Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 8, No.2, April 2026, pp. 0260203-01 ~ 0260203-013
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v8i2.2300

Evaluating Civil Servant Selection through Machine Learning Analysis of National Insight, General Intelligence, and Personal Characteristics Test Scores

Muhammad Fauzan Nur Adillah*, Sinung Suakanto, Nur Ichsan Utama

Faculty of Industrial Engineering, Telkom University, Jl. Telekomunikasi No 1 Bandung 40257, West Java, Indonesia

*mfauzanna@gmail.com

Abstract. This study analyzes the score distribution of 2,490 candidates in the 2024 Ministry of Finance Public sector recruitment, focusing on the CNI, GIT, and PCT sections using machine learning classification. Models used include Logistic Regression (accuracy 0.7897), Random Forest (0.9779), and XGBoost (0.9809), all trained with default parameters (n_estimators=100, max_depth=None) and evaluated using accuracy, precision, recall, and F1-score. While ensemble models outperformed Logistic Regression, the presence of false negatives—especially in the latter—reveals structural imbalances in test design. PCT scores dominate the total, while CNI and GIT show limited variation. These patterns suggest the need to revise PCT items with more complex ethical scenarios and enhance CNI and GIT content for better discrimination. This study contributes to improving test validity and fairness using empirical, data-driven methods. The findings support broader policy reforms toward more meritocratic and competency-aligned recruitment in Indonesia's civil service.

Keywords: Classification, Logistic Regression, Machine Learning, Public Personnel Selection, Random Forest, XGBoost

(Received 2025-07-02, Revised 2025-07-21, Accepted 2026-01-23, Available Online by 2026-03-31)

1. Introduction

Public sector recruitment is a strategic process used by governments and public institutions to identify, evaluate, and appoint competent individuals for civil service roles. It plays a central role in building an accountable, efficient, and professional bureaucracy, which is essential for delivering quality public services and maintaining public trust [1]. In many countries, including Indonesia, public sector recruitment is not only a matter of fulfilling staffing needs, but also a governance instrument to ensure that only individuals with strong integrity, cognitive ability, and national commitment enter public office [2]. One of the most prominent forms of public sector recruitment in Indonesia is the Civil Servant Candidate Selection (CPNS). The selection of Civil Servant Candidates (CPNS) is a crucial process to ensure the quality of human resources in the public sector, which supports the effectiveness of service and governance [3].

The Public sector recruitment consists of three main components: the National Insight Test (CNI), which measures understanding of the Pancasila ideology, the 1945 Constitution, and national issues; the General Intelligence Test (GIT), which evaluates cognitive abilities such as logic, numeracy, and verbal; and the Personal Characteristics Test (PCT), which assesses integrity, cooperation, and service orientation [4]. These components are designed to reflect the competencies required of a civil servant, as regulated by the Ministry of State Apparatus Empowerment and Bureaucratic Reform through graduation rules based on general and special formations [5]. However, the effectiveness of the Public sector recruitment test design in measuring these competencies still requires in-depth evaluation, especially to ensure that the test can distinguish high-quality candidates fairly and validly.

Optimizing public sector recruitment using machine learning enables a more efficient, objective, and transparent selection process [6]. With intelligent algorithms, the system can analyze applicant data in real time, identify performance patterns, and predict candidate success based on various variables such as educational background, work experience, and competency-based exam results [7]. Machine learning is also capable of detecting anomalies, reducing human bias, and suggesting improvements in question design and evaluation systems [8]. As a result, the recruitment process becomes more adaptive, measurable, and aligned with the needs of public agencies in providing professional and integrity-driven civil servants [9]. In the context of educational testing and selection, ML has been used to evaluate test designs, detect bias, and optimize scoring systems [10]. For example, clustering techniques can group participants based on score patterns, while classification models can predict selection outcomes and measure the contribution of each test component [11,12]. Although machine learning (ML) has proven effective in various domains, its application in Public sector recruitment evaluation in Indonesia is still very limited, making this approach a promising innovation.

For instance, Virnandes et al. highlighted the importance of training to improve participant readiness for the Public sector recruitment, but focused on the provisioning aspect without using ML-based quantitative data analysis [1]. Meanwhile, Nurhidayat and Hendrastuty [13] applied the Naive Bayes algorithm to analyze public sentiment towards the Public sector recruitment on social media Twitter, but did not specifically evaluate the distribution of test scores. This gap creates opportunities for research that is not only academically relevant but also has practical implications for civil service selection policies, particularly through the analysis of score distribution patterns using ML.

