
Ying Ping * and Zhuolin Li

Abstract: The convergence of digitization and greening is an unavoidable path of modern economic progress. Nonetheless, the digital economy does not consistently align with the principles of green development, potentially leading to a rebound effect in urban digitalization initiatives. To investigate the correlation between the digital rebound effect and urban green development, this study utilizes panel data from Chinese prefecture-level cities spanning from 2011 to 2019. By examining the dual impact of the digital economy on green development, the paper posits a theoretical hypothesis regarding the nonlinear marginal effect of the digital economy. This research demonstrates an inverted U-shaped correlation between the digital economy and urban green development via empirical analyses employing the random forest algorithm and partial dependency plots. It supports the existence of a moderate digital resiliency effect, which eventually reaches a state of stability rather than greatly diminishing the degree of green development in urban areas. In addition, the heterogeneity analysis reveals that the positive effects of the digital economy are more popular in cities located in the eastern and central regions, as well as in the National Comprehensive Pilot Zone for Big Data. However, these effects do not vary significantly among different ranks of cities. The mechanism test found that the information effect and the capital allocation effect are the mechanisms by which the digital economy affects green development, and there is a “U-shaped” relationship between the digital economy and information asymmetry and capital mismatch. According to the study’s results, improving the digital economy’s governance structure continues to make more sense than merely increasing the number of digital inputs.

Keywords: digital economy; green development; machine learning; digital-green convergence; digital rebound

1. Introduction

The development and use of digital technologies, such as artificial intelligence, visual computing, and global positioning systems, have paved the way for the promising future of self-driving cars, which are expected to become widely accessible. These vehicles offer significant environmental advantages over traditional transportation methods. They offer novel prospects for enhancing the coordination of vehicles and optimizing traffic patterns by linking nearby vehicles to the infrastructure via a “digital cloud”. This has the potential to contribute to environmental conservation by reducing carbon dioxide emissions and decreasing oil consumption [1,2]. Furthermore, the smaller size and lighter weight of self-driving cars, along with the promotion of car sharing, are expected to further diminish resource consumption and enhance environmental performance [1].

Nevertheless, the impact does not stop there. The appeal of autonomous vehicles will result in increased automobile usage as they may effectively substitute other forms of mobility [3]. This encompasses an increased distance covered or elevated driving velocities, resulting in a greater consumption of oil [4]. Moreover, individuals who were previously...
excluded, such as the elderly and children, are predicted to begin using cars, resulting in a surge in consumption and associated environmental challenges [4].

This scenario is frequently seen in the progress of the digital economy. It illustrates that the development of the digital economy does not consistently align with the principles of sustainable development, and the application of digital technology may lead to heightened consumption, thereby intensifying the challenges of achieving sustainable development. Is the occurrence of digital rebound widespread in the context of urban economic development? What are the underlying reasons for the emergence of this economic phenomenon?

Several studies have shown the occurrence of the “digital rebound” effects and have raised concerns about the increased pressure on resources and the environment caused by the progress of the digital economy [5]. The rebound effect, initially an established concept in the field of energy economics, refers to the notion that improvements in energy efficiency eventually reduce the potential energy savings and could possibly end in an increase in energy consumption. Coroamă and Mattern (2019) have extended the general concept of the rebound effect by introducing the concept of the “digital rebound,” suggesting that this rebound effect may vary in magnitude [6]. Lange et al. (2020) have examined the connection between the digital economy and the consumption of energy from a theoretical perspective. They argue that the growth of the digital economy leads to enhancements in energy efficiency and changes in industrial sectors or trade, ultimately leading to a decrease in energy consumption. Still, they assert that it results in an energy rebound effect and an expansion of economic scale, subsequently amplifying energy consumption. Avom et al. (2020) [8], as well as Kunkel and Matthess (2020) [9], take a broader perspective to recognize the digital empowerment effect and the rebound effect. They present empirical data which supports the theory that the digital economy could negatively impact green development strategies. Further empirical evidence indicates the presence of a nonlinear correlation between the digital economy and green growth [10,11], suggesting that the digital rebound effect emerges during the later phases of digital economy development.

Still, there are also studies which advocate for the belief that the digital economy keeps promoting sustainable growth. In their research, Zhang and Wang (2023) highlight the significant impact of the digital economy on enhancing China’s green total factor productivity, asserting that it represents a novel driver for green development [12]. Zhang et al. identified the capacity of digital economic advancement to reduce carbon emissions. Their study determined that digital economic progress enhances the efficiency of carbon emission reduction and generates a non-linear spillover effect that can support neighboring cities in attaining their objectives for sustainable development [13]. The present discussion on the relationship between the digital economy and sustainable development has attracted considerable interest, leading numerous scholars to study how digitalization may support environmentally friendly activities. Researchers have conducted both micro and macro level analyses on this subject. For instance, Sun et al. have posited that the convergence of digital economy and traditional economy is unlocking substantial potential for green innovation. This collaboration is expected to improve environmentally friendly innovation by reducing the financial constraints on enterprises, promoting the adoption of digital technology, and encouraging companies to take responsibility for their impact on society [14]. Similarly, Hu et al. have discovered a similar mechanism by analyzing provincial panel data in China [15]. In addition, Huang and Ni has emphasized the significance of the digital economy in fostering high-quality growth in the Yangtze River Economic Belt by rectifying disparities in capital, labor, and land resources [16].

Connected to these research investigations, our research consists of two primary components. First of all, we explore how the digital economy may assist promote green development in urban areas, which is a topic that falls in line with existing academic studies. This aspect holds particular significance, as the absence of the positive impact of the digital economy on urban green development would naturally preclude the existence of a digital rebound. Furthermore, we clarify that the level of urban green development will decrease in the latter stages of the digital economy as digitalization keeps expanding. The phenomenon
known as digital rebound may be attributed to three factors: the scale effect of the digital economy, the energy rebound effect, and the law of diminishing marginal efficiency. To investigate the prevalence of the digital rebound effect in urban economic development, we gather panel data from Chinese cities spanning from 2011 to 2019. We subsequently employ the random forest algorithm, bagging algorithm, or gradient boosting decision tree (GBDT) algorithm to train the data and develop the model. Afterwards, we employ partial dependency plot (PDP) to visually illustrate the non-linear correlation between the digital economy and urban green growth, with a specific emphasis on identifying the presence of digital rebound. In summary, our findings validate the widespread occurrence of the digital rebound effect in urban development, particularly at the later phases of digital economy progress. Moreover, it is important to acknowledge that the digital rebound effect has limited impact on diminishing the level of sustainable growth of the city, and ultimately reaches a state of stability. The digital economy possesses the capability to accelerate environmentally friendly progress on a broader magnitude.

It is beneficial to position our research within the existing body of literature concerning the interplay between the digital economy and sustainable development. A significant portion of the current literature investigates strategies for harnessing the environmentally friendly potential of the digital economy, often operating under the assumption that the digital economy consistently fosters sustainable development, a perspective commonly referred to as the “mainstream view”. Our research acknowledges the environmental consequences associated with the rapid growth of the digital economy, while also delving into the concept of digital rebound that emerges in the later stages of this economic sector. Consequently, our study seeks to bridge the mainstream perspective with the notion of digital rebound, aiming to present empirical evidence to ascertain the prevalence of the digital rebound phenomenon in the historical urban economic development of China. Furthermore, we endeavor to offer a theoretical framework to elucidate the economic incentives driving the emergence of this phenomenon, with the ultimate goal of assisting policymakers in formulating informed decisions regarding urban economic development.

The possible marginal contributions of this paper are as follows. This study recognizes the occurrence of the digital rebound effect, which stems from the scale effect, the energy rebound effect, and the law of declining marginal efficiency in the digital economy. Furthermore, this research delves into the underlying factors contributing to the digital rebound phenomenon from a novel standpoint within the realm of the capital market. The evidence presented in this study indicates a nonlinear correlation between the digital economy and capital misalignment and information asymmetry, resulting in the occurrence of digital rebound. Beyond that, this study employs an ensemble learning method (EL) and PDP to estimate and explain nonlinear relationships among economic variables. Traditional modeling concepts are limited by the structure of the model configuration and the challenge of visually analyzing non-linear connections among economic factors. Hence, this paper employs the random forest algorithm and other EL algorithms that are based on tree models for data training. Additionally, it illustrates these non-linear relationships by applying PDP. This contributes to the progression of utilizing machine learning methods in analyzing the nonlinear correlations among economic factors. More importantly, this research offers novel viewpoints on the developments of the digital economy. It highlights the importance of changing the approach to digital economic growth, improving the digital economic governance system and institutional structure, and speeding up digital innovation, in response to the digital rebound impact. In the present deployment of digital economic growth, these aspects are considered more essential than merely augmenting the incorporation of digital elements.
2. Theoretical Analysis and Research Hypotheses

2.1. The Non-Linear Relationship between Digitization and Urban Green Development

2.1.1. Empowering Effects of the Digital Economy

The emergence of the digital economy triggered substantial transformations in the micro, meso, and macro elements that impact the endeavor to establish environmentally friendly cities. At the micro level, the digital economy has caused a significant transformation in manufacturing procedures, affected changes in government regulatory approaches, and influenced consumer purchasing preferences and behaviors. By reducing the expenses linked to obtaining information on green innovations, companies can facilitate the introduction and progression of novel information technologies, surpass temporal and spatial constraints on information distribution, and incentivize the development of an entirely new generation of green technologies [17,18]. The process of digitalization is also expected to enhance the acquisition of environmentally conscious human resources and the longevity of environmentally friendly technological advancements, driven by the substantial impact of information dissemination [19]. In addition, digital technology may facilitate the environmentally friendly evolution of corporate supply chains by monitoring and optimizing energy usage throughout the entire supply chain. It can also predict maintenance and recycling of machinery and equipment, thereby improving energy efficiency and achieving the corporate goal of reducing costs and increasing efficiency. In terms of consumption, consumers embrace information technology to support green consumption principles and undertake a new consumption trend [13]. Additionally, it may augment consumers’ comprehension of the concept of sustainable development and promote the public’s engagement in environmental affairs [5,20]. The introduction of algorithmic big data analysis and digital environmental monitoring systems can reduce information restrictions in environmental control and improve regulatory efficiency within the context of government regulation [21,22].

