Document downloaded from:

http://hdl.handle.net/10251/190053

This paper must be cited as:

García-Rodríguez, MJ.; Rodríguez-Montequín, V.; Ballesteros-Pérez, P.; Love, PED.; Signor, R. (2022). Collusion detection in public procurement auctions with machine learning algorithms. Automation in Construction. 133:1-13. https://doi.org/10.1016/j.autcon.2021.104047

The final publication is available at

https://doi.org/10.1016/j.autcon.2021.104047

Copyright Elsevier

Additional Information

Collusion Detection in Public Procurement Auctions with Machine Learning Algorithms

Manuel J. García Rodríguez[1*], Vicente Rodríguez-Montequín[2], Pablo Ballesteros-Pérez[3], Peter E.D. Love[4], and Regis Signor[5]

[1*] *Corresponding author*: Project Engineering Area, University of Oviedo, 33012 Oviedo, Spain. manueljgarciar@gmail.com

[2] Project Engineering Area, University of Oviedo, 33012 Oviedo, Spain. montequi@uniovi.es

[3] Departamento de Proyectos de Ingeniería, Universitat Politècnica de València, 46022 Valencia, Spain. pabbalpe@dpi.upv.es

[4] School of Civil and Mechanical Engineering, Curtin University, GPO Box U1987, Perth, Western Australia 6845, Australia. p.love@curtin.edu.au

[5] Brazilian Federal Police, Rua Paschoal Apóstolo Pítsica, 4744, Florianópolis, Brazil. regis.rs@pf.gov.br

Collusion Detection in Public Procurement Auctions with Machine Learning Algorithms

Abstract

Collusion is an illegal practice by which some competing companies secretly agree on the prices (bids) they will submit to a future auction. Worldwide, collusion is a pervasive phenomenon in public sector procurement. It undermines the benefits of a competitive marketplace and wastes taxpayers' money. More often than not, contracting authorities cannot identify non-competitive bids and frequently award contracts at higher prices than they would have in collusion's absence. This paper tests the accuracy of eleven Machine Learning (ML) algorithms for detecting collusion using collusive datasets obtained from Brazil, Italy, Japan, Switzerland and the United States. While the use of ML in public procurement remains largely unexplored, its potential use to identify collusion are promising. ML algorithms are quite information-intensive (they need a substantial number of historical auctions to be calibrated), but they are also highly flexible tools, producing reasonable detection rates even with a minimal amount of information.

Keywords: Auction; collusion; contracting; construction; machine learning; procurement.

1. Introduction

Public procurement is a common form of public spending whose purpose is to provide works, goods or services to a purchasing entity [1]. Within the context of procuring capital works, companies compete to be awarded a contract to build, improve or maintain a capital asset. Such contracts can vary in nature and may require the construction of new civil (e.g., roads and bridges) and social (e.g., schools and hospitals) infrastructures, the modification of existing assets or require maintenance [2].

Public procurement can be an intensive and complex process and thus can consume significant resources. For example, the European Union spends around 16% of its Gross Domestic Product on public procurement [3]. Collusion in these auctions (also called bid-rigging) refers to various illegal agreements among competing firms that aim to increase their profit margins. These collusive practices usually take the form of coordinated (non-competitive) price increases that are set between the companies (commonly referred to as *cartels*) [4]. Collusion is a recurring problem confronting the public sector, particularly when procuring capital works, with some being the most expensive items to be acquired [4]. Criminal investigations are regularly initiated to combat collusive activity, but being able to prosecute and obtain a conviction is challenging [5].

A major issue that stymies public institutions (e.g. contracting authorities, police bodies, competition commissions and courts of justice) from obtaining a conviction is detecting and proving that collusion has occurred [6]. However, the secrecy surrounding illegal agreements between firms tends to be underpinned by a carefully coordinated and sophisticated strategy, which is difficult to expose. In stark contrast, procurement authorities adhere to transparent and relatively stable purchasing patterns whereby they reuse awarding procedures, purchase standard products, resort to similar service specifications and the like. The predictability of such procurement practices can facilitate illicit market sharing and coordinated action among collusive firms [7–9]. Against this contextual backdrop, it can be said that a reliable method for detecting the presence of collusion in public procurement auctions would significantly help procurement authorities and other institutions mitigate the adverse economic and social effects of collusion.

A plethora of models for detecting collusion has been propagated in the normative literature. In this paper, the most relevant models, which we will review, have proven to flag long term collusive patterns among bidding cartels [10]. They have also helped in discover how these cartels dissuade companies from submitting competitive bids in markets dominated by them [11,12]. However, while the models have been able to detect collusion, their accuracy often comes into question as to the data that underpins them can contain noise or insufficient detail. It is common, for example, for developed models to rely on information from the bidder's (private) costs structures and/or pre-tender cost estimates (PTE), though such information is generally confidential (and collusive firms are obviously not willing to share it) or simply does not exist [13,14].

Machine Learning (ML), a branch of artificial intelligence that focuses on building an application that can automatically learn and improve from experience, analyze and draw patterns of inference from auction information, even when it is scant (i.e., just the bid values and winning bidder from each auction) [15,16,17]. Yet, ML algorithms usually require a significant amount of reliable information obtained from previous auctions to calibrate them [16].

This paper aims to examine the ability of various ML algorithms to detect collusive auctions accurately. Each algorithm is tested under different conditions (e.g., with access to more or less information and with/without the input of Screening Variables, SV). We refer to SV as those statistical indices (directly calculated from the bid values) whose preprocessing may help ML algorithms to increase their level of detection [17].

To test the performance of ML algorithms, we will analyze six procurement datasets from five different countries (i.e., Brazil, Italy, Japan, Switzerland and the United States, US). Access to such auction data is generally unavailable to researchers as it is deemed sensitive (e.g., contract cost estimates) [18], but access and permission have been given for the collusion detection research presented in this paper. Thus, our research demonstrates that ML algorithms can detect collusion and produce representative performance results by applying them to a wide variety of datasets from different countries boasting different types of data. To the best of the authors' knowledge, this is the first time a transversal study of this nature has been undertaken in the domain of collusion detection.

The paper commences reviewing the literature and identifying the research gap to be examined (Section 2). Then, the procurement datasets, the screening variables, the ML algorithms being compared and the error metrics adopted are described (Section 3). We next summarize the major quantitative results of the experimental analysis for identifying collusive auctions (Section 4). This summary is followed by identifying the significance and contribution of our study (Section 5). Finally, we conclude this paper by explicitly identifying the limitations and avenues for future research (Section 6).

2. Literature Review

Many studies in auction theory have proven that bidders' cost structures strongly condition their competitive and/or collusive strategies [1,10,19–22]. McAfee and McMillan [23] were the first to analyze collusion in static bid rotation schemes when no compensation payments existed between cartel members. In McAfee and McMillan's [23] auction model, the awardee is independent of previous (past) auctions. Building on the work of McAfee and McMillian [23], Aoyagi [24] and Skrzypacz and Hopenhayn [25] extended their model by considering repeated collusion in dynamic bid rotation schemes.

Studies have also analyzed collusion's occurrence and effect in real procurement auctions [26,27]. However, empirical-based collusion detection models are limited. One of the first attempts to develop an empirically-based model was Porter and Zona's [19], who sought to measure the probability of a bidder winning when some observable cost factors are known. However, that model aimed not to determine collusion, *per se*, but rather to anticipate the range of prices of future (competitive) bids. Other empirical-based models have been proposed since the propagation of Porter and Zona's [19] work. We will now summarize the four most relevant models in the remainder of this section.

The first seminal model in collusion detection is also known as *econometric screening* and was proposed by Bajari and Ye [28]. This model attempts to anticipate how a standard (competitive) distribution of bids should look based on the participating bidders' cost parameters. Unfortunately, these cost parameters constitute private data, which is generally difficult to gather and often disclosed by the bidders themselves. As a result, most data needs to be directly inferred by industry experts, resulting in a loss in accuracy. Bajari and Ye's [28] model does flag systematic deviations from a reference scenario. In this instance, the industry experts have to anticipate the reference scenario as they can assume the bidders submitting competitive bids want to be awarded the contract and will not cooperate with the cartel.

Bajari and Ye [29] model was initially tested in highway repair contract auctions in the US Midwest in 1994-1998. It was implemented as a functional reduced-form of linear regression where additional pieces of information such as bidders' past bidding history and pre-tender cost estimates (PTE) were needed (besides bidders' financial data). As a result of including this additional information, Bajari and Ye [28] could make valid comparisons with the reference scenario. However, Bajari and Ye's [28] model also has some important limitations:

- over-reliance on the functional form chosen when implementing the regression analysis;
- high sensitivity to missing information; and
- it is easy to cheat when the cartel knows 'how' it works (e.g., coordinated cover bids).

Considering the limitations above, the most important is the need for detailed data from each bidder and auction. The absence of such data precludes the model from being applicable in real bidding contexts. Fortunately, since Bajary and Ye's [28] study, more public data is available on public contracts and competitors, which can be used in the near future to improve collusion detection with ML.

The second model we examine is developed by Ballesteros-Pérez *et al*. [29], which focuses on analyzing possible abnormal dispersions in the distribution of bids, assuming they follow a Uniform distribution. In essence, the Ballesteros-Pérez *et al*. [29] model is an approximated collusion detection method used in conjunction with other approaches. It uses a simplified order statistics approach where the bids absolute order of magnitude is neglected and only the relative distances are considered. This approach, of course, leaves the possibility of cheating the method by submitting cover bids that 'emulate' a uniformly distributed pattern, no matter they are still abnormally high on average.

The third model has been proposed by Signor *et al*. [30], which is a *Probabilistic method* [2,34]. Signor *et al*.'s [30] model analyses submitted bids at two levels. Firstly, it analyses whether the bids overall distribution conform to a reference scenario (e.g., a Lognormal distribution). Additionally, the location of this distribution (i.e., absolute order of magnitude of the bids) can be closely approximated by historical auctions whenever data about their pre-tender estimates (PTE) is available. Hence, the model scrutinizes the distance of submitted bids from the PTE.

Secondly, Signor *et al*.'s [30] probabilistic method analyze the lowest bid's dispersion by drawing on order statistics theory. Put simply, it compares the probability of the lowest bid (i.e., the theoretical winner) being materialized as if it had been generated from the same reference distribution of the previous step. Hence, in Signor *et al*.'s [30] method, the actual winning bid observed is compared against the lowest order statistic (i.e., the minimum draw of *n* artificially generated bids) from a calibrated reference distribution. If the statistical deviation is significant, we can be confident that such a bid is unlikely to be truly competitive. Thus, the probabilistic method is robust, but it has the limitation of being strongly dependent on the availability and reliability of a PTE for a number of previous honest auctions and the auction being tested.

Finally, the fourth model is that developed by Imhof [17,35]. This model has been the first to examine the application of ML to bidding and the detection of collusion by applying a small set of Screening Variables (SV) in a Swiss dataset of roads construction. We will use those SV and the same dataset in our study but assuming different levels of access to auction data.

Additionally, Imhof [17,35] utilized two ML algorithm types: (1) the Lasso regression and an 'Ensemble method' consisting of a weighted average of several algorithms; and (2) bagged regression trees, random forests, and neural networks. In this research, we will consider a wider range of algorithmic options and various datasets to understand better the conditions leading to SV and ML algorithms performing better.