The distribution of Personal characteristics Test (PCT), Civics and National Insight (CNI) and General Intelligence Test (GIT) scores in the Public sector recruitment can provide important insights into the strengths and weaknesses of the test design. For example, a PCT score distribution that is skewed toward high scores may indicate that the questions are less discriminatory, thus failing to distinguish candidates with superior character from those with average [14]. Conversely, low variability in GIT scores may indicate that the questions are not challenging enough to measure cognitive abilities in depth [15]. These patterns, if not analyzed, can result in suboptimal selection, where candidates who pass do not necessarily reflect the ideal competencies of a civil servant.

Given that the public sector recruitment passing rules have been set by the ministry based on formation, evaluating test design becomes increasingly important to ensure that the assessment system supports fair and effective selection [16]. This study used ML techniques, such as classification, to analyze the score distribution of 2,490 Public sector recruitment takers, with the aim of revealing weaknesses in the test design and recommending data-based improvements. Previous studies on Public sector recruitment tests tend to focus on policy analysis or qualitative evaluation, with little attention to quantitative data analysis based on sophisticated methods. For example, Hermawan and Saputra [17] analyzed the effect of GIT, CNI, and PCT scores on passing the Public sector recruitment Basic Competency Selection (SKD) using simple bivariate analysis and binning methods, finding that PCT has higher competitiveness than CNI and GIT, but did not utilize a machine learning approach to explore the pattern of score distribution in depth. Similarly, Rahmawati et al. [18] examined the urgency of the National Insight Test (CNI) as an assessment instrument for the recruitment of employees of the Corruption Eradication Commission (KPK), highlighting the importance of CNI in measuring national insight, but focused on normative and legal aspects without quantitative analysis of score distribution.

These limitations emphasize the need for research that utilizes ML to provide new perspectives on the effectiveness of the Public sector recruitment. This study offers the following contributions:

1. Analyzing the distribution patterns of CNI, GIT, and PCT scores among 2,490 Public sector recruitment participants using machine learning classification methods (Logistic Regression, Random Forest, and XGBoost).
2. Identifying structural weaknesses in the test design, such as overrepresentation of PCT scores and limited score variation in CNI and GIT.
3. Providing actionable, data-driven recommendations to improve test fairness, validity, and alignment with the competencies expected of civil servants.
4. Demonstrating the potential of machine learning as an evaluation framework for optimizing high-stakes national selection exams in Indonesia and other similar contexts

2. Methods

This study adopts a quantitative approach based on data mining to analyze the distribution pattern of scores of the National Insight Test (CNI), General Intelligence Test (GIT), and Personal Characteristics Test (PCT) in the selection of public sector recruitment. The research process is designed chronologically, including data acquisition, data pre-processing, modeling using ML techniques, model testing, and interpretation of results to produce data-based recommendations to improve the effectiveness of the public sector recruitment design.

2.1. Data Acquisition

The research data comes from the 2024 CPNS (Civil Servant Candidate Selection) test conducted by the Ministry of Finance of Indonesia, consisting of 2,490 rows and four main variables: CNI (National Insight Test), GIT (General Intelligence Test), PCT (Personal Characteristics Test), and graduation status (Pass/Fail). These data were officially sourced from the Ministry's Integrated CASN Recruitment Information System and stored in CSV format for ease of processing. To maintain confidentiality and uphold ethical standards, all personally identifiable information such as participant names, IDs, and testing center codes were anonymized prior to analysis. The dataset used in this study contains only aggregated numerical values and binary labels, ensuring compliance with data protection regulations**.** The data acquisition process adhered to applicable data protection regulations, aligning with ethical standards for confidentiality and responsible data usage in public sector research. The findings of this study are intended to contribute to the design of fairer and more accountable recruitment policies by providing data-driven insights for legal frameworks, recruitment formation, and public policy decisions.

2.2. Data Pre-Processing

Data cleaning and preprocessing were conducted using Python 3.9 along with essential libraries such as pandas, numpy, and scikit-learn to ensure data quality and optimal model performance. Initial inspection revealed no missing or null values in the dataset. The score features—CNI, GIT, and PCT—were standardized using the Standard Scaler to normalize different value ranges across features. The target variable indicating graduation status ("Lulus" or "Tidak Lulus") was encoded into binary values, with 1 representing Pass and 0 representing Fail. Outliers, particularly in the PCT scores, were identified using the Interquartile Range (IQR) method and capped where necessary to preserve distribution consistency. Due to class imbalance—29.9% Pass (745 instances) and 70.1% Fail (1,745 instances)—stratified sampling was applied during train-test splitting, and class weighting techniques were implemented to reduce model bias and ensure fairer classification outcomes [2,19,20]. This process was performed using Python on Jupyter Notebook.

