHAP values to quantify feature contributions at the prediction level. Across models, time-related variables (most notably *tender duration*) and project and contract characteristics such as infrastructure projects and longer contract durations repeatedly emerge as important drivers. Moreover, different models point to the same drivers, which increases confidence in the overarching explanatory story. At the same time, agreement is primarily assessed in terms of direction and ranking, because coefficient magnitudes and SHAP values are not directly comparable in scale across methods.

To complement the cross-validation results, a small shadow prediction on the five most recent projects provided an additional qualitative check of forward-looking plausibility. In a reflection session, the model predictions were compared with the bid/no-bid outcomes stated by C&C business leads; given the very small sample ($n = 5$), this exercise is not a statistical validation but a small experiment to see if the model can support decision-making. Random Forest and linear SVM matched the business-lead outcomes in three out of five projects, while Logistic Regression matched in two out of five projects. The mismatches aligned with decision drivers that are currently not encoded in the dataset: one case reflected a policy change over time, and the other reflected a consortium and location constraint. This suggests that the models perform reasonably when decisions are driven by stable tender characteristics and clear scope fit, but they are more likely to diverge from practice in cases where policy or organisational constraints play a role.

The main scientific contribution of this thesis is methodological: it shows a practical and reproducible way to translate early-stage tender information of Dutch complex construction projects into a machine-learning-ready dataset. First, it provides a transparent, codebook-based approach with clear coding rules to keep subjectivity as low as possible. Second, it documents a complete modelling and validation workflow for bid/no-bid prediction, including comparing multiple model types and using interpretability methods to understand what the models are doing.

Empirically, the results suggest that bid/no-bid outcomes can be predicted to a moderate extent using early-available tender characteristics. This indicates that the features extracted from historic tender files can be systematically associated with the eventual bid/no-bid outcome. At the same time, the results also show that an important part of the decision remains outside the current dataset, because several internal and strategic drivers are not consistently documented. In particular, factors such as strategic fit, current workload, and client capability are likely to be of influence, yet could not be reconstructed reliably from the available data.

Taken together, the results provide not only evidence that explainable machine learning is meaningful for studying and predicting the bid/no-bid decision, but also a concrete starting point for future work on improved data capture. The current results should therefore be treated as a proof-of-concept rather than a deployable decision tool. By improving structured tender documentation (including internal factors and decision rationale), expanding the dataset over time, introducing time-based evaluation, and incorporating costs-of-errors, future work can move from modelling insights towards more practical decision support.

References

Aje, I., Olatunji, O., & Makanjuola, S. (2017). Bid or no-bid decision factors of indigenous contractors in nigeria. *Engineering Construction & Architectural Management*, *24*, 378–392. https://doi.org/10.1108/ECAM-01-2016-0029

Ayodele, T. O. (2010). Types of machine learning algorithms. In Y. Zhang (Ed.), *New advances in machine learning*. IntechOpen. https://doi.org/10.5772/9385

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation [Data Mining for Software Trustworthiness]. *Information Sciences*, *191*, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70–83. https://doi.org/10.1016/j.csda.2017.11.003

Bochenek, J. (2014). The contractor selection criteria in open and restricted procedures in the public sector in selected eu countries. *Procedia Engineering*, *85*, 69–74.

Brainial. (2025). *5 challenges in the construction tender process (and how to fix them)*. https://nl.brainial.com/tendermanagement/5-challenges-in-the-construction-tender-process-and-how-to-fix-them

Burzykowski, T., Geubbelmans, M., Rousseau, A.-J., & Valkenborg, D. (2023). Validation of machine learning algorithms. *American Journal of Orthodontics and Dentofacial Orthopedics*, *164*(2), 295–297. https://doi.org/10.1016/j.ajodo.2023.05.007

Carpenter, J., & Lynch, A. (1999). Survivorship bias and attrition effects in measures of performance persistance. *Journal of Financial Economics*, *54*, 337–374.

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, *8*(8). https://doi.org/10.3390/electronics8080832

Chauhan, V., Dahiya, K., & Sharma, A. (2019). Problem formulations and solvers in linear svm: A review. *Artif Intell Rev*, *52*, 803–855. https://doi.org/10.1007/s10462-018-9614-6

Cheng, M.-Y., Hsiang, C.-C., Tsai, H.-C., & Do, H.-L. (2011). Bidding decision making for construction company using a multi-criteria prospect model. *JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT*, *17*, 424–436. https://doi.org/10.3846/13923730.2011.598337

Creswell, J. W. (2009). *Qualitative, quantitative, and mixed methods approaches*. Research Design. Third ed. Thousand Oaks: Sage, chapter 1.

CROW kennisplatform. (n.d.). *Uav-gc 2025*. https://www.crow.nl/kennisproducten/uav-gc-2025/

de Sousa, L. J., Martins, J. P., & Sanhudo, L. (2024). Predicting construction project compliance with machine learning model: Case study using portuguese procurement data. *Engineering, Construction and Architectural Management*, *31*(13), 285–302. https://doi.org/10.1108/ECAM-09-2023-0973

Directive 2004/18/EC. (2004). *Directive 2004/18/EC of the European Parliament and of the Council of 31 March 2004 on the coordination of procedures for the award of public works contracts, public supply contracts and public service contracts*. https://eur-lex.europa.eu/eli/dir/2004/18/oj/eng

Directive 2014/24/EU. (2014). *Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC Text with EEA relevance*. https://eur-lex.europa.eu/eli/dir/2014/24/oj/eng

Egemen, M., & Mohamed, A. (2007). A framework for contractors to reach strategically correct bid/no bid and mark-up size decisions. *Building and Environment*, *42*(3), 1373–1385. https://doi.org/10.1016/j.buildenv.2005.11.016

El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In I. El Naqa, R. Li, & M. J. Murphy (Eds.), *Machine learning in radiation oncology: Theory and applications* (pp. 3–11). Springer International Publishing. https://doi.org/10.1007/978-3-319-18305-3_1

Fuentes-Bargues, J. L., González-Cruz, M. C., & González Gaya, C. (2017). Environmental criteria in the spanish public works procurement process. *International Journal of Environmental Research and Public Health*, *14*(2). https://doi.org/10.3390/ijerph14020204

Gao, Z. (2020). Researcher biases. In *The wiley encyclopedia of personality and individual differences* (pp. 37–41). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118970843.ch76

Gunduz, M., & Lutfi, H. M. A. (2021). Go/no-go decision model for owners using exhaustive chaid and quest decision tree algorithms. *Sustainability (Switzerland)*, *13*, 1–24. https://doi.org/10.3390/su13020815

Han, H., & Jiang, X. (2014). Overcome support vector machine diagnosis overfitting. *Cancer Informatics*, *13s1*, CIN.S13875. https://doi.org/10.4137/CIN.S13875

Hardeman, S. (2012). Aanbestedingsgedrag opdrachtgevers - aanbestedingen en transactiekosten 2009-2011. *Economisch Instituut voor de Bouw*.

Jang, W., Lee, J. K., Lee, J., & Han, S.-H. (2015). Naïve bayesian classifier for selecting good/bad projects during the early stage of international construction bidding decisions. *Mathematical Problems in Engineering*, *2015*, 1–12. https://doi.org/10.1155/2015/830781

Ketaren, K., & Sianturi, N. M. (2017). Decision making modelling with logistic regression approach. *International Journal of Applied Engineering Research*, *12*, 9067–9073.

Knoope, M., Faber, R., & Francke, J. (2022). Trendprognose wegverkeer 2022-2027 (tech. rep.). *Kennisinstituut voor Mobiliteitsbeleid*.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Stanford University*.

Koppenjan, J., Verweij, S., & van Marrewijk, A. (2024). An overview op ppps in the netherlands. In *Handbook on ppps in international infrastructure development: A critical perspective* (pp. 214–243). Edward Elgar Publishing. https://doi.org/10.4337/9781839102769.00014

Lenferink, S., Tillema, T., & Arts, J. (2013). Public-private interaction in contracting: Governance strategies in the competitive dialogue of dutch infrastructure projects. *Public Administration*, *91*(4), 928–946. https://doi.org/10.1111/padm.12033

Lesniak, A., Kubek, D., Plebankiewicz, E., Zima, K., & Belniak, S. (2018). Fuzzy ahp application for supporting contractors' bidding decision. *Symmetry*, *10*. https://doi.org/10.3390/sym10110642

Li, Z. (2022). Extracting spatial effects from machine learning model using local interpretation method: An example of shap and xgboost. *Computers, Environment and Urban Systems*, *96*, 101845. https://doi.org/10.1016/j.compenvurbsys.2022.101845

Liu, Q., Chen, C., Zhang, Y., & Hu, Z. (2011). Feature selection for support vector machines with rbf kernel. *Artificial Intelligence Review*, *36*(2), 99–115. https://doi.org/10.1007/s10462-011-9205-2

Louhichi, M., Nesmaoui, R., Mbarek, M., & Lazaar, M. (2023). Shapley values for explaining the black box nature of machine learning model clustering. *Procedia Computer Science*, *220*, 806–811. https://doi.org/10.1016/j.procs.2023.03.107

Lowe, D. J., & Parvar, J. (2004). A logistic regression approach to modelling the contractor's decision to bid. *Construction Management and Economics*, *22*, 643–653. https://doi.org/10.1080/01446190310001649056

Mahamid, I. (2022). Critical factors influencing the bid / no-bid decision in the palestinian construction industry. *Engineering, Technology & Applied Science Research*, *12*, 8096–8100. https://doi.org/10.48084/etasr.4538

Marinelli, M., & Antoniou, F. (2019). Improving public works' value for money: A new procurement strategy. *International Journal of Managing Projects in Business*, *13*(1), 85–102. https://doi.org/10.1108/IJMPB-04-2018-0084

Molinaro, A., Simon, R., & Pfeiffer, R. (2005). Prediction error estimation: A comparison of resampling methods. *Bioinformatics*, *21*(15), 3301–3307. https://doi.org/10.1093/bioinformatics/bti499

Murel, J., & Kavlakoglu, E. (n.d.). What is ensemble learning? *IBM*. https://www.ibm.com/think/topics/ensemble-learning

National Coordinator for Counterterrorism and Security. (2024). Cybersecurity assessment netherlands 2024. *Ministry of Justice and Security*.

Navia-Vázquez, A., & Parrado-Hernández, E. (2006). Support vector machine interpretation. *Neurocomputing*, *69*(13), 1754–1759. https://doi.org/10.1016/j.neucom.2005.12.118

NOS. (2025). *Renovatie van brienenoordbrug minstern twee keer zo duur als gedacht*. https://nos.nl/artikel/2566542-renovatie-van-brienenoordbrug-minstens-twee-keer-zo-duur-als-gedacht

Nti, I. K., Nyarko-Boateng, O., Aning, J., *et al.* (2021). Performance of machine learning algorithms with different k values in k-fold cross-validation. *International Journal of Information Technology and Computer Science*, *13*(6), 61–71.

