

Article

Quantitative Analysis of Satisfaction with Chinese Local Government Digital Public Service Policies Using XGBoost Algorithm

Qin Hu ^{1,†}, Bin Yang ^{2,†} and Shengli Dai ^{1,*}¹ School of Public Administration and Emergency Management, Jinan University, Guangzhou 510632, China; qinhu@stu2022.jnu.edu.cn² School of Government, Sun Yat-sen University, Guangzhou 510275, China; yangb256@mail.sysu.edu.cn

* Correspondence: daishengli0610@jnu.edu.cn

† These authors contributed equally to this work.

Abstract

With the development of digital technology, although local governments have been using digital means to improve the quality of public services, traditional statistical methods have limitations in processing complex, high-dimensional data and revealing factors influencing policies. This paper used the XGBoost algorithm to construct a satisfaction prediction model, leveraging its advantages in handling nonlinear relationships and feature interactions to assist government decision-making through prediction and feature analysis. This study is based on questionnaire surveys and public data, and the optimal configuration of the model was determined through preprocessing and parameter tuning. Experiments showed that the proposed model outperforms other models in terms of prediction accuracy, robustness, efficiency, and cross-scenario applicability. Through empirical analysis, this study shows that the XGBoost model has significant advantages in predicting local government digital public service policy satisfaction. Its mean square error (MSE) is only 0.056, which is 37.1% lower than the traditional linear regression model. This means that XGBoost can more accurately capture the complex nonlinear relationships that influence public satisfaction.

Keywords: XGBoost algorithm; digital public service policy; satisfaction analysis; gradient boosting decision tree; model optimization



Academic Editor: Vladimír Bureš

Received: 29 July 2025

Revised: 12 September 2025

Accepted: 12 September 2025

Published: 15 September 2025

Citation: Hu, Q.; Yang, B.; Dai, S. Quantitative Analysis of Satisfaction with Chinese Local Government Digital Public Service Policies Using XGBoost Algorithm. *Systems* **2025**, *13*, 808. <https://doi.org/10.3390/systems13090808>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As digital technology develops rapidly, the digital transformation of local governments providing public services has been an important means to enhance governance capabilities [1,2]. Digital public services are increasingly significant in the modern governance system by optimizing resource allocation, improving service efficiency, and enhancing public experience [3–5]. However, scientific evaluation of policy implementation effectiveness has always been a core issue in this field [6]. At present, most local governments obtain public satisfaction data through questionnaire surveys and other means, but traditional analysis methods mostly rely on simple statistics or linear models, which makes it difficult to reveal complex data relationships and key influencing factors [7,8]. This deficiency may not only lead to one-sided evaluation results but also limit the pertinence and effectiveness of policy optimization. Therefore, it is urgent to apply more refined and data-driven analysis methods to effectively explore the deep factors affecting policy satisfaction and provide reliable support for policy optimization [9,10].

In response to the above problems, this study aims to build a practical and efficient satisfaction prediction model to provide data-driven decision support for policy optimization. Although generative artificial intelligence methods have shown strong capabilities in many fields, in the specific scenario of digital public service policy satisfaction prediction, the XGBoost model still has irreplaceable advantages due to its excellent interpretability, efficiency, and stability. This study not only empirically verifies the effectiveness of the XGBoost model in this scenario, but also deeply explores the core factors that affect satisfaction, providing a key basis for policymakers to implement precise policies.

The main contributions of this paper are as follows: (1) Innovative application of the XGBoost algorithm in the field of public management. The XGBoost algorithm was introduced into the prediction of local government digital public service policy satisfaction, effectively processing high-dimensional complex data, making up for the shortcomings of traditional public management analysis methods in data processing, improving the scientific rigor and accuracy of policy evaluation, and providing reliable data support for public management decision-making. (2) Policy factor mining based on feature importance analysis. The feature importance analysis function of the XGBoost algorithm was used to deeply mine and sort the key policy factors that affect public service satisfaction, which helps to clarify the core influencing factors in public management, provide a scientific basis for policy optimization and resource allocation, and enhance the pertinence and effectiveness of public management decisions. (3) Model robustness and applicability verification and guidance for public management practice. By verifying the robustness and applicability of the XGBoost model under different noise interference and data subsets, it is proven that this model has high stability and adaptability in complex and changing public management scenarios. This provides a theoretical guarantee for the widespread application of this model in the field of public management and also provides a powerful decision-making tool for local governments to formulate and adjust public service policies under different circumstances. Compared with international research, this study is innovative in three aspects. First, although XGBoost has been applied in fields such as finance and healthcare, this is its first systematic application in the analysis of digital public service policy satisfaction. European scholars such as Kuhlmann and Heuberger mainly use traditional statistical methods to analyze digital transformation, while Scupola focuses on qualitative research methods. Second, the feature engineering framework constructed in this study integrates multi-dimensional indicators such as policy implementation strength, fiscal investment, and service coverage. It is more comprehensive than existing research methods that only focus on a single dimension. Third, this study not only focuses on prediction accuracy, but also provides specific guidance for policy optimization through feature importance analysis. This method of combining high-precision prediction with policy guidance has not yet been systematically applied in international digital government research.

The core problem that this study aims to solve is how to effectively identify and quantify the key factors that affect the satisfaction with local government digital public service policies and to build a high-precision prediction model to guide policy optimization. There are three obvious gaps in existing research: first, it is difficult to handle nonlinear relationships in high-dimensional complex data with traditional statistical methods; second, existing research lacks in-depth analysis of the interactions among policy elements; and third, most studies only focus on the surface factors affecting satisfaction and fail to reveal the deep-seated mechanism of action. To address these gaps, this study proposes the following research objectives: (1) to construct a satisfaction prediction model based on the XGBoost algorithm to improve prediction accuracy; (2) to identify key policy elements through feature importance analysis; and (3) to verify the robustness and applicability of the model in different scenarios. This paper first reviews related research and then

introduces in detail the construction process of the satisfaction prediction model based on XGBoost, including data sources, preprocessing methods, and key parameter settings; then, the experimental design and result analysis are presented; and finally, the theoretical and practical significance of the research findings are discussed.

