Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: https://www.science-gate.com/IJAAS.html

Applying decision tree models to identify key factors influencing voluntary military enlistment: A case study of Thailand's 23rd Military District

Thotsaporn Phanngoen, Kriangsak Chanthinok *, Palan Jantarajaturapath

Department of Digital Business and Information System, Mahasarakham Business School, Mahasarakham University, Maha Sarakham Province, Thailand

ARTICLE INFO

Article history:
Received 18 May 2026
Received in revised form
28 August 2026
Accepted 23 September 2026

Keywords:
Military enlistment
Voluntary recruitment
Decision tree models
Random forest
Machine learning

ABSTRACT

Military enlistment decisions are shaped by complex, multidimensional factors, including socioeconomic status and personal attitudes. This study develops a predictive decision model to identify the key factors that distinguish voluntary enlistment from conscription by lottery in Thailand's 23rd Military District. A dataset of 200 military personnel (147 volunteers and 53 drafted personnel) was analyzed using data mining techniques in Weka. The Random Tree and Random Forest algorithms were employed alongside three baseline classifiers: logistic regression, Naive Bayes, and J48. The models were evaluated based on accuracy, precision, recall, F1 score, and kappa statistics. Random Forest achieved the best balance between overall accuracy and sensitivity to the minority class (Accuracy = 92.86%, Kappa = 0.814, Weighted F1 = 0.926, Lottery-class F1 = 0.861; 10-fold cross-validation accuracy = 95.95% ± 0.44), outperforming all baseline models, particularly in classifying the minority (lottery/drafted) class, for which logistic regression achieved an F1 score of only 0.706. The most influential demographic predictors were occupation, province, and income. When attitudinal variables were included, cultural factors became the most important factor, followed by risk and legal perceptions. However, predictive accuracy did not improve compared with using demographic variables alone. These findings show that machine learning can effectively model enlistment behavior. The Royal Thai Army should develop recruitment policies that target specific occupations and geographical areas while strengthening cultural messaging and risk communication to increase voluntary enlistment rates.

© 2026 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The recruitment and selection of personnel is a critical mission for national security and military operational readiness. In Thailand, the process is governed by the Military Service Act B.E. 2497 (1954), which requires all Thai males reaching the age of 21 to undergo selection for military service. The resulting system is a hybrid mechanism that combines voluntary enlistment with a conscription lottery (drawing black and red cards) to meet the manpower needs of the Royal Thai Army. According to Binková and Štěpánková (2023), the decision to enlist in the military is often influenced by internal motivations such as patriotism and self-improvement, as well as external factors like economic incentives, educational opportunities, and social pressure. Understanding these motivations is crucial for military recruiters to design effective strategies to attract suitable candidates. Although the legal framework compels participation, the choice to enlist voluntarily rather than risk the lottery is a multidimensional behavioral decision shaped by economic, social, cultural, and psychological factors (Baker and Phongpaichit, 2017). Understanding how these factors combine, and which of them dominate, requires analytical tools able to capture nonlinear and hierarchical relationships rather than conventional statistical models alone.

Three gaps remain in this body of work. First, existing scholarship examines enlistment either under fully voluntary systems (All-Volunteer Force, AVF) or under full conscription, leaving hybrid systems such as Thailand's—where the same individual must choose between volunteering and

* Corresponding Author.
Email Address: kriangsak.c@acc.msu.ac.th (K. Chanthinok)
https://doi.org/10.21833/ijaas.2026.09.013
Corresponding author's ORCID profile:
https://orcid.org/0000-0003-2448-4575
2313-626X/© 2026 The Authors. Published by IASE.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

entering a lottery—largely unmodeled. Second, machine learning applications in security research have concentrated on macro-level structural indicators such as GDP, terrain characteristics, and political instability rather than on the micro-level attitudinal determinants of individual enlistment. Third, evidence on localized feature importance within specific military districts is limited, so national-level explanations may conceal substantial regional variation.

This study addresses these gaps by developing a data-driven decision model of enlistment behavior using Random Tree and Random Forest algorithms implemented in WEKA. Micro-level Likert-scale measures across six attitudinal dimensions—economic, social, psychological, cultural, legal/policy, and risk factors—are combined with demographic data collected from personnel in Thailand's 23rd Military District. Root node and splitting-criterion analysis is then used to identify which dimensions act as primary predictors of voluntary enlistment versus lottery-based conscription, and whether cultural values, economic incentives, or risk perceptions function as dominant decision nodes.

The study contributes theoretically by extending ensemble learning methods to hybrid military service systems and methodologically by integrating behavioral data into predictive security modeling. Because decision tree–based algorithms yield transparent and interpretable classification rules, the approach also responds to the requirement for explainability in AI-assisted public-sector decision-making, where policy decisions must be traceable and defensible. Practically, the results give the Royal Thai Army an evidence base for recruitment strategy, targeted public relations, and welfare policies aligned with the behavioral profiles of potential recruits.

However, unlike conventional statistical approaches such as logistic regression or structural equation modeling, decision tree–based algorithms offer transparent and interpretable classification rules that are particularly well suited to public-sector decision-making contexts. In AI-assisted governance, explainability is now a must-have. Policy decisions must be clear, easy to follow, and fair to all parties involved. In response to this need, the present study adopts an explainable machine learning framework that systematically integrates data-driven analysis with policy-oriented insights. As illustrated in Fig. 1, the proposed research workflow follows the Knowledge Discovery in Databases (KDD) paradigm, progressing from data collection and preprocessing to model development using the WEKA platform, predictive evaluation, and the translation of analytical findings into actionable recruitment policies. By explicitly linking machine learning outputs to strategic decision-making, this study contributes not only to the literature on military recruitment but also to the broader field of digital governance, demonstrating how understandable AI can effectively support evidence-based policy formulation.

Fig. 1 illustrates the integrated research workflow adopted in this study, which follows the Knowledge Discovery in Databases (KDD) paradigm and connects data-driven analytics with policy-oriented decision-making. The framework systematically links data collection, machine learning model development, predictive evaluation, and policy formulation to support evidence-based military recruitment strategies.

The workflow begins with the Knowledge Discovery (KDD) phase, which involves data collection and preprocessing. Survey data and demographic information were collected from military personnel in Thailand's 23rd Military District. This phase includes data cleaning, handling missing values, encoding categorical variables, and aggregating attitudinal constructs. These steps are essential to ensure data quality and consistency prior to machine learning analysis, as emphasized in established KDD and data mining literature (Fayyad et al., 1996; Han et al., 2011).

