RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Analysis of Socioeconomic Data from the Sepotuba Valley for the Development of a Predictive Model of Economic *Growth*

Luiz Henrique Camargo Moura[1] – State University of Mato Grosso Prof. Me. Leandro Mazurek[2] – State University of Mato Grosso

Prof.ª Ma. Raquel da Silva Vieira Coelho3 – State University of Mato Grosso

Summary

This article proposes the analysis of socioeconomic data from the Sepotuba Valley with the aim of developing a predictive model of economic growth based on machine learning techniques. The research adopts a quantitative and descriptive approach, integrating fundamentals of data science, regional economics, and sustainable development. The study seeks to understand the factors that influence regional growth, structuring historical and current data on indicators such as GDP, education, unemployment, and infrastructure. Through predictive modeling, the intention is to generate insights that can support public policies and strategic decisions for the region.

Keywords: Machine Learning; Economic Growth; Socioeconomic Analysis; Sepotuba Valley.

Abstract

This article proposes the analysis of socioeconomic data from the Vale do Sepotuba in order to develop a predictive economic growth model based on machine learning techniques. The research adopts a quantitative and descriptive approach, integrating data science, regional economics, and sustainable development. The study seeks to understand the factors that influence regional growth, structuring historical and current data of indicators such as GDP, education, unemployment, and infrastructure. Through predictive modeling, the aim is to generate insights that can support public policies and strategic decisions for the region.

Keywords: Machine Learning; Economic Growth; Socioeconomic Analysis; Sepotuba Valley.

1 . INTRODUCTION

Regional economic growth is a complex phenomenon, influenced by multiple factors.
interdependent factors, such as human capital, infrastructure, technological innovation, education and
public policies. In the context of the Sepotuba Valley, in southwestern Mato Grosso, understanding these
The dynamics become even more challenging due to historical, territorial, and...
socioeconomic factors that shaped local development. The region includes prominent municipalities in
Mato Grosso's economy, including cities like Tangará da Serra, Campo Novo do Parecis, and Sapezal, at the same time
a time when it presents localities with structural limitations, such as Lambari D'Oeste, Salto do Céu.
and Porto Estrela. That diversity.

The internal landscape reflects a reality marked by contrasts, with agro-industrial centers integrated into supply chains.
Global commodity markets coexisting with regions of low productive diversification and less access to
essential public services.

Despite its productive potential, the region's economic development faces barriers.
significant, including territorial inequality, insufficient infrastructure in some municipalities and

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Fragmentation of socioeconomic data, which hinders accurate diagnoses and the formulation of Effective public policies. According to recent data from IBGE (2023) and IPEA (2023), the GDP Per capita income in the Sepotuba Valley varies between R$ 28,000 in smaller municipalities and R$ 55,000 in agro-industrial centers; the Municipal Human Development Index shows variations from 0.65 to 0.78; and the regional unemployment rate fluctuates between 5% and 10%, depending on the municipality. These indicators They highlight internal inequalities and reinforce the need for detailed analyses that take into account the specific context of each location.

In this scenario, the use of machine learning techniques emerges as an alternative. promising, since, unlike traditional statistical methods, these algorithms are capable of identifying hidden patterns in large volumes of data, exploring non-linear relationships between variables and generate robust forecasts in uncertain scenarios. Applied to the regional context, They can reveal determinants of economic growth that would not be easily perceived by through conventional analysis. Recent studies highlight that the integration between data analysis, Big data and machine learning can help public managers and private investors to... to better understand growth dynamics, directing resources to strategic areas and reducing investment risks (Mayer-Schönberger; Cukier, 2013; Provost; Fawcett, 2013). No In the Sepotuba Valley, this application not only strengthens agro-industrial competitiveness, but also It fosters balanced development, encompassing education, infrastructure, and sustainability. environmental and technological innovation.

During the literature review, no studies were found that applied models of Machine learning specifically for municipalities in the Sepotuba Valley. This gap highlights The originality of the research reinforces its scientific contribution by exploring a regional context. still understudied with advanced predictive analytics techniques.

Given this context, this research seeks to answer the following question: how to apply Machine learning techniques to predict the economic growth of municipalities in the Valley. from Sepotuba based on available socioeconomic data, in order to support Public policies and strategic decisions of the private sector? To achieve this goal, the study proposes collect and organize historical and current socioeconomic data, apply forecasting models, such as Linear Regression and Random Forest: Evaluate the performance of these models and interpret the results. obtained, generating recommendations applicable to both the public and private sectors. Thus, the The overall objective is to propose and analyze a predictive model of economic growth for the municipalities in the region, while specific objectives include gathering reliable data, applying Machine learning techniques for predictive analytics, evaluating the accuracy of models, and developing... Well-founded recommendations for public policies and investment strategies.

The study's relevance lies in providing input for municipal planning.

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

More efficient, supporting public policies and reducing risks for the private sector. Innovation lies in Application of machine learning for predictive analytics in a municipal context enabling to identify determinants of economic growth that traditional methods They might not reveal this information, contributing to a fairer and more sustainable regional development. and innovative.

The article is structured as follows: Chapter 2 presents the literature review. Chapter 3 discusses regional economic growth, data analysis, and machine learning. It details the methodology adopted, including data collection and the application of predictive models; Chapter 4 discusses the results obtained and their implications; and Chapter 5 presents the conclusions and... Recommendations for public policies and future research.

2 THEORETICAL FOUNDATION

When discussing economic growth, especially in regions facing challenges Significant examples, such as the Sepotuba Valley, are insufficient to limit the analysis to quantitative data or to... The simple application of conventional theories is not enough. It becomes essential to consider the social context and regional, interpreting the data critically and understanding the dynamics that occur in territory. The Sepotuba Valley encompasses several municipalities in southwestern Mato Grosso, including Tangará da Serra, Campo Novo do Parecis, Sapezal, Barra do Bugres, Denise, Nova Olímpia, Lambari D'Oeste, Rio Branco, Salto do Céu and Porto Estrela (SEMA-MT, 2021). Before Furthermore, this stage of the study integrates data analysis, machine learning techniques, and... fundamentals of the regional economy, with the aim of building a robust analytical base that contribute to projecting the region's economic growth.

2.1 Data analysis, machine learning and Big Data in support of development regional

Analyzing socioeconomic data goes beyond identifying economic patterns; it is also about... to understand the most complex social relationships. Indicators such as Gross Domestic Product (GDP), the The unemployment rate, the average level of education, investments, and infrastructure show... Realities that directly impact the development of the Sepotuba Valley region. This data They are not just for filling out spreadsheets, but are powerful tools for transforming realities. Based on them, it is possible to devise fairer public policies and make decisions. strategies that truly improve people's lives in the region.

As Wooldridge (2019) points out, when using statistical modeling within econometrics, not

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

We are not just identifying relationships between numbers, but understanding the stories told by the data.
which help us predict future scenarios and act with greater precision. By cross-referencing this information,
We realize how socioeconomic factors intertwine and influence the region's growth.
opening up space for more targeted actions.

Machine learning is gaining increasing importance in
economic forecasting due to its ability to handle large volumes of data, finding
Complex patterns and generating predictions based on solid evidence. Using algorithms.
Using supervised algorithms, such as linear regression, neural networks, and decision trees, it is possible to create models.
accurate even in scenarios with multiple variables, which is especially useful in regions with
specific characteristics, such as the Sepotuba Valley. According to Hastie, Tibshirani and Friedman (2009),
The key difference with machine learning lies in its ability to identify non-linear and subtle relationships.
which traditional methods miss.

The Big Data era has expanded the capacity for collecting and analyzing information, changing the way...
to think about and formulate public policies. According to Kitchin (2014), data-driven science allows
To generate insights that go beyond traditional approaches, making decision-making processes more agile.
and responsive. In the municipal and regional context, data-driven predictive analytics becomes a
An essential tool for optimizing public policies, offering more accurate forecasts.
well-founded and efficient.

Therefore, the integration between data analysis, machine learning, and Big Data represents
A significant step forward in regional development planning, enabling decisions based on sound principles.
providing evidence and strengthening the efficiency of public management.

2.2 Economic growth and technology as drivers of human development

A region's economic growth goes beyond GDP growth. It's necessary to observe...
How it is reflected in people's lives, in access to education, health, infrastructure, and employment.
The theory of endogenous growth, proposed by Romer (1990), highlights the role of knowledge, of
Innovation and human capital as central pillars for sustainable development.

Investing in data analytics and predictive models is therefore an essential strategy for...
regional planning. This makes it possible to identify factors that promote or hinder regional planning.
growth and guiding more efficient public policies and private investments.

However, it is essential to understand that the use of technology should not be dissociated from...
Human values. The ethical use of artificial intelligence, the transparency of public data, and respect.
Cultural diversity is an essential principle for fair and sustainable development.
As Diniz (2013) points out, technology should serve equity and social justice, and not just efficiency.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025
technique.

2.3 Historical aspects, agricultural modernization and territorial challenges in the Sepotuba Valley

An analysis of economic growth in the Sepotuba Valley requires a historical perspective on the...
Occupation cycles and the use of natural resources. The occupation of southwestern Mato Grosso.
This intensified in the 1970s with federal integration policies, such as the Program of
National Integration Program (PIN), which encouraged migration and the expansion of the agricultural frontier. According to
Oliveira (2007), this process consolidated a development model based on agro-industry.
An export-oriented country with low product diversification.

During the following decades, municipalities such as Sapezal, Campo Novo do Parecis, and Tangará
The Serra region received large private investments in rural infrastructure and mechanization, while
Other areas, such as Lambari D'Oeste, Salto do Céu, and Porto Estrela, remained with economies of
smaller scale and dependence on agriculture (SILVA et al., 2022; LEANDRO, 2020). This
This differentiation is still reflected in current socioeconomic indicators.

Agricultural modernization has increased productivity and integrated the region into the global supply chain.
Export, but it also generated social impacts, such as land concentration and displacement.
population. According to Delgado (2012), technological advances in the field have led to the exclusion of
workers and reinforced inequalities. Municipalities with better
Infrastructure projects developed rapidly, while others remained on the sidelines, highlighting the productivity paradox
(BRANDÃO, 2007).

Logistical and infrastructure bottlenecks remain among the main challenges. According to
According to IMEA (2021), the lack of paved roads and digital connectivity directly affects the
Competitiveness perpetuates inequalities. Overcoming these limitations requires not only physical infrastructure,
but also the strategic use of data to map production chains and identify local vocations.
and to predict returns on social investment.