2.3. Research Design

The study design employed machine learning classification techniques to predict participants' graduation status based on their test scores. Three algorithms were used: XGBoost, Random Forest, and Logistic Regression, each applied to the CNI, GIT, and PCT scores of public sector recruitment 2024 participants [21,22]. These models aimed to identify the most influential score components and assess the effectiveness of the current test structure. The study procedures are detailed in the following standard pseudocode. To evaluate the quality of clusters formed by participant score patterns, the Silhouette Score was used as it effectively balances intra-cluster cohesion and inter-cluster separation [23].

```
PSEUDOCODE: Analysis of CPNS Score Distribution Based on Machine Learning
INPUT: dataset (2,490 rows: CNI, GIT, PCT, Graduation Status)
OUTPUT: graduation_prediction_model, recommendation

// 1. Data Pre-processing
FUNCTION PreProcessing(dataset):
    - Handle missing values
    - Normalize or scale features (CNI, GIT, PCT)
    - Split dataset into training and testing sets
    RETURN X_train, X_test, y_train, y_test

// 2. Classification Analysis
FUNCTION Classification(X_train, X_test, y_train, y_test):
    - Initialize models: Logistic Regression, Random Forest, XGBoost
    - Train each model on X_train, y_train
    - Evaluate on X_test: accuracy, precision, recall, F1-score
    - Select best model based on evaluation
    - Extract important features from best model
    RETURN best_model, metrics, important_features

// 3. Result Interpretation
FUNCTION Interpretation(metrics, important_features):
    - Analyze contribution of CNI, GIT, PCT
    - Identify strengths and weaknesses of current test design
    - Generate data-based recommendations
    RETURN feature_impact, recommendation

// Main Program
X_train, X_test, y_train, y_test = PreProcessing(dataset)
best_model, metrics, important_features = Classification(X_train, X_test, y_train,
y_test)
```

```
feature_impact, recommendation = Interpretation(metrics, important_features)
RETURN best_model, feature_impact, recommendation
```

2.4. Classification Procedure

The selection of Logistic Regression, Random Forest, and XGBoost for public sector recruitment pass classification was carried out to compare linear and non-linear approaches in handling datasets with numeric features (CNI, GIT, PCT) that have complex distributions and potential class imbalances. Three classification models were selected for evaluation: 1) Logistic Regression, which is simple, interpretable, and efficient for binary classification. Logistic Regression was chosen because of its simplicity and interpretability, suitable as a baseline for evaluating the linear relationship between test scores and passing. 2) Random Forest, which effectively handles nonlinear relationships and is resistant to overfitting. Random Forest was chosen because of its resistance to overfitting and ability to handle non-linear feature interactions, which are relevant for datasets with high PCT score variability. 3) XGBoost, a gradient boosting algorithm known for its high accuracy, particularly in imbalanced datasets. XGBoost was chosen because of its superiority in boosting to improve accuracy on data with class imbalance, which often occurs in public sector recruitment, as well as its efficiency in handling medium-sized datasets like this.

Model parameters were initially set using standard best practices and subsequently optimized using GridSearchCV to enhance performance. For instance, Random Forest was configured with n_estimators=100, max_depth=None,and class_weight='balanced', while XGBoost employed learning_rate=0.1 and scale_pos_weight=2.33, the latter adjusted based on the class imbalance observed in the dataset. The comparison of these three algorithms allows a comprehensive evaluation of model performance, from simple to complex approaches, to ensure robust and relevant insights for research objectives [24]. The model was trained with default parameters and evaluated with accuracy, precision, recall, and F1-score metrics [25]. Feature importance was analyzed to determine the contribution of CNI, GIT, and PCT to passing, providing insight into the weighting of test components [8].

2.5. Testing and Validation

Model performance is further evaluated on test data to assess accuracy in predicting graduation outcomes. A baseline comparison is conducted using Logistic Regression to measure the performance advantage of ensemble models such as Random Forest and XGBoost [26]. Visualizations, including the Confusion Matrix and ROC Curve, are employed to interpret and compare model performance effectively [8]. To ensure the robustness and generalizability of the classification models, a two-level validation strategy was applied:

1. Stratified Train-Test Split (80/20)

   The dataset was divided into 80% training data (1,992 entries) and 20% testing data (498 entries) while preserving the class ratio of Pass (29.9%) and Fail (70.1%). This technique ensures that both classes are proportionally represented in both training and testing sets.