At the mesoscopic level, the digital economy facilitates the optimization and enhancement of the industrial structure, thereby supporting the attainment of environmentally sustainable development objectives. The internal structure of the information and communication technology (ICT) sector is evolving towards a service-oriented model, as evidenced by the growing share of the service industry and the declining share of the manufacturing industry [7]. Given that the service industry is generally more environmentally friendly than the manufacturing industry, these shifts within the ICT sector are expected to expedite the progression towards a greener economy. Furthermore, the increasing proportion of the ICT sector and its growing integration with other economic sectors are anticipated to empower the transformation, optimization, and enhancement of the industrial structure. This is likely to result in reduced pollutant emissions, lower energy consumption, and overall contributions to green development.

At the macro level, the process of digitalization has significantly enhanced the allocation of resources. The popularity of online marketplaces and digital platforms has brought forward a new method of resource allocation called the data mechanism. This mechanism works in conjunction with or as a substitute for the traditional price mechanism, ultimately contributing to the enhancement of resource allocation efficiency [23]. Additionally, information technology plays a crucial role in facilitating the exchange of data between producers and consumers, converting it into valuable information. This, in turn, fosters the development of feedback mechanisms between production and consumption, leading to an overall improvement in resource allocation efficiency.

In conclusion, it would be imprudent to overlook the facilitating impact of the digital economy on sustainable development, as it serves as a catalyst for promoting environmental sustainability.

2.1.2. Digital Rebound Effects

The digital economy plays an essential part in fostering sustainable development, but it has limitations and cannot continue expanding infinitely. The eventual digital rebound
is influenced by factors such as the scale effect, energy rebound effect, and the limits imposed by the law of decreasing marginal efficiency. The scale effect implies that the growth of the digital economy is unlikely to result in a decrease in total energy consumption. Ecological economists contend that there is an interconnected relationship between physical capital and energy consumption, implying that an increase in physical capital stock would unavoidably result in an elevation in energy consumption [7]. Additionally, due to the physical constraints of the real world, the energy produced from consuming a unit of energy cannot grow infinitely. Consequently, energy consumption cannot be completely detached from economic growth, and the economic expansion driven by the digital economy is expected to lead to increased energy consumption [7]. The empirical findings of Ren et al. (2021) provide support for this claim, demonstrating that the growth of the ICT sector in China leads to an overall increase in energy consumption, but also results in a decrease in energy consumption per unit of output and an improvement in the composition of sources of energy [24]. In contrast, Zhou et al. (2018) present actual data demonstrating that the growth of the ICT industry results in a rise in both overall energy consumption and energy intensity [25].

Secondly, the energy rebound effect may skew the potential influence of the digital economy on promoting environmental sustainability. The Khazzoom–Brookes hypothesis suggests that any advancements in energy efficiency may be offset by an increase in energy consumption due to the behavioral responses of economic agents [26]. Advances in energy efficiency result in decreased energy expenses, which in turn encourages individuals to consume more energy as a result of the substitution effect caused by the decrease in pricing. In addition, the drop in prices generates an income effect, causing people’s income to relatively grow, which in turn leads to a rise in energy consumption. Empirical research shows that the rebound effects on residential energy consumption in France varies from 38% to 86%, with notable differences related to household consumption [27]. According to Galvin (2015), the rebound effect across multiple industrialized nations is estimated to range from 115% to 161%, indicating a potential backfire effect [28]. Han et al. (2019) calculate that the direct rebound effect and spillover effect for residential consumption of electricity in China are 37% and 13%, respectively [29]. This suggests that the main factor impacting the implementation of energy efficiency is the direct rebound effect, while the spillover effect is negligible. As a result, the energy efficiency advances brought about by the digital economy may not produce the expected energy savings. This might result in higher environmental pollution due to increasing energy usage [8].

Finally, it is crucial to take into account the decreasing marginal efficiency while analyzing the facilitating function of the digital economy. From a social efficiency standpoint, the growth of the digital economy comes with expenditures, and the addition of a digital element may not always result in higher production or progress in sustainable development if marginal cost outweighs the marginal benefits. Security of information [30], digital monopolies [31,32], and data governance [33] provide challenges in the real-world implementation of digital economy growth, impeding progress towards greater economic efficiency and hindering the promotion of sustainable green development. Consequently, although the optimal digital economy is in accordance with Metcalfe’s law, the actual extent of digital economic empowerment is constrained by disparities in institutional contexts [10] and regional developmental frameworks [34].

The scale effect, energy rebound effect, and the law of diminishing marginal efficiency become apparent once the digital economy has advanced to a certain stage. When the scale of the digital economy is limited and digital technology innovation and application have not yet led to reduced energy prices, or when the marginal cost of digital factor inputs remains lower than their marginal benefit, the aforementioned effects are less pronounced. Consequently, we propose the following hypothesis.

Hypothesis 1: There exists a curvilinear relationship between the advancement of the digital economy and the progress of urban green development, characterized by an inverted U-shaped
pattern. This indicates that as the digital economy matures, there is a rebound effect on urban green development, leading to a non-linear evolutionary trend of initial growth followed by decline.

2.2. Mechanism of Digital Rebound Generation

2.2.1. Information Asymmetry

The existence of information asymmetry, especially when it comes to environmental factors, presents a major challenge to companies who are involved in green innovation initiatives. The mentioned aspect also exerts an influence on the degree of advancement in urban green development [35]. Information asymmetry, according to the theory of information economics, leads to the presence of (ex-ante) hidden information and (ex-post) hidden actions. The progression of the digital economy contributes to the reduction in challenges associated with adverse selection and moral hazard that arise from discrepancies in information among economic actors.

The issue of adverse selection has several and intricate origins. Enterprises may engage in opportunistic behaviors, such as excessive consumption during work, short-term incentives, and tax avoidance [36]. They may also have important private information about their environmental responsibilities, which they could decide to conceal or not promptly disclose. Furthermore, organizations may choose to maintain confidentiality regarding specific data, such as advancements in digital technology, commitment to environmental sustainability, and financial information, as a result of expenses associated with revealing such information or concerns related to competitive market dynamics [37]. Investors and other stakeholders may encounter external constraints while attempting to access information or incur substantial search expenses, so impeding their capacity to acquire pertinent enterprise information. The digital economy possesses the capacity to address these difficulties through the reduction in expenses associated with search, copy, transportation, tracking, and verification [38]. Consequently, it enables the facilitation of information sharing across different entities involved in financing activities. The adoption of digital reputation procedures [39] and blockchain technology [40] serves to augment organization openness and diminish the motivation to withhold facts. Hence, the digital economy contributes to mitigating information limitations and resolving the issue of adverse selection.

Moral hazard is an undesirable consequence that arises from information asymmetry. It occurs when an organization, having obtained financing, may misappropriate funds obtained for environmental initiatives into other endeavors. This is because investors lack the ability to verify the true scope of the company’s green initiatives. This situation highlights the difficulty posed by moral hazard. The explosion of the digital economy has expanded the opportunities for sharing environmental information, resulting in a favorable exposure impact that attracts the interest of investors and motivates companies to actively participate in environmental stewardship [35]. The proliferation of the digital economy has given rise to a wide range of civilian social platforms, such as social media, which are distinguished by their promptness, cost-efficiency, ease of use, swift distribution, and extensive interactivity [41]. These platforms have the potential to enable investors to effectively monitor and address the environmental responsibility of enterprises, bring attention to activities that harm the environment, and recognize environmentally favorable actions through efficient information dissemination. As a result, this procedure serves as a catalyst for enterprises to adhere to their environmental responsibilities, thus mitigating the problem of moral hazard.

While the advancement of the digital economy has mitigated green information asymmetry to a significant extent, various challenges persist during the digitization process that give rise to information barriers, hindering the enhancement of green development. Enterprises and other environmentally responsible entities may have motives and capabilities to manipulate environmental information, as noted by Zhang et al. [42]. The advancement of digital technology, while environmentally neutral in essence, is contingent upon the adaptation and utilization by its creators and users. Enterprises are driven to distort envi-
ronmental data for various reasons such as regulatory compliance, enhancing reputation, reducing costs, improving efficiency, and securing financing to meet specific economic goals [42]. Similarly, local governments may also have incentives to manipulate environmental information, particularly under the scrutiny of central government performance evaluations, prompting officials to whitewash environmental performance in pursuit of political advancement. China’s institutional setting and linguistic and textual attributes offer mechanisms for concealing environmental information. The absence of a uniform structure for disclosing environmental information in China grants environmentally accountable entities significant autonomy in determining the content of such information [43]. The intricacy of China’s language and writing provides environmentally conscious persons with enough opportunity to conceal or minimize unfavorable environmental information, or to magnify positive data [44].

Furthermore, digital technology has the potential to be utilized by corporations as a means to manipulate environmental data. As highlighted by Li et al. (2021), enterprises possess a significant advantage in terms of data control, scale, and access to advanced algorithmic technology for information extraction compared to consumers and investors [45]. This advantage enables enterprises to potentially distort economic signals for their own benefit using digital technology, thereby contributing to public confusion. Modern digital tools such as large language models and AI writing technologies further facilitate the dissemination of misleading environmental information. For instance, some individuals committed to environmental responsibility may leverage AI writing tools to create vague or inaccurate environmental content, which can then be shared online, exacerbating the disparity in green information dissemination.

Hence, digital technology serves a dual purpose in addressing information asymmetry related to environmental sustainability. Nevertheless, at present, the advantageous impacts of digital technologies outweigh any negative implications. Following the analysis presented, the subsequent hypothesis is posited.

