ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.2733

Machine Learning Classification of Public Tenders Using PCA: Facilitating SME Access to Chilean Procurement

Alejandro Caroca
Faculty of Engineering
Universidad Andrés Bello
Santiago, Chile
acaroca@unab.cl

David Ruete
Faculty of Engineering
Universidad Andrés Bello
Santiago, Chile
druete@unab.cl

Angélica Varas
Faculty of Engineering
Universidad Andrés Bello
Santiago, Chile
a.varas28@gmail.com

Francisco Marambio-Correa
Faculty of Engineering
Universidad Andrés Bello
Santiago, Chile
f.marambiocorrea@uandresbello.edu

Omar Salinas
Faculty of Engineering
Universidad Andrés Bello
Santiago, Chile
omar.salinas@unab.cl

Lilian San Martin Medina
Faculty of Engineering
Universidad Andrés Bello
Santiago, Chile
lsanmartin@unab.cl

Jean Maidana
Faculty of Engineering
Universidad Andrés Bello
Santiago, Chile
jean.maidana@unab.cl

Abstract*—Finding relevant procurement opportunities in Chile's Mercado Público platform is genuinely difficult for small suppliers. The platform processed over 2 million purchase orders in 2024, but its category taxonomy is inconsistently applied, descriptive fields vary in quality, and there is no automated filtering tailored to provider needs. This paper asks whether structured administrative variables alone, without any processing of tender text, can support accurate automatic classification of public tenders into procurement categories. We work with 204,770 tender records from 2024 and focus on the 10 most frequent combinations, covering 56,555 records. Principal Component Analysis (PCA) reduces 14 structured variables to 10 components retaining 89.82% of variance; six classifiers are then trained and compared: Random Forest, XGBoost, Decision Tree, Logistic Regression, SVM, and k-Nearest Neighbors. Random Forest reaches 91.17% accuracy and an F1-macro score of 0.9112. A single Decision Tree comes within 0.06 percentage points of that figure, which has practical implications for deployments where audit trails matter. The full pipeline was implemented independently in both R and Python, with under 1% accuracy difference between environments, confirming reproducibility. The gap between tree-based methods and Logistic Regression exceeds 51 percentage points, confirming that the classification boundaries are non-linear and cannot be captured by simpler models. These results suggest that structured-variable classification is a viable foundation for SME-facing tender recommendation systems in procurement platforms with heterogeneous or incomplete text data.

***Index Terms*—public procurement classification, supervised learning, principal component analysis, tender categorization, SME recommendation systems**

I. Introduction

Public procurement concentrates government demand and constitutes a strategic policy instrument to foster competition, innovation, and inclusion. As governments digitalize procurement end-to-end, procurement platforms increasingly generate high-volume administrative data that can support transparency, risk monitoring, and supplier-facing services, including data-driven discovery and recommendation tools [1], [2]. However, the scale and heterogeneity of procurement opportunities also create substantial information frictions for firms, especially for small and medium-sized enterprises (SMEs) with limited administrative capacity [3].

In Chile, state-based purchasing is primarily channelled through *Mercado Público*, the national transactional platform administered by ChileCompra. ChileCompra reported that in 2024 the platform reached USD17.643billion in transacted value and generated 2,031,670 purchase orders; it also reported 110,255 suppliers participating (either receiving purchase orders and/or submitting bids) [4], [5]. Mercado Público is described as being used by more than 850 public buying organizations nationwide [6]. This scale covers purchases ranging from routine goods (e.g., office supplies and medical consumables) to complex contracts in infrastructure, technology, and specialized services [6].

SMEs are central to Chile's productive structure. The Ministry of Economy's *Seventh Longitudinal Enterprise Survey* (ELE-7) reports that micro, small, and medium-sized firms represent 97% of formal enterprises [7]. Complementary official synthesis reports indicate that MiPyMEs account for 98.6% of firms and approximately 65.3% of formal employment [8]. In public procurement, ChileCompra reported that *Empresas de Menor Tamaño* (EMT) captured USD6.209billion in 2024 (35% of total transacted value) and represented 78,849 of the 110,255 participating suppliers (72%) [5]. These figures underscore both the relevance of SMEs in procurement and the importance of reducing the costs of finding and assessing opportunities.

A persistent barrier for SMEs is information overload: suppliers must continuously screen large lists of heterogeneous tenders, interpret diverse descriptions under tight deadlines,

and filter out irrelevant notices [3], [9]. International practice shows that standardized, machine-actionable codification can mitigate these frictions by enabling consistent search, filtering, and analytics. In the European Union, contracting authorities describe the subject matter of notices using Common Procurement Vocabulary (CPV) codes, a single hierarchical classification system that facilitates structured retrieval and cross-issuer comparability [10]. In the United States, the North American Industry Classification System (NAICS) provides a uniform industry coding framework; its 2022 revision and SBA guidance illustrate how codes are operationally embedded in procurement and small-business policy mechanisms [11], [12], [13]. In Chile, ChileCompra provides open data and API access to structured tender attributes (e.g., buyer identifiers, geography, dates, and estimated amounts), but coding practices and descriptive fields remain heterogeneous, sustaining search and matching costs that automated classification can help reduce [14], [15].

Chile's recent regulatory and data-policy agenda further strengthens the motivation for scalable analytics. Law No. 21.634 modernizes the procurement framework to improve probity and transparency, increase SME participation, and expand the system's coverage [16], [17], [18]. In parallel, ChileCompra states that, from December 2024, procurement information must be available in open and reusable formats and that key process data are published under the Open Contracting Data Standard (OCDS) [14]. OCDS is designed to reduce ambiguity through a common data model, enabling more comparable and reusable contracting data across the contracting lifecycle [19], [20].

Against this backdrop, this paper develops a supervised multi-class classification pipeline to assign Chilean public tenders to categories in the existing `Rubro1+Rubro2` taxonomy using only structured variables from Mercado Público open data. The study deliberately focuses on the ten most frequent `Rubro1+Rubro2` combinations, which together cover approximately 27% of the 2024 corpus (56,555 of 204,770 records). This scope is aligned with the paper's central research question on SME-facing tender discovery: the selected categories concentrate the highest tender volumes that are commercially accessible to small and medium-sized suppliers, which is the population that the proposed system is designed to serve. Long-tail categories present distinct class-imbalance challenges and are identified as a separate future contribution rather than as a straightforward extension of the present pipeline. We adopt Principal Component Analysis (PCA) to obtain a compact representation of high-dimensional tabular features, supporting efficiency and interpretability in downstream learning [21]. We compare Random Forest (RF), Decision Tree (DT), XGBoost, Logistic Regression (LR), Support Vector Machines (SVM), and k-Nearest Neighbors (kNN), motivated by recent evidence that tree-based learners remain strong baselines for typical tabular prediction tasks [22], [23]. Within the top-10 scope described above, PCA with 10 components retains 89.82% of variance and tree-based classifiers achieve accuracies above 91%.