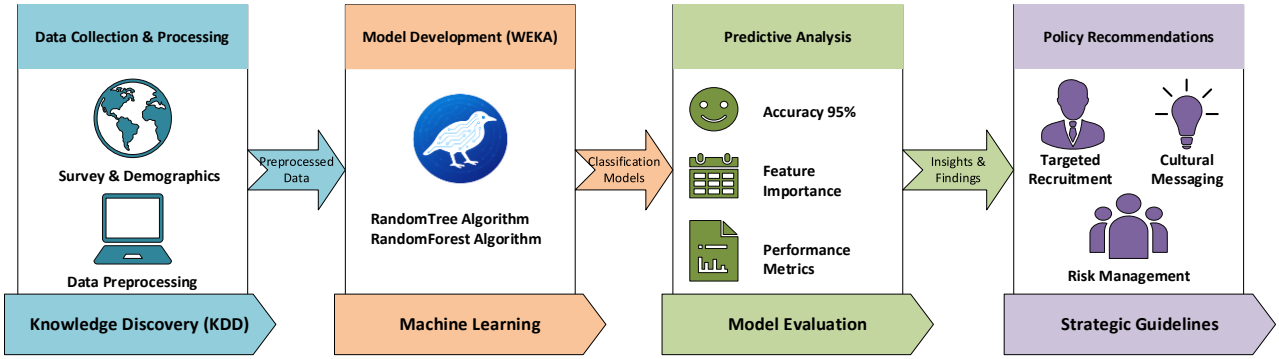


Fig. 1: The research workflow framework for data-driven military enlistment analysis

Next, the Model Development phase is implemented using the WEKA machine learning platform. WEKA was selected due to its robustness, reproducibility, and wide adoption in academic research for classification tasks. In this study, decision tree–based algorithms—namely Random Tree and Random Forest—were employed to classify enlistment outcomes (voluntary enlistment versus lottery-based conscription). These algorithms were chosen because of their ability to model nonlinear relationships and their suitability for handling mixed data types commonly found in social and behavioral

datasets (Breiman, 2001). The third stage, predictive analysis and model evaluation, focuses on assessing model performance using standard classification metrics, including accuracy, precision, recall, and the Kappa statistic. Beyond predictive accuracy, this stage emphasizes model interpretability, particularly through feature importance and tree structure analysis. The use of decision tree–based models enables transparent identification of key predictors, such as occupation, province, and cultural perception, aligning with the growing emphasis on explainable artificial intelligence (XAI) in public-sector analytics.

Finally, the workflow culminates in the policy recommendation phase, where analytical insights are translated into actionable strategic guidelines. The extracted decision rules and hierarchical factor structures inform targeted recruitment strategies, culturally oriented communication, and risk management policies. This final stage reflects the principles of data-driven governance, demonstrating how machine learning outputs can be systematically transformed into practical policy interventions for public organizations.

Overall, Fig. 1 highlights the contribution of this research in bridging machine learning techniques with policy-oriented applications, illustrating how understandable models can enhance transparency, accountability, and effectiveness in military recruitment decision-making.

2. Literature review

2.1. Determinants of military enlistment

Scholarship on military recruitment centers on the tension between state coercion and individual agency. Political institutions, colonial legacies, and security threats shape whether states rely on conscription or all-volunteer forces (AVFs). In a 200-year cross-national analysis, Asal et al. (2017) found that democracies are statistically less likely than autocracies to implement the draft and that states with a British colonial heritage show a strong historical aversion to conscription, whereas interstate rivalry or war significantly raises the probability of enforcement. Cohn and Toronto (2017) added that those structural economic variables—particularly labor market regulation—are more decisive than threat levels, since states with more regulated labor markets rely significantly more on conscripts.

At the individual level, socio-economic status is central to the enlistment decision. Cameron et al. (2000) showed that selection bias in conscription disproportionately affects lower socio-economic groups, framing service as a burden placed on the poor rather than a universal civic duty. Under all-volunteer systems these framing shifts: Strand and Berndtsson (2015) described a “marketization” of recruitment in Sweden and the UK, in which service is presented not as a duty but as a path to self-fulfillment—the “enterprising soldier.” Attitudinal dispositions matter as well. In a study directly relevant to the Thai context, Bjarnegård et al. (2023) found that patriarchal values drive voluntary participation in paramilitary service in southern Thailand while conscription itself does not produce such values, indicating self-selection on pre-existing attitudes. Béhague (2008) reported a parallel dynamic in Brazil, where marginalized youth seek military training as a form of institutionalized therapy and social mobility. Woodruff (2017) further demonstrated that intrinsic motives such as altruistic service and self-enhancement generate stronger organizational commitment and pro-social behavior than extrinsic economic incentives, with effects persisting throughout military careers, and Woodruff et al. (2025) stressed the importance of understanding how motivational factors interact amid declining propensity to serve.

2.2. Machine learning in military and security research

Machine learning has demonstrated strong performance across military and security applications. Random Forest outperforms standard and penalized logistic regression in predicting rare events such as civil war onset (Muchlinski et al., 2016) and crash risk (Omrani et al., 2026), owing to its ability to model complex nonlinear interactions without restrictive parametric assumptions; Omrani et al. (2026) also confirmed its superiority over Multi-Layer Perceptron (MLP) and Ordinal Logistic Regression (OLR), particularly when combined with SMOTE (Synthetic Minority Oversampling Technique) to address class imbalance. Related applications include NLP-based classification of military aircraft failure data (de Almeida et al., 2026), feature selection for admission into national military high schools (Plăcintă, 2022), forecasting success in US Air Force pilot training (Jenkins et al., 2022), and predicting health perception among military recruits with 95.4% accuracy using decision tree and random forest classifiers (Chen et al., 2025), while Maathuis and Cools (2025) emphasized the broader role of AI in supporting data-driven military decision-making and operational readiness.

Among these techniques, decision tree models are particularly effective for classification problems involving multiple variables. They generate interpretable structures of nodes and branches that prioritize factors according to their influence, making them well suited to analyzing the hierarchical relationship between demographic attributes, personal attitudes, and enlistment decisions.

2.3. Theoretical framework

This study is guided by an integrated framework that links established behavioral theories to the six attitudinal dimensions examined. Maslow’s (1943) hierarchy of needs underpins the economic dimension, positing that income stability and welfare

benefits address fundamental security needs. Social Learning Theory informs the social and cultural dimensions, in which family influence, social recognition, and traditional values shape behavior through the social environment. The Theory of Reasoned Action (Fishbein and Ajzen, 1975) frames the psychological dimension, whereby attitudes toward military service and subjective norms jointly determine enlistment intention. Herzberg et al.'s (1993) motivation–hygiene theory distinguishes intrinsic motivators (honor, self-improvement) from hygiene conditions represented by the legal/policy and risk dimensions, whose perceived inadequacy can deter voluntary enlistment. This framework guided both the construction of the measurement instrument (Section 3.3) and the interpretation of the decision tree structures in the Discussion (Section 5).

Taken together, prior work explains enlistment either under fully voluntary or fully conscription-based systems, relies largely on macro-level structural indicators when applying machine learning, and offers little evidence on localized feature importance. Predictive modeling of hybrid systems, in which the same individual weighs voluntary enlistment against a lottery, and of the micro-level attitudinal determinants behind that choice, therefore remains underexplored. The present study addresses this gap by combining an explainable decision tree approach with attitudinal survey data from a single military district, as set out in the methodology below.