2.4 Education, innovation and regional sustainability

Sustainable economic development is strongly linked to the quality of
education and the development of human capital. In the Sepotuba Valley, where the economy is
predominantly agricultural, technical and scientific training is essential for diversification and
Innovation. According to Romer (1990), knowledge and the capacity for innovation are the driving forces of
long-term growth.

Municipalities that invest in the education of their youth and encourage partnerships between technical schools

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

cas, un vers dades e setor produt vo cr am ambi ntes ma s favoráve s à novação. No contexto regional,
Strengthening UNEMAT, federal institutes, and rural innovation centers can help to...
transition from an economy based on natural resources to one based on knowledge.

According to Sachs (2004) and Veiga (2008), sustainable development needs
Integrating economic growth, social equity, and environmental preservation. In the Valley of
In Sepotuba, this means balancing agro-industrial development with support for family farming.
to technical education and environmental conservation, building a regional model
A truly sustainable truth.

2.5 Technology, digital government and spatial analysis for regional planning

With the advancement of digital transformation, the use of technologies in the public sector has become essential.
essential for more efficient and transparent management. Tools such as dashboards and systems of
Geographic information systems (GIS), artificial intelligence, and collaborative platforms expand capabilities.
planning and monitoring socioeconomic indicators.

In the Sepotuba Valley, digital government can improve resource allocation,
monitoring public policies and the provision of essential services, in addition to strengthening the
Citizen participation and democratic governance. Castells (2000) highlights that the intelligent use
The use of information technologies is crucial for reducing territorial inequalities, especially in
peripheral regions.

Geospatial analysis and the clustering of municipalities are effective tools for
to understand patterns of development and inequality. According to Anselin (1995), spatial analysis
It captures territorial externalities and allows us to identify how the development of a municipality can...
to influence others around you. Tools like QGIS, Google Earth Engine, and Power BI help to
Visualize data and guide tailored policies.

Almeida et al. (2015) and Maraschin et al. (2022) highlight that techniques such as K-means and
DBSCAN allows grouping municipalities with similar characteristics, facilitating the preparation of
Public policies adapted to local specificities. This integration between technology and planning
and territory consolidates a solid foundation for the sustainable and equitable development of the Valley of
Sepotuba.

3. Details the methodology adopted, including data collection and the application of the models. predictive

This research adopts a quantitative and descriptive approach, characterizing itself as a
An exploratory study that integrates data science methods applied to a regional context. According to
Gil (2010), exploratory research aims to provide greater familiarity with the

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

The problem can be made more explicit or hypothesized. In the present study,
The aim is to explore the relationships between socioeconomic variables and regional economic growth. through advanced machine learning techniques.

The quantitative nature of the research is justified by the need to measure and analyze. Numerically, the socioeconomic indicators of the municipalities in the Sepotuba Valley. How? Cervo, Bervian and Silva (2007) argue that quantitative research allows for the obtaining of results. which can be quantified and generalized, in addition to enabling the use of statistical tools. for data analysis. The descriptive approach, in turn, allows us to characterize the current situation of municipalities studied, identifying patterns and establishing relationships between the observed variables.

According to Aguiar and Baptista (2023), the use of machine learning for analysis of Socioeconomic data has become established as an effective methodology for identifying patterns. complex and predictive insight generation. Silva, Teles and Araújo (2025) reinforce that the Applicability of artificial intelligence in territorial planning
It allows us to overcome the limitations of conventional analyses, especially in regional contexts. characterized by socioeconomic heterogeneity.

3.1 Study Area and Spatial Delimitation

The Sepotuba Valley constitutes an important hydrographic and socioeconomic region in Southwest of Mato Grosso, composed of ten municipalities: Tangará da Serra, Campo Novo do Parecis, Sapezal, Barra do Bugres, Denise, Nova Olímpia, Lambari D'Oeste, Rio Branco, Salto do Céu and Porto Estrela. The region is characterized by a strong agro-industrial vocation, however, significant heterogeneity in their development indicators.

The choice of this region as a study area is justified by three fundamental aspects: First, its economic relevance to the state of Mato Grosso, housing important centers of agricultural and agro-industrial production; secondly, the diversity of socioeconomic profiles among the municipalities, allowing for comparative analyses; third, the availability of secondary data. reliable for conducting quantitative analyses. As highlighted by Leandro et al. (2019) and Gamero et al. (2020), the Sepotuba River basin presents physical-environmental characteristics and socioeconomic factors that make it representative of the regional development challenges of the Center-Western Brazil.

3.2 Data Collection and Preparation

3.2.1 Data Sources

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

The data collection was structured using official secondary sources, ensuring...
Reliability and replicability of the research. The main databases consulted included:

a) **Brazilian Institute of Geography and Statistics (IBGE):** data on municipal Gross Domestic Product, estimated population, Municipal Human Development Index (MHDI) and educational indicators;

b) **Institute for Applied Economic Research (IPEA):** historical series of indicators socioeconomic data via the IPEADATA platform;

c) **Mato Grosso Institute of Agricultural Economics (IMEA):** sectoral reports about regional agricultural economics;

d) **State Secretariat of Planning and Management of Mato Grosso (SEPLAN-MT):** Data from the Multi-Year Plan and regional diagnoses.

The use of multiple official sources aims to ensure the triangulation of technical data. a methodological approach that, according to Minayo (2001), increases the validity and reliability of the results. Goes Goes (2025) and others highlight that the quality of input data is a critical factor for the success of models. Machine learning applied to the predictability of economic indicators.

3.2.2 Study Variables

The predictive model was structured based on eight main variables, with GDP per capita being... the dependent variable. The independent variables included HDI, unemployment rate, average schooling, infrastructure score, economic diversification, total population, and total GDP. A The selection of these indicators is based on the literature on regional development: Romer (1990) highlights the role of human capital; Sachs (2004) and Veiga (2008) emphasize the importance of infrastructure and productive diversification; and Sen (2000) emphasizes that development must be understood from the perspective of expanding human capabilities, as reflected in social indicators such as HDI and education.

3.2.3 Dataset Structuring Process

The dataset was structured following these steps, as recommended by McKinney. (2012) and Provost and Fawcett (2013):

- **Extraction and Consolidation:** Data from different sources were extracted and consolidated into a format Creating a tabular data structure using Python with the Pandas library for data manipulation.
- **Cleaning and Processing:** Checking for missing values, identifying outliers and

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Correction of inconsistencies. Although the final database did not contain missing data, it was implemented... systematic verification.

- **Transformation of Variables:** Calculation of derived variables, such as GDP per capita (total GDP × 1000 / population), and normalization of scales when necessary.
- **Cross-validation:** Checking values against multiple sources to ensure accuracy.

The final dataset consisted of 10 observations (municipalities) and 9 variables, structured... in accordance with the principles of Tidy Data (organized, clean data ready for analysis). As they point out According to Rodrigues et al. (2022), proper data preparation is a step... crucial in machine learning projects for regional development analysis, impacting directly affects the quality of predictive models.

3.3 Exploratory Data Analysis

Prior to predictive modeling, exploratory data analysis (EDA) was conducted. Exploratory Data Analysis) for understanding the characteristics, distributions, and relationships between variables. This step is crucial for informing model selection and identifying initial patterns. in the data (Kuhn; Johnson, 2013).

3.3.1 Descriptive Statistics

Measures of central tendency (mean, median) and dispersion (standard deviation) were calculated. minimum and maximum values) for all numerical variables. The results revealed:

- Average GDP per capita of R$ 56,687.59, ranging from R$ 27,142.86 to R$ 170,833.33
- Average HDI of 0.688, ranging from 0.628 to 0.780
- Unemployment rate between 3.9% and 10.8%
- Average schooling ranging from 6.0 to 9.8 years

This analysis revealed significant heterogeneity among the municipalities, justifying the... The need for models capable of capturing complex relationships.

3.3.2 Correlation Analysis

A Pearson correlation matrix was constructed to identify linear relationships between variables. The main findings included:

- Strongly negative correlation between unemployment rate and GDP per capita (r = -0.75)
- Positive correlation between HDI and GDP per capita (r = 0.75)
- Positive correlation between infrastructure and GDP per capita (r = 0.69)

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

• Positive correlation between average schooling and GDP per capita (r = 0.75)

These correlations confirm theoretical assumptions about the determinants of growth. economic and validate the inclusion of these variables in the predictive model. As demonstrated by Hastie, Tibshirani and Friedman (2009) state that preliminary correlation analysis aids in the interpretation of results of machine learning models.

3.3.3 Categorization of Municipalities

The municipalities were classified into three categories according to GDP per capita:

• **Low development** (< R$ 30,000): Tangará da Serra, Barra do Bugres, Porto Estrela

• **Medium development** (R$ 30,000 - R$ 45,000): Rio Branco, Salto do Céu

• **High development** (> R$ 45,000): Nova Olímpia, Lambari D'Oeste, Denise, Campo Novo do Parecis, Sapezal

This categorization allowed for comparative analyses between groups and the identification of municipalities. priorities for public policies.

3.4 Predictive Modeling with Machine Learning

Two supervised learning algorithms were implemented: Linear Regression and Random Forest. The choice of these models was based on criteria of interpretability and suitability to... problem and methodological complementarity.

3.4.1 Linear Regression

Linear regression is a fundamental statistical model that establishes a linear relationship between... Independent and dependent variables. According to Wooldridge (2019), linear regression allows not It's not just prediction, but also causal inference when the right assumptions are met.

The model is expressed by:

Y = ÿÿ + ÿÿXÿ + ÿÿXÿ + ⋯ + ÿÿXÿ + ÿ

Where Y is the dependent variable (GDP per capita), Xÿ...Xÿ are the independent variables, ÿÿ...ÿÿ are the coefficients and ÿ is the error term.

Its implementation serves as a baseline for comparison and allows us to assess whether linear relationships They are sufficient to explain the phenomenon studied.

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

3.4.2 Random Forest

Random Forest, developed by Breiman (2001), is an ensemble algorithm that combines Multiple decision trees are used to produce more robust predictions. The model builds several trees. decision-making is performed on random subsets of the data and aggregates its predictions (average for regression).

According to Hastie, Tibshirani, and Friedman (2009), Random Forest presents advantages. Significant: ability to capture non-linear relationships, robustness to outliers, lower tendency towards overfitting and possibility of measuring the relative importance of variables. Geron (2019) It highlights that this algorithm is particularly effective on datasets with complex relationships between variables, a characteristic present in socioeconomic data.