3. Materials and Methods

This section describes the research methods adopted to detect collusion in auctions of public sector capital works. In Figure 1, we present a summary of the research process used in this study.

3.1. Datasets

To assess the collusive detection capabilities of ML algorithms under different conditions (e.g., countries, types of auctions, time period, and the availability of data per auction), we acquired six public procurement datasets. These datasets are derived from five countries covering periods between 1980 to 2013.

All datasets can be found in the *Supplementary file* attached to this paper so that others can replicate our results. A quantitative description of the datasets is presented in Table 1. At this juncture, no study that has examined collusion has had access to such an extensive dataset, which enables the suitability of ML to be explored as a detection approach.

It is worth noting that all six datasets have been investigated and/or provided by public institutions [e.g. *Swiss Competition Commission* (COMCO), *Brazilian Federal Police*, *Japanese Fair-Trade Commission* (JFTC) and two courts of justice from the USA and Italy].

Hence, we assume the data are reliable and trustworthy. While the datasets may contain minor contradictions, we are unable to judge the auctions' bidding consistency. Actually, the datasets' owners are also unable due to the secret nature of the agreements. For example, there are instances where an auction's winning bidder was classified as collusive while other higher (not awarded) bids were not. Clearly, in the context of capital works procurement, collusion generally involves being awarded contracts at a higher-than-usual price. In the example above, all bidders may have facilitated this outcome. However, we can only assume the awarded bidder was flagged with a consistent abnormal bidding pattern through a series of auctions. Thus, without criminal proof, other companion bidders might have avoided being flagged as collusive and consequently avoided conviction, or even being honest competitors unwittingly involved in a case of partial collusion.

Alternatively, these non-awarded bids may have been the result of estimation errors or were competitive bids with intentionally high mark-ups where evidence of coordinated action among bidders either did not exist or could not be determined. Coordinated action is a necessary condition for collusion to occur being the most difficult to prove. Despite some minor inconsistencies with the data, all auctions are treated being uniform in our study. Indeed, due to differing formats for collecting data the ability to ensure its calibration poses a challenge. However, it needs to be acknowledged this is the most comprehensive study undertaken to date that examines the detection of collusion in real-life auctions. We now proceed to briefly describe the datasets, whose main features are summarized in Table 1.

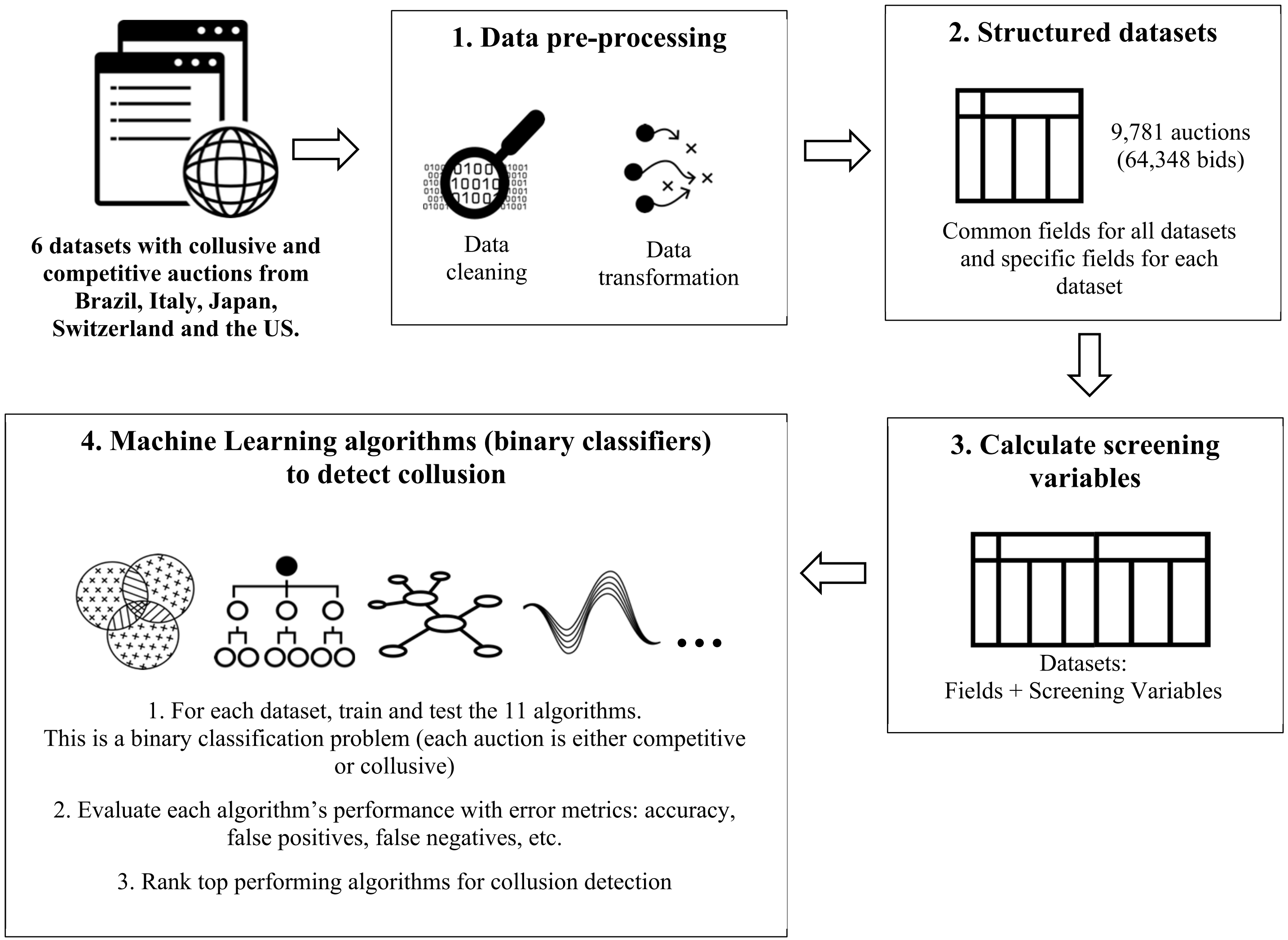


Figure 1. Flowchart summarizing the research approach for collusion detection

Table 1. Description of the collusive datasets

Topic	Description	Brazil	Italy	Japan	Swiss–Ticino	Swiss–SG&GR	US
General information	**Scope**	Oil infrastructure projects	Road construction	Building constr. and civil eng.	Road construction	Road construction and civil engineering	School milk market
	Time period	2002 - 2013	2000 - 2003	2003 - 2007	1999 - 2006	14 years (over 2005)	1980 - 1990
	Nº auctions	101	278	1,080	224	4,344	3,754
	Nº bids	683	20,286	13,515	1,629	21,231	7,004
	Awarding criteria	Lowest bid	Average Bid Method	Lowest bid	Lowest bid	Lowest bid	Lowest bid
	Avg. nº of bids per auction	6.76	72.97	12.51	7.27	4.89	1.91
Available information per dataset	**Common fields**	Auction code, bid values, winning bidder and number of bids per auction					
	Auction date	Yes	N/A	Yes	N/A	Yes	Yes
	Pre Tender Estimate (PTE)	Yes	Yes	Yes	N/A	N/A	N/A
	Identity of bidders	Yes. 272	Yes. 821	Yes. 1,665	N/A	N/A	Yes. 120
	Nº of different awardees	80 (29.41%)	19 (2.31%)	690 (41.44%)	N/A	N/A	91 (75.83%)
	Other fields (additional information)	Location and Brazilian State	Location, legal company type and economic size	Location	Consortium composition	Contract type	Inflation adjusted bid and inflation raw milk price adjusted bid
Collusive vs competitive data	**Collusive auctions**	N/A	N/A	N/A	184 (82.14%)	N/A	N/A
	Competitive auctions	N/A	N/A	N/A	40 (17.86%)	N/A	N/A
	Collusive bids	128 (18.74%)	8,085 (39.86%)	1,093 (8.09%)	1,332 (81.77%)	12,501 (58.88%)	866 (12.36%)
	Competitive bids	555 (81.26%)	12,201 (60.14%)	12,422 (91.91%)	297 (18.23%)	8,730 (41.12%)	6,138 (87.64%)
	Collusive bidders	47 (17.28%)	195 (23.75%)	230 (13.81%)	N/A	N/A	11 (9.17%)
	Competitive bidders	225 (82.72%)	626 (76.25%)	1,435 (86.19%)	N/A	N/A	109 (90.83%)
Bids per auction	**1 ≤ bids ≤ 4**	42 (41.58%)	0	0	29 (12.95%)	2,315 (53.29%)	3,727 (99.28%)
	5 ≤ bids ≤ 10	38 (37.62%)	5 (1.80%)	474 (43.89%)	171 (76.34%)	1,897 (43.67%)	27 (0.72%)
	11 ≥ bids	21 (20.79%)	273 (98.20%)	606 (56.11%)	24 (10.71%)	132 (3.04%)	0
Awarding price	**Aggregated total**	€12,170,309,780	€11,520,750,772	€402,195,427	€514,972,754	€2,136,031,656	N/A (Bid values are unit price per half a pint of milk)
	Aggregated collusive	€7,918,003,543 (65.06%)	€7,911,773,729 (68.67%)	€91,405,888 (22.73%)	€458,103,059 (88.96%)	€908,666,894 (42.54%)	
	Aggregated competitive	€4,252,306,237 (34.94%)	€3,608,977,044 (31.33%)	€310,789,539 (77.27%)	€56,869,695 (11.04%)	€1,227,364,760 (57.46%)	

Note: datasets used in this paper, apart from the Italian dataset, adopt the lowest bid wins awarding criterion

3.1.1 Brazil

Between 2002 and 2013, the Brazilian Oil Company Petrobras (a publicly traded, State-controlled company) was subjected to significant bid-rigging during the procurement of infrastructure projects. The dataset has been previously analyzed and made available by Signor *et al*. [18,30,33,34]. In 2014, a routine investigation by the *Brazilian Federal Police* into money laundering quickly turned into a very important anticorruption operation called "Operation Car Wash". Signor *et al*.'s [18,30,33,34] dataset form part of an ongoing investigation where several collusive companies confessed to price-fixing and bid-rigging. It was shown that 16 of the largest Brazilian construction companies (a cartel referred to as the "*Club of 16*") colluded in many of Petrobras's auctions.

3.1.2 Italy

The Italian dataset comprises road construction auctions from the municipality of Turin [36]. The legal office of Turin collected the dataset as part of a legal case against several firms accused of bid-rigging between 2000 and 2003. This dataset employs the Average Bid Auction (ABA) method: the awardee is the bid closest to a trimmed average [36]. The ABA can be used to create incentives to coordinate bids among bidders with the intention of manipulating the bids distribution. In 2008, the *Court of Justice of Turin* convicted 95 construction firms that operated in eight cartels that had been successfully awarded contracts (<10% of the firms won >80% of the auctions).