3. Methodology

3.1. Research design

This study employed a quantitative research design utilizing data mining techniques to analyze factors influencing military enlistment decisions. The primary objective was to develop a predictive decision model to classify personnel into two categories: voluntary enlistment and conscription by lottery (drafted). The research utilized the Knowledge Discovery in Databases (KDD) process, incorporating data collection, preprocessing, and model development using machine learning algorithms within the Weka (Waikato Environment for Knowledge Analysis) version 3.9.7 software environment.

3.2. Population and data collection

The population of this study consisted of Thai males who passed the military service selection process in the 23rd Military District, which covers the provinces of Khon Kaen, Kalasin, Sakon Nakhon, and Bueng Kan.

Purposive sampling was used to select 200 military personnel currently serving in the district (Campbell et al., 2020). The participants were divided into two groups according to their enlistment method. The first group consisted of 147 volunteers (73.5%) who had voluntarily chosen to enlist, while the second group included 53 drafted personnel (26.5%) who entered military service through the lottery system and received a red card.

Before data collection, formal permission was obtained from the commander of the 23rd Military District. Data were collected on-site at military units between November 2024 and January 2025. The researcher distributed the questionnaires directly to military personnel during assembly periods to help ensure that the questions were understood and the responses were complete.

3.3. Research instruments

A closed-ended questionnaire was used as the main research instrument to collect data on the different factors that influenced the decision to enlist in military service. The questionnaire consisted of two main sections. The first section collected demographic information, including age, province of residence, marital status, number of children, education level, occupation before military service, and income.

The second section measured participants' opinions on six dimensions using a five-point Likert scale, ranging from 1 (least) to 5 (most). Economic factors (EF) included six items related to income stability and welfare. Social factors (SF) consisted of six items covering family influence and social recognition. Psychological factors (PF) included four items related to patriotism and self-improvement. Cultural factors (CF) comprised five items addressing traditional values and respect. Legal and policy factors (LPF) included five items concerning the fairness of conscription laws. Finally, risk factors (RiF) consisted of six items related to safety during military training.

The validity and reliability of the questionnaire were also assessed. Content validity was evaluated by experts, with the item-objective congruence (IOC) index ranging from 0.67 to 1.00. Reliability was assessed using Cronbach's alpha coefficient, which ranged from 0.84 to 0.90 across the different factors. These results indicated good internal consistency.

3.4. Data preprocessing

Prior to modeling, data cleaning was performed to handle missing values and outliers. Missing data in attitudinal dimensions (if less than 20%) were replaced with the mean value of the respective dimension. Variables were encoded for compatibility with Weka; specifically, the target variable Enter_way was converted to a nominal type (Voluntary/Lottery), and demographic variables were grouped to reduce sparsity. Class Imbalance Consideration: The dataset exhibits a moderate class imbalance, with 147 voluntary enlistees (73.5%) and 53 lottery-drafted individuals (26.5%), yielding an imbalance ratio of approximately 2.77:1.

While techniques such as SMOTE have been applied in prior studies to address severe imbalance (Omrani et al., 2026), the present imbalance ratio was considered moderate and did not require external resampling. Instead, the Random Forest algorithm's internal bootstrap aggregation mechanism was relied upon, which naturally produces diverse training subsets across the ensemble, mitigating the effect of class imbalance without introducing synthetic data artifacts. Model evaluation was conducted using per-class precision, recall, F1-score, and the Kappa statistic—metrics that are robust to class imbalance—in addition to overall accuracy. To verify this assumption empirically, a sensitivity analysis using cost-sensitive learning (a 3:1 penalty on misclassifying lottery-drafted cases) and class-balanced resampling was additionally performed under 10-fold cross-validation (see Section 4.3).

3.5. Model development and analysis

The study employed classification algorithms to predict enlistment behavior. Five algorithms were evaluated and compared: Logistic Regression and Naïve Bayes served as statistical and probabilistic baselines; the J48 decision tree provided a deterministic benchmark, and Random Tree and Random Forest were the primary tree-based models of interest. Random Forest was designated as the primary model given its ensemble architecture and its established robustness in handling class imbalance and nonlinear feature interactions.

Model configuration: The models were configured with specific parameters to optimize performance and prevent overfitting, as detailed in Table 1.

Based on Table 1, in this study, a decision tree model was developed using the WEKA machine learning platform. The parameter settings were carefully configured to balance model accuracy, interpretability, and generalization capability.

The minimum number of instances per leaf was set to 15 WEKA option -M 15). This parameter determines the smallest number of observations allowed in a terminal node. Setting a minimum threshold helps prevent the formation of overly specific rules that may result from noise in the data. Previous studies have shown that increasing the minimum number of instances per leaf can effectively reduce overfitting and improve model robustness, particularly when dealing with moderate sample sizes.

The minimum number of instances required to perform a split was set to approximately *30*. This constraint ensures that sufficient data are available at each decision node before allowing further partitioning. By enforcing this condition, the model avoids unnecessary splits that could degrade predictive performance and interpretability.

The maximum depth of the decision tree was limited to four levels. Controlling tree depth is a common strategy for reducing model complexity and enhancing interpretability. Shallow trees tend to generalize better to unseen data, especially in social science and behavioral studies, where excessive model complexity may obscure meaningful patterns.

The confidence factor for pruning was set to *0.25*, which is the default value in WEKA. This parameter governs the extent of post-pruning applied to the tree, with lower values leading to more aggressive pruning. Pruning plays a crucial role in mitigating overfitting by removing branches that do not contribute significantly to predictive accuracy.

The number of folds used for pruning was set between three and five. Cross-validation-based pruning allows the model to evaluate its performance across multiple subsets of the data, thereby enhancing stability and reducing variance. This approach is particularly effective in datasets where class imbalance or sampling variability may affect model outcomes.

A random seed value of 1 was specified to ensure the reproducibility of the experimental results. Fixing the random seed allows other researchers to replicate the model-building process and verify the findings, which is a key requirement in empirical research.

No explicit Bonferroni adjustment was applied, as WEKA internally manages statistical significance during tree construction. Since the primary objective of this study was predictive modeling rather than hypothesis testing, additional multiple comparison corrections were deemed unnecessary.

Overall, the selected parameter configuration reflects a trade-off between predictive accuracy and model interpretability. This is particularly important in the context of voluntary military enlistment, where decision rules must be transparent and actionable for policy analysis and decision-making.

- Validation Method: The dataset was split into a training set (70%) and a testing set (30%) to evaluate predictive performance.
- Performance Metrics: The models were evaluated based on accuracy, Kappa statistic, precision, recall, F1-score, and confusion matrix analysis.
- Variable Analysis: The study initially tested models using only demographic variables (e.g., occupation, province) and subsequently integrated the rating scale factors (meanCulture, meanRisk, etc.) to assess their impact on the decision tree structure.

