Benchmarking Ensemble Learning vs. Deep Learning for GovTech Indices: A Comparative Analysis of GBR and MLP with a Focus on the UAE

Mohammad Zhair Swalha
Rochester Institute of Technology
Dubai, United Arab Emirates
ariful1251996@gmail.com

Dr. Ayman Ibrahim
Department of Graduate Programs & Research
Rochester Institute of Technology
Dubai, United Arab Emirates
ayman.ibrahim@rit.edu

https://doi.org/10.63962/OOWH6236

Abstract—As governments increasingly rely on composite indices to align national strategies with global best practices, the need to understand and validate the structural composition of these metrics has grown. This study evaluates the ability of four machine learning algorithms Decision Tree (DT), Random Forest (RF), Gradient Boosting Regressor (GBR), and Multi-Layer Perceptron (MLP) to model the World Bank's GovTech Maturity Index (GTMI) within a controlled analytical setting and capture the underlying non-linear relationships embedded within its composite structure.

Rather than treating the task as a conventional predictive problem, the study adopts a structural modeling and benchmarking perspective to assess how different machine learning approaches approximate complex governance indices under consistent input conditions. Using a dataset of 198 economies, the results demonstrate that ensemble learning methods significantly outperform deep learning approaches in structured tabular governance datasets [1].

These findings highlight the importance of aligning model selection with data characteristics and provide insights into the suitability of ensemble methods for modeling composite public sector indices.

Keywords— GovTech Maturity Index, Structural Modeling, Machine Learning Benchmarking, Ensemble Learning, Gradient Boosting Regressor, Deep Learning, Public Sector Analytics

I. INTRODUCTION

A. Background

Governments worldwide are accelerating digital transformation initiatives to enhance public service delivery, administrative efficiency, and transparency. As digital government capabilities expand, international organizations increasingly rely on composite indices [2] to benchmark national progress in digital governance. One of the most widely recognized frameworks is the World Bank's GovTech Maturity Index (GTMI), which evaluates digital government capabilities across multiple dimensions including Core Government Systems, Public Service Delivery, Digital Citizen Engagement, and GovTech Enablers.

Digital governance indicators play an increasingly important role in shaping national digital transformation strategies and guiding policy reforms. Countries with higher levels of digital maturity tend to demonstrate stronger institutional capacity, improved service delivery mechanisms, and greater integration of digital technologies within public sector operations.

At the same time, advances in artificial intelligence and machine learning have opened new opportunities for analyzing complex governance datasets. Machine learning techniques are increasingly applied in public administration research to model structured relationships in multidimensional data and support evidence-based policy analysis..

B. Research Gap

Despite the growing importance of digital governance indices such as the GovTech Maturity Index, relatively few studies have examined the internal structural relationships of these composite indicators using machine learning techniques. Most existing research treats digital governance indices as descriptive benchmarking tools rather than analytical systems whose underlying relationships can be systematically modeled using data-driven approaches.

Recent research highlights the growing role of artificial intelligence and machine learning in assessing digital government maturity and supporting data-driven governance decisions, particularly in structured public sector datasets. [3]

Furthermore, the rapid expansion of deep learning architectures has led to the widespread assumption that neural networks outperform traditional machine learning algorithms across most analytical tasks. While deep learning models have achieved remarkable success in domains such as computer vision and natural language processing, their advantages are not universal and may depend heavily on dataset characteristics.

Governance datasets typically consist of structured tabular indicators with relatively small sample sizes, which differ significantly from the large-scale datasets commonly used in deep learning applications. In such environments, ensemble learning methods such as Random Forest and

Gradient Boosting often provide more reliable modeling performance compared to deep learning approaches..

C. Research Contribution

This study addresses the identified research gap by applying machine learning models to analyze and model the World Bank's GovTech Maturity Index (GTMI) within a controlled analytical framework, focusing on evaluating the modeling capability of different algorithms in structured governance datasets.

