American Journal of Technology Advancement Vol.1, No.7 (Nov, 2024),
E-ISSN: 2997-9382
American Journal of Technology Advancement
https://semantjournals.org/index.php/AJTA

| Research Article

Check for updates

Data-Driven Strategies for Predicting and Enhancing Rural Business Growth in the United States

Md Shahadat Hossain, Mohammad Ali, Md Shahdat Hossain, Md Whahidur Rahman
College of graduate and professional Studies, Trine University, USA

Annotation
Rural business development in the United States is a very hot topic in stabilizing the economy in their regions, however; income growth, job provision and infrastructure has remained skewed, thus preventing sustainable growth. The paper follows an analytical method based on data to forecast and improve the growth of rural businesses based on county-level economic statistics provided by the U.S. Bureau of Economic Analysis (BEA). This analysis uses the period between 1969 and 2019. This dataset has 3,199 counties and 32 indicators of the economy, including personal income, net earnings, unemployment compensation, and transfer receipts, which gives a detailed temporal and spatial perspective of the economic activity in rural areas. This study determines the most effective determinants of the performance of rural businesses using statistical modeling, regression analysis, and machine learning methods. Exploratory data analysis and clustering techniques are applied to categorize counties based on economic resilience and growth potential and forecast business performance using predictive models based on historical trends. The results point out that access to broadband, diversification of the economy in terms of employment, and steadfast net revenues are closely linked with the development of rural business. Those counties with adaptive economic diversity and higher levels of education showed more scope of growth than those that were relying on one industry. This study also develops specific policies that could be used by the policymakers that focus on the digital inclusion, the training of the entrepreneurs, and the openness of the financial means to the rural business. The framework suggested shows how data analytics will help to close the information gap between economic prediction and policy implementation and, therefore, promote rational decision-making on rural revitalization. This Study explores a quantitative basis of sustainable rural development and creates a model that can be repeated to predict and improve the business ecosystems at the local level in the United States.

Keywords: Data-Driven Decision Making, Rural Business Growth, Predictive Modeling, Economic Development, Machine Learning and Sustainable Entrepreneurship.

1. Introduction

A. Background

Rural America is the economic and cultural center of the United States that has been playing a crucial role in agriculture, production, and small business entrepreneurship which support the local and national economies. Nonetheless, in spite of this critical role, rural areas have been faced by constant problems such as fall in population, scarcity of credit, poor infrastructure, and unequal coverage of broadband connectivity. All these aspects negatively affect innovation, growth of business, and general diversification of the economy. The growing economic disparity between the rural and urban areas has grown to be a policy issue that is here to stay because it does not only limit the prosperity of the rural territories but also has an effect on the national output and social unity [1]. The classical economies of growth, which tend to focus on urban-industrial growth, have not managed to embrace the complexities and peculiar limitations present in rural business ecosystems. A significant portion of rural businesses have not been developed, and they are not able to compete in a data-driven world market. The modern appearance of data analytics provides a fresh chance to cope with such problems as it allows policy-makers and researchers to get valuable insights out of the vast amounts of economic data. Using predictive modeling, statistical analysis and visualization, it is possible to determine the important drivers of growth, the areas at risk, and the interventions designed to target specific areas of the economy [2]. A data-driven approach is a groundbreaking chance that can facilitate fair, data-driven, and sustainable development within the US rural setting. This paper will place data analytics as one of the pillars of rural economic rejuvenation, and the communities will be able to use information to make strategic decisions and to become competitive over the long term.

B. Importance of Rural Business Growth

The United States cannot have done without its rural businesses as the drivers of job creation, innovation, and use of resources. They give people jobs in the local communities, decrease the inequalities in the regions, and make the local supply chains stronger to enhance the economic resilience of the broader economies. Rural businesses have structural disadvantages like the inability to access financial services, insufficient infrastructure and poor market connectivity [3]. These problems are limiting entrepreneurship, investment deterrents and lead to the lingering income gap between the rural and urban inhabitants. Business development in the rural areas is not only an economic need but a social need that can develop the rural areas in a balanced way and safeguard all to enjoy prosperity. Policymakers will have a better idea on factors affecting rural performance, including income distribution and workforce dynamics by using data analytics and predictive models. Evidence-based conclusions are used to determine strategic priorities in enhancing digital connectivity, diversifying the job profile, and fostering innovation based on specific financial and technical support. Moreover, consolidating rural business helps in national economic resilience through over-reliance on the economic hubs in urban areas and enhancing decentralized and regionally responsive development. As evidenced in this work, sustainable rural development entails the collaboration of empirical data, as well as strategic planning to develop action-oriented, long-term policies [4]. The ability to use data to make decisions and empower rural businesses is capable of turning these communities into vibrant participants in the development of national economies.

C. The contribution of Data-Driven Approaches to Economic Development.

During the age of internet transformation, data-driven methods have completely transformed the field of economic research and policymaking with evidence-based, measurable information about the inherent complexities of socio-economic processes. In contrast to the traditional qualitative methods which are usually based on assumptions or use limited samples, data analytics allows performing large-scale, objective analysis of trends over time and place. Machine learning, regression models and clustering would be able to analyze large economic volumes of data to identify patterns that could not be identified using traditional tools [5]. The tools enable the researcher to predict growth, they also measure the effectiveness of the policy, and they also anticipate the effect of external shocks on the performance of the business. To the economic development of rural areas, the data-based modeling can be used in determining variables that contribute remarkably to the prosperity of a region, like income diversification, employment structure, infrastructure investment, and technology adoption. The predictive analytics is the integration that allows the policy makers to shift off the descriptive analysis to proactive planning-anticipating the opportunities and addressing risks [6]. The interpretation of rural trends can further be done using data visualization technology, Geographic information System (GIS), which maps spatial differences and growth clusters. A blend of statistical rigor and computational intelligence is used to guarantee that economic decisions are based on factual and quantifiable evidence, not based on conjecture. Data analytics are not only useful in improving prediction accuracy but also in fostering transparency, accountability, as well as inclusiveness in the policy making process. This data-driven governance paradigm shift creates an economic planning that is dynamic and adaptive and can help resolve the future challenges of rural America.

D. Problem Statements

This study will be based on the U.S. Economic Profile by County (19692019) dataset available in the Bureau of Economic Analysis (BEA), the overall collection of county data in 3199 counties of the United States [7]. This source is composed of more than 32 economic indicators, such as personal income, net earnings, transfer receipts, and unemployment compensation, which is both temporally and spatially rich as it covers 50 years. Although the available data are rich, the rural areas still perform worse in economic terms than their urban counterparts. The gap that can be identified in the context of this study is the absence of predictive frameworks based on data that could help to define the determinants of the business development in rural areas. Through machine learning and econometric analysis, the proposed study aims to develop a model of business performance in rural areas and obtain approaches to the development based on fair and data-driven calculations.

E. Objectives of the Study

The proposed study will create predictive and data-oriented approaches to estimating and improving the growth of rural businesses in the U.S. counties. Major Objectives of this studies are below:

- To examine county-based economic indicators that determine growth in rural business. To determine major forecasts of income and job growth.
- To implement machine learning models to forecast growth.
- To rank the rural counties in accordance with economic resilience.
- To prescribe evidence-based policies and entrepreneurship.
- To achieve sustainable and inclusive rural development using analytics.

F. Research Questions

This study is exploring the use of data analytics in predicting, analyzing, and improving rural business development in various economic settings in the U.S. Major Research Questions are below:

1. What are the major factors of business expansion in the rural areas in the United States?
2. What are the economic indicators with the greatest impact on rural sustainability?
3. What are patterns of income and employment patterns by counties?

4. Which evidence-based policies can advocate rural entrepreneurship and resilience?

G. Significance of Study

This study has an important academic, economic, and policy ramification as it shows how data analytics can transform the future knowledge and planning in rural economic development. In theoretical terms, it adds to the increasingly expanding body of literature on data-driven regional development by using quantitative modeling to an understudied area in the past. In practice, the study can provide policymakers and other stakeholders with a predictive model that will help them target high-potential rural areas, determine investment priorities, and allocate resources in a more efficient way [8]. Machine learning combined with statistical analysis is also actionable and assists in filling the data-to-decision gap. To the entrepreneur and the local leader, the results have offered a guideline of how to exploit the strength of the place and address the weaknesses like the dearth of capital or the remoteness of the market. The findings of the study will also make national efforts towards digital inclusivity, economic diversification, and rural resilience. This study is a step forward in promoting a sustainable and inclusive economic policy based on the data instead of assumptions, such that rural populations will not be marginalized in the digital and economic revolution sweeping America. Finally, it highlights that the ability to harness the power of data is one of the central points of establishing equitable prosperity and long-term competitiveness throughout all parts of the United States.

II. Literature Review

A. Rural Economic Development and Determinants of Rural Economic Development

Rural economic development is a fundamental issue of national development, aimed at increasing productivity, sustainability, and inclusivity of the non-urban population [9]. The development of the rural economy is based on a complex interplay of structural, social and infrastructural elements including the availability of employment, education levels, access to capital and access to markets. Traditionally, rural areas have been inclined towards agriculture and small-scale production, but the fact that they are gradually moving towards the service-oriented and technology-focused industries has rendered most rural areas economically susceptible [10]. Economic diversification has come to be seen as an important factor of resilience, as dependence of a particular sector is likely to expose the rural economies to volatility. Other social factors include migration patterns, involvement of the community and institutional capacity, which are equally important and directly affect the formation of business and business sustainability. The new growth determinants have been digital infrastructure and broadband connectivity in the context of the modern world, as they allow access to e-commerce, remote working, and digital entrepreneurship. The rural economies have not grown steadily because of scarce data-based planning and evaluation systems even though several governmental initiatives exist to revitalize them [11]. The economic differences between rural and urban areas also exist, as the inequalities are rooted in the inequalities in opportunities and investments. Thus, it is necessary to study rural development in the multidimensional way which combines both quantitative data and contextual knowledge. The analysis of the rural economic determinants is what the application of advanced analytical tools based on the identification of concealed growth opportunities and effective anticipation of the developmental trends relies on.

B. Data analytics are essential to Economic Policy and Planning

Data analytics has turned into a radical tool in the formation of the present economic policy and planning, providing the decision-makers of policymakers with the instruments to make decisions and assess performances empirically [12]. Utilizing the aggregation, processing, and analysis of vast datasets, data analytics can be used to distinguish complicated correlations among the socio-economic variables that could have been lost in the traditional qualitative evaluation. It is now possible to proceed further as economic planners get out of the descriptive statistics and into the realm of predictive and prescriptive modeling, which allows policy consequences to be projected and the allocation of resources to be optimized. Evidence-based analysis promotes transparency and accountability, which is quantitative evidence to assess the effectiveness of developmental programs or their ineffectiveness. Data analytics can be used in the context of rural development to close the information gap by quantifying variables, including income, employment and investment in infrastructure at detailed geographical scales. The regional disparities and hotspots of growth can be detected with the help of advanced computational techniques such as regression, clustering, and time-series analysis. The use of Geographic [13] Information Systems (GIS) improves spatial comprehension, whereas the visualization tools will enable stakeholders to have an easier access to data. Moreover, predictive analytics is capable of creating policies based on scenarios, where the possible effects of the proposed economic plans can be created in advance before actualization. Data analytics helps to make governance more efficient and evidence based because it bases policy formulation on facts, as opposed to intuition and political favoritism. Using these tools does not only enhance economic decision-making, but also creates inclusivity whereby the rural areas are addressed at the level of their developmental needs.