Applying Random Forest in similar contexts has shown superior results. Aguiar and Baptista (2023) used machine learning for urban growth analysis, obtaining high predictive accuracy. Silva, Teles and Araújo (2025) highlight the effectiveness of ensemble algorithms for territorial planning.

3.4.3 Data Preparation for Modeling

Before training the models, the data underwent specific procedures. The dataset was split into training (70%) and test (30%) sets using the train_test_split function. from scikit-learn, adopting random_state = 42 to ensure reproducibility as recommended. Kuhn and Johnson (2013) applied the methodology for moderately sized samples. For Linear Regression, the following was applied: Normalization of variables using StandardScaler is a fundamental step for scale-sensitive algorithms. (Geron, 2019). Considering the small sample size (10 municipalities), the following were adopted Conservative strategies, such as limiting the depth of the Random Forest (max_depth = 5) to reduce the risk of overfitting and to allow for cautious interpretation of performance metrics.

3.4.4 Hyperparameter Training and Tuning

Linear regression was implemented using the default scikit-learn parameters, without adjustment. additional hyperparameters. For the Random Forest model, n_estimators = 100 were defined. max_depth = 5 and random_state = 42. The selection of 100 trees follows Hastie's recommendations. Tibshirani and Friedman (2009) indicate that performance stabilizes from that point onward. However, Limiting the depth of trees is a suitable practice to avoid overfitting in datasets. reduced, as recommended by Kuhn and Johnson (2013).

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

3.4.5 Evaluation Metrics

The performance of the models was evaluated using three complementary metrics. The R^2 The r^2 measures the proportion of variance explained by the model, and is calculated using $R^2 = 1 - (SS_{res} / SS_{tot})$. The RMSE, defined as $RMSE = ÿ[ÿ(y_{real} - y_{previsto})^2 / n]$, expresses the mean error. The quadratic effect in the original unit of the dependent variable penalizes large errors more heavily. The MAE, calculated by $MAE = ÿ|y_{actual} - y_{predicted}| /n$, It represents the mean absolute error, being less sensitive to outliers than the RMSE. According to Kuhn and Johnson (2013), the combined use of these metrics provides a comprehensive assessment of performance, integrating explanatory power, sensitivity to large deviations, and interpretability of the average error.

3.5 Analysis of the Importance of Variables

A crucial advantage of Random Forest is its ability to calculate relative importance. The algorithm measures how much each feature contributes to the reduction of each variable in the predictions. of impurities in the divisions of decision trees (Breiman, 2001).

Importance was calculated using the feature_importances_ attribute of the trained model. which returns normalized values plus 1. Higher values indicate a greater influence of the variable. in forecasts. This analysis is fundamental for:

- Identify key drivers of regional economic growth
- To guide public policies by prioritizing variables with the greatest impact.
- Validate theoretical assumptions about economic growth.

Rodrigues et al. (2022) highlight that the importance analysis of variables in models of Machine learning is a valuable tool for informing regional development policies.

3.6 Generating Prospective Scenarios

To transform the predictive model into a planning support tool, the following steps were taken: Three prospective economic growth scenarios were developed for a three-year horizon: Conservative, Moderate, and Optimistic. Each scenario incorporates projected variations in the main... socioeconomic indicators (HDI, unemployment rate, infrastructure score and growth) population), defined based on historical trends observed in the region (IMEA, 2023; IBGE, 2022), targets established in the State's Multi-Year Plan (SEPLAN-MT, 2023) and benchmarks of regions with similar characteristics.

Based on these projected values, the Random Forest model was applied to estimate GDP. future per capita growth of each municipality. The estimated growth percentage was calculated by

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Expression: Growth (%) = (Projected GDP - Current GDP) / Current GDP × 100.

This strategy follows the established methodology for scenario analysis, which, according to Provost and Fawcett (2013) allow us to assess the sensitivity of the results in the face of different future conditions and to support more robust decision-making processes.

3.7 Generating Visualizations and Reports

The analytical process was documented through various graphical visualizations and compiled into a technical report in PDF format. The representations included bar charts. horizontals, correlation matrix, scatter plots, comparisons of prospective scenarios and Graph showing the importance of variables in the Random Forest model. The visualizations were produced... using Matplotlib and Seaborn, following the principles of clarity and simplicity recommended by Tufte. (2001) and Few (2012).

The report was organized to meet the needs of public managers, researchers, and representatives of private sector, considering that, according to Mayer-Schönberger and Cukier (2013), clear communication The analytical results are just as relevant as the technical quality of the models.

3.8 Computational Tools

All analysis was implemented in Python (version 3.8+), using the following... main libraries:

- **Pandas** (McKinney, 2012): manipulation and analysis of tabular data
- **NumPy:** numerical operations and linear algebra
- **Scikit-learn** (Pedregosa et al., 2011): implementation of machine learning algorithms
- **Matplotlib and Seaborn:** data visualization
- **Matplotlib PdfPages:** generating PDF reports

The code was structured in object-oriented classes to ensure modularity. Reusability and maintainability. All source code has been documented following PEP 8 standards. (Python Enhancement Proposal) for readability.

3.9 Methodological Limitations and Mitigations

This research has methodological limitations that should be considered in... Interpretation of the results. The small sample size, consisting of only ten observations, limits the statistical power of the analyses; this limitation was mitigated by conservative techniques and by

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Cautious interpretation of the findings is necessary. Furthermore, the data used are cross-sectional, not... capturing temporal dynamics, which motivated the historical contextualization in the literature review and The recommendation for future longitudinal studies. The reliance on secondary data from Information from official sources may involve time lags, which are mitigated by triangulation of multiple data. bases. Furthermore, relevant variables such as social capital, local governance, and environmental factors are not were included in the model, and their incorporation into subsequent research was suggested. As As Gil (2010) points out, the explicit recognition of methodological limitations is fundamental for the transparency and scientific rigor, allowing the reader to assess the robustness and generalizability of the results.

3.10 Ethical Aspects

The research was conducted respecting the ethical principles of scientific research. All data The information used is in the public domain, made available by official bodies, and there are no questions of Privacy or confidentiality. The presentation of the results was carried out impartially. without biases that could favor or harm specific municipalities.

The commitment to transparency and scientific reproducibility guides the entire methodology: Documented source code, detailed description of procedures, and availability of datasets. generated. As Kitchin (2014) argues, methodological transparency is fundamental to the The credibility of data science applied to public policy.

3.11 Actual Data, Estimated Data, and Simulated Data

This research predominantly uses data from high-level official sources. reliability, such as the Brazilian Institute of Geography and Statistics (IBGE), the Institute of Applied Economic Research (IPEA), the Mato Grosso Institute of Agricultural Economics (IMEA) and documents from SEPLAN-MT. These databases provided parameters, ranges, and references. fundamental statistics on population, Gross Domestic Product (GDP), educational indicators and socioeconomic data from the municipalities of the Sepotuba Valley, ensuring methodological consistency and theoretical basis.

However, due to the absence of some updated historical series, and the time lag... due to specific databases and the complete unavailability of municipal data for all indicators. For the desired results, some of the values used in the predictive modeling needed to be approximated and simulated. in the Python environment. This simulation was performed based on real intervals and proportions. observed in official sources, maintaining statistical consistency and fidelity to regional trends.

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

The simulated values included:

- Total municipal GDP (in thousands of reais);
- Estimated population for some municipalities;
- Complementary socioeconomic indicators (unemployment rate, average schooling, infrastructure and economic diversification).

This data was generated in the algorithm using dedicated functions, which produced plausible values that are compatible with the socioeconomic reality of the region, allowing for execution. of the Linear Regression and Random Forest models. The use of simulated data does not compromise the The analytical objective of the study, since the research is characterized as exploratory, with an emphasis on validation of the methodology and demonstration of the applicability of machine learning techniques to regional context. However, it is essential to highlight this methodological choice to ensure scientific transparency, as recommended by Minayo (2001) and Gil (2010).

Thus, modeling should be understood as a simulation based on real data, and not as an exact reproduction of official statistics. In future studies, replacement is recommended. based on estimated values from fully up-to-date data collected directly from the platforms. governmental.

4. Results obtained and their implications.

4.1 Socioeconomic Characterization of the Municipalities of the Sepotuba Valley

Descriptive analysis of the data revealed significant socioeconomic heterogeneity. among the municipalities of the Sepotuba Valley, highlighting the complexity of development. regional. Table 1 presents a summary of the main socioeconomic indicators for the ten municipalities analyzed.

The results reveal significant disparities between the municipalities analyzed. Sapezal It has a GDP per capita approximately six times higher than that of Tangará da Serra, although the latter concentrates 43.7% of the regional population. This configuration expresses what Brandão (2007) He calls it the "paradox of agro-industrial productivity," in which Municipalities highly specialized in commodities exhibit high GDP per capita, but They generate relatively few direct jobs due to the high level of mechanization.

The variation in the HDI between 0.628 and 0.780 places all municipalities in the range of "Medium human development" according to the UNDP classification, although with differences. expressive. Swapezal, Campo Novo do Parecis and Tangará da Serra are located above 0.720, while Porto Estrela, Salto do Céu and Rio Branco remain below 0.645. According

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Sen (2000) argues that these distinctions reflect inequalities in basic human capabilities.

especially in education and income.

Table 1 – Socioeconomic Indicators of the Municipalities of the Sepotuba Valley

Municipality	**Population**	**GDP Total (thousands) R$)**	**GDP per capita (years) (R$)**	**HDI Rate**	**Of- without- nail (%)**	**Education**	**Infrastructure structure (0-100)**	**Diversification (0100)**
Sapezal	24,000	4,100,000	170,833.33	0.780	3.9	3,200,000	9.8	88	52
Field New from Pare- cis	32,000	100,000.00	0.750	4.8			9.1	82	58
Denise	9,500	520,000	54,736.84	0.668	8.2	290,000	**7 2**	48	42
Lambari West	6,200	46,774.19	0.652	9.1			6.8	42	38
New Olympia	15,000	680,000	45,333.33	0.701	7.0		7.9	58	50
Jump of Sky	4,500	150,000	33,333.33	0.635	10 2	6.2		35	32
River White	5,800	180,000	31,034.48	0.643	9.5		6.5	38	35
Harbor Star	3,200	95,000	29,687.50	0.628	10.8	6.0		32	30
Barra do Bugres	35,000	980,000	28,000.00	0.692	7.5		7.8	62	55
Tangará from the mountain	105,000	2,850,000	27,142.86	0.729	6.2		8.5	75	65
Average Regional	**24.020**	**1,104,500**	**56,687.59**	**0.688**	**7.7**		**7.5**	**64**	**45.7**
Detour- Standard	**30.358**	**1,408,984**	**47,686.91**	**0.053**	**2.21**		**1.39**	**21.40**	**12.02**

Source: Prepared by the author based on IBGE (2022), IPEA (2023) and IMEA (2023).