2. Related Work

The study of satisfaction with digital public service policies has become an important topic in public administration [11,12]. As digital technology develops, it has become a global trend for local governments to improve public service quality through digital means [13,14]. At present, the evaluation of satisfaction with digital public services mainly relies on traditional methods such as questionnaires and interviews [15,16] and focuses on the impact of factors such as service quality, information disclosure, and technological innovation on satisfaction [17–19]. Traditional methods mainly rely on linear regression and simple statistical analysis, but they are unable to cope with high-dimensional data and nonlinear relationships [20,21] and cannot fully reveal the complex interactions between influencing factors, resulting in insufficient precision of policy improvement recommendations [22–24]. Currently, research analyzing digital public service policy satisfaction primarily focuses on Europe, the United States, and some Asian countries. In the United States, some studies have used meta-analysis to examine the relationship between citizen satisfaction and public service quality. In Europe, Gasco-Hernandez et al. analyzed the transformation of public services in smart cities through case studies. In Asia, some studies have examined the evaluation methods for digital public services in China. These studies primarily use traditional statistical methods such as linear regression and structural equation modeling. While these methods can reveal some influencing factors, they struggle to capture complex nonlinear relationships and feature interactions. When the interaction effect between service efficiency and policy transparency is considered, the explanatory power of traditional linear models decreases significantly. In recent years, a few studies have begun to apply machine learning methods, using sentiment analysis techniques to assess user satisfaction with government mobile apps and supervised learning methods to evaluate public service applications. However, these studies are mostly limited to specific service areas, lack a systematic analysis of overall policy satisfaction, and fail to fully utilize the advantages of advanced algorithms such as XGBoost in feature importance analysis. Therefore, systematically applying the XGBoost algorithm to analyze satisfaction with local government digital public service policies remains a relatively novel research direction.

To overcome these limitations, machine learning, especially decision trees and ensemble learning methods, has gradually become a research hotspot in satisfaction analysis [25,26]. Machine learning can process high-dimensional data, mine nonlinear relationships, and reveal complex feature interactions, providing more precise policy optimization recommendations. As an efficient ensemble learning method, the XGBoost algorithm has been widely applied to various data analysis tasks due to its advantages in processing complex data and high prediction precision [27]. Despite this, the application of XGBoost in digital public service policy satisfaction analysis is still relatively rare, and existing research is mostly concentrated in finance, medical care, education, etc. [28–30]. Notably, recent research indicates that digital evaluation is reshaping the landscape of public policy evaluation, and traditional evaluation methods appear insufficient in the complex and dynamic digital public service environment. Effective digital evaluation requires the ability to process multi-source heterogeneous data, capture nonlinear relationships, and provide interpretable analytical results [31], which is highly consistent with the goals of this study.

XGBoost is based on the principle of gradient boosting decision tree (GBDT). It optimizes the model by iteratively training decision trees, which can effectively process

high-dimensional data, adapt to nonlinear relationships, and avoid overfitting [32]. Compared with traditional linear models, XGBoost performs better in processing complex data and provides more transparent and actionable analysis results through built-in feature importance analysis [33,34]. The XGBoost algorithm has been widely applied to many fields and achieved remarkable results. In the analysis of public service policy satisfaction, the application potential of XGBoost is also particularly prominent. First, XGBoost can process multi-dimensional data from questionnaires, government data, and third-party data effectively and reveal the key factors affecting public satisfaction through feature selection and importance analysis [35]. Second, XGBoost has obvious advantages in prediction precision and processing nonlinear relationships, which can provide accurate satisfaction prediction in a complex policy environment [36]. This makes XGBoost the core algorithm for satisfaction analysis in this paper. Satisfaction with digital public services is not only influenced by service quality and policy implementation but also closely linked to users' socio-psychological factors. Recent research indicates that user trust, digital literacy, and motivation to participate play a key role in shaping satisfaction [37]. Furthermore, motivations for participation (such as instrumental and social motivations) also moderate satisfaction ratings. Users with instrumental motivations prioritize efficiency, while those with social motivations prioritize interactive experiences. Incorporating these individual factors into analysis helps fully understand the mechanisms that shape satisfaction, avoiding the limitations of analyzing solely through the lens of policy implementation.

3. Construction of Satisfaction Prediction Model Based on XGBoost Algorithm

3.1. Data Source and Preprocessing

3.1.1. Data Source

This study was conducted in 30 regions across eastern, central, and western China from March to August 2023. Stratified random sampling was used, with a 7:3 online to offline questionnaire ratio, covering a diverse group in terms of age, education level, and digital literacy. The questionnaire was reviewed and pre-tested by experts and the final version includes 25 core questions covering policy transparency, service efficiency, user experience, and problem-solving skills. The questions are written using a five-point Likert scale and multiple-choice options. To ensure quality, the questionnaire incorporates logical verification and attention checks, and eliminates anomalous responses. All respondents signed informed consent forms to ensure anonymity and data security, and this study adhered to strict academic ethics.

The survey framework is based on a stratified random sampling design to ensure representativeness. Regions are categorized by GDP level into three tiers: high, medium, and low, for a total of nine sampling strata, from which three to four prefecture-level cities are randomly selected, ultimately covering 30 prefecture-level cities. Within-city sample allocation is systematic, based on the age, gender, and education structure of the population. The sample design is based on national census data, with weighting applied to gender, age, and education level to ensure that the sample structure is consistent with the population. Despite the rigorous sampling, there may be biases due to insufficient digital literacy among the elderly and a bias in the online questionnaire toward digitally proficient individuals. To mitigate this issue, this study used inverse probability weighting in data preprocessing.

In terms of missing data handling, variables with a missing rate below 5% are imputed using multiple imputation; variables with a missing rate between 5% and 20% are imputed using random forests; and variables with a missing rate exceeding 20% are eliminated. Ultimately, no duplicate data was found, resulting in 100,000 samples generated through re-

gional aggregation, time expansion, and feature engineering, rather than simple replication, ensuring the authenticity and validity of the data.

Policy implementation data comes from government public data and third-party agencies, covering implementation strength, financial investment, service coverage, and resource allocation.

3.1.2. Data Preprocessing

During data preprocessing, missing values are first checked. Some respondents in the questionnaire data do not complete or skipped irrelevant questions, resulting in missing values. For satisfaction rating items, mean imputation is used to replace missing values with the average of other respondents' responses for that question. Data related to policy implementation is interpolated based on statistical yearbooks, and samples with a missingness rate exceeding 30% are removed to avoid biased results.

In terms of feature construction, multiple satisfaction indicators are generated based on the multiple-choice and rating questions in the questionnaire. Rating questions are weighted to generate a comprehensive satisfaction score, while multiple-choice questions are converted into binary variables and weighted according to importance. Policy implementation data is weighted to construct composite features based on different indicators. Furthermore, four core indicators are constructed: service efficiency, policy transparency, user experience, and problem-solving ability. These indicators are summed using a five-point scale and weighted equally, and the scores are then normalized to a range of 0–1. Policy implementation data is further integrated into comprehensive features.

Under a unified weighted scoring framework, raw scores from the questionnaire are first normalized to a range of 0–1 and then weighted to calculate a composite indicator. A similar approach is used for policy data, but weights are determined based on expert opinion and historical data to ensure the integration of theory and practice. Feature selection is initially conducted based on domain knowledge and correlation analysis, eliminating variables with correlation coefficients below 0.2. Recursive feature elimination (RFE) combined with XGBoost is then used to rank features by importance and progressively select them, ultimately retaining the 10 features with the greatest impact on satisfaction. The dataset is ultimately divided into training, test, and validation sets in an 8:1:1 ratio to improve the model's generalization capabilities.