4. Results

4.1. Demographic characteristics of the sample

The study analyzed data from 200 military personnel in the 23rd Military District. The analysis of the enlistment method revealed that 147 individuals (73.5%) entered the service voluntarily (request), while 53 individuals (26.5%) were drafted through the lottery process (red card). Regarding general characteristics, the majority of the sample were 21 years old (47.0%), followed by those who were 18 years old (14.0%). In terms of domicile, the

largest group originated from Khon Kaen province (48.5%), followed by Kalasin province (20.5%). The education level analysis indicated that the majority held a high school diploma or equivalent (34.5%), followed by a junior high school education (28.0%). Prior to enlistment, general labor was the most common occupation (46.5%), and the majority of personnel reported a monthly income between 5,001 and 10,000 THB (27.5%).

Table 1: The parameters of the decision tree model were set in Weka

Item(s)	Weka setting
Minimum number of instances per leaf	-M 15
Minimum number of instances per split	~30
Maximum depth of a tree	4
Confidence factor for pruning	0.25
Number of folds for pruning	3-5
Seed for randomization	1

4.2. Descriptive analysis of attitudinal factors

The study evaluated six key factors influencing enlistment decisions using a 5-point Likert scale. Table 2 presents the mean scores and standard deviations for these factors. The results indicate that psychological factors had the highest mean score ($\bar{x}$ = 3.80, S.D. = 0.81), classified as a "high" level of importance. This suggests that internal motivations, such as self-improvement and a sense of honor, play a critical role in the decision-making process. Risk Factors ranked second ($\bar{x}$ = 3.60, S.D. = 0.75), indicating that safety concerns and risk management are significant considerations for recruits. Economic Factors received the lowest mean score ($\bar{x}$ = 3.00, S.D. = 0.80), suggesting that while financial stability is relevant, it may not be the primary driver for this specific demographic compared to psychological and safety factors.

Table 2: Descriptive statistics of factors influencing enlistment decisions

Factors	Mean ($\bar{x}$)	S.D.	Interpretation
Psychological factors	3.80	0.81	High
Risk factors	3.60	0.75	High
Legal and policy factors	3.41	0.70	Moderate
Social factors	3.19	0.80	Moderate
Cultural factors	3.18	0.81	Moderate
Economic factors	3.00	0.80	Moderate
Overall	3.36	0.77	Moderate

4.3. Classification model performance

To predict the method of enlistment (voluntary vs. lottery), five classification algorithms were evaluated using Weka under identical experimental conditions (30%/70% train–test split: training n = 60, hold-out test set n = 140; random seed = 1): logistic regression, Naive Bayes, random tree, random forest, and J48 decision tree. The primary models, random tree and random forest, utilized demographic variables (province, marital status, education, occupation, and income) as predictors. The comparative results are presented in Table 3. In addition, all classifiers were re-evaluated using stratified 10-fold cross-validation repeated ten times (random seeds 1–10), and a set of robustness analyses—cost-sensitive learning, class-balanced resampling, Random Forest feature importance, and a geographic ablation test—was performed; these results are reported at the end of this section.

The Random Forest algorithm achieved the strongest balance of overall accuracy and minority-class sensitivity, with an accuracy of 92.86%, a Kappa statistic of 0.814, and a weighted F1 score of 0.926. The Random Tree model performed comparably, achieving 92.14% accuracy (Kappa = 0.790, Weighted F1 = 0.917), though with notably lower sensitivity to the lottery class (Recall = 0.783 vs. 0.870). The high Recall for the Voluntary class (0.982) in both tree-based models confirms that the demographic profile of volunteers is well-defined and consistently captured by the feature set.

The F1 score for the lottery/drafted class is the primary evaluation criterion in this study, as it represents both the minority class and the policy-critical classification target—individuals whose enlistment pathway is less predictable and more consequential for recruitment planning. Random Forest achieved an F1 of 0.861 for this class (recall = 0.775), compared to 0.841 for the random tree (recall = 0.725), confirming the ensemble model's superior robustness under class imbalance.

Among the three baseline classifiers, Logistic Regression performed the weakest (Accuracy = 85.71%, Kappa = 0.615, Lottery F1 = 0.706), reflecting its sensitivity to the class imbalance and its inability to capture nonlinear interactions among demographic variables. Naïve Bayes improved substantially (Accuracy = 90.71%, Kappa = 0.752, Lottery F1 = 0.812) yet remained below both tree-based models. J48 recorded the highest raw accuracy (97.14%, Kappa = 0.929, Lottery F1 = 0.949); however, inspection of its tree structure revealed that Province served as the sole splitting criterion across all branches—a deterministic geographic rule that exploits the strong provincial stratification in this dataset but contributes limited interpretive or generalizable insight beyond geographic location. Taken together with the robustness analyses reported below (cross-validation, confusion matrices, class-imbalance sensitivity, feature importance, and a geographic ablation test), these results support the selection of Random Forest as the primary model on the grounds of ensemble stability, a distributed and interpretable decision structure, and robustness to class imbalance.

4.3.1. Cross-validated performance

To obtain more reliable estimates than a single hold-out split, all five classifiers were re-evaluated using stratified 10-fold cross-validation repeated ten times (random seeds 1–10) under identical conditions (Table 4). The results confirm strong generalization: Random Forest achieved a mean accuracy of 95.95% (S.D. = 0.440) with a lottery-class F1 of 0.923 (S.D. = 0.008), closely matching or exceeding its hold-out performance, while Naive

Bayes improved markedly relative to the small hold-out training partition (96.70% ± 0.42).

The close agreement between the two evaluation schemes, together with the low fold-to-fold variance of all classifiers, indicates that the models generalize beyond the training partition and that the tree complexity reported in Section 4.4 does not reflect overfitting. Notably, the single Random Tree exhibited the highest variance (S.D. = 1.76), whereas the Random Forest ensemble reduced this to 0.44, empirically confirming the variance-reduction benefit of bootstrap aggregation.