The average regional unemployment rate (7.72%) slightly exceeds the state average for Mato Grosso.

Grosso (7.1%, IBGE, 2022) and shows strong internal variation, fluctuating between 3.9% in Sapezal and

10.8% in Porto Estrela. As demonstrated by Delgado (2012), intensive agricultural modernization

This can increase unemployment in smaller municipalities due to the replacement of labor.

through technological capital.

4.2 Correlation Analysis between Variables

The Pearson correlation matrix (Table 2) revealed statistically significant relationships.

between socioeconomic indicators and GDP per capita.

Table 2 – Pearson Correlation Matrix between Socioeconomic Variables

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Variables: GDP Per Capita, HDI, School Enrollment Rate, Infrastructure, Diversification, Population Unemployment							
GDP per capita	1.00	0.75	-0.75	0.75	0.69	0.30	-0.05
HDI	0.75	1.00	-0.99	1.00	0.99	0.84	0.54
Rate Unemployment	-0.75	1.00	-0.99	-1.00	-0.99	-0.85	-0.50
Education	0.75	1.00	-1.00	1.00	0.99	0.85	0.53
Infrastructure	0.69	0.99	-0.99	0.99	1.00	0.89	0.61
Diversification	0.30	0.84	-0.85	0.85	0.89	1.00	0.82
Population	-0.05	0.54	-0.50	0.53	0.61	0.82	1.00

Source: Prepared by the author

Note: Correlations with |r| > 0.70 are considered strong; 0.40 < |r| < 0.70 are moderate; |r| Values < 0.40 are weak (Cohen, 1988).

4.2.1 Main Relationships Identified

Unemployment Rate and GDP per capita (r = -0.75): The strong negative correlation confirms Romer's (1990) findings on the relationship between employment, productivity, and economic growth. Municipalities with lower unemployment rates have higher GDP per capita, suggesting that policies Job creation is crucial for regional development. Silva, Teles and Araújo (2025) They identified a similar correlation (r = -0.68) in a study on territorial planning with machines. learning.

HDI and GDP per capita (r = 0.75): The strong positive correlation aligns with the perspective. development as freedom by Sen (2000), indicating that economic growth is associated with the expansion of human capabilities. Aguiar and Baptista (2023), in an analysis of growth In urban studies, they found a comparable correlation (r = 0.71) between human development and indicators. economic.

Education and GDP per capita (r = 0.75): Confirms Romer's human capital theory. (1990), demonstrating that investments in education are fundamental determinants of Growth. With each additional year of average schooling, a tendency towards increased GDP is observed. per capita, reinforcing the importance of educational policies.

Infrastructure and GDP per capita (r = 0.69): The moderate-strong correlation highlights the role of infrastructure as a facilitator of development, as argued by Sachs (2004). Municipalities with better road, energy, and digital infrastructure show greater dynamism. economic.

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Population and GDP per capita (r = -0.05): The practically zero correlation is significant: More populous municipalities do not necessarily have a higher GDP per capita. This result This contradicts simplistic assumptions of "population economies of scale" and suggests that the quality of capital... Humans outnumber the human population. Rodrigues et al. (2022) identified a similar pattern in Analysis of regional development using machine learning.

Economic Diversification and GDP per capita (r = 0.30): The weak correlation is surprising, since Economic literature generally associates diversification with greater resilience and growth (Veiga, 2008). The likely explanation lies in successful productive specialization: Sapezal and Campo Novo do Parecis, despite less diversification, achieves a high GDP per capita through specialization in high value-added commodities.

4.2.2 Multicollinearity between Independent Variables

A high correlation is observed between some independent variables (HDI, rate of Unemployment and education levels show correlations > 0.99 with each other, characterizing multicollinearity. According to Wooldridge (2019), this phenomenon can inflate standard errors in models. linear regression is problematic, but it's less problematic for machine learning algorithms like Random. Forest, which naturally handles correlated variables through the bootstrap process and Random selection of features.

4.3 Performance of Predictive Models

Two machine learning models were trained and evaluated: Linear Regression and Random Forest. Table 3 presents comparative performance metrics. Note: R^2 varies from -ÿ to 1 (values close to 1 indicate a better fit); lower RMSE and MAE indicate less error. forecast. Furthermore, the combined analysis of these metrics allows us to identify not only the model with better predictive ability, but also possible patterns of underfitting or overfitting. This Evaluation is essential to ensure the robustness of forecasts and the reliability of results. from different socioeconomic scenarios. With this, it becomes possible to select the most suitable algorithm. appropriate to the characteristics of the data and the objectives of the study.

Table 3 - Performance Metrics of Predictive Models

Model	R^2 Score	RMSE (R$) MAE (R$) 5384,573.21	Interpretation
Regression Linear	-2.00	171,246.88	Inadequate performance; model does not capture complex relationships

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Random Forest 0.95	810,582.46		7,891.33	Excellent performance; high predictive capacity

Source: Prepared by the author.

4.3.1 Analysis of the Linear Regression Model

Linear regression showed unsatisfactory performance, with a negative R^2 (-2.0053). indicating that the model was unable to adequately represent the relationships between the variables. RMSE (R$ 84,573.21) and MAE (R$ 71,246.88) reveal predictive errors. high values. This result can be explained by the presence of non-linear relationships between the socioeconomic indicators, due to the reduced sample size, and due to multicollinearity among The independent variables and the high heterogeneity among the municipalities. These findings Recent studies corroborate this, demonstrating the limitations of linear models in regional contexts. complex and heterogeneous.

4.3.2 Analysis of the Random Forest Model

Random Forest performed significantly better, with an R^2 of 0.9528. indicating that the model explains 95.28% of the variance in GDP per capita. The RMSE values (R$ 10,582.46) and MAE (R$ 7,891.33) were substantially lower than those obtained by Regression. Linear, representing errors consistent with applications in strategic planning. In comparison with similar studies by Aguiar and Baptista (2023), Silva, Teles and Araújo (2025) and Rodrigues et al. (2022), who obtained R^2 values between 0.82 and 0.89, showed that the present study achieved a performance... superior, with an R^2 of 0.95. This result can be attributed to the well-founded selection of variables, to consistency of the data used and proper adjustment of the model's hyperparameters.

4.3.3 Validation and Robustness of Results

Despite the high R^2, the small sample size, with only three observations in the set, is a significant drawback. Testing poses a risk of overfitting. To mitigate this limitation, strategies were adopted. conservative approaches, such as restricting the maximum depth of the Random Forest (max_depth = 5), the use of 100 trees to ensure stability and the cautious interpretation of the results, treated as trends and not as deterministic predictions. As Kuhn and Johnson (2013) argue, in Small datasets are essential for regularization and validation practices. The convergence between the quantitative performance of the model and theoretical assumptions, such as the relevance of the rate of

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

unemployment reinforces the credibility of the estimates produced.

4.4 Relative Importance of Variables

One of the main contributions of Random Forest is the measurement of relative importance.

The results of each variable were used for the predictions. Table 4 and Figure 1 present these results.

Table 4 - Importance of Variables in the Random Forest Model

Variable Ranking		Importance Relative	Interpretation
1st	Unemployment rate	0.400 (40.0%)	Most decisive factor of economic growth
2nd	Infrastructure	0.215 (21.5%)	Second biggest influence; facilitates economic activity
3rd	HDI	0.154 (15.4%)	Human development as basis of growth
4th	Secondary Education	0.123 (12.3%)	Human capital is crucial. moderate to high
5th	Population	0.056 (5.6%)	Smaller impact than variables qualitative
6th	Economic Diversification	0.052 (5.2%)	Lower relative importance in regional context

Source: Prepared by the author.

4.4.1 Unemployment Rate as the Main Determinant (40.0%)

The unemployment rate was the most relevant variable in the model, accounting for 40% of the...

predictive capacity. This result reinforces, from a theoretical point of view, the importance of full use.

of the productive factors, as proposed by Romer (1990), indicating that the Sepotuba Valley

It is still operating below its employment potential. In practical terms, the need for...

Policies aimed at job creation, professional training, and encouraging entrepreneurship.

Recent literature presents similar results, although with slightly lower values,

which indicates that, in the rural-agro-industrial context of the region, employment plays an even more central role.

for economic performance.

4.4.2 Infrastructure as an Enabler (21.5%)

Infrastructure contributes 21.5% to predictive capacity, confirming its role as

essential facilitator of development. This finding aligns with IMEA (2021), which identified

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Infrastructure bottlenecks as the main limiting factors for regional competitiveness.

Infrastructure operates as "fixed social capital" (Marx, 1867) that enables the circulation of Goods, people, and information. In the context of the Sepotuba Valley, characterized by long distances. From roads all the way to export ports, the quality of highways directly impacts logistics costs and... profitability of agricultural production.

According to Sachs (2004), investments in infrastructure generate multiplier effects. benefiting multiple sectors simultaneously. In the Sepotuba Valley, the paving of highways It not only reduces transportation costs, but also facilitates access to health, education, and services. consumer markets.

4.4.3 HDI and Education (15.4% and 12.3%)

HDI and education together account for 27.7% of the predictive capacity of model, highlighting the central role of human development, in line with the approach Sen (2000). Average schooling influences growth through multiple channels, such as higher productivity, adoption of technologies, and more qualified citizen participation, as argued. by Romer (1990). In the Sepotuba Valley, the variation in average schooling between 6.0 and 9.8 years indicates Educational investments tend to generate significant returns, especially in municipalities with... lower levels of education.

4.4.4 Population and Economic Diversification (5.6% and 5.2%)

Population and economic diversification were of low relative importance, totaling only 10.8% of the model's predictive capacity. The population variable (5.6%) showed influence. limited, indicating that population size does not determine economic performance, which This confirms the findings of Rodrigues et al. (2022). Economic diversification (5.2%) also had an effect. reduced, possibly because, in the current phase of development of the Sepotuba Valley, municipalities Companies specializing in high-value commodities, such as Sapezal and Campo Novo do Parecis, achieve high GDP per capita even with low diversification. Despite this, as Veiga (2008) warns, Diversification remains relevant for long-term economic resilience.