3.2. Model Construction

3.2.1. Introduction to the XGBoost Algorithm and Its Applicability

In the process of model construction, this study selected the XGBoost algorithm as the core model [38,39]. To improve performance, this study draws on the feature engineering ideas of generative AI to generate new features to enrich the dataset and enhance the model's diversity and predictive accuracy. Figure 1 shows its workflow.

The objective function of XGBoost consists of a loss function and a regularization term [40,41], and its general form can be expressed as follows:

$$\mathcal{L}(t) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (1)$$

where $l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))$ represents the loss function between the true value and the predicted value, $f_t(x_i)$ represents the predicted value of the t -th tree for sample x_i , and $\Omega(f_t)$ is the model's regularization term used to control the complexity of the tree. The regularization term can be specifically expressed as follows:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (2)$$

where T represents the number of leaf nodes in the tree, w_j represents the weight of the leaf node, and γ and λ represent hyperparameters that control the complexity of the model. Through the regularization term, XGBoost can effectively suppress model overfitting.

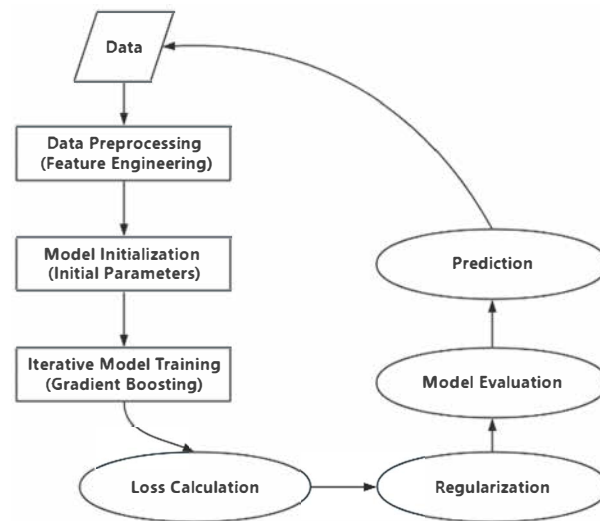


Figure 1. XGBoost algorithm workflow.

In each iteration, XGBoost approximates the optimization objective function through the second-order Taylor expansion to speed up the calculation. It is assumed that the loss function is expanded in the current i -th tree as follows:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \quad (3)$$

In Formula (3), g_i is the first-order derivative, which is defined as follows:

$$g_i = \frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i} \quad (4)$$

h_i is the second-order derivative:

$$h_i = \frac{\partial^2 l(y_i, \hat{y}_i)}{\partial \hat{y}_i^2} \quad (5)$$

3.2.2. Key Parameter Settings and Optimization Strategies of the Model

During the model construction phase, this paper balances predictive power and generalization performance by appropriately setting the learning rate, tree depth, subsampling rate, feature sampling rate, and regularization parameters. A two-stage optimization of key parameters is then performed using grid search combined with five-fold cross-validation. A coarse-grained search is first used to determine the range, followed by a fine-grained search to optimize the results. The average RMSE is used as the evaluation metric. The optimal parameter combination ultimately determined is learning rate 0.05, tree depth 4, subsampling rate 0.9, feature sampling rate 0.8, L2 regularization parameter 1.5, and minimum split loss 0.1.

In the optimization phase, a grid search combined with a cross-validation strategy is used to adjust the above key parameters. Grid search searches for the parameter configuration that optimizes model performance by traversing all possible combinations in the parameter space. Figure 2 shows the search process.

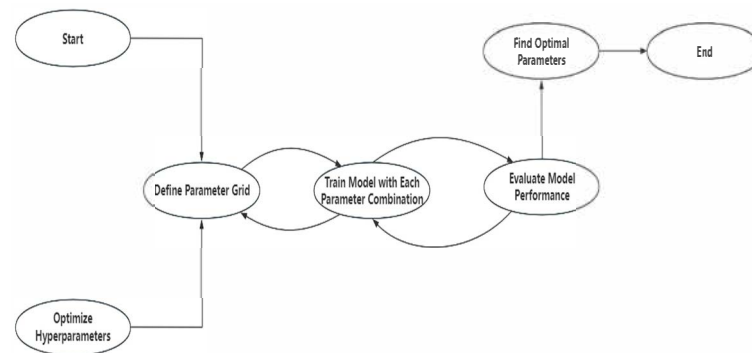


Figure 2. Grid search process.

This study uses K-fold cross-validation to divide the training set into five mutually exclusive subsets, using four for training and one for validation each time, and repeating the cycle five times to ensure that each subset is used as a validation set. Finally, the average RMSE of the five validation results is used as the performance indicator, and the optimal parameter combination is determined based on this. To avoid data leakage, the model architecture and hyperparameters are fixed before validation, and the validation data is strictly derived from the training set partitioning, without involving the test set. External validation uses independent data from the first quarter of 2024 for forward-looking testing. This data is completely invisible during the training and parameter adjustment process to further test the model's generalization ability. RMSE is selected as the evaluation indicator, and the calculation formula of RMSE is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

The smaller the RMSE, the higher the model's prediction precision.

To prevent data leakage, this study strictly adheres to the principle of separating feature engineering from model training. A “pipeline” approach is used, ensuring that steps such as feature selection, missing value handling, and normalization are based solely on training set data. During recursive feature elimination (RFE), feature ranking is based solely on training set performance. During cross-validation, feature selection and parameter optimization are performed independently for each fold, thus preventing information leakage from the validation set. Controlled experiment results show that the RMSE increases from 0.215 to 0.236 using the pipeline approach, confirming the presence of data leakage. Furthermore, noise features are not selected as significant features, further demonstrating the reliability of the feature selection process.

3.3. Analytical Framework

3.3.1. Analysis Objectives

This research aims to build an XGBoost-based satisfaction prediction framework to provide quantitative support for local government policy optimization. Leveraging the model's feature mining and generalization capabilities, it accurately predicts public service satisfaction and identifies key influencing factors. Reliability validation ensures the robustness of predictions, while incorporating feature importance analysis to provide data-driven optimization recommendations for policymaking.

3.3.2. Analysis Steps

The analysis process of this study includes three stages: data input and model training, model prediction, and feature importance mining. First, the training set is input into the

XGBoost model, and the gradient boosting algorithm is used for iterative optimization. The optimal parameters (learning rate 0.05, tree depth 4, subsampling rate 0.9, feature sampling rate 0.8, and L2 regularization parameter 1.5) are determined by combining grid search and cross-validation. Subsequently, five-fold cross-validation is used to improve model stability, and the average result is finally used as the performance indicator. After training is completed, the validation set is used for independent evaluation to ensure that the model has good generalization ability, providing a basis for subsequent prediction and feature importance analysis. Figure 3 shows the overall process of this stage.