Oo, B. L., Nguyen, A. T., Ahn, Y., & Lim, B. T. H. (2025). Predicting the number of bidders in construction competitive bidding using explainable machine learning models. *Construction Innovation*, *25*, 158–188. https://doi.org/10.1108/CI-10-2024-0325

Pargent, F., Schoedel, R., & Stachl, C. (2023). Best practices in supervised machine learning: A tutorial for psychologists. *Advances in Methods and Practices in Psychological Science*, *6*(3), 1–35. https://doi.org/10.1177/25152459231162559

Park, H.-A. (2013). An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain. *jkan*, *43*(2), 154–164. https://doi.org/10.4040/jkan.2013.43.2.154

Petruseva, S., Sherrod, P., Zileska Pancovska, V., & Petrovski, A. (2016). Predicting bidding price in construction using Support Vector Machine. *TEM Journal*, *5*(2), 143–151. https://doi.org/10.18421/TEM52-04

PIANOo. (n.d.-a). *Aanbestedingswet 2012*. https://www.pianoo.nl/nl/regelgeving/aanbestedingswet-2012

PIANOo. (n.d.-b). *Contracteren in de gww*. https://www.pianoo.nl/nl/sectoren/gww/inkopen-gww/contracteren-de-gww

PIANOo. (n.d.-c). *Drempelbedragen europees aanbesteden*. https://www.pianoo.nl/nl/regelgeving/drempelbedragen-europees-aanbesteden

PIANOo. (n.d.-d). *Selectiecriteria*. https://www.pianoo.nl/nl/inkopen-het-kort/hoe-ga-ik-met-de-regels-om/selectiecriteria

PIANOo. (n.d.-e). *Uniform europees aanbestedingsdocument (uea)*. https://www.pianoo.nl/nl/regelgeving/uniform-europees-aanbestedingsdocument-uea

PIANOo. (n.d.-f). *Uniforme administratieve voorwaarden (uav en uav-gc)*. https://www.pianoo.nl/nl/regelgeving/voorwaarden/uniforme-administratieve-voorwaarden-uav-en-uav-gc

Pinto, J. M., & Marçal, E. F. (2019). Cross-validation based forecasting method: A machine learning approach. *FGV EESP – Escola de Economia de Sao Paulo*.

Plebankiewicz, E. (2024). Procedures for awarding work contracts in europe. *Buildings*, *14*(4), 883. https://doi.org/10.3390/buildings14040883

Prasetyo, M. L., Peranginangin, R. A., Martinovic, N., Ichsan, M., & Wicaksono, H. (2025). Artificial intelligence in open innovation project management: A systematic literature review on technologies, applications, and integration requirements. *Journal of Open Innovation: Technology, Market, and Complexity*, *11*(1), 100445. https://doi.org/10.1016/j.joitmc.2024.100445

Rampini, L., & Cecconi, F. R. (2022). Artificial intelligence in construction asset management: A review of present status, challenges and future opportunities. *Journal of Information Technology in Construction (ITcon)*, *27*, 884–913. https://doi.org/10.36680/j.itcon.2022.043

Rijksdienst voor Ondernemend Nederland (RVO). (2024). *Aanbestedingsregels*. https://ondernemersplein.overheid.nl/aanbestedingsregels/

Rijkswaterstaat. (2025). *A16: Onderhoud en vernieuwen van brienenoordbrug*. https://www.rijkswaterstaat.nl/wegen/projectenoverzicht/a16-onderhoud-en-vernieuwen-van-brienenoordbrug

Ryo, M., & Rillig, M. C. (2017). Statistically reinforced machine learning for nonlinear patterns and variable interactions. *Ecosphere*, *8*(11), e01976. https://doi.org/10.1002/ecs2.1976

Schrijfgroep Gids Proportionaliteit. (2022). Gids proportionaliteit, 3e herziening. *Ministerie van Economische zaken en Klimaat*.

Scikit-learn. (n.d.-a). *1.1.11. logistic regression*. https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

Scikit-learn. (n.d.-b). *1.10. decision trees*. https://scikit-learn.org/stable/modules/tree.html

Scikit-learn. (n.d.-c). *1.11. ensembles: Gradient boosting, random forests, bagging, voring, stacking*. https://scikit-learn.org/stable/modules/ensemble.html

Scikit-learn. (n.d.-d). *1.4. support vector machines*. https://scikit-learn.org/stable/modules/svm.html

Shokri-Ghasabeh, M., & Chileshe, N. (2016). Critical factors influencing the bid/no bid decision in the australian construction industry. *Construction Innovation*, *16*. https://doi.org/10.1108/CI-04-2015-0021

SIGMA. (2016). Brief 10 - public procurement procedures. *OECD and EU*.

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 1310–1315.

Sonmez, R., & Sözgen, B. (2017). A support vector machine method for bid/no bid decision making. *Journal of Civil Engineering and Management*, *23*, 641–649. https://doi.org/10.3846/13923730.2017.1281836

Taboada, I., Daneshpajouh, A., Toledo, N., & Vass, T. (2023). Artificial intelligence enabled project management: A systematic literature review. *Applied Sciences*, *13*, 5014. https://doi.org/10.3390/app13085014

Thaler, R. (1991). *The winner's curse: Paradoxes and anomalies of economic life*. The Free Press.

Tirataci, H., & Yaman, H. (2023). Estimation of ideal construction duration in tender preparation stage for housing projects. *Organization, Technology and Management in Construction: an International Journal*, *15*(1), 192–212. https://doi.org/10.2478/otmcj-2023-0014

Trouw. (2024). *Bam wil meer groene woningen bouwen en denkt dat er genoeg ruimte is*. https://www.trouw.nl/duurzaamheid-economie/bam-wil-meer-groene-woningen-bouwen-en-denkt-dat-er-genoeg-ruimte-is~b7a86628/

Van Winckelen, G., & Blockeel, H. (2012). On estimating model accuracy with repeated cross-validation. *KU Leuven*.

van Ham, J. C., & Koppenjan, J. F. M. (2002). Port expansion and public-private partnership: The case of rotterdam. *WIT Transactions on the Built Environment*, *62*, 13–22.

Verweij, S., & van Meerkerk, I. (2021). Do public–private partnerships achieve better time and cost performance than regular contracts? *Public Money & Management*, *41*(4), 286–295. https://doi.org/10.1080/09540962.2020.1752011

Wanous, M., Boussabaine, A. H., & Lewis, J. (1998). Tendering factors considered by syrian contractors. *14th annual ARCOM Conference, Association of Researchers in Construction Management*, *2*, 535, 42.

xgboost developers. (n.d.). *Xgboost python package api reference*. https://xgboost.readthedocs.io/en/latest/index.html

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, *1168*(2), 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

A

Tendering context - extra information

A.1. Dutch procurement instruments

This appendix complements the EU procurement laws within the Dutch legal framework of §2.1.1. In addition to the EU procurement obligations, the Netherlands has instruments that help with the implementation and approach of procurement, and with the legal framework. These are the Aanbestedingsreglement Werken 2016 (ARW 2016), Gids Proportionaliteit, and the UEA/ESPD (Uniform Europees Aanbestedingsdocument/European Single Procurement Document).

ARW (Works Procurement Regulations) sets out standardised procedures for public works tenders. It is mandatory for works below the EU thresholds; for higher values or for supplies/services, use is highly recommended. It provides a consistent way to run the four EU-law procedures for works (§2.1.2) while aligning with the Dutch legal framework.

The Gids Proportionaliteit (Proportionality Guide) is mandatory guidance that ensures all requirements and criteria are proportionate to the scope and risk of the contract. It addresses selection requirements (e.g., experience, financial capacity), the design of award criteria and their weighting, the proportional use of guarantees, insurances and certifications, and reasonable demands on turnover and references (Schrijfgroep Gids Proportionaliteit, 2022).

The UEA/ESPD is a self-declaration of compliance with exclusion grounds and suitability criteria (e.g., financial and technical capacity) and is mandatory for tenders above EU thresholds (PIANOo, n.d.-e). By standardising the self-declaration of compliance, the administrative costs during the selection phase will be reduced.

A.2. Four main procurement procedures

This appendix complements §2.1.2. Figure A.1 visualises these routes side by side so their main steps and difference can be observed. Below, each procedure is explained in more detail.

Open procedure: Used when the works are well specified and of low to moderate complexity, and the authority wants broad competition without pre-shortlisting. It is a single-stage route: the contracting authority publishes a notice and makes the full tender documents available to all interested suppliers; bidders submit both selection (qualification) information and their tenders in one go. Selection checks and award evaluation are applied in sequence; no negotiations are permitted (only clarifications). Tenders may be evaluated on either lowest price or the most economically advantageous tender (MEAT/EMVI) (Marinelli & Antoniou, 2019; Plebankiewicz, 2024; SIGMA, 2016).

Restricted procedure: Chosen when capability should be vetted before pricing and the client wants to manage the number of full bids. It runs in two stages: after the notice, economic operators submit requests to participate (UEA/ESPD) for the selection stage; the authority may then draw up a shortlist and invite only those bidders to submit tenders with the full invitation-to-tender documents. This limits the number of tenders and reduces wasted effort. No negotiations are permitted (only clarifica-

tions). Award may be on lowest price or MEAT/EMVI (Marinelli & Antoniou, 2019; Plebankiewicz, 2024; SIGMA, 2016).

Competitive Procedure with Negotiation (CPwN): Suitable where project specifications are not set in stone, and/or the type of work is complex. After selection (requests to participate and shortlisting), the authority invites shortlisted bidders and issues initial tender documents; bidders submit initial tenders and then enter one or more negotiation rounds on negotiable elements. When negotiations close, an invitation to submit final tenders is issued and final tenders are evaluated. This provides structured interaction, but requires careful governance and equal-treatment records. Award typically be on MEAT/EMVI, or sometimes lowest price (Plebankiewicz, 2024; SIGMA, 2016).

Competitive Dialogue (CD): Used for high-complexity or innovative projects where the solution cannot be fully defined upfront and outcomes must be shaped with market input. After selection and shortlisting, the authority issues descriptive/dialogue documents and conducts one or more dialogue rounds to develop the solution; the number of solutions may be reduced during the process. Once the dialogue is formally closed and the requirements are fixed, the authority issues invitations for the final tenders. Under CD, award can only be given on the most economically advantageous tender (MEAT/EMVI); no negotiations on the final tenders are permitted beyond clarifications (Lenferink *et al.*, 2013; Marinelli & Antoniou, 2019; Plebankiewicz, 2024; SIGMA, 2016).

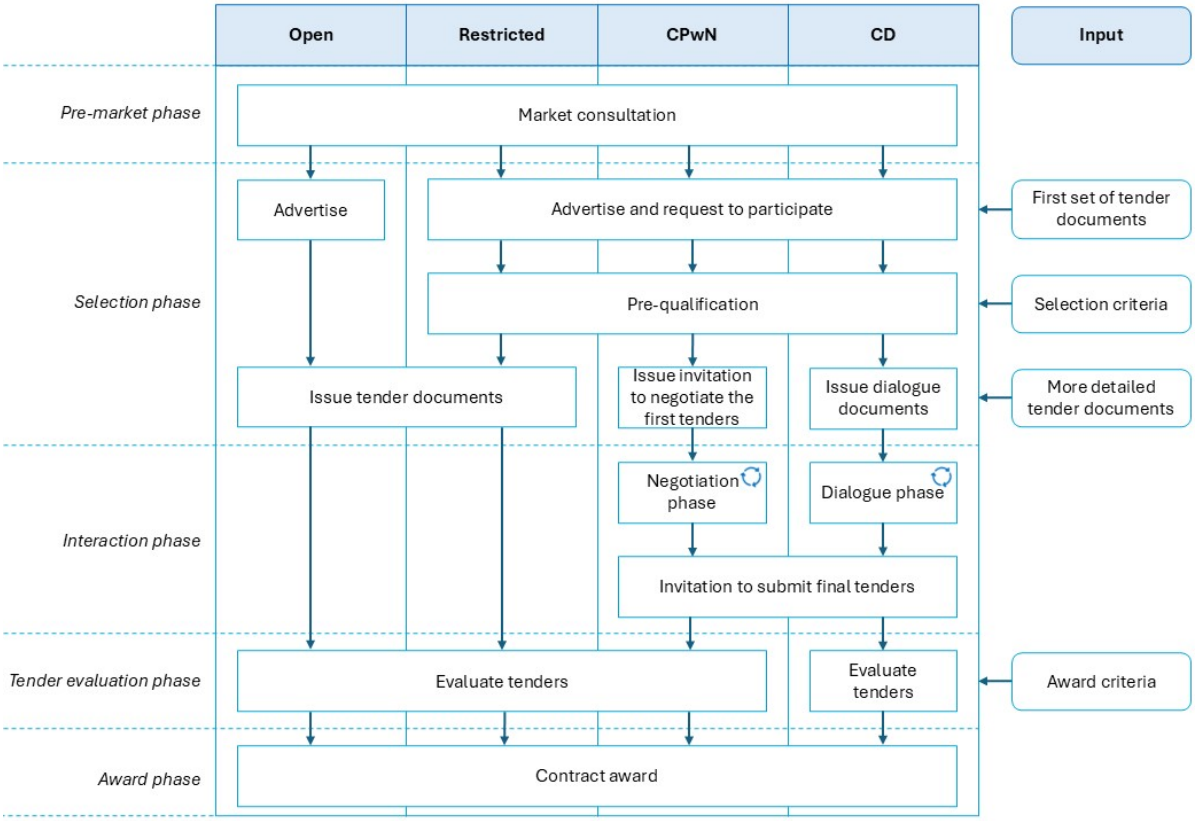


Figure A.1: Flowchart visualising the main competitive procedures side by side.