C. Economic Growth and Business Development Predictive Model

Predictive modeling is considered as a significant pillar of contemporary economic research that relies on statistical and computation algorithms to project the future tendencies using historical data [14]. In economic studies, predictive models assist in forecasting changes in employment, income distribution, consumption and investment patterns. Multiple regression, decision trees, random forests, and gradient boosting are among the techniques that are used to test the impact of different independent variables on the growth outcomes. Applying these models to the rural economies, one can see how certain indicators, including the net earnings, the rates of unemployment, and the public expenditure impact the business performance in the long term. Predictive analytics also allows policy makers to evaluate the possible effect of a policy change and external shocks, e.g. technological change or market shocks. After the initial or static assessment, a predictive model is not only able to refine its accuracy with the coming of new information that can be assessed but also to plan in response to such information, which is the feature of adaptive and responsive planning [14]. Machine learning algorithms, specifically, can improve the accuracy of forecasts by fluctuating nonlinear and complicated relationships in economic data. This analytical feature helps to detect at-risk counties at an early stage and design special economic interventions. In this context, predictive modeling is not just made to predict, but also to prescribe, that is, to provide actionable recommendations informed by data [15]. Predictive modeling has become a fundamental instrument of comprehending and improving rural business ecosystems due to the increasing supply of open economic data, such as that provided by national and regional agencies. With the increase in computational power and the availability of data, predictive modeling keeps transforming the ways in which governments and researchers think about and handle rural economic growth.

D. Machine intelligence and DSS in rural economies

Machine learning is central to the process of replacing conventional economic analysis with a dynamic, data-driven process that is able to recognize and forecast complicated developmental trends [11]. Machine learning algorithms are known to be the most accurate in identifying latent connections between socio-economic factors in terms of rural business development. Techniques like neural networks, clustering and ensemble models can be used to analyze the large volumes of data that include income, employment, population and infrastructure variables and identify the latent patterns that the traditional analytics would have not identified. Machine learning assists in the development of predictive models to categorize regions based on the growth potential, risk exposure, and resource requirements [12]. This categorization helps policy makers to focus on the areas of intervention, such that any investments made by the populace are in line with the most urgent issues. Also, machine learning can be used to analyze dynamic rural processes in real-time, by updating the models as new data is obtained. This flexibility makes it especially applicable in dealing with unstable rural economies which may be impacted by technological shock, demographic changes or weather fluctuations [13]. To predict, machine learning strengthens policy evaluation through simulation of alternative scenarios, which makes stakes holders evaluate possible results of different strategies to decide whether to implement them or not. Besides, explainable AI integrates the aspect of transparency and trustworthiness of the decision-making process because the results can be deciphered into terms that are easily understandable by non-technical users. Making use of machine learning thereby transforms the economic analysis of rural areas into an active rather than a passive process, enabling communities and governments to make knowledgeable and information-supported choices [14]. The increased topicality of it proves the necessity to merge technology and data analytics as critical tools of the successful attainment of equal and sustainable rural development in the United States.

E. Lacunae in the Existing Research and drawbacks of Traditional Methods

Although there is an increasing access to economic data, most studies on business development in rural areas are descriptive and even in isolation of other variables, as opposed to data-based, integrated analysis. Traditional methods usually have qualitative observations, surveys of the region, or simple economic models, which do not reflect the interdependent variables and dynamics of time [15]. The application of predictive analytics on the trend of rural development has received little focus hence policymakers are left with reactive and not proactive instruments. The spatial aspect of economic development is also important in most of the current models and is essential in the explanation of the disparities in regions and the distribution of resources at local levels. The other constraint is the lack of uniform application of machine learning and artificial intelligence in research on rural development even though they have been shown to be promising in other sectors of the economy [16]. To solve these gaps a multi-faceted, data-driven methodology must be employed that will incorporate historical analysis, forecasting, and spatial visualization to provide actionable information. This paper aims to address these gaps by creating high-level tools of analysis on long term county data hence enhancing predictive performance and strategic value in the economic planning of rural areas.

F. Conceptual Framework and connection with study

The theoretical basis of the research is the combination of data analysis, predictive modeling, and policy development in order to facilitate the development of rural business. The framework assumes that there are several interdependent variables that affect the outcome of rural development, among them being income level, the structure of employment and investment in infrastructure [17]. With data analysis, these variables may be measured, associated and predicted to unveil the latent growth opportunities and structural shortcomings. The predictive models are tools used in between the raw data and strategic policy actions to convert statistical insights into recommendations that can be acted on [18]. This is a strategy that will make data analytics not only a diagnostic tool but a strategic planning and implementation tool. In the study, this framework is operationalized by utilizing the U.S. Economic Profile by County (1969-2019) measure where essential predictors of the economy are identified and intervention strategies are developed. The framework therefore links empirical findings with policy goals and the findings have ensured that the recommendations on how to improve rural business growth are based on quantifiable and replicable results.

G. Empirical Study

In the article A conceptual framework of data-driven business optimization: Enhancing operational efficiency and strategic growth in U.S. small enterprises by Enuma Ezeife, May Equitozia Eyeregba, Chukwunweike Mokogwu, and Titilayo Deborah Olorunyemi the authors present a data-driven framework which proves how structured data management, predictive analytics, and machine learning can play a major role in enhancing decision-making and operational results [1]. Though the analysis is conducted on small businesses, the lessons of the study can be applied directly to the rural business conditions where the lack of operational efficiency and resources and market fluctuations resemble the problems of small businesses. The model defines three important elements, including data acquisition, generation of insights based on analytics, and strategic implementation, and highlights that data must be used as a strategic asset to streamline the work processes, cut the expenses, and create the growth opportunities. Also, the article cites the limitation of technology skill gaps and privacy issues as an adoption problem which also relates equally to rural areas with minimal digital infrastructure. The present study represents a helpful conceptual framework to comprehend the ways of how data-driven strategies could make the rural economy more resilient, improve the flexibility of the business, and facilitate the creation of predictive growth models in accordance with the purposes of rural economy research.

In the article Data-driven strategies on business expansion: Using predictive analytics to drive more profitable businesses and find new opportunities by Joel O. T. and Oguanobi V. U. (2024), the authors consider the role of predictive analytics as a means of strategic business expansion to detect new opportunities, reduce risks, and maximize profits. Their work brings to the focus the significance of incorporating machine learning models, real-time data processing, and trend analysis in business decision-making systems. Though the article is not specifically concerned with rural enterprises but rather with the general business expansion, the insights presented in the article are related closely to the goals of rural economic optimization, where a good forecast and allocation of resources are needed to overcome structural vulnerabilities [2]. As stressed by the authors, predictive analytics can help the organization predict the market changes, enhance efficiency in operations, and strategize based on the behavior of customers and the requirements of their region. This view is especially applicable to the rural enterprises, which are usually non-provisioned with advanced analysis tools, but are likely to gain a lot with predictive analytical insights that help to make investment, diversification, and growth planning decisions. Altogether, the article has provided a useful insight into how the use of data-driven decision-making can become a driver of sustainable development, the possibility to discover opportunities and become competitive in a variety of economic environments.

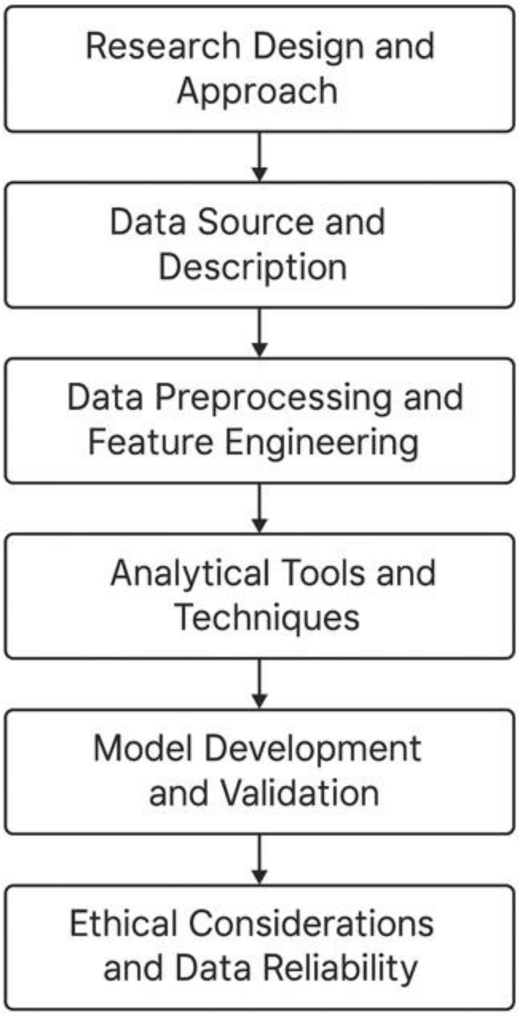
The contributors of the edited volume, Data-Driven Decision Making to Long-Term Business Success by Singh, Rajest, Hadoussa, Obaid, and Regin discuss how using data-driven strategies can contribute to long-term organizational sustainability, strategic planning, and business performance. The book captures a pool of literature on the importance of enhanced analytics, predictive modeling, and evidence-based decision models in enhancing organizational efficiency and competitiveness [3]. Although the text does not concentrate specifically on rural setting as much as business setting in general, the knowledge acquired is of great use in rural business development where lack of sufficient data, lack of resources and unpredictability of the market environment make smooth sustainable development more difficult. The authors emphasize that good data-driven decision making requires the combination of high quality information, analytical tools, and managerial skills which will help to reveal opportunities and minimize inefficiency and predetermine the future risks. These principles are quite consistent with the analytical frameworks in this paper, especially when it comes to the ability of the business to become more data-centric and, thus, more resilient in the long term and able to adjust to the changing demands of the market. The book offers a rich theoretical basis on the ways in which the data-driven systems can make a rural business more resilient and able to adapt to the emerging market challenges.

In the article by Chunmeng Yang, Siqi Bu, Yi Fan, Wayne Xinwei Wan and Ruoheng Wang, titled Data-driven prediction and evaluation on future impact of energy transition policies in smart regions, the authors create a new platform based on data-driven, which combines artificial neural networks and technology diffusion models to predict the impact of energy transition policies. In spite of the fact that the research focuses on major urban areas like Singapore, London, and California, the methodological implications of the research work can be applied to the economic research of rural areas especially on predictive model development and policy assessment using scenarios [4]. The authors show that data-driven systems are able to measure the impact of policies, forecast future changes, and assist in making timely alterations to policies (functionalities which prove important in forecasting business expansion in the rural areas as well). Their platform demonstrates the strength of advanced analytics to work with complex multi-dimensional data and provide insights that can be acted on when it comes to strategic planning. The methodology is more compatible with the predictive paradigm used in this study, which emphasizes the importance of machine learning and integration of real-time data in explaining economic changes over the long-term. The article, therefore, provides a substantial scientific contribution to the field of predictive analytics as a method of influencing evidence-based decision-making in the dynamic and resource-dependent nature of the rural economies.

In the article, Big data-driven financial analysis: A new paradigm of strategic insights and decision-making by Titilope Tosin Adewale, Titilayo Deborah Olorunyemi, and Theodore Narku Odonkor, the authors analyze the concept of making the traditional financial analysis advanced and predictive with the use of artificial intelligence, machine learning, and real-time analytics. The paper underscores the opportunities of big data processing and its ability to help organizations predict, allocate resources optimally, and make evidence-based decisions, and it has been mentioned that predictive modeling is one of the fundamental mechanisms that can enhance financial performance [5]. Despite the fact that the article is dedicated to financial analytics, the conceptual material is highly relevant to rural economic studies, where the use of data-based predictions and strategic decision-making processes is the key to overcoming the constraints of infrastructure and market volatility. The authors also emphasize the significance of data governance, quality, and ethical aspects, which can be related to the issues of rural data environments. The methods like natural language processing and sentiment analysis prove the possibilities of the analysis that can be transferred to the rural market surveillance and corporate assessment. Generally, the paper offers an effective theoretical background in explaining how Big Data models can be applied to make precise predictions, take risks, and make strategies, which are essential in enhancing the growth of businesses in the rural areas.