Table 3: Classification performance comparison across five algorithms

Algorithm	Accuracy	Kappa	Precision (lottery)	Recall (lottery)	F1 (lottery)	Weighted F1
Logistic regression	85.71%	0.615	0.857	0.600	0.706	0.849
Naive Bayes	90.71%	0.752	0.966	0.700	0.812	0.902
Random tree	92.14%	0.790	1.000	0.725	0.841	0.917
Random forest*	92.86%	0.814	0.969	0.775	0.861	0.926
J48 decision tree**	97.14%	0.929	0.974	0.925	0.949	0.971

*: Primary model: Selected for ensemble stability, distributed decision structure, and robustness to class imbalance; **: J48 relies on Province as the sole splitting criterion; All models used identical experimental conditions (30%/70% train–test split: training n = 60, hold-out test set n = 140; random seed = 1; Weka 3.9.7); Metrics for the lottery/drafted class are reported separately, as it represents the minority and policy-critical classification target

Table 4: Ten-fold cross-validation performance

Algorithm	Accuracy (%)	Kappa	Recall (lottery)	F1 (lottery)
Logistic regression	92.00 ± 0.85	0.797 ± 0.021	0.868 ± 0.025	0.852 ± 0.015
Naive Bayes	96.70 ± 0.42	0.917 ± 0.010	0.960 ± 0.006	0.939 ± 0.008
Random tree	93.10 ± 1.76	0.819 ± 0.046	0.834 ± 0.044	0.865 ± 0.035
Random forest	95.95 ± 0.44	0.895 ± 0.011	0.909 ± 0.008	0.923 ± 0.008
J48 decision tree	97.00 ± 0.00	0.923 ± 0.000	0.943 ± 0.000	0.943 ± 0.000

mean ± S.D; over ten repetitions; demographic model

4.3.2. Confusion matrices and class-imbalance sensitivity

Confusion matrices for all five classifiers on the identical hold-out test set (n = 140; 100 voluntary, 40 lottery-drafted) are presented in Table 5. Misclassifications were concentrated in lottery-drafted cases predicted as voluntary (e.g., 9 of 40 for Random Forest and 16 of 40 for Logistic Regression), consistent with the moderate class imbalance.

To verify that this imbalance did not distort the conclusions, a sensitivity analysis was conducted for the Random Forest under 10-fold cross-validation using (i) cost-sensitive learning with a 3:1 penalty on lottery-to-voluntary errors and (ii) class-balanced resampling. Both adjustments raised lottery recall from 0.906 to 0.943 without reducing overall accuracy (95.50% and 96.00%, respectively, vs. 95.50% at baseline), indicating that the substantive findings are robust to the class distribution.

Table 5: Confusion matrices on the hold-out test set

Algorithm	Actual V → Pred. V	Actual V → Pred. L	Actual L → Pred. V	Actual L → Pred. L
Logistic regression	96	4	16	24
Naive Bayes	99	1	12	28
Random tree	100	0	11	29
Random forest	99	1	9	31
J48 decision tree	99	1	3	37

n = 140; V = Voluntary; L = Lottery-drafted

4.3.3. Feature importance

Attribute importance for the Random Forest, computed as the average impurity decrease, is reported in Table 6. Province emerged as the dominant predictor in both variable sets, followed by marital status, education, occupation, and income in the demographic model; an information-gain ranking produced a consistent ordering (Province = 0.681, income = 0.089, occupation = 0.035). In the combined model, economic (0.48), legal (0.46), and risk (0.42) perceptions ranked immediately after the core demographic variables, converging with the tree-structure analysis in Section 4.4, in which cultural, risk, and legal perceptions occupied the upper hierarchy of the integrated tree.

4.3.4. Geographic ablation analysis

Finally, an ablation test was conducted to quantify the contribution of geography. When Province was removed, accuracy fell sharply for both J48 (97.14% → 71.43%; Kappa = 0.00, equivalent to majority-class prediction) and Random Forest (92.86% → 71.43%), confirming that geographic separability constitutes the dominant signal in this dataset for all tree-based models.

Table 6: Random Forest attribute importance

Rank	Demographic model	Combined model (12 variables)
1	Province (0.73)	Province (0.71)
2	Marital status (0.41)	Marital status (0.57)
3	Education (0.30)	Education (0.48)
4	Occupation (0.28)	Economic perception (0.48)
5	Income (0.28)	Legal perception (0.46)
6	–	Occupation (0.45)

Average impurity decrease; –: Not applicable

The preference for Random Forest as the primary model therefore rests not on independence from geography but on three properties: (i) markedly lower fold-to-fold variance than a single tree (S.D. = 0.44 vs. 1.76 in Table 4); (ii) a distributed decision structure in which Province interacts with marital status, education, occupation, and income across

hundreds of nodes of the ensemble (407 Province-related nodes) rather than serving as a single deterministic split, thereby enabling the feature-importance analysis above; and (iii) native compatibility with cost-sensitive extensions for the policy-critical minority class. J48's numerically higher accuracy is acknowledged, but its shallow Province-based structure offers limited interpretive value for recruitment policy beyond geographic targeting.

4.4. Impact of attitudinal factors on decision tree structure

In the final phase of analysis, attitudinal rating scale variables (mean scores for Economic, Social, Psychological, Cultural, Legal, and Risk factors) were integrated into the Random Tree model to assess their structural impact. While the overall predictive accuracy did not significantly exceed the demographic-only models (correlation coefficient = 0.539), the structure of the decision tree changed fundamentally. As shown in the generated tree structure, Cultural Factors (meanCulture) emerged as the root node.

This indicates that cultural perception is the primary splitting criterion for distinguishing between volunteers and draftees when attitudinal data is available. Subsequent nodes in this integrated model included Risk Factors (meanRisk) and Legal Factors (meanLaw), along with the original demographic variables such as occupation and province. This structural shift suggests that while demographic data (occupation, province) provides the most robust statistical prediction, the decision-making process is deeply layered with cultural values and risk perceptions.

The resulting decision tree structure, presented in Fig. 2, reveals a significant shift in the classification logic compared to models relying solely on demographic data.

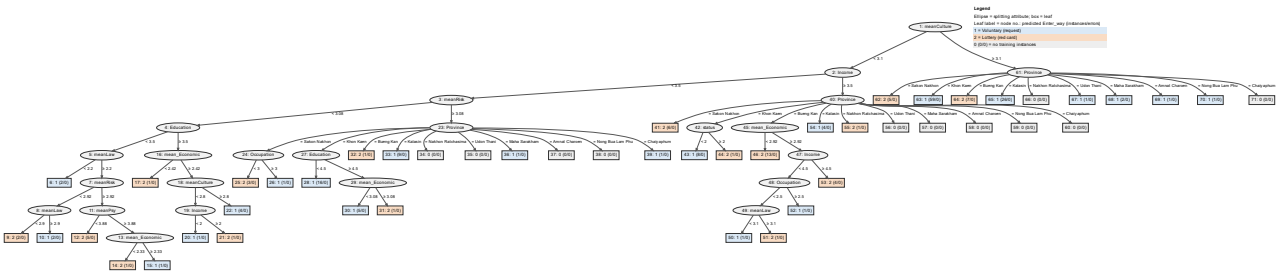


Fig. 2: Diagram of the classification structure for military enlistment results using the random tree model (after adding rating scale variables)

As illustrated in Fig. 2, the integration of attitudinal variables altered the decision tree's topology, reducing the tree size to 71 nodes compared to larger demographic-only trees. The structural analysis highlights the following key findings:

- The Dominance of Cultural Factors: The model selected Cultural Factors (meanCulture) as the root node. This indicates that the perception of military service as a cultural duty, or the influence of traditional values, is the primary splitting criterion with the highest information gain. It suggests that, before economic or risk considerations, an individual's cultural attitude is the strongest determinant of whether they voluntarily enlist or wait for the lottery.
- The Role of Risk and Legal Perception: Following the initial split by culture, the model utilizes Risk Factors (meanRisk) and Legal Factors (meanLaw) as the subsequent branching nodes. This implies that once the cultural baseline is established, individuals differentiate their decisions based on their perception of safety (risk management) and their understanding of conscription laws.
- Interaction with Demographic Variables: While attitudinal factors dominate the upper hierarchy of the tree, demographic variables such as occupation and province reappear at deeper levels (leaf nodes). This demonstrates that while attitudes drive the broad decision-making strategy, socio-economic context remains relevant for fine-tuning the classification of specific subgroups.