Specifically, the study benchmarks four machine learning models: Decision Tree (DT) [4], Random Forest (RF) [5], Gradient Boosting Regressor (GBR), and Multi-Layer Perceptron (MLP). By comparing ensemble learning techniques with a neural network architecture, the study provides empirical evidence regarding the suitability of different machine learning approaches for analyzing structured tabular data in the public sector context.

The study makes three primary contributions. First, it provides an empirical benchmark comparing ensemble learning and deep learning models in modeling digital government maturity indicators within a consistent analytical setting. Second, it demonstrates that Gradient Boosting effectively captures the structural relationships underlying the GovTech Maturity Index across 198 economies, highlighting its suitability for structured governance datasets. Third, it offers methodological insights for researchers and policymakers on selecting appropriate models for structured data environments in digital governance analytics.

Accordingly, this study addresses three key questions: how effectively ensemble learning models the structural composition of the GTMI, which digital pillars represent the primary structural drivers within the index, and how the UAE's digital maturity profile aligns with these structural characteristics.

The study is conducted within a controlled modeling framework, where the objective is to evaluate algorithmic performance under consistent input conditions rather than to generate fully independent predictive outcomes. This enables fair comparison of model behavior and isolates their capability to capture structured relationships in governance data.

Extending this framework to incorporate independent exogenous variables (e.g., socio-economic and governance indicators) represents an important direction for future research to enhance generalization and support real-world GovTech forecasting applications.

II. METHODOLOGY

A. Dataset Description

This study utilizes the World Bank GovTech Maturity Index (GTMI) dataset, which evaluates digital government development across 198 economies worldwide. The GTMI framework measures national digital governance capabilities across four core dimensions: Core Government Systems (CGSI), Public Service Delivery (PSDI), Digital Citizen Engagement (DCEI), and GovTech Enablers (GTEI) [1].

These dimensions collectively capture the technological infrastructure, institutional capacity, and service delivery mechanisms that support digital government transformation. The dataset includes multiple indicators representing the digital maturity of national governments and serves as a comprehensive benchmark for assessing digital governance readiness.

For the purposes of this study, the GTMI score is modeled within a controlled analytical framework, where the underlying indicators representing the four GTMI pillars are used as input features. This setup enables the evaluation of how different machine learning models capture structured relationships within composite governance data under consistent input conditions. The dataset provides a structured tabular environment suitable for benchmarking the performance of different machine learning algorithms in modeling governance-related indicators.

figure 1 provides a structured overview of the analytical workflow adopted in this study

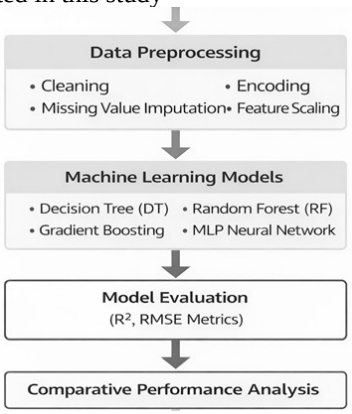


Figure 1. Structured Analytical Workflow

B. Data Preprocessing

Prior to model training, basic preprocessing steps were applied to ensure data consistency and analytical robustness. Non-informative attributes such as country identifiers were removed, while categorical variables were encoded using label encoding techniques. Missing values were handled using standard imputation methods to maintain dataset completeness.

Feature scaling was applied where required, particularly for neural network models that are sensitive to feature magnitude. These preprocessing steps ensure that the dataset is suitable for structured modeling and benchmarking across different machine learning algorithms, and support stable and consistent model training [2].

C. Machine Learning Models

To evaluate the modeling capability of machine learning algorithms in structured governance data, four supervised models were implemented: Decision Tree (DT), Random Forest (RF), Gradient Boosting Regressor (GBR), and Multi-Layer Perceptron (MLP).

Decision Tree models represent one of the simplest machine learning approaches for regression tasks. They partition the dataset into hierarchical decision rules and are capable of capturing nonlinear relationships between variables in a transparent and interpretable manner.