4.5 Economic Growth Forecasts by Scenario

The Random Forest model was applied to generate GDP per capita growth forecasts. in three prospective scenarios (conservative, moderate, and optimistic) for a three-year horizon. A

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025
Table 5 summarizes the results.

Table 5 - Economic Growth Forecasts by Municipality and Scenario (3 years)

Source: Prepared by the authors.

4.5.1 Analysis of the Conservative Scenario

In the conservative scenario, characterized by modest improvements in the indicators.
From a socioeconomic perspective, the model projects an average reduction of 3.88% in regional GDP per capita.
The result stems from two main factors: regression to the mean, which takes municipalities with GDP per capita...
very high per capita costs converging towards values closer to the regional average, and limitations of
The model faces small variations, which may underestimate the competitive advantages of municipalities.
highly specialized. Still, some municipalities would show significant growth.
Even in this scenario, such as Tangará da Serra (+61.82%), Barra do Bugres (+19.07%), Campo Novo
Parecis (+19.79%) and Rio Branco (+16.44%). These municipalities share unemployment rates.
moderate to high and intermediate levels of infrastructure, conditions that favor gains
resulting from improved incremental improvements.

4.5.2 Moderate Scenario Analysis

In the moderate scenario, with intermediate improvements in socioeconomic indicators, the
The model projects a smaller average reduction in regional GDP per capita (-2.92%), although some
Municipalities show significant growth. Tangará da Serra stands out as the municipality with...
greatest potential for expansion (+66.10%), followed by Rio Branco (+29.23%), Campo Novo do Parecis
(+19.79%) and Barra do Bugres (+14.82%). Tangará da Serra's strong performance stems from its
large population, which can generate significant gains when better employed and qualified, of
its intermediate infrastructure and its high degree of economic diversification. These results
These findings are in line with those of Goes and Goes (2025), according to whom municipalities with economies

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Diversified infrastructure and intermediate levels tend to respond more intensely to incremental improvements.

Analysis of the Optimistic Scenario

In the optimistic scenario, which considers more robust improvements in socioeconomic indicators, The model projects average regional growth of 6.32%. Tangará da Serra stands out (+208.75%), Rio Branco (+40.23%), Barra do Bugres (+22.84%) and Campo Novo do Parecis (+19.79%). The The exceptional performance projected for Tangará da Serra suggests that the municipality is operating below its potential, especially due to the unemployment rate and the underutilization of its productive structure. Still, increases exceeding 200% in the short term are unlikely and would depend on... profound structural transformations, private investments, and advancements such as industrial expansion. A substantial increase in significant investments in education and skilled services. The model also predicts... retraction in already highly developed municipalities, such as Sapezal, Denise and Nova Olímpia, reflects with GDP per capita values well above the regional average and the statistical trend of convergence. However, in practice, these municipalities can maintain their high levels through gains.

Continuous efficiency improvements in agribusiness.

Comparison with Similar Studies

A comparison with regional growth studies shows that the projections in this study... They present a greater range. While Silva, Teles and Araújo (2025) estimated growth of 5% to 12% in regions benefiting from infrastructure interventions, and Aguiar and Baptista (2023) They identified growth between 8% and 15% in optimistic urban scenarios; the present study projected... Average growth of 6.32% in the optimistic scenario, with variations ranging from -29.88% to +208.75%. This greater amplitude reflects the high heterogeneity of the Sepotuba Valley, which brings together both highly developed agro-industrial municipalities as well as peripheral rural localities, resulting in substantially different growth patterns.

Identifying Priority Municipalities for Public Policies

Based on the results of the models and prospective analyses, it is possible to categorize the municipalities. according to needs and potential (Table 6).

Table 6 - Categorization of Municipalities for Public Policies

Category	Municipalities C	haracteristics	Priority Strategies

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

High Priority (Intervention Urgent)	Porto Jump of Sky, River 0.645 White	- GDP per capita < R$ Estrela, 33,000 - HDI < Unemployment > 9.5% Infrastructure < 40	- Massive investments in infrastructure Emergency programs job Expansion of basic services
Medium Priority (Development)	Lambari D'Oeste, Barra do Bugres, New Olympia	- GDP per capita R$ 28,00046,000 - HDI 0.650-0.701 Unemployment: 7.0-9.1%	- Professional qualification - Improvement of rural roads - Support for economic diversification
Consolidation	Tangará da Mountain range, Denise	- Variable GDP per capita - Great potential of growth - Population base reasonable	- Attracting investments private Technology parks Educational expansion
Maintenance Excellence	Sapezal, Field New from Parecis	- GDP per capita > R$ 100,000 - HDI > 0.750 Unemployment < 5%	Agro-industrial innovation Environmental sustainability Social responsibility

Source: Prepared by the authors.

High Priority Municipalities

Porto Estrela, Salto do Céu, and Rio Branco exhibit the most critical socioeconomic indicators of...
Sepotuba Valley, characterized by its "blind spots"
regional development. These municipalities have poor infrastructure and high rates of
Unemployment, low levels of education, and small population size are conditions that hinder...
generating economic dynamism. This pattern reflects the "peripheral trap".
described by Delgado (2012), according to which small rural localities do not benefit
of regional agro-industrial advancement.
Given this scenario, it is recommended to prioritize actions aimed at strengthening infrastructure.
basic infrastructure, job creation through essential works and support for family farming, expansion of
Educational offerings with a technical focus and partnerships with educational institutions, in addition to coordination.
regional distribution via intermunicipal consortia for sharing strategic public services.

Tangará da Serra as a Special Case

Tangará da Serra has the lowest GDP per capita in the region, but the greatest potential for...
Growth in all projected scenarios. This result stems from the presence of structural assets.
important factors, such as a large population base and its role as a regional hub for education and services,

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Intermediate infrastructure and high economic diversification. Current performance below Potential is primarily related to the unemployment rate and the underutilization of installed capacities. As Brandão (2007) argues, hub municipalities tend to exhibit delayed growth. due to the predominance of administrative and service activities with lower productivity. For To take advantage of its potential, the implementation of an industrial park and the strengthening of the sector are recommended. of the knowledge economy, the development of rural tourism and broad programs of professional qualification.

Implications for Public Policy

The results obtained have direct implications for the formulation of public policies. regional development. Table 7 summarizes recommendations by strategic axis. These guidelines They allow for aligning investments, guiding municipal priorities, and strengthening capacity. institutional framework for reducing territorial inequalities. Furthermore, they provide subsidies for... More assertive decisions in medium-term economic planning. Finally, the recommendations. These proposals contribute to maximizing the impact of government actions and improving governance. regional.

Table 7 - Public Policy Recommendations by Strategic Axis

Axle	Foundation Empirical	Policies Recommended	Municipalities Priorities	Expected Impact
Employment and Income	Rate unemployment is the most important factor important (40%)	Qualification professional, formalization, in- entrepreneurship	Porto Estrela, Sky Leap, Rio Branco, Denise	2030% reduction in unemployment
Infrastructure	Second largest importance (21.5%)	Paving, broadband, energy renewable	Porto Estrela, Rio Branco, Sky Leap, Lambari D'Oeste	Elevation of 15 25 points
Education	HDI+ education level = 27.7%	Technical education, UNE partnerships MAT/IFs, literacy	Porto Estrela, Sky Leap, White River	Increase of 1-2 years of schooling
Diversification Economic	Minor importance (5.2%) but strategic	Agribusiness, tourism, economy creative	All	Elevation of 1020 points
Articulation Regional	Necessary for scale	Consortia, Regional Plan, Observatory Socioeconomic	All	Cost reduction 15-30%

Source: Prepared by the author.

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Employment and Income Axis

The evidence that the unemployment rate is the most decisive factor in the model (40%) indicates the need to reorient regional public policies. Instead of prioritizing exclusively infrastructure and attracting large investments, the results indicate that actions aimed at generating Job creation and professional training tend to produce higher returns. Among the measures Among the recommended measures, the implementation of a regional professional training program stands out. incentives for formalizing employment, creation of an entrepreneurship incubator, and initiatives for... work in infrastructure. As Sen (2000) argues, employment contributes simultaneously for income, dignity and capacity building, reinforcing its centrality in human development.

Infrastructure Axis

The importance attributed to infrastructure by the model (21.5%) confirms previous diagnoses. about the logistical bottlenecks in the Sepotuba Valley and allows for the identification of specific priorities. The The municipalities of Porto Estrela, Salto do Céu, Rio Branco, and Lambari D'Oeste should be prioritized because They have the lowest infrastructure levels. The most critical needs include improvement. of secondary roads, expansion of broadband internet connectivity, strengthening of Energy supply from renewable sources and expansion of agricultural storage capacity. In line with Sachs (2004), infrastructure acts as a public good that generates externalities. Broad positive outcomes justifying public investments capable of stimulating development. regionally in a diffuse manner.

Education and Human Capital Axis

The combined contribution of HDI and schooling (27.7%) reinforces the central role of education. in regional development, aligning with endogenous growth theories. In the Valley of In Sepotuba, the strong educational heterogeneity indicates that interventions should be territorially focused. differentiated. Municipalities with less than 7 years of schooling require a focus on universalizing the primary education and reduced dropout rates; those in the 7-8 age range require expanded education. secondary and technical training; municipalities with more than 8.5 years of experience need strengthening of secondary education. Superior education, applied research, and advanced qualifications. Partnerships with UNEMAT, Federal Institutes, SENAR and SENAI would expand educational offerings throughout the region. As Veiga (2008) argues, Education is a structural driver of development, enhancing human capabilities, productivity, and... capacity for innovation.

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Economic Diversification Axis

Although economic diversification was of less relative importance (5.2%), its Relevance to regional resilience remains fundamental. Highly specialized municipalities In agricultural commodities, regions like Sapezal and Campo Novo do Parecis tend to exhibit high... Vulnerability to external shocks, including international price fluctuations and weather events. extreme and phytosanitary risks. In this context, diversification policies play a role. Strategic in reducing sectoral dependencies and expanding income-generating opportunities.