Although the inclusion of these variables resulted in a correlation coefficient of 0.539 and did not statistically outperform the accuracy of the demographic-only Random Forest model (92.86%; 96.55% ± 0.16 under 10-fold cross-validation), Fig. 2 provides critical explanatory power. It empirically confirms that the decision-making process is multidimensional, where "soft" factors like cultural values and risk perception act as the primary filters for the decision to enlist.

5. Discussion

5.1. Model performance and classification efficiency

The application of data mining techniques in this study demonstrated that the Random Forest algorithm is highly effective in predicting military enlistment behavior, achieving an accuracy of 92.86% (95.95% ± 0.44 under repeated 10-fold cross-validation) and a Kappa statistic of 0.814. These metrics indicate a strong consistency between the model's predictions and actual enlistment data,

surpassing the performance of the single decision tree (random tree) approach. The high precision and recall values, particularly for the voluntary enlistment class, suggest that the decision-making patterns of volunteers are systematic and distinct, rather than random. This structural clarity confirms that the dataset contains robust explanatory variables capable of distinguishing between those who volunteer and those who wait for the draft lottery.

The baseline comparison further reinforces this conclusion. Although J48 recorded the highest raw accuracy (97.14%), its classification structure collapsed entirely onto Province as the sole splitting criterion, rendering the model informationally equivalent to a simple geographic lookup table. Such a model offers no actionable insight into the behavioral or socio-economic drivers of enlistment and would fail to generalize beyond districts with similar provincial compositions. Indeed, the geographic ablation analysis reported in Section 4.3.4 shows that geographic separability in fact underlies all tree-based models in this dataset; the advantage of Random Forest lies in distributing this signal across many interacting nodes rather than eliminating it, preserving interpretive richness while retaining predictive stability.

At the opposite end of the performance spectrum, logistic regression, despite its widespread use in social science research, achieved the lowest minority-class F1 score (0.706), reflecting its inability to capture nonlinear interactions among demographic variables under conditions of class imbalance. Naïve Bayes performed moderately (Lottery F1 = 0.812), benefiting from its probabilistic treatment of categorical features but remaining constrained by its conditional independence assumption, which is unlikely to hold across correlated demographic variables such as province and occupation.

These results collectively affirm that Random Forest strikes the optimal balance between predictive accuracy, minority-class sensitivity, and interpretive transparency—qualities that are essential when model outputs are intended to inform recruitment policy rather than merely optimize a classification metric.

5.2. The role of socio-economic determinants

The analysis of the demographic-based decision tree revealed that occupation served as the primary root node, followed by income and province. This finding aligns with Maslow's (1943) Hierarchy of Needs, suggesting that economic security is a fundamental driver for enlistment. The prominence of occupation as a splitting criterion reflects the socio-economic reality where individuals with less stable employment or lower income streams view military service as a viable career path offering financial stability and welfare benefits. Furthermore, the significance of the province variable supports the Social Learning Theory, indicating that environmental contexts—such as local economic conditions and regional attitudes toward the military—play a crucial role in shaping individual decisions. This pattern is consistent with findings from the Czech context, where Holcner et al. (2021) demonstrated that military recruitment outcomes are significantly mediated by national labor market conditions—the military must compete with civilian employers for qualified personnel, making occupational and income variables central determinants of enlistment. From a human resource perspective, Ryan and Tippins (2004) emphasized that effective recruitment strategies must align selection tools with organizational goals and applicant perceptions, a principle equally applicable to military recruitment, where occupational fit and geographic targeting are essential for maximizing enlistment rates.

5.3. The influence of cultural and psychological factors

When attitudinal variables were integrated into the model, the decision tree structure shifted significantly, identifying cultural factors (meanCulture) as the new root node. This indicates that while economic factors are foundational, the decision to enlist is ultimately filtered through a lens of values and social norms. This aligns with the Theory of Reasoned Action by Fishbein and Ajzen (1975), which posits that an individual's intention is driven by their attitude toward the behavior and subjective norms. The study suggests that recruits who perceive military service as a distinct honor or a duty to the nation (high cultural value) are more likely to volunteer.

Additionally, the emergence of risk perception (meanRisk) and legal/policy understanding (meanLaw) as subsequent nodes supports Herzberg et al.'s (1993) Motivation-Hygiene Theory. The findings imply that while cultural values act as "motivators," the perception of safety (risk) and the fairness of conscription laws act as "hygiene factors"—their perceived inadequacy (e.g., high risk or unfair laws) could deter voluntary enlistment. It should be emphasized that the contribution of the attitudinal variables is explanatory rather than predictive. Following the well-established distinction between explanatory and predictive modeling (Shmueli, 2010), a variable set may add substantial explanatory value even when it does not raise predictive accuracy.

Although the attitudinal factors did not improve accuracy beyond the demographic-only model, their inclusion fundamentally restructured the decision tree, revealing cultural perception as the primary decision node—an insight that is central to the policy aims of this study. This pattern is unlikely to reflect measurement error, as the instrument demonstrated acceptable content validity (IOC = 0.67–1.00) and high internal consistency (Cronbach's alpha = 0.84–0.90). A more plausible explanation is informational redundancy:

demographic variables such as occupation, province, and income act as partial proxies for the underlying attitudes, absorbing much of the variance that the Likert-scale measures would otherwise contribute.

5.4. Consistency with previous studies

This research extends prior knowledge by empirically demonstrating the hierarchical relationship between these factors through decision tree modeling. It highlights that decision-making is a multidimensional process where "soft" factors (culture, risk perception) interact with "hard" factors (occupation, income) to determine the final enlistment outcome.

Accordingly, all empirical claims in this study are scoped to the 23rd Military District. A clear distinction should be drawn between the analytical framework and the district-specific findings: the framework—combining demographic and attitudinal predictors within explainable decision tree models—is transferable to other regions of Thailand and to other countries operating hybrid or voluntary military recruitment systems, particularly in Southeast Asia and emerging economies. The specific parameter estimates and factor hierarchies reported here, however, reflect the socio-economic and cultural composition of this district and should be re-validated before being applied elsewhere.