A.3. Selection and award criteria - extra information

This section compliments §2.1.3 and provides a more in-depth explanation of the selection and award criteria.

Selection criteria

There are two main elements within selection criteria: exclusion grounds and suitability requirements. Exclusion grounds address issues such as legal or tax non-compliance and serious professional misconduct, and lead to a binary decision on admissibility. Suitability requirements concern legal standing,

financial and economic capacity, and technical or professional capacity (PIANOo, n.d.-a; SIGMA, 2016). A selection criterion could, for example, be that a contractor must have designed and constructed a movable concrete bridge that could handle at least 500 tons in the last ten years.

Proportionality is a core principle: thresholds for turnover, the number and relevance of references, certifications, or key personnel qualifications must be calibrated to the contract's size, risk profile, and allocation of responsibilities (Schrijfgroep Gids Proportionaliteit, 2022). In the open procedure, selection checks are performed after the submissions; in the other three procedures, they occur at the request-to-participate stage and may be followed by shortlisting. Shortlisting, where used, limits the number of invited bidders after selection has been passed (SIGMA, 2016).

Award criteria

Award criteria determine which compliant tender is the most economically advantageous. This is done under the MEAT (Most Economically Advantageous Tender, sometimes also called EMAT; and in Dutch it is called *EMVI*), combining qualitative criteria with price or cost according to weights and a stated calculation method. Typical qualitative elements include the plan of approach, risk management, planning/phasing, stakeholder and environmental management, and sustainability. Each element is graded according to predefined criteria, and the scores are combined with the price/cost component to yield an overall ranking (Fuentes-Bargues *et al.*, 2017; PIANOo, n.d.-a; Plebankiewicz, 2024; SIGMA, 2016). For simple, well-specified works, lowest price can be used as an alternative route. However, that could result in imperfect competition on the market by means of artificially depressing the prices, which may lead to unforeseen costs (Bochenek, 2014). In Competitive Dialogue the award must be based on MEAT, which is why, in Figure A.1, CD has a stand alone *Evaluate Tenders* (SIGMA, 2016).

Across procedures, the timing differs but the logic remains the same: in open and restricted procedures, award is based on final tenders as submitted (with clarifications where appropriate); in the competitive procedure with negotiation, initial tenders may be refined through negotiations before final tenders are evaluated; and in competitive dialogue, the solution is developed during the dialogue phase and only the final tenders are evaluated for award under MEAT (Plebankiewicz, 2024; SIGMA, 2016). Selection and award criteria shape how bidders enter and progress through competitions, and thus influence contractor decision-making.

A.4. Contract forms descriptions

This section compliments §2.1.4 by providing a description for each contract form.

Cost-Plus: The client pays for the actual costs plus a fee. Cost overrun remains with the client. Tendering focuses on rates, caps, and capacity rather than a detailed solution.

RAW/Bestek: The client provides drawings and specifications; the contractor builds to them, usually on a lump-sum (fixed total price) or unit-rate basis (per-unit pricing). The client holds responsibility for requirements and design definition; the contractor carries out the work preparation and execution. Tendering concentrates on quantities, constructable viability, and change handling.

Bouwteam: Client and contractor jointly develop the design, target price, and risk picture before a separate execution agreement. Responsibilities are shared in the development phase and shifted towards the contractor during work preparation and execution. Tendering emphasises the collaboration approach and early engagement.

Design & Build: This contract form moves the design responsibility and coordination to the bidder. Tendering therefore requires front-end engineering, a clear plan for managing interfaces, and early commitment from the designer and critical subcontractors.

Turnkey: The contractor is responsible for both the design and construction of a facility. The basic concept is that in a Turnkey Contract the contractor shall provide the works ready for use at the agreed price and by a fixed date. Beyond design and build, the contractor is also accountable for integrating systems and proving performance at handover. Tendering centres on integration and testing strategies and the credibility of guarantees.

DBFM(O): Design, build, financing and long-term maintenance (and sometimes operation) are bundled. The consortium carries out the whole life-cycle of the project, including risks. The contractor is paid for

availability over time. Tendering adds life-cycle modelling, an operations/maintenance partner, and a full financial plan (and often a management consultancy). The selection is strict.

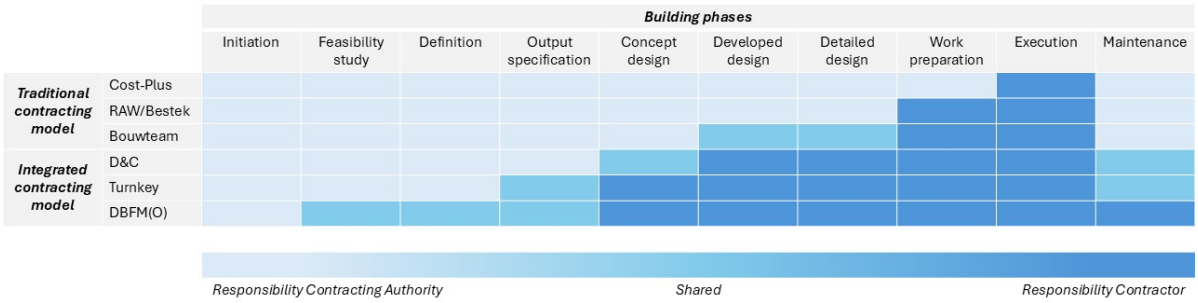


Figure A.2: Distribution of responsibility per contract form (Source: adjusted from Count & Cooper).

The more integrated the contract form, the more responsibility transfers to the contractor (see Figure A.2). Under traditional forms, the bid primarily demonstrates conformity with the client's design, and an accurate view of quantities and productivity. Under design-and-build and turnkey forms, the bid must also show how the contractor will shape and validate the solution, control interfaces, and prove that the finished asset will perform. Under long-term forms, the bid must show how the solution will remain reliable in use and how it will be maintained.

Contract conditions (UAV and UAV-GC): In addition to these contract form, Dutch public works contracts typically refer to a standard set of contract conditions that shapes the legal relationship between client and contractor (PIANOo, n.d.-f). For traditionally specified works, this is commonly the UAV 2012 (Uniform Administrative Conditions), which is written for execution contracts where the client provides the design basis and the contractor is primarily responsible for work preparation and execution. For integrated contracts, clients can also use UAV-GC. UAV-GC 2025 is made for project with a design, construct, and/or maintain component (CROW kennisplatform, n.d.). As a result, two projects with the same "contract form" label (e.g., D&C) may still differ in risk allocation depending on whether UAV or UAV-GC applies. For this reason, this thesis treats contract form and contract conditions as separate analytical constructs and separate variables in the dataset.

A.5. Decision points of contractors in the tendering process

There are six identified decision points for contractors during the entire procurement process. These are listed below in chronological order. During the *pre-market phase*, a contractor can decide whether or not to participate in the market consultation; in this research this is called the **early scan** (1) decision. During the *selection phase*, the four main decisions must be made: (2) the **pre-qualification decision**, (3) the **teaming/consortium strategy**, (4) the **solution strategy**, and (5) the **bid economics** decision. Bid economics (5) may also happen during the *interaction phase*, depending on the chosen procurement process. Lastly, a final decision must be made whether to submit the final tender or not: the **final submit/no-submit** decision (6). Below, the six decision points are explained in more detail.

Early scan: Many contractors use third party tender tools to scan for potential projects that are interested for them (e.g. TenderNed or Brainial). Contractors can apply filters to search for projects in their expertise. A selection of these projects are then further examined.

Pre-qualification decision: The next decision is whether to enter the *selection phase*. This depends on the selection criteria, project characteristics, partnering, etc. This choice is heavily based on expertise and experience of the management.

Teaming/consortium strategy: When deciding to tender, different teaming strategies can be chosen: e.g., joint venture, subcontract, consortium, or partnerships. This will have impact on the selection criteria qualification, bid strategy, and earnings.

Solution strategy: This defines what the team will propose and how it will score on quality, and what concepts, methods, variants, and assumptions will be used. This strategy will be the base for the bid

economics.

Bid economics: This decision will be based on the award criteria that the contracting authority has implied. It decides what the project should cost, what margins are used, and it will guide the price-quality trade-off. The content of the tender is based on this decision.

Final submit/no-submit: The final decision whether to submit the tender or not. The last check that reassesses the risks, margins, planning, costs, etc. to see if it is still feasible. In the Netherlands, not submitting rarely happens, since contractors will not receive any tender compensation when deciding not to submit.

B

Machine learning model principles

This appendix introduces the machine learning models used to predict the bid/no-bid decision and motivates their inclusion in the comparative analysis. These different models enable a balanced comparison between models that prioritise interpretability and those that prioritise more on predictive performance. For each model, the learning principle is explained and its expected strengths and limitations are discussed in the context of the dataset characteristics (limited sample size and a mixture of categorical and numerical predictors). This part complements the machine learning selection for the comparative analysis, explained in Section 4.1.

B.1. Selected models

Logistic Regression (baseline)

Logistic Regression serves as the baseline model for this study. It is a widely used statistical classification method that estimates the probability of a binary outcome based on a set of independent variables. The algorithm applies a logistic function to model the relationship between explanatory variables and the target variable, producing a probability value between 0 and 1 (Park, 2013). Logistic Regression provides the baseline for evaluating the added value of more complex models, as it has high interpretability and a probabilistic output, but assumes a linear relationship between predictors; it also suffers from multicollinearity and may have unstable results if the dataset is small (Lowe & Parvar, 2004; Park, 2013; Singh *et al.*, 2016).

Decision Tree

The Decision Tree (DT) algorithm represents a classification method that splits the data into subsets based on the most informative input variables. Each internal node of the tree corresponds to a decision rule, while each leaf node represents a predicted class. This hierarchical structure enables the model to capture nonlinear relationships and variable interactions in a transparent and interpretable way (Carvalho *et al.*, 2019). Decision Trees can handle a variety of data types (nominal, numeric, textual), missing values, and redundant attributes; have good generalisation ability; and are robust to noise. However, it is difficult to handle high dimensional data with DTs, as it is prone to overfitting. As the tree grows, the number of records in the leaf nodes may be too small to make statistically significant decisions about the class representation. This is called the *Data Fragmentation Problem* (Singh *et al.*, 2016). It can be avoided by disallowing further splitting when numbers of records falls below a certain threshold. Because Decision Trees can easily overfit, ensemble methods such as Random Forest were developed to improve predictive stability and generalisation.