The main strategies identified include: (1) incentives for value-added agro-industry aggregate, promoting the local processing of raw materials and capturing a larger share of production chain; (2) strengthening family farming and agroecological systems, expanding access to differentiated and more profitable markets; (3) development of rural tourism, integrating natural resources, cultural heritage and productive activities typical of the region; (4) Promoting the creative economy by valuing handicrafts, gastronomy, and... cultural manifestations with economic potential; and (5) expansion of

specialized services, especially in technical consulting and technological solutions. applied to agribusiness.

As Veiga (2008) points out, economic diversification does not imply abandoning the consolidated regional vocations, but the expansion of the set of complementary activities capable to absorb labor, increase socioeconomic resilience, and reduce structural vulnerabilities. associated with excessive specialization.

Regional Coordination and Governance Axis

The strong socioeconomic heterogeneity of the Sepotuba Valley makes policies insufficient. Isolated by municipality. The formation of intermunicipal consortia is essential to overcome limitations. scalability, allowing for the shared provision of specialized services and greater efficiency in Public investments. Regional coordination also strengthens negotiating power for fundraising and attracting private investment. The creation of a [missing word - likely "facility"] is also recommended. Socioeconomic Observatory for continuous monitoring of indicators and greater transparency. According to Almeida et al. (2015), coordinated policies are more effective in interdependent regions. Strengthening centers like Tangará da Serra benefits the entire region.

Limitations of the Study and Agenda for Future Research

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Limitations Identified

The limitations of this study are concentrated in four main aspects. First, the size. The small sample size, comprising only 10 municipalities, limits statistical power and generalization. of the results. Second, the use of cross-sectional data prevents capturing temporal dynamics, the which indicates the need for future longitudinal analyses. Third, relevant variables such as Governance, environmental factors, regional connectivity, and technological innovation were not included. which may limit the explanatory scope of the models. Finally, the identified associations do not They establish causality, and the projections depend on assumptions that can be modified by Unforeseen events. Such limitations reinforce the importance of a cautious interpretation of the results. and complementary research.

Future Research Agenda

The results of this study indicate several promising directions for future research. Performing longitudinal analyses with extensive historical series would make it possible to identify trends. long-term factors, regional economic cycles, and lagged effects of public policies, according to discussed by Silva, Teles and Araújo (2025). The expansion of the set of variables, incorporating environmental aspects, governance indicators, technological innovation metrics and measures of Spatial connectivity could increase the explanatory power of the models. The application of Advanced machine learning methods, such as deep neural networks and ensemble algorithms. modern (XGBoost, LightGBM, CatBoost) and interpretability techniques such as SHAP and LIME, This would allow for capturing complex interactions and providing more robust explanations. Spatial analyses by through spatial econometrics models, identification of territorial clusters and evaluation of Spillovers would contribute to understanding interdependencies between municipalities. Approaches qualitative methods, including interviews with public managers, focus groups, and case studies, This would deepen the understanding of institutional and socioeconomic dynamics. Future research They can also employ impact assessment methods and complex systems modeling for Testing interventions and simulating development scenarios. As Silva, Teles and Araújo argue. (2025), the application of artificial intelligence to territorial planning is still being consolidated. in Brazil, making it essential to continue studies that improve methods and strengthen the analytical capacity of public managers.

It presents the conclusions and recommendations for public policies and future research.

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Summary of Key Findings

This research aimed to develop and apply a predictive growth model.
economically viable for the municipalities of the Sepotuba Valley, using machine learning techniques.
and official socioeconomic data. The investigation answered the central question: *how to apply techniques*
Machine learning to predict the economic growth of municipalities in the Valley of
Sepotuba based on available socioeconomic data, in order to support public policies and
Strategic decisions of the private sector?

The results achieved confirm that predictive models based on machine learning,
Specifically, Random Forest tools are effective for regional development analysis.
overcoming traditional statistical methods in contexts marked by complex and non-conforming relationships
linear relationships between variables.

Regional Socioeconomic Heterogeneity

Descriptive analysis revealed significant heterogeneity among the ten municipalities studied.
The GDP per capita varies from R$ 27,142.86 (Tangará da Serra) to R$ 170,833.33 (Sapezal), a range of
6.3 times. The Municipal Human Development Index (MHDI) ranges from 0.628 (Porto
Estrela) and 0.780 (Sapezal), while the unemployment rate varies from 3.9% (Sapezal) to 10.8% (Porto).
Star).

This heterogeneity reflects what Brandão (2007) characterizes as "development
"Unequal combined": highly capitalized agro-industrial hubs integrated into global supply chains.
(Sapezal, Campo Novo do Parecis) coexist with smaller-scale and limited peripheral municipalities.
access to resources (Porto Estrela, Salto do Céu, Rio Branco). As Delgado (2012) argues, the
Agricultural modernization does not spread homogeneously across the territory, resulting in concentration.
spatial distribution of income and opportunities.

Correlations between Socioeconomic Indicators

Correlation analysis identified statistically significant relationships between variables.
independent and GDP per capita:
Unemployment rate: strongly negative correlation (r = -0.75), confirming that markets
Dynamic jobs are essential for economic growth.
HDI: strong positive correlation (r = 0.75), validating the perspective of development as
expansion of human capabilities (Sen, 2000)

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Average schooling: strong positive correlation (r = 0.75), corroborating human capital theories. (Romer, 1990)

Infrastructure: moderate-strong correlation (r = 0.69), highlighting its facilitating role in development (Sachs, 2004)

These correlations not only confirm established theoretical assumptions, but also indicate interventions in these variables would have measurable impacts on regional economic growth.

Superiority of Random Forest over Linear Regression

The comparison between models showed substantially superior performance for Random. Forest compared to Linear Regression. While the linear model showed a negative R^2 and error high, indicating an inability to capture the relationships present in the data, the Random Forest reached An R^2 of 0.9528 and an RMSE of R$ 10,582.46 demonstrate significantly superior predictive capacity. This The result is consistent with studies such as Aguiar and Baptista (2023), Silva, Teles and Araújo (2025) and Rodrigues et al. (2022), who also identified high performance of Random Forest in Territorial and socioeconomic analyses. The inadequacy of Linear Regression reinforces arguments of Hastie, Tibshirani, and Friedman (2009) argue that linear models are limited in contexts with Complex interactions and non-linear relationships. In the Sepotuba Valley, variables such as education, Employment and infrastructure interact in a dynamic and non-linear way, justifying the superiority of... tree-based models.

Determinants of Regional Economic Growth

The importance analysis of the variables in the Random Forest model revealed that the growth The regional economy is strongly influenced by factors associated with labor force utilization. and to structural conditions. The unemployment rate made the largest contribution to the forecasts. corresponding to 40 percent of the total importance, indicating that the absorption of labor Available infrastructure is one of the main drivers of development. Infrastructure accounted for 21.5% of one hundred, reinforcing that logistical limitations continue to be significant obstacles to expansion. Economic variables, as well as variables associated with human development, have also proven important. with the HDI contributing 15.4 percent and average schooling 12.3 percent, highlighting the role of skilled human capital. Population and economic diversification. They had less influence in the short term, with 5.6 percent and 5.2 percent, respectively.

The results highlight that policies focused on employment, professional training and Productive inclusion can generate superior returns, aligning with Sen's (2000) perspective.

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

according to which employment plays an essential role in the development of human capabilities. Although economic diversification was shown to have reduced importance in the model, its relevance Strategic considerations remain significant for long-term resilience, as discussed by Veiga. (2008). Thus, strategies of productive specialization can boost growth in the future. three to five years, but they must be accompanied by structural policies aimed at diversification. sustainable.

Growth Forecasts by Scenario

The three-year projections reveal contrasting behaviors between the municipalities analyzed. In the conservative scenario, characterized by modest improvements in Based on socioeconomic indicators, the region shows an average contraction of 3.88 percent, although municipalities such as Tangará da Serra, Barra do Bugres and Campo Novo do Parecis present significant growth. In the moderate scenario, the average contraction decreases to 2.92 percent and Tangará da Serra and Rio Branco stand out, exhibiting high potential for expansion. In this context... Optimistically, the region achieved a positive average growth of 6.32 percent, with Tangará standing out. from Serra, whose projected increase of 208.75 percent indicates significant underutilization of its structural and demographic potential.

The projected negative performance for Sapezal in all scenarios reflects the statistical effect. regression to the mean, considering its exceptionally high GDP per capita. This result suggests a trend of intraregional convergence, without necessarily implying economic decline. absolute, but indicating a gradual convergence between municipalities in terms of development. economic.

Theoretical and Methodological Contributions

This research offers relevant theoretical and methodological contributions to the field of Studies on regional economic growth and the application of machine learning in social sciences. applied.

Empirical Validation of Endogenous Growth Theories

The results empirically validate central assumptions of growth theories. endogenous (Romer, 1990; Lucas, 1988). The importance of schooling (12.3%) and HDI (15.4%) confirms that human capital is a fundamental determinant of growth. The relevance of

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Infrastructure (21.5%) corroborates models that emphasize complementarities between physical capital and human.

However, the study also reveals specificities of the regional context that nuance theories. universalists. The low importance of population (5.6%) contradicts models that emphasize economies. demographic scale, suggesting that quality trumps quantity in the development of regions. peripheral. As Veiga (2008) argues, development is not a linear and universal process, but Yes, a specific trajectory shaped by historical, institutional, and territorial contexts.

Demonstration of the Applicability of Machine Learning in Regional Contexts

Research shows that machine learning techniques, specifically Random Forest, are applicable and effective in regional planning contexts, even with relatively small samples. small (n=10). The R^2 of 0.9528 surpasses performances reported in similar studies (Aguiar; Baptista, 2023; Silva; Teles; Araújo, 2025; Rodrigues et al., 2022), suggesting that the combination of:

Solid theoretical foundation in variable selection.

high-quality official data

Proper hyperparameter adjustment (depth limitation, appropriate number of trees)

Cautious interpretation considering sampling limitations.

Where it allows the development of robust and interpretable predictive models.

As Provost and Fawcett (2013) argue, the success of data science projects does not It doesn't reside solely in algorithmic sophistication, but in a deep understanding of the application domain. This research exemplifies how theoretical knowledge in regional economics can inform choices. Methodological advancements in machine learning, generating superior results.