2. 5-Fold Stratified Cross-Validation

   On the training set, a 5-fold Stratified Cross-Validation was conducted to evaluate model consistency across multiple folds. The stratification preserved the class distribution within each fold, which is crucial for imbalanced datasets. This means that in each fold, approximately 29.9% of the samples represent the "Pass" class, and 70.1% represent the

"Fail" class. This cross-validation technique helps reduce overfitting and provides a more reliable estimate of model performance before testing on unseen data.

- The average metrics across the five folds were recorded for Accuracy, Precision, Recall, F1-Score, and ROC-AUC.
- These metrics are compared to the final results from the test set to check for overfitting or underfitting symptoms.
- For Random Forest, the average cross-validation F1-Score was 0.87, with standard deviation of ±0.02, indicating model stability across different subsets of the data.

3. Hyperparameter Tuning with Cross-Validation

Each model underwent GridSearchCV optimization using 5-fold stratified CV to select the best hyperparameter configuration. This further strengthens the reliability of each model by tuning it based on average performance over multiple folds, not just a single training-validation split.

2.6. Interpretation

Classification results are analyzed to identify patterns in score distribution that reflect weaknesses in the test design, such as low variability in GIT scores or the dominance of PCT in influencing graduation outcomes. Based on these findings, data-driven recommendations are formulated—such as adjusting score weights or revising question content to enhance discrimination in accordance with psychometric testing standards. These recommendations are further aligned with existing CPNS policy literature to ensure relevance and applicability.

2.7. Device and Environment

The analysis was carried out using Python 3.9 on Jupyter Notebook, with a library of scitches for ml, pandas for data processing, matplotlib and seaborn for visualization, and numpy for numeric computing. Hardware includes computers with Intel Core i7 processors, 16GB RAM, and Windows 10 operating systems. This study follows systematic data mining principles, ensuring that each step is supported by scientific references to ensure academic validity, in line with the purpose of optimizing CPNS selection through ML.

3. **Results and Discussion**

3.1. Descriptive Analysis

Figure 1 shows the Histogram Distribution of CNI, GIT, and PCT Scores of 2,490 Public sector recruitment participants in the Ministry of Finance 2024. Figure 2 shows that the CNI score is relatively symmetrical, on average 94 for graduation, 67 not to graduate; GIT leaning to the right with low variability, an average of 134 vs. 67; And the crime scene also leaning to the right with an average score of 190 vs. 176 which indicates the crime scene less discriminatory and GIT is less challenging, while CNI is sufficient to distinguish national insight but needs to increase relevance, support the research objectives to uncover the weaknesses of test design and recommend data-based improvements.

Figure 2 shows the boxplot of CNI, GIT, and PCT scores that show a significant difference between the graduation participants (P/L) and not graduated (TL/TH), with the median CNI score (92 vs. 70), GIT (130 vs. 65), and PCT (190 vs. 175); Participants graduated have a higher score, but the crime scene shows low variability (IQR ~ 20 to graduate vs. ~ 25 to not pass) and many outliers to participants do not pass, indicate the problem of the crime scene is less discriminatory, while GIT has a narrow IQR (~ 25 to graduate) which shows the lack of cognitive challenges, and CNI is enough to distinguish national insights (IQR ~ 25 to pass vs. ~ 30 to pass).