Random Forest

Random Forest (RF) is an ensemble method that operates by training a number of Decision Trees and returning the class with the majority over all the trees in the ensemble. An ensemble model combines several individual models to produce more accurate predictions than a single model alone (Murel & Kavlakoglu, n.d.). RFs are robust to noise and typically generalise well due to averaging over many

trees. However, RF is less interpretable than a single tree (feature importance helps, but it is not transparent), and as the number of trees increases, the model can become slower (Oo *et al.*, 2025; Singh *et al.*, 2016).

XGBoost

Extreme Gradient Boosting (XGBoost) is an ensemble learning method based on the principle of gradient boosting, where multiple weak learners (typically, Decision Trees) are sequentially trained to correct the residual errors of preceding trees (Li, 2022). Each new tree is fitted on the gradient of the loss function, which allows the model to iteratively minimise prediction error and improve overall accuracy (Murel & Kavlakoglu, n.d.). Compared to traditional boosting algorithms, XGBoost introduces several optimisations, including regularisation to prevent overfitting, parallel computation, and efficient handling of missing values.

Support Vector Machine

The Support Vector Machine (SVM) is a supervised classification algorithm that seeks to find the optimal hyperplane that best separates data points belonging to different classes (Han & Jiang, 2014; Sonmez & Sözgen, 2017). The optimal hyperplane is defined as the one that maximises the margin between the closest points of the two classes, known as support vectors (Ayodele, 2010). SVM avoids overfitting, has flexible selection of kernels for nonlinearity, and it can be effective when there are many features compared to the number of observations (Singh *et al.*, 2016).

In this research, SVM is likely to perform well, because of the high-dimensional spaces and its effectiveness when the number of features is high relative to the number of observations (Singh *et al.*, 2016). However, SVMs are sensitive to parameter choices and kernel selection, and they offer limited interpretability compared to tree-based models, making them primarily useful as a performance benchmark rather than for explanatory insights (Navia-Vázquez & Parrado-Hernández, 2006).

B.2. Main differences between the models

Together, these algorithms enable a comparison of modelling approaches for the bid/no-bid decision. Although they are trained on the same dataset and evaluated with the same metrics, the models have different approaches in reaching their optimum result. These differences have impact on (i) transparency, (ii) nonlinearity, and (iii) prone to overfitting. Table B.1 summarises these differences; the paragraphs below discuss the implications for the present dataset and study objectives.

Table B.1: Comparison of the selected models.

Model	Interpretability	Nonlinearity	Overfitting risk
Logistic Regression	High (coefficients)	Only linear	Low
Decision Tree	High (pathways)	Captures nonlinearity	High overfitting risk without constraints
Random Forest	Medium (only feature importance)	Strongly captures nonlinearity	Medium overfitting risk without constraints
XGBoost	Medium (feature importance/SHAP possible)	Very strong modelling of complex patterns	Can overfit if not regularised
SVM	Low/high (depends on kernel)	Depends on kernel	Low overfitting risk

Sources: Carvalho *et al.* (2019), Li (2022), Navia-Vázquez and Parrado-Hernández (2006), and Park (2013), Chauhan *et al.* (2019), Han and Jiang (2014), Ryo and Rillig (2017), and Ying (2019)

Interpretability

Interpretability (or explainability) describes how easily a model's predictions can be explained in terms of the input variables. In this research, only post-model interpretability will be regarded (Carvalho *et al.*, 2019). Logistic Regression and Decision Trees are most transparent: coefficients provide the influence

of each variable, while a tree provides a set of human-readable decision rules (Park, 2013). Ensemble methods (Random Forest and XGBoost) typically have a higher predictive rate, but reduce transparency because predictions result from many trees (Li, 2022). Decision taken by Support Vector Machines (SVM) are hard to interpret from a human perspective (Navia-Vázquez & Parrado-Hernández, 2006). In this study, interpretability is relevant because the model is also intended to support the contractor's decision-making.

Nonlinearity & interactions

The models differ in the complexity of relationships they can represent. Logistic Regression is linear and therefore represents additive effects unless interaction terms or nonlinear transformations are added by hand (Park, 2013). In contrast, tree-based models naturally capture threshold effects, nonlinearity, and interactions through recursive splitting (Ryo & Rillig, 2017). SVMs can represent nonlinear decision boundaries when using nonlinear kernels, whereas linear kernels behave similarly to linear classifiers (Chauhan *et al.*, 2019).

Overfitting risk

With a limited number of observations, models are prone to overfitting if not contained. Decision Trees use features to make decisions, and if there are relatively many features, the model can overfit. Network-reduction constraints such as maximum depth and minimum samples per leaf are means to reduce such risks (Ying, 2019). Random Forest reduces this risk by averaging across many decorrelated trees, typically improving stability (Singh *et al.*, 2016). XGBoost can achieve strong performance, but is also capable of overfitting small datasets if the trees are too deep (de Sousa *et al.*, 2024). Logistic Regression tends to be stable, but can suffer when predictors are highly correlated (common with one-hot encoded categories) (Park, 2013). SVMs can perform well in high-dimensional settings, but their generalisation depends strongly on appropriate regularisation and kernel settings (Han & Jiang, 2014).

These algorithms enable a systematic comparison of modelling approaches for the bid/no-bid decision. By applying both traditional statistical and ensemble methods under identical conditions, the analysis can assess how different models capture the complex interactions among project, organisational, and institutional factors. This comparative setup acts an extra validity of the overall research design.

C

Operationalisation codebook

This appendix documents the detailed operationalisation, diagnostic analysis, and preprocessing steps underlying the modelling dataset. This part complements Section 4.2 in the main text by providing the full dataset overview, data-quality assessment, bid/no-bid crosstabs, and data-cleaning steps that were used to construct the final modelling dataset.

C.1. Codebook

Table C.1 presents the codebook used to construct the quantitative dataset for the bid/no-bid analysis. For each feature, it specifies the corresponding variable name used in the dataset, the primary data source(s) (tender notice, tender documents, and/or Count & Cooper experts), the measurement type (nominal, ordinal, or ratio), and the applied coding scheme. Together, these definitions ensure that all project-, contract-, tender-, and client-related characteristics are operationalised consistently across cases and can be directly used as input for the subsequent diagnostic analysis and model development. Please find the description of each variable in Section 3.2. Before the codebook is discussed, two important data gathering aspects must be addressed: subjectivity and internal assessment.

Subjectivity

The following variables are based on tender documents but require interpretive coding: tender documents quality and selection criteria. *Tender document quality* captures the completeness and clarity of the project to the contractors. The assessment focuses on whether the project scope and requirements are sufficiently specified, whether all appendices are available (e.g., technical specifications, drawings, annexes, draft contract), whether quantities/volumes/planning are described. Based on these aspects, tenders are coded on an ordinal scale where higher values reflect a more complete and less ambiguous information set. This means that the researcher's subjectivity has influence of the rating a project receives. In addition to the four quality levels in Table C.1, a separate category *No guidelines* exists, it sometimes happens that a market consultation was held, but the project got cancelled before officially publishing. This category is treated as the lowest-information condition and is therefore coded as 0 in the dataset to keep the ordinal interpretation consistent.

Selection criteria reflects how demanding the participation requirements are for contractors. It is derived from suitability and capability requirements stated in the tender notice or documents, such as the required company size, the number and type of reference projects, the assessed quality of those references, and the quality of the submitted CVs for key personnel (PIANOo, n.d.-d). These requirements are mapped to an ordinal scale (low/medium/high) that prioritises the strictness of the requirements rather than any single criterion. This means that the researcher's subjectivity also has influence on this mapping. When the documentation does not describe participation requirements clearly enough to apply these rules, the value is coded as *unknown/unclear*.

Internal assessment

The following variables would normally be known internally, but could not be reconstructed during the making of the dataset: *strategic potential*, *current workload*, and *client capability*. These factors depend on time-specific internal context (e.g., strategy, resource availability, and prior client experiences) that was not documented by Count & Cooper. This means this data is unavailable at the time of constructing the dataset. In the codebook, a coding is provided to assess this factor if information is available.

The next step is to describe the dataset, examine distributions and class balance, and assess data quality issues such as missing or "unknown" values. This will happen in the next section (Appendix C.2.1). The insights from this diagnostic analysis inform any final preprocessing choices and the specification of the final dataset used for model development.

Table C.1: Codebook used to operationalise the features identified in Section 3.2.

Feature	Variable name	Source	Type	Coding
Project size	proj_value_bucket	Tender notice Tender documents C&C experts	Ordinal	1= <€50m; 2= €50-100m; 3= €100-250m; 4= €250-500m; 5= >€500m; 0= unknown
Project type	proj_type	Tender documents	Nominal	Infrastructure Marine Energy Modern cities
Project duration	contract_duration	Tender notice Tender documents C&C experts	Ordinal	1= <2 years 2= 2-<3 years 3= 3-<4 years 4= 4-<6 years 5= >6 years 0= unknown
Contract type (see §2.1.4)	contract_type	Tender notice Tender documents	Nominal	Cost-plus RAW/Bestek D&C Turnkey DBFM(O) Framework agreement unknown
Contract conditions (see §2.1.4)	contract_conditions	Tender notice Tender documents	Nominal	UAV 2005/2012 UAV-GC 2005/2025 Other standard Project specific
Payment structure	payment_structure	Tender documents	Nominal	Lump-sum/fixed price Cost-plus/regie Other unknown
Tender duration	tender_duration	Tender documents	Ratio	[final_submission_date] - [date_official_publication] or [selection_decision]

Continued on the next page...

Table C.1: Codebook used to operationalise the features defined in Section 3.2. *(continued)*

Feature	Variable name	Source	Type	Coding
Documents quality	tender_docs_quality	Tender documents	Ordinal	1= Very poor/incomplete 2= Poor/unclear 3= Acceptable 4= Good/minor ambiguities 0= No guidelines
Procurement procedure	procurement_procedure	Tender notice Tender documents	Nominal	Open Restricted CPwN CD
Selection criteria	selection_criteria	Tender notice Tender documents	Ordinal	1= Low requirements 2= Medium requirements 3= High requirements 0= unknown/unclear
Award criteria	award_criteria	Tender notice Tender documents	Nominal	Lowest price MEAT/EMVI
Price/quality	%price_quality	Tender notice Tender documents	Mixed	Fictitious discount % price quality
New projects potential	strategic_potential	C&C experts	Ordinal	0= No/low strategic potential 1= Medium strategic potential 2= High strategic potential
Current workload	current_workload	C&C experts	Ordinal	1= Low available workload 2= Medium available workload 3= High available workload
Previous experience	previous_experience	C&C experts	Binary	0= No 1= Yes

C.1.1. Expert validation of factor operationalisation

Expert input is used to assess the practical relevance of the operationalisation choices, particularly the bucketing of ordinal variables and the information availability at the bid/no-bid decision moment. Because Count & Cooper focuses on large and complex projects, experts confirmed that the project-size buckets should reflect different project ”calibre” rather than equal monetary increments. In this context for example, EUR 150 million and EUR 250 million are relatively the same projects. However, EUR 250 and 350 million are different calibres. The same counts for megaprojects that are larger than EUR 500 million.