**Hypothesis 2:** A U-shaped correlation exists between the advancement of the digital economy and information asymmetry, with information asymmetry levels expected to experience a moderate rebound during the later phases of digital economy development.

2.2.2. Capital Misallocation

Improving the efficiency of capital allocation is widely acknowledged to be a vital strategy for fostering sustainable development [46]. Efficient capital allocation involves reallocating capital resources among different economic entities, either through market forces or government interventions, with the goal of optimizing overall social output or achieving a Pareto-optimal equilibrium [47]. On the other hand, capital mismatch indicates a deviation from the Pareto-optimal state.

One reason for capital mismatch is the distortion of the capital market, which leads to capital prices not accurately reflecting the marginal product returns of capital. This hinders the optimal flow of capital towards achieving maximum output. The advancement of the digital economy has the potential to enhance the movement of capital factors by reducing information costs and effectively adjusting capital prices. The utilization of big data mining and analysis technology expands the information channels available to economic entities, facilitating a more efficient process of information acquisition, search matching, communication, and sharing. This contributes to a reduction in information costs [48]. The decrease in information search and acquisition expenses results in a wealth of information, leading to increased transparency. This transparency enables both the supply and demand sides of capital to thoroughly assess risk and return, thereby facilitating the dynamic adjustment of various capital prices. Consequently, this improves the price mechanism for allocating capital factors.

The second reason for capital mismatch is attributed to obstacles and limitations hindering the movement of capital factors. The advancement of the digital economy plays a
crucial role in dismantling these barriers to factor mobility. The establishment of digital financing platforms and the enhancement of regulatory frameworks have facilitated a more streamlined financing process, making it easier for individuals to participate in financing activities. Consequently, this leads to a reduction in transaction costs and principal-agent expenses [49]. The decreased transaction costs significantly broaden the geographical reach of capital factor movements, facilitating inter-regional capital flows. Furthermore, the inherent openness of the digital economy fosters efficient connections and reorganization of goods and services across different regions, thereby diminishing regional market segmentation and facilitating the unrestricted movement of capital factors between regions [50].

The efficiency of capital factor allocation may not always be enhanced by digitization. As the digital economy’s development model progresses, certain adverse factors that impede the enhancement of allocation efficiency begin to emerge, resulting in a non-linear correlation between digitization and capital mismatch. First of all, the advancement of digital finance is heavily reliant on the utilization of extensive data inputs. As market competition continues to evolve dynamically, the distribution of data resources among economic agents fluctuates, leading to the emergence of financial monopolies due to the Matthew effect (i.e., strengths tend to create additional strengths). In the contemporary digital landscape, the integration of capital and data is profound, with capital leveraging the network effect through digital platforms to achieve amplified returns to scale [51]. Disparities in the level of digital transformation and the capacity to innovate and implement digital technologies across financial institutions result in the reconfiguration of data components within the sector. The ongoing evolution of market competition may give rise to data asymmetry, culminating in a competitive model where a single entity dominates, known as a “winner-takes-all” scenario. Certain prominent digital financial platforms possess a significant share of data resources, utilizing their core data assets, transactional data, and financial data to establish a stronghold through data monopoly and digital division, thereby significantly altering the competitive landscape of the financial market and fostering a trend towards financial monopolization. The monopolization process is likely to lead to inefficiencies in resource allocation [51].

Moreover, the degree of information asymmetry within capital markets is intricately linked to the effectiveness of allocating capital resources. It is important to note that the advancement of the digital economy does not consistently diminish information asymmetry. Throughout the process of digitalization, the presence of specific interests among environmentally conscious individuals and deficiencies within the institutional framework may prompt environmentally responsible individuals to manipulate or withhold environmental protection information using digital tools. This manipulation can intensify information asymmetry, hindering the optimal flow of capital resources towards maximizing factor rewards and exacerbating the mismatch of capital resources.

Following the analysis presented above, the subsequent hypothesis is posited:

Hypothesis 3: A U-shaped correlation exists between the advancement of the digital economy and capital mismatch, with the degree of capital mismatch experiencing a moderate resurgence during the later phases of digital economy development.

3. Methods and Data
3.1. Algorithms and Data Training

A proposed approach involves the utilization of a random forest algorithm to examine the nonlinear correlation between the digital economy and urban green development, with the aim of confirming the presence of the digital rebound effect.

Random forest and other ensemble learning techniques offer distinct advantages compared to conventional methods. Firstly, prior studies have frequently relied on explanatory models that prioritize the selection of theoretical frameworks or emphasize particular value dimensions. This approach can introduce bias by incorporating preconceived theories, potentially leading to flawed causal inferences. In contrast, random forest algorithms do
not necessitate the establishment of a predefined model structure, thereby mitigating the influence of preconceived theories on the model configuration and regression process. Secondly, the conventional approach to panel data analysis is centered on assessing the collective marginal influence of particular explanatory factors on the variables being studied. This approach lacks the adaptability to effectively address fluctuations in marginal impact. In contrast, partial dependency graphs offer a visual representation that distinctly illustrates alterations in the marginal impact of explanatory variables on the explained variables. This visual aid enables policymakers to monitor shifts in economic performance and formulate appropriate policy interventions. Thirdly, conventional panel data analysis techniques may not offer a comprehensive evaluation of the significance of explanatory variables, as the regression coefficients' relative magnitudes in these models could primarily stem from differences in the units of the variables. Expanding upon the findings from the ensemble learning algorithm experiments, Shapley additive explanations can be employed to elucidate the model outcomes and assess the relative importance of the variables.

In accordance with the overarching principles of modeling in machine learning algorithms, the training data are randomly divided into an 80% training set and a 20% test set during the modeling process. Some modeling method is developed to mitigate problems around “overfitting” or “under-fitting.”

Initially, the approach of “random search+ grid search” is utilized for hyperparameter exploration. The process begins with a random search within the hyperparameter space. This method employs random sampling to approximate the optimal hyperparameter set during the selection process, potentially resulting in a suboptimal solution. Subsequently, a grid search is conducted within the vicinity of each previously identified hyperparameter set to approach the optimal hyperparameter set. To strike a balance between search accuracy and computational efficiency, both the randomized search and the grid search are subjected to triple-folded cross-validation.

Secondly, the hyperparameter search process incorporates multiple score evaluation criteria. Specifically, three score evaluation matrices, namely $R^2$, negative mean square error, and negative root mean square error, are concurrently utilized. It is important to note that higher scores in $R^2$, negative mean square error, and negative root mean square error signify superior model fitting.

Thirdly, the process of selecting hyperparameters and loss functions is crucial. In order to prevent overfitting, the random forest algorithm employs minimum cost complexity pruning and restricts the maximum depth of each decision tree. The range of the hyperparameter search encompasses the number of decision trees, the maximum depth of the tree, the number of features to consider when looking for the best split, the minimum number of samples required to split an internal node, the minimum number of samples required to be at a leaf node, and the cost complexity pruning parameter.

To investigate the relationship between feature variables and the response variable, Molnar (2022) [52] utilized a partial dependence plot (PDP) as a tool for elucidating machine learning models. The PDP illustrates the marginal influence of individual or paired features on the predictive outcomes of the machine learning model, thereby providing a more comprehensive representation of the nonlinear association between the feature variables and the response variables. Its expression is:

$$ \hat{f}_S(x_s) = \mathbb{E}_{X_c} \left[ \hat{f}_S(x_s, X_c) \right] = \int \hat{f}_S(x_s, X_c) dP(X_c) $$

where $\hat{f}_S$ denotes the partial function and $x_s$ is the feature variable of interest, which is $DE$ in this article. $X_c$ are all the other feature variables. The relationship between $x_s$ (i.e., $DE$) and the predicted outcome can be obtained by integrating over the other feature variables (the set of variables of no interest, $X_c$). In practice, the Monte Carlo method can be used to
estimate $\hat{f}_S$, which denotes the magnitude of the average marginal effect predicted by a given value of $x_s$, i.e.:

$$f_S(x_s) = \frac{1}{n} \sum_{i=1}^{n} f_S(x_s, x_c^{(i)})$$  \hspace{1cm} (2)$$

where $n$ denotes the number of samples and $x_c^{(i)}$ is the eigenvalue of the $i$th sample in $X_c$. It is usually assumed that $x_s$ is uncorrelated with $x_c^{(i)}$, and unrealistic data points may occur if the assumption is violated. Therefore, in the robustness test section, we use the cumulative local effects plot to relax this assumption.

Furthermore, we examine the magnitude and direction of the impact of each feature on the response variable using the Shapley value within the framework of the expression [53]:

$$\phi_j(v) = \sum_{S \subseteq x_1, x_2, \ldots, x_p} \frac{|S|!(p - |S| - 1)!}{p!} (v(S \cup x_j) - v(S))$$  \hspace{1cm} (3)$$

where $\phi_j(v)$ is the Shapley value of feature $j$, i.e., the contribution to the result, $S$ is the subset of features used in the model, $x_j$ is the vector of feature values to be interpreted, $p$ is the total number of features, $|S|!(p - |S| - 1)!/p!$ denotes the weight, $v(S)$ is the prediction of $S$ based on the marginal distribution of the features that are not in the set of $S$, and $v(S \cup x_j) - v(S)$ is the difference in prediction with the inclusion of the subset of features $S$. The Shapley value matrix of feature $j$ and the data matrix of the feature are computed to determine the correlation coefficients for judgment, which indicate the positive or negative effects of the feature on the response variable. A positive correlation coefficient indicates a positive effect, while a negative correlation coefficient indicates a negative effect.