3.1.3 Japan

The Japanese dataset comprises building construction and civil engineering contracts from Okinawa. Initially, the data was published in Ishii [37], and it was later analyzed in Imhof [38]. The dataset was obtained from the Okinawa Prefectural Government (OPG), covering the period between 2003 and 2007. The construction market in Okinawa exhibits several features facilitating collusion: (1) geographic conditions (islands); (2) restricted invitation procedure (the buyer chooses those companies allowed to bid); and (3) contracts and bidders segmented into ranks. In June 2005, the *Japanese Fair-Trade Commission* (JFTC) filed a bid-rigging investigation against many firms involved in the auctioning process. The dataset covers three periods:

1. *Pre-inspection period*: auctions before the opening of the JFTC investigation (June 2005). These auctions can be collusive or competitive, according to JFTC resolutions.
2. *Post-inspection period*: auctions between the opening of the JFTC investigation (June 2005) and the amendment of Japanese competition laws in January 2006. These auctions are not used in our analysis as it was a transition period without information from the JFTC.
3. *Post-amendment period*: auctions after the amendment of Japanese competition laws. The JFTC sentenced and sanctioned the involved cartel participants at the beginning of the post-amendment period in March 2006. Therefore, all these auctions can be considered competitive as there has not been any proof of collusion ever since.

3.1.4 Swiss – Ticino

The Swiss dataset comprises road construction projects from the Canton of Ticino in Switzerland [35,39,40]. The cartel operating in this area of Switzerland had existed since the 50s, but it was not until the mid-90s that collusion became more frequent. By then, competition pressure within cartel companies started to grow, reaching its peak in 1998. This motivated cartel members to reach a tacit agreement in 1998 to which they adhered until 2005. During this period, all cartel firms in the road construction sector rigged nearly all procurement contracts. Therefore, this is undoubtedly one of the most severe bid-rigging cartels. As a result, local politicians went to the *Swiss Competition Commission* (COMCO) to investigate how awarding prices were exaggeratedly high in Ticino compared to other country regions.

3.1.5 Swiss – St Gallen and Graubünden

The next Swiss dataset covers the period between 2004 and 2010. It comprises the operations of two cartels specialized in road construction, asphalt paving, and civil engineering works in the Swiss cantons of St. Gallen and Graubünden [40]. In the first canton, eight firms participated in bid-rigging conspiracies. They met once or twice per month until 2009, when the *Swiss Competition Commission* (COMCO) launched house searches in the neighbor canton. In the second canton, another cartel was made up of a local trade association for road construction and asphalt paving operated until 2010. Both cartels were well organized and were awarded a very large share of auctions. As a result, the COMCO opened an investigation after the statistical anomalies identified in the procurement data until 2010.

3.1.6 United States

The US dataset was published in Porter and Zona [19] and also used in the study of Wachs and Kertész [41]. The dataset involves school milk procurement contracts in the State of Ohio between 1980 and 1990. School district officials independently solicited bids on annual supply contracts for milk and other products to regional milk producers (dairies). Typically, the lowest bidder was selected to supply milk in half pints to the schools during the following school year. In 1993 representatives of two dairies in Ohio confessed having bid-rigged these auctions during the 1980s. Thus, all bidding data were collected by the *United States District Court of Ohio* in 1994, and 30 dairies were charged with collusion. After careful analysis of these auctions, it was concluded that the estimated average effect of collusion on this market resulted in a 6.5% price increase. The dataset is non-construction-related, but it is useful to analyze it as it serves as a frame of reference to better understanding bidding behaviors and patterns in other markets.

3.2 Screening Variables

Screening Variables, or just *Screens*, are specific indices derived from each auction's bid values distribution (prices offered by bidders). These screens can help ML algorithms process auction information more efficiently to detect collusion [17]. However, there have been limited studies that have investigated the performance of different screens in collusive datasets.

Screens can be useful, not just for flagging possible collusion in a given auction but also for identifying sustained collusive patterns among specific bidders. Screens frequently consist of statistical indices calculated directly from the bid values of each auction (e.g. the bids standard deviation, skewness or kurtosis) or after removing or selecting some of the bids (e.g. the lowest and highest bid in an auction, or the lowest and second-lowest). They are generally easy to calculate and have proven to produce higher performance in ML algorithms. As a result, screens are usually beneficial when combined with ML algorithms and in our case, for detecting abnormally high bids.

The process to create a screen commences by letting t be the t-th auction in a dataset. We will not use an additional subscript to refer to each of the six datasets for the sake of clarity. Let sd_t be the (economic) bids standard deviation in auction t; $\bar{b}_t$ the mean (average) of all bids submitted to auction t; $b_{max,t}$ the maximum (most expensive) bid; $b_{min,t}$ the minimum (lowest, cheapest) bid; b_{2t} is the second-lowest bid; $sd_{losingbids,t}$ is the standard deviation of the non-awarded bids (all but the winning bid); n_t is the number of bids submitted to auction t; and b_{it}

is the *i-th* bid in auction *t* when ordered from lowest to highest. With this notation, the following screens are initially proposed to detect collusion better:

$$CV_t = \frac{sd_t}{\bar{b}_t} \quad \text{[Eq.1]}$$

$$SPD_t = \frac{b_{max,t} - b_{min,t}}{b_{min,t}} \quad \text{[Eq.2]}$$

$$DIFFP_t = \frac{b_{2t} - b_{min,t}}{b_{min,t}} \quad \text{[Eq.3]}$$

$$RD_t = \frac{b_{2t} - b_{min,t}}{sd_{losingbids,t}} \quad \text{[Eq.4]}$$

$$SKEW_t = \frac{n_t}{(n_t - 1)(n_t - 2)} \sum_{i=1}^{n_t} \left(\frac{b_{it} - \bar{b}_t}{sd_t}\right)^3 \quad \text{[Eq.5]}$$

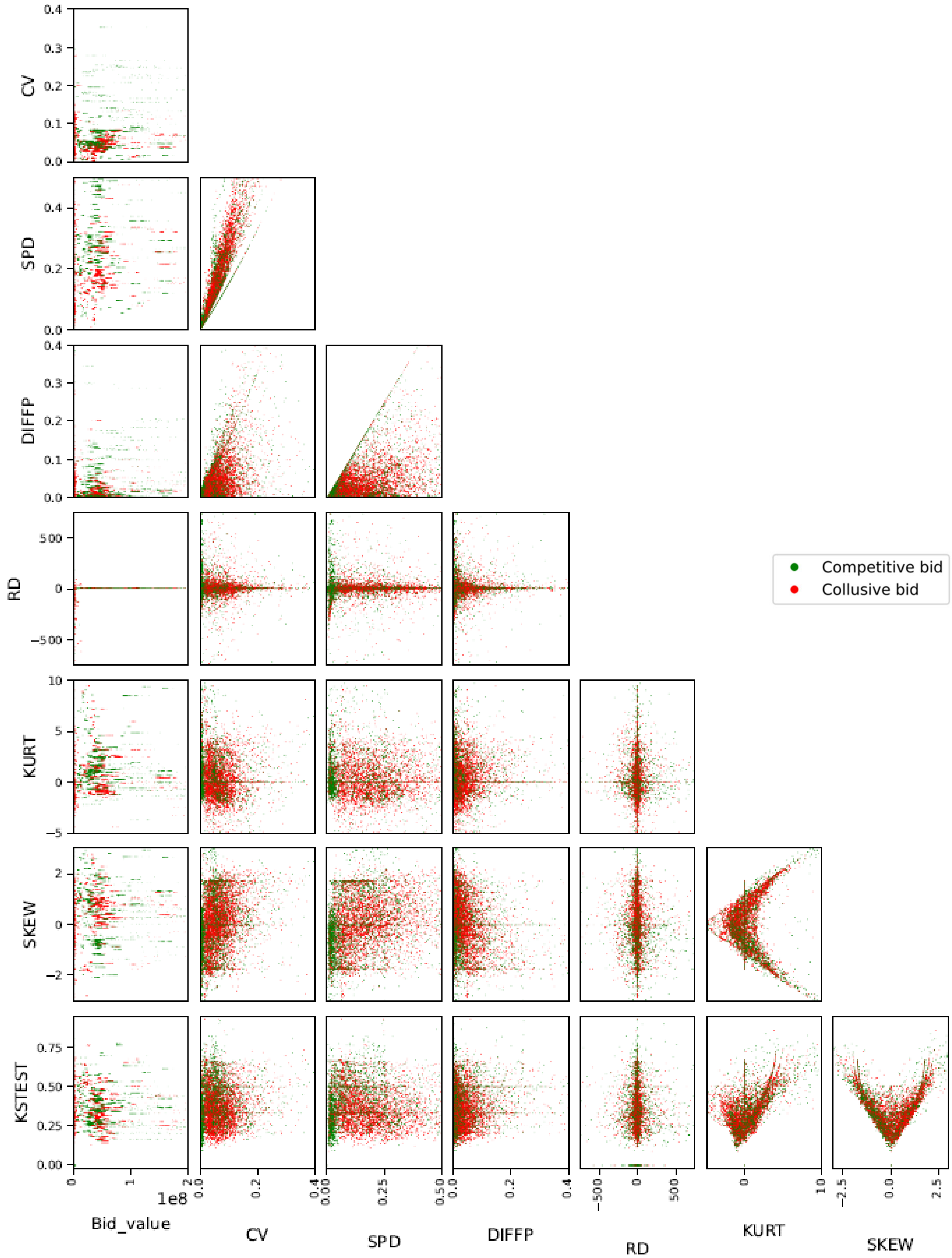
$$KURT_t = \frac{n_t(n_t + 1)}{(n_t - 1)(n_t - 2)(n_t - 3)} \sum_{i=1}^{n_t} \left(\frac{b_{it} - \bar{b}_t}{sd_t}\right)^4 - \frac{3(n_t - 1)^3}{(n_t - 2)(n_t - 3)} \quad \text{[Eq.6]}$$

$$KSTEST_t = max(D_t^+, D_t^-) \; with \; D_t^+ = max_i\left(\frac{b_{it}}{sd_t} - \frac{i_t}{n_t + 1}\right), D_t^- = max_i\left(\frac{i_t}{n_t + 1} - \frac{b_{it}}{sd_t}\right) \quad \text{[Eq.7]}$$

All previous screening variables have been proposed by different researchers in the context of collusion detection (e.g. [35,38–40,42]). The first screen is the *Coefficient of Variation* called CV_t (1), a scale-invariant statistic calculated as the ratio of the bids' standard deviation divided by the average of the bids. The second screen is the *Spread* (SPD_t) represented in Equation 2. Equation 3 measures the relative difference between the two lowest bids in the auction ($DIFFP_t$). An alternative screen to the latter is the *Relative Distance* (RD_t) which replaces the term in the denominator by the losing bids standard deviation (equation 4). Finally, the last three screens refer to the bid values' *Skewness* ($SKEW_t$), *Excess Kurtosis* ($KURT_t$) and *Kolmogorov-Smirnov test* ($KSTEST_t$). These three screens allow identifying possible bid distribution asymmetries (Equation 5), the condensation of bid values next to (or too far from) the average of the bids (Equation 6), and the similarity of the bid values for a uniform distribution (Equation 7), respectively. As the Excess Kurtosis requires at least four bids per auction to its calculation and our datasets contain a significant number of auctions with less than four bids (see Table 1), this screen will not be adopted in our study.