Second, the expert input clarified several patterns of missingness and limited observability in the dataset. For payment structure, experts indicated that it is normal for the information to be unknown at the bid/no-bid moment, as payment terms are typically negotiated after award; consequently, high levels of missingness are expected and the variable is rarely a decisive driver in C&C’s decision-making. In contrast, the client/strategy characteristics (e.g., strategic fit, client capability, potential for future work, and current workload) were described as highly influential in practice but were not documented because they are internal and time dependent “snapshots”. These factors are important, but not observable in the available documentation.

C.2. Initial dataset, diagnostic results, and preprocessing

This section presents the detailed diagnostic analysis and preprocessing procedure that followed the initial codebook-based operationalisation. First, the initial dataset is described through class-balance tables and numerical summary statistics. Second, data quality is assessed by examining missing values,

skewed categories, and descriptive associations with the bid/no-bid outcome. Finally, the data-cleaning and encoding steps used to construct the final modelling dataset are documented.

C.2.1. Initial dataset overview

After applying the codebook, an initial dataset was constructed in which each observation represented one tender opportunity for which the bid/no-bid outcome was known. The resulting dataset contained $N = 101$ observations and 16 variables. Because class balance can affect model learning and interpretation, the category distributions of the target variable and all categorical and ordinal predictors are reported in Table C.2. Numerical variables are summarised separately in Table C.3.

Table C.2: Class balance of target, nominal, and ordinal variables in the initial dataset ($N = 101$).

Variable	Category	n	%
bid/no-bid	Yes	48	47.52
	No	53	52.48
proj_type	Infrastructure	52	51.49
	Modern cities	20	19.80
	Marine	16	15.84
	Energy	13	12.87
contract_form	Framework agreement	23	22.77
	RAW/Bestek	22	21.78
	D&C	19	18.81
	Bouwteam	17	16.83
	DBFM(O)	8	7.92
	Cost-plus	5	4.95
	Unknown	5	4.95
	Stabu	2	1.98
contract_conditions	UAV-GC	49	48.51
	UAV	30	29.70
	Unknown	11	10.89
	Other standard	9	8.91
	Project-specific	2	1.98
payment_structure	Unknown	80	79.21
	Cost-plus/reimbursement	8	7.92
	Other	7	6.93
	Lump-sum/fixed price	6	5.94
procurement_procedure	CD	36	35.64
	Restricted	35	34.65
	Open	18	17.82
	CPwN	9	8.91
	Unknown	2	1.98
	Other	1	0.99
award_criteria	MEAT	95	94.06
	Lowest price	6	5.94
proj_value_bucket	<€50m	41	40.59
	€100–250m	18	17.82
	€50–100m	15	14.85
	€250–500m	10	9.90
	Unknown	10	9.90
	>€500m	7	6.93
contract_duration	>6 years	31	30.69

Continued on the next page...

Table C.2: Class balance of target, nominal, and ordinal variables in the initial dataset ($N = 101$). *(continued)*

Variable	Category	n	%
	<2 years	22	21.78
	4–<6 years	20	19.80
	2–<3 years	18	17.82
	3–<4 years	10	9.90
tender_docs_quality	Acceptable	46	45.54
	Good/minor ambiguities	38	37.62
	Poor/unclear	8	7.92
	No guidelines	6	5.94
	Very poor/incomplete	3	2.97
selection_criteria	High requirements	78	77.23
	Medium requirements	15	14.85
	Unknown	5	4.95
	Low requirements	3	2.97

The merged overview shows that the target variable is reasonably balanced, whereas several predictors contain pronounced class imbalance. In particular, Infrastructure projects dominate `proj_type`, the `selection_criteria` variable is heavily concentrated in the *high requirements* category, and `payment_structure` is dominated by *unknown*. These patterns are important because they can affect the stability of model learning and the interpretability of category-specific effects.

The dataset contains two numerical variables: `tender_duration` and %`price_quality`. Their summary statistics are reported in Table C.3.

Table C.3: Summary statistics of numerical variables.

(a) Tender duration (days)

Statistic	Value
Count	96
Mean	132.23
Std	93.70
Min	28
25%	70
50% (Median)	112
75%	166
Max	554

(b) % price-quality (ratio)

Statistic	Value
Count	43
Mean	0.5779
Std	0.3247
Min	0.0000
25%	0.4500
50% (Median)	0.6000
75%	0.8000
Max	1.0000

C.2.2. Data quality assessment

Missingness is a key data-quality concern because it reduces the effective sample size and may bias estimates when the absence of information is systematic. In this dataset, missingness often appears as *unknown* values, reflecting information that is not available in tender documentation at the time of the bid/no-bid decision. Table C.4 reports the number and percentage of missing values per variable.

Table C.4: Missing values per variable in the initial dataset ($N = 101$).

Variable	Missing (n)	Missing (%)
strategic_potential	101	100.00
current_workload	101	100.00
client_capability	101	100.00
payment_structure	80	79.21
contract_conditions	11	10.89
proj_value_bucket	10	9.90
%price_quality	7	6.93
tender_duration	5	4.95
contract_form	5	4.95
selection_criteria	5	4.95
procurement_procedure	2	1.98
proj_type	0	0.00
contract_duration	0	0.00
tender_docs_quality	0	0.00
award_criteria	0	0.00
bid/no-bid	0	0.00

The results show substantial variation across variables. Some variables are fully observed, whereas others contain moderate or severe missingness. In most cases, the observed missingness aligns with what can realistically be retrieved from tender documentation at the time of a bid/no-bid decision. For example, `payment_structure` is missing for 80 tenders (79.21%), which is plausible because payment arrangements are often not explicitly specified in the announcement or are only clarified later in the contracting phase. Similarly, `proj_value_bucket` is missing in 10 cases (9.90%) and `contract_conditions` in 11 cases (10.89%), which is consistent with the fact that clients do not always disclose budget ranges or provide complete contractual details in the released tender documents. The `%price_quality` variable is missing for 7 tenders (6.93%), which also makes sense because award weights are not always published in a structured way.

However, some missing values are less expected because the information is typically known at the start of the tender process. `Tender_duration` is missing for 5 tenders (4.95%), even though publication and deadline dates are generally available that should allow this variable to be derived. Similarly, `selection_criteria` has 5 missing values (4.95%), and `procurement_procedure` is missing for 2 tenders (1.98%), despite both being standard elements of tender announcements. Finally, three variables – `strategic_potential`, `current_workload`, and `client_capability` – are fully missing. While these factors would normally be known internally during the bid/no-bid decision, they could not be reconstructed reliably in this post-hoc dataset creation.

C.2.3. Bid/no-bid crosstabs

To further explore the descriptive structure of the initial dataset, Table C.5 reports crosstabs between each nominal and ordinal predictor and the bid/no-bid outcome. These patterns are purely descriptive and should not be interpreted as causal effects.

Table C.5: Crosstabs for nominal and ordinal variables versus bid/no-bid ($N = 101$).

Variable	Category	No (n)	No (%)	Yes (n)	Yes (%)	Total
proj_type	Energy	9	69.2	4	30.8	13
	Infrastructure	19	36.5	33	63.5	52
	Marine	8	50.0	8	50.0	16
	Modern cities	17	85.0	3	15.0	20
contract_form	Stabu	2	100.0	0	0.0	2

Continued on the next page...

Table C.5: Crosstabs for nominal and ordinal variables versus bid/no-bid ($N = 101$). *(continued)*

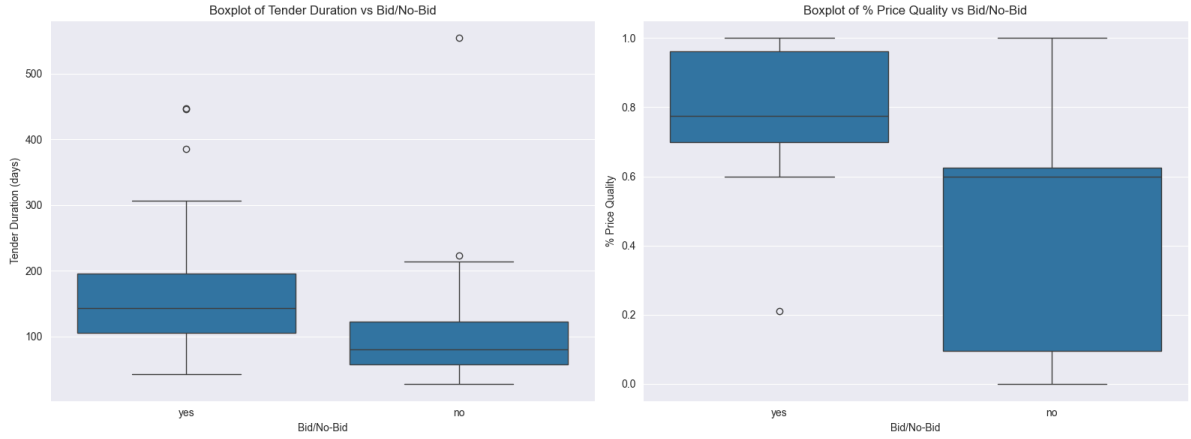
Variable	Category	No (n)	No (%)	Yes (n)	Yes (%)	Total
	Bouwteam	9	52.9	8	47.1	17
	Cost-plus	2	40.0	3	60.0	5
	D&C	4	21.1	15	78.9	19
	DBFM(O)	1	12.5	7	87.5	8
	Framework agreement	11	47.8	12	52.2	23
	RAW/Bestek	19	86.4	3	13.6	22
	Unknown	5	100.0	0	0.0	5
contract_conditions	Other standard	4	44.4	5	55.6	9
	Project-specific	1	50.0	1	50.0	2
	UAV	22	73.3	8	26.7	30
	UAV-GC	18	36.7	31	63.3	49
	Unknown	8	72.7	3	27.3	11
payment_structure	Cost-plus/reimbursement	4	50.0	4	50.0	8
	Lump-sum/fixed price	6	100.0	0	0.0	6
	Other	4	57.1	3	42.9	7
	Unknown	39	48.8	41	51.2	80
procurement_procedure	CD	13	36.1	23	63.9	36
	CPwN	3	33.3	6	66.7	9
	Open	15	83.3	3	16.7	18
	Other	0	0.0	1	100.0	1
	Restricted	20	57.1	15	42.9	35
	Unknown	2	100.0	0	0.0	2
award_criteria	Lowest price	6	100.0	0	0.0	6
	MEAT	47	49.5	48	50.5	95
proj_value_bucket	<€50m	25	61.0	16	39.0	41
	>€500m	1	14.3	6	85.7	7
	Unknown	10	100.0	0	0.0	10
	€100–250m	9	50.0	9	50.0	18
	€250–500m	2	20.0	8	80.0	10
	€50–100m	6	40.0	9	60.0	15
contract_duration	2–<3 years	11	61.1	7	38.9	18
	3–<4 years	6	60.0	4	40.0	10
	4–<6 years	10	50.0	10	50.0	20
	<2 years	17	77.3	5	22.7	22
	>6 years	9	29.0	22	71.0	31
tender_docs_quality	Acceptable	31	67.4	15	32.6	46
	Good/minor ambiguities	14	36.8	24	63.2	38
	No guidelines	3	50.0	3	50.0	6
	Poor/unclear	4	50.0	4	50.0	8
	Very poor/incomplete	1	33.3	2	66.7	3
selection_criteria	High requirements	38	48.7	40	51.3	78
	Low requirements	2	66.7	1	33.3	3
	Medium requirements	9	60.0	6	40.0	15
	Unknown	4	80.0	1	20.0	5

The crosstabs show that some variables contain potentially informative differences across bid and no-bid cases, whereas others are dominated by one large category with relatively little separation. For example, Infrastructure tenders show a markedly higher bid share than Modern cities tenders, while higher project-value buckets appear descriptively associated with a greater likelihood of bidding.