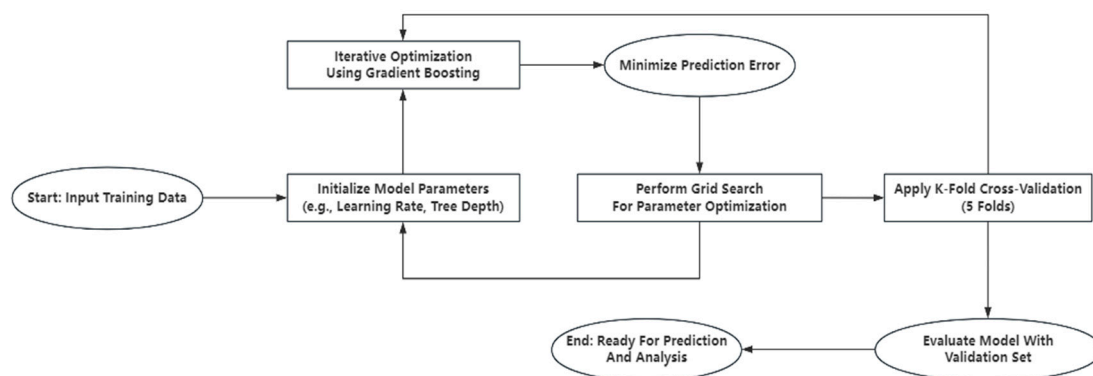


Figure 3. Data input and model training process.

In the model prediction stage, the test set is used to comprehensively evaluate the performance of the model in the satisfaction prediction task, focusing on calculating RMSE and MAE as evaluation indicators. Among them, RMSE is used to quantify the overall deviation between the predicted value and the actual value. MAE provides the average level of absolute error between the predicted value and the actual value. The combination of the two can fully reflect the model's prediction performance. The formula is as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

In addition, to verify the reliability of the prediction results, the error range between the predicted value and the actual value is analyzed, including calculating the standard deviation and skewness of the error distribution, evaluating whether the model has systematic deviations, and exploring the performance differences of the model in certain specific scenarios based on the error analysis results. Figure 4 shows the workflow of the model prediction stage.

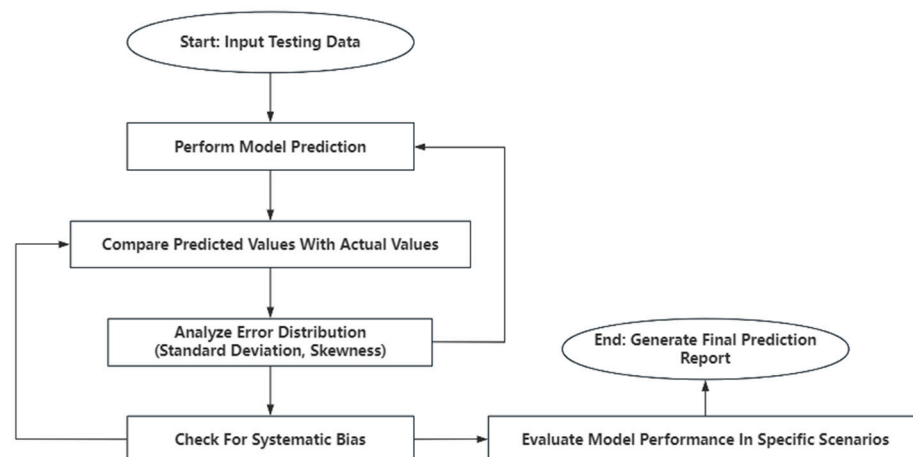


Figure 4. Model prediction workflow.

In the feature importance mining stage, the built-in feature importance analysis module of the XGBoost model is utilized to quantify the contribution of each input feature. By counting the cumulative contribution of each feature to the optimization of the objective function in the model decision tree split, the feature importance ranking is generated and presented in a visual form, such as a bar chart or heat map. This intuitive display method not only reveals the core factors that affect satisfaction but also provides a clear basis for subsequent policy optimization. Figure 5 shows the process.

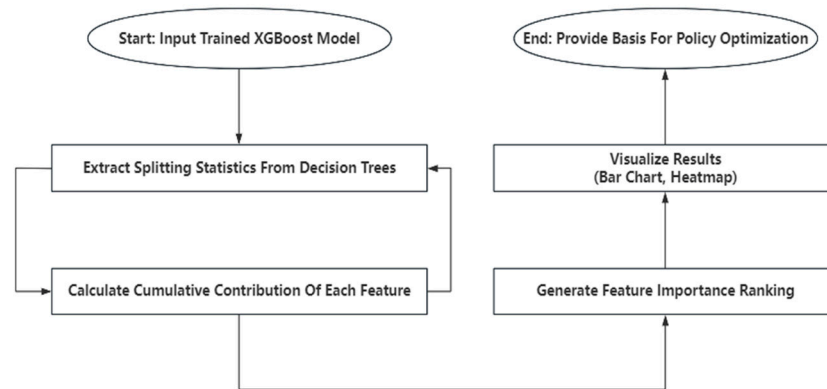


Figure 5. Feature importance mining process.

3.3.3. Scientific Basis for Policy Optimization

This study, combined with feature importance analysis, provides a basis for local governments to formulate targeted policies. If “service efficiency” has the highest weight, digital government initiatives should be prioritized, leveraging artificial intelligence and big data to optimize approval processes and increase response speed. If “policy transparency” plays a significant role, information disclosure and public participation should be strengthened to enhance trust. If “infrastructure” is key, investment in transportation, water, electricity, and healthcare should be increased. If “social participation” is important, questionnaires and opinion collection channels should be expanded to facilitate the alignment of policies with public opinion. If “people’s livelihood security” has a high weight, the social security system should be improved, and access to education, healthcare, and other services should be enhanced to strengthen people’s sense of gain.

4. Experimental Design and Results Analysis

4.1. Experimental Setup

This research experiment was completed on a high-performance computer running the Windows 11 Professional 64-bit operating system, using Python 3.9 and Jupyter Notebook 6.5.2 for modeling and analysis.

To validate the XGBoost model’s advantages, this study employed four control groups: linear regression (Control Group 1), support vector regression (Control Group 2), random forest regression (Control Group 3), and gradient boosted regression tree (Control Group 4). All models use a continuous satisfaction score as the dependent variable, employ the same training/test set split and evaluation metrics, and perform parameter optimization to ensure fairness.

4.2. Experiment and Results Analysis

4.2.1. Comparative Experiment of Model Prediction Precision

This study randomly selects 20,000 data from the experimental data set to conduct experiments so as to evaluate the model’s prediction precision. The experimental group and control group models are used for prediction. During the experiment, the predicted

values of each model for the data are recorded and compared with the true values, and the MSE, RMSE, and MAE of each group are calculated. The formula for MSE is as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

Table 1 lists the MSE, RMSE, and MAE calculated for each group.

Table 1. Data obtained from experimental calculations.