3.2. Variables
3.2.1. Response Variable: Green Total Factor Productivity (GMLPI)

According to the study by Pang and Wang (2023) [5], the concept of green total factor productivity is employed as a surrogate measure for the extent of urban green development. In practice, we utilize the comprehensive Malmquist–Luenberger productivity index for measurement [54–56]. Firstly, we assume that the production technology is capable of producing $M$ desired outputs $y \in R^M_+$ and $J$ undesired outputs $b \in R^J_+$, and that the set of output $P(x)$ requires $N$ inputs as $x \in R^N_+$. Consider a $t = 1, 2, 3, \ldots, T$ period panel data with $k(= 1, 2, 3, \ldots, K)$ DMUs; thus, the production technology can be mathematically expressed as follows:

$$P^t(x^t) = \{ (y^t, b^t) : x^t \text{produce } (y^t, b^t) \}$$ \hspace{1cm} (4)$$

The global production technology set is defined as a union of all contemporaneous technology set, i.e., $P^G(x) = P^1(x^1) \cup P^2(x^2) \cup P^3(x^3) \cup \ldots \cup P^T(x^T)$. Building on the axioms of weak disposability, null-jointness, and the strong disposability of the desirable outputs. If we define $g = (g_y, g_b), \in R^M_+ \times R^J_+$ is the direction vectors in which the outputs should be scaled, and the corresponding directional distance function can be expressed as follows:

$$\overrightarrow{D^G_0} \left( x^t, y^t, b^t; g_y^*, g_b^* \right) = \max \{ \beta : (y^t + \beta g_y^*, b^t - \beta g_b^*) \in P^G(x^t) \}$$  \hspace{1cm} (5)$$

Furthermore, GML is defined as:

$$GML^{t+1} = \frac{1 + \overrightarrow{D^G_0} \left( x^{t+1}, y^{t+1}, b^{t+1}; g_y^{t+1}, g_b^{t+1} \right)}{1 + \overrightarrow{D^G_0} \left( x^{t+1}, y^{t+1}, b^{t+1}; g_y^{t+1}, g_b^{t+1} \right)}$$  \hspace{1cm} (6)$$
We assume a constant returns to scale (CRS) technology set $P^G(x)$, and the directional distance function can be solved by the following linear equation:

$$D^G_G(x^s, y^s, b^s, s^y, s^b) = \max \beta$$

(7)

subject to

$$\sum_{t=1}^{T} \sum_{k=1}^{K} z^t_k y^t_{km} \geq (1 + \beta) y^m_{m}$$

(8)

$$\sum_{t=1}^{T} \sum_{k=1}^{K} z^t_k b^t_{kj} = (1 - \beta) b^j_{j}$$

$$\sum_{t=1}^{T} \sum_{k=1}^{K} z^t_k x^t_{kn} \leq x^s_{n}$$

$$z^t_k \geq 0$$

$m = 1, \ldots, M; j = 1, \ldots, J; n = 1, \ldots, N; k = 1, \ldots, N$

where $z^t_k$ is the intensity variable.

In the measurement, the selected measure of output is the real GDP using 2010 as the baseline year. Undesired outputs are identified as sulfur oxide emissions from urban industry, industrial wastewater emissions, and industrial smoke (dust) emissions [57,58]. Based on the research conducted by Zhang (2004) [59], a depreciation rate of 9.6% is chosen for the physical capital. The perpetual inventory method is employed to compute the physical capital stock representing the physical capital inputs for the baseline year of 2010. Subsequently, the labor input is determined by the number of employees of each year, while the energy input is represented by urban energy consumption. As data on urban energy consumption are not readily available in the statistical yearbook, they are estimated based on the city’s contribution to the gross regional product of its respective province [60].

The index system for measuring urban green total factor productivity, as developed in this study, is presented in Table 1.

<table>
<thead>
<tr>
<th>Level I Indicators</th>
<th>Level II Indicators</th>
<th>Level III Indicators</th>
<th>Unit (of Measure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outputs</td>
<td>Desired outputs</td>
<td>Real GDP</td>
<td>Ten thousand yuan</td>
</tr>
<tr>
<td></td>
<td>Undesired outputs</td>
<td>Industrial sulfur dioxide emissions</td>
<td>Ton</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industrial wastewater emissions</td>
<td>10,000 Tons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industrial smoke (dust) emissions</td>
<td>Ton</td>
</tr>
<tr>
<td>Inputs</td>
<td>Physical inputs</td>
<td>Physical capital stock</td>
<td>1,000,000 Yuan</td>
</tr>
<tr>
<td></td>
<td>Labor inputs</td>
<td>Number of employed</td>
<td>10,000 people</td>
</tr>
<tr>
<td></td>
<td>Energy inputs</td>
<td>City energy consumption</td>
<td>10,000 tons of standard coal</td>
</tr>
</tbody>
</table>

3.2.2. Feature Variable of Most Interest: Digital Economic (DE)

Our measurement approach stems from the framework proposed by Zhao et al. (2020) [61], and we further extend it to evaluate the degree of digital economy development at the urban level across five key aspects: the establishment and application of local digital infrastructure, the workforce engaged, the output of relevant industries, the level of digital financial development, and digital innovation and entrepreneurship. The measurement index system is outlined in Table 2, and the entropy-weight-based TOPSIS method is utilized to calculate the overall development level index of the digital economy. This approach has the benefit of successfully dealing with concerns associated with intense subjectivity, restricted data, and inadequate information.
Table 2. The index system for measuring urban digital economic development.

<table>
<thead>
<tr>
<th>Level I Indicators</th>
<th>Level II Indicators</th>
<th>Unit (of Measure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The establishment and application of local digital</td>
<td>Cell phone penetration rate (cell phone subscribers per 100 population)</td>
<td>User</td>
</tr>
<tr>
<td>infrastructure</td>
<td>Mobile internet user size</td>
<td>User per person</td>
</tr>
<tr>
<td></td>
<td>Mobile switch capacity</td>
<td>10,000 households</td>
</tr>
<tr>
<td></td>
<td>Size of internet broadband access subscribers</td>
<td>User per person</td>
</tr>
<tr>
<td>Digital industry practitioners</td>
<td>Percentage of employees in the information transmission, computer services, and software industry</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>Percentage of employees in the transportation, storage, post, and telecommunications industry</td>
<td>%</td>
</tr>
<tr>
<td>Digital industry outputs</td>
<td>Total postal operations</td>
<td>10,000 Yuan</td>
</tr>
<tr>
<td></td>
<td>Telecommunications services per capita</td>
<td>Yuan per person</td>
</tr>
<tr>
<td>Level of development of digital finance</td>
<td>Digital inclusive finance index</td>
<td>-</td>
</tr>
<tr>
<td>Digital innovation and entrepreneurship</td>
<td>Digital innovation and entrepreneurship index</td>
<td>-</td>
</tr>
</tbody>
</table>

3.2.3. The Remaining Feature Variables

We have implemented a filter method to select feature variables in order to enhance the estimation scores of the integrated learning model. This approach involves eliminating highly correlated variables by utilizing the Pearson correlation coefficient. The chosen feature variables encompass population density (PD), financial development level (Fin, institutional deposit and loan balances/gross regional product), service-oriented industrial structure (PropSer, value added of tertiary industry/gross regional product), foreign direct investment (FDI, total foreign direct investment expressed in RMB), government’s emphasis on science and education (PropRE, the proportion of government’s expenditure on science and education), green innovation (PropGI, proportion of green patents/total patents), government environmental concern (EcoVoc, frequency of environmental protection words in municipal government work reports), and provincial innovation index (IntSc). Simultaneously, the unmeasurable factors throughout time and region are managed by transforming the province and year string type to an integer. Both the province and year are consistently transformed into integer feature variables, with zero as the starting point. Furthermore, as the decision tree splits using only one feature per split as the criterion, it is not affected by differences in variable magnitudes, and therefore there is no need to apply logarithmic processing to all feature variables.

3.2.4. Mechanism Variables

Capital mismatch index (KMI). As shown in Equation (9), we refer to the methods of Chen et al. (2011) [62] and Bai et al. (2018) [47] to measure the capital mismatch index,

\[ KMI_{it} = \frac{1}{\gamma_{K}^{it}} - 1 \]  \hspace{1cm} (9)

where \( \gamma_{K}^{it} \) is the absolute factor price distortion coefficient, which is generally replaced in the measurement by the relative factor price distortion coefficient \( \hat{\gamma}_{K}^{it} \).

\[ \hat{\gamma}_{K}^{it} = \left( \frac{K_{it}}{K_{t}} \right) f \left( \frac{s_{it}^{K}}{\beta_{t}^{K}} \right) \]  \hspace{1cm} (10)

In Equation (10), \( s_{it} = y_{it}/Y \) denotes the share of output \( y_{it} \) of region \( i \) in period \( t \) in total economy-wide output \( Y \). \( K_{it}/K_{t} \) denotes the share of capital used by region \( i \) in period \( t \) of the total capital used by the economy as a whole in period \( t \), which means the proportion of capital actually used by the city. And \( \beta_{t}^{K} = \sum_{i} s_{it}^{K} \beta_{t}^{K} \) indicates the value of output-weighted capital contributions, whereas \( \beta_{t}^{K} \) is the output elasticity of capital factors.
s_i(t)\beta^K_i / \beta^K_t is the theoretical proportion of capital used by city i when capital is efficiently allocated. The coefficient of relative distortion of capital factor prices \((K_{it} / K_t) / (\delta_i \beta^K_i / \beta^K_t)\) is determined by comparing the actual proportion of use to the theoretical proportion of use.

According to Equation (10), the determination of \(\hat{\gamma}^{K}_{it}\) necessitates an initial estimation of \(\beta^{K}_{it}\). At this juncture, the approach outlined by Zhao et al. (2006) [63] is utilized, wherein a panel randomly varying coefficient model is employed for the estimation of Equation (11).

\[
\ln(Y_{it} / L_{it}) = \beta_0 + \beta_1 \ln(K_{it} / L_{it}) + \epsilon_{it}
\]

In Equation (11), \(Y_{it}\) represents the real GDP, while the variables \(K_{it}\) and \(L_{it}\) denote the quantities of capital inputs and labor inputs, respectively, \(L_{it}\) being measured in relation to urban employment. Once the estimation of \(\beta^{K}_{it}\) has been completed, proceed to determine \(\gamma^{K}_{it}\) and then substitute it into Equation (9) in order to calculate \(KMI\). If the value of \(KMI\) is positive, it signifies a situation of under-allocation of capital resources, whereas a negative value indicates over-allocation. Absolute values of \(KMI\) are calculated to represent the total resource mismatch without considering the direction of the difference.