By contrast, variables such as award_criteria and selection_criteria show limited discriminatory value within this sample because most observations fall into one dominant category.

Figure C.1 presents box plots of the two numerical variables split by the bid/no-bid outcome. For tender_duration, bid cases tend to be associated with longer preparation time. A similar upward pattern is visible for %price_quality, suggesting that tenders with a stronger quality component are more likely to be pursued in this sample. These differences are descriptive associations only.

Figure C.1: Box plots of the continuous variables *tender_duration* and *%price_quality*.



C.2.4. Data-cleaning and preprocessing steps

This subsection documents the data-cleaning and preprocessing steps that were applied to construct the final modelling dataset. The rules below are presented in the same order as implemented.

To avoid modelling on variables that could not be reconstructed reliably or that contained excessive missingness, the following variables were removed from the dataset:

- strategic_potential, current_workload, and client_capability (fully missing post hoc).
- payment_structure (majority missing).
- selection_criteria (highly imbalanced; dominated by the *high requirements* category and therefore of limited discriminatory value).

Missing values in contract_conditions (coded as unknown) were imputed using rule-based mappings based on common pairings between contract_form and contract_conditions:

- contract_form = d&c and contract_conditions = unknown → uav-gc
- contract_form = bouwteam and contract_conditions = unknown → uav-gc
- contract_form = framework agreement and contract_conditions = unknown → uav
- contract_form = cost-plus and contract_conditions = unknown → project-specific
- contract_form = unknown and contract_conditions = unknown → project-specific

After these imputations, the remaining contract_conditions categories unknown and other standard were recoded to project-specific to reduce sparsity:

$$\texttt{contract\_conditions} \in \{\texttt{unknown}, \texttt{other standard}\} \rightarrow \texttt{project-specific}.$$

To reduce sparsity in low-frequency categories and limit the number of dummy variables during one-hot encoding, the following contract_form categories were grouped into other:

$$\texttt{contract\_form} \in \{\texttt{Stabu}, \texttt{cost-plus}, \texttt{unknown}\} \rightarrow \texttt{other}.$$

First, unknown values in procurement_procedure were imputed with the modal category. Second, the categories cpwn and other were consolidated into a single other category to reduce sparsity:

$$\texttt{procurement\_procedure} \in \{\texttt{cpwn}, \texttt{other}\} \rightarrow \texttt{other}.$$

The variable tender_duration was converted to numeric. Missing values were imputed with the median tender_duration observed in the dataset:

$$\texttt{tender\_duration} \leftarrow \texttt{median-imputation for missing values}.$$

Because award_criteria was dominated by MEAT, and because MEAT can be operationalised either through an explicit price–quality weighting or through a fictitious discount mechanism, award_criteria was transformed into a binary indicator:

- fictitious_discount = yes if award_criteria = MEAT and %price_quality = fictitious discount
- fictitious_discount = no otherwise

Additionally, lowest price tenders were assigned %price_quality = 0 before removing award_criteria. The original award_criteria column was then dropped. Finally, %price_quality was converted to numeric; values such as fictitious discount and other non-numeric strings were coerced to missing (NaN) for numeric modelling purposes. These missing values were then imputed as the median of the %price_quality variable.

To reduce dimensionality relative to the limited sample size, selected ordinal variables were encoded as numeric scores and treated as approximately continuous during modelling. The following mappings were applied:

Project value bucket (proj_value_bucket)

- 1 = <€50m
- 2 = €50--100m
- 3 = €100--250m
- 4 = €250--500m
- 5 = >€500m
- unknown → treated separately according to the preprocessing script

Contract duration (contract_duration)

- 1 = <2 years
- 2 = 2--<3 years
- 3 = 3--<4 years
- 4 = 4--<6 years
- 5 = >6 years

Tender documents quality (tender_docs_quality)

- 0 = no guidelines / unknown
- 1 = very poor/incomplete
- 2 = poor/unclear
- 3 = acceptable
- 4 = good/minor ambiguities

After applying the above steps, remaining missing values were re-evaluated to confirm which variables still contained missing entries. At this stage, missingness remained primarily concentrated in %price_quality due to the fictitious-discount mechanism and other non-numeric representations that cannot be converted to a ratio value.

C.2.5. Final modelling dataset

After data cleaning and feature engineering, the final modelling dataset consisted of $N = 101$ observations, 10 retained variables, and 19 model input features after dummy encoding. The retained variables include ordinal variables encoded as numeric scores, ratio variables, and one-hot encoded nominal predictors.

The final feature list is:

- proj_value_bucket
- contract_duration
- tender_duration
- tender_docs_quality
- %price_quality
- proj_type_infrastructure
- proj_type_marine
- proj_type_modern cities
- contract_form_d&c
- contract_form_dbfm(o)
- contract_form_framework agreement
- contract_form_other
- contract_form_RAW/Bestek
- contract_conditions_uav
- contract_conditions_uav-gc
- procurement_procedure_open
- procurement_procedure_other
- procurement_procedure_restricted
- fictitious_discount_yes

The only variable still containing substantial missingness in the final modelling dataset was the adjusted %price_quality ratio. This remaining missingness was retained because it resulted directly from the way fictitious-discount tenders were operationalised.

Figure C.2 reports the correlation matrix for selected predictors after encoding choices were applied. It is included as a final diagnostic to assess potential multicollinearity and clustering of related variables.

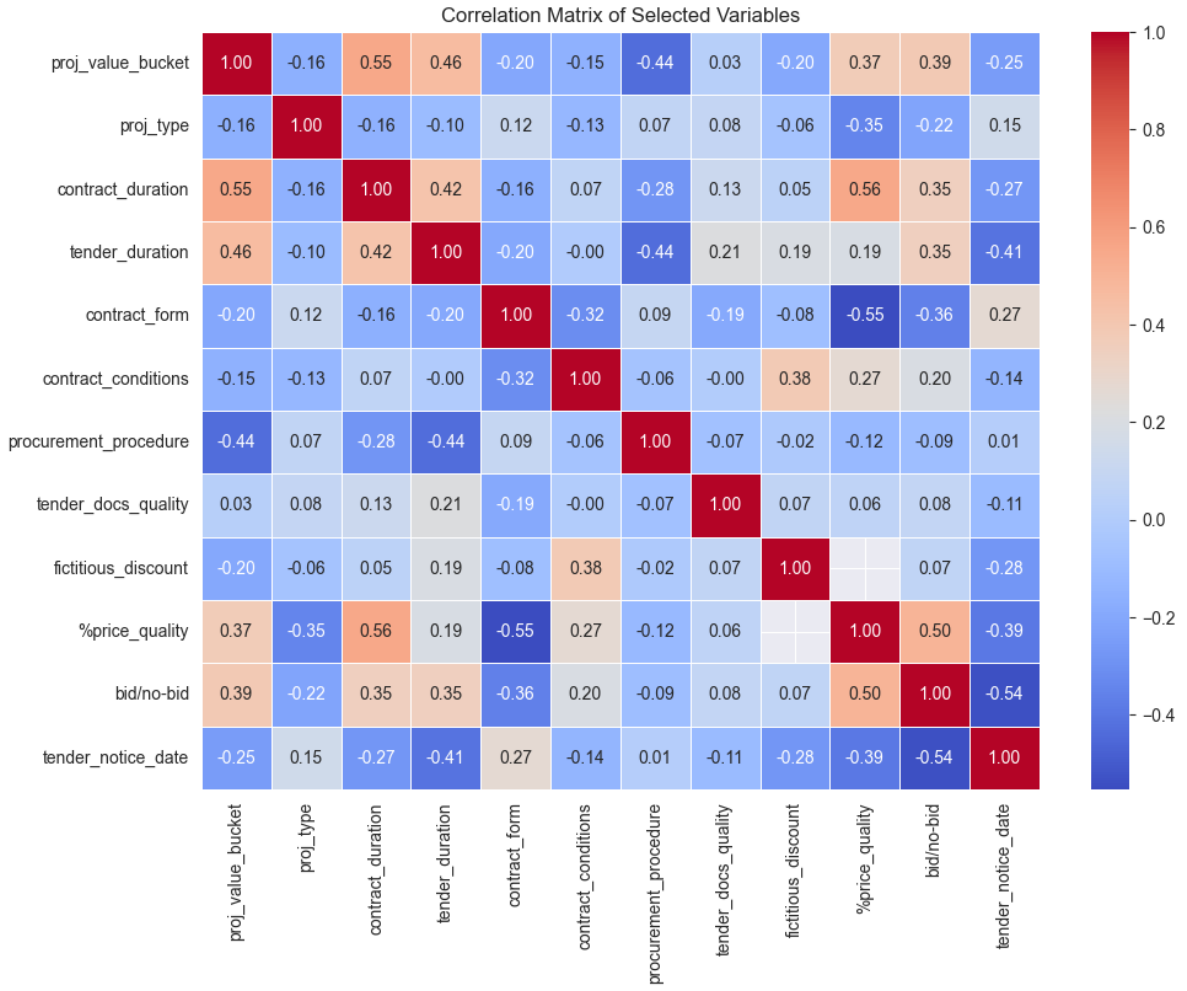


Figure C.2: Correlation matrix of selected variables.

D

Time-based validation results

This appendix complements Section 4.3 by providing the time-based validation results. First, multiple chronological splits are analysed (Appendix D.1). Second, a rolling (growing) split analysis is performed (Appendix D.2). Lastly, the Post-August 2024 subset is described and analysed (Appendix D.3). The results show that analyses, to predict the future with past data, is unstable and therefore unsuitable for validation.

D.1. Chronological splits

The dataset is sorted chronologically by `tender_notice_date`. For each split, the model is trained on the earliest observations and tested on the most recent observations. The test sets contain very few ”Yes” cases (between 1-5), so the resulting metrics (especially accuracy) are sensitive to class imbalance and sampling noise. Table D.1 reports performance under several chronological hold-out splits to illustrate time-sensitivity of the results. This complements repeated cross-validation by showing how performance varies when training strictly precedes testing in time.

Table D.1: Chronological results across four splits.

Split	Model	N_{test}	Yes in test	Test Acc	Test AUC	Train Acc	Train AUC
50/50	Decision Tree	51	5	0.098	0.707	0.860	0.854
50/50	Logistic Regression	51	5	**0.255**	**0.826**	0.840	0.874
50/50	Random Forest	51	5	0.098	0.804	0.860	0.887
50/50	SVM (linear)	51	5	0.098	0.761	0.860	0.937
50/50	SVM (RBF)	51	5	0.098	0.765	0.860	0.924
50/50	XGBoost	51	5	0.098	0.500	0.860	0.500
60/40	Decision Tree	41	5	**0.512**	0.694	0.833	0.861
60/40	Logistic Regression	41	5	0.439	0.817	0.850	0.865
60/40	Random Forest	41	5	0.122	**0.844**	0.717	0.875
60/40	SVM (linear)	41	5	0.488	0.828	0.850	0.871
60/40	SVM (RBF)	41	5	0.415	**0.844**	0.817	0.848
60/40	XGBoost	41	5	0.122	0.500	0.717	0.500
70/30	Decision Tree	31	2	0.548	0.914	0.771	0.808
70/30	Logistic Regression	31	2	0.548	0.879	0.757	0.842
70/30	Random Forest	31	2	0.387	0.897	0.757	0.847
70/30	SVM (linear)	31	2	**0.677**	0.897	0.800	0.863
70/30	SVM (RBF)	31	2	0.516	**0.983**	0.757	0.804
70/30	XGBoost	31	2	0.065	0.810	0.657	0.700
80/20	Decision Tree	21	1	**0.809**	0.950	0.750	0.787

Continued on next page...