Random Forest and Gradient Boosting [6] represent ensemble learning techniques, which combine multiple decision trees to improve modeling performance and reduce overfitting. Random Forest constructs multiple trees using bootstrap sampling, while Gradient Boosting sequentially builds trees that correct the errors of previous models, enabling it to capture complex structured relationships in tabular data.

In contrast, the Multi-Layer Perceptron represents a neural network architecture capable of modeling nonlinear relationships between variables through layered transformations of input features. However, such models typically require larger datasets and may be less suited to structured tabular environments with limited sample sizes.

The inclusion of both ensemble learning models and a neural network architecture enables a comparative evaluation of model behavior and suitability across different machine learning paradigms within a controlled analytical setting.

While advanced ensemble variants such as XGBoost and LightGBM were considered, the Gradient Boosting Regressor was selected as a representative implementation of boosting techniques, given the structured and relatively small size of the GTMI dataset (N=198).

D. Model Evaluation

To evaluate the modeling performance of the implemented machine learning models, two standard regression evaluation metrics were used: the Coefficient of Determination (R^2) and the Root Mean Squared Error (RMSE). The R^2 metric measures the proportion of variance in the dependent variable captured by the model, indicating how well the model approximates the observed data within the analytical setting.

In contrast, RMSE quantifies the average magnitude of errors, providing a measure of the difference between model outputs and observed values. These metrics are widely used in regression-based machine learning studies to assess model performance and enable consistent comparison across different algorithms [7].

To ensure robust and reliable model evaluation, the dataset was divided into training and testing subsets (75%/25%). In addition, 5-fold cross-validation was applied during model training to assess model stability and reduce potential bias in performance estimation.

The evaluation focuses on comparing model performance using standardized metrics (R^2 and RMSE) under consistent conditions, ensuring a fair basis for benchmarking different algorithms.

The interpretation of results emphasizes comparative model behavior and relative performance rather than absolute predictive capability.

All models were implemented under consistent experimental settings to ensure fair comparison.

III. RESULTS

A. Model Performance Comparison

The experimental results provide a comparative evaluation of the modeling performance of the implemented machine learning models in analyzing the World Bank's GovTech Maturity Index (GTMI) within a controlled analytical setting. The objective of this analysis is to assess how effectively different algorithms capture the structured relationships between GTMI components and the overall digital maturity score.

To compare model performance, two standard regression evaluation metrics were used: the Coefficient of Determination (R^2) and the Root Mean Squared Error (RMSE). These metrics provide complementary insights into model performance, enabling a consistent comparison across the evaluated algorithms..

Table 1. Comparative Performance of Machine Learning Models

Model	R^2	RMSE
Decision Tree (DT)	0.96	0.021
Random Forest (RF)	0.989	0.012
Gradient Boosting Regressor (GBR)	0.993	0.009
Multi-Layer Perceptron (MLP)	0.942	0.028

As shown in Table 1, the Gradient Boosting Regressor (GBR) achieved the highest and most stable modeling performance among all evaluated models, producing an R^2 value of 0.993 and the lowest error level. This result indicates that the model effectively captures the structured relationships underlying the GovTech Maturity Index within the analytical framework.

The model performance remained consistent across cross-validation folds, indicating stable and robust behavior.

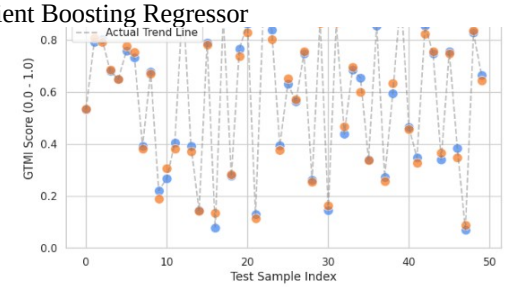
These findings provide empirical evidence that ensemble learning methods may outperform deep learning architectures when applied to structured tabular governance datasets.

The Random Forest model also demonstrated strong modeling performance, confirming the effectiveness of ensemble learning techniques in structured data environments. In contrast, the Decision Tree model produced moderate results due to its relatively simple structure.