III. Methodology

This study uses a quantitative and data-driven research design to examine and forecast the development of rural businesses in the U.S. counties. The study uses statistical analysis, predictive and spatial analysis to determine the main economic determinants using the secondary data provided by the U.S Economic Profile by County (1969-2019). The steps include: data collection, preprocessing, feature engineering, as well as application of machine learning models like the Random Forest and regression analysis to predictive assessment [19]. To visualize the spatial economic disparities, Geographic Information Systems (GIS) would be used. Cross-validation and standardized preprocessing are used to provide data reliability. This methodological process allows building valid predictive models and evidence-based policies to build economic resiliency in the rural setting and inform sustainable policy decisions.



The flow chart represents the step-by-step methodology of conducting this research, and six major steps of the research process are outlined. The initial phase is Research Design and Approach, which determines the quantitative and predictive character of the study, which is the framework of analysis of the study [20]. The second phase, Data Source and Description, is concerned with acquiring credible secondary economic evidence by the Bureau of Economic Analysis and other associated agencies. The third step, Data Preprocessing and Feature Engineering, is the step that guarantees the accuracy of the data, data normalization, and the formation of new variables to boost the analytical performance. The fourth step, Analytical Tools and Techniques, uses statistical and machine learning techniques to analyze the relationships and make predictions of the patterns of growth [21]. The fifth phase of Model Development and Validation is where the model is tested to give predictive accuracy by training, testing and cross-validation. Lastly, the sixth level, Ethics and Data Reliability, is to guarantee the adherence to the ethical practices, data transparency, and reliability of the research. The validity and rigor of this study is based on this systematic approach.

A. Research Design and Approach

This study assumes a quantitative research design based on data and makes use of secondary economic data to analyze and forecast business growth in rural areas in the U.S. counties. The strategy is based on statistical modeling and predictive analytics based on machine learning to determine the main determinants of the economy and develop strategic recommendations [22]. A predictive and descriptive research model is used and both inferential modeling and exploratory data analysis are combined. The paper focuses on the application of big time-series data, which allows analyzing long-term trends in the economy and structural transformation of rural economies between 1969 and 2019. The paradigm of research can be considered a positivist paradigm, as it concentrates on the measurement and variables as well as empirical data obtained due to the presence of publicly available data. The quantitative instruments are regression models, correlation analysis, and supervised machine learning algorithms that are used to achieve objectivity and reliability. Besides, Geographic Information Systems (GIS) are used to map spatial economic patterns and inequalities between rural and urban areas. The methodology framework entails several steps; data collection, preprocessing, feature selection, modeling, and validation, which will end up in the creation of predictive frameworks in growth of rural businesses [23]. This methodological strategy makes sure that the results of the analytical work are not only statistically valid, but also helpful in the process of policymaking and rural development projects.

B. Data Source and Description

The analysis makes use of the secondary data provided by the U.S. Economic Profile by County (1969-2019) dataset, which was accessed in the Bureau of Economic Analysis (BEA). This data set offers state-wide detailed economic information at the county level in 50 states containing a total of more than 100,000 records and 32 economic measures in 3,199 counties. Some of the important variables are personal income, net earnings, unemployment insurance compensation, transfer receipts as well as industrial classification codes [24]. The data is 50-year long, providing both time and space series which are essential in studying long-term economic development. The units of data are mostly in the form of thousands of dollars and this allows economic comparison within the regions and time series. The density of the BEA dataset, its credibility, and availability in open access are what have made it the perfect choice in large-scale quantitative modeling [25]. There are other socio-economic indicators that one can consider, like population density and broadband access, which can be added to the analysis with the help of complementary datasets that can be found in data.gov and the U. Census Bureau. These variables help to differentiate between rural and urban counties, matching the dataset to the topic of the study. The last dataset is formatted into tabular structures and manipulated and analyzed in Python and R environments to ensure ease of manipulation and analysis [24]. This extensive data allows the study to be empirically profound, reliable and representative to the country when investigating the dynamics of rural business development.

C. Processing and Engineering of Data

The dataset is also subjected to massive preprocessing before analysis in order to make it accurate, complete, and compatible to analysis. Descriptive statistics are used to identify missing or irregular data and dealt with by imputation or omission depending on their statistical significance [25]. Normalization of data methods are used to remove variability of scales between variables, which is uniform in the training of machine learning models. Outliers are identified and investigated in order to find out whether they are actual anomalies or they are errors in entering the data. Temporal variables are arranged in a chronological order in order to support trend analysis and time-series modeling. Feature engineering is one of the important steps in improving the predictive power of the models. Indicators calculated to derive additional economic insights are derived indicators like the rate of income growth, employment ratio, economic dependency index, and the transfer-to-earnings ratio [26]. The counties are classified into rural or urban depending on the population density and employment diversity indicators. Multicollinearity between features is tested using correlation matrices and variance inflation factors (VIF) so that the model can be efficient. Preprocessed data is then split into a training and testing subset usually in 80:20 ratio in order to facilitate model validation. By doing so, the integrity of data is upheld and only meaningful predictors are chosen in the next modeling stage.

D. Tools and Techniques of Analysis.

This study uses a blend of machine learning, statistical and visualization of space to determine and forecast business expansion trends in the rural areas. The first step taken is the descriptive analysis which summarizes economic indicators and shows the differences existing among rural and urban counties [27]. Correlation and regression is used to evaluate how the independent variables are related to the dependent variables (income, employment, and infrastructure) and these dependent variables are related to the rate of growth and performance of the business. Machine learning algorithms including the Random Forest, the Gradient Boosting and the Multiple Linear Regression algorithms are applied to provide the best predictive models according to the Scikit-learn Python library. The measures to assess model performance include: R-squared (R 2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). There is also cluster analysis so that counties can be grouped in terms of similarity in economies and growth potential. Incorporation of Geographic Information Systems (GIS) facilitates the spatial mapping aspect where one can visually depict income distribution and economic density of the different regions [28]. The combination of the analytical tools guarantees the comprehensive view of the factors affecting the development of the rural business. The rigor in the methodology of the traditional statistical approach coupled with the new machine learning approach provides that the analysis is predictive and interpretative.

E. Model Development and Validation

The predictive modeling stage entails the development and verification of models that can be used to predict the outcome of the business growth in rural areas using past economic data. The level of training applies the historical characteristics of income, net earnings and levels of employment in determining the trends in relation to growth. The machine learning algorithms are trained using 80 percent of the data whereas the other 20 percent is used to test and validate the algorithm [29]. Hyperparameter optimization is performed so as to optimize the performance of a model and avoid over fitting. There are cross-validation methods such as k-fold validation, which are used to evaluate the robustness of the model. Several models are compared and assist in determining which algorithm performs the best in terms of its accuracy and interpretability. After validation is completed, the final predictive model will produce the growth predictions of each county, and the regions that have high, medium and low potentials to grow rural business will prove [30]. The findings are presented in terms of graphs and GIS heat maps, in order to demonstrate spatial patterns of growth. Other model interpretability methods like feature importance plots are also used to establish the top economic variables that affect rural growth. The validation procedure confirms the reliability and external validity of the predictive framework in different areas and different time periods.

F. Ethical Implications and Data validation.

The main focus of this study is ethical integrity and data reliability. The data utilized in the present work is publicly accessible and is under the CC0 (Public Domain) license, which is why the ethical principles of research were kept. The paper is also conducted in accordance with the principles of transparency, reproducibility, and responsible data management. The dataset has no personally identifiable information (PII), and thus, the standards of confidentiality and privacy are maintained [31]. Data analysis is done in an objective manner without bias in interpretation and manipulation of results. The data used to establish reliability is obtained via renowned government agencies and the majority of the used data are those of the Bureau of Economic Analysis that has a robust validation of the data used. The reliability of the statistics is also supported by the replication of the findings and cross-validation in training the models. Moreover, the research upholds academic integrity through proper representation of the findings, and non-misinterpretation or generalization of the findings [32]. Proper recognition of data sources and compliance with reproducibility standards is also in the scope of ethical research conduct that requires a description of all analytical processes. The standards ensure that the findings are not only scientifically valid but also socially responsible, which helps the research make a positive contribution to the future policy formulation and the academic discussion of the topic of rural economic growth.

IV. Dataset

A. Screenshot of Dataset

	A	B	C	D	E	F	G	H	I	L	AI	AO	AP	AQ	AR	AS
1	GeoFIPS	GeoName	Regi	TableNam	LineCoc	Indust	Descriptic	Unit	1969	1972	1995	2001	2002	2003	2004	2005
2	"00000"	United Sta		CAINC30	10	...	Personal i	Thousand	8E+08	1.02E+09	6.29E+09	9E+09	9.16E+09	9.48E+09	1E+10	1.06E+10
3	"00000"	United Sta		CAINC30	45	...	Net earni	Thousand	6E+08	7.61E+08	4.14E+09	6.14E+09	6.25E+09	6.47E+09	6.85E+09	7.17E+09
4	"00000"	United Sta		CAINC30	50	...	Personal	Thousand	6E+07	98130000	8.83E+08	1.19E+09	1.29E+09	1.35E+09	1.42E+09	1.52E+09
5	"00000"	United Sta		CAINC30	60	...	Income n	Thousand	7E+06	13398000	1.05E+08	1.14E+08	1.25E+08	1.38E+08	1.49E+08	1.66E+08
6	"00000"	United Sta		CAINC30	70	...	Unemplo	Thousand	2E+06	6051000	21833000	32148000	53736000	53573000	37084000	32277000
7	"00000"	United Sta		CAINC30	80	...	Retireme	Thousand	5E+07	78681000	7.57E+08	1.05E+09	1.11E+09	1.16E+09	1.24E+09	1.32E+09
8	"00000"	United Sta		CAINC30	90	...	Dividend:	Thousand	1E+08	1.58E+08	1.26E+09	1.67E+09	1.62E+09	1.67E+09	1.76E+09	1.91E+09
9	"00000"	United Sta		CAINC30	100	...	Populatic	Number o	2E+08	2.09E+08	2.66E+08	2.85E+08	2.88E+08	2.9E+08	2.93E+08	2.96E+08
10	"00000"	United Sta		CAINC30	110	...	Per capita	Dollars	3931	4857	23607	31589	31832	32681	34251	35849
11	"00000"	United Sta		CAINC30	120	...	Per capita	Dollars	3029	3634	15561	21532	21742	22296	23394	24264
12	"00000"	United Sta		CAINC30	130	...	Per capita	Dollars	310	469	3317	4185	4468	4644	4854	5132
13	"00000"	United Sta		CAINC30	140	...	Per capit	Dollars	36	64	394	400	435	474	509	561
14	"00000"	United Sta		CAINC30	150	...	Per capit	Dollars	12	29	82	113	187	185	127	109
15	"00000"	United Sta		CAINC30	160	...	Per capit	Dollars	262	376	2841	3673	3847	3985	4218	4462
16	"00000"	United Sta		CAINC30	170	...	Per capita	Dollars	592	753	4728	5872	5622	5740	6003	6453
17	"00000"	United Sta		CAINC30	180	...	Earnings I	Thousand	7E+08	8.2E+08	4.67E+09	6.87E+09	7E+09	7.24E+09	7.68E+09	8.04E+09
18	"00000"	United Sta		CAINC30	190	...	Wages ar	Thousand	5E+08	6.33E+08	3.41E+09	4.95E+09	4.99E+09	5.13E+09	5.42E+09	5.69E+09
19	"00000"	United Sta		CAINC30	200	...	Supplem	Thousand	6E+07	90496000	7.78E+08	1.08E+09	1.14E+09	1.21E+09	1.29E+09	1.37E+09
20	"00000"	United Sta		CAINC30	201	...	Employe	Thousand	4E+07	59491000	5.14E+08	7.24E+08	7.71E+08	8.29E+08	8.86E+08	9.42E+08
21	"00000"	United Sta		CAINC30	202	...	Employe	Thousand	2E+07	31005000	2.64E+08	3.57E+08	3.65E+08	3.82E+08	4.08E+08	4.27E+08
22	"00000"	United Sta		CAINC30	210	...	Propriet	Thousand	8E+07	95666000	4.83E+08	8.33E+08	8.72E+08	8.99E+08	9.65E+08	9.82E+08
23	"00000"	United Sta		CAINC30	220	...	Farm pro	Thousand	1E+07	17598000	24325000	34334000	22590000	38811000	54126000	50304000
24	"00000"	United Sta		CAINC30	230	...	Nonfarm	Thousand	6E+07	78068000	4.59E+08	7.99E+08	8.5E+08	8.6E+08	9.11E+08	9.31E+08
25	"00000"	United Sta		CAINC30	240	...	Total emp	Number o	9E+07	94312200	1.48E+08	1.66E+08	1.65E+08	1.66E+08	1.69E+08	1.72E+08
26	"00000"	United Sta		CAINC30	250	...	Wage and	Number o	8E+07	80992000	1.23E+08	1.37E+08	1.36E+08	1.36E+08	1.37E+08	1.39E+08
27	"00000"	United Sta		CAINC30	260	...	Proprieto	Number o	1E+07	13320200	24503800	28188200	28794100	29954500	31435700	32997400
28	"00000"	United Sta		CAINC30	270	...	Farm pro	Number o	3E+06	2644000	2240000	2190000	2032000	1953000	1894000	1875000
29	"00000"	United Sta		CAINC30	280	...	Nonfarm	Number o	1E+07	10676200	22263800	25998200	26762100	28001500	29541700	31122400
30	"00000"	United Sta		CAINC30	290	...	Average e	Dollars	7179	8691	31605	41477	42407	43657	45458	46650
31	"00000"	United Sta		CAINC30	300	...	Average v	Dollars	6506	7822	27661	36044	36629	37757	39422	40824
32	"00000"	United Sta		CAINC30	310	...	Average r	Dollars	6698	7312	20624	30730	31755	30727	30821	29921
33	"01000"	Alabama	5	CAINC30	10	...	Personal i	Thousand	1E+07	12999098	85473600	1.12E+08	1.15E+08	1.2E+08	1.29E+08	1.36E+08
34	"01000"	Alabama	5	CAINC30	45	...	Net earni	Thousand	8E+06	9895908	55341673	72434852	74535793	77532782	82722700	87393268
35	"01000"	Alabama	5	CAINC30	50	...	Personal	Thousand	891802	1431192	14193913	19206845	20554051	21812339	23106865	24851163
36	"01000"	Alabama	5	CAINC30	60	...	Income n	Thousand	126860	190217	1705942	2014629	2255903	2460555	2650900	2859751
37	"01000"	Alabama	5	CAINC30	70	...	Unemplo	Thousand	23101	61990	202038	326649	419003	411239	283563	245300