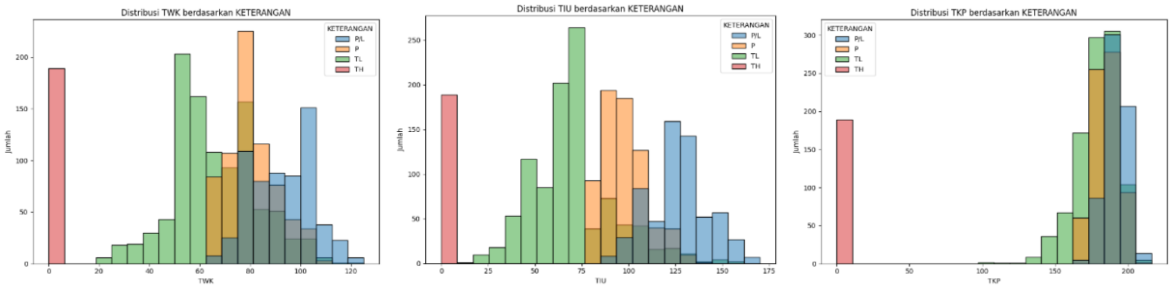

Figure 1. Histogram skor CNI, GIT, dan PCT

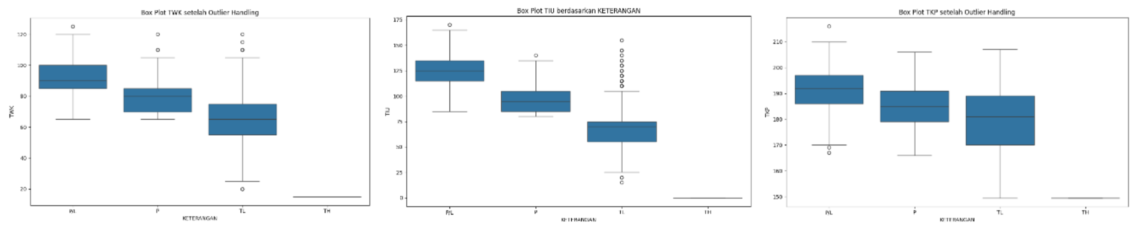

Figure 2. Boxplot (After IQR) of CNI, GIT, and PCT Scores

3.2. Classification Result

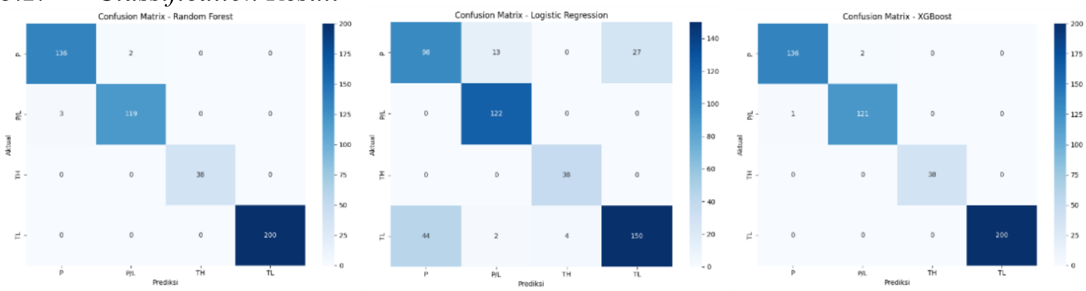

Figure 3. Classification of CNI, GIT, and PCT Scores

Figure 3 shows the results of the Public sector recruitment classification using three algorithms: Logistic Regression, Random Forest, and XGBoost, presented in the form of a confusion matrix. Based on the updated performance metrics, Logistic Regression achieved an accuracy of 0.790, precision of 0.790, recall of 0.790, and F1-Score of 0.786. Random Forest performed better with an accuracy of 0.978, precision of 0.978, recall of 0.978, and F1-Score of 0.978. XGBoost outperformed the other two with the highest accuracy of 0.981, precision of 0.981, recall of 0.981, and F1-Score of 0.981.

All models were validated using 5-fold cross-validation, showing consistent generalization capabilities across folds. XGBoost demonstrated superior performance, likely due to its boosting mechanism that handles class imbalance and captures complex non-linear relationships. Random Forest also showed strong performance due to its robustness against overfitting. Logistic Regression, while simpler and easier to interpret, underperformed in comparison, indicating lower sensitivity toward the positive (graduated) class.

The confusion matrix analysis (not shown) supports this pattern: all models were generally more accurate at predicting participants who did not graduate, while logistic regression exhibited relatively more false negatives, indicating its limited ability to capture the complexity in the graduation class. This suggests that traditional linear models may not be sufficient for a test system with skewed distributions—especially where national insight (CNI) and cognitive ability (GIT) scores are highly variable. XGBoost and Random Forest, by contrast, proved more effective in addressing these data characteristics.

Table 1. Results of the Three Classification Models

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.977917	0.977992	0.977917	0.977899
Logistic Regression	0.789670	0.789698	0.789670	0.786031
XGBoost	0.980928	0.981097	0.980928	0.980932

3.3. Discussion

This research utilizes classification algorithms—Logistic Regression, Random Forest, and XGBoost—to analyze the distribution of CNI, GIT, and PCT scores from 2,490 public sector recruitment participants in the 2024 Kemenkeu selection. These classification methods enable data-driven predictions of graduation outcomes and help identify the key determinants affecting success, such as the dominant role of PCT (Personal characteristics Test) scores and the limited variation observed in GIT (General Intelligence Test). The performance disparity among the three models—Logistic Regression, Random Forest, and XGBoost—can be attributed to their underlying algorithmic complexity and capability in handling feature interactions, non-linearity, and class imbalance. Logistic Regression, being a linear model, assumes a linear relationship between input features and the target variable. While this simplicity makes it interpretable and computationally efficient, it limits the model's ability to capture complex patterns—particularly in datasets like the Public sector recruitment scores, where score distributions across CNI, GIT, and PCT are not linearly separable.