Information asymmetry index (ASY_{it}): the information asymmetry index of A-share listed companies in each city is employed as a measure to indicate the level of information asymmetry present in the local capital market. This involves computing the liquidity ratio index (LR), illiquidity ratio index (ILL), and yield return reversal index (GAM) based on the methodologies outlined in the works of Yu et al. (2012) [64] and Wu and Chang (2021) [65], utilizing daily trading data frequency from A-share listed companies in China’s stock market.

\[
LR_{kt} = \frac{1}{D_{kt}} \sum_{d=1}^{D_{kt}} \frac{V_{kd}^d}{r_{kd}^d}
\]

\[
ILL_{kt} = \frac{1}{D_{kt}} \sum_{d=1}^{D_{kt}} \frac{r_{kd}^d}{V_{kd}^d}
\]

\[
GAM_{kt} = |\delta_{kt}|
\]

In these equations, the variable \(D_{kt}\) represents the aggregate count of trading sessions within a given calendar year denoted as \(t\) for a publicly traded company identified as \(k\). \(V_{kd}^d\) denotes the volume of \(d\) trading day. \(r_{kd}^d\) represents the stock return on the \(d\) trading day. \(\delta_{kt}\) can be obtained by estimating \(\gamma_{kt}^d = \alpha_0 + \alpha_1 r_{kt}^d + \delta_{kt} V_{kd}^{d-1} \times \text{sign}(r_{kt}^{d-1}) + \epsilon_{kt}\), where \(\text{sign}\) is a sign function that takes the value 1 for greater than 0, -1 for less than 0, and 0 for equal to 0. Excess return \(r_{kt}^d = r_{kt}^d - r_{mt}^d\), where \(r_{mt}^d\) is the market return weighted by the market capitalization outstanding as a weight. The three aforementioned indicators underwent principal component analysis, resulting in the extraction of the initial principal component to establish a comprehensive assessment index for information asymmetry (ASY_{kt}). The rationale for using the above indicator to measure information asymmetry is that the higher the intensity of information asymmetry between the supply and demand sides of capital regarding the value of a firm’s assets, the more “lemon premiums” will be demanded by uninformed traders, resulting in more illiquidity of the stock, more price changes per unit of turnover, and more over-representation of the order flow, resulting in more reversal of yields. Following the acquisition of firm-level indicators, this study calculates the average of the ASY_{kt} indicators across cities and years to derive the city-level information asymmetry index denoted as ASY_{it}.

3.3. Sample and Data

This study employs unbalanced panel data from 267 prefecture-level cities in China, covering the period from 2011 to 2019, as a sample for the research. The digital inclusive finance index is derived from a cooperative endeavor between the Digital Finance...
Research Center of Peking University and Ant Technology Group [66]. The digital economy innovation and entrepreneurship index is derived from Peking University’s open research data platform [67]. The province innovation index is calculated based on the results of the 2011–2020 Evaluation of China’s Regional Innovation Capability Report. Governmental environmental concern is assessed by examining the frequency of certain terms in the research reports of prefecture-level cities, using a vocabulary produced by Dong and Wang (2021) [68]. And the other data are gathered from EPSDATA, CNRDS, China Urban Statistical Yearbook, and province statistical yearbooks.

During the data processing phase, any missing data are addressed through linear interpolation. In cases where an observation sample contains a significant amount of missing characteristic variables, it is excluded from the analysis. Additionally, the dataset is trimmed by 1% on both ends to mitigate the impact of outliers on the outcomes. The modeling and interpretation processes are conducted using the Python 3.9 interpreter, with algorithmic support from packages including sklearn, scipy, matplotlib, and shap.

4. Empirical Results and Analysis

4.1. Descriptive Statistics and Correlation Analysis

The results of the descriptive statistics of the variables are provided in Table 3, and the Pearson correlation coefficients between the characteristic variables are shown in Table 4, with the correlation coefficients between most of the characteristics being less than 0.3.

<table>
<thead>
<tr>
<th>Table 3. Descriptive statistics for variables.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Count</strong></td>
</tr>
<tr>
<td><strong>year</strong></td>
</tr>
<tr>
<td><strong>province</strong></td>
</tr>
<tr>
<td><strong>GMLPI</strong></td>
</tr>
<tr>
<td><strong>DE</strong></td>
</tr>
<tr>
<td><strong>PD</strong></td>
</tr>
<tr>
<td><strong>Fin</strong></td>
</tr>
<tr>
<td><strong>PropSer</strong></td>
</tr>
<tr>
<td><strong>FDI</strong></td>
</tr>
<tr>
<td><strong>PropRE</strong></td>
</tr>
<tr>
<td><strong>PropGI</strong></td>
</tr>
<tr>
<td><strong>EcoVoc</strong></td>
</tr>
<tr>
<td><strong>InvSc</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4. Pearson correlation coefficients.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year</strong></td>
</tr>
<tr>
<td><strong>year</strong></td>
</tr>
<tr>
<td><strong>province</strong></td>
</tr>
<tr>
<td><strong>DE</strong></td>
</tr>
<tr>
<td><strong>PD</strong></td>
</tr>
<tr>
<td><strong>Fin</strong></td>
</tr>
<tr>
<td><strong>PropSer</strong></td>
</tr>
<tr>
<td><strong>FDI</strong></td>
</tr>
<tr>
<td><strong>PropRE</strong></td>
</tr>
<tr>
<td><strong>PropGI</strong></td>
</tr>
<tr>
<td><strong>EcoVoc</strong></td>
</tr>
<tr>
<td><strong>InvSc</strong></td>
</tr>
</tbody>
</table>

4.2. Analysis of Random Forest Algorithm Results

The results of the estimation using the random forest method are shown in Figures 1 and 2, while the associated scores can be found in Table 5. Figure 1 illustrates the impact of various features on the outcome of the model using the Shapley value. This
analysis reveals both the positive and negative effects of each feature variable on green
development, as well as the ranking of their contributions. Figure 2 displays a partial
dependency plot (PDP) illustrating the relationship between the feature variable $DE$ and
the predicted (average) green total factor productivity. The horizontal axis represents the
distribution of $DE$, while the vertical axis shows the predicted (average) green total factor
productivity. Figure 2 visualizes the non-linear marginal impact of the digital economy on
green development.

![Figure 1. Impact of features on response variable calculated using Shapley values.](image1)

![Figure 2. Partial dependence between response variable and the digital economy.](image2)

### Table 5. Scores of the random forest algorithm.

<table>
<thead>
<tr>
<th>Method</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>0.195</td>
<td>0.063</td>
<td>0.040</td>
</tr>
</tbody>
</table>

Note: All scores are calculated based on the test set.

According to Figure 1, the digital economy has a positive impact on the increase
in green total factor productivity, and is the second most influential factor. This finding
aligns with previous empirical research. In addition, the estimation findings demonstrate
a positive connection between the regional innovation index and green development.
Furthermore, the government’s focus on environmental conservation and advancements
in science and education, along with the degree of service-oriented industrial structure, durability of green innovation capabilities, foreign direct investment, and population density all exhibit a positive correlation with increased green total factor productivity. However, the degree of financial development has not made the anticipated beneficial impact on green development, indicating that the full potential of green finance has not been completely realized.

Figure 2 illustrates an inverted U-shaped link between the digital economy and green development. However, the latter half of the inverted U-shape does not exhibit a significant fall, but rather demonstrates a converging tendency. This suggests that the digital rebound effect will occur during the later phase of the digital economy’s growth, but the rebound will be small in magnitude. More precisely, when the digital economy (DE) is below 0.065, it has the potential to significantly enhance green growth, demonstrating a clear facilitative impact. At the coordinates [0.065, 0.08], the digital economy has a negligible impact on the degree of urban green growth. When the digital economy (DE) surpasses a threshold of 0.08, the degree of green development experiences a substantial decline. Additionally, there is a tendency for convergence or stabilization.

4.3. Robustness Checks

An accumulated local effects (ALE) plot, is a graphical representation that shows the cumulative effect of a variable on an algorithm’s predictions. The PDP implies that there is no connection between the features. However, if the feature variables are associated with each other, and they may result in unrealistic data points during the computation process. This, in turn, might lead to incorrect causal inference. Thus, the assumption of correlation among feature variables is relaxed, and an ALE plot is generated to confirm the validity of the results. Figure 3 illustrates that the horizontal axis of the figure represents the distribution of the feature variables, while the vertical axis represents the ALE value. The ALE value measures the impact of a certain value of DE relative to the average prediction of the data. For example, an ALE value of −0.002 suggests that the prediction value is 0.002 lower compared with the average prediction around DE = 0.055. The trend of the ALE plot closely aligns with that of the PDP graph, providing evidence for the accuracy of the aforementioned conclusion.

![Figure 3. ALE plots for the urban green development prediction model by digital economics.](image-url)

Due to the concealment of variability among individuals in PDP, the process of averaging individual effects may be susceptible to the influence of extreme values. Hence, a composite PDP and individual conditional expectation (ICE) graph is depicted in Figure 4. Each solid gray line in the illustration represents the correlation between a single characteristic variable (DE) and the forecasted value for a randomly selected individual. As can
be seen from the Figure, it is evident that the majority of the observed sample follows the economic rules mentioned above, indicating the reliability of the results.