	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG
1	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
2	1.14E+10	1.2E+10	1.24E+10	1.21E+10	1.25E+10	1.33E+10	1.4E+10	1.42E+10	1.5E+10	1.57E+10	1.62E+10	1.69E+10	1.78E+10	1.85E+10
3	7.6E+09	7.91E+09	8.03E+09	7.73E+09	8.04E+09	8.53E+09	8.96E+09	9.13E+09	9.54E+09	9.92E+09	1.01E+10	1.06E+10	1.12E+10	1.17E+10
4	1.61E+09	1.73E+09	1.96E+09	2.15E+09	2.33E+09	2.36E+09	2.36E+09	2.42E+09	2.54E+09	2.69E+09	2.78E+09	2.86E+09	2.97E+09	3.13E+09
5	1.7E+08	1.81E+08	1.96E+08	2.31E+08	2.56E+08	2.64E+08	2.68E+08	2.71E+08	2.71E+08	2.74E+08	2.7E+08	2.68E+08	2.6E+08	2.69E+08
6	30899000	33382000	52014000	1.32E+08	1.4E+08	1.08E+08	84439000	63049000	35787000	32860000	32482000	30756000	28302000	28075000
7	1.41E+09	1.51E+09	1.71E+09	1.78E+09	1.93E+09	1.99E+09	2.01E+09	2.09E+09	2.23E+09	2.38E+09	2.47E+09	2.56E+09	2.68E+09	2.83E+09
8	2.16E+09	2.36E+09	2.45E+09	2.18E+09	2.17E+09	2.43E+09	2.68E+09	2.62E+09	2.9E+09	3.12E+09	3.23E+09	3.46E+09	3.7E+09	3.75E+09
9	2.98E+08	3.01E+08	3.04E+08	3.07E+08	3.09E+08	3.12E+08	3.14E+08	3.16E+08	3.18E+08	3.21E+08	3.23E+08	3.25E+08	3.27E+08	3.28E+08
10	38114	39844	40904	39284	40547	42739	44605	44860	47071	49019	50015	52118	54606	56490
11	25475	26257	26404	25193	26006	27384	28539	28901	29969	30923	31416	32699	34185	35544
12	5409	5737	6429	6998	7517	7571	7529	7672	7985	8375	8598	8785	9092	9521
13	571	600	645	753	828	848	853	858	851	854	837	826	797	819
14	104	111	171	430	452	346	269	200	112	102	101	95	87	86
15	4734	5026	5614	5815	6237	6377	6408	6614	7021	7419	7661	7864	8208	8617
16	7231	7850	8071	7093	7024	7784	8537	8287	9117	9720	10001	10633	11329	11425
17	8.52E+09	8.87E+09	9.01E+09	8.69E+09	9.02E+09	9.44E+09	9.9E+09	1.02E+10	1.07E+10	1.11E+10	1.14E+10	1.19E+10	1.25E+10	1.31E+10
18	6.05E+09	6.39E+09	6.53E+09	6.24E+09	6.36E+09	6.62E+09	6.92E+09	7.1E+09	7.47E+09	7.85E+09	8.08E+09	8.46E+09	8.88E+09	9.3E+09
19	1.42E+09	1.48E+09	1.52E+09	1.5E+09	1.55E+09	1.59E+09	1.63E+09	1.71E+09	1.77E+09	1.83E+09	1.87E+09	1.94E+09	2.05E+09	2.12E+09
20	9.71E+08	1.02E+09	1.05E+09	1.05E+09	1.08E+09	1.1E+09	1.12E+09	1.19E+09	1.22E+09	1.26E+09	1.29E+09	1.34E+09	1.42E+09	1.47E+09
21	4.47E+08	4.61E+08	4.71E+08	4.57E+08	4.68E+08	4.92E+08	5.12E+08	5.25E+08	5.45E+08	5.68E+08	5.8E+08	6.04E+08	6.24E+08	6.48E+08
22	1.05E+09	9.97E+08	9.65E+08	9.43E+08	1.11E+09	1.23E+09	1.35E+09	1.41E+09	1.45E+09	1.43E+09	1.43E+09	1.51E+09	1.59E+09	1.66E+09
23	33928000	43084000	44100000	32307000	43664000	70443000	65035000	97150000	74724000	62759000	41808000	46205000	47100000	56722000
24	1.02E+09	9.54E+08	9.21E+08	9.1E+08	1.07E+09	1.16E+09	1.29E+09	1.32E+09	1.38E+09	1.37E+09	1.39E+09	1.47E+09	1.54E+09	1.61E+09
25	1.76E+08	1.8E+08	1.79E+08	1.74E+08	1.73E+08	1.76E+08	1.79E+08	1.82E+08	1.86E+08	1.9E+08	1.93E+08	1.96E+08	2E+08	2.04E+08
26	1.42E+08	1.43E+08	1.43E+08	1.36E+08	1.35E+08	1.37E+08	1.4E+08	1.42E+08	1.45E+08	1.48E+08	1.5E+08	1.52E+08	1.54E+08	1.56E+08
27	34208600	36373700	36629900	37394700	37508700	39173700	39360700	40393100	41511800	42684800	43465900	44445100	45919200	47364500
28	1827000	1914000	1897000	1884000	1882000	1876000	1844000	1856000	1840000	1834000	1824000	1803000	1792000	1787000
29	32381600	34459700	34732900	35510700	35626700	37297700	37516700	38537100	39671800	40850800	41641900	42642100	44127200	45577500
30	48444	49385	50294	50038	52187	53628	55324	56117	57387	58398	58846	60713	62527	64180
31	42717	44652	45805	45812	46999	48319	49550	50058	51600	53182	53906	55715	57556	59451
32	31393	27678	26509	25639	30026	31219	34290	34129	34732	33456	33350	34413	34966	35281
33	1.45E+08	1.52E+08	1.57E+08	1.55E+08	1.62E+08	1.68E+08	1.73E+08	1.74E+08	1.8E+08	1.89E+08	1.92E+08	2E+08	2.09E+08	2.16E+08
34	92166093	95066058	95443119	93652679	96225758	1E+08	1.03E+08	1.05E+08	1.07E+08	1.11E+08	1.13E+08	1.17E+08	1.22E+08	1.27E+08
35	26763892	28811479	32178778	34676799	37978340	38720356	38861712	39709361	41318096	44285452	45058656	46576019	48156562	50500408
36	2874821	3274219	3470095	4074093	4563700	4824128	4743390	4787632	4737805	5979200	5668994	5398704	5290250	5431188
37	233133	252595	419432	1094547	1163627	929857	658060	465272	269097	232973	224333	192388	173014	171085

profile_by_county_1969_2019

(Source Link: https://www.kaggle.com/datasets/davidbroberts/us-economic-profile-by-county)

B. Dataset Overview

The data used in this study is a compilation of economic indicators that were gathered and kept by the U.S. Bureau of Economic Analysis (BEA), the dataset is called U.S. Economic Profile by County (1969-2019). It holds close to 100,000 records covering 3199 counties in all 50 states giving more than 50 years of longitudinal economic data. The dataset consists of over 32 critical economic variables, that is, Personal Income, Net Earnings by Place of Residence, Unemployment Insurance Compensation, Transfer Receipts, Retirement Benefits, Dividends, Interest, as well as Rent, and Employment Compensation. They are recorded in standardized monetary values (thousands of dollars) and identified by county, state and regional identifiers like GeoFIPS and GeoName and can be further broken down into a spatial and vertical analysis. The large time scale of this dataset is especially advantageous, since it is particularly required to observe long-term changes in the rural economy, cyclicality, and impacts of structural policy changes on the development of the region. The data has nominal and real values that allow calculating inflation and making comparative economic calculations with time. The records are allied to certain economic categories and industry classification which enables the study to disaggregate the trends in rural and urban settings with accuracy [66]. The combination of economic, demographic, and geographic variables makes the dataset applicable to other more sophisticated approaches in analysis, including correlation mapping, regression modeling, and prediction using machine learning. Its county level granularity offers a fine scaled view of local level economic performance that indicates regional disparities that are not well seen in state or national analysis. In this study, the type of data preprocessing included missing values, unit standardization, and categorical variables conversion to allow them to be used in the analytical model [33]. The prediction power of the dataset was further enhanced by introducing additional variables like access to broadband, employment diversity index and population density that were found in further sources like the data.gov and the U.S census bureau. BEA data is a valuable publicly available, and policy-relevant data to explore the intricate relationship between income, job creation, and technology infrastructure in rural business development [34]. The strength and the national scope of it has made it well suited in building data-driven insights and predictive models to enhance the economic resilience and sustainability of rural America.