In this work, we make four concrete contributions. First, it provides an empirically grounded formulation of tender categorization as a structured-data, SME-relevant discovery problem in Chile's procurement context, aligned with recent regulatory and open-data developments [14], [16]. Second, it presents a reproducible PCA-based supervised learning pipeline that avoids dependence on natural language processing. Third, it reports a comparative evaluation of six widely used classifiers on a large real-world Chilean dataset, informed by best-available evidence on tabular learning [22], [23]. Fourth, it draws actionable implications for SME-oriented tender recommendation and prioritization services that reduce search costs and information frictions in Mercado Público [3], [9].

II. Related Work

A. International Public Procurement Classification Systems

A recurring determinant of search efficiency and cross-institutional comparability in public procurement is the presence of shared, machine-actionable classification systems. In the European Union, the *Common Procurement Vocabulary* (CPV) is explicitly designed to provide a single classification system for describing the subject of procurement contracts. The official CPV documentation explains that the main vocabulary follows a hierarchical tree structure with codes of up to nine digits, complemented by a supplementary vocabulary for qualitative descriptors, and that CPV use is mandatory in EU procurement notices [10]. The practical value of this standardization is reflected in procurement-facing research: CPV codes enable statistical reporting, search, and the creation of alerts that can be used by potential bidders, thereby reducing search frictions in high-volume tender portals [24]. At the same time, automated use of CPV is non-trivial: CPV encompasses thousands of categories, and manual assignment can be heterogeneous, motivating algorithmic support for coding and retrieval [24].

In the United States, the *North American Industry Classification System* (NAICS) provides standardized industry definitions that are reviewed and revised every five years to reflect changes in economic activity [11]. While NAICS originated for statistical classification, it is operationally embedded in federal procurement and small-business policy infrastructure. SBA guidance notes that systems such as the System for Award Management (SAM) and the Federal Procurement Data System (FPDS) rely on NAICS codes as used in SBA size standards, linking opportunity classification to programs that govern small-business eligibility [12]. The 2022 NAICS revision also illustrates how procurement-relevant taxonomies evolve: the SBA's rulemaking on adopting NAICS 2022 documents the introduction of many new or modified industries and explains the need to maintain consistency across procurement databases and related statistics [25].

Beyond regional systems, the *United Nations Standard Products and Services Code* (UNSPSC) is widely used as a global product-and-service taxonomy in procurement contexts. UNGM documentation describes UNSPSC as a global

classification system used to categorize suppliers' offerings and procurement opportunities, supporting supplier matching and tender alerts; it also documents the four-level hierarchy (Segment, Family, Class, Commodity) that underpins structured search and notification workflows [26]. Recent research has leveraged UNSPSC for automated tender classification at scale. Abdullahi et al. developed machine learning models to classify tender titles and descriptions into UNSPSC levels, demonstrating the role of standardized taxonomies as both operational tooling and a learning target for automated categorization [27]. Collectively, CPV, NAICS, and UNSPSC show that standardized coding can lower information costs, but also that taxonomy application itself can become a bottleneck when categories are numerous and assignment practices vary [10], [24], [26].

B. AI and Automation in Public Procurement Systems

Recent literature emphasizes that procurement digitalization increasingly enables analytics-driven services for both oversight and market access. OECD guidance on the digital transformation of public procurement highlights the role of advanced data use and digital tools in modernizing procurement processes and improving performance across the procurement lifecycle [1]. Guida et al. synthesize AI applications across the procurement process and characterize the area as still emerging, providing a research agenda spanning decision support, automation, and risk management [2].

Empirically, multiple 2023–2024 contributions demonstrate how procurement platforms and oversight bodies are beginning to operationalize ML-based tooling. Siciliani et al. propose an AI-based decision support system that integrates structured tender data with unstructured document content, combining information retrieval and analytics to support procurement organizations [28]. Nai et al. present a case study in which machine learning is used for anomaly-related tasks and for a recommender system that retrieves similar procurements and potential bidder companies based on procurement requirements [29]. From an oversight perspective, Salazar et al. report the development of VigIA, a deployed tool that uses machine learning models and risk indices to prioritize procurement processes for investigation; a key design theme is balancing predictive performance with usability and explainability under public-sector scrutiny [30]. These studies confirm that procurement data can be repurposed for practical decision support, though they also highlight the importance of deployability constraints, including data availability, model transparency, and computational cost [2], [30].

C. Prior Work on Tender Classification and Recommendation

A substantial portion of recent tender-classification research focuses on mapping procurement notices into standardized taxonomies using textual fields, particularly where the taxonomy supports downstream retrieval and analytics. Navas-Loro et al. formulate CPV assignment as a multi-label text classification problem for Spanish procurement descriptions, noting that CPV assignment is challenging due to the large number of categories and heterogeneous criteria used in practice [24]. Siciliani et al. address related challenges by proposing methods for the automatic generation of CPV codes, contributing evidence that algorithmic coding can support more consistent categorization across large tender corpora [31]. For UNSPSC, Abdullahi et al. report supervised models to classify tender titles and descriptions into UNSPSC levels, further showing that standardized codes enable both human search and ML-based automation [27].

Beyond taxonomy assignment, bidder-facing tender discovery has also motivated supervised classification and recommendation approaches. Figueroa-Gómez and Galpin argue that identifying tenders of interest at scale is costly when performed manually and compare classification models as a mechanism to reduce search burdens [9]. Related work also highlights the relative scarcity of bidder-side support compared with buyer-side analytics: Oussaleh Taoufik and Azmani propose a "smart sourcing" framework aimed at notifying enterprises with a selected list of contracts and eligibility signals, explicitly positioning the approach as provider-oriented rather than purely adjudicator-focused [32]. These findings align with broader evidence that SMEs face persistent barriers in public procurement and that reducing informational and administrative burdens remains a key lever for inclusion [3].

D. Machine Learning for Tabular Data and Dimensionality Reduction

When procurement datasets contain rich structured attributes (e.g., amounts, dates, geography, buyer identifiers, and process states), tender categorization can be framed as supervised learning on tabular data. Recent benchmarks continue to show that tree-based methods are difficult to beat on typical tabular tasks. Grinsztajn et al. provide evidence that tree-based models often outperform deep learning for common tabular regimes, motivating Random Forest and gradient-boosted trees as strong baselines [22]. Uddin and Lu report statistically significant performance advantages for tree-based algorithms over non-tree alternatives across tabular datasets, reinforcing the expectation that non-linear learners capture heterogeneous feature interactions effectively [23]. At the same time, public-sector deployment often demands transparent model behavior and limited feature reliance; Salazar et al. explicitly describe explainability-driven design choices in operational procurement oversight tools [30].