**Figure 4.** Composite PDP and ICE plots of predicted urban green development by digital economics.

Elimination of municipality samples. Due to the unique administrative status of the municipality directly under the central government, its advanced level of digital economy development, and its larger population and land area, the municipality directly under the central government sample is excluded from the algorithm program’s rerun. The estimation findings are displayed in Figure 5, indicating that the influence of the digital economy on the city’s green development exhibits an initial quick growth followed by a gradual leveling off in the later phase.

**Figure 5.** Partial dependence between response variable and the digital economy without municipality samples.

In addition, we re-evaluated the digital rebound effect using the Bagging algorithm and the gradient-boosting decision tree (GBDT) algorithm, and compared three major ensemble learning algorithms based on tree models, and the related results are shown in Appendix A. The findings presented in the appendix demonstrate that the random forest algorithm effectively addresses issues related to overfitting and underfitting while exhibiting a high level of predictive accuracy. Furthermore, the PDP generated from the data using alternative integrated learning algorithms closely resemble those produced by the random forest algorithm. This suggests that the conclusions drawn in this study are reliable and consistent.
4.4. Analysis of Heterogeneity

In addition, this study updates the random forest model using the three characteristic variables of regional digital policy (PolicySup), city rank (Rank), and geographic location (Region), which may contribute significantly to the geographic heterogeneity of the digital economy’s impact on green development. Figures 6–8 display the respective 3D PDP, while Table 6 presents the related model scores.

![Figure 6](image1.png)

**Figure 6.** Partial dependence between response variable and the digital economy, geographic location.

![Figure 7](image2.png)

**Figure 7.** Partial dependence between response variable and the digital economy, city rank.

![Figure 8](image3.png)

**Figure 8.** Partial dependence between response variable and the digital economy, regional digital policy.
Table 6. Scores of the random forest algorithm for heterogeneity analysis.

<table>
<thead>
<tr>
<th></th>
<th>$R^2$</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locational</td>
<td>0.192</td>
<td>0.063</td>
<td>0.040</td>
</tr>
<tr>
<td>heterogeneity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank heterogeneity</td>
<td>0.175</td>
<td>0.064</td>
<td>0.041</td>
</tr>
<tr>
<td>Policy heterogeneity</td>
<td>0.200</td>
<td>0.063</td>
<td>0.040</td>
</tr>
</tbody>
</table>

Note: All scores are calculated based on the test set.

The geographic locations of the eastern, central, western, and northeastern regions are marked as 1, 2, 3, and 4 in accordance. The outcomes of the random forest method after rerunning are presented in Figure 6. The phenomena of digital rebound are evident in all four major regions of China. However, the impact of the digital economy varies among areas, resulting in a significant disparity in the level of convergence. When $Region = 1$, the urban green development exhibits the maximum degree of partial dependency on $DE$. As the value of $Region$ grows, the degree of partial dependence steadily diminishes, and the point of convergence is positioned at a lower level. These findings indicate that the eastern area experiences the most significant beneficial impact of digital empowerment, whereas the central, western, and northeastern regions see a decreasing effect in that respective order. The reason for this is the well-established digital economy foundation in the eastern and central regions, which can effectively promote the green development of cities through digital technologies. In contrast, cities in the western and northeastern regions are still in the early stages of digital economy growth, and the benefits of digital empowerment have not been fully realized. As a result, there is still significant untapped potential for green development in these areas.

Beyond that, a city's rank is marked as 1 if it is a provincial capital, sub-province, or municipality, and as 0 if it is any other type of city. The outcomes achieved by rerunning the algorithm are displayed in Figure 7. Empirical evidence indicates that the degree of partial dependency remains rather consistent regardless of the value assigned to the feature variable $Rank$. Furthermore, the impact of the digital economy on the heterogeneity of cities at different levels is not prominent. This phenomenon may be attributed to the spillover effect of the capital city on other cities. Moreover, the digital economy has the potential to expedite the mutually beneficial growth of both central and non-central cities [69].

In the context of regional digital policy support, the National Big Data Comprehensive Pilot Zone city is marked as 1, while other cities are marked as 0. The findings are displayed in Figure 8. Despite the variety of regional digital policies, the digital rebound effect exists, but the ultimate convergence position does not exhibit considerable difference. In general, the partial dependence level of $PolicySup = 1$ is marginally higher than $PolicySup = 0$. This indicates that establishing a comprehensive big data pilot area or supporting digital policies may speed up short-term digital development and the urban green transformation process. Given the districts supported by policy and those not supported by policy ultimately converge at the same place, it may be inferred that the influence of digital policy appears to be less apparent in the long term. This may be attributed to the short-sightedness of local governments, or it could be due to the incomplete recognition of the long-term consequences of the policies.

4.5. Mechanism Analysis

Based on theoretical analysis, it is posited that a reduced level of capital mismatch and information asymmetry leads to accelerated growth in urban green development. This study employs a partial dependency graph to investigate the potential mechanism by assessing the marginal influence of the digital economy on capital mismatch and information asymmetry. The outcomes of the estimation process conducted through the random forest algorithm are depicted in Figures 9 and 10, with the associated model scores presented in Table 7.
The outcomes of conducting hypothesis 2 testing are depicted in Figure 9, with the x-axis representing the degree of advancement in the digital economy and the y-axis, indicating the anticipated urban information asymmetry index. The findings indicate a notable reduction in information asymmetry as a result of the progress in digital economy development. Simultaneously, it is important to highlight that surpassing the threshold of 0.07 for variable $DE$ results in a marginal rise in the information asymmetry index. Moreover, when $DE$ exceeds 0.1, there is no discernible association between the digital economy and information asymmetry. Given the negative relationship between information asymmetry and the extent of urban green progress, altering the urban information asymmetry mechanism gives rise to a scenario where an “inverted U-shape” connection between the
digital economy and urban green development exists. Consequently, hypothesis 2 has been validated.

The outcomes of testing hypothesis 3 are illustrated in Figure 10, with the x-axis representing the degree of advancement in digital economy and the y-axis indicating the anticipated urban capital mismatch index. The findings indicate a general reduction in the capital mismatch index due to the development of the digital economy. Simultaneously, it is important to highlight that once $DE$ surpasses the threshold of 0.115, the impact of $DE$ in alleviating capital mismatch becomes less discernible, and could potentially marginally elevate the capital mismatch index. Given that a reduced degree in capital mismatch facilitates the attainment of an elevated level of urban green development, altering the urban capital mismatch mechanism results in the emergence of an “inverted U-shaped” correlation between the digital economy and urban green development, thereby confirming hypothesis 3.

5. Discussion

5.1. The Relationship between Digitalization and Urban Green Development

Real-world economists have taken a keen interest in the more intricate relationship between digitization and the process of greening cities, which has emerged as a significant factor influencing the latter. The findings of our investigation offer fresh perspectives on this issue: the process of digitization does not consistently contribute to the promotion of green practices in cities. Instead, the phenomenon of digital rebound is prevalent in the deployment of urban development. To test this idea, we acknowledged the constraints of conventional econometric techniques and instead utilized an ensemble learning algorithm to estimate the correlation between the two variables. Additionally, we employed a partial dependence plot (PDP) to visually represent the relationship. The empirical findings indicate that when the level of digitization in the city is below 0.08, there is a rapid rise in the city’s green development. However, when the digitization level exceeds 0.08, a digital rebound occurs, leading to a significant decline in the city’s green development as digitization progresses. This decline eventually converges to a certain level. This provides compelling evidence that the phenomena of digital rebound are prevalent in China’s urban development process.

A discussion of these results would be interesting. Our results offer evidence that supports the theory of digital rebound. Coroamă and Mattern (2019) argue that digitization is causing a transformation in societal metabolism [6]. The increased efficiency or reduced resource costs linked to digitization are stimulating a surge in consumption, which could potentially surpass the advantages of efficiency improvements. We affirm the existence of the digital rebound phenomena, specifically in the advanced phases of digital advancement. No indication of digital rebound appears during the pre-digital growth period of cities. Moreover, our empirical evidence substantiates the concept of a ‘modest’ digital rebound, as suggested by Coroamă and Mattern (2019) [6]. This implies that the extent of digital rebound is not substantial enough to cause a noteworthy decline in the degree of green growth in cities. This could be associated with the concurrent effects of digitization, as suggested by Lange et al. (2020) [7] or Avom et al. (2020) [8], who exhibit resemblances. Although digitalization has the potential to enhance energy efficiency, instigate structural changes, and modernize industries, it also results in resource consumption rising and economic scope expanding. The digitization process has both negative and positive effects. Negative elements result in digital rebound, while positive aspects contribute to the digital economy’s ability to promote green development and reduce the extent of digital rebound.

Secondly, there is not much empirical evidence on the phenomenon of digital rebound, and to the best of our knowledge, we find that the above estimation results are similar to those of Ahmadova et al., who, using global panel data on firms in 47 countries over the period 2014–2019, conclude that there is an inverted U-shape relationship between home country digitization and environmental performance [10]. Differently from their study, we argue that this inverted U-shaped relationship is incomplete; in other words, our results
find that the second half of the inverted U-shape does not decline significantly, but rather converges gradually. This implies that the rebound effect is not large enough to reduce the level of green development to a significant extent.

Still, some scholars contend that digitalization does not diminish the extent of environmentally friendly development. For instance, Liu and Wang (2022) posit that the digital economy is fostering green and high-quality urban development, with a more pronounced positive effect observed in regions with advanced levels of digitization [20]. This perspective contrasts with our research findings, which indicate that a digital rebound is more probable in highly digitized areas. Additionally, Zhang and Wang (2023) assess the degree of inter-provincial digital economy advancement in China and evaluate the influence of the digital economy on green total factor productivity [12]. They assert that the digital economy, within China’s new developmental phase, notably boosts provincial green total factor productivity without any adverse repercussions. This stance also diverges significantly from our study outcomes.