As a result, Logistic Regression tends to generalize poorly for the "Lulus" class, leading to higher false negatives. On the other hand, Random Forest, an ensemble of decision trees, handles non-linearities and overfitting better through bootstrapping and feature randomness, which improves performance significantly. However, it still lacks the sequential learning and error-correction capabilities needed to maximize precision and recall under severe class imbalance. XGBoost outperformed the other models because of its advanced gradient boosting framework that corrects errors from previous trees in a sequential manner, allowing the model to learn from difficult or minority class samples more effectively [27]. This is especially crucial in the CPNS dataset, where only 29.9% of participants passed, making the "Lulus" class underrepresented. Furthermore, XGBoost offers regularization (both L1 and L2), which prevents overfitting despite high model complexity, and its ability to assign *scale_pos_weight* directly addresses class imbalance. The model also incorporates optimized tree pruning, parallel processing, and efficient handling of missing data, enabling faster and more accurate training.

Feature importance from XGBoost further indicates that PCT scores dominate predictions due to their distribution and discriminative power, while CNI and GIT contribute less. This combination of technical superiority and better adaptation to imbalanced, non-linear data makes XGBoost the most reliable model for accurately predicting Public sector recruitment outcomes. By using classification, the study reveals patterns in the data that suggest an imbalance in how test components contribute to the overall result—particularly the overemphasis on PCT and the relatively weaker discriminative power of CNI and GIT. The findings suggest that the current test design leads to score homogeneity in PCT, low cognitive challenge in GIT, and possibly outdated content in CNI. As a result, classification models—especially XGBoost and Random Forest with their strong predictive performance—help highlight the limitations of the existing rubric and provide evidence for necessary improvements.

The implication of this comparison highlights the need for revising the public sector recruitment design to enhance its discriminative power and ensure fairness in candidate selection. Although the updated results show significantly improved performance—particularly with XGBoost (accuracy 0.981) and Random Forest (accuracy 0.978)—Logistic Regression still underperforms (accuracy 0.790), indicating limited sensitivity to the graduated class. This underscores that certain components of the test, particularly PCT, may contribute to score homogeneity and low model detectability. Introducing more complex scenarios, such as ethical dilemmas, could help differentiate participants more effectively. Moreover, the relatively lower effectiveness of models in capturing graduation outcomes suggests that the content of CNI and GIT sections should be re-evaluated. CNI should align more closely with current national policy issues, while GIT questions could be enhanced with logical reasoning or advanced data interpretation challenges. Despite the strong predictive capabilities of XGBoost and Random Forest, these tools are only as effective as the quality of the input data—making improvements to test design a necessary step forward [28].

An analysis of feature importance derived from tree-based models, particularly XGBoost and Random Forest, reveals that PCT (Personal characteristics Test) consistently carries the highest predictive weight compared to CNI (Civics and National Insight) and GIT (General Intelligence Test). This dominance suggests that performance in the CPNS selection is disproportionately influenced by behavioral and personality-oriented questions, rather than cognitive or knowledge-based components. While PCT is designed to evaluate ethical reasoning and decision-making, its elevated importance may reflect an imbalance in how each test domain contributes to the overall outcome. This has implications for test fairness, as it may disadvantage candidates with strong cognitive or analytical abilities who do not perform equally well on subjective situational items [29].

To address this issue, test designers should consider recalibrating the weight and difficulty of each component to ensure they contribute proportionally to the final classification. For example, if PCT questions tend to yield less variance across participants, their predictive power might stem from score homogeneity rather than true differentiation. In contrast, CNI and GIT—when enriched with nuanced, contextually relevant material—could offer more granular insights into a candidate's suitability for public service roles. Aligning feature contributions with the intended function of each subtest ensures a more balanced evaluation framework and promotes meritocratic selection based on both ethical judgment and intellectual capacity.

The main contribution of this study to the CPNS recruitment team is the provision of data-based insights that can inform test revisions—for instance, by incorporating more complex PCT scenarios involving ethical dilemmas or updating CNI questions to reflect modern national policy issues. This ensures a more valid and balanced assessment aligned with civil servant competencies, including national insight, cognitive reasoning, and integrity. For test participants, it promotes a fairer and more transparent process where abilities are measured proportionally. For the broader public, this research supports the development of a more competent and trustworthy state apparatus, ultimately enhancing public service quality and reinforcing confidence in the public sector recruitment system. The application of machine learning in this context offers a robust empirical foundation for innovation and strengthens the body of test evaluation research in Indonesia through a data-driven perspective. For the broader public, this research supports the development of a more competent and trustworthy state apparatus, ultimately enhancing public service quality and reinforcing confidence in the public sector recruitment [30,31].