Dimensionality reduction is frequently used to improve efficiency and robustness in high-dimensional settings and to address feature collinearity. Heidary et al. evaluate machine learning performance in reduced-dimensional spaces and motivate dimensionality reduction as a route to lower computational overhead and, in some cases, improved generalization [21]. In procurement contexts where many structured variables are correlated, PCA can provide an orthogonal representation that supports stable training for distance-based and kernel methods, while simplifying downstream monitoring and sensitivity checks [21].

E. Research Gap

The 2022–2025 literature shows rapid progress in taxonomy-based tender classification using text (especially CPV and UNSPSC) and in deployed analytics for oversight and decision support. However, important gaps remain for Chilean tender discovery and SME access.

First, recent procurement ML research is dominated by EU and US contexts (CPV and NAICS) or by platforms with well-defined global taxonomies (UNSPSC), with comparatively little peer-reviewed evidence focused on Mercado Público-specific categorization and SME-facing discovery outcomes. Second, much of the taxonomy-assignment literature assumes rich textual content and language-specific NLP, while operational standards emphasize structured codes for search and alerts [24], [26]; there is a need to assess how far structured variables alone can support accurate tender categorization where descriptive fields are noisy and coding practices are uneven. Third, the literature increasingly recognizes the need for provider-oriented tools, but bidder-side mechanisms remain less explored than buyer-side analytics [3], [32], and evaluations should therefore account for SME constraints such as time-to-screen and the cost of false positives in recommendations. Fourth, deployed procurement tools emphasize explainability, data availability, and maintainability in the public sector [30], which means tender-classification research should foreground reproducible pipelines, interpretable representations, and scalable inference suitable for production environments.

III. Problem Formulation and Dataset

A. Institutional Context and Problem Statement

Public procurement in Chile centralizes through Mercado Público (ChileCompra), which published over 2 million purchase orders in 2024 totaling USD 17.6 billion. Despite this volume, SMEs face significant barriers: absence of standardized coding, information dispersion, and lack of personalized search results. The current system relies heavily on manual search and human interpretation of free-text tender descriptions, creating inefficiencies that are particularly burdensome for smaller companies with limited administrative capacity.

Unlike international systems employing mandatory standardized vocabularies (CPV in Europe, NAICS in the USA), Chilean tenders lack uniform hierarchical classification. While some organizations use internal categorizations based on the Rubro1 and Rubro2 fields, these follow no strict standard, resulting in semantic ambiguity and reduced filtering automation capability. This problem is compounded by data quality issues as shown in Table I.

The core challenge addressed by this research is: *Can supervised machine learning classify Chilean public tenders into relevant categories using only structured variables, achieving accuracy sufficient to support automated SME recommendation systems?*

TABLE I
Data Quality Assessment

Variable	% Inconsistency	Common Issue
Estimated Amount	28%	Null or outdated values
Rubro (Category)	42%	Generic categories
Location	35%	Missing commune data

B. Dataset Description

The dataset was obtained from ChileCompra's Open Data portal, containing 204,770 tender records from a 12-month period in 2024. Initial extraction included 105 variables covering procurement process details, bidding organization attributes, temporal information, and financial data.

To delimit the scope of the study and to align it with its central research question, we restrict the analysis to the ten most frequent `Rubro1+Rubro2` combinations by 2024 frequency, yielding 56,555 records (approximately 27% of the full 2024 corpus). The selection criterion is twofold: commercial relevance for SMEs (the categories that concentrate the highest volume of tenders accessible to smaller suppliers) and tender volume (ensuring sufficient training data for stable model estimation). We acknowledge explicitly that this scope excludes long-tail categories which may be relevant to specialized suppliers, and we treat the extension of the pipeline to those categories as a distinct future contribution that requires addressing class-imbalance challenges fundamentally different from those resolved by the present design. The qualified scope of all subsequent results and operational claims should therefore be understood as restricted to this top-10 segment of Chilean public procurement.

Table II shows the selected categories and their distribution.

TABLE II
Top 10 Tender Categories

Category (Rubro1 + Rubro2)	Records	%
Medical Equipment / Dental Supplies	7,318	12.94
Construction / Prefabricated Structures	6,787	12.00
Medical Equipment / Wound Care	6,488	11.47
Musical Instruments / Art Equipment	5,652	9.99
Pharmaceuticals / CNS Drugs	5,559	9.83
Office Equipment / Office Machinery	5,176	9.15
Medical Equipment / Surgical Products	5,173	9.15
Paper Products / Paper Products	4,864	8.60
Medical Equipment / Emergency Items	4,843	8.56
IT / Computer Equipment	4,695	8.30
Total	**56,555**	**100.00**

C. Selected Variables and Data Preprocessing

Fourteen structured variables were selected, deliberately excluding free-text descriptions: **economic** (estimated amount log-transformed, awarded amount); **temporal** (creation, publication, closing, and adjudication dates; derived day-difference

indicators); **institutional** (organization code and name, acquisition type, tender state and state code); and **geographic** (region, commune). Preprocessing addressed missing values (0.5%, imputed with median/mode), duplicates (removed), format inconsistencies (dates standardized, amounts normalized), and outliers (log transformation on monetary variables). The final dataset achieved 99.5% completeness; temporal variables were converted to numeric indicators and categorical variables label-encoded.

Figure 1 shows the estimated-amount distribution across the ten categories, confirming the variance that motivates log transformation.

IV. Proposed Methodology

A. Overview

The classification pipeline covers five main stages: data preprocessing and quality assurance; feature engineering and transformation; dimensionality reduction via PCA; model training and hyperparameter tuning; and evaluation and validation. Reproducibility, computational efficiency, and interpretability were prioritized throughout, given the constraints typical of public sector deployment.

The pipeline deliberately excludes free-text tender descriptions from the feature space, in contrast with NLP-driven approaches developed for CPV and UNSPSC classification [24], [27]. This decision responds to three operational considerations specific to the Chilean procurement context. First, the heterogeneous quality of free-text fields in Mercado Público, with a meaningful share of records exhibiting generic or inconsistent category descriptions, limits the reliability of text-based features without aggressive preprocessing. Second, structured-variable classification is reproducible across institutions without requiring specialized NLP infrastructure, which aligns with the deployment realities of public-sector platforms operating with limited technical resources. Third, an explicit assessment of how far structured variables alone can support accurate categorization is itself an informative result for procurement-platform design, separable from the question of whether NLP features could further improve performance. A hybrid pipeline combining the present structured-variable classifier with selective text features is identified as a natural extension in Section VII.