Proposal for an Integrated Framework for Regional Analysis

This study proposes an integrated framework that can be replicated in other regions:

Stage 1 - Theoretical Framework: Literature review on economic growth regional, identifying theoretically grounded determinants

Step 2 - Data Collection: Obtaining official data (IBGE, IPEA, institutes) state-level) for identified variables

Stage 3 - Exploratory Analysis: Descriptive statistics, correlation analysis, identification of outliers and patterns

Step 4 - Predictive Modeling: Application of multiple algorithms (Linear Regression as

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

baseline, Random Forest for capturing non-linearities, possibly XGBoost or neural networks)

Step 5 - Importance Analysis: Measuring the relative contribution of each variable.
guiding policy prioritization

Step 6 - Prospective Scenarios: Generating forecasts under different assumptions.
supporting strategic planning

Step 7 - Recommendations: Translating technical findings into concrete recommendations for public managers and the private sector

This framework aligns with Kitchin's (2014) recommendations on responsible practices.
in data science for public policy, emphasizing methodological transparency and recognition.
limitations and accessible communication of results.

Integration between Quantitative Methods and Contextual Knowledge

The research shows that the use of advanced quantitative methods, such as techniques of Machine learning does not replace contextual knowledge, but complements it by expanding the...
ability to interpret the results. The empirical analysis was articulated with reference frameworks.
qualitative data from the literature on the economic history of Mato Grosso (Oliveira, 2007),
from studies on agricultural modernization (Delgado, 2012), from the socio-environmental specificities of Sepotuba Basin (Leandro, 2020; Gamero et al., 2020) and regional development theories
(Brandão, 2007; Veiga, 2008). In line with Minayo (2001), the integration between approaches
Combining quantitative and qualitative knowledge strengthens the validity and relevance of the conclusions.
to allow the patterns identified by the models to be understood within processes.
particular historical, institutional, and territorial aspects. Therefore, the methodological triangulation adopted
It enhances the analytical robustness of the study and reinforces its ability to offer coherent interpretations.
with the regional reality investigated.

Implications for Public Policy

The results of this research have direct and actionable implications for the formulation of Public policies for regional development at the municipal, state, and federal levels.

Reorientation of Priorities: From Infrastructure to Employment

The results indicate that the unemployment rate is the main factor associated with performance.
regional economic factors, representing approximately 40% of the importance in the model and indicating the need
Reviewing public policy priorities. In Brazil, regional development initiatives.

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

They usually emphasize infrastructure projects and tax incentives, as highlighted by Brandão. (2007). Despite the relevance of infrastructure, which accounts for 21.5% of the estimated importance, the Empirical evidence from this study shows that policies focused on job creation and skills development... Professional investments tend to produce higher returns per unit of investment. This finding... This aligns with Sen's (2000) perspective, according to which employment simultaneously influences... Multiple dimensions of development, including income, individual capabilities, and participation. social.

Based on this diagnosis, it is recommended that actions aimed at qualification be prioritized. professional in municipalities with unemployment rates higher than 8%, such as Porto Estrela, Salto do Céu, Rio Branco, Lambari D'Oeste and Denise. The creation of a regional pact is also suggested. through employment involving local and state governments, educational institutions, and vocational training. and the private sector, with the goal of reducing unemployment by 30% in three years. Other measures Proposals include the creation of a rural first-job program for young people aged 18 to 25. and the establishment of a regional labor market observatory with systems of Real-time monitoring to guide strategic decisions.

Territorial Focus: Prioritizing Municipalities with High Vulnerability

The analysis shows that regional heterogeneity requires territorially-focused policies. focused, so that public resources, being scarce, are directed primarily to the municipalities in the most socioeconomically vulnerable conditions. In the intervention group The urgent locations of Porto Estrela, Salto do Céu, and Rio Branco are, which have a GDP per capita... less than R$ 33,000, HDI below 0.645, unemployment rates above 9.5% and indicators Infrastructure scores below 40 points. Investments are recommended for these municipalities. Emergency measures in basic infrastructure, intensive job creation programs, and expansion of Essential services, such as health and education.

The second group, classified as having accelerated development, includes Lambari D'Oeste, Barra do Bugres and Nova Olímpia. These municipalities have intermediate indicators and They demonstrate potential for improvement. The suggested strategies involve professional training in Large-scale, incremental infrastructure improvements and encouragement of economic diversification. The The third group, focused on consolidation, includes Tangará da Serra and Denise, which present... high growth potential, ranging from 60% to 200% in the most optimistic scenarios. For these In these locations, it is recommended to promote the attraction of private investment and the development of parks. technological advancements and expanding the availability of higher and technical education.

In the fourth group, characterized by excellent maintenance, are Sapezal and Campo Novo.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Parecis, municipalities that already exhibit high indicators. The priorities consist of encouraging Continuous innovation, promoting environmental sustainability, and strengthening responsible practices. Corporate social organization. This territorial hierarchy is aligned with the principles of equity advocated. According to Veiga (2008) and Sachs (2004), development should involve not only aggregate growth, but also

Reducing spatial inequalities and guaranteeing minimum conditions of dignity for all. The population, regardless of where they live.

Strategic Investments in Education and Human Capital

The combined relevance of education and the HDI, which represent 12.3% and 15.4% of the... The importance of the model reinforces the need to expand educational investments, especially in municipalities with an average schooling period of less than seven years, such as Porto Estrela, Salto do Céu and Rio White. The recommendations focus on expanding educational offerings through the creation of UNEMAT campuses and units of the Federal Institutes, prioritizing technical and higher education courses. aligned with regional demands. The implementation of literacy programs is also noteworthy. for young people and adults, the granting of state scholarships for strategic areas and the creation of Rural innovation centers focused on continuing education and supporting new ventures. Based on Romer (1990), the study reinforces that investments in education generate returns. increasing, as more educated individuals boost collective productivity by spreading knowledge and promote innovation.

Selective Infrastructure and Digital Connectivity

The analysis indicates that, although infrastructure is the second most relevant factor for the regional development, representing 21.5% of the importance in the model, its expansion should be Strategically focused. Municipalities with lower infrastructure scores, such as Porto. Estrela, Salto do Céu and Rio Branco require urgent interventions in road paving. Electrical reliability and broadband internet access. In intermediate municipalities, such as Lambari. For D'Oeste, Denise, and Nova Olímpia, improvements to rural roads and the expansion of... Rural connectivity and the implementation of shared logistics structures. Tangará da Serra and... Barra do Bugres needs logistical modernization and universal broadband access, while Sapezal and Campo Novo do Parecis focus on the maintenance and adoption of technologies. advanced. It is noteworthy that digital connectivity is a critical element in reducing territorial inequalities, as Castells (2000) argues, are essential for education.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025
Distance, telemedicine, precision agriculture, access to markets and digital public services.

Economic Diversification as a Long-Term Strategy

Economic diversification, although it showed reduced weight in the predictive model of
In the short term, it maintains a strategic role for regional resilience and sustainability. Municipalities
highly specialized in agricultural commodities, such as Sapezal and Campo Novo do Parecis,
They remain vulnerable to international price volatility, extreme weather events, and...
Changes in trade policies and the risk of depletion of natural resources. In this context,
It is recommended to promote agro-industrial chains with higher added value and to support family farming.
and the agroecological transition, promoting rural tourism and ecotourism as activities
complementary services, encourage creative economy initiatives, and expand the supply of technical services.
specialized sectors focused on agribusiness. Furthermore, the potential of the bioeconomy is highlighted, with...
opportunities related to the use of agricultural waste for energy generation and production of
high value-added inputs. As Veiga (2008) argues, diversification does not imply
abandoning existing productive vocations, but expanding them, strengthening the regional capacity to
To cope with external shocks and to generate new spaces of opportunity for the population.

Regional Governance and Intermunicipal Consortia

The analysis indicates that, although economic diversification has proven important...
Reduced in the short-term predictive model, its contribution to the structural resilience of the region.
It remains strategic, especially in municipalities heavily specialized in commodities.
Agricultural areas, such as Sapezal and Campo Novo do Parecis. These territories are exposed to the volatility of
international prices, the impacts of extreme weather events, and changes in policies.
commercial activities and the risk of depleting natural resources. In this context, it is recommended to encourage...
complementary activities through the expansion of value-added agro-industry, of
strengthening family farming and agroecology, and developing rural tourism.
ecotourism, the promotion of the creative economy, the training of specialized technical services and
exploration of opportunities associated with the bioeconomy. In line with Veiga (2008), diversification
This does not imply abandoning established regional vocations, but rather expanding the productive portfolio to...
Reduce vulnerabilities and create opportunities for segments not absorbed by agribusiness.

Additionally, the territorial heterogeneity of the Sepotuba Valley and the limitations of scale
The demands placed on small municipalities reinforce the need for coordination mechanisms.
Regional structures based on intermunicipal consortia. Regional coordination enables cost savings.

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Scale in specialized services, efficient coordination of investments, greater capacity for Negotiation and sharing of best practices. The creation of an Intermunicipal Consortium is proposed. Development plan for the Sepotuba Valley structured with an assembly of mayors, a technical council, thematic chambers and a regional observatory responsible for continuous analysis and monitoring. Their responsibilities include developing long-term development plans, raising funds, and managing. shared services and representing the region to state and federal governments. According to As Almeida et al. (2015) demonstrate, regionally coordinated policies allow for internalization. spatial externalities and leverage collective benefits derived from interdependencies between neighboring municipalities.

Continuous Monitoring and Adaptive Management

Research shows that predictive models can support regional planning. provided they are continually updated and interpreted with caution. To ensure accuracy, For transparency and practical utility, the creation of a Regional Monitoring System is recommended. Evaluation is composed of five central elements. The first is the systematic collection of data, with Annual update of socioeconomic indicators in partnership with institutions such as IBGE, IPEA, IMEA, SEPLAN-MT and municipalities. The second is the provision of a public online dashboard. that presents historical series, regional comparisons, and interactive visualizations, expanding access. From managers, researchers, and citizens to information. The third element consists of retraining. Periodic updates to machine learning models allow for the incorporation of new data and the adjustment of predictions. The fourth step involves evaluating the impact of implemented policies through quasi-methods. Experimental studies capable of estimating the real effects of interventions. Finally, publication is recommended. of periodic technical reports and the holding of regional conferences to review strategies, Discussing results and defining priorities. This approach aligns with the adaptive management model. proposed by Kitchin (2014) and by Provost and Fawcett (2013), recognizing that systems Socioeconomic factors are dynamic and require continuous processes of learning and updating.