5.2. Why Does Digital Rebound Exist?

Digitization’s enhanced economic efficiency prompts human beings to consume a greater quantity of resources, resulting in the occurrence of digital rebound. This is the primary definition and the simplest explanation of digital rebound. Nevertheless, we provide novel perspectives on this issue. We contend that the digital rebound phenomenon may arise due to alterations in certain factors influencing green development as a result of the digitization process, resulting in a shift in the level of green development. The prevailing economic perspective posits that reducing capital mismatch and addressing information asymmetry about green initiatives can facilitate the advancement of sustainable development. Our research indicates that digitization initially decreases the extent of capital mismatch and green information asymmetry in urban areas. However, as digitization progresses, some dynamics change happen and give rise to the phenomenon known as digital rebound.

From an economic perspective, it is believed that the presence of green information asymmetry has a significant role in determining the extent of green development. Reducing information asymmetry has the potential to facilitate green development. Empirical research indicates that digitalization has the ability to decrease the degree of information asymmetry. For instance, Du (2023) [36] demonstrates that the digitalization of companies enhances internal governance and reinforces external monitoring, hence diminishing the extent of information asymmetry within and outside the organization. The information effects of the digital economy are documented in our study, representing a significant facet of digitization. At the same time, we acknowledge the information barriers arising from the digital economy. The empirical orientation of our investigation is guided by the theoretical research conducted by Li et al. (2021) [45]. Our findings illustrate the impact of digitization on information and the challenges that arise from the growth of the digital economy. Figure 9 demonstrates that the correlation between digital economic development and information asymmetry is expected to follow an U-shaped pattern, rather than a linear one.

Our study stands out from other studies by addressing the topic of whether digitalization consistently reduces the degree of capital mismatch. Multiple studies have discovered that the digital economy has the ability to rectify discrepancies in capital allocation. Sun et al. (2023) discovered that the process of enterprise digital transformation yields benefits such as enhancing the information environment, improving corporate governance, and alleviating financing constraints. These outcomes serve to enhance the enterprise’s investment decision-making processes and facilitate the efficient allocation of capital within the organization [48]. Zhang and Wang (2023) find that through fostering widespread market competition and advancing technological innovation, the digital economy assists in tackling resource mismatch at the municipal level [70]. Our study recognizes that digitalization has the capacity to reduce the capital mismatch, yet the challenges that arise in the subsequent phases of digital development need to be taken into account. We con-
tend that certain ecologically conscientious entities would employ digital technology to manipulate environmental data in order to conceal or downplay its negative impact, as a result of divergent economic objectives. Furthermore, the issue of monopoly resulting from digitalization should not be overlooked. Ye and Fan (2023) [71] conducted a study that supports our perspective. They discovered that when the digital economy is at a low level of development, its growth can greatly alleviate the capital mismatch. However, as the digital economy continues to develop, the effectiveness of this improvement steadily diminishes. Our study employs a ensemble learning approach for regression and acquires empirical validation that aligns with the findings of Ye and Fan (2023) [71].

5.3. Policy Implications

Based on the primary findings outlined in this research, we propose the following policy suggestions.

5.3.1. Reforming the Governance Framework of the Digital Economy

The digital rebound phenomenon serves as a cautionary signal to governments against solely prioritizing digital inputs and unchecked growth of digital productivity. Instead, it is imperative to enhance governance of the digital economy and direct its development in alignment with green development principles through institutional regulations. These measures are crucial strategies amidst the ongoing digital transformation.

Initially, it is imperative for the government to establish a transparent and inclusive structure for overseeing the digital economy and to promote broader engagement of various stakeholders in the policy-making procedures. This involvement should encompass governmental bodies, businesses, consumers, and non-governmental organizations. Through the implementation of a multilateral communication and consultation mechanism, the government can guarantee that the perspectives of all involved parties are taken into account, thereby facilitating the development of more holistic and equitable policies.

Furthermore, in addressing the challenges associated with the increased energy consumption and emission of pollutants resulting from the expansion of the digital economy, it is recommended that the government actively advocate for the advanced restructuring of the industrial sector. This can be achieved by endorsing the growth of innovative technologies such as artificial intelligence, blockchain, and the Internet of Things, while also phasing out energy-intensive and environmentally harmful enterprises in a timely manner. These actions are essential for fostering the improvement of green total factor productivity. Additionally, the government can play a role in reducing the utilization of high-energy-consuming and high-emission goods and services by promoting energy-efficient alternatives. This approach can facilitate the evolution and enhancement of consumption patterns and consumer behaviors, ultimately aiding in the reduction in pollutant emissions from the demand side.

In addressing the issue of energy rebound stemming from the utilization of digital technologies, governmental intervention can be employed to promote the adoption of energy-efficient technologies. This can be achieved through the establishment of rigorous energy efficiency standards and regulations, as well as by incentivizing individuals and businesses to curtail energy usage via the implementation of carbon levies or the provision of subsidies for energy-saving initiatives. Furthermore, public awareness campaigns and educational initiatives can be utilized to elucidate the energy rebound phenomenon, emphasizing the necessity for alterations in consumption patterns and lifestyles alongside technological advancements to mitigate energy consumption challenges. Additionally, the government may consider implementing a variable pricing mechanism that adjusts energy costs based on demand fluctuations, thereby encouraging consumers to utilize energy during off-peak periods and effectively managing energy demand.
5.3.2. Effectively Leveraging the Information and Capital Allocation Impacts of the Digital Economy

Initially, it is imperative for the government to implement a robust data management framework. The government ought to provide direction to businesses and institutions in establishing effective data management systems to uphold data accuracy, integrity, and security. This entails formulating explicit data management guidelines, procedures, and benchmarks, alongside implementing suitable technological measures for data safeguarding. The government should foster collaborative data sharing and innovation, facilitate international cooperation and information sharing, and expedite advancements in science, technology, and societal development. Simultaneously, emphasis should be placed on safeguarding intellectual property rights and individual privacy.

Moreover, enhancing pertinent legislation, such as data security and personal information protection laws, is essential to promote the judicious utilization of digitized data. These measures are crucial for overseeing the acquisition, retention, transfer, and utilization of data, safeguarding individual privacy, and upholding intellectual property rights.

Finally, governmental authorities need to monitor the influence of the digital economy on monopolistic practices, enforce anti-monopoly measures promptly, regulate industry competition levels, and offer tailored support to foster the growth of small- and medium-sized enterprises. The government ought to direct digital finance towards facilitating green innovation, advancing scientific and technological progress, and bolstering small- and medium-sized enterprises (SMEs). This can be accomplished by promoting a varied and multi-layered approach to digital financial services, and creating a complete digital service platform to enhance the efficiency of allocating resources.

5.4. Limitation and Future Work

This research is subject to intrinsic restrictions. The study was constrained by the limited availability of sources of data and the methodologies employed for data collecting, which hindered a significant expansion in the size of the study sample. This constraint encompassed both temporal and spatial dimensions, making it more challenging to extend and contrast the research across various time periods and locations. Future investigations should expand the temporal and spatial scopes when constructing datasets, and apply comparable techniques to comparisons that include several countries and regions. It is important to focus on geographies and urban areas outside of China. These investigations will provide significant information for investigating the extent to which the digital rebound phenomena can be applied to other situations and for developing a more profound understanding of its root causes.

In addition, it is acknowledged that measurement inaccuracies are an inherent aspect of indicator assessment. The exact influence of these errors on study results cannot be conclusively determined. Hence, future research efforts should prioritize improving the measuring techniques and indicator frameworks related to the development level of the digital economy or the green development level of cities. Improvements in measurement methodologies are crucial for strengthening the monitoring of economic performance and investigating the significance of economic concepts.

6. Conclusions

Theoretical and practical research have increasingly highlighted the detrimental effects of the digital economy on green development as urban digitalization develops. Several scholars have introduced and formulated the notion of digital rebound, arguing that the digital economy does not perpetually contribute to the advancement of environmentally friendly development. They argue that once digitization surpasses a specific threshold, the digital rebound effect emerges and undermines the enforcement of green development. The objective of our research is to enhance comprehension of the digital rebound effect. To achieve this, we employ panel data from Chinese cities spanning the years 2011 to 2019. We utilize an ensemble learning (EL) algorithm to train the data, and employ a partial
dependency plot (PDP) to illustrate the non-linear correlation between the digital economy and the green development of cities. The primary findings of our study are as follows:

- The relationship between the digital economy and urban green growth follows a non-linear pattern that suggests an inverted U shape. As digitalization advances, the degree of green development first increases, but then gradually drops and eventually becomes stable after surpassing a specific threshold. This implies the presence of a digital rebound impact in urban economic growth, yet the magnitude of the rebound is not large. Nevertheless, the digital economy might still contribute to enhancing the overall level of green development.

- An empirical study of regional capital markets found that the information effect and the capital allocation effect are the mechanisms by which the digital economy affects green development, and there is a “U-shaped” relationship between the digital economy and information asymmetry and capital mismatch. As the urban digitization progresses, the Random Forest algorithm anticipates a gradual decrease in information asymmetry and capital mismatch levels, followed by a slight rebound and eventual convergence. At a sufficiently high level of digitization, the correlation between digital economy advancement and information asymmetry or capital mismatch becomes less discernible. These observations suggest that in the advanced phases of digital economy evolution, the facilitative impact of digitization on the city’s green development diminishes, potentially leading to outcomes contrary to the initial expectations.

- The phenomenon of the digital rebound effect is present in most areas of China, while there are variations in the magnitude of the digital empowerment impact and exactly where urban green developments ultimately converge. The impact of the digital economy on green development degrees is more pronounced in eastern and central China compared to western and northeastern China. Consequently, the final convergent green development degrees in the former regions are significantly greater. The presence of spillover effects minimizes the variation in the impact of the digital economy’s development on the green development degree between cities of different tiers. Comprehensive big data pilot zones exhibit more green growth in the short term, but in the long run, they have comparable levels of convergence in green development to cities that receive no assistance from digital policies.