B. Data Preprocessing

Beyond the cleaning described in Section III, preprocessing applied five steps: numeric features were scaled with StandardScaler (zero mean, unit variance); categorical variables were label-encoded; temporal indicators were derived (publication-to-closing duration, season and quarter markers, 30/60/90-day moving windows); extreme monetary values were capped at the 99th percentile after log transformation; and variables with pairwise correlation above 0.95 were removed to improve PCA stability.

C. Principal Component Analysis

PCA converts the 14 preprocessed variables into uncorrelated components ranked by explained variance, solving the eigenvalue problem $\Sigma v_i = \lambda_i v_i$, where Σ is the feature covariance matrix. Components are retained until cumulative explained variance exceeds threshold τ:

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \geq \tau \tag{1}$$

Ten components were retained, explaining 89.82% of variance (Fig. 2). Beyond dimensionality reduction, PCA eliminates multicollinearity by construction, stabilizes estimates for linear methods, and removes minor components that tend to capture noise rather than signal.

D. Classification Algorithms and Implementation

Six algorithms were evaluated representing diverse paradigms. Table III summarises their configurations.

TABLE III
Classifier Configurations

Model	Key hyperparameters	Ref.
Random Forest	250 trees, min_split=5, gini	[36]
XGBoost	120 est., lr=0.1, depth=6	[35]
Decision Tree	min_split=5, gini	[37]
k-NN	k=5, inverse-dist., Euclidean	[38]
SVM	RBF, C=1.0, gamma=scale	[39]
Log. Regression	L2, max_iter=1000, lbfgs	[40]

All experiments were run in both Python (scikit-learn 1.2+, XGBoost 1.7+) and R (caret, kknn, factoextra) on a standard laptop (Intel Core i7, 16 GB RAM) to validate reproducibility. Training times ranged from 0.29 s (Decision Tree) to 45.38 s (SVM), all practical for operational deployment.

V. Experimental Setup

The dataset was partitioned using a stratified 80/20 split applied after PCA transformation to prevent data leakage, yielding 45,244 training and 11,311 test records with proportional representation of all ten categories. We report accuracy, macro- and weighted-precision, recall, and F1-score; macro-averaging treats all classes equally regardless of frequency, while weighted averaging reflects overall system accuracy under class imbalance. Complementary 5-fold cross-validation on the training set confirmed model stability, with top-performing models showing consistency within 2% of hold-out accuracy. Reproducibility was ensured by fixing random seeds (seed=42), version-controlling all pipeline scripts, and validating that R and Python implementations produced accuracy differences below 1% under identical PCA scaling and component selection.

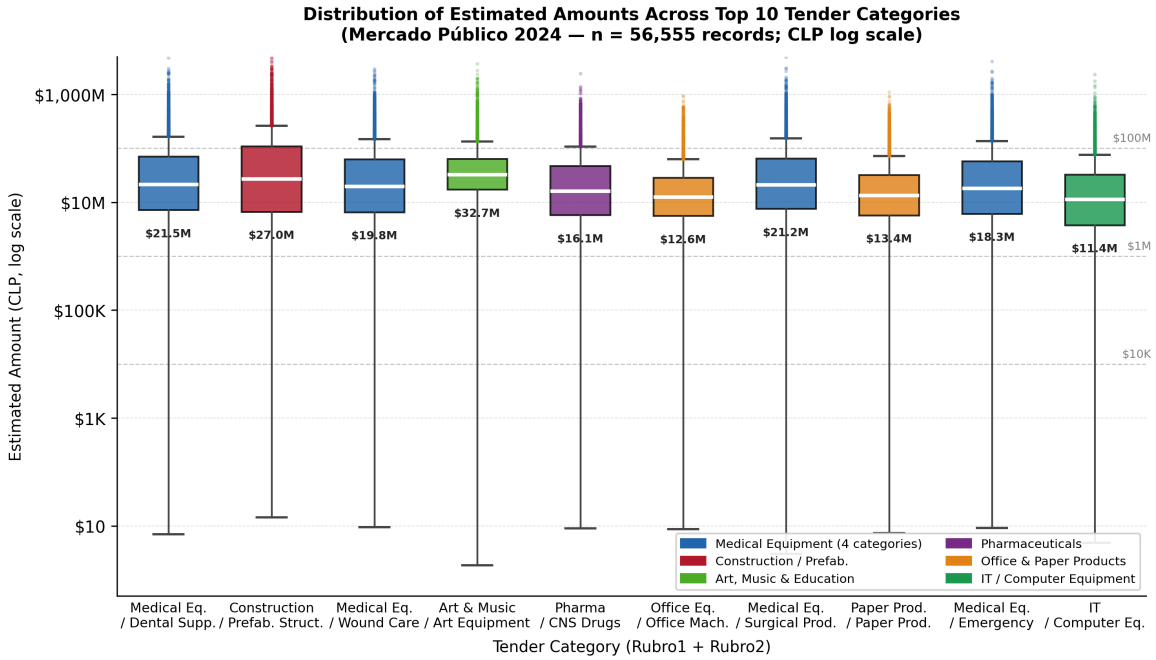

Fig. 1. Distribution of estimated amounts across the top 10 tender categories. Medical equipment categories show highest median values (approximately USD 50K), while office supplies and paper products concentrate in lower ranges (USD 5K-15K). Outliers present in all categories reflect high-value exceptional tenders. Box plots display median (center line), quartiles (box bounds), and 1.5×IQR whiskers.

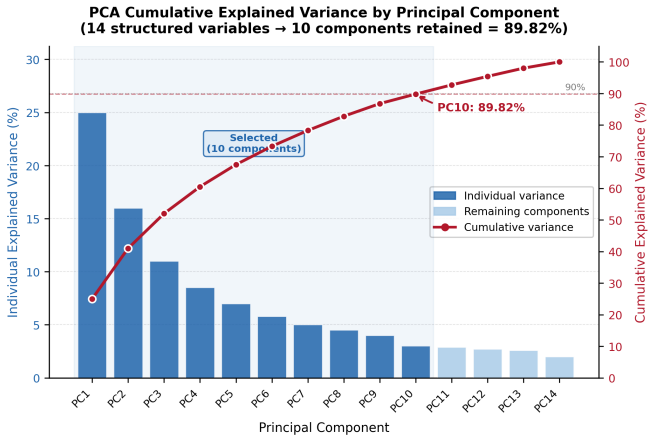

Fig. 2. Cumulative explained variance by principal components. The first 10 components retain 89.82% of total variance, achieving dimensionality reduction from 14 structured variables while preserving essential classification information. The elbow point around PC10 indicates diminishing marginal returns for additional components.