6. Conclusions

This study successfully developed a data-driven predictive model to identify the key factors distinguishing voluntary military enlistment from conscription by lottery in Thailand's 23rd Military District. Using data mining techniques implemented in Weka, five classification algorithms were evaluated on a dataset of 200 military personnel, incorporating both demographic variables and six attitudinal constructs (economic, social, psychological, cultural, legal/policy, and risk).

The comparative analysis across five algorithms demonstrated that Random Forest achieved the best overall balance of predictive accuracy and minority-class sensitivity, with an accuracy of 92.86% (95.95% ± 0.44 under repeated 10-fold cross-validation), a Kappa statistic of 0.814, a weighted F1-score of 0.926, and a lottery-class F1-score of 0.861. Although J48 recorded the highest raw accuracy (97.14%), its classification structure relied exclusively on Province as the splitting criterion, offering limited interpretive value beyond geographic stratification. Logistic regression, as the statistical baseline, recorded the weakest minority-class performance (Lottery F1 = 0.706), confirming that nonlinear ensemble methods are better suited to this classification task. These results collectively validate Random Forest as the primary model for enlistment prediction, particularly given its robustness in correctly identifying the lottery/drafted group—the policy-critical minority class. The structural analysis of the decision trees revealed that Occupation and Province are the most influential demographic predictors, reflecting the socio-economic reality that individuals with unstable employment or lower income are more likely to view military service as a viable career path. When attitudinal variables were integrated, cultural factors (meanCulture) emerged as the primary root node, followed by risk and legal perceptions. This structural shift confirms that while demographic factors provide the strongest statistical predictability, the decision-making process is fundamentally filtered through cultural values, risk perception, and legal awareness—factors that operate as primary cognitive determinants before economic considerations come into play.

These findings carry direct implications for Royal Thai Army recruitment policy. Recruitment strategies should be geographically and occupationally targeted, focusing on provinces and occupational groups with historically lower voluntary enlistment rates. Simultaneously, culturally oriented communication campaigns that emphasize national honor, military identity, and risk management transparency are likely to strengthen intrinsic motivation toward voluntary enlistment. The study further demonstrates that explainable machine learning frameworks are well-suited to public-sector behavioral analysis, offering decision-makers interpretable, evidence-based tools for policy formulation in hybrid military service systems.

Despite the satisfactory performance of the predictive models, this research has certain limitations that should be acknowledged:

- Sample Size and Generalization: The study was conducted with a sample size of 200 subjects specifically from the 23rd Military District. Consequently, the findings may reflect specific regional characteristics and may not be fully generalizable to the entire population of Thai males eligible for conscription across different regions.
- Data Characteristics: The inclusion of ordinal data from Likert scale rating questionnaires (attitudinal factors) into the decision tree algorithms can introduce bias in numerical analysis. While useful for interpretation, these variables did not significantly improve the statistical predictive power (accuracy) compared to using demographic variables alone.
- Model Complexity: The generated decision tree models, particularly when including all factors, resulted in a complex structure with approximately 71 nodes. This level of complexity poses a potential risk of overfitting, where the model may be too tailored to the training data, although pruning techniques were applied.

Based on the empirical findings and acknowledged limitations of this study, several policy implications and directions for future research can be proposed. First, given that cultural

factors emerged as the primary decision node in the decision tree structure, the Royal Thai Army should prioritize culturally oriented public relations strategies that emphasize honor, national pride, and the positive societal value of military service in order to strengthen intrinsic motivation toward voluntary enlistment.

In parallel, the results indicate that risk perception and legal understanding play critical roles in shaping enlistment decisions; therefore, transparent communication regarding training safety measures, risk management protocols, and the fairness of conscription laws should be systematically enhanced to reduce uncertainty and anxiety among potential recruits. Furthermore, the strong influence of occupation and province as key demographic predictors suggests that recruitment policies should adopt a more targeted approach by tailoring incentives, such as vocational training programs or career development opportunities, to specific occupational groups and geographic areas identified as having lower voluntary enlistment rates.

Concretely, the decision rules identify general laborers and unemployed youth, jointly about 61% of the sample (46.5% and 14.5%, respectively), as the primary target groups. For these groups, incentive packages should combine vocational skill certification earned during service, guaranteed post-service career pathways, and income-security measures benchmarked against the 5,001–10,000 THB monthly income bracket most common among recruits.

Geographically, recruitment campaigns should be prioritized in provinces with lower voluntary enlistment shares within the district, with the provincial voluntary enlistment rate adopted as a measurable key performance indicator. To address risk perception (meanRisk), the Royal Thai Army should publicize training-safety statistics and simulation-based training protocols through channels popular among men aged 18–25, and to address legal perception (meanLaw), it should communicate the fairness and transparency of conscription procedures at the sub-district level prior to each annual selection period.

From an academic perspective, future research should extend this study by applying and comparing more advanced machine learning techniques, such as neural networks or XGBoost, to validate the robustness and generalizability of the identified predictors across different modeling frameworks. Alzate Vanegas et al. (2022) demonstrated the comparative utility of multiple predictive methods, including logistic regression, Random Forest, and neural networks, for predicting attrition among US Marine Corps personnel and found that model selection substantially affects predictive accuracy. Moreover, extending cross-national comparative analysis would benefit from the work of Margulies (2021), who identified that foreign patron relationships significantly shape military recruitment policies, suggesting that geopolitical dependencies interact with internal labor market dynamics to determine enlistment systems. Additionally, organizational reputation has been identified as a critical factor in recruitment contexts; Cable and Turban (2003) found that organizational reputation serves as a signal of job attributes and generates anticipated pride in membership, which should be considered in designing military branding and public communication strategies. Additionally, the inclusion of broader explanatory variables reflecting contemporary influences, such as trust in military institutions, social media exposure, and political attitudes, may provide deeper insight into the evolving decision-making processes of young adults.

Finally, to enhance external validity, subsequent studies should expand the sample size and geographic coverage beyond the 23rd Military District to capture more diverse socio-economic and cultural contexts, thereby improving the generalizability of data-driven models for military recruitment policy formulation.

List of abbreviations

AI	Artificial intelligence
AVF	All-volunteer force
CF	Cultural factors
EF	Economic factors
F1	F1 score
GDP	Gross domestic product
IOC	Item-objective congruence
J48	J48 decision tree algorithm
KDD	Knowledge discovery in databases
LPF	Legal and policy factors
ML	Machine learning
MLP	Multilayer perceptron
NLP	Natural language processing
OLR	Ordinal logistic regression
PF	Psychological factors
RF	Random forest
RiF	Risk factors
S.D.	Standard deviation
SF	Social factors
SMOTE	Synthetic minority oversampling technique
THB	Thai baht
WEKA	Waikato environment for knowledge analysis
XAI	Explainable artificial intelligence

Acknowledgment

This research project was financially supported by Mahasarakham Business School, Mahasarakham University, Thailand. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

Compliance with ethical standards

Ethical considerations

This study was reviewed and approved by the Mahasarakham University Ethics Committee for Research Involving Human Subjects, Thailand

(Certificate of Approval No. 110-066/2025). Participation was entirely voluntary; all participants were informed of the objectives, procedures, and intended use of the data before data collection, and informed consent was obtained. No personally identifiable information was collected, and all responses were anonymized, treated as strictly confidential, and reported only in aggregate form for academic purposes.