Group	MSE	RMSE	MAE
Experimental Group	0.056	0.236	0.180
Control Group 1	0.089	0.299	0.237
Control Group 2	0.078	0.279	0.221
Control Group 3	0.062	0.249	0.192
Control Group 4	0.071	0.267	0.204

Note: The lower the MSE, RMSE, and MAE values, the higher the prediction accuracy. The experimental group used the XGBoost model, and the control groups 1–4 used linear regression, support vector regression, random forest, and gradient boosting tree models, respectively.

According to the data in Table 1, the experimental group shows the best prediction effect in the three evaluation indicators MSE, RMSE, and MAE. Specifically, in the experimental group, MSE is 0.056, RMSE is 0.236, and MAE is 0.180, all of which are lower than the values of the other control groups. This indicates that the error between its predicted and true values is the smallest, and the fitting effect is the best. In contrast, control group 1 performs poorly in all indicators. Its MSE is 0.089, RMSE is 0.299, and MAE is 0.237, indicating that the prediction precision of the model is relatively low. Although control groups 2 and 4 improve the prediction precision to a certain extent, they still fail to reach the level of the experimental group. Control group 3 also shows a relatively reasonable prediction effect, but it is still lower than the experimental group in various indicators. Therefore, the XGBoost model of the experimental group shows the strongest prediction ability in this experiment.

In addition, this study randomly selects the same number of predicted and true values from each group and draws a scatter plot to evaluate the model's fitting ability, as shown in Figure 6:

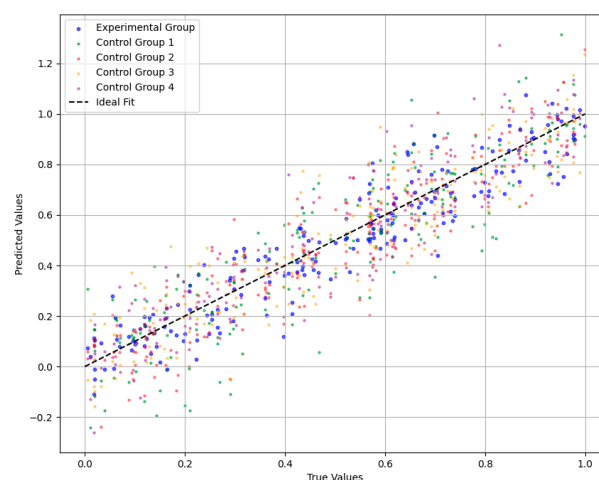


Figure 6. Scatter plot comparing four groups of predicted values with the true values.

In Figure 6, the predicted values of the experimental group are evenly distributed on both sides of the diagonal line, which are relatively close to the diagonal line, with only a few predicted values deviating from it, indicating that the experimental group model

achieves a better fitting effect. In contrast, although the predicted values of the other control group models are also on both sides of the diagonal line, they deviate slightly further, indicating that their fitting effect is worse than that of the experimental group.

4.2.2. Model Robustness Test

This study randomly extracts 20,000 sample data from the experimental data set and divides it into two data sets containing 10,000 samples each to comprehensively evaluate the model's robustness. In the experiment, two different types of noise are added to the two data sets: (1) 10%, 20%, and 30% of the feature data are randomly deleted, and the mean is filled; and (2) random disturbances that conform to the normal distribution are added, and the disturbance intensity is set to 5%, 10%, and 15%, respectively. The experimental group uses the XGBoost model constructed in this paper, and the control groups include models based on linear regression, logistic regression, random forest, and support vector machine. All models are run under different noise intensities to evaluate their robustness performance. During the experiment, the prediction results of each model are systematically recorded, and the MSE, RMSE, and MAE of each group are calculated. Figures 7 and 8 present the experimental results with the two types of noise.

The experimental results of feature deletion noises show that with the increase in noise intensity (from 10% to 30%), the MSE, RMSE, and MAE of each group of models all show different degrees of increase, indicating that the prediction precision decreases with the increase of noise (Figure 7). However, the performance of the experimental group is always better than that of other control groups under all noise intensities. For example, under 30% noise intensity, the MSE of the experimental group is 0.085, the RMSE is 0.291, and the MAE is 0.217, which are all lower than those of the other control groups, indicating that it can still maintain high precision in a noisy environment.

In the experiment with disturbance noises added (Figure 8), the robustness of the experimental group is still the strongest. Although the errors of all models increase as the disturbance intensity increases, the error increase of the experimental group is always small, suggesting that it has a strong ability to adapt to noises. Overall, the experimental group model shows the best robustness under two different types of noise interference. Whether it is feature deletion noise or adding disturbance noise, it can effectively reduce the decline in prediction precision and show strong anti-interference ability.

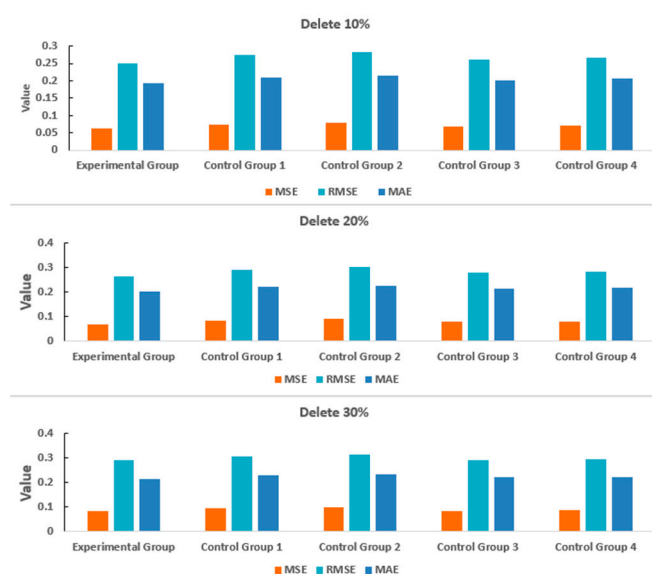


Figure 7. Prediction results under different noise intensities.

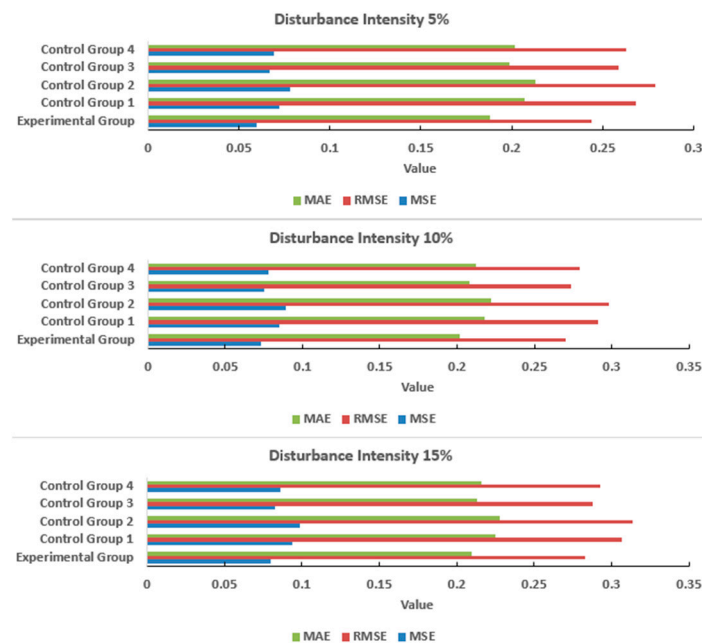


Figure 8. Prediction results under different interference noises.