Other screening variables could have also been proposed, but a detailed exploration of their potential use remains outside the scope of this investigation. The ones used are the most common in other ML applications that work with statistically distributed values. Of note, it has been observed that the statistical distribution of bids is expected to become explicit when taking the *log* bids instead of their natural values (i.e., a lognormal distribution) [43,44]. In our experiments, we also tested the performance of these screens with log bids besides natural bid values. However, we found no improvement in the algorithms detection rates. Thus, a bids log transformation is not to be considered in this paper.

The *Scatter matrix* of the screening variables above (Eq. 1 to 7) for all the datasets (64,348 bids in total) is shown in Figure 2. This matrix is frequently generated in ML applications to identify correlations between the screening variables. It is also useful for detecting the screens that differentiate between competitive and collusive bids. However, we can see from Figure 2 that it does not show any distinct relationship between the space dispersion of competitive (green dots) versus collusive bids (red dots). That is, we cannot find separated clusters of red versus green dots in any subgraph of Figure 2. This finding indicates that we will need to rely on each algorithm's learning process (training) and performance (with and without the help of screens).



Coefficient of variation (CV), spread (SPD), two lowest bids differences in percentage (DIFFP), relative distance (RD), Skewness (SKEW), Excess Kurtosis (KURT) and Kolmogorov-Smirnov test (KSTEST).

Figure 2. Screening variables scatter matrix from all datasets

3.3 Machine Learning Algorithms Settings

The collusion detection capability of 11 algorithms is tested in this paper under different scenarios of information availability. It is assumed that each auction could be classified as either ‘collusive’ or ‘competitive’. Hence, the algorithms have to perform a binary classification for each auction *t*. The following algorithms are utilized to perform this task:

- Linear models: SGD (Stochastic Gradient Descent) [45];
- Ensemble methods: Extra Trees (Extremely Randomized Trees) [46], Random Forest [47], Ada Boost [48] and Gradient Boosting [49];
- Support Vector Machines: SVC (C-Support Vector Classification) [50];
- Nearest Neighbors: K Neighbors [51];
- Neural network models: MLP (Multi-Layer Perceptron) [52];
- Naive Bayes: Bernoulli Naive Bayes and Gaussian Naive Bayes [52]; and
- Gaussian Process [53].

Ensemble methods are the top-performing algorithms in our study as shown later. They combine several models (multiple learning algorithms) that produce a single optimal predictive model. This model is also generally more robust from the prediction point of view. Decision tree is usually one of those learning algorithms integrated in the Ensemble methods. This algorithm resembles a flowchart-like structure where each node implements a test on an attribute. Hence, each branch represents the outcome of a test, and each leaf node represents a class label. Two families of ensemble methods are usually distinguished:

- *Averaging methods*. The principle is to build several estimators independently and then to average their predictions. On average, the combined estimator is usually better than any of the single base estimator because its variance is reduced. For examples, Extra Trees and Random Forest.

- *Averaging methods;* they encompass several independent estimators and then average their predictions. On average, the combined estimator is usually better than any of the single base estimator because its variance is lower. Examples of averaging methods are *Extra Trees* and *Random Forest* algorithms.
- *Boosting methods;* their base estimators are implemented sequentially which reduces the bias of a combined estimator in some cases. Broadly speaking, the objective of Boosting methods is to combine several weak models to produce a single, more powerful model. *Ada Boost* and *Gradient Boosting* are some examples of Boosting methods.

These are common ML algorithms that have produced satisfactory results in many engineering applications, construction sector [54,55] and public procurement [56,57] included. All datasets and the algorithms' code can be found in the *Supplementary files* (csv format) we have provided. This will facilitate the future replicability of our results. The Python (3.0) programming language and the ML library *scikit-learn* have been used in this research [58]. Details about the eleven ML algorithms have not been provided but they are freely available from the *scikit-learn* library. However, we do provide some additional information at the end of this section about the numerical settings (parameter values) adopted for those algorithms that performed better. For those readers interested in extending their knowledge on the inner workings of each algorithm, we suggest resorting to the references provided in the list above and referring to the *Supplementary material* we have provided.

All the ML algorithms we have identified require calibration (training) before they are capable of differentiating collusive from competitive auctions. In conventional ML applications, training datasets typically comprise of thousands of entries. Algorithms generally use 80% of the data for training purposes and the remaining 20% to test their performance [59]. However, in our study and even though some of these datasets are large compared to most auction datasets reported in the construction bidding literature [60–62], many are too small to train all algorithms properly (i.e., they ‘only’ comprise 9,781 auctions with 64,348 bids).

To avoid collusion detection results being biased by the particular choice of training and test subsets, we performed 500 iterations with each algorithm. Thus, for each algorithm and dataset, we tested their detection performance while changing the specific subset of auctions used for training and testing (random choices). Noteworthy, the bids of each auction were either all used for training or testing; that is, they were not split for different purposes. This avoids the transfer of knowledge (rendering collusion detection harder for the algorithms, as they cannot use the same auction ID to flag an auction as collusive later), but provides a realistic scenario (as the bids of the same auction are generally known at once, not in different stages). Hence, our algorithms classify an auction as collusive or competitive based on each of the specific bids it contains. Markedly, all bids from the same auction were used as a single group of analyses.

The performance of the algorithms was analyzed under four different settings (scenarios). Each setting represents access to different pieces of data per auction. We named these pieces of information as *fields* in Table 1. Naturally, a higher amount of data per auction should lead to better collusion detection results. However, in actual practice, some data is not always available. Yet, it is equally valuable for anticipating the detection rates of each algorithm in the absence of data. Hence, the algorithms were trained and tested individually for each dataset under the following settings:

- *Setting 1 (all fields).* In this scenario, the algorithms used all the available data with one exception: the bidders’ identity (see Table 1 to identify the specific fields that were available in each dataset). The 'identity of bidders' was not used to avoid the potential risk of a bidder being easily catalogued upfront as collusive in the training process, and later classify as collusive almost all the auctions where it was involved (during the testing stage).

- *Setting 2 (all fields + screens).* Algorithms had the same data available as in setting 1 but with the assistance of the screening variables (CV, SPD, DIFFP, RD, SKEW and KSTEST). Theoretically, this should correspond to the scenario where ML algorithms perform better.
- *Setting 3 (common fields only).* In this scenario, the algorithms were only allowed to use the data shared among all datasets: that is, the auction code, bid values, winning bidder and number of bids per auction.
- *Setting 4 (common fields only + screens).* As in setting 2, this scenario assumed the data availability of setting 3 plus the aid of the screen variables described earlier.

Finally, we summarize the configuration adopted for the four ensemble methods as they were the top-performing algorithms in our study. A preliminary exploratory analysis was conducted to set the values of the algorithm parameters. Namely, we fine-tuned them based on data from related algorithm [35,38–40] and our first implementation results. With this, the best detection results were obtained for this parameters configuration:

- *Extra Trees and Random Forest*: The number of trees was 300; the function to measure the quality of a split was Gini; and the maximum depth of tree was until all leaves were pure or contained less than two samples.

- *Ada Boost*: The maximum number of estimators at which boosting terminated was 300. The base estimator was a decision tree classifier with 1 as the maximum depth of the tree with a learning rate also of 1.
- *Gradient Boosting*: The number of boosting stages to perform was 300; deviance was the loss function; and the learning rate was 0.1.

3.4 Error Metrics

To compare the performance of the proposed algorithms for classification problems, it is necessary to initially define some error metrics. The most common error metrics in ML are *accuracy*, *precision*, *recall*, *balanced accuracy* and *F1 score* [63]. Each metric was calculated in our research, though all of them are reported in the manuscript.

In this study, we are dealing with a binary classification performed at the auction level. This focus on auctions rather than bids was chosen to compare previous studies, which also classify auctions as collusive or not (as a full-colluded auction is more harmful than a small percentage of collusive bids among honest ones). However, as the algorithms must first analyze every bid, every auction will be classified as collusive or competitive. This classification depends on the ratio between its collusive and competitive bids. In our study, the minimum percentage of collusive bids to classify an auction as collusive was established as follows: Brazil (≥11%), Italy (≥44%), Japan (≥11.5%), Swiss – Ticino (≥10%), Swiss – SG&GR (≥10%), and US (≥10%). As stated earlier, most of these percentages correspond to those used by the courts of justice and/or researchers who published the datasets. We only increased the Italian percentage to present good results for two reasons: the average number of bids per auction was considerably high (72.92, which is about ten times higher than the average value of the other datasets), and it has a different awarding criterion (ABA). Overall, adhering to previous percentages of collusive versus competitive bids allows us to benchmark the improvement of detection rates against previous research.

Thus, let $\hat{y}_i$ be the predicted value of the $i-th$ sample $(1 \leq i \leq n)$, y_i is the corresponding true value, and L is the set of classes $(1 \leq l \leq L)$. In our case, L=2 has two possible classes: (1) collusive or (2) competitive bid. In this instance, the *accuracy* error metric is defined as the proportion of correct predictions over n samples and expressed as:

$$Accuracy = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} 1(\hat{y}_i = y_i) \qquad \text{[Eq.8]}$$

where $1(\hat{y}_i)$ is the indicator function. The equation returns 1 if the classes match and 0 otherwise.

Precision, also called positive predictive value, is intuitively the ability of the classifier not to label as positive (collusive bid) a sample that is negative (competitive bid). *Recall*, also called sensitivity or true positive rate, represents the ability of the classifier to find all positive samples. Let y_l be the subset of true values with class l, and $\hat{y}_l$ the subset of true predicted values in the same class l:

$$Precision_l = \frac{|y_l \cap \hat{y}_l|}{|y_l|} \qquad \text{[Eq.9]}$$

$$Recall_l = \frac{|y_l \cap \hat{y}_l|}{|\hat{y}_l|} \qquad \text{[Eq.10]}$$

The balanced accuracy avoids biased performance estimates in imbalanced datasets. Our collusion datasets are imbalanced as the number of competitive auctions in most datasets outnumber the number of collusive auctions (refer to Table 1 for the exact percentage of collusive and competitive bids in each dataset). This means, one of the two classes appears is more frequent than the other. Hence, the balanced accuracy can be defined as the average of the true positive rates (recall) of each class, that is:

$$Balanced\ Accuracy = \frac{1}{L}\sum_{l=1}^{L} recall_l \ = \frac{1}{L}\sum_{l=1}^{L}\frac{|y_l \cap \hat{y}_l|}{|\hat{y}_l|} \qquad [Eq.11]$$

Finally, the *F1 score* can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal and expressed as:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \qquad [Eq.12]$$

The aforementioned error metrics can be adapted to our specific problem. Our study involves a binary classification (two classes), thus a *True Positive (TP)* is a correctly identified collusive bid. Additionally, a *True Negative (TN)* is a competitive bid that has also been correctly identified. A *False Positive (FP)* implies the ML algorithm flags a bid as collusive even though it was competitive. Conversely, a *False Negative (FN)* implies that the method does not classify a bid as collusive when it is so. The *FP* and *FN* have worse consequences depending on the type of public institution being involved. From the perspective of police agencies and courts of justice, *FP* is the worst type of prediction error, as it could induce an unjustified investigation in a competitive (honest) bidder. From the perspective of contracting authorities, a high percentage of *FN* is worse as there are many collusive bidders that go unnoticed. Summarizing, we have TN = Correct (not collusion), FP = Unexpected collusion, FN = Missing collusion and TP = Correct (collusion), with:

$$TN+FP+FN+TP = \text{Total number of bids} \qquad [Eq.13]$$

Hence, the previous error metrics can be expressed into our binary classification problem as:

$$Accuracy = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} 1(\hat{y}_i = y_i) = \frac{TP + TN}{\mathrm{n}} \quad \text{[Eq.14]}$$

$$Precision = \frac{TP}{TP+FP} \quad \text{[Eq.15]}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{[Eq.16]}$$

$$Balanced\ Accuracy = \frac{1}{\mathrm{L}} \sum_{l=1}^{L} \frac{|y_l \cap \hat{y}_l|}{|\hat{y}_l|} = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \quad \text{[Eq.17]}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad \text{[Eq.18]}$$

Hence, the eleven ML algorithms were trained and tested to detect collusion in six datasets from five countries. As mentioned earlier, each algorithm was run 500 times while randomly changing the training subset (80%) and the test subset (20%) from each dataset. For each repetition (run), the previous error metrics were calculated and recorded. The error metric values reported below correspond to the average values obtained from those 500 repetitions.