The most notable finding is the performance of the Multi-Layer Perceptron (MLP) [8]. Despite representing a more complex deep learning architecture, the model produced comparatively lower performance. This result highlights an

important methodological insight: in structured tabular datasets with limited observations, ensemble learning models may be more suitable than deep neural network architectures.

Figure 2. Actual vs. Predicted GTMI Scores using the Gradient Boosting Regressor



The Figure 2 illustrates the strong modeling capability of the Gradient Boosting model by comparing the observed GTMI values with the model-generated outputs. The close alignment between modeled and observed values indicates that the model effectively captures the underlying nonlinear relationships within the structured governance dataset [9].

This alignment should be interpreted as the model's ability to approximate the structural composition of the index rather than its capability to predict unseen outcomes.

Overall, these results provide empirical support for the suitability of ensemble learning approaches in modeling structured governance datasets and demonstrate the effectiveness of Gradient Boosting in capturing complex relationships within composite digital governance indicators.

B. Structural Drivers of Digital Maturity

Beyond model performance, the feature importance analysis represents a form of global interpretability, capturing the relative structural importance of different components within the GTMI as modeled by the Gradient Boosting Regressor (GBR). The results indicate that the Core Government Systems Index (CGSI) is the most influential feature within the model context, accounting for 51.09% of the relative importance score. The Public Service Delivery Index (PSDI) accounts for 25.37%, followed by Digital Citizen Engagement (DCEI) at 16.54%, and GovTech Enablers (GTEI) at 6.75%.

These findings highlight the relative contribution of different dimensions within the modeled structure of the index. In particular, foundational system-level components—such as digital identity frameworks and interoperability platforms—are more strongly represented within the model compared to citizen-facing engagement metrics.

It is important to note that these results reflect statistical associations within the dataset and should not be interpreted as causal relationships.

C. Application: The UAE Context

Applying the GBR-based modeling framework to the United Arab Emirates (UAE) provides a contextual interpretation of its position as a leading digital governance performer. The UAE achieves an overall GTMI score of 0.9609, placing it fourth globally out of 198 economies.

The UAE's performance shows strong alignment with the model-derived feature importance distribution, with high scores in the most influential components identified within the analytical framework, particularly Core Government Systems (0.9751) and Public Service Delivery (0.9660).

Deviation analysis indicates that the UAE outperforms the global average across these core dimensions, reflecting a strong emphasis on foundational digital infrastructure and integrated government systems. This alignment supports the interpretation that infrastructure-focused digital strategies—such as those implemented by national digital transformation entities—are associated with higher levels of digital maturity within the context of the GTMI framework.

IV. DISCUSSION

A. Interpretation of Model Performance

The results demonstrate clear differences in modeling performance among the evaluated machine learning models. Ensemble learning approaches, particularly the Gradient Boosting Regressor, achieved the highest performance in modeling the GovTech Maturity Index within the analytical framework. This finding indicates that the structured relationships between the GTMI components and the overall maturity score can be effectively captured using tree-based ensemble models.

The Random Forest model also demonstrated strong modeling performance, further confirming the suitability of ensemble learning methods for structured governance datasets. In contrast, the Multi-Layer Perceptron produced comparatively lower performance despite its higher model complexity.

The relatively weaker performance of the Multi-Layer Perceptron may be attributed to the limited size of the dataset and the structured tabular nature of the indicators, which are better suited for tree-based ensemble algorithms, as deep neural networks generally require larger datasets to achieve optimal performance.

This observation is consistent with prior findings indicating that neural networks tend to underperform in small structured datasets due to their sensitivity to data size and feature representation.

These results suggest that, in structured tabular datasets such as digital governance indicators, ensemble learning models may provide more stable and reliable modeling outcomes than deep neural network architectures.

B. Implications for Machine Learning in Digital Governance

The findings contribute to the growing literature on artificial intelligence applications in public sector analytics.