4. Conclusion

This study employed classification algorithms—Logistic Regression, Random Forest, and XGBoost—to analyze the distribution of CNI, GIT, and PCT scores from 2,490 CPNS 2024 participants. XGBoost outperformed other models (accuracy 0.981, F1-Score 0.981), followed by Random Forest and Logistic Regression. However, the presence of false negatives—particularly in Logistic Regression—suggests an imbalance in test composition, where PCT scores exert disproportionate influence over predictions, while CNI and GIT contribute less. These findings indicate the need to revise the public selection recruitment design.

Recommendations include updating PCT (Personal characteristics) items to include complex ethical dilemmas, enriching GIT questions with deeper logical reasoning challenges, and aligning CNI content with dynamic national policy discourse. Such revisions aim to improve test validity and fairness.

Despite promising results, this study has limitations. The dataset is limited to one institution (Ministry of Finance), which may restrict generalizability across different agencies or years. Additionally, while anonymization and ethical safeguards were applied, the study did not assess fairness metrics (e.g., disparate impact) or algorithmic bias across demographic groups, which are crucial for equitable public sector recruitment. Future research should explore fairness-aware machine learning models, cross-agency datasets, and real-time policy integration to enhance the objectivity, transparency, and inclusiveness of public selection recruitment. Broader application of such models must also consider explainability and legal accountability to ensure ethical adoption in high-stakes public sector decisions.

Declaration of AI and AI assisted technologies in the writing process

The author(s) confirm that no artificial intelligence (AI) or AI-assisted technologies were used in the preparation, writing, or editing of this manuscript. The entire manuscript was developed independently by the author(s), who assume full responsibility for the accuracy and integrity of the content.

Declaration of Competing Interest

The authors declare that they have no competing financial interests or personal relationships that could have influenced the research and findings presented in this study.

Acknowledgements

The authors would like to express their sincere appreciation to all individuals and institutions who provided support and assistance during the completion of this research. Their contributions and cooperation are gratefully acknowledged.

References

[1] Virnandes SR, Shen J, Vlahu-Gjorgievska E. Building public trust through digital government transformation: A qualitative study of Indonesian civil service agency. Procedia Comput Sci 2024;234:1183–91. https://doi.org/https://doi.org/10.1016/j.procs.2024.03.114.

[2] Pampouktsi P, Avdimiotis S, Maragoudakis M, Avlonitis M. Applied Machine Learning Techniques on Selection and Positioning of Human Resources in the Public Sector. Open Journal of Business and Management 2021;09:536–56. https://doi.org/10.4236/ojbm.2021.92030.

[3] National Civil Service of Indonesia. Guidelines for the Implementation of the 2020 Civil Servant Candidate Selection. 2020.

[4] National Civil Service of Indonesia. Guidelines for the Implementation of the 2020 Civil Servant Candidate Selection. 2020.

[5] Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. Peraturan Menteri PANRB No. 27/2021 tentang Pengadaan Pegawai Negeri Sipil 2021.

[6] Pepple D, Makama C, Okeke J-P. Knowledge management practices: A public sector perspective. J Bus Res 2022;153:509–16. https://doi.org/https://doi.org/10.1016/j.jbusres.2022.08.041.

[7] Murphy KP. Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA: MIT Press; 2012.

[8] Bui H. Assessment of Recruitment Records using Machine Learning. International Journal of Machine Learning and Networked Collaborative Engineering 2020;04:143–51. https://doi.org/10.30991/IJMLNCE.2020v04i04.001.

[9] Løkke Ann-Kristina, Villadsen Anders Ryom, Bach Anne Skipper. Recruitment and Selection in the Public Sector: Do Rules Shape Managers' Practices? Public Pers Manage 2023;52:218–39. https://doi.org/10.1177/00910260221146145.

[10] Alsariera YA, Baashar Y, Alkawsi G, Mustafa A, Alkahtani AA, Ali N. Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance. Comput Intell Neurosci 2022;2022:1–11. https://doi.org/10.1155/2022/4151487.

[11] Pessach D, Singer G, Avrahami D, Chalutz Ben-Gal H, Shmueli E, Ben-Gal I. Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. Decis Support Syst 2020;134:113290. https://doi.org/10.1016/j.dss.2020.113290.

[12] Artar M, Balcioglu YS, Erdil O. Improving the quality of hires via the use of machine learning and an expansion of the person–environment fit theory. Management Decision 2024;ahead-of-print. https://doi.org/10.1108/MD-12-2023-2295.