4 Results

Table 2 shows four of the most relevant error metrics (accuracy, FP, FN and balanced accuracy) when each dataset is used independently to detect collusion under the setting 1 (all fields) and 2 (all fields + screens). Results from the other error metrics (precision, recall and F1 score) are included later and in our *Supplementary material*. Table 3 presents the same four error metrics but applying settings 3 (common fields) and 4 (common fields + screens). Additionally, and only because settings 3 and 4 share the same input parameters, it was also possible to aggregate all datasets and analyze them as a whole. These aggregated results are presented in the bottom rows of each error metric in Table 3 (values highlighted in bold).

Tables 2 and 3 show our major results - in facilitating the process of interpreting the results presented in Tables 2 and 3, we summarise key issues in Table 4. We also would like to point out that no single algorithm performs best in all datasets. Yet, we find the ensemble methods (Extra Trees, Random Forest, Ada Boost and Gradient Boosting) are generally better among the top performers.

The screens improve the accuracy of collusion detection and decrease the rate of false positives (FP) and false negatives (FN) in almost every situation. The screens are especially effective when used with the ensemble methods. This can be readily appreciated when comparing the results of ‘setting 2 versus setting 1’ (Table 2) and ‘setting 4 versus setting 3’ (Table 3). A simple summary of this increase can be seen in the central block of Table 4. For example, Setting 2 (all fields + screens) provides evidence of the best percentages of balanced accuracy. This was expected as this is the scenario where ML algorithms have access to more auction information. For the best four algorithms (the ensemble methods) in setting 2, it is possible to see that:

- accuracy is usually higher than 80%;
- FP and FN are generally lower than 10%; and
- balanced accuracy is usually higher than 70%.

Comparing the top-performing algorithms’ detection rates and results reported in the literature (bottom row of Table 4), we can see some of our algorithms have outperformed previous empirical models’. We also reveal that the US dataset was the most difficult for detecting collusion as it shows the worst percentages of balanced accuracy (about 60%) for almost all settings and algorithms. This may have arisen due to the dataset containing the lowest number of collusive bidders (11 bidders, 9.17% of the total). Similarly, the Swiss–SG&GR dataset had a low balanced accuracy (about 70%). This situation may have arisen due to the extremely high proportion of collusive versus competitive bids (59% vs 41%), rendering it difficult for the ML to differentiate between the varying bids. However, results are satisfactory when all the datasets are trained together (results in bold text in setting 4). The best algorithm, in this case, is the Extra Trees, which reaches a balanced accuracy of 86%. For this algorithm, the rate of FP is 8%, and the rate of FN is 6%.

The worst performing algorithms (SGD, SVC, K Neighbors, MLP, Bernoulli and Gaussian Naive Bayes and Gaussian Processes) hardly improve their detection results with the help of the screens. The implemented neural network algorithm (MLP, Multi-Layer Perceptron) has shown low percentages of balanced accuracy in all datasets and settings. Our MLP adopted four hidden layers with 240, 120, 70 and 35 neurons, respectively. However, a better combination of hidden layers and neurons might have reached better detection results. It should be acknowledged that combining hidden layers and neurons is an uphill task and is thus outside of the scope of this research.

Table 2. Average error metrics (accuracy, FP, FN and balanced accuracy) for each dataset in settings 1 (all fields) and 2 (all fields + screens).

		Algorithm																						
		SGD		Extra Trees		Random Forest		Ada Boost		Gradient Boosting		SVC		K Neighbors		MLP		Bernoulli Naive Bayes		Gaussian Naive Bayes		Gaussian Process		Colour legend
Error metrics	Setting / Dataset	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	
Accuracy (%)	Brazil	65.2	65.1	84.9	91.2	84.9	89.8	82.4	88.1	85.2	92.4	79.3	82.7	83.2	83.5	84.0	83.5	76.9	76.5	78.6	79.6	78.5	75.9	100%
	Italy	51.3	51.1	84.4	87.4	82.5	83.1	79.5	80.8	76.1	80.2	50.8	52.4	57.2	57.5	57.4	57.1	57.4	64.8	54.4	54.4	58.1	58.5	75%
	Japan	87.8	87.8	94.7	94.5	93.1	93.0	93.5	93.1	90.5	89.2	87.8	87.9	92.5	92.5	88.7	88.8	88.7	88.6	94.6	94.6	89.5	88.9	60%
	Swiss - Ticino	74.7	74.5	79.4	90.8	77.4	86.7	73.8	91.4	77.4	87.6	60.2	55.1	75.6	76.0	81.5	81.3	81.6	80.0	81.5	81.2	19.2	16.3	45%
	Swiss - SG&GR	68.5	68.9	83.4	85.3	82.7	84.7	84.1	85.0	78.6	74.2	50.0	49.0	77.8	77.6	80.1	80.2	80.2	80.1	75.1	41.4	20.0	20.1	30%
	US	70.3	72.6	84.1	84.8	83.5	83.9	83.0	82.4	77.1	76.1	46.3	45.4	79.4	79.4	82.2	82.3	82.3	77.9	81.8	79.1	73.6	75.2	0%
False positives (FP) (%)	Brazil	23.6	23.8	4.7	2.6	6.7	5.5	8.1	6.3	4.7	4.0	12.1	10.5	6.3	5.9	4.8	5.0	1.7	6.5	14.2	13.8	0.0	0.0	100%
	Italy	23.8	24.1	9.7	6.2	9.0	8.5	11.7	9.3	13.2	11.3	37.6	39.7	17.0	17.2	0.0	0.0	2.2	8.2	34.1	32.3	0.1	0.1	75%
	Japan	5.6	5.6	0.9	0.8	2.6	2.6	2.4	2.5	4.8	5.8	9.4	9.3	3.7	3.6	0.1	0.0	0.0	0.0	0.7	0.7	0.0	0.0	60%
	Swiss - Ticino	16.0	16.0	13.9	7.5	13.1	7.7	14.6	5.8	14.3	8.5	2.9	2.8	13.6	13.5	18.5	18.7	18.4	12.8	18.5	18.8	0.0	0.1	45%
	Swiss - SG&GR	15.1	15.5	9.2	9.7	9.3	9.0	9.4	8.8	9.9	10.9	7.5	7.4	17.9	18.3	19.9	19.8	19.8	19.4	15.8	5.4	0.0	0.0	15%
	US	15.4	12.5	4.0	1.7	4.3	3.1	2.3	3.9	12.6	13.2	48.1	48.9	4.0	3.7	0.0	0.0	0.0	8.9	3.8	8.0	11.8	10.1	0%
False negatives (FN) (%)	Brazil	11.2	11.1	10.4	6.3	8.5	4.7	9.5	5.7	10.1	3.6	8.6	6.8	10.5	10.6	11.2	11.5	21.4	17.0	7.1	6.5	0.0	24.1	100%
	Italy	24.9	24.9	6.0	6.4	8.5	8.4	8.7	9.9	10.7	8.5	11.6	7.9	25.8	25.3	42.6	42.9	40.3	27.1	11.5	13.3	41.8	41.5	75%
	Japan	6.6	6.6	4.4	4.7	4.3	4.5	4.1	4.4	4.7	5.0	2.8	2.7	3.8	3.9	11.3	11.2	11.3	11.4	4.7	4.7	10.5	11.1	60%
	Swiss - Ticino	9.3	9.4	6.7	1.7	9.4	5.5	11.6	2.9	8.3	3.8	36.9	42.1	10.8	10.6	0.0	0.0	0.0	7.2	0.0	0.0	80.8	83.6	45%
	Swiss - SG&GR	16.5	15.7	7.4	5.0	7.9	6.3	6.6	6.2	11.4	14.9	42.5	43.6	4.3	4.1	0.0	0.0	0.0	0.5	9.1	53.2	80.0	79.9	15%
	US	14.3	14.9	11.9	13.6	12.2	13.0	14.6	13.6	10.4	10.7	5.6	5.7	16.7	16.9	17.8	17.7	17.7	13.1	14.4	12.9	14.7	14.7	0%
Balanced accuracy (%)	Brazil	59.5	59.8	74.0	84.6	77.0	86.0	74.3	83.7	75.5	90.6	74.2	77.5	72.9	72.7	71.3	71.5	48.9	58.7	74.0	75.3	50.0	50.0	100%
	Italy	50.5	50.4	84.7	87.2	82.3	82.7	79.5	80.3	76.4	80.3	53.5	56.4	55.5	55.6	50.0	50.0	50.8	61.0	57.2	56.7	50.7	50.4	75%
	Japan	67.6	67.9	79.8	78.7	79.3	78.6	80.4	79.2	76.3	75.3	82.9	83.1	80.7	80.7	50.0	50.1	50.0	50.1	78.3	78.7	50.0	50.0	60%
	Swiss - Ticino	50.0	50.0	57.7	78.6	58.2	75.6	52.7	82.7	61.3	76.9	69.7	66.8	56.2	56.5	50.0	50.0	50.0	61.5	50.0	50.0	50.3	50.0	45%
	Swiss - SG&GR	51.8	51.4	72.4	72.6	71.5	73.4	72.3	74.0	67.6	63.6	54.6	54.4	52.0	51.6	50.0	50.0	50.0	50.9	54.5	53.3	50.0	50.0	30%
	US	50.7	50.5	64.2	60.7	63.1	61.6	57.3	59.2	62.9	61.6	55.0	54.9	50.5	50.3	50.0	50.0	50.0	57.6	57.4	58.6	50.4	51.4	0%

Table 3. Average error metrics (accuracy, FP, FN and balanced accuracy) for each dataset in settings 3 (common fields) and 4 (common fields + screens).