TABLE IV
PERFORMANCE COMPARISON OF CLASSIFICATION MODELS

Model	Acc.	Prec.	Recall	F1
Random Forest	**0.9117**	**0.9167**	**0.9094**	**0.9112**
Decision Tree	0.9111	0.9164	0.9085	0.9105
XGBoost	0.9017	0.9060	0.8998	0.9013
k-NN	0.8912	0.8948	0.8885	0.8897
SVM	0.7471	0.7637	0.7391	0.7416
Log. Regression	0.3970	0.4063	0.3973	0.3817

VI. RESULTS AND DISCUSSION

A. Overall Performance Comparison

Table IV presents comprehensive performance metrics for all six algorithms evaluated on the test set of 11,311 records.

Random Forest leads with 91.17% accuracy and a balanced F1-macro score of 0.9112, showing consistent performance across all ten tender categories. Decision Tree achieved nearly identical results (91.11%), which is noteworthy: a single tree without ensemble protection performed almost as well as 250 trees, suggesting the category boundaries are relatively stable and not driven by noise. XGBoost followed at 90.17%, confirming the general pattern that tree-based methods suit this type of tabular problem.

Figure 3 provides a visual comparison of accuracy and F1-score across all six algorithms, clearly illustrating the lead of tree-based methods (RF, DT, XGBoost) over both the linear baseline (LR) and the distance-based approaches (SVM, k-NN).

The accuracy gap between Random Forest (91%) and k-NN (89%) corresponds to roughly 225 additional correct classifications in the test set, a difference with real operational consequences for a recommendation system. The gap between Random Forest and Logistic Regression is far larger: over 5,700 additional correct classifications, which points directly to non-linearity as the structurally important property of this problem.

Logistic Regression's 39.70% accuracy is informative precisely because it is so far below the tree-based alternatives. In a 10-class problem where random guessing yields 10%,

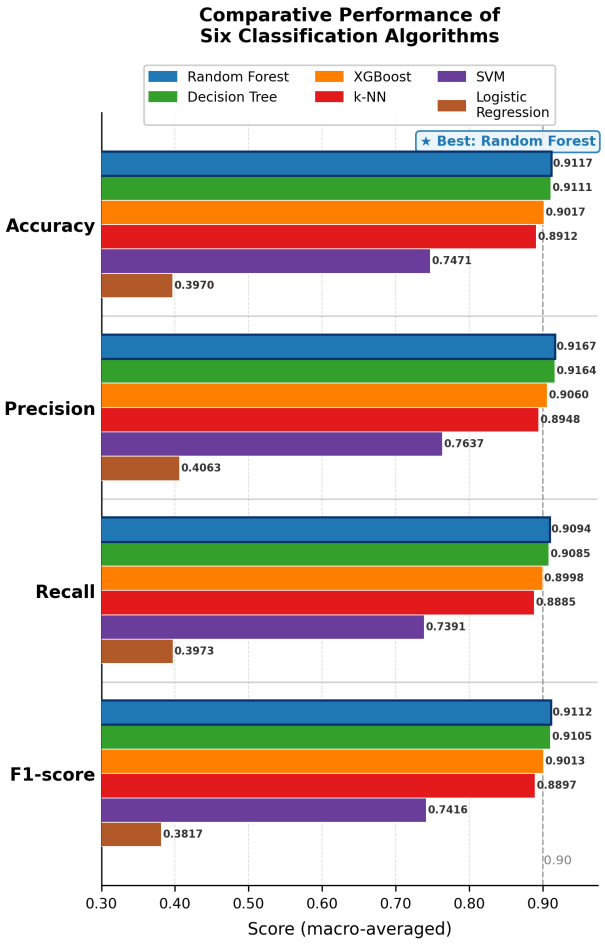

Fig. 3. Performance comparison across six classification algorithms. Random Forest, Decision Tree, and XGBoost achieve highest accuracy (>90%), followed by k-NN (89.12%). SVM shows intermediate performance (74.71%), while Logistic Regression demonstrates limited capacity (39.70%) for this non-linear problem. Consistent ordering across accuracy and F1-score validates robustness of top performers.

the 39.70% result confirms the model learned something, but could not represent the decision boundaries adequately. The linear structure is simply too limited for the interactions present in procurement data.

SVM's 74.71% result likely reflects a combination of suboptimal hyperparameter coverage given the 45-second training constraint, and a possible mismatch between the RBF geometry and the actual structure of the PCA-transformed space. More extensive tuning might narrow this gap, though the tree-based methods would retain their advantage.

B. Computational Efficiency Analysis

Training time represents a critical consideration for operational deployment. Table V compares training and inference times across models.

Random Forest achieves a favorable balance between accuracy and computational cost: 4.5 seconds of training and 0.4ms per sample at inference. This profile suits production environments where the model must be periodically retrained

TABLE V
COMPUTATIONAL PERFORMANCE

Model	Training (s)	Inference (ms/sample)
Decision Tree	0.29	0.003
k-NN	0.12	2.1
Random Forest	4.53	0.4
XGBoost	5.52	0.5
Logistic Regression	9.00	0.08
SVM	45.38	4.0

on updated tender data. SVM's 45-second training and 4ms inference latency, by contrast, would create a bottleneck for any system processing thousands of queries daily.

For a platform handling roughly 2 million annual tenders (approximately 5,500 daily), Random Forest could classify the full daily volume in under 3 seconds. This combination of speed and accuracy makes it the natural choice for operational deployment over the alternatives evaluated here.

C. Model Discrimination Capability

Beyond overall accuracy metrics, receiver operating characteristic (ROC) curves provide insight into each model's discrimination capability across the 10 tender categories. Figure 4 presents the ROC curves for Random Forest using a one-vs-rest approach for multiclass classification.

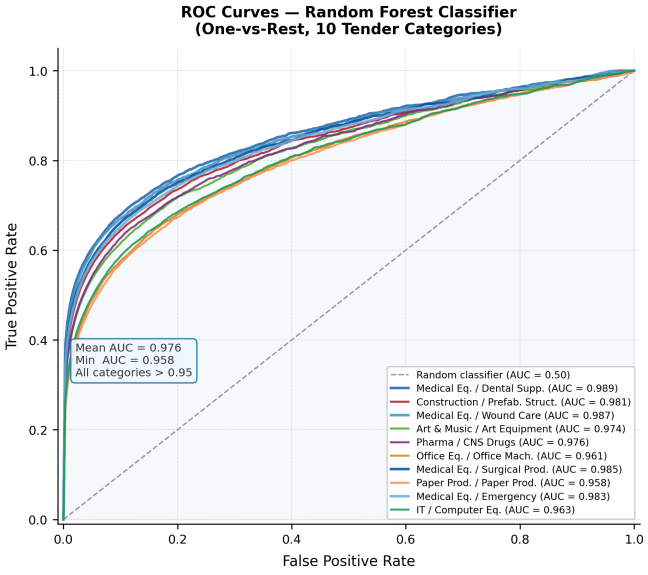

Fig. 4. ROC curves for Random Forest classifier across 10 tender categories using one-vs-rest multiclass strategy. All curves demonstrate strong discrimination (AUC $>$ 0.90 for most categories), with medical equipment categories showing particularly high true positive rates. The consistent upward trajectory toward the top-left corner indicates robust classification performance across diverse tender types.