While deep learning models have shown strong performance in large-scale and unstructured datasets, their advantages may not fully extend to structured governance datasets with limited observations.

In such contexts, ensemble learning techniques provide a more suitable analytical approach for modeling digital government maturity indicators and analyzing structured governance data. These results emphasize the importance of aligning model selection with dataset characteristics rather than relying solely on model complexity.

C. Policy Implications

From a policy perspective, the feature importance results provide insights into the relative contribution of different dimensions within the GTMI framework. While contemporary public sector discourse often emphasizes citizen-facing applications and digital engagement (DCEI), the modeling results indicate a stronger representation of foundational system-level components (CGSI) within the analytical structure of the index.

In this context, investments in integrated core systems—such as national digital identity platforms and interoperability frameworks—are associated with higher levels of digital maturity. The UAE's high global ranking (0.9609) reflects strong performance in these core dimensions, particularly in Core Government Systems and Public Service Delivery.

This observation suggests that strengthening foundational digital infrastructure, alongside citizen-facing services, may support the advancement of digital governance maturity. However, these interpretations are based on observed associations within the dataset and should not be interpreted as causal relationships.

V. CONCLUSION

This study examined the application of machine learning techniques to model the World Bank's GovTech Maturity Index and evaluate the performance of different algorithms in analyzing digital governance indicators within a controlled analytical framework. Using a dataset covering 198 economies, four machine learning models were compared: Decision Tree, Random Forest, Gradient Boosting Regressor, and Multi-Layer Perceptron.

The results demonstrate that ensemble learning approaches, particularly the Gradient Boosting Regressor, achieved the highest modeling performance in capturing structured relationships within the GTMI framework. In contrast, the neural network model showed comparatively lower performance despite its higher algorithmic complexity. These findings highlight that model effectiveness in digital governance analytics depends strongly on dataset characteristics, particularly when working with structured tabular data.

Furthermore, applying this modeling framework to the UAE highlights its strong global standing (ranked fourth worldwide), reflecting high performance in core government systems and public service delivery. This alignment provides useful insights for policymakers seeking to strengthen foundational digital capabilities within the broader context of digital transformation.

This study contributes to the growing body of research on artificial intelligence applications in digital government by providing an empirical comparison between ensemble learning and deep learning models in a structured governance dataset.

Future research will extend this work by incorporating independent exogenous variables and applying advanced explainable artificial intelligence techniques to further investigate the structural characteristics of digital government maturity and enhance model generalization..

To further enhance interpretability, future research will incorporate SHAP (SHapley Additive exPlanations) to analyze feature-level contributions at the country level beyond the global structural importance identified in this study.

REFERENCES

[1] World Bank, "GovTech Maturity Index 2022 Update: Trends in Public-Sector Digital Transformation," World Bank Group, Washington, DC, USA, 2023. [Online]. Available: https://www.worldbank.org/en/programs/govtech

[2] OECD, "Digital Government Index: 2019 Results," OECD Public Governance Policy Papers, Paris, France, 2020. doi: https://doi.org/10.1787/4de9f5bb-en

[3] M. Alfadhli, et al., "AI-driven digital maturity assessment models for government organizations: A cross-country analysis," Government Information Quarterly, 2025: https://www.worldbank.org/en/programs/govtech/gtmi

[4] T. Janowski, "Digital government evolution: From transformation to contextualization," Government Information Quarterly, vol. 32, no. 3, pp. 221–236, 2015. doi: https://doi.org/10.1016/j.giq.2015.07.001

[5] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. doi: https://doi.org/10.1023/A:1010933404324

[6] H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001. doi: https://doi.org/10.1214/aos/1013203451

[7] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY, USA: Springer, 2009. doi: https://doi.org/10.1007/978-0-387-84858-7

[8] Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org

[9] United Nations Department of Economic and Social Affairs (UN DESA),

[10] "United Nations E-Government Survey 2022: The Future of Digital Government," New York, NY, USA: United Nations, 2022.

[11] [Online]. Available: https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2022