4.2.3. Feature Importance Interpretation Ability Experiment

This study evaluates the model's feature importance interpretation ability by comparing the performance of each group of models via feature contribution analysis. During the experiment, 30,000 sample data are randomly selected from the experimental data set for analysis. The experimental group uses XGBoost's built-in feature importance analysis module to generate feature importance rankings. The control groups use regression coefficients, information gain, and feature weights to evaluate feature importance. During the experiment, the scores of the top 10 features of each model are recorded. Regarding the control groups' scores, average scores of each feature are statistically obtained and ranked. Figure 9 presents the experimental results.

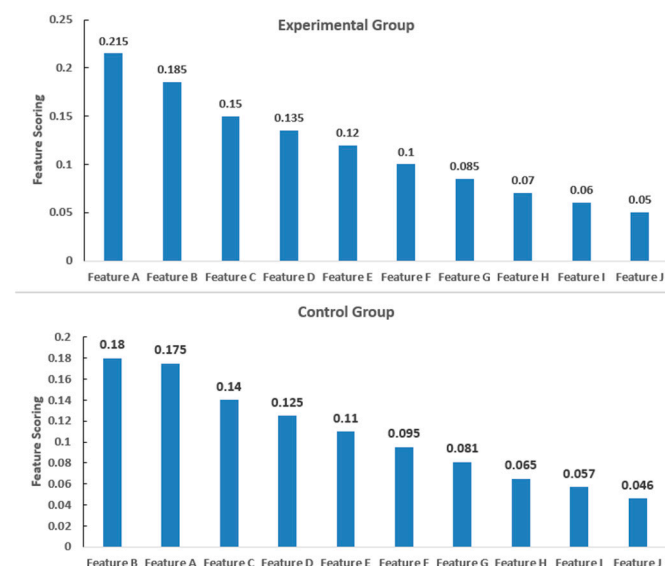


Figure 9. Histogram of feature importance rankings for the models.

According to the comparison between the experimental group's scores and the average scores of the control groups shown in Figure 9, the importance scores of the first 10 features of the XGBoost model in the experimental group are evenly distributed and high. The

importance scores of features A and B of the model are 0.215 and 0.185, respectively, and the contribution of other features is also significant. This suggests that the XGBoost model can effectively distinguish the influence of each feature and generate a feature ranking with high credibility. The performance of the model in feature importance analysis provides a solid theoretical basis for subsequent model optimization and feature selection.

In contrast, although the average scores of features B and A of the control groups are 0.18 and 0.175, respectively, their scores are generally lower than those in the experimental group. This indicates that the control group models are relatively weak in distinguishing the importance of features to the prediction results, thus limiting the interpretability of the models. Overall, the experimental group's XGBoost model demonstrates a stronger ability to evaluate feature importance, helping to improve the model's interpretability and application value.

4.2.4. Model Processing Efficiency Test

This study randomly obtains 10,000, 30,000, and 50,000 data from the experimental data set, respectively, to form three data sets of small to large sizes to evaluate the processing efficiency of each model. The data in these three data sets is predicted by the experimental and control group models, respectively. During the experiment, the prediction time and processing speed of each model in different scale data sets are recorded. Table 2 lists the experimental results.

Table 2. Model processing efficiency results.

Data Set Size	Group	Prediction Time (Seconds)	Processing Speed (Items/Second)
10,000	Experimental Group	13.1	763.4
	Control Group 1	17.9	558.7
	Control Group 2	15.6	641.0
	Control Group 3	21.9	456.6
	Control Group 4	18.8	531.9
30,000	Experimental Group	42.7	701.8
	Control Group 1	61.5	487.6
	Control Group 2	52.5	571.5
	Control Group 3	82.8	362.4
	Control Group 4	65.4	458.7
50,000	Experimental Group	73.7	678.3
	Control Group 1	123.3	405.6
	Control Group 2	102.4	488.2
	Control Group 3	173.7	287.8
	Control Group 4	134.6	371.5

In Table 2, the computational efficiency of the experimental group is consistently better than that of the control group in datasets of varying sizes. Processing speed = dataset size/prediction time, which indicates the number of data items that the model can process per second. A larger value indicates higher processing efficiency. Specifically, when the dataset size is 10,000 items, the total computation time of the experimental group is 13.1 s and the computational efficiency is 763.4 items/second, compared to the control group's 17.9 s, 15.6 s, 21.9 s, and 18.8 s, respectively, showing an advantage in computational time. In addition, as the dataset size increases, the computation time of the experimental group on datasets of 30,000 and 50,000 items is 42.7 s and 73.7 s, respectively, with computational efficiencies of 701.8 items/second and 678.3 items/second, respectively, maintaining a high processing efficiency.

In contrast, the computational efficiency of control group 4 is significantly lower than that of the experimental group. This is particularly true for the 50,000-item dataset, where the control group's computational efficiency is generally low. Specifically, control group 3 achieves a computational efficiency of 287.8 items per second on the 50,000-item dataset, significantly lower than the experimental group. This demonstrates that the experimental group possesses higher computational efficiency when processing large-scale data, is better able to adapt to datasets of varying sizes, and demonstrates strong scalability and efficiency.

4.2.5. Cross-Scenario Applicability Verification Experiment

To verify the model's cross-scenario applicability, this study randomly samples different regions and populations from the experimental dataset to construct subsets, including underdeveloped regions, special economic zones, autonomous regions, municipalities, and remote mountainous areas. Each region contains 5000 records, covering characteristics such as income, education, and living conditions. Predictions are made using the experimental and control group models, and the Pearson correlation coefficient is used to measure the linear correlation between the predicted and actual values, a metric suitable for evaluating continuous satisfaction data. Table 3 lists the experimental results.

Table 3. Cross-region prediction correlation coefficient results.

Region	Experimental Group	Control Group 1	Control Group 2	Control Group 3	Control Group 4
Underdeveloped Regions	0.85	0.70	0.72	0.75	0.73
Special Economic Zones	0.90	0.78	0.80	0.82	0.81
Autonomous Regions	0.83	0.68	0.71	0.74	0.72
Municipalities Directly under the Central Government	0.88	0.76	0.78	0.80	0.79
Remote Mountainous Areas	0.80	0.65	0.68	0.72	0.70

Table 3 shows the Pearson correlation coefficient between the predicted and true values, measuring the model's consistency across different regions. The experimental group demonstrates significant advantages across all regions, and the correlation coefficient reaches 0.85 in underdeveloped regions, significantly higher than the four control groups (0.70–0.75). The correlation coefficients are 0.90 and 0.88 in special economic zones and municipalities, respectively. The results also remain stable in autonomous regions and remote mountainous areas. These results demonstrate that the XGBoost model not only demonstrates outstanding overall predictive power but also exhibits good cross-regional adaptability, effectively addressing regional differences.