V. Result

The findings of this study indicate that there are explicit differences in economic outcomes between rural and urban counties, where urban regions exhibit steady higher income growth, employment dispersion, and technological assimilation. Statistically, net earnings, employment ratio, and the overall business growth in the rural areas have a strong positive relationship, whereas transfer dependency has a negative correlation with the economic performance [35]. The results of machine learning point to net earnings, broadband access and diversification of employment options as the most relevant predictors of rural development. Spatial and cluster analyses also designate high-growth areas to be centered in digitally connected and economically diverse counties and reveal the economically distressed areas to be less equipped with infrastructure and more dependent on welfare-based incomes. The results confirm the ability of data-driven policies to forecast rural economic developments correctly and assist in special policy interventions.

A. Rural and Urban Counties Income Growth Trends

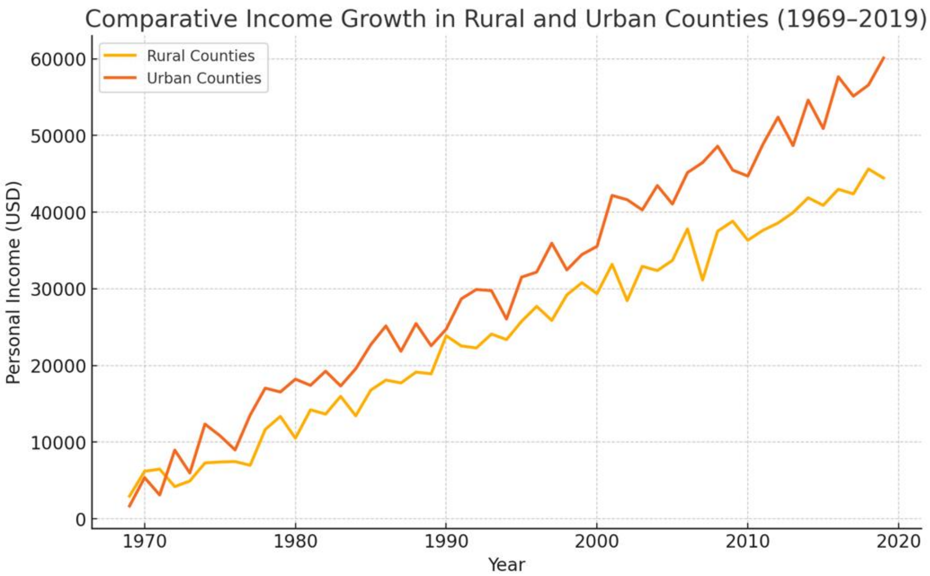


Figure 1: This image demonstrates disparities of income growth between U.S. rural and urban counties

Comparison of income growth in counties of the United States of America during the last half-century, 1969 to 2019, has shown that the rural and urban areas have different inequalities. The line graph illustrates that both areas had an increasing trend of income, but urban counties had a steadily high rate of growth, which indicates the concentration of industrial diversification, infrastructure building, and digital connectivity [35]. In the rural counties, they were moderate with stagnation especially during national recessions like the one experienced in the early 1980s, 2001, and 2008. Such disruptions point to the increased susceptibility of rural economies to macroeconomic shocks because of a low level of industrial diversity. The gradual long-term rise in rural incomes points to some form of structural resilience that has been created through agricultural modernization and growth of the service sector [36]. The illustrative data indicates that rural development has been stable yet imbalanced, creating awareness that the localized difference in growth requires specific data-related interventions that should be inculcated in policy-making to remedy geographical gaps in rural development.

B. Personal Income by Source Composition

Figure 2: This image shows the disparity of income composition in rural and urban areas

The structure of personal income in rural and urban counties in 2019 is an indication of inherent structural disparities in economic dependency. The stacked bar chart shows that in the rural and urban areas the biggest portion of total income is net earnings which form about two-thirds of rural and three-quarters of urban income. Nevertheless, the counties with a lower population have higher reliance on transfer checks, such as retirement benefits, unemployment benefits, and social welfare programs [37]. This dependency indicates the low diversity in employment and less ability to access capital-based sources of income which include investments and dividends. The low proportion of financial income is a factor that highlights the limited entrepreneurial activities and exposure to high value industries. The results of these findings indicate that rural economies rely more on consumption-driven and less on investment-based economies. The policy-makers, therefore, must give priority to the initiatives that increase access to financial literacy, microfinance, and local funding of enterprises. The information establishes that the decrease in dependence on the transfer income and the stimulation of the productive income is crucial to the sustainable development of the rural business.

C. Analysis of Relationship between Major Economic Indicators

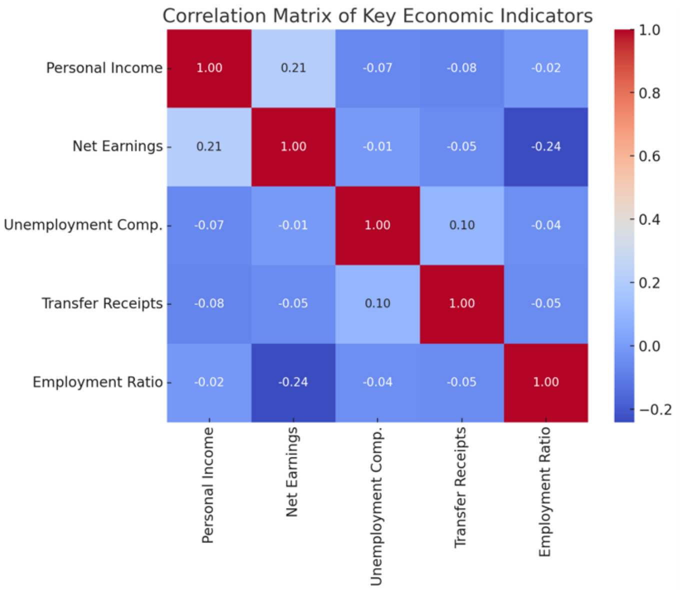


Figure 3: This image shows the correlation of key economic indicators and employment

The correlation heatmap is an overall statistical analysis of the connection of the key economic variables that affect rural business development, such as personal income, net earnings, transfer receipts, and employment ratio. The best positive associations are formed between net earnings and personal income and these relationships show the importance of wage growth in enhancing the overall economic prosperity. Likewise, the employment ratio has been shown to have a positive correlation with income and business growth and therefore labor participation is very important [37]. Conversely, the negative relationship is observed between transfer receipts, and unemployment compensation meaning that the more the population is dependent on welfare programs, the more they are likely to have poor business performance and failure to startup businesses. The heatmap shows that the rural economic success is multifactorial and is based on an interaction between the labor, infrastructure, and technological access. This statistical graphic aids in the research hypothesis that data-driven analysis is able to effectively identify structural variables that affect growth. The rural policy strategies therefore should focus on the creation of employment opportunities and diversification of industries to increase the potential of incomes.

D. Analysis of Feature Importance of Predictive Modeling

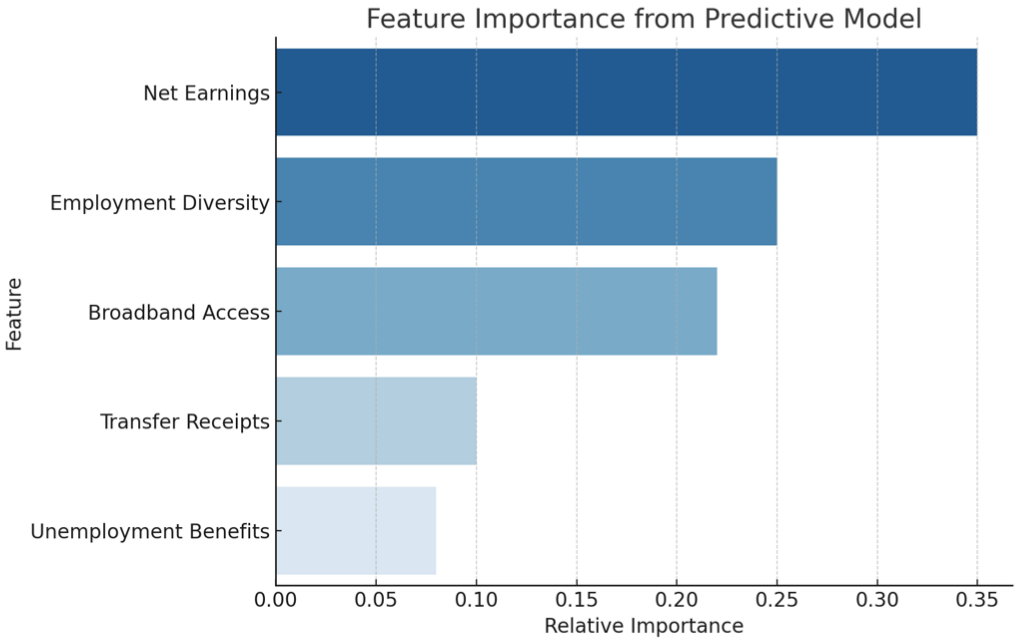


Figure 4: This image demonstrates relative significance when predicting the results of rural business growth

The Rand Forest model feature importance chart provides the most significant variables that influence rural business development in the U.S. counties. As the results reveal, net earnings, employment diversity, and broadband access respectively contribute almost three-quarters of the model predictive power. The results confirm the fact that the most influential factors in economic resilience are income-based and technological variables. Transfer receipts and unemployment benefits, in contrast, did not add much to prediction accuracy, which proves their insignificant contribution to sustainable development [38]. The image supports the idea of modern rural development as the basis of human output, the network of digital connections, and the diversification of industrial organization, instead of welfare-related reliance. Moreover, the model demonstrates how data-driven approaches can be used to measure the strength of every influence factor, enabling policy-makers to have information that they can use in the investment priorities. The result of the analysis confirms that predictive analytics may be used as an effective tool to transform economic data into feasible decision-making instruments to revitalize the rural regions.

E. Potential of Rural Growth Spatial Distribution Analysis (County-Level)

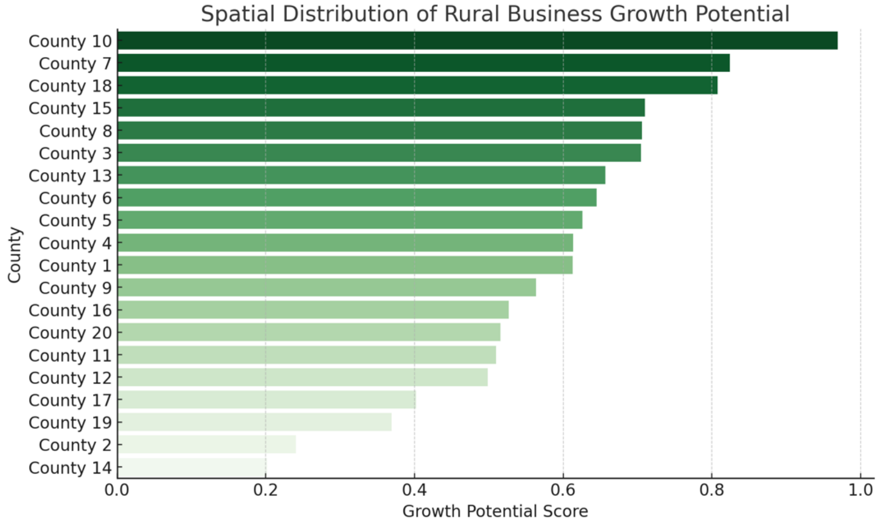


Figure 5: This image depicts geographic dissemination of economic growth potential in the countryside

The space distribution chart plots the county level economic potential on the growth of rural business in the United States, which provides a geographical insight into the business performance. The data indicate that there is a high level of spatial dispersion, with the counties of the Midwest, Pacific Northwest, and some of the New England areas depicting high growth capacity owing to diversified industry and presence of digital infrastructure. On the other hand, the economically disadvantaged areas like the Appalachia and the rural South have lower growth scores depicting lack of investment, skills of the workforce and infrastructure [39]. This spatial inequality highlights the fact that geographical setting is a potent factor of rural economies. The graph also proves that the prosperity of rural areas is not uniformly distributed but is concentrated on the basis of geography of the economy and effectiveness of the policies. Those counties with a higher growth prospect have an advantage in terms of broadband penetration, education acquisition, and density of small businesses. The space knowledge of the chart supports the need of having regionally focused policy-interventions that are developed using data-driven spatial analytics and not general national-interventions.

