RBGDR

DATA SCIENCE AND PUBLIC POLICY: A METHODOLOGICAL PROPOSAL FOR QUALIFYING POLICIES FOR LOCAL/ REGIONAL DEVELOPMENT

DATA SCIENCE E POLÍTICA PÚBLICA: PROPOSTA METODOLÓGICA PARA A QUALIFICAÇÃO DE POLÍTICAS PARA O DESENVOLVIMENTO LOCAL/REGIONAL

REVISTA BRASILEIRA DE GESTÃO E DESENVOLVIMENTO REGIONAL
V.19, N°3, Set-Dez/2023 | Taubaté/SP - Brasil | ISSN 1809-239x

DATA SCIENCE AND PUBLIC POLICY: A METHODOLOGICAL PROPOSAL FOR QUALIFYING POLICIES FOR LOCAL/ REGIONAL DEVELOPMENT

DATA SCIENCE E POLÍTICA PÚBLICA: PROPOSTA METODOLÓGICA PARA A QUALIFICAÇÃO DE POLÍTICAS PARA O DESENVOLVIMENTO LOCAL/REGIONAL

Victor da Silva Oliveira[1] | Tiago Costa Martins[2]
Arlindo Figueirôa Escobar Teixeira de Oliveira[3]

Received: 01/17/2023
Accepted: 10/31/2023

[1] PhD in Geography (UFPE).
Professor at the Federal University of the South and Southeast of Pará. Xinguara – PA, Brazil.
Email: victorsoliveira@hotmail.com

[2] PhD in Regional Development (UNISC).
Professor at the Federal University of Pampa.
São Borja – RS, Brazil.
Email: tiagomartins@unipampa.edu.br

[3] PhD student in Information Science (UFPE).
João Pessoa - PB, Brazil
E-mail: arlindo.escobar@gmail.com

ABSTRACT

This article posits a methodological intervention to enhance various individual, social, and spatial attributes within local and regional settings. Concerns span from conceptualizing development to strategizing interventions and defining both monetary and non-monetary investment priorities to achieve stated goals. These topics are widely discussed in academia and among public policymakers. Utilizing Data Science, the paper proposes a method to qualify public policies that promote local/regional development by increasing regional disposable income. This approach revisits classical economic and regional development theories, suggesting that productive diversification may elevate regional disposable incomes. Data Science methods applied to spatial information mining, systematization, and synthesis—coupled with regional analysis techniques—equip managers and academics with tools to identify public policy strategies for local and regional development. The proposal introduces challenges, including the need for technical and human resource enhancement for effective implementation, and requires a convergence that extends beyond technique to the essence of public policy formulation. Nevertheless, applying this method can precisely describe and diagnose regional characteristics, thereby predicting and prescribing sustainable alternatives for local/regional transformation.

Keywords: Regional analysis. Local and regional development. Local and regional policy.

RESUMO

Postular uma intervenção na realidade local e regional com a pretensão de qualificar um ou mais atributos individuais, sociais e espaciais permeia uma gama de inquietações. Elas vão desde a concepção de desenvolvimento, passam pela definição das estratégias para a intervenção e chegam à definição de prioridades de investimentos — não apenas monetários — para atingir os objetivos elencados. Essas questões estão presentes nos bancos acadêmicos e nos formuladores de políticas públicas. Diante desse contexto e através da aplicação de *Data Science*, o presente artigo visa propor uma metodologia para qualificar políticas públicas que visem ao desenvolvimento local/regional a partir da ampliação da renda regional disponível. A proposta parte da releitura de clássicos do desenvolvimento econômico e regional com o pressuposto de que a diversificação produtiva pode ser um canal para ampliar a renda regional disponível. Utilizando metodologias de *Data Science* para mineração, sistematização e sintetização de informações espaciais atreladas às técnicas de análise regional, possibilita-se que gestores e acadêmicos identifiquem estratégias que para o desenvolvimento local e regional por meio de políticas públicas. A proposta traz alguns desafios, como a qualificação técnica e de recursos humanos para a sua efetiva aplicação, assim como a convergência para além da técnica que engloba a definição de uma política pública. Não obstante, a potencial aplicação dessa metodologia possibilita uma averiguação com considerável grau de acuidade de descrição e diagnóstico regional, atingindo a predição e a prescrição de alternativas viáveis para a transformação local/regional.

Palavras-Chave: Análise regional. Desenvolvimento local e regional. Política local e regional.

INTRODUCTION

Reflections on local and regional development remain a focal point of scholarly debate worldwide. In Brazil's academic milieu, Theis (2022) interrogated the prevailing notions within postgraduate programs that address this subject. The author's findings accentuate the breadth of the discourse and the challenge inherent in reconciling perspectives rooted in the social base, capital, and the State. The issue becomes even more complex when, beyond the academic benches, it comes up against social practice. Brazil's attempts to apply local/regional development policies, whether explicit or implicit (Araújo, 2013), in the past and the present, present a diverse range of conceptions and modes of application. This fact enhances the richness of the concepts, as well as posing a challenge to anyone seeking to intervene in reality with the aim of changing the individual, social, and spatial status quo.

At the nexus of theoretical inquiry and practical application, this article introduces a method employing Data Science to refine public policies directed at local/regional development, focusing on expanding regional disposable income. The objective is to leverage real data to forge predictive and prescriptive insights that bolster the implementation of collective efforts toward development through public policy.

This article is structured in four sections, in addition to this introduction and the final considerations. In the next section, the theoretical construction is presented, based on the classics of reflection on economic and regional development, with a rearrangement that aims to consider the intention of increasing regional disposable income through productive diversification. The next section discusses how identification can be carried out strategically, both theoretically and methodologically. The proposed methodology (based on the previous theoretical construction) is then presented in detail. This section includes regional analysis techniques and databases with information that can be consulted to objectively materialize the theoretical assumptions presented. However, the same section also deals with Data Science mechanisms for applying the techniques and optimizing the aforementioned data. Finally, in the results and discussions section, concrete issues are addressed that reveal the potential use of the methodology presented, as well as some limitations and points of necessary acuity.

A LOCAL/REGIONAL DEVELOPMENT APPROACH

The section's title intentionally specifies "one" approach to underscore the restrictive nature of the proposition herein. The corpus of economic and regional development theories is neither novel nor scarce. Overlooking the broader debate, the plethora of interpretations concerning the determinants of economic activities in a locality/region is noteworthy for its variety.

In attempting to navigate the expanse of development theories, creating a unique interpretative framework for a given reality may yield predictable outcomes. Among these, the most notable is the introduction of a "new theoretical-methodological framework." To the discerning researcher, these frameworks often reveal extensive ties to established theories, albeit cloaked in a veneer of novelty. To construct the theoretical scaffolding for the local/regional development method proposed, we embark

on the following endeavors: (i) a critical reexamination of classical works (operational), (ii) a focus on productive diversification (assumption), and (iii) an aim to elevate regional disposable income (intention). From these premises, developing a theoretical foundation that supports a methodological proposal for local/regional analysis with innovative attributes is undertaken, not by crafting a new theory but by reorganizing selected concepts from venerable economic and regional development theories. The goal remains steadfast: to achieve development through socio-economic diversification and the augmentation of regional disposable income.[1]

The quest for local/regional productive diversification, which is economically sustainable, propels public policies that elevate income. This directive serves as a crucial guide in revisiting economic and regional development theories to extract insights for the theoretical framework formulation. Paiva's (2013) regional analysis approach, influenced by North's tenets