		Algorithm																						
		SGD		Extra Trees		Random Forest		Ada Boost		Gradient Boosting		SVC		K Neighbors		MLP		Bernoulli Naive Bayes		Gaussian Naive Bayes		Gaussian Process		Colour legend
Error metrics	**Setting / Dataset**	**3**	**4**	**3**	**4**	**3**	**4**	**3**	**4**	**3**	**4**	**3**	**4**	**3**	**4**	**3**	**4**	**3**	**4**	**3**	**4**	**3**	**4**	
Accuracy (%)	**Brazil**	65.4	64.8	87.8	89.6	86.7	89.1	87.9	86.5	85.6	89.3	84.2	80.6	84.8	85.1	86.3	85.5	81.1	77.3	56.0	46.5	81.2	80.5	100%
	Italy	51.3	50.7	78.9	86.8	79.9	81.9	77.3	79.5	74.7	72.4	54.5	50.8	56.6	56.5	57.7	57.0	57.4	65.0	53.8	53.4	57.5	60.5	75%
	Japan	83.9	83.7	94.5	94.5	93.2	93.4	93.3	92.3	90.7	87.9	85.5	82.5	92.3	92.4	88.2	88.7	88.8	88.8	94.0	94.3	88.7	88.9	60%
	Swiss - Ticino	73.8	73.3	78.1	90.9	77.0	86.9	73.7	91.4	74.4	90.3	53.6	55.1	76.0	75.6	81.8	81.9	81.9	79.9	82.0	81.4	18.9	18.0	45%
	Swiss - SG&GR	69.3	70.0	76.6	81.1	75.8	80.3	79.4	79.2	70.5	69.4	49.8	48.3	77.8	77.7	80.2	80.2	80.1	80.2	75.5	42.2	19.3	19.8	30%
	US	70.7	70.7	83.8	83.7	82.9	83.0	82.5	81.9	77.0	74.7	47.9	47.5	79.1	79.4	82.1	82.3	82.2	78.0	82.1	79.1	72.8	74.5	15%
	All datasets	**48.7**	**48.5**	**82.0**	**86.3**	**80.5**	**84.0**	**81.6**	**81.8**	**75.6**	**72.0**	**48.1**	**47.8**	**59.2**	**59.5**	**52.5**	**52.6**	**53.7**	**58.8**	**53.6**	**53.1**	**53.3**	**52.6**	0%
False positives (FP) (%)	**Brazil**	23.2	23.5	2.7	3.9	4.4	4.9	4.5	6.7	4.0	4.0	10.2	12.3	3.9	5.1	3.4	3.8	0.1	3.5	39.2	48.7	0.0	0.0	100%
	Italy	22.4	23.7	11.9	6.7	10.4	9.2	12.2	10.2	14.5	14.5	38.1	40.9	16.7	16.6	0.0	0.0	0.6	7.2	27.8	28.0	0.2	0.2	75%
	Japan	8.2	8.2	1.1	0.6	2.5	2.2	2.2	3.0	5.3	7.7	11.5	14.5	3.7	3.6	0.0	0.0	0.0	0.0	0.3	0.4	0.0	0.0	60%
	Swiss - Ticino	15.7	16.5	14.9	7.7	13.1	7.6	14.8	5.7	17.6	5.9	4.3	3.0	13.5	13.7	18.2	18.1	18.1	12.4	18.0	18.6	0.0	0.2	45%
	Swiss - SG&GR	15.2	15.7	15.2	16.4	15.1	14.6	17.9	15.6	12.6	12.5	7.4	7.3	17.9	18.1	19.8	19.8	19.9	19.3	16.0	5.6	0.0	0.0	30%
	US	14.8	15.1	3.8	1.7	4.6	3.3	2.2	4.1	12.2	14.2	45.7	47.1	3.9	3.8	0.0	0.0	0.0	8.8	3.8	8.0	11.5	10.9	15%
	All datasets	**25.2**	**24.8**	**9.7**	**8.0**	**9.7**	**8.7**	**10.2**	**9.9**	**9.7**	**9.6**	**42.3**	**45.8**	**18.5**	**18.6**	**22.7**	**21.8**	**0.0**	**24.4**	**10.5**	**3.9**	**1.3**	**1.4**	0%
False negatives (FN) (%)	**Brazil**	11.4	11.7	9.5	6.6	8.9	6.1	7.6	6.7	10.4	6.7	5.6	7.1	11.3	9.8	10.3	10.7	18.9	19.2	4.8	4.9	18.8	19.5	100%
	Italy	26.4	25.7	9.2	6.6	9.8	8.9	10.4	10.3	10.8	13.1	7.5	8.3	26.7	26.9	42.3	43.0	42.0	27.8	18.3	18.7	42.4	39.3	75%
	Japan	7.9	8.1	4.5	4.9	4.3	4.3	4.5	4.7	4.0	4.4	3.0	3.0	4.1	3.9	11.2	11.3	11.2	11.2	5.6	5.4	11.3	11.1	60%
	Swiss - Ticino	10.5	10.1	7.0	1.5	9.9	5.6	11.5	2.9	7.9	3.8	42.0	41.9	10.5	10.8	0.0	0.0	0.0	7.7	0.0	0.0	81.1	81.9	45%
	Swiss - SG&GR	15.5	14.3	8.2	2.5	9.1	5.1	2.7	5.2	16.9	18.1	42.9	44.3	4.3	4.1	0.0	0.0	0.0	0.5	8.5	52.2	80.6	80.2	30%
	US	14.5	14.2	12.4	14.6	12.5	13.8	15.3	14.1	10.8	11.1	6.4	5.3	16.9	16.8	17.9	17.7	17.8	13.2	14.0	12.9	15.6	14.6	15%
	All datasets	**26.1**	**26.7**	**8.2**	**5.7**	**9.8**	**7.3**	**8.2**	**8.3**	**14.7**	**18.4**	**9.6**	**6.4**	**22.3**	**21.9**	**24.8**	**25.6**	**46.3**	**16.7**	**35.9**	**43.0**	**45.6**	**46.0**	0%
Balanced accuracy (%)	**Brazil**	57.1	58.6	73.9	83.7	74.3	83.6	78.0	81.2	71.0	83.6	78.4	73.6	69.5	74.2	71.3	72.4	50.0	54.1	64.0	58.4	50.0	50.0	100%
	Italy	50.0	50.0	78.9	86.5	79.6	81.5	77.1	79.1	74.3	72.6	58.8	55.4	54.4	54.3	50.0	50.0	50.2	61.1	55.2	55.1	51.3	51.7	75%
	Japan	60.5	59.6	79.6	78.2	79.5	79.4	78.9	77.3	77.4	75.8	80.2	78.4	80.2	80.5	50.1	50.0	50.0	50.0	74.4	75.7	50.2	50.0	60%
	Swiss - Ticino	50.0	50.0	55.4	78.3	58.1	75.8	52.6	82.5	55.9	82.3	62.1	67.8	56.7	56.7	50.0	50.0	50.0	61.3	50.0	50.0	50.3	49.8	45%
	Swiss - SG&GR	51.8	51.5	56.5	56.9	56.2	59.9	53.1	57.3	57.1	57.5	54.6	53.7	51.9	51.6	50.0	50.0	50.0	50.9	54.3	53.4	50.0	50.0	30%
	US	50.3	50.5	62.7	58.0	62.0	59.2	55.6	58.0	62.0	59.9	54.1	56.3	50.2	50.3	50.0	50.0	50.0	57.6	57.9	58.7	50.3	50.8	15%
	All datasets	**48.4**	**48.1**	**82.1**	**86.4**	**80.4**	**84.0**	**81.7**	**81.8**	**75.1**	**70.8**	**49.7**	**50.7**	**58.7**	**59.1**	**53.0**	**53.0**	**50.0**	**59.3**	**51.6**	**50.1**	**49.3**	**49.1**	0%

Note: In this table, an extra row named "All datasets" is included as settings 3 and 4 only use fields shared among all datasets. Hence, it is possible to combine the auctions from all datasets into one.

Table 4. Summary of collusion detection average results with ML algorithms.

		Datasets						
Topic	**Description**	**Brazil**	**Italy**	**Japan**	**Swiss - Ticino**	**Swiss - SG&GR**	**US**	**All datasets**
Fields	**Common fields**	Auction code, bid values, winning bid and number of bids per auction						
	All fields in the dataset	Common fields, PTE, difference Bid/PTE, location, Brazilian State and date	Common fields, PTE, difference Bid/PTE, location, type and size of bidding companies	Common fields, PTE, difference Bid/PTE, location and date	Common fields and consortium composition	Common fields, contract type and date	Common fields, bid value with and without inflation and date	Common fields only
	Num. of variables	9	9	8	5	6	7	4
	Screens	Coefficient of variation (CV), spread (SPD), percentage difference between the two lowest bids (DIFFP), relative distance (RD), skewness statistic (SKEW) and Kolmogorov–Smirnov test (KSTEST)						
Results. Best accuracy and top-performing algorithm	**Setting 1** All fields from each dataset	85.2% Gradient Boosting	84.4% Extra Trees	94.7% Extra Trees	81.6% Bernoulli Naive Bayes	84.1% Ada Boost	84.1% Extra Trees	N/A
	Setting 2 All fields from each dataset + screens	92.4% Gradient Boosting	87.4% Extra Trees	94.6% Gaussian Naive Bayes	91.4% Ada Boost	85.3% Extra Trees	84.8% Extra Trees	N/A
	Setting 3 Common fields	87.9% Ada Boost	79.9% Random Forest	94.5% Extra Trees	82.0% Gaussian Naive Bayes	80.2% MLP	83.8% Extra Trees	82.0% Extra Trees
	Setting 4 Common fields + screens	89.6% Extra Trees	86.8% Extra Trees	94.5% Extra Trees	91.4% Ada Boost	81.1% Extra Trees	83.7% Extra Trees	86.3% Extra Trees
Average accuracy increase on including screens (for the four top-performing algorithms)	**Best algorithms**	Ensemble methods: Extra Trees, Random Forest, Ada Boost and Gradient Boosting						
	Setting 2 from 1	+6.0%	+2.3%	-0.5%	+12.1%	+0.1%	-0.1%	N/A
	Setting 4 from 3	+1.6%	+2.5%	-0.9%	+14.1%	+1.9%	-0.7%	+1.1%
Detection rates reported in the literature	**Paper/s**	[30,34]	[36]	[38]	[40]	[40]	[19]	N/A
	Method	Probabilistic methods	Standard hierarchical clustering algorithm	ML methods: Random Forest & Ensemble Method	ML method: Random Forest	ML method: Random Forest	N/A	N/A
	Accuracy	81% - 96%	N/A	88% - 93%	77% - 86%	61% - 84%	N/A	N/A

Finally, Figure 3 identifies three error metrics (precision, recall and F1 score) for settings 3 and 4 for the dataset called '*All datasets*' (auctions from all datasets merged into one). For each algorithm, the error metrics are denoted by a cross. The cutting point of the cross is the median of the precision and the recall. The endpoints of the cross are the minimum and maximum values of the precision and the recall (remember, we performed 500 iterations with each algorithm, so there are 500 values of precision and the other 500 values of recall). As a result, the precision, recall, and F1 score values remain inside the rectangle formed by the cross with a high degree of confidence. The algorithms with <50% of precision and <50% recall are not shown in the figure. By comparing setting 3 (left graph) with setting 4 (right graph), it is seen how the screens improve the precision slightly and recall for the four top-performing algorithms (ensemble methods). Summarizing the graphical results from setting 4, we observe:

- Extra Trees: 83%-86% precision, 86%-89% recall and 84%-87% F1 score.
- Random Forest: 80%-84% precision, 82%-86% recall and 81%-85% F1 score.
- Ada Boost: 78%-82% precision, 80%-84% recall and 79%-83% F1 score.
- Gradient Boosting: 73%-81% precision, <50%-76% recall and <78% F1 score.