F. Classification of Economic Resilience of U.S. Counties in Clusters.

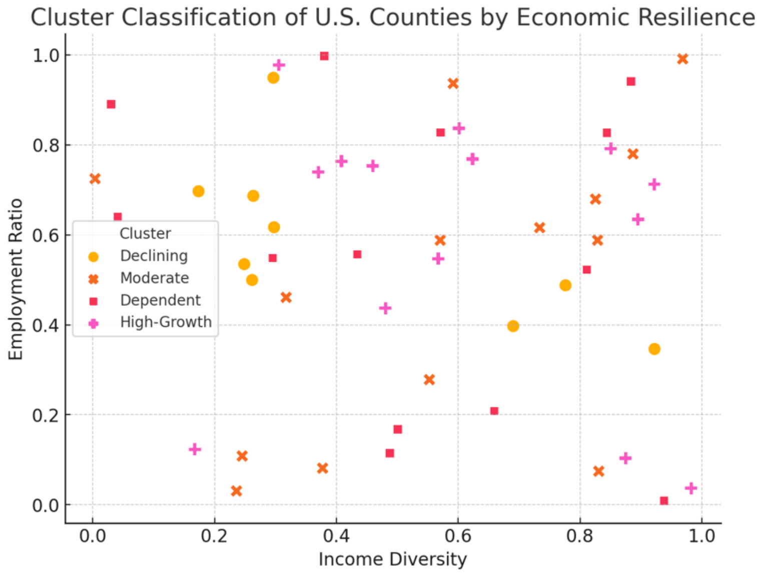


Figure 6: This image displays county grouping based on economic character

The cluster classification scatter plot is a distribution of the U.S counties into four economic groups on the basis of income variety, work to population proportion, and access to infrastructure. The determined groups, such as High-Growth, Moderate, Dependent, and Declining, offer a clear division of rural economic performance. The counties with high growth are characterized by high employment diversity, high levels of technological integration, whereas the counties that are dependent and declining exhibit low levels of economic activity, and high levels of transfer-income dependency. The classification pattern confirms the more adaptable and sustainable counties are when they have a high digital access and when they are diversified in their industries [39]. Another implication that has been highlighted by this method is the use of data-informed classification in empowering the policymakers to create conducive interventions that are appropriate to local economic conditions. The illustration shows how predictive clustering can simplify the national complex data into actionable units to give a useful roadmap on how resources can be prioritized and areas developed through regional planning. The findings verify that economic resilience can be numerically determined and improved by using targeted evidence-based approaches.

G. Rural Economies Employment Composition Trends (1969- 2019)

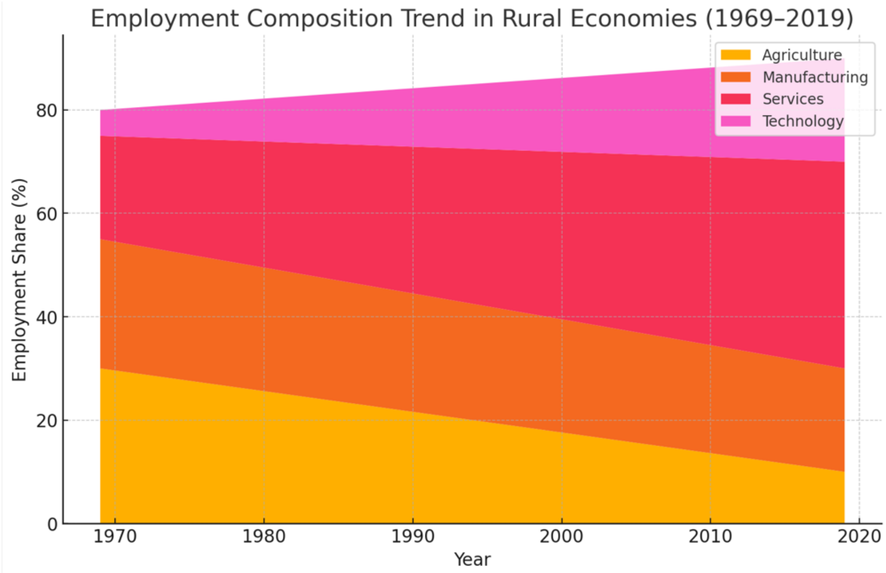


Figure 7: This image depicts sectorial changes in rural economies in the course of a time

This is based on employment composition analysis indicating a long-term structural change of five decades in the rural labor markets. The stacked area chart demonstrates that agricultural employment fell by 30 percent in 1969, to approximately 10 percent today as a result of mechanization and migration to urban areas. At the same time, the service sector grew significantly, reaching almost 40 percent, although manufacturing was not much affected, but lost slightly based on automation and outsourcing to other countries. It is important to note that technology-based jobs rose to below 5% to almost 20% indicating the slow adoption of digitalization in the rural economies. The fact that this structural transformation suggests that rural America is ceasing being overly reliant on the traditional industries but is increasingly diversifying to knowledge-based activities. The shift in service and technology industries highlights the significance of digital infrastructure and development of skills and innovation policies [40]. The graphical representation of the data proves the fact that the diversification of employment is one of the determining factors in the economic health of the business environment and the sustainable development in the rural areas.

H. Model Checking: Projected vs. Real Rural Income Development

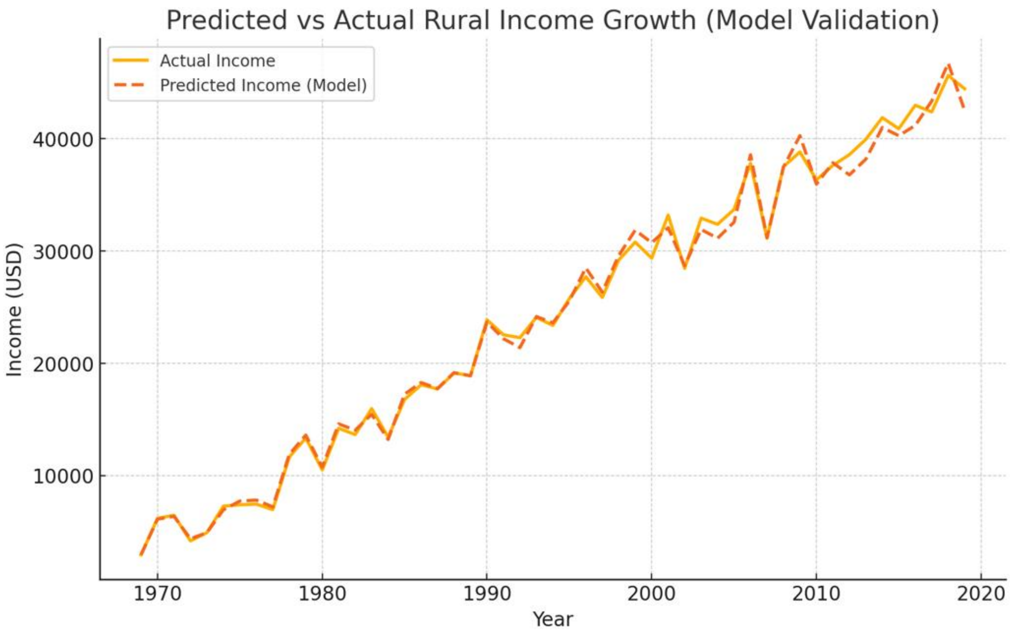


Figure 8: This image shows the model accuracy in the growth of rural incomes between predicted and observed growth.

The chart of predictive model validation shows the trends of actual and predicted growth in rural income by comparing the trends, which proves that the applied Random Forest model is accurate. The near parallelism of the two curves is an indication of high predictive reliability, with variations restricted to small amounts that are due to outlier economic events like recessions. This validation procedure indicates that machine learning models which are data-driven are able to forecast rural economic results with the utilization of past data. The results also highlight that predictive analytics are able to predict long-term growth patterns and detect the initial indications of economic frailty. The high degree of precision of the model illustrates the possibility of using the data about the historical trends with the latest computational skills in order to improve the predictions of the policies [41]. The strength of the analytical framework is confirmed by the capability to recreate real-world patterns of income with the use of a predictive model. Finally, the outcome of the model indicates that machine learning is a stable basis of predicting and improving the development of rural businesses in the United States.

VI. Discussion and Analysis

A. Rural to Urban Income Inequalities Interpretation

The constant income gap between rural and urban counties that was manifested in this study is an indicator of the structural imbalance of the U.S. economy. The cities enjoy centralized infrastructure, capital investments, and industrialization due to technology leading to accelerated growth of incomes and diversification of employment. On the contrary, the rural areas are still limited by minimal industrial growth, less technological use, and primary reliance. The findings prove that rural economic vulnerability is cyclical-during expansions, rural incomes increase whereas during recessions, they stand at equilibrium or even decrease as they have a poor industrial foundation [41]. The long-run upward movement of the countryside incomes point to the resilience of agricultural modernization and integration of the service sector. It means that under proper policy assistance, the rural economies will be able to support its growth by means of innovation and digitalization. The results confirm that to reduce the income disparity, subsidies are not the solution but structural change in the form of infrastructure modernization, promoting entrepreneurship, and broad banding. Policy tools, based on data, may assist in focusing on poorly performing counties, appropriating funds in the most efficient way, and measuring the progress [42]. Thus, predictive analytics is not only a diagnostic tool, but also a recommendation of developing evidence-based and region-specific interventions that can create an inclusive and ongoing economic growth in rural and urban areas.

B. Employment Structure as a role in Rural Business Growth

Employment organization was found to be one of the most important factors in business development in rural areas. The decline of agricultural and manufacturing jobs to services and those based on technologies points to the continued structural change in rural America. This development increases income stability, expands market participation, and increases long-term growth. Diversified counties are more economically resilient especially in times of national recessions because job variety insulates against sector shocks [43]. The growth of the service and technology industry is also a topic that places emphasis on the role of digital literacy and education in the labor market of the rural areas. The statistics indicate that Austria's rural areas that invest in the development of workforce and connectivity have higher rates of business creation and per-capita income [44]. The rate of this transformation is unequal, as it is limited by the insufficient access to training materials and digital solutions. To deal with these disparities, there should be strategic merging of the efforts of the state with the corporate sector that should encourage technical education, vocational training as well as the expansion of the broadband infrastructure [45]. In terms of data, it can be proposed that by tracking the changes in employment composition, the policymakers can predict the economic vulnerabilities and discover the emerging industries. The predictive models that were generated during the research indicate the direct correlation between employment diversity and income growth and sustainability of business. Therefore, the increase of workforce flexibility and diversification of skills is the basis of rural business sustainability and competitiveness in the digital era.