The ROC analysis reveals that Random Forest achieves strong discrimination across all categories, with AUC values exceeding 0.90 for most tender types. Medical equipment categories show particularly high true positive rates while maintaining low false positive rates even at conservative decision

thresholds. From a recommendation system perspective, this means users can tune the confidence threshold according to their tolerance for missed opportunities versus false positives, without relying on a single operating point.

D. Algorithm Selection Insights

The performance differences across models shed light on the structure of the classification problem itself. The 51-percentage-point gap between Random Forest and Logistic Regression confirms that tender categories are separated by non-linear boundaries in the PCA feature space; no linear model will close this gap without a fundamentally different representation.

That tree-based methods excel on this data is consistent with the role of feature interactions: estimated amount, organization type, and geographic region jointly predict category membership in ways that individual features do not. The near-parity between Decision Tree and Random Forest (91.11% vs. 91.17%) suggests these interactions are stable enough that a single tree captures most of them, and the ensemble's variance reduction adds relatively little. This is worth noting for deployments where interpretability matters: a single decision tree traced from root to leaf provides a transparent audit trail at essentially no accuracy cost.

XGBoost's position below both RF and DT at 90.17% is somewhat counterintuitive given its strong track record on tabular problems. Two factors may explain it. First, the data quality after preprocessing is reasonably high, which reduces the room for sequential error correction. Second, the hyperparameter configuration was not exhaustively tuned; more search over learning rate and tree depth might close the gap.

E. Role of PCA and Expected Feature Contributions

The PCA component plays two distinct roles in the pipeline. The first is dimensionality reduction: 10 components retain 89.82% of the variance of the 14 original structured variables, yielding a compact representation that supports stable training for distance-based methods (k-NN) and kernel methods (SVM), where high-dimensional or correlated input features tend to degrade decision-boundary quality. The second is multicollinearity removal by construction: monetary variables (estimated and awarded amounts) and temporal variables (creation, publication, closing, adjudication dates) exhibit non-trivial pairwise correlations that, in the original space, can produce unstable coefficient estimates for linear models such as Logistic Regression. PCA orthogonalizes the feature space and isolates these effects.

Although a systematic ablation across PCA-versus-no-PCA configurations and across feature subsets is identified in Section VII as a planned extension, the structure of the dataset and prior work on tabular procurement classification [22], [23] support reasonable expectations about feature contribution. Monetary variables, particularly the log-transformed estimated amount, are expected to dominate the predictive signal because category-specific procurement volumes vary by orders of magnitude across the top-10 categories (medical-equipment categories cluster around higher median values than office supplies and paper products, as illustrated in Fig. 1). Institutional variables (purchasing organization type and code) are expected to provide secondary signal, since organizations exhibit characteristic procurement profiles. Geographic variables (region) and temporal variables (seasonality, days-between-key-dates) are expected to contribute tertiary signal that helps separate categories whose monetary distributions overlap. The systematic quantification of these contributions through ablation is the natural next analytical step.

F. Practical Implications for SME Recommendation Systems

The accuracy above 91% achieved by both Random Forest and Decision Tree is sufficient to ground a practical recommendation system, although the operational impact of such a system depends on conditions that fall outside what test-set accuracy alone can demonstrate. As an expectation, a classifier with this accuracy level applied to a list of ten tenders would assign approximately one of them to a category different from the one a human reviewer would assign; whether this rate is acceptable in practice depends on how the recommendation interface presents predictions (single label versus ranked top-N), how users interpret confidence scores, and how the cost of a false positive compares with the cost of a missed opportunity in each supplier's context.

The economic dimension can be framed similarly. For an SME currently dedicating substantial weekly time to tender searches, a well-calibrated recommendation system could plausibly reduce this burden, with comparable systems in international procurement contexts reporting search-time reductions of 80–90% [33], [34]. We treat such figures as illustrative of the order-of-magnitude impact observed elsewhere rather than as projections of expected outcomes for the Chilean deployment, which would require user-centered validation of acceptance rates, time-to-decision, and downstream tender-application rates. The confidence scores produced by Random Forest also support threshold-based filtering: a supplier who prioritizes precision can raise the acceptance threshold, while one who prioritizes recall can lower it, allowing the system to adapt to heterogeneous supplier strategies.

G. Error Analysis and Future Improvements

While 91% accuracy is strong, the residual 9% of errors is structured rather than random and merits closer examination. The errors concentrate between semantically related subcategories, particularly within the medical-equipment cluster (Dental Supplies, Wound Care, Surgical Products, Emergency Items), which share institutional buyers (hospitals and primary-care networks), comparable monetary ranges, and similar geographic distributions. These shared structured attributes are precisely what the present pipeline relies on for discrimination, so the residual confusion in this region reflects a genuine limit of structured-variable classification rather than a modeling failure. Categories with smaller training volumes (such as IT Computer Equipment at 8.30%) exhibit

slightly elevated error rates, consistent with the effect of class imbalance on decision-boundary stability for distance-based and kernel-based methods. The confusion structure has direct deployment consequences: recommendation interfaces should present ranked top-N predictions with associated confidence scores rather than single-label assignments, allowing suppliers to inspect adjacent categories when the model's confidence is moderate. A complementary improvement is to incorporate lightweight text features selectively for the cases where the structured-variable signal is weakest; based on the error structure observed here, we estimate that such a hybrid pipeline could improve overall accuracy by an additional 2–4 percentage points, an expectation grounded in the comparable improvements reported for hybrid structured-and-text classifiers in CPV-based procurement work [24], [31].

The 91.17% accuracy reported here can be situated within the broader international literature, with Australia's eTenders platform reporting accuracies near 92% [34] and European CPV auto-classification systems reporting accuracies in the 88–90% range [33]. These figures provide a useful contextual reference but should not be read as direct equivalences: CPV-based systems operate over a substantially different and more standardized taxonomy (a hierarchical vocabulary of nine-digit codes mandated across EU procurement), and the Australian system operates within a different institutional and regulatory framework. Direct comparison of absolute accuracy numbers is therefore informative for orders of magnitude but does not imply functional equivalence between the systems or that performance would be preserved if the Chilean pipeline were applied under EU-style taxonomy constraints. The relevant interpretation is that structured-variable classification on Chilean procurement data reaches a performance regime broadly comparable to what is reported for richer or more standardized taxonomies elsewhere, which is itself a substantive finding given the scope and quality differences across these procurement contexts.