To control overfitting, this study implements several measures. By comparing the loss curves of the training and validation sets, 150 trees are determined to be the optimal size. Early stopping is employed, terminating training after five consecutive rounds of validation without performance improvement. Regularization parameters λ and γ are increased to control complexity, ultimately selecting $\lambda = 1.5$. Experiments show that XGBoost achieves a RMSE difference of only 1.08 between the training and test sets, lower than the control group, demonstrating strong resistance to overfitting. Shapley value analysis verifies that feature contributions are consistent with domain knowledge, with the RMSE difference between the reduced and full feature sets being only 0.003, further demonstrating that the model relies on meaningful features and exhibits good robustness and interpretability.

The model's external validity and forward-looking predictive capabilities were verified through two experiments. First, using 5000 independent datasets for testing, the model achieves an MSE of 0.061, an RMSE of 0.247, and an MAE of 0.191 on the new data. While

performance is slightly lower than on the original test set, it remains stable, demonstrating strong temporal generalization. Second, in collaboration with a provincial government, the model was applied to predict policy satisfaction in the first quarter of 2024. The actual score, 83.2, is highly consistent with the predicted score of 82.7, with an error of only 0.5 points, demonstrating the model's accuracy and practical value.

5. Conclusions

This study, for the first time, systematically constructed a digital public service satisfaction assessment framework based on XGBoost. This theoretical approach overcomes the limitations of traditional linear models in capturing nonlinear relationships and feature interactions, expanding the “efficiency–transparency” dual-core driving model in digital governance theory. In practice, by ranking feature importance, it provides local governments with precise policy optimization paths, such as prioritizing digital process reengineering and strengthening policy feedback loops and differentiating regional resource allocation. The model's implementation in cross-regional, noise-resistant, and efficient scenarios was validated, providing an empirical foundation and algorithmic tool for establishing unified national digital government evaluation standards. In terms of model performance, XGBoost's MSE (0.056), RMSE (0.236), and MAE (0.180) all outperform the control model, demonstrating its ability to more accurately capture nonlinear relationships. The model is robust to missing features and noise, and its interpretability demonstrates its ability to identify and rank key factors, providing a basis for policy optimization. Current research is limited by regional response bias, inadequate control of digital literacy, and insufficiently representative data from the pandemic period. Future work should: (1) introduce a hierarchical mixed-effects model to distinguish the satisfaction mechanisms of different digital literacy groups; (2) extend the longitudinal data collection period to cover both the normal and emergency policy stages; and (3) combine field experiments with behavioral data to calibrate the model's generalization capabilities in areas with significant digital divides.

Author Contributions: Conceptualization, Q.H. and B.Y.; methodology, Q.H. and B.Y.; software, Q.H. and B.Y.; formal analysis, Q.H. and B.Y.; investigation, B.Y.; resources, Q.H.; data curation, Q.H. and B.Y.; writing—original draft preparation, Q.H. and B.Y.; writing—review and editing, Q.H. and B.Y.; visualization, Q.H. and B.Y.; supervision, S.D.; project administration, S.D.; funding acquisition, S.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research and APC was funded by the Major Project of Philosophy and Social Science Research of the Ministry of Education of China, grant number No. 23JZD019.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kuhlmann, S.; Heuberger, M. Digital transformation going local: Implementation, impacts and constraints from a German perspective. *Public Money Manag.* **2023**, *43*, 147–155. [\[CrossRef\]](#)
2. Scupola, A.; Mergel, I. Co-production in digital transformation of public administration and public value creation: The case of Denmark. *Gov. Inf. Q.* **2022**, *39*, 101650–101661. [\[CrossRef\]](#)
3. Gasco-Hernandez, M.; Nasi, G.; Cucciniello, M.; Hiedemann, A.M. The role of organizational capacity to foster digital transformation in local governments: The case of three European smart cities. *Urban Gov.* **2022**, *2*, 236–246. [\[CrossRef\]](#)
4. Andersson, C.; Hallin, A.; Ivory, C. Unpacking the digitalisation of public services: Configuring work during automation in local government. *Gov. Inf. Q.* **2022**, *39*, 101662–101672. [\[CrossRef\]](#)
5. Shen, Y.; Cheng, Y.; Yu, J. From recovery resilience to transformative resilience: How digital platforms reshape public service provision during and post COVID-19. *Public Manag. Rev.* **2023**, *25*, 710–733. [\[CrossRef\]](#)