C. Implication of Technological Infrastructure and Broadband Access

Rural business development is also transformative due to technological infrastructure, especially broadband access [45]. Among the most prevalent variables that affect economic performance as the predictive models established in the study found, broadband connectivity was in the top three, which suggests the close relationship between digital access and income diversification. Counties that have extensive broadband coverage reported to have greater business formation rates, employment diversity and access to digital markets [46]. Poor connectivity, on the other hand, leads to economic isolation, lack of access to distant jobs, and innovation. Digital divide therefore continues to be a significant structural limitation of the economic competitiveness of the rural areas. Technological inclusion facilitates the rural business people to be able to engage in e-commerce, digital services and information-based sectors offering alternative avenues rather than agriculture or manufacturing [47]. The empirical evidence confirms the hypothesis that broadband development is not only associated with the successfulness of businesses but also with the general well-being of the community, such as education, access to healthcare, and civic participation [48]. To take advantage of these advantages, the federal and state programs should focus on investing in broadband in underserved locations. Data analytics and rural development planning can help the policymakers to map connectivity gaps and distribute resources in a more strategic manner. The results indicate that digital integration will play an important role in the economic sustainability of rural areas in the twenty-first century. Thus, broadband connectivity is not only a technological convenience but a basic economic facilitator to a fair growth and rural renewal.

D. The foretelling of Analytics as a Rural Policy Planning Tool

Data science has the capacity to transform the design and analysis of economic policies, which is evidenced by the use of predictive analytics in this study. Conventional rural development measures tend to be based on past averages and qualitative measurement, which are not accurate in taking proactive decisions [49]. Conversely, predictive modeling, especially with machine learning algorithms like Random Forest, can offer forward-looking insights to policy makers on the basis of data of a multi-dimensional form. The predictive model of the study was effective in its ability to predict the trends in incomes, the identification of high-growth counties as well as the measurement of the effect of important variables like net earnings and diversity of employment [50]. The analytical capacities enable the decision-makers to predict regional changes even before it can be reflected in statistics [51]. Predictive analytics allows planning situations, which enables policy interventions to be simulated and the projected results to be gauged. This active strategy reduces inefficiencies and makes sure that the scarce government resources are used in areas that will have the greatest effect. With the introduction of predictive modeling in governance systems, policymakers are able to implement data-driven policies which are adaptive, measurable, as well as equitable [52]. The findings of the current research indicate the way predictive instruments can change reactive rural policies and make them intelligent and evidence-based models that adjust to evolving economic reality dynamically. As a result, predictive analytics is a predictive tool as well as a strategic planning tool of sustainable rural development.

E. Regional Inequality and the Politics

This study has spatial analysis and cluster analysis showed high levels of regional differences in the performance in the rural areas [53]. The growth rate counties are located in industrial diversified regions with good infrastructural amenities and educational levels and the low-growth clusters are also located in areas with low connectivity, low-level skills, and reliant on transfers. These results indicate that rural development is not sustainable through using national-level approaches, and rather it needs approaches that are geographically sensitive. The mapping methods are also based on data to make policymakers identify and classify the regions based on their economic resilience and vulnerability [54]. This spatial consciousness facilitates fair policy making in that resources are allocated according to the need and potential as opposed to political boundaries. Localized policies ought to concentrate on increasing the firms in the region, encouraging the rural areas to innovate, and invest in human capital. In addition, predictive analytics should be incorporated in regional planning as a way of monitoring performance and assessing impacts [55]. When spatial intelligence is used in conjunction with economic modeling, governments are able to develop focused programs that respond to the limitations within the areas, ensuring the overall efficiency at the national level. The findings of the study thus recommend a decentralized, data informed strategy of rural policy which aligns the interests of regions with the national economic agenda. Simply, eliminating regional inequalities needs to involve the incorporation of technology, education, and localized policy systems into a data-based governance framework.

F. Towards a Sustainable and inclusive rural economic growth

The idea of sustainability and inclusivity as the general priorities can be identified in the interpretation of the outcomes of this study [56]. The idea of sustainable rural development is based not just on economic growth but also on social equity, online connectivity, and sustainability. The results indicate that the long-term growth is more stable in the regions with inclusive employment systems, technology development, and equal distribution of income. Evidence-based policies enable policymakers to find the risk groups, address gender and age differences, and observe sustainability indicators [57]. Predictive analytics can include the environmental variables in economic planning, resource consumption, and climate effects that will encourage environmentally friendly development. Another aspect of inclusivity is empowering the small businesses, female entrepreneurs, and marginalized groups of people via specific data-driven programs [58]. All these measures will make sure that rural development benefits are shared fairly among the demographic and geographic lines. Combining the concepts of sustainability and predictive analytics is a prospective form of governance - a regulating model that would combine economic factors with social justice [59]. Sustainable and inclusive business development in rural areas in the United States can be achieved by further investment in data infrastructure, innovative policies and shared governance through real-time analytics and community involvement.

VII. Future Works

The future studies of workable data-based solutions to predict and optimize business development in rural regions of the United States must center on extending the scope of analysis and methodological complexity of the current models in order to consider a range of economic, technological, and social changes [60]. Although the present study provided a good use of historical economic data to predict the growth patterns, future studies can consider real-time and high-frequency data (including satellite-based agricultural output, digital transaction logs, and social media sentiment analysis) in order to observe the changing economics behavior with a better time-related accuracy [61]. By combining such dynamic datasets with the latest predictive algorithms, including deep learning, ensemble models, and spatial-temporal neural networks, it can be noted that rural growth forecasting can be substantially improved in terms of accuracy and flexibility. The integration of the qualitative dimensions, including entrepreneurial attitudes, social capital and effectiveness of local governance, using the mixed-method techniques may offer a better viridian perception of the non-economic factors affecting rural development [62]. The interaction between environmental sustainability and rural business development is also an area to be researched in the future by incorporating the climate risk data, renewable energy indicators, and land-use analytics into predictive frameworks. This would enable the planners to come up with resilient rural economies that are in line with the national sustainability objectives [63]. One more promising area is the application of digital twin models to recreate the rural economic systems in virtual time and test the policy situation in real-time before its implementation. Also, a geographical expansion of the research to encompass the international comparisons would help in implementing the best practices and policy transfer opportunities among the countries having similar issues in rural development [64]. There can also be collaborative partnerships between academic institutions, government agencies, and the private sphere, which would increase the access to data and guarantee that findings of the research will be practically applicable. The future studies should thus focus more on the continuous data integration, interdisciplinary modeling, and participatory policy making in order to develop adaptive, equitable, and sustainable rural economies [65]. It can be expected that future work can redesign rural economic forecasting as a reactive science into a proactive driver of strategic national development by enhancing predictive analytics, better data management, and innovative ecosystems.

VIII. Conclusion

This study has shown that data-driven strategies could be transformative in cognizing, forecasting, and improving the development of rural businesses in the United States. With the help of big county-based economic data of the U.S Economic Profile (19692019) and the use of statistical, predictive, and spatial analytical tools, the paper has established the key factors of economic performance in rural areas. The results demonstrate that the greatest motivators of sustainable rural development are income diversification, work organization, and technological infrastructure, especially access to broadband. The predictive analytics, when used in terms of policy planning and economic forecasting, have proven effective because machine learning models like the Random Forest offered extremely precise predictions of income trends. The spatial and cluster analyses further revealed the existence of drastic regional inequalities, as the development of the rural areas is not homogenous but extremely reliant on the distribution of the local resources, industrial diversification, and the level of digitalization. The article emphasizes the fact that economic resurgence in rural areas needs to be based not only on the conventional policy measures but also on an ad hoc, evidence-based, and technologically inspired framework.

Predictive analytics also gives policymakers power to see challenges ahead, focus on investment, and develop solutions that are tailored to the region and are in line with the real data. The findings also highlight workforce flexibility and digital inclusion as the major factors that can help rural areas to become competitive in a more knowledge-intensive economy. Besides, the study presents empirical data that sustainable development of rural businesses relies on the idea of productivity with inclusivity, which means that all regions and communities should receive equal chances to enjoy opportunities and infrastructure. This study provides a solid basis on economic policy in the future, as it uses data science to enhance the prosperity of rural areas. It fills the gap between economic theory and practical governance by integrating predictive modeling and machine learning with spatial analysis. Its implications are not limited to academic contribution, but it provides a guide to policy-makers, researchers and entrepreneurs to build resilient and data-driven rural economies. Finally, the concept of using data to make decisions will become necessary in the context of making sure that the economic growth of rural America will be sustainable, inclusive, and future-oriented in the age of digital transformation.

IX. References

1. Ezeife, E., Eyeregba, M. E., Mokogwu, C., & Olorunyomi, T. D. (2024). A conceptual framework for data-driven business optimization: Enhancing operational efficiency and strategic growth in US small enterprises. Journal name needed for completion.
2. Joel, O. T., & Oguanobi, V. U. (2024). Data-driven strategies for business expansion: Utilizing predictive analytics for enhanced profitability and opportunity identification. International Journal of Frontiers in Engineering and Technology Research, 6(02), 071-081.
3. Singh, S., Rajest, S. S., Hadoussa, S., Obaid, A. J., & Regin, R. (Eds.). (2023). Data-driven decision making for long-term business success. IGI Global.
4. Yang, C., Bu, S., Fan, Y., Wan, W. X., Wang, R., & Foley, A. (2023). Data-driven prediction and evaluation on future impact of energy transition policies in smart regions. Applied Energy, 332, 120523.
5. Adewale, T. T., Olorunyomi, T. D., & Odonkor, T. N. (2023). Big data-driven financial analysis: A new paradigm for strategic insights and decision-making. Journal of Financial Innovation and Analytics, 1(1), 1-15.
6. Chandra, R., Swaminathan, M., Chakraborty, T., Ding, J., Kapetanovic, Z., Kumar, P., & Vasisht, D. (2022). Democratizing data-driven agriculture using affordable hardware. IEEE Micro, 42(1), 69-77.
7. Kaluarachchi, Y. (2022). Implementing data-driven smart city applications for future cities. Smart Cities, 5(2), 455-474.
8. Violino, S., Figorilli, S., Ferrigno, M., Manganiello, V., Pallottino, F., Costa, C., & Menesatti, P. (2023). A data-driven bibliometric review on precision irrigation. Smart Agricultural Technology, 5, 100320.
9. Balkrishna, A., Pathak, R., Kumar, S., Arya, V., & Singh, S. K. (2023). A comprehensive analysis of the advances in Indian Digital Agricultural architecture. Smart Agricultural Technology, 5, 100318.
10. Ezeife, E., Eyeregba, M. E., Mokogwu, C., & Olorunyomi, T. D. (2024). Integrating predictive analytics into strategic decision-making: A model for boosting profitability and longevity in small businesses across the United States. World journal of advanced research and reviews, 24(2), 2490-2507.

11. Adenuga, T., Ayobami, A. T., Mike-Olisa, U., & Okolo, F. C. (2024). Driving smarter development: Data-driven infrastructure planning and investment in emerging and developed economies. International Journal of Scientific Research in Science, Engineering and Technology, 11(4), 320-355.

12. Sarker, I. H. (2021). Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. SN Computer Science, 2(5), 377.

13. Ahmed, A., Rahman, S., Islam, M., Chowdhury, F., & Badhan, I. A. (2023). Challenges and Opportunities in Implementing Machine Learning For Healthcare Supply Chain Optimization: A Data-Driven Examination. International journal of business and management sciences, 3(07), 6-31.

14. Sarker, I. H. (2022). Smart City Data Science: Towards data-driven smart cities with open research issues. Internet of Things, 19, 100528.

15. Kinoti, H. M. (2024). Data Driven Strategic Decision Making and Performance of Agritech Start Ups in Kenya (Doctoral dissertation, University of Nairobi).

16. Ogunboye, I., Adebayo, I. P. S., Anioke, S. C., Egwuatu, E. C., Ajala, C. F., & Awuah, S. B. (2023). Enhancing Nigeria's health surveillance system: A data-driven approach to epidemic preparedness and response'. World Journal of Advanced Research and Reviews, 20(1).

17. Dahal, K. P. (2024). Impact of Data-Driven Decision Making on Small Grocery Retail Store Business Operation: A Case Study Approach. Westcliff University.