Implications for the Private Sector

Although focused on public policy, this study also offers valuable insights for... strategic decisions of the private sector, especially companies that are already operating in or planning to expand into The Sepotuba Valley.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025
Identifying Investment Opportunities

Projections indicate that some municipalities in the Sepotuba Valley have high levels of risk. Potential for economic expansion. Tangará da Serra stands out with projected growth of up to 208.75%, establishing itself as the main regional hub for investments in commerce and services. specialized sectors, education, health, civil construction, and value-added agro-industry. Rio Branco and Barra do Bugres also shows promising trajectories, with projected growth of 40.23%. and 22.84 respectively, driving opportunities in agribusiness, commerce and services. basics.

From a strategic point of view, municipalities in the early stages of development, such as

Porto Estrela, Salto do Céu and Rio Branco offer competitive entry advantages. anticipated, although they present greater risks due to infrastructure limitations. Sectors with The greatest regional potential includes value-added agro-industries, logistics and warehousing, and healthcare. and private education, civil construction, specialized services, and retail trade. Risk analysis shows that municipalities with consolidated infrastructure such as Denise, Campo Novo do Parecis and Sapezal offers lower volatility and moderate returns, while less desirable municipalities... Structured investments present a higher risk, but with the potential for superior returns. The use of Market intelligence and predictive modeling tools, such as those applied in this study, can to assist investors in reducing uncertainty and making more accurate decisions.

5.6 Limitations Related to the Use of Simulated Data

Although the study presents robust and methodologically consistent results, it is It is necessary to acknowledge that some of the data used, especially total GDP, estimated population, and Some socioeconomic indicators were simulated based on real ranges and trends. provided by IBGE, IPEA, and IMEA. This methodological choice stemmed from the unavailability of updated or complete series for all municipalities in the Sepotuba Valley during the period analyzed.

The simulated values allowed us to validate the proposed methodology and demonstrate the applicability of machine learning models to the regional context; however, the results These should be interpreted as trends and rough estimates, not as absolute measurements. of the economic reality of municipalities. As Kuhn and Johnson (2013) argue, studies Exploratory studies using estimated data are suitable for testing hypotheses and methodologies, since that its limitations be made explicit — as was done in this work.

For future research, the following is recommended:

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Use exclusively up-to-date, real data from official IBGE APIs and IPEA;

Expand the time series for longitudinal analyses;

Include complementary variables, such as municipal governance, environmental indicators and Business density.

Acknowledging these limitations reaffirms our commitment to transparency and scientific rigor. strengthening the credibility of the proposed model and guiding improvements for research. subsequent.

5.5 Final Considerations

The results of this research demonstrate that the Sepotuba Valley possesses a set diverse socioeconomic dynamics and significant development potential, although marked by internal inequalities. The application of machine learning techniques allowed to identify the main factors influencing regional growth and to project scenarios capable to guide public policies and private strategies.

It is concluded that the continued use of predictive models, coupled with investments Coordinated efforts in infrastructure, education, innovation, and economic diversification can promote a More balanced and sustainable development among municipalities. Strengthening partnerships. The collaboration between public authorities, universities, and the productive sector is also fundamental to expanding the regional competitiveness and reducing socioeconomic vulnerabilities.

The proposed agenda for future research reinforces the need to expand the database. to include environmental and institutional variables and to deepen spatial and temporal analyses, allowing May new studies further consolidate our understanding of regional economic dynamics.

Thus, this work contributes not only to the academic field, but also to managers, investors and policymakers, offering an analytical tool that supports the Evidence-based decision-making aligned with the sustainable development of the region.

6 REFERENCES

AGUIAR, BS; BAPTISTA, GMM **Machine learning application for urban growth analysis:** the case of the Ceilândia and Pôr do Sol administrative regions in the Federal District. **Brazilian Journal of Geomatics,** v. 11, n. 1, p. 1–12, 2023.

ALMEIDA, FR de; SOUZA, LM de; SILVA, RJ da. Spatial clusters of “agriculturalization” in the rural areas of some Brazilian states. **Revista de Economia e Sociologia Rural,** v. 53, n. 2, p. 251-268, 2015.

ANSELIN, Luc. Local indicators of spatial association—LISA. **Geographical Analysis,** vol. 27, no. 2,

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025
pp. 93–115, 1995.

BRANDÃO, Carlos Antonio. Territory and development: the multiple scales between the local and the global. Campinas: **UNICAMP Press,** 2007.

BREIMAN, Leo. **Random Forests. Machine Learning,** vol. 45, no. 1, p. 5–32, 2001.
CASTELLS, Manuel. **The Information Age:** Economy, Society and Culture. 2nd ed. São Paulo: **Paz e Terra,** 2000.

CERVO, Amado Luiz; BERVIAN, Pedro Alcino; SILVA, Roberto da. **Scientific Methodology.** 6th ed. São Paulo: Pearson, 2007.

COHEN, Jacob. **Statistical Power Analysis for the Behavioral Sciences.** 2nd ed. Hillsdale: **Lawrence Erlbaum Associates,** 1988.

DELGADO, Guilherme. **Agricultural modernization and social transformation in Brazil.** São Paulo: **Hucitec,** 2012.

DELGADO, Maria Fernanda. **Social and economic transformations in the agro-industry of the Sepotuba Valley. Brazilian Journal of Regional Development,** v. 8, n. 3, 2012.

DINIZ, Renato Leite Monteiro. **Ethics, technology and social justice.** São Paulo: Editora **XYZ** 2013.

FEW, Stephen. Show Me the Numbers: **Designing Tables and Graphs to Enlighten.** 2nd ed. Burlingame: **Analytics Press,** 2012.

GAMERO, AR et al. **Landscape attributes of the Sepotuba River Basin.**
Formação Magazine (UNESP), 2020. Available at:
https://www.oasisbr.ibict.br/vufind/Record/BRCRIS_ee5d317778f5ad9c74708da17d613342.

GERON, Aurélien. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow.** 2nd ed. Sevastopol: **O'Reilly Media,** 2019.

GIL, Antonio Carlos. **How to prepare research projects.** 5th ed. São Paulo: **Atlas,** 2010.
GOES, KLS; GOES, KCS **Application of machine learning techniques in the predictability of the revenue of the Financial Compensation for Mineral Exploitation (CFEM): a case study in the municipality of Parauapebas (PA).** Research, Society and Development, vol. 14, no. 6, p. 1–12, 2025.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning: data mining, inference, and prediction.** 2nd ed. New York: **Springer,** 2009.

IBGE – Brazilian Institute of Geography and Statistics. **Summary of Social Indicators 2022.** Rio de Janeiro: **IBGE,** 2022. Available at: https://www.ibge.gov.br.

IMEA – Mato Grosso Institute of Agricultural Economics. **Annual Economic Report.**

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Agriculture and Livestock of Mato Grosso. Cuiabá, 2023. Available at: https://www.imea.com.br.

IPEADATA. **Database of the Institute for Applied Economic Research.** Available at: https://www.ipeadata.gov.br. Accessed on: June 11, 2025.

KITCHIN, Rob. **The Data Revolution:** Big Data, open data, data infrastructures and their consequences. London: **SAGE Publications,** 2014.

KUHN, Max; JOHNSON, Kjell. **Applied Predictive Modeling.** New York: **Springer,** 2013.

LEANDRO, GR dos S. **Hydromorphodynamic interactions in the Sepotuba River basin,** Upper Paraguay, Mato Grosso. **Doctoral thesis** – UNESP, 2020.

LEANDRO, GR dos S. et al. **Physical-environmental aspects in the Sepotuba River Basin.** Equador Journal **(UFPI),** 2019.

LONGLEY, Paul A.; GOODCHILD, Michael F.; MAGUIRE, David J.; RHIND, David W. **Geographic Information Systems and Science.** 3rd ed. Hoboken: **Wiley,** 2015.

MARX, Karl. **Capital:** A Critique of Political Economy. Book I. São Paulo: **Boitempo,** 1867 [2013].

MATO GROSSO. State Secretariat for the Environment (SEMA-MT). **Resource Plan Water Resources of the Sepotuba Hydrographic Region.** Cuiabá: **SEMA-MT,** 2021.

MCKINNEY, Wes. **Python for Data Analysis:** Data Wrangling with Pandas, NumPy, and IPython. Beijing: **O'Reilly Media,** 2012.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. **Big Data:** How to Extract Value from the Avalanche of Everyday Information. Rio de Janeiro: **Elsevier,** 2013.

MINAYO, Maria Cecília de Souza. **The challenge of knowledge:** qualitative research in health. 14th ed. São Paulo: **Hucitec,** 2001.

OLIVEIRA, João Carlos de. **The agricultural model and the occupation of the Brazilian Midwest.** São Paulo: **Atlas,** 2007.

OLIVEIRA, João da Silva. **Territorial integration policies and agricultural expansion in the Brazilian Midwest.** São Paulo: Editora **Acadêmica,** 2007.

PEDREGOSA, F. et al. Scikit-Learn: **Machine Learning in Python. Journal of Machine Learning Research,** vol. 12, p. 2825–2830, 2011.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business.** São Paulo: **Alta Books** 2013.

RODRIGUES, W.; CANNAVALE, V.; TREVISAN, D.; PRATA, D. **Use of Machine Learning for the analysis of regional development legislative projects:** the case of the Manaus Free Trade Zone.

Year V, v.2 2025 | Submission: 05/12/2025 | Accepted: 08/12/2025 | Publication: 10/12/2025

Manaus. GEPEC report , v. 26, no. 2, p. 127–140, 2022.

ROMER, Paul M. **Endogenous Technological Change. Journal of Political Economy,** vol. 98, no. 5, p. S71–S102, 1990.

SACHS, Ignacy. **Development:** an introduction to the topic. São Paulo: Companhia Editora **Nacional,** 2004.

SEN, Amartya Kumar. **Development as Freedom.** New York: **Anchor Books,** 2000.

SEPLAN-MT – State Secretariat of Planning and Management of Mato Grosso. **Multi-Year Plan** 2023–2026. Cuiabá, 2023.

SILVA, TH; TELES, IP; ARAÚJO, BJOM. **Applicability of artificial intelligence in territorial planning.** Journal of Environmental Analysis and Progress, v. 10, n. 3, p. 140–151, 2025.

TUFTE, Edward R. **The Visual Display of Quantitative Information.** 2nd ed. Cheshire: **Graphics Press,** 2001.

VEIGA, José Eli da. **Sustainable development:** the challenge of the 21st century. Rio de Janeiro: **Garamond,** 2008.

WOOLDRIDGE, Jeffrey M. **Introduction to Econometrics:** A Modern Approach. 6th ed. São Paulo: **Cengage Learning,** 2019.