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Alzate Vanegas JM, Wine W, and Drasgow F (2022). Predictions of attrition among US Marine Corps: Comparison of four predictive methods. Military Psychology, 34(2): 147-166. https://doi.org/10.1080/08995605.2021.1978754 **PMid:38536332 PMCid:PMC10013448**

Asal V, Conrad J, and Toronto N (2017). I want you! The determinants of military conscription. Journal of Conflict Resolution, 61(7): 1456-1481. https://doi.org/10.1177/0022002715606217

Baker C and Phongpaichit P (2017). A history of Ayutthaya: Siam in the early modern world. Cambridge University Press, Cambridge, UK. https://doi.org/10.1017/9781108120197

Béhague DP (2008). Psychiatry and military conscription in Brazil: The search for opportunity and institutionalized therapy. Culture, Medicine, and Psychiatry, 32: 194-218. https://doi.org/10.1007/s11013-008-9090-6 **PMid:18347961 PMCid:PMC2841340**

Binková K and Štěpánková E (2023). Understanding the motivation of young adults to enlist and pursue military leadership studies to become professional soldiers. In: Rich M (Ed.), Proceedings of the 19th European Conference on Management Leadership and Governance: 46-54. Academic Conferences International Limited, Reading, UK. https://doi.org/10.34190/ecmlg.19.1.1901

Bjarnegård E, Engvall A, Jitpiromsri S, and Melander E (2023). Armed violence and patriarchal values: A survey of young men in Thailand and their military experiences. American Political Science Review, 117(2): 439-453. https://doi.org/10.1017/S0003055422000594

Breiman L (2001). Random forests. Machine Learning, 45: 5-32. https://doi.org/10.1023/A:1010933404324

Cable DM and Turban DB (2003). The value of organizational reputation in the recruitment context: A brand-equity perspective. Journal of Applied Social Psychology, 33(11): 2244-2266. https://doi.org/10.1111/j.1559-1816.2003.tb01883.x

Cameron S, Dorling G, and Thorpe A (2000). It could be you! Military conscription and selection bias in rural Honduras. Revista Europea de Estudios Latinoamericanos y del Caribe/European Review of Latin American and Caribbean Studies, 68: 47-63.

Campbell S, Greenwood M, Prior S, Shearer T, Walkem K, Young S, Bywaters D, and Walker K (2020). Purposive sampling: Complex or simple? Research case examples. Journal of Research in Nursing, 25(8): 652-661. https://doi.org/10.1177/1744987120927206 **PMid:34394687 PMCid:PMC7932468**

Chen KH, Chiu PL, and Chen MH (2025). Application of machine learning in predicting health perception through military personnel's sense of empowerment. Applied Psychology: Health and Well-Being, 17(1): e12619. https://doi.org/10.1111/aphw.12619 **PMid:39479890**

Cohn LP and Toronto NW (2017). Markets and manpower: The political economy of compulsory military service. Armed Forces & Society, 43(3): 436-458. https://doi.org/10.1177/0095327X16667086

de Almeida YH, Aros SK, and Cherobini D (2026). Use of artificial intelligence to classify failure data from a military aircraft fleet. International Journal of Quality & Reliability Management, 43(2): 453-474. https://doi.org/10.1108/IJQRM-05-2024-0146

Fayyad U, Piatetsky-Shapiro G, and Smyth P (1996). From data mining to knowledge discovery in databases. AI Magazine, 17(3): 37-54.

Fishbein M and Ajzen I (1975). Belief, attitude, intention, and behavior: An introduction to theory and research. Addison-Wesley Publishing Company, Reading, USA.

Han J, Kamber M, and Pei J (2011). Data mining: Concepts and techniques. 3rd Edition, Elsevier, Amsterdam, The Netherlands.

Herzberg F, Mausner B, and Snyderman BB (1993). The motivation to work. Transaction Publishers, New Brunswick, USA.

Holcner V, Davidová M, Neubauer J, Kubínyi Ľ, and Flachbart A (2021). Military recruitment and Czech labour market. Prague Economic Papers, 30(4): 489-505. https://doi.org/10.18267/j.pep.778

Jenkins PR, Caballero WN, and Hill RR (2022). Predicting success in United States Air Force pilot training using machine learning techniques. Socio-Economic Planning Sciences, 79: 101121. https://doi.org/10.1016/j.seps.2021.101121

Maathuis C and Cools K (2025). The role of AI in military cyber security: Data insights and evaluation methods. Procedia Computer Science, 254: 191-200. https://doi.org/10.1016/j.procs.2025.02.078

Margulies MZ (2021). Patrons and personnel: The foreign determinants of military recruitment policies. Security Studies, 30(3): 354-384. https://doi.org/10.1080/09636412.2021.1951832

Maslow AH (1943). A theory of human motivation. Psychological Review, 50(4): 370–396. https://doi.org/10.1037/h0054346 **PMCid:PMC8689106**

Muchlinski D, Siroky D, He J, and Kocher M (2016). Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. Political Analysis, 24(1): 87-103. https://doi.org/10.1093/pan/mpv024

Omrani F, Shad R, Ghaemi M, Abdulzahra SH, and Ziaee SA (2026). Predicting pedestrian crash risk around bus stops: A multi-city random forest approach. Transportation Research Interdisciplinary Perspectives, 36: 101854. https://doi.org/10.1016/j.trip.2026.101854

Plăcintă DD (2022). A machine learning approach to identify the feature importance for admission in the national military high schools. Journal of Social and Economic Statistics, 11(1-2): 118-131. https://doi.org/10.2478/jses-2022-0007

Ryan AM and Tippins NT (2004). Attracting and selecting: What psychological research tells us. Human Resource Management, 43(4): 305-318. https://doi.org/10.1002/hrm.20026

Shmueli G (2010). To explain or to predict? Statistical Science, 25(3): 289-310. https://doi.org/10.1214/10-STS330

Strand S and Berndtsson J (2015). Recruiting the "enterprising soldier": Military recruitment discourses in Sweden and the United Kingdom. Critical Military Studies, 1(3): 233-248. https://doi.org/10.1080/23337486.2015.1090676

Woodruff T, Kelty R, and Segal DR (2025). Revisiting propensity to serve and motivations to enlist: Insights and implications for contemporary military recruitment challenges and research. Armed Forces & Society, 51(2): 482-491. https://doi.org/10.1177/0095327X241259465

Woodruff TD (2017). Who should the military recruit? The effects of institutional, occupational, and self-enhancement enlistment motives on soldier identification and behavior. Armed Forces & Society, 43(4): 579-607. https://doi.org/10.1177/0095327X17695360