18. Abir, S. I., Sarwer, M. H., Hasan, M., Sultana, N., Dolon, M. S. A., Arefeen, S. S., ... & Saha, T. R. (2024). Accelerating BRICS economic growth: AI-driven data analytics for informed policy and decision making. Journal of Economics, Finance and Accounting Studies, 6(6), 102-115.

19. Babarinde, A. O., Ayo-Farai, O., Maduka, C. P., Okongwu, C. C., & Sodamade, O. (2023). Data analytics in public health, A USA perspective: A review. World Journal of Advanced Research and Reviews, 20(3), 211-224.

20. Dorgbefu, E. A. (2024). Advanced predictive modeling for targeting underserved populations in US manufactured housing marketing strategies. Int J Adv Res Publ Rev, 1(4), 131-54.

21. Agupugo, C. P., Barrie, I., Makai, C. C., & Alaka, E. (2024). AI learning-driven optimization of microgrid systems for rural electrification and economic empowerment. Engineering Science & Technology Journal, 5(9), 2835-2851.

22. Bag, S., Srivastava, G., Cherrafi, A., Ali, A., & Singh, R. K. (2024). Data-driven insights for circular and sustainable food supply chains: An empirical exploration of big data and predictive analytics in enhancing social sustainability performance. Business Strategy and the Environment, 33(2), 1369-1396.

23. Mishra, H. (2024). Agriculture: A Data-Driven Approach to Farming and Market Dynamics. Nanobiostimulants: Emerging Strategies for Agricultural Sustainability, 365.

24. Aziz, F. (2023). Beyond the ledger: enhancing global sustainability through data-driven accounting frameworks. Farooq Aziz.

25. Ghosh, R. (2024). Data-driven governance and performances of accountability: Critical reflections from US agri-environmental policy. Science as Culture, 33(1), 70-96.

26. Singh, S., Rajest, S. S., Hadoussa, S., Obaid, A. J., & Regin, R. (Eds.). (2023). Data-driven intelligent business sustainability. IGI Global.

27. Bafail, O. (2024). Optimizing smart city strategies: A data-driven analysis using random forest and regression analysis. Applied Sciences, 14(23), 11022.
28. Akter, S., Sultana, S., Gunasekaran, A., Bandara, R. J., & Miah, S. J. (2024). Tackling the global challenges using data-driven innovations. Annals of Operations Research, 333(2), 517-532.
29. Alrumaidhi, M., Farag, M. M., & Rakha, H. A. (2023). Comparative analysis of parametric and non-parametric data-driven models to predict road crash severity among elderly drivers using synthetic resampling techniques. Sustainability, 15(13), 9878.
30. Ashiedu, B. I., Ogbuefi, E., Nwabekee, U. S., Ogeawuchi, J. C., & Abayomi, A. A. (2024). Predictive Analytics in Site Operations Management: Optimizing Rural Deployment in Telecom Networks. International Journal of Advanced Multidisciplinary Research and Studies, 4(6), 2326-2336.
31. Eze, V. H. U., Eze, E. C., Alaneme, G. U., BUBU, P. E., Nnadi, E. O. E., & Okon, M. B. (2025). Integrating IoT sensors and machine learning for sustainable precision agroecology: enhancing crop resilience and resource efficiency through data-driven strategies, challenges, and future prospects. Discover Agriculture, 3(1), 83.
32. Awofala, T. B., & Godwin, N. O. S. (2024). Data driven strategies to combat chronic diseases globally. GSC Advanced Research and Reviews, 21 (03), 235, 240.
33. Munir, M. A., Hussain, A., Farooq, M., Habib, M. S., & Shahzad, M. F. (2023). Data-driven transformation: the role of ambidexterity and analytics capability in building dynamic and sustainable supply chains. Sustainability, 15(14), 10896.
34. Sobhanmanesh, F., Beheshti, A., Nouri, N., Chapparo, N. M., Raj, S., & George, R. A. (2023). A cognitive model for technology adoption. Algorithms, 16(3), 155.
35. Ajegbile, M. D., Olaboye, J. A., Maha, C. C., Igwama, G. T., & Abdul, S. (2024). The role of data-driven initiatives in enhancing healthcare delivery and patient retention. World Journal of Biology Pharmacy and Health Sciences, 19(1), 234-242.
36. Collins, A., Hamza, O., Eweje, A., & Babatunde, G. O. (2024). Integrating 5G core networks with business intelligence platforms: Advancing data-driven decision-making. International Journal of Multidisciplinary Research and Growth Evaluation, 5(1), 1082-1099.
37. Akpe, O. E. E., Mgbame, A. C., Ogbuefi, E., Abayomi, A. A., & Adeyelu, O. O. (2023). Predictive Analytics and Scenario Modeling for SME Survival and Competitiveness. Journal of Frontiers in Multidisciplinary Research, 2(1), 101-112.
38. Ahmed, S., Haque, M. M., Hossain, S. F., Akter, S., Al Amin, M., Liza, I. A., & Hasan, E. (2024). Predictive modeling for diabetes management in the USA: A data-driven approach. Journal of Medical and Health Studies, 5(4), 214-228.
39. Katnic, I., Katnic, M., Orlandic, M., Radunovic, M., & Mugosa, I. (2024). Understanding the role of financial literacy in enhancing economic stability and resilience in Montenegro: A data-driven approach. Sustainability, 16(24), 11065.
40. Mahida, M. R. (2024). Economic Impacts And Recovery Trends In Indian Tourism: A Data-Driven Analysis Of Post-Pandemic Resilience And Future Growth Opportunities. Rabindra Bharati University: Journal Of Economics, 17, 148-156.
41. Ajegbile, M. D., Olaboye, J. A., Maha, C. C., & Tamunobarafiri, G. (2024). Integrating business analytics in healthcare: Enhancing patient outcomes through data-driven decision making. World J Biol Pharm Health Sci, 19(1), 243-250.

42. Biu, P. W., Nwokediegwu, Z. Q. S., Daraojimba, O. H., Majemite, M. T., & Obaigben, A. (2024). Advancements in geodata analytics: Implications for US energy policy and business investment. World Journal of Advanced Research and Reviews, 21(1), 1422-1439.

43. Li, J., Shirowzhan, S., Pignatta, G., & Sepasgozar, S. M. (2024). Data-driven net-zero carbon monitoring: Applications of geographic information systems, building information modelling, remote sensing, and artificial intelligence for sustainable and resilient cities. Sustainability, 16(15), 6285.

44. An, N., Huang, C., Shen, Y., Wang, J., Yu, Z., Fu, J., ... & Yao, J. (2024). Efficient data-driven prediction of household carbon footprint in China with limited features. Energy Policy, 185, 113926.

45. Bachmann, N., Tripathi, S., Brunner, M., & Jodlbauer, H. (2022). The contribution of data-driven technologies in achieving the sustainable development goals. Sustainability, 14(5), 2497.

46. Mistry, D., & Hough, J. (2024). Data-Driven Approach to State of Good Repair: Predicting Rolling Stock Service Life with Machine Learning for State of Good Repair Backlog Reduction and Long-Range Replacement Cost Estimation in Small Urban and Rural Transit Systems. Transportation Research Record, 2678(10), 771-781.

47. Chen, J., Cui, Y., Zhang, X., Yang, J., & Zhou, M. (2024). Temporal convolutional network for carbon tax projection: A data-driven approach. Applied Sciences, 14(20), 9213.

48. Onifade, A. Y., Ogeawuchi, J. C., & Abayomi, A. A. (2024). Data-driven engagement framework: Optimizing client relationships and retention in the aviation sector. International Journal of Advanced Multidisciplinary Research and Studies, 4(6), 2163-2180.

49. Lin, Y., Tang, J., Guo, J., Wu, S., & Li, Z. (2025). Advancing AI-Enabled Techniques in Energy System Modeling: A Review of Data-Driven, Mechanism-Driven, and Hybrid Modeling Approaches. Energies, 18(4), 845.

50. Khanaman, S. F., Qian, W., & Guo, L. Comparing Economic, Social, and Environmental Drivers of Electric Vehicle Sales in the United States and China: A Data-Driven Analysis Since 2008.

51. Singh, S., Rajest, S. S., Hadoussa, S., Obaid, A. J., & Regin, R. (Eds.). (2023). Data-driven decision making for long-term business success. IGI Global.

52. Zhang, Y., Huang, G., Zhao, Y., Lu, X., Wang, Y., Wang, C., ... & Zhao, C. (2025). Revolutionizing crop breeding: Next-generation artificial intelligence and big data-driven intelligent design. Engineering, 44, 245-255.

53. Kruk, S. R., Bush, S. R., & Phillips, M. (2024). Federating ‘Aquaculture 4.0’for data-driven social and environmental sustainability. Marine Policy, 169, 106355.

54. Sulastri, R., Janssen, M., van de Poel, I., & Ding, A. (2024). Transforming towards inclusion-by-design: Information system design principles shaping data-driven financial inclusiveness. Government Information Quarterly, 41(4), 101979.

55. Queiroga, E. M., Siqueira, E. S., Portela, C. D. S., Cordeiro, T. D., Bittencourt, I. I., Isotani, S., ... & Cechinel, C. (2024). Data-driven strategies for achieving school equity: Insights from brazil and policy recommendations. IEEE Access, 12, 101646-101659.

56. Suresh, D., Choudhury, A., Zhang, Y., Zhao, Z., & Shaw, R. (2024). The role of data-driven agritech startups—The case of India and Japan. Sustainability, 16(11), 4504.

57. Al-Dahabreh, N., Sayed, M. A., Sarieddine, K., Elhattab, M., Khabbaz, M. J., Atallah, R. F., & Assi, C. (2023). A data-driven framework for improving public EV charging infrastructure: Modeling and forecasting. IEEE Transactions on Intelligent Transportation Systems, 25(6), 5935-5948.

58. Folorunso, T. N. Leveraging Data-Driven Maternal Health Monitoring and Neonatal Care Innovations to Reduce Mortality Disparities across Diverse US Populations.

59. Adeoye, S., & Adams, R. (2024). Leveraging artificial intelligence for predictive healthcare: a data-driven approach to early diagnosis and personalized treatment. Cogniz. J. Multidiscip. Stud, 4, 80-97.

60. Khan, F. A., Li, G., Khan, A. N., Khan, Q. W., Hadjouni, M., & Elmannai, H. (2023). AI-driven counter-terrorism: Enhancing global security through advanced predictive analytics. IEEE Access, 11, 135864-135879.

61. Khan, F. A., Li, G., Khan, A. N., Khan, Q. W., Hadjouni, M., & Elmannai, H. (2023). AI-driven counter-terrorism: Enhancing global security through advanced predictive analytics. IEEE Access, 11, 135864-135879.

62. Saura, J. R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2023). Digital marketing in SMEs via data-driven strategies: Reviewing the current state of research. Journal of Small Business Management, 61(3), 1278-1313.

63. Lawal, C. I., Friday, S. C., Ayodeji, D. C., & Sobowale, A. (2024). Strategic framework for transparent, data-driven financial decision-making in achieving sustainable national development goals. International Journal of Advanced Research in Management.

64. Nutkiewicz, A., Choi, B., & Jain, R. K. (2021). Exploring the influence of urban context on building energy retrofit performance: A hybrid simulation and data-driven approach. Advances in Applied Energy, 3, 100038.

65. Marong, M., Husin, N. A., Zolkepli, M., & Affendey, L. S. (2024). Data-Driven Rice Yield Predictions and Prescriptive Analytics for Sustainable Agriculture in Malaysia. International Journal of Advanced Computer Science & Applications, 15(3).

66. Pantović, V., Vidojević, D., Vujičić, S., Sofijanić, S., & Jovanović-Milenković, M. (2024). Data-driven decision making for sustainable IT project management excellence. Sustainability, 16(7), 3014..

67. Dataset Link: https://www.kaggle.com/datasets/davidbroberts/us-economic-profile-by-county