6. Mu, R.; Wang, H. A systematic literature review of open innovation in the public sector: Comparing barriers and governance strategies of digital and non-digital open innovation. *Public Manag. Rev.* **2022**, *24*, 489–511. [\[CrossRef\]](#)
7. Poluan, M.S.; Pasuhuk, L.S.; Mandagi, D.W. The role of social media marketing in local government institution to enhance public attitude and satisfaction. *J. Ekon.* **2022**, *11*, 1268–1279.
8. Zhang, J.; Chen, W.; Petrovsky, N.; Walker, R.M. The expectancy-disconfirmation model and citizen satisfaction with public services: A meta-analysis and an agenda for best practice. *Public Adm. Rev.* **2022**, *82*, 147–159. [\[CrossRef\]](#)
9. Kim, S.; Andersen, K.N.; Lee, J. Platform government in the era of smart technology. *Public Adm. Rev.* **2022**, *82*, 362–368. [\[CrossRef\]](#)
10. Schmidhuber, L.; Hilgers, D.; Hofmann, S. International Public Sector Accounting Standards (IPSASs): A systematic literature review and future research agenda. *Financ. Account. Manag.* **2022**, *38*, 119–142. [\[CrossRef\]](#)
11. Chien, N.B.; Thanh, N.N. The impact of good governance on the people's satisfaction with public administrative services in Vietnam. *Adm. Sci.* **2022**, *12*, 35–47. [\[CrossRef\]](#)
12. Crucke, S.; Kluijtmans, T.; Meyfrootd, K.; Desmidt, S. How does organizational sustainability foster public service motivation and job satisfaction? The mediating role of organizational support and societal impact potential. *Public Manag. Rev.* **2022**, *24*, 1155–1181. [\[CrossRef\]](#)
13. Fleischer, J.; Wanckel, C. Job satisfaction and the digital transformation of the public sector: The mediating role of job autonomy. *Rev. Public Pers. Adm.* **2024**, *44*, 431–452. [\[CrossRef\]](#)
14. Ochoa Rico, M.S.; Vergara-Romero, A.; Subia, J.F.R.; Del Rio, J.A.J. Study of citizen satisfaction and loyalty in the urban area of Guayaquil: Perspective of the quality of public services applying structural equations. *PLoS ONE* **2022**, *17*, e0263331. [\[CrossRef\]](#)
15. Romero-Subia, J.F.; Jimber-del Rio, J.A.; Ochoa-Rico, M.S.; Vergara-Romero, A. Analysis of citizen satisfaction in municipal services. *Economies* **2022**, *10*, 225–249. [\[CrossRef\]](#)
16. Wang, C.; Ma, L. Digital transformation of citizens' evaluations of public service delivery: Evidence from China. *Glob. Public Policy Gov.* **2022**, *2*, 477–497. [\[CrossRef\]](#)
17. Noori, M. The effect of e-service quality on user satisfaction and loyalty in accessing e-government information. *Int. J. Data Netw. Sci.* **2022**, *6*, 945–952. [\[CrossRef\]](#)
18. Benmohamed, N.; Shen, J.; Vlahu-Gjorgievska, E. Public value creation through the use of open government data in Australian public sector: A quantitative study from employees' perspective. *Gov. Inf. Q.* **2024**, *41*, 101930–101944. [\[CrossRef\]](#)
19. Junaidi, A.; Basrowi, B.; Sabtohad, J.; Wibowo, A.; Wibowo, S.; Asgar, A.; Yenti, E. The role of public administration and social media educational socialization in influencing public satisfaction on population services: The mediating role of population literacy awareness. *Int. J. Data Netw. Sci.* **2024**, *8*, 345–356. [\[CrossRef\]](#)
20. Zhang, Z.; Li, A.; Xu, Y.; Liang, Y.; Jin, X.; Wu, S. Understanding citizens' satisfaction with the government response during the COVID-19 pandemic in China: Comprehensive analysis of the government hotline. *Libr. Hi Tech* **2023**, *41*, 91–107. [\[CrossRef\]](#)
21. Alkrajji, A.; Ameen, N. The impact of service quality, trust and satisfaction on young citizen loyalty towards government e-services. *Inf. Technol. People* **2022**, *35*, 1239–1270. [\[CrossRef\]](#)
22. Ye, X.; Su, X.; Yao, Z.; Dong, L.A.; Lin, Q.; Yu, S. How Do Citizens View Digital Government Services? Study on Digital Government Service Quality Based on Citizen Feedback. *Mathematics* **2023**, *11*, 3122–3146. [\[CrossRef\]](#)
23. Yang, L.; Wang, J. Factors influencing initial public acceptance of integrating the ChatGPT-type model with government services. *Kybernetes* **2024**, *53*, 4948–4975. [\[CrossRef\]](#)
24. Agostino, D.; Saliterer, I.; Steccolini, I. Digitalization, accounting and accountability: A literature review and reflections on future research in public services. *Financ. Account. Manag.* **2022**, *38*, 152–176. [\[CrossRef\]](#)
25. Hadwan, M.; Al-Sarem, M.; Saeed, F.; Al-Hagery, M.A. An improved sentiment classification approach for measuring user satisfaction toward governmental services' mobile apps using machine learning methods with feature engineering and SMOTE technique. *Appl. Sci.* **2022**, *12*, 5547–5572. [\[CrossRef\]](#)
26. Mustaqim, I.Z.; Puspasari, H.M.; Utami, A.T.; Syalevi, R.; Ruldeviyani, Y. Assessing public satisfaction of public service application using supervised machine learning. *Int. J. Artif. Intell.* **2024**, *2252*, 1609–1620.
27. Chandrasekaran, G.; Dhanasekaran, S.; Moorthy, C.; Arul Oli, A. Multimodal sentiment analysis leveraging the strength of deep neural networks enhanced by the XGBoost classifier. *Comput. Methods Biomech. Biomed. Eng.* **2024**, *28*, 777–799. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Ben Jabeur, S.; Stef, N.; Carmona, P. Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering. *Comput. Econ.* **2023**, *61*, 715–741. [\[CrossRef\]](#)
29. Chen, R.; Zhang, S.; Li, J.; Guo, D.; Zhang, W.; Wang, X.; Wang, X. A study on predicting the length of hospital stay for Chinese patients with ischemic stroke based on the XGBoost algorithm. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, 49. [\[CrossRef\]](#)
30. Asselman, A.; Khaldi, M.; Aammou, S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interact. Learn. Environ.* **2023**, *31*, 3360–3379. [\[CrossRef\]](#)
31. Fantozzi, I.C.; Olhager, J.; Johnsson, C.; Schiraldi, M.M. Guiding organizations in the digital era: Tools and metrics for success. *Int. J. Eng. Bus. Manag.* **2025**, *17*, 18479790241312804. [\[CrossRef\]](#)

32. Choudhury, A.; Mondal, A.; Sarkar, S. Searches for the BSM scenarios at the LHC using decision tree-based machine learning algorithms: A comparative study and review of random forest, AdaBoost, XGBoost and LightGBM frameworks. *Eur. Phys. J. Spec. Top.* **2024**, *233*, 2425–2463. [[CrossRef](#)]
33. Zhang, P.; Jia, Y.; Shang, Y. Research and application of XGBoost in imbalanced data. *Int. J. Distrib. Sens. Netw.* **2022**, *18*, 15501329221106935–15501329221106945. [[CrossRef](#)]
34. Deng, Y.; Lumley, T. Multiple imputation through XGBoost. *J. Comput. Graph. Stat.* **2024**, *33*, 352–363. [[CrossRef](#)]
35. Liu, C.; Li, Y.; Fang, M.; Liu, F. Using machine learning to explore the determinants of service satisfaction with online healthcare platforms during the COVID-19 pandemic. *Serv. Bus.* **2023**, *17*, 449–476. [[CrossRef](#)]
36. Feng, Y.; Park, J. Using machine learning-based binary classifiers for predicting organizational members' user satisfaction with collaboration software. *PeerJ Comput. Sci.* **2023**, *9*, e1481–e1503. [[CrossRef](#)]
37. Potluka, O.; Harten, S.; Kocks, A.; Dvorak, J. Digitalization in evaluations and evaluations of digitalization: The changing landscape of evaluations. *Evaluation* **2025**, *31*, 13563890251357650. [[CrossRef](#)]
38. Shi, Y.; Ke, G.; Chen, Z.; Zheng, S.; Liu, T.Y. Quantized training of gradient boosting decision trees. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 18822–18833.
39. Kavzoglu, T.; Teke, A. Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). *Arab. J. Sci. Eng.* **2022**, *47*, 7367–7385. [[CrossRef](#)]
40. Deng, X.; Li, M.; Deng, S.; Wang, L. Hybrid gene selection approach using XGBoost and multi-objective genetic algorithm for cancer classification. *Med. Biol. Eng. Comput.* **2022**, *60*, 663–681. [[CrossRef](#)]
41. Suresh, G.V.; Reddy, S. Uncertain data analysis with regularized XGBoost. *Webology* **2022**, *19*, 3722–3740. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.