Our research addresses the ongoing academic debate surrounding the digital rebound effect and contributes significantly to the existing knowledge on this subject. Contrary to popular belief among scholars, digitalization is not without its environmental drawbacks, and the expansion of urban digitalization may lead to outcomes akin to the Jevons Paradox (i.e., enhancements in energy efficiency have the potential to elevate energy consumption levels beyond their original projections when real energy prices remain constant). However, our study indicates that concerns regarding a substantial decline in green development in the advanced stages of urban digitalization may be unwarranted. Therefore, we recommend that city policymakers avoid placing undue emphasis on digital inputs in their future planning, particularly in the later phases of digital development. Instead, they should prioritize the environmentally conscious application of digital technologies and carefully consider the potential risks associated with excessive digital infrastructure expansion.

**Author Contributions:** Methodology, Z.L.; Software, Z.L.; Validation, Z.L.; Formal analysis, Z.L.; Investigation, Y.P.; Resources, Y.P.; Data curation, Z.L.; Writing—original draft, Z.L.; Writing—review & editing, Y.P.; Supervision, Y.P.; Project administration, Y.P.; Funding acquisition, Y.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.
Data Availability Statement: Data supporting the results reported in this article can be found at https://www.stats.gov.cn/sj/ (accessed on 1 November 2023).

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

Appendix A

Three aspects—learning curves, model scores, and partial dependency plot (PDP)—are compared in this paper so as to illustrate the differences between the estimation results and learning process of the bagging algorithm, random forest algorithm, and gradient-boosting decision tree (GBDT) algorithm. During the modeling process, the GBDT method employs the Huber loss function as a substitute for the absence of a single mean square error (MSE) and root mean square error (MAE) loss function. In contrast, the remaining algorithms utilize the squared loss function to minimize the L2 loss.

To get things started, a comparison is made between the differences in learning curves. Figures A1–A3 display the learning curves for the bagging method, random forest algorithm, and GBDT algorithm, respectively. The diagram illustrates a horizontal axis representing the size of training samples, with the number of instances progressively increasing from left to right. The vertical axis represents the MSE, which measures the proximity of the predicted value to the actual value. A lower MSE implies a better-trained model. The picture displays the trend of MSE in the cross-validation set using an upper triangle solid line, while the trend of MSE in the training set is represented by a lower square-dashed line. Typically, the MSE of the cross-validation set and the training set increasingly get closer to each other as training continues. A high position of convergence MSE suggests a significant bias issue, indicating that the model is “under fitting”. A large gap between the cross-validation set and the training set, as depicted by the distance between the two lines in the picture, suggests a high variance issues and the possibility of “overfitting”. This implies that the model’s capacity to generalize is weak. When operating an algorithm, it is crucial to carefully consider the bias-variance dilemma. The MSE of the three methods mentioned above converges to a value lower than 0.0040. Additionally, the MSE of the cross-validation set reduces progressively during the training process, demonstrating an improvement in the model’s generalization capacity.

![Figure A1. Learning curve for the bagging algorithm.](image-url)
Figure A1. Learning curve for the bagging algorithm.

Figure A2. Learning curve for the random forest algorithm.

Figure A3. Learning curve for the GBDT algorithm.

Upon comparing the bias, it is obvious that all three algorithms exhibit a low level of bias, with minimal variation among them. When considering variance, the GBDT method exhibits the least variance, followed by the random forest algorithm with the second least variance, and the bagging algorithm with the greatest variance (although it may still have less variance than the decision tree algorithm). When considering the complexity of models or tuning difficulties, the bagging method is less complicated than the random forest method, which in turn is less complex than the GBDT algorithm. Upon comparing the bias, it is obvious that all three algorithms exhibit a low level of bias, with minimal variation among them. When considering variance, the GBDT method exhibits the least variance, followed by the random forest algorithm with the second least variance, and the bagging algorithm with the greatest variance (although it may still have less variance than the decision tree algorithm). When considering the complexity of models or tuning difficulties, the bagging method is less complicated than the random forest method, which in turn is less complex than the GBDT algorithm.
decision tree algorithm). When considering the complexity of models or tuning difficulties, the bagging method is less complicated than the random forest method, which in turn is less complex than the GBDT algorithm.

Subsequently, $R^2$, root mean square error (RMSE), and MAE serve as evaluation criteria to assess and contrast the disparities in model performance. A higher $R^2$ value and lower RMSE and MAE values imply superior model performance. The scores for the three methods are displayed in Table A1. In terms of performance, the GBDT algorithm achieves the greatest score and exhibits the smallest gap between predicted and true values. The random forest method ranks second in terms of performance, while the bagging algorithm has the lowest score.

Table A1. Scores of the bagging, random forest, and GBDT algorithm.

<table>
<thead>
<tr>
<th>Method</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bagging</td>
<td>0.161</td>
<td>0.064</td>
<td>0.041</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.195</td>
<td>0.063</td>
<td>0.040</td>
</tr>
<tr>
<td>GBDT</td>
<td>0.207</td>
<td>0.062</td>
<td>0.040</td>
</tr>
</tbody>
</table>

Note: All scores are calculated based on the test set.

Ultimately, the PDP of the three algorithms is reviewed. Figures A4 and A5 display the PDP generated by applying the bagging method and GBDT algorithm, respectively. The PDP of the random forest approach is depicted in Figure 2. Through observation and comparison, it is evident that the trend of partial dependence produced from running the three methods is basically the same. The digital economy’s effect on green development is largest between the range of 0.07 and 0.08. If the digital economy exceeds 0.08, it would hinder green growth. However, there is a possibility of a rebound process, eventually stabilizing at a particular level. Nevertheless, there are disparities in the PDP of the three techniques. The random forest method produces smoother curves in comparison to the bagging and GBDT algorithms. The reason for this is that each tree in the random forest algorithm is trained using the random sampling of data, and the predictions of several decision trees are then averaged. Introducing randomization aids in diminishing noise and volatility, leading to a more consistent trend in the PDP.

Figure A4. Partial dependence between response variable and digital economy using the bagging algorithm.
is largest between the range of 0.07 and 0.08. If the digital economy exceeds 0.08, it would hinder green growth. However, there is a possibility of a rebound process, eventually stabilizing at a particular level. Nevertheless, there are disparities in the PDP of the three techniques. The random forest method produces smoother curves in comparison to the bagging and GBDT algorithms. The reason for this is that each tree in the random forest algorithm is trained using the random sampling of data, and the predictions of several decision trees are then averaged. Introducing randomization aids in diminishing noise and volatility, leading to a more consistent trend in the PDP.

Figure A4. Partial dependence between response variable and digital economy using the bagging algorithm.

Figure A5. Partial dependence between response variable and digital economy using the GBDT algorithm.

To summarize, the GBDT method achieves the best scores and effectively balances bias and variance. However, it comes with a larger model complexity and modeling challenges, resulting in more volatile partial dependence. The bagging method exhibits the lowest scores and shows a relatively limited capacity for preventing overfitting. However, the modeling technique associated with it is reasonably straightforward. The random forest algorithm is characterized by its intermediate nature, commonly showing a more equitable distribution and yielding a smoother PDP. This makes it particularly well-suited for investigating the relationship between the features and the response variables. Hence, the GBDT method is well-suited for situations requiring precise prediction accuracy, while the bagging algorithm is more appropriate for modeling analysis during the data exploration period. On the other hand, the random forest algorithm offers a comparably balanced solution.

References

10. Ahmadova, G.; Delgado-Márquez, B.L.; Pedauga, L.E.; Leyva-de la Hiz, D.I. Too good to be true: The inverted U-shaped relationship between home-country digitalization and environmental performance. Ecol. Econ. 2022, 196, 107393. [CrossRef]


36. Du, K. Enterprise Digital Transformation, Information Asymmetry and the Risk of a Stock Price Crash. *China Circ. Econ.* **2023**, *141–144*. Available online: [https://kns.cnki.net/kcms2/article/abstract?v=_TL_exAJKC-MFHsXnKzZCcMv3xyuxPaLgDm-ks9hvFQaLrZoZH8ioLMzFMS_nOV-goWY23RhA6rZcQgFZuaOkhupHHmpwCPWoMvYyg8I2EFA6aOob-R3hJxkgpEg8CceEVD6PdK4984ppC7fI1XCig==&uniplatform=N2ZKPTl&language=CHS](https://kns.cnki.net/kcms2/article/abstract?v=_TL_exAJKC-MFHsXnKzZCcMv3xyuxPaLgDm-ks9hvFQaLrZoZH8ioLMzFMS_nOV-goWY23RhA6rZcQgFZuaOkhupHHmpwCPWoMvYyg8I2EFA6aOob-R3hJxkgpEg8CceEVD6PdK4984ppC7fI1XCig==&uniplatform=N2ZKPTl&language=CHS) (accessed on 16 May 2024). [CrossRef]


38. Goldfarb, A.; Tucker, C. Digital Economics. *J. Econ. Lit.* **2019**, *57*, 3–43. [CrossRef]

40. Catalini, C.; Gans, J.S. Some simple economics of the blockchain. *Commun. ACM* 2020, 63, 80–90. [CrossRef]

41. Lu, S.; Wu, C.; Jiang, Y.; Du, R. Can Social Media Information Disclosure of Companies Reduce Information Asymmetry? Comparative Study of GEM Market and Main Board Market. *Econ. Probl.* 2021, 121–129. Available online: https://kns.cnki.net/kcms2/article/abstract?v=TI_exAJK69p2uoEL3SeaGCHY3Rt32ZLBUPY_JRgs0UCBIO4hIgNB_Pgw9iy-gNwlonUd2q6hN5HN52V1sgsYs8D9mn2hsozUG3MFSUv_wSZ3uDF1L8moWuvcGRDU_0TyU2ILBproSMyyUVadsOXmUv=\&uniplatform=NZKPT&language=CHS (accessed on 16 May 2024). [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.