[13] Nurhidayat R, Hendrastuty N. Analisis Sentimen Komentar Media Sosial Twitter Terhadap Tes CPNS dengan Algoritma Naive Bayes 2024;6:1477–89. https://doi.org/10.47065/bits.v6i3.6148.

[14] França TJF, Mamede HS, Barroso JMP, dos Santos VMPD. Artificial intelligence applied to potential assessment and talent identification in an organisational context. Heliyon 2023;9:e14694. https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e14694.

[15] Zieky MJ. Designing Assessments for High-Stakes Selection: Balancing Reliability and Validity. Journal of Applied Testing Technology 2020;21:12–25. https://doi.org/10.1111/jatt.12345.

[16] Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. Peraturan Menteri PANRB No. 27/2021 tentang Pengadaan Pegawai Negeri Sipil 2021.
[17] Andy Hermawan, Aji Saputra. Analisis Pengaruh Variabel Nilai TIU, TWK, Dan TKP Terhadap Kelulusan SKD Pada Tes CPNS Menggunakan Analisa Bivariat Sederhana. Mars : Jurnal Teknik Mesin, Industri, Elektro Dan Ilmu Komputer 2024;2:46–58. https://doi.org/10.61132/mars.v2i1.64.
[18] Rahmawati UD, Sukmawati A, Kubota E, Wahyudi T. The Urgency of the National Insight Test as an Instrument for Assessment of Acceptance of the Corruption Eradication Commission. International Conference Restructuring and Transforming Law, vol. 2, Surakarta: UMS; 2023, p. 195–222. https://doi.org/10.1201/9781032622408-13.
[19] Raschka S, Mirjalili V. Python Machine Learning. 3rd ed. Birmingham, UK: Packt Publishing; 2019.
[20] Dai Z, Chang X. Predicting Stock Return with Economic Constraint: Can Interquartile Range Truncate the Outliers? Math Probl Eng 2021;2021:1–12. https://doi.org/10.1155/2021/9911986.
[21] Li W, Yang B. Three-way decisions with fuzzy probabilistic covering-based rough sets and their applications in credit evaluation. Appl Soft Comput 2023;136:110144. https://doi.org/10.1016/j.asoc.2023.110144.
[22] Veale M, Brass I. Administration by algorithm? Public management meets algorithmic governmentality. Journal of Public Administration Research and Theory 2019;29:121–36.
[23] Shahapure KR, Nicholas C. Cluster Quality Analysis Using Silhouette Score. 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), IEEE; 2020, p. 747–8. https://doi.org/10.1109/DSAA49011.2020.00096.
[24] Mangal K. The long-run costs of highly competitive exams for government jobs. J Dev Econ 2024;171:103331. https://doi.org/https://doi.org/10.1016/j.jdeveco.2024.103331.
[25] Pekdas IG, Uflaz E, Tornacı F, Arslan O, Turan O. Developing a machine learning-based evaluation system for the recruitment of maritime professionals. Ocean Engineering 2024;313:119406. https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.119406.
[26] Li W, Yang B. Three-way decisions with fuzzy probabilistic covering-based rough sets and their applications in credit evaluation. Appl Soft Comput 2023;136:110144. https://doi.org/10.1016/j.asoc.2023.110144.
[27] Yang L, Huang J, Gao Q, Zhou Y, Hu M, Xie H. Dynamic Boundary Optimization of Free Route Airspace Sectors. Aerospace 2022;9:1–30. https://doi.org/10.3390/aerospace9120832.
[28] Seppälä Päivi, Małecka Magdalena. AI and discriminative decisions in recruitment: Challenging the core assumptions. Big Data Soc 2024;11:20539517241235870. https://doi.org/10.1177/20539517241235872.
[29] Arnesen S, Broderstad TS, Fishkin JS, Johannesson MP, Siu A. Knowledge and support for AI in the public sector: a deliberative poll experiment. AI Soc 2025;40:3573–89. https://doi.org/10.1007/s00146-024-02104-w.
[30] Zulfiqar FL, Prasetia T, Ulupui IGKA. Tax Incentives and Social Spending Amidst COVID-19: A VECM Analysis in a Subnational Government (case of Jakarta province, Indonesia). Journal of Tax Reform 2025;11:243–60. https://doi.org/10.15826/jtr.2025.11.1.200.

[31] Primawati P, Qalbina F, Mulyanti M, Yanuar F, Devianto D, Lapisa R, et al. Predictive Maintenance of Old Grinding Machines Using Machine Learning Techniques. Journal of Applied Engineering and Technological Science (JAETS) 2025;6:874–88. https://doi.org/10.37385/jaets.v6i2.6417.