Table D.1 continued from previous page

Split	Model	N_{test}	Yes in test	Test Acc	Test AUC	Train Acc	Train AUC
80/20	Logistic Regression	21	1	0.667	0.950	0.787	0.847
80/20	Random Forest	21	1	0.524	0.900	0.762	0.860
80/20	SVM (linear)	21	1	0.714	0.850	0.838	0.870
80/20	SVM (RBF)	21	1	0.619	0.900	0.713	0.807
80/20	XGBoost	21	1	0.762	**0.975**	0.725	0.778

So when looking at the 60/40 split, we see that Logistic Regression has relatively high test accuracy and AUC. Table D.2 shows the confusion matrix for a representative 60/40 chronological split using Logistic Regression. The matrix is included to clarify the types of errors (false positives vs false negatives) made under this time-based evaluation. The training set contains way more ”bids” than the test set, meaning the model will also predict more ”bids”.

Table D.2: Confusion matrix for Logistic Regression 60/40 split.

	Predicted No	**Predicted Yes**
Actual No	TN = 13	FP = 23
Actual Yes	FN = 0	TP = 5

D.2. Rolling (growing) split

The rolling-origin (expanding window) evaluation trains on the earliest observations and tests on the next chronological block of tenders. Starting from an initial training set of 60 observations, the training set grows by 10 observations per run while the test window remains fixed at 10 observations, resulting in four test windows in total. Because each test block is small, some blocks contain only a single class; in those cases ROC AUC is undefined, which explains why the AUC statistics are based on three evaluable windows for several models.

Table D.3 shows that performance varies substantially across windows, which is consistent with the limited test sizes and class imbalance in the most recent observations. Based on the pooled out-of-sample selection rule used for this appendix, Logistic Regression is selected for the rolling split confusion matrix. Its confusion matrix (TN = 19, FP = 16, FN = 0, TP = 5) indicates that the model captures all positive cases in the rolling test windows, but does so with a relatively high number of false positives, implying a tendency to over-predict the positive class under this temporal evaluation setting.

Table D.3: Rolling-origin evaluation: mean (standard deviation) performance over the four test windows.

Model	**Acc (mean)**	**Acc (std)**	**AUC (mean)**	**AUC (std)**
Decision Tree	0.575	0.171	0.741	0.225
Logistic Regression	0.600	0.141	0.847	0.073
Random Forest	0.375	0.171	**0.884**	0.119
SVM (linear)	**0.650**	0.173	0.847	0.073
SVM (RBF)	0.500	0.115	0.852	0.170
XGBoost	0.500	0.356	0.759	0.251

Table D.4 reports the pooled confusion matrix for the selected best model in the rolling-origin evaluation. It provides an interpretable overview of aggregated error behaviour across all rolling test windows.

Table D.4: Confusion matrix for the selected best model in the rolling-origin evaluation (Logistic Regression).

	Predicted No	**Predicted Yes**
Actual No	TN = 19	FP = 16
Actual Yes	FN = 0	TP = 5

D.3. Post-August 2024 subset

The following analysis repeats the chronological hold-out validation on a subset of the data that only includes tenders after the Brainial migration (August 2024). The goal is to assess whether the temporal evaluation results change when focusing on the most recent period, which is less affected by the historic underrepresentation of “no-bid” decisions. As in the main time-based split, the data are sorted by `tender_notice_date`; models are trained on the earliest observations in the subset and tested on the most recent observations using four train/test ratios (50/50, 60/40, 70/30, and 80/20).

It should be noted that the post-August 2024 subset is smaller and remains strongly imbalanced in the test sets (only 1–3 positive cases per split). As a result, the reported metrics are sensitive to small changes in the test composition. The results are therefore included as a robustness check rather than as a primary basis for model selection. Something else to notice is that the XGBoost model does not train at all (Test AUC = 0.500). The reason is most likely the limited sized dataset. Even without training it does score the highest test accuracy, although the model only guesses ”no-bid”.

Table D.5 repeats the chronological hold-out evaluation for the post–August 2024 subset, which better reflects the most recent decision environment. It is included to check whether model behaviour changes in the more recent period.

Table D.5: Chronological hold-out results across four splits (post–August 2024 subset).

Split	Model	N_{test}	Yes in test	Test Acc	Test AUC	Train Acc	Train AUC
50/50	Decision Tree	32	3	0.656	0.810	0.750	0.818
50/50	Logistic Regression	32	3	0.812	0.851	0.875	0.936
50/50	Random Forest	32	3	**0.906**	**0.920**	0.688	0.934
50/50	SVM (linear)	32	3	0.812	0.862	0.875	0.964
50/50	SVM (RBF)	32	3	0.844	0.885	0.812	0.905
50/50	XGBoost	32	3	**0.906**	0.500	0.688	0.500
60/40	Decision Tree	26	2	0.846	**0.917**	0.763	0.842
60/40	Logistic Regression	26	2	0.769	0.875	0.816	0.919
60/40	Random Forest	26	2	**0.923**	0.833	0.711	0.926
60/40	SVM (linear)	26	2	0.808	0.875	0.842	0.936
60/40	SVM (RBF)	26	2	0.846	**0.917**	0.816	0.896
60/40	XGBoost	26	2	**0.923**	0.500	0.711	0.500
70/30	Decision Tree	20	1	0.850	0.763	0.818	0.875
70/30	Logistic Regression	20	1	0.750	0.947	0.795	0.930
70/30	Random Forest	20	1	**0.950**	0.947	0.727	0.922
70/30	SVM (linear)	20	1	0.750	0.947	0.864	0.943
70/30	SVM (RBF)	20	1	0.800	**1.000**	0.841	0.898
70/30	XGBoost	20	1	**0.950**	0.500	0.727	0.500
80/20	Decision Tree	13	1	0.846	0.917	0.824	0.875
80/20	Logistic Regression	13	1	0.846	0.917	0.765	0.904
80/20	Random Forest	13	1	**0.923**	**1.000**	0.765	0.904
80/20	SVM (linear)	13	1	0.846	0.833	0.843	0.932
80/20	SVM (RBF)	13	1	0.846	0.917	0.784	0.889
80/20	XGBoost	13	1	**0.923**	0.500	0.765	0.500

Table D.6 reports the confusion matrix for the selected best-performing model in the 60/40 chronological split on the post–August 2024 subset (Decision Tree). The model is trained on the earliest 60% of observations in this subset and evaluated on the most recent 40%. The confusion matrix summarises the distribution of correct and incorrect classifications in the test set by distinguishing between true negatives/positives and false negatives/positives.

Table D.6: Confusion matrix (counts) for the 60/40 chronological split on the post–August 2024 subset (Decision Tree).

	Predicted No	Predicted Yes
Actual No	TN = 20	FP = 4
Actual Yes	FN = 0	TP = 2

E

Model evaluation metrics and hyperparameters

E.1. Model evaluation metrics

The models will be validated through a 5-fold CV, of which the mean of those five folds will used as final results. This will be done ten times to give the final mean of fifty runs. The k-fold CV performance of the selected ML algorithms will be based on these five metrics, each of these metrics will be explained afterwards:

$$\textit{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{E.1}$$

$$\textit{Precision} = \frac{TP}{TP + FP} \tag{E.2}$$

$$\textit{Recall} = \frac{TP}{TP + FN} \tag{E.3}$$

$$\textit{F1} = \frac{2 \cdot \textit{Precision} \cdot \textit{Recall}}{\textit{Precision} + \textit{Recall}} = \frac{2TP}{2TP + FP + FN} \tag{E.4}$$

$$\textit{AUC} = \int_0^1 \frac{TP}{TP + FN}\, d\frac{FP}{FP + TN} = \int_0^1 \frac{TP}{P}\, d\frac{FP}{N} \tag{E.5}$$

where, TP $=$ true positive, TN $=$ true negative, FP $=$ false positive, and FN $=$ false negative.

Accuracy (6.1) is the percentage of correctly predicted observations among all observations. It provides the general performance measure, but it can mask whether the model mainly makes false positives or false negatives.

Precision (6.2) measures how reliable positive predictions are: it is the fraction all the positive prediction that is truly positive. A high precision indicates few false positives.

Recall (6.3) measures how many true positive cases are identified among all truly positive cases, it is the fraction that the model predicts as positive. A high recall indicates few false negatives.

F1-score (6.4) is the mean of precision and recall. It is high only when both precision and recall are high. Note: it is not truly the mean of the two; if one is lower, the F1-score will be penalised more. It can be seen as the mean when precision and recall are close to each other.

AUC (6.5) AUC is the Area Under the Curve, which plots the true positive rate against the false positive rate across all possible decision. Suppose you randomly pick one truly positive case and one truly negative case and compare their model scores. The AUC equals the probability that the model assigns a higher score to the positive case than to the negative case. It is a bit more forgiving than accuracy,

because a ”Bid” might score higher than a ”No-bid”, but still could get a score that is lower than the decision threshold and would therefore be predicted as a ”No-bid”.

E.2. Model optimisation

In this section, the model settings (hyperparameters) will be explained. Because the dataset is small, the goal is not to search for the best possible settings, but to use stable and conservative settings that reduce overfitting. This makes the comparison between models fair and clear. The setup is the same for each model, there will be no adjustments to class weights (more on this in Chapter 6), nor will be decision threshold be adjusted (default is 0.5). This means that the impact of false negatives (FN) and false positives (FP) are treated equally. In the bid/no-bid decision, a FN can result in missed opportunity costs, and FP can result in excessive business costs. However in this research, no monetary value will be added towards FN and FP. Again, more about this in Chapter 6.

Parameter settings and explanation

The final hyperparameter settings were selected with the aim of avoiding overfitting rather than maximising predictive performance. Given the limited dataset size, some of the chosen models were susceptive to overfitting. Therefore, each algorithm was first trained using standard, baseline settings. These baseline runs typically resulted in test-accuracy between 65-74% (with the exclusion of the Decision Tree model, which resulted in 60%). However, when looking at the training-accuracies, they ranged between 80–90%, which suggested that several models showed signs of overfitting.

For this reason, the models were restricted by limiting their training complexity (e.g., shallower trees, or larger leaf sizes for tree-based models). The final settings were chosen such that the resulting gap between the training-accuracy and test-accuracy remained below 10%p; this threshold was chosen by the researcher. Since overfitting cannot be completely avoided (Ying, 2019), constant parameter adjusting was applied until desired results were reached (meaning: train_acc $-$ test_acc < 10).