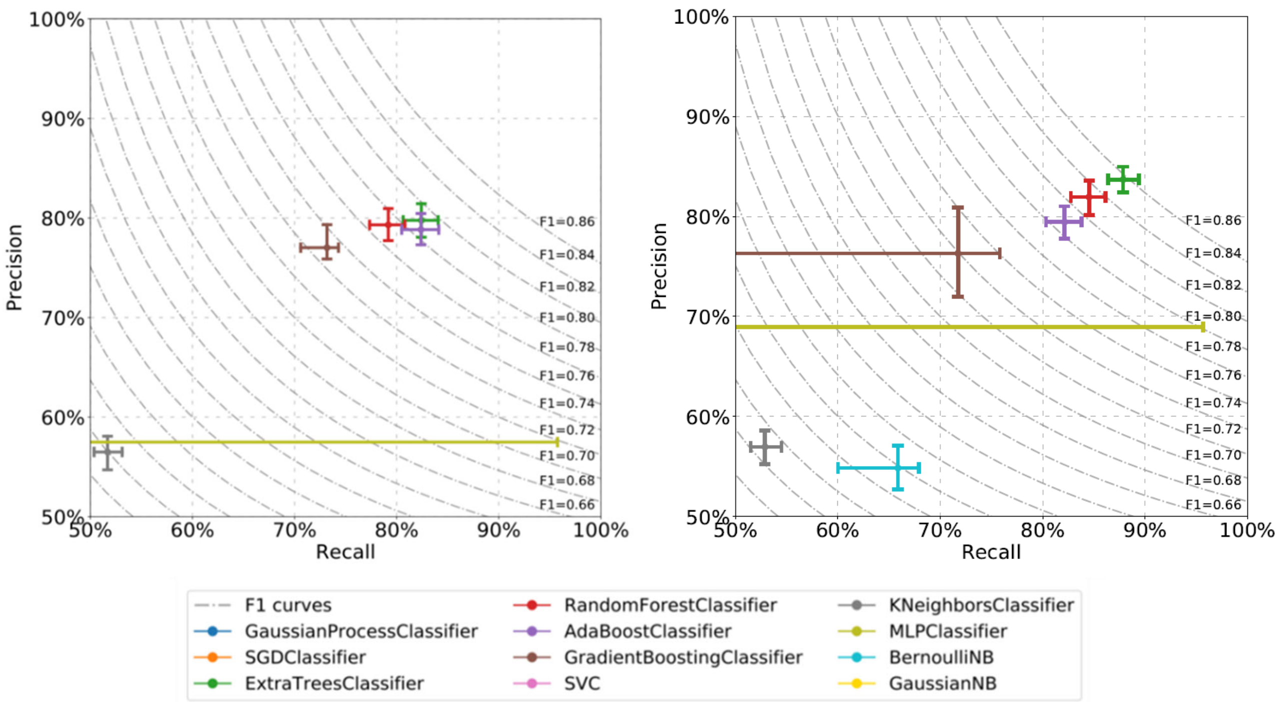


Figure 3. Error metrics (precision, recall and F1 score) for the '*All datasets*' combination in setting 3 (common fields) on the left and setting 4 (common fields + screens) on the right.

For additional detail with regard to the screens boxplot and the precision, recall and F1 scores of other settings and specific datasets, we refer the readers to our *Supplementary material*.

5 Discussion

Our research demonstrates that the amount of data available per auction is positively correlated with a higher collusion detection balanced accuracy in the majority of the tested ML algorithms. Yet, even with limited access to primary data, the ML algorithms were able to achieve satisfactory collusion detection rates. To this end, the research empirically demonstrates that ML tools can be implemented and be useful even when few pieces of information are available from a large number of auctions. In this case, this basic information was the bid values and the winning bid from each auction.

The eleven ML algorithms have been tested extensively with four different settings (input data configurations). They have been analyzed with standard error metrics for binary classification problems: accuracy, false positive, false negative, balanced accuracy, precision, recall and F1 score. The results from the previous section highlight that the four ensemble methods are the top-performing algorithms for the six collusive datasets. If the field ‘identity of bidders’ had also been considered in settings 1 and 2, the error metrics would have also significantly improved.

Yet, we have observed some minor differences in the screen’s effectiveness across datasets. In this regard, the US dataset (non-construction) and (but to a lesser extent) the Japanese dataset did not augment their average accuracy when screens were applied. Still, it is expected that screens in construction datasets will help boost collusion detection rates. Furthermore, there are no significant differences between the two awarding criteria (lowest bid versus the average bid method), at least not in accuracy for the top-performing algorithms or screens. Even though we only counted on a single dataset with different awarding criteria (the Italy dataset), hardly any differences have been found with other datasets results.

Another interesting analysis would involve training the algorithms in all but one country and then predicting collusion in the excluded country [38]. Basically, one could iteratively change the country excluded from the training data but later use it for testing purposes. This analysis would provide additional evidence on how well the methods work in terms of transferability across countries. Still, this would be a highly time-consuming, and it can only be implemented when all datasets share the same *fields*. Instead, we performed a similar analysis thanks to the so-called '*All datasets*' combination (combining the auctions from all datasets into one) with promising results. This combination was only possible for settings 3 and 4, though, as they were the only ones using shared information across all datasets.

6 Conclusions

Collusion has malevolent effects on public procurement, diminishes the confidence in a competitive market, and dissuades truly competitive competitors from submitting realistic bids. Research in collusion detection in construction has focused on producing both theoretical and empirical methods. However, theoretical models have been restricted to simple applications with few bidders and under the assumption of perfect information. In contrast, the accuracy of those of an empirical nature has come into question. Our research contributes to those based on empirical models and has used a comparison of ML algorithms to demonstrate their potential for improving the accuracy of detecting collusion.

The increasing availability of public procurement information and the recent development of ML techniques has made it much easier to develop alternative empirical models to detect collusion. While ML algorithms require large amounts of data for training, they can provide robust results with fewer input variables. Recognizing the potential of ML, we have compared the performance of eleven algorithms to detect collusion. We have provided evidence that these algorithms can work with a lot of limited pieces of information. We have also shown how detection rates can be improved with the help of some screening variables. The eleven ML algorithms were tested using an extensive dataset acquired from six public procurement datasets (a total of 9,781 auctions) from five countries: (1) Brazil; (2) Italy; (3) Japan; (4) Switzerland; and (5) the US.

Our analyses' three top-performing ML algorithms have been the Extra Trees, Random Forest and Ada Boost (ensemble methods). In the scenario where all auction information was available, these algorithms' accuracy (detection rates) ranged between 81% and 95%, with a balanced accuracy generally above 73% (excluding the US dataset). The algorithms can also be used with limited data, which poses a significant advantage over existing empirical methods. Once the algorithms are trained, they can be automatically updated with the latest auctions, and the user needs to make little effort in supervising their outcomes.

The research has limitations, which also need to be acknowledged. It is widely known that ML algorithms are akin to a black box from which it is difficult to explain the inherent complexity of the problem being analyzed (at least not in a straightforward manner). Moreover, they need a substantial amount of reliable historical data, some of which (especially the collusion-related) may not always be made available by competition commissions or law enforcement agencies – this problem is shared by other detection methods. Future research is needed to address the shortcomings of ML, specifically examining different algorithm types and fine-tuning their parameters. Access to data is critical for improving detection accuracy. A promising path for future research is to combine auction and company data (e.g., annual operating income, backlog, earnings before interest, taxes, depreciation, and amortization). By merging ML concepts with the economic theory first explored by Bajari and Ye [28] (driven by currently available data mining/scraping tools), we hope that the results will be even more accurate and their explanation better substantiated. Whereas the use of ML to detect collusion is in its infancy, we hope the research presented in this paper can foster future studies in this fertile and unexplored area.

Data Availability: All auction datasets (in csv format) and algorithms code (in Python) are included as a *Supplementary file.*

Acknowledgements: The authors are grateful to the Swiss Competition Commission (COMCO) and Dr David Imhof for their valuable comments and sharing some of the collusive datasets used in this paper.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] F. Curtis, P. Maines, Closed competitive bidding, Omega. 1 (1973) pp. 613–619. https://doi.org/10.1016/0305-0483(73)90049-2.

[2] R. Signor, P.E.D. Love, A. Oliveira, A.O. Lopes, P.S. Oliveira, Public Infrastructure Procurement: Detecting Collusion in Capped First-Priced Auctions, Journal of Infrastructure Systems. 26 (2020) pp. 05020002. https://doi.org/10.1061/(asce)is.1943-

555x.0000543.

[3] Public procurement contracts, European Parliament, Fact Sheets on the European Union. (2021). https://www.europarl.europa.eu/factsheets/en/sheet/34/public-procurement-contracts (accessed May 21, 2021).

[4] Algorithms and Collusion: Competition Policy in the Digital Age, OECD. (2017). www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.htm (accessed May 20, 2021).

[5] OECD, Fighting Bid Rigging in IMSS Procurement: Impact of OECD Recommendations, (2018) pp. 1–256. http://www.oecd.org/daf/competition/IMSS-procurement-impact-OECD-recommendations2018-ENG.pdf (accessed May 20, 2021).

[6] R.C. Marshall, L.M. Marx, The Vulnerability of Auctions to Bidder Collusion, Quarterly Journal of Economics. 124 (2009) pp. 883–910. https://doi.org/10.1162/qjec.2009.124.2.883.

[7] E.J. Anderson, T.D.H. Cau, Implicit collusion and individual market power in electricity markets, European Journal of Operational Research. 211 (2011) pp. 403–414. https://doi.org/10.1016/j.ejor.2010.12.016.

[8] R. Ishii, Collusion in Repeated Procurement Auction: A Study of a Paving Market in Japan, Discussion Paper No. 710, Institute of Social and Economic Research, Osaka University. (2008). https://doi.org/10.2139/ssrn.1148064.

[9] R. Ishii, Favor exchange in collusion: Empirical study of repeated procurement auctions in Japan, International Journal of Industrial Organization. 27 (2009) pp. 137–144. https://doi.org/10.1016/j.ijindorg.2008.05.006.

[10] R.H. Porter, J.D. Zona, Detection of Bid Rigging in Procurement Auctions, Journal of Political Economy. 101 (1993) pp. 518–538. https://doi.org/10.1086/261885.

[11] A. Blume, P. Heidhues, Modeling Tacit Collusion in Auctions, Journal of Institutional

and Theoretical Economics JITE. 164 (2008) pp. 163–184. https://doi.org/10.1628/093245608783742101.

[12] A. Hu, T. Offerman, S. Onderstal, Fighting collusion in auctions: An experimental investigation, International Journal of Industrial Organization. 29 (2011) pp. 84–96. https://doi.org/10.1016/j.ijindorg.2009.06.003.

[13] P. Bajari, G. Summers, Detecting collusion in procurement auctions, Antitrust Law Journal. 70 (2002) pp. 143–170. www.jstor.org/stable/40844085.

[14] K. Hendricks, R. Porter, G. Tan, Bidding rings and the winner's curse, The RAND Journal of Economics. 39 (2008) pp. 1018–1041. https://doi.org/10.1111/j.1756-2171.2008.00048.x.