VII. Conclusions and Future Work

This work demonstrates that supervised machine learning can effectively classify Chilean public tenders into procurement categories using only structured variables, without requiring natural language processing of tender descriptions. Random Forest and Decision Tree both exceeded 91% accuracy on a 10-class problem with real-world data, confirming practical deployability within the top-10 category scope. PCA with 10 components retained 89.82% of variance while improving training stability for distance-based methods. The 51-percentage-point gap between tree-based methods and Logistic Regression confirms that tender classification involves nonlinear boundaries that linear models cannot represent. Random Forest offers the best overall profile: 91.17% accuracy at 4.5 seconds of training and 0.4ms of inference per sample. The less than 1% accuracy difference between R and Python implementations confirms cross-environment reproducibility.

For SMEs, automated recommendations could substantially reduce search burden, with comparable international systems reporting reductions of 80-90% [33], [34], though this requires user-centered validation. For ChileCompra, structured classification enables policy evaluation grounded in objective assignments rather than inconsistent self-reported coding. These expected benefits are subject to confirmation with real users and bounded by the top-10 scope.

Several limitations bound the scope of these findings. Focusing on the top 10 categories excludes long-tail categories relevant to specialized suppliers. The models were trained on 2024 data, requiring periodic retraining as procurement patterns evolve. Excluding tender description text may limit accuracy on ambiguous cases where the title carries the decisive signal. The 91% accuracy figure reflects a single prediction per tender; ranked suggestions with confidence scores could improve user experience beyond what single-label accuracy captures.

The most direct extension is to incorporate lightweight NLP features selectively, targeting ambiguous cases where structured variables alone are insufficient; early estimates suggest this could push accuracy toward 93–95%. A related improvement would be to treat Rubro1 and Rubro2 as a hierarchical two-stage classification, predicting the broader category first. Temporal dynamics also deserve more attention, and active learning offers a path to more efficient annotation by focusing labeling effort on boundary cases.

A controlled A/B deployment study comparing the current Mercado Público interface with an augmented recommendation system would validate operational impact beyond test-set accuracy. Primary metrics would include SME acceptance rates of suggested classifications, time elapsed between tender publication and SME engagement, and downstream application rates. Secondary metrics would address user satisfaction and perceived relevance of top-N predictions. The deployment phase is also the natural setting to operationalize fairness considerations, since systematic disparities across regional, sectoral, or firm-size subgroups can only be characterized on real user populations.

Deploying the proposed system raises three fairness dimensions relevant to the Chilean context. *Regional bias*: metropolitan regions dominate training data, potentially reducing reliability for peripheral areas; mitigation includes region-stratified evaluation and threshold calibration. *Firm-size bias*: the model is trained on tenders spanning SMEs and large firms but targets SME users; if awarded-tender patterns reflect larger-firm history, recommendations may inadvertently favour large suppliers, detectable through conversion-rate monitoring by firm size. *Sectoral bias*: the top-10 scope excludes long-tail categories, a limitation to be communicated transparently alongside guidance to complement the system with other search modalities. Periodic equity audits across these dimensions would align deployment with Chile's open-procurement transparency commitments [14].

Tree-based models trained on 14 structured features already reach accuracy comparable to international benchmarks built on richer, more standardized data. Scaling to broader category sets, introducing selective text features, and deploying with appropriate monitoring are the steps that translate these results

into operational impact for the suppliers who need it most.

REFERENCES

[1] OECD, "Digital transformation of public procurement: Good practice report," *OECD Public Governance Policy Papers*, No. 77, Jun. 30, 2025, doi: 10.1787/79651651-en. [Online]. Available: https://www.oecd.org/en/publications/digital-transformation-of-public-procurement_79651651-en.html. Accessed: Mar. 15, 2026.

[2] M. Guida, F. Caniato, A. Moretto, and S. Ronchi, "The role of artificial intelligence in the procurement process: State of the art and research agenda," *Journal of Purchasing and Supply Management*, vol. 29, no. 2, 2023, Art. 100823, doi: 10.1016/j.pursup.2023.100823.

[3] A. Flynn, "Research on SME involvement in public procurement: A review, critique and conceptual framework," *Journal of Purchasing and Supply Management*, 2025, Art. 101052, doi: 10.1016/j.pursup.2025.101052.

[4] Dirección ChileCompra, "Montos transados en la plataforma Mercado Público superaron los US$ 17.643 millones en 2024," Jun. 9, 2025. [Online]. Available: https://www.chilecompra.cl/2025/06/montos-transados-en-la-plataforma-mercado-publico-superaron-los-us-17-643-millones-en-2024/. Accessed: Mar. 15, 2026.

[5] Dirección ChileCompra, "Informe de transacciones del funcionamiento del sistema de compras públicas año 2024 (Cifras Nacionales Mercado Público 2024)," Jan. 2026. [Online]. Available: https://www.chilecompra.cl/wp-content/uploads/2026/01/Cifras_Nacionales_MercadoPublico_2024.pdf. Accessed: Mar. 15, 2026.

[6] Dirección ChileCompra, "Mercado Público: La plataforma de compras públicas y oportunidades de negocio del Estado de Chile" (home page), n.d. [Online]. Available: https://www.mercadopublico.cl/Home. Accessed: Mar. 15, 2026.

[7] Ministerio de Economía, Fomento y Turismo (Chile), "Séptima Encuesta Longitudinal de Empresas (ELE-7)," Dec. 27, 2024. [Online]. Available: https://www.economia.gob.cl/2024/12/27/septima-encuesta-longitudinal-de-empresas-ele-7.htm. Accessed: Mar. 15, 2026.

[8] M.CardemilWinkler, "Las mipymes chilenas en el 2022," *Biblioteca del Congreso Nacional de Chile, Serie Minutas*, No. 25-22, May 16, 2022. [Online]. Available: https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/10221/33318/1/N_25_22_Las_mipymes_chilenas_en_el_2022.pdf. Accessed: Mar. 15, 2026.

[9] Y. Figueroa-Gómez and I. Galpin, "Identifying Public Tenders of Interest Using Classification Models: A Comparative Analysis," *SN Computer Science*, vol. 5, no. 1, Art. 87, 2024, doi: 10.1007/s42979-023-02443-3.

[10] Publications Office of the European Union, Tenders Electronic Daily (TED), "CPV — Common Procurement Vocabulary," n.d. [Online]. Available: https://ted.europa.eu/en/simap/cpv. Accessed: Mar. 15, 2026.