Below (Listing E.1), the implementation of the models and their settings are given. The priority was not reaching the highest accuracy, but to find balance between avoiding overfitting while remaining decently accurate.

```
LR_model = LogisticRegression(max_iter=2000, C=0.1)
DT_model = DecisionTreeClassifier(random_state=19, max_depth=2, min_samples_leaf=10)
RF_model = RandomForestClassifier(n_estimators=300, max_depth=2, min_samples_leaf=10,
    random_state=19)
GB_model = XGBClassifier(objective="binary:logistic", n_estimators=500, max_depth=2,
    min_child_weight=7, random_state=19, eval_metric='logloss', subsample=0.8,
    colsample_bytree=0.8)
SVM_model = SVC(kernel="linear", C=0.1)
SVM_model_non = SVC(kernel="rbf", C=100, gamma="scale")
```

Listing E.1: Model creation

The hyperparameters in Listing E.1 were selected to control the models complexity in this small-sample setting. For the linear models (Logistic Regression and the SVMs), regularisation is governed by the parameter C (Scikit-learn, n.d.-a). In the LR and the linear SVM, smaller values of C correspond to stronger regularisation, which shrinks model coefficients and reduces the risk of overfitting (Scikit-learn, n.d.-d). Therefore, C=0.1 was chosen to encourage a relatively simple decision boundary. However, during testing the non-linear SVM was underperforming, so a larger C-value was chosen in combination with gamma=”scale”. Gamma defines how much influence a single training example has. The larger gamma is, the closer other examples must be to be affected (Scikit-learn, n.d.-d).

For tree-based models, complexity is primarily controlled by restricting tree growth. The parameter max_depth sets the maximum depth of a tree; smaller depths prevent the model from learning highly specific rules that only fit the training data. In addition, min_samples_leaf specifies the minimum number of observations required in a leaf node. Larger values enforce more robust splits and reduce the chance of modelling noise (Scikit-learn, n.d.-b). Accordingly, conservative values were used for both the Decision Tree and Random Forest (max_depth=2 and min_samples_leaf=10). For the Random Forest, n_estimators=300 was selected to reduce variance by averaging across many trees, without increasing the complexity of individual trees (Scikit-learn, n.d.-c).

For XGBoost, a similarly conservative configuration was applied. The parameter max_depth=2 limits the complexity of each boosted tree. The parameter min_child_weight=7 acts as an additional regularisation control by preventing splits that are supported by only a small effective number of observations (Scikit-learn, n.d.-c). Furthermore, subsample=0.8 and colsample_bytree=0.8 were used to train each tree on a random 80% subset of observations and features, respectively, which helps mitigate overfitting and improves generalisation (xgboost developers, n.d.). Finally, n_estimators=500 specifies the number of boosting iterations; combined with the regularisation controls above, this provides sufficient learning capacity while keeping the model conservative.

F

Supplementary final results

F.1. Error behaviour

Table F.1 compares the average confusion matrices (mean ± standard deviation across folds) for four classifiers: a linear SVM, an RBF SVM, a decision tree, and XGBoost. Overall, the linear SVM and XGBoost achieve the strongest balance between correctly identifying negatives and positives, with the highest true negatives (TN = 38.9) and relatively low false positives (FP = 14.1). The RBF SVM shows a small drop in performance, producing fewer true negatives and true positives (TN = 36.0, TP = 31.2) alongside more false positives. The decision tree performs worst in terms of specificity, with substantially more false positives (FP = 21.7) and fewer true negatives (TN = 31.3), indicating a greater tendency to misclassify negatives as positives. The reported standard deviations reflect fold-to-fold variability; within each class, TN and FP (and likewise FN and TP) exhibit identical standard deviations because they are constrained to sum to the fixed number of samples in that class for each fold.

Table F.1: Confusion matrices (mean and std). Rows = true class (Neg, Pos), columns = predicted class (Neg, Pos).

Model	Mean CM	Std CM
SVM (linear)	$\begin{bmatrix} 38.9 & 14.1 \\ 15.0 & 33.0 \end{bmatrix}$	$\begin{bmatrix} 1.58 & 1.58 \\ 2.49 & 2.49 \end{bmatrix}$
SVM (RBF)	$\begin{bmatrix} 36.0 & 17.0 \\ 16.8 & 31.2 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 1.00 \\ 1.17 & 1.17 \end{bmatrix}$
Decision Tree	$\begin{bmatrix} 31.3 & 21.7 \\ 14.8 & 33.2 \end{bmatrix}$	$\begin{bmatrix} 3.13 & 3.13 \\ 3.97 & 3.97 \end{bmatrix}$
XGBoost	$\begin{bmatrix} 38.9 & 14.1 \\ 16.0 & 32.0 \end{bmatrix}$	$\begin{bmatrix} 1.04 & 1.04 \\ 1.18 & 1.18 \end{bmatrix}$

F.2. Model explainability summary

Table F.2 shows the feature importance of the Random Forest classifier model. Table F.3 reports permutation feature importance for the RBF SVM, quantified as the change in predictive performance when a single feature is randomly permuted while all others are kept fixed (mean and standard deviation over repeated permutations). Positive values indicate that permuting the feature degrades performance (i.e., the feature is informative), whereas values close to zero indicate limited contribution within the evaluated sample; small negative values may occur due to sampling noise. Table F.4 summarises the shallow decision tree (max_depth=2). Finally, Tables F.5 and F.6 provide complementary views of XGBoost feature relevance: gain reflects the average improvement in the training objective attributable to splits on a feature, while mean absolute SHAP values summarise the average magnitude of each feature's contribution to the model output across the evaluation data.

Table F.2: Random Forest feature importance based on mean absolute SHAP values.

Feature	Mean |SHAP|
tender_duration	0.055 475
contract_duration	0.028 485
contract_form_RAW/Bestek	0.026 806
proj_type_infrastructure	0.025 261
proj_type_modern cities	0.017 031
proj_value_bucket	0.016 770
contract_conditions_uav-gc	0.013 306
contract_conditions_uav	0.013 220
contract_form_d&c	0.012 771
procurement_procedure_open	0.009 690
tender_docs_quality	0.008 839
%price_quality	0.008 551
fictitious_discount_yes	0.002 981
procurement_procedure_restricted	0.000 968
contract_form_framework agreement	0.000 805
proj_type_marine	0.000 144
contract_form_dbfm(o)	0.000 000
contract_form_other	0.000 000
procurement_procedure_other	0.000 000

Table F.3: SVM (RBF) permutation feature importance (mean and standard deviation).

Feature	Mean	Std
contract_duration	0.1536	0.0419
tender_duration	0.1464	0.0927
procurement_procedure_other	0.0891	0.0421
contract_form_framework agreement	0.0836	0.0371
proj_type_infrastructure	0.0673	0.0299
%price_quality	0.0655	0.0709
tender_docs_quality	0.0594	0.0587
contract_form_dbfm(o)	0.0118	0.0504
procurement_procedure_restricted	−0.0009	0.0512
proj_type_modern cities	−0.0058	0.0092
contract_form_d&c	−0.0094	0.0213
contract_conditions_uav-gc	−0.0109	0.0282
contract_conditions_uav	−0.0136	0.0208
procurement_procedure_open	−0.0145	0.0265
proj_type_marine	−0.0148	0.0242
fictitious_discount_yes	−0.0279	0.0528
contract_form_RAW/Bestek	−0.0397	0.0234
proj_value_bucket	−0.0576	0.0539

Table F.4 shows impurity-based feature importance for the Decision Tree model. Although this method has known biases, it provides a simple baseline view of which splits drive the tree's decisions.

Table F.4: Decision tree impurity-based feature importance (non-zero only).

Feature	Importance
tender_duration	0.6886
proj_value_bucket	0.1870
contract_conditions_uav	0.1244

Table F.5 reports XGBoost feature importance based on gain. This highlights which predictors contribute most to reducing loss across the boosted trees.

Table F.5: XGBoost feature importance (gain; non-zero only).

Feature	Importance
tender_duration	1.5450
proj_type_infrastructure	0.8183
contract_duration	0.7274
contract_conditions_uav-gc	0.2982
fictitious_discount_yes	0.1821
proj_value_bucket	0.0315

Table F.6 complements Table C.5 by reporting XGBoost global importance using mean absolute SHAP values. This supports cross-method comparison by using the same SHAP framework applied to the Random Forest.

Table F.6: XGBoost global importance by mean absolute SHAP value (non-zero only).

Feature	Mean|SHAP|
proj_type_infrastructure	0.4609
tender_duration	0.4347
contract_duration	0.3314
contract_conditions_uav-gc	0.2141
fictitious_discount_yes	0.0626
proj_value_bucket	0.0149

G

Explanatory results of the shadow predictions

This appendix is supplementary to §6.2.1 about the recommendation for Count & cooper. Table G.1 summarises the explanatory results of the five-project shadow prediction. For each project, the “Items” column lists the reported rows, and the model columns provide the corresponding values. For Logistic Regression and linear SVM, “Score” is the summed linear predictor (feature contributions plus intercept): positive implies a bid prediction and negative a no-bid prediction under the default threshold. For Random Forest, a single linear score is not defined, so Score and Intercept are shown as “–”; instead, the column reports the per-project importance values for the listed variables.

Table G.1: Shadow prediction explanations per project.

Project	True	Items	RF	LR	SVM
Project 1	T	Prediction	**T**	F	**T**
		Score	–	-0.072	0.056
		Intercept	–	-2.329	-2.683
		tender_duration	32.885	0.695	0.580
		contract_duration	0.433	0.568	0.917
		proj_value_bucket	0.220	0.668	0.536
		tender_docs_quality	0.112	0.129	0.482
		procurement_procedure_restricted	0.079	0.333	0.437
		%price_quality	0.037	0.059	0.109
		contract_form_other	–	-0.195	-0.323
Project 2	F	Prediction	**F**	**F**	**F**
		Score	–	-0.437	-0.117
		Intercept	–	-2.329	-2.683
		tender_duration	36.199	0.765	0.638
		contract_duration	0.325	0.426	0.688
		proj_value_bucket	0.110	0.334	0.268
		tender_docs_quality	0.149	0.172	0.643
		proj_type_infrastructure	0.079	0.333	0.437
		procurement_procedure_restricted	0.010	0.142	0.173
		contract_form_uav/raw	0.098	-0.222	-0.296
		contract_conditions_uav	0.043	-0.072	-0.013
		%price_quality	0.009	0.014	0.026

Continued on next page...

Project	True	Items	RF	LR	SVM
Project 3	T	Prediction	F	F	F
		Score	–	-0.752	-0.534
		Intercept	–	-2.329	-2.683
		tender_duration	19.884	0.420	0.351
		contract_duration	0.325	0.426	0.688
		proj_value_bucket	0.220	0.668	0.536
		tender_docs_quality	0.149	0.172	0.643
		proj_type_marine	0.000	0.134	0.054
		procurement_procedure_open	0.040	-0.196	-0.156
		contract_conditions_uav	0.043	-0.072	-0.013
		%price_quality	0.016	0.025	0.047
Project 4	F	Prediction	T	T	T
		Score	–	0.599	0.767
		Intercept	–	-2.329	-2.683
		tender_duration	45.886	0.970	0.809
		contract_duration	0.325	0.426	0.688
		proj_value_bucket	0.220	0.668	0.536
		tender_docs_quality	0.149	0.172	0.643
		proj_type_marine	0.000	0.134	0.054
		contract_form_d&c	0.063	0.288	0.447
		contract_conditions_uav-gc	0.039	0.220	0.313
		fictitious_discount_yes	0.004	0.051	-0.040
Project 5	F	Prediction	**F**	**F**	**F**
		Score	–	-0.630	-0.360
		Intercept	–	-2.329	-2.683
		tender_duration	17.335	0.366	0.306
		contract_duration	0.541	0.710	1.147
		proj_value_bucket	0.329	1.002	0.805
		tender_docs_quality	0.112	0.129	0.482
		proj_type_modern cities	0.060	-0.298	-0.291
		procurement_procedure_open	0.040	-0.196	-0.156
		contract_form_framework agreement	0.003	-0.000	-0.065
		contract_conditions_uav	0.043	-0.072	-0.013
		%price_quality	0.037	0.059	0.109