[15] Y. Torres Berru, V.F. López Batista, P. Torres-Carrión, M.G. Jimenez, Artificial Intelligence Techniques to Detect and Prevent Corruption in Procurement: A Systematic Literature Review, in: Springer (Ed.), Communications in Computer and Information Science, 2020: pp. 254–268. https://doi.org/10.1007/978-3-030-42520-3_21.

[16] P. Razmi, M. Oloomi Buygi, M. Esmalifalak, A Machine Learning Approach for Collusion Detection in Electricity Markets Based on Nash Equilibrium Theory, Journal of Modern Power Systems and Clean Energy. 9 (2021) pp. 170–180. https://doi.org/10.35833/MPCE.2018.000566.

[17] M. Huber, D. Imhof, Machine learning with screens for detecting bid-rigging cartels, International Journal of Industrial Organization. 65 (2019) pp. 277–301. https://doi.org/10.1016/j.ijindorg.2019.04.002.

[18] R. Signor, P.E.D. Love, L.A. Ika, White Collar Crime: Unearthing Collusion in the Procurement of Infrastructure Projects, IEEE Transactions on Engineering Management. (2020) pp. 1–12. https://doi.org/10.1109/TEM.2020.2994636.

[19] R.H. Porter, J.D. Zona, Ohio School Milk Markets: An Analysis of Bidding, The RAND

Journal of Economics. 30 (1999) pp. 263-288. https://www.jstor.org/stable/2556080.

[20] L.H. Baldwin, R.C. Marshall, J. Richard, Bidder Collusion at Forest Service Timber Sales, Journal of Political Economy. 105 (1997) pp. 657–699. https://doi.org/10.1086/262089.

[21] E. Maskin, J. Riley, Asymmetric Auctions, Review of Economic Studies. 67 (2000) pp. 413–438. https://doi.org/10.1111/1467-937X.00137.

[22] M. Pesendorfer, A Study of Collusion in First-Price Auctions, Review of Economic Studies. 67 (2000) pp. 381–411. https://doi.org/10.1111/1467-937X.00136.

[23] R.P. McAffe, J. McMillan, Bidding Rings, The American Economic Review. 82 (1992) pp. 579–599. https://www.jstor.org/stable/2117323. (accessed May 20, 2021).

[24] M. Aoyagi, Bid rotation and collusion in repeated auctions, Journal of Economic Theory. 112 (2003) pp. 79–105. https://doi.org/10.1016/S0022-0531(03)00071-1.

[25] A. Skrzypacz, H. Hopenhayn, Tacit collusion in repeated auctions, Journal of Economic Theory. 114 (2004) pp. 153–169. https://doi.org/10.1016/S0022-0531(03)00128-5.

[26] J.E. Harrington, Detecting Cartels, Working Paper, No. 526, The Johns Hopkins University, Department of Economics. (2005). http://hdl.handle.net/10419/72037.

[27] J. Paha, Empirical methods in the analysis of collusion, Empirica. 38 (2011) pp. 389–415. https://doi.org/10.1007/s10663-010-9160-1.

[28] P. Bajari, L. Ye, Deciding Between Competition and Collusion, Review of Economics and Statistics. 85 (2003) pp. 971–989. https://doi.org/10.1162/003465303772815871.

[29] P. Ballesteros-Pérez, M. Skitmore, R. Das, M.L. del Campo-Hitschfeld, Quick Abnormal-Bid-Detection Method for Construction Contract Auctions, Journal of Construction Engineering and Management. 141 (2015) pp. 04015010. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000978.

[30] R. Signor, P. Ballesteros-Perez, P.E.D. Love, Collusion Detection in Infrastructure

Procurement: A Modified Order Statistic Method for Uncapped Auctions, IEEE Transactions on Engineering Management. (2021) pp. 1–14. https://doi.org/10.1109/TEM.2021.3049129.

[31] R. Signor, P.E.D. Love, O. Olatunji, Determining Overpricing in Brazilian Infrastructure Projects: A Forensic Approach, Journal of Construction Engineering and Management. 142 (2016) pp. 06016001. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001156.

[32] R. Signor, P.E.D. Love, O. Olatunji, F.F. Marchiori, W.G. Gripp, A probabilistic method for forensic cost estimating of infrastructure projects, Applied Mathematical Modelling. 40 (2016) pp. 9197–9211. https://doi.org/10.1016/j.apm.2016.05.025.

[33] R. Signor, P.E.D. Love, J.J.C.B. Vallim, A.B. Raupp, O. Olatunji, It is not collusion unless you get caught: the case of 'Operation Car Wash' and unearthing of a Cartel, Journal of Antitrust Enforcement. 7 (2019) pp. 177–202. https://doi.org/10.1093/jaenfo/jnz009.

[34] R. Signor, P.E.D. Love, A.T.N. Belarmino, O. Alfred Olatunji, Detection of Collusive Tenders in Infrastructure Projects: Learning from Operation Car Wash, Journal of Construction Engineering and Management. 146 (2020) pp. 05019015. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001737.

[35] D. Imhof, Y. Karagök, S. Rutz, Screening for Bid Rigging - Does it works?, Journal of Competition Law & Economics. 14 (2018) pp. 235–261. https://doi.org/10.1093/joclec/nhy006.

[36] T.G. Conley, F. Decarolis, Detecting Bidders Groups in Collusive Auctions, American Economic Journal: Microeconomics. 8 (2016) pp. 1–38. https://doi.org/10.1257/mic.20130254.

[37] R. Ishii, Bid Roundness Under Collusion in Japanese Procurement Auctions, Review of Industrial Organization. 44 (2014) pp. 241–254. https://doi.org/10.1007/s11151-013-

9408-6.

[38] M. Huber, D. Imhof, R. Ishii, Transnational machine learning with screens for flagging bid-rigging cartels, Working Papers SES 519, Faculty of Economics and Social Sciences, University of Fribourg. (2020). https://doc.rero.ch/record/329575/files/WP_SES_519.pdf. (accessed May 30, 2021).

[39] D. Imhof, Simple Statistical Screens to Detect Bid Rigging, Working Papers SES 484, Faculty of Economics and Social Sciences, University of Fribourg. (2017). http://doc.rero.ch/record/289133/files/WP_SES_484.pdf. (accessed May 30, 2021).

[40] H. Wallimann, D. Imhof, M. Huber, A Machine Learning Approach for Flagging Incomplete Bid-rigging Cartels, Working papers SES 513, Faculty of Economics and Social Sciences, University of Fribourg. (2020). http://doc.rero.ch/record/328358/files/WP_SES_513.pdf. (accessed May 30, 2021).

[41] J. Wachs, J. Kertész, A network approach to cartel detection in public auction markets, Scientific Reports 9, 10818 (2019). https://doi.org/10.1038/s41598-019-47198-1.

[42] D. Imhof, Empirical Methods for Detecting Bid-rigging Cartels, Université Bourgogne Franche-Comté, 2018. https://tel.archives-ouvertes.fr/tel-01963076. (accessed May 31, 2021).

[43] P. Ballesteros-Pérez, M. Skitmore, On the distribution of bids for construction contract auctions, Construction Management and Economics. 35 (2017) pp. 106–121. https://doi.org/10.1080/01446193.2016.1247972.

[44] P. Ballesteros-Pérez, M.C. González-Cruz, J.L. Fuentes-Bargues, M. Skitmore, Analysis of the distribution of the number of bidders in construction contract auctions, Construction Management and Economics. 33 (2015) pp. 752–770. https://doi.org/10.1080/01446193.2015.1090008.

[45] T. Zhang, Solving large scale linear prediction problems using stochastic gradient

descent algorithms, in: Twenty-First International Conference on Machine Learning - ICML ’04, ACM Press, New York, New York, USA, 2004: pp. 116. https://doi.org/10.1145/1015330.1015332.

[46] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Machine Learning. 63 (2006) pp. 3–42. https://doi.org/10.1007/s10994-006-6226-1.

[47] L. Breiman, Random forests, Machine Learning. 45 (2001) pp. 5–32. https://doi.org/10.1023/A:1010933404324.

[48] Y. Freund, R.E. Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, Journal of Computer and System Sciences. 55 (1997) pp. 119–139. https://doi.org/10.1006/jcss.1997.1504.

[49] J.H. Friedman, Greedy function approximation: A gradient boosting machine., The Annals of Statistics. 29 (2001) pp. 1189–1232. https://doi.org/10.1214/aos/1013203451.

[50] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning. 20 (1995) pp. 273–297. https://doi.org/10.1007/BF00994018.

[51] N.S. Altman, An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression, The American Statistician. 46 (1992) pp. 175–185. https://doi.org/10.1080/00031305.1992.10475879.

[52] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Second, Springer New York, New York, NY, 2009. ISBN 978-0-387-84858-7 https://doi.org/10.1007/978-0-387-84858-7.

[53] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006. ISBN: 978-0-262-18253-9. www.GaussianProcess.org/gpml.

[54] H. Anysz, A. Foremny, J. Kulejewski, Comparison of ANN Classifier to the Neuro-Fuzzy System for Collusion Detection in the Tender Procedures of Road Construction Sector, IOP Conference Series: Materials Science and Engineering. 471 (2019) pp.

112064. https://doi.org/10.1088/1757-899X/471/11/112064.

[55] H. Anysz, Ł. Brzozowski, Long Short-Term Memory (LSTM) neural networks in predicting fair price level in the road construction industry, IOP Conference Series: Materials Science and Engineering. 1015 (2021) pp. 012060. https://doi.org/10.1088/1757-899X/1015/1/012060.

[56] M.J. García Rodríguez, V. Rodríguez Montequín, F. Ortega Fernández, J.M. Villanueva Balsera, Public Procurement Announcements in Spain: Regulations, Data Analysis, and Award Price Estimator Using Machine Learning, Complexity. 2019 (2019). https://doi.org/10.1155/2019/2360610.

[57] M.J. García Rodríguez, V. Rodríguez Montequín, F. Ortega Fernández, J.M. Villanueva Balsera, Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain, Complexity. 2020 (2020) pp. 1–20. https://doi.org/10.1155/2020/8858258.

[58] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research. 12 (2011) pp. 2825–2830. http://arxiv.org/abs/1201.0490. (accessed June 5, 2021).

[59] A. Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 1st ed., O'Reilly, 2017. ISBN 978-1-491-96229-9.

[60] P. Ballesteros-Pérez, M. Skitmore, E. Pellicer, X. Zhang, Scoring Rules and Competitive Behavior in Best-Value Construction Auctions, Journal of Construction Engineering and Management. 142 (2016) pp. 04016035. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001144.

[61] P. Ballesteros-Pérez, M. Skitmore, E. Sanz-Ablanedo, P. Verhoeven, Forecasting the Number and Distribution of New Bidders for an Upcoming Construction Auction, Journal of Construction Engineering and Management. 145 (2019) pp. 04019056. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001694.

[62] P. Ballesteros-Pérez, M. Skitmore, E. Pellicer, M.C. González-Cruz, Scoring rules and abnormally low bids criteria in construction tenders: a taxonomic review, Construction Management and Economics. 33 (2015) pp. 259–278. https://doi.org/10.1080/01446193.2015.1059951.

[63] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Information Processing and Management. 45 (2009) pp. 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.