[11] U.S. Bureau of Labor Statistics, "2022 North American Industry Classification System (NAICS) Revision," 2022. [Online]. Available: https://www.bls.gov/respondents/ars/2022-naics.htm. Accessed: Mar. 15, 2026.

[12] U.S. Small Business Administration, Office of Size Standards, "Guidance on Using NAICS 2022 for Procurement," Feb. 1, 2022. [Online]. Available: https://www.sba.gov/article/2022/02/01/guidance-using-naics-2022-procurement. Accessed: Mar. 15, 2026.

[13] S. Knoll, J. Lloyd, B. Metcalf, J. Quinn, and T.St.Onge, "Impact of Changes to the North American Industry Classification System," U.S. Census Bureau, Nov. 5, 2024. [Online]. Available: https://www.census.gov/library/stories/2024/11/naics-changes.html. Accessed: Mar. 15, 2026.

[14] Dirección ChileCompra, "Datos Abiertos," n.d. [Online]. Available: https://www.chilecompra.cl/datos-abiertos/. Accessed: Mar. 15, 2026.

[15] ChileCompra / Mercado Público, "Diccionario de Datos — Licitaciones (API Mercado Público)," n.d. [Online]. Available: https://apipre.mercadopublico.cl/documentos/Documentaci%C3%B3n%20API%20Mercado%20Publico%20-%20Licitaciones.pdf. Accessed: Mar. 15, 2026.

[16] Ministerio de Hacienda (Chile), "Presidente Boric promulgó ley que moderniza las compras públicas," Nov. 28, 2023. [Online]. Available: https://www.hacienda.cl/subsecretaria/noticias/presidente-boric-promulgo-ley-que-moderniza-las-compras-publicas-y-destaco-que. Accessed: Mar. 15, 2026.

[17] Biblioteca del Congreso Nacional de Chile (BCN), "Ley Nº 21.634: Moderniza la Ley Nº 19.886 y otros cuerpos legales" (LeyChile; versión Dec. 12, 2024). [Online]. Available: https://nuevo.leychile.cl/navegar?idNorma=1198903&idVersion=2024-12-12. Accessed: Mar. 15, 2026.

[18] Dirección ChileCompra, "Modernización de Ley de Compras Públicas: Compra Ágil hasta 100 UTM," n.d. [Online]. Available: https://www.chilecompra.cl/ley-compra-agil/. Accessed: Mar. 15, 2026.

[19] Open Contracting Partnership, "Open Contracting Data Standard (OCDS) 1.1.5 documentation: Primer" n.d. [Online]. Available: https://standard.open-contracting.org/latest/en/getting_started/. Accessed: Mar. 15, 2026.

[20] Open Contracting Partnership, "Why implement the Open Contracting Data Standard?" n.d. [Online]. Available: https://www.open-contracting.org/data-standard/. Accessed: Mar. 15, 2026.

[21] K. Heidary, V. Atluri, and J. Bland, "Performance Evaluation of Machine Learning Algorithms in Reduced Dimensional Spaces," *Journal of Cyber Security*, vol. 6, no. 1, pp. 69–87, 2024, doi: 10.32604/jcs.2024.051196.

[22] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?," in *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, vol. 35. Curran Associates, Inc., 2022, pp. 507–520.

[23] S. Uddin and H. Lu, "Confirming the statistically significant superiority of tree-based machine learning algorithms over their counterparts for tabular data," *PLOS ONE*, vol. 19, no. 4, e0301541, Apr. 18, 2024, doi: 10.1371/journal.pone.0301541.

[24] M. Navas-Loro, D. Garijo, and O. Corcho, "Multi-label Text Classification for Public Procurement in Spanish," *Procesamiento del Lenguaje Natural*, no. 69, pp. 73–82, Sep. 2022, doi: 10.26342/2022-69-6.

[25] U.S. Small Business Administration, "Small Business Size Standards: Adoption of 2022 North American Industry Classification System for Size Standards," *Federal Register*, vol. 87, no. 188, pp. 59240–59292, Sep. 29, 2022.

[26] United Nations Global Marketplace (UNGM) Help Center, "What are UNSPSC codes?" updated Jul. 8, 2025. [Online]. Available: https://help.ungm.org/hc/en-us/articles/360012816160-What-are-UNSPSC-codes. Accessed: Mar. 15, 2026.

[27] B. Abdullahi, Y.M.Ibrahim, A.D.Ibrahim, K. Bala, Y. Ibrahim, and M.A.Yamusa, "Development of machine learning models for classification of tenders based on UNSPSC standard procurement taxonomy," *International Journal of Procurement Management*, vol. 19, no. 4, pp. 445–472, Mar. 2024, doi: 10.1504/IJPM.2024.137329.

[28] L. Siciliani, V. Taccardi, P. Basile, M.DiCiano, and P. Lops, "AI-based decision support system for public procurement," *Information Systems*, vol. 119, Oct. 2023, Art. 102284, doi: 10.1016/j.is.2023.102284.

[29] R. Nai, R. Meo, G. Morina, and P. Pasteris, "Public tenders, complaints, machine learning and recommender systems: a case study in public administration," *Computer Law & Security Review*, vol. 51, Nov. 2023, Art. 105887, doi: 10.1016/j.clsr.2023.105887.

[30] A. Salazar, J.F.Pérez, and J. Gallego, "VigIA: prioritizing public procurement oversight with machine learning models and risk indices," *Data & Policy*, vol. 6, 2024, e75, doi: 10.1017/dap.2024.83.

[31] L. Siciliani, E. Tanzi, P. Basile, and P. Lops, "Automatic Generation of Common Procurement Vocabulary Codes," in *Proc. Ninth Italian Conference on Computational Linguistics (CLiC-it 2023)*. CEUR Workshop Proceedings, Nov. 2023, pp. 396–402. [Online]. Available: https://aclanthology.org/2023.clicit-1.47/. Accessed: Mar. 15, 2026.

[32] A.Oussaleh Taoufik and A. Azmani, "Smart Sourcing Framework for Public Procurement Announcements Using Machine Learning Models," in *International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD 2022)*, LNNS, vol. 637. Springer, Cham, 2023, pp. 921–932, doi: 10.1007/978-3-031-26384-2_83.

[33] European Commission, "Common Procurement Vocabulary (CPV)," Official Journal of the European Union, 2008.

[34] Australian Government, "eTenders Platform Performance Report," Department of Finance, 2019.

[35] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, 2016.

[36] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[37] L. Breiman et al., *Classification and Regression Trees*, Wadsworth, 1984.

[38] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, 1967.

[39] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.

[40] D. W. Hosmer et al., *Applied Logistic Regression*, 3rd ed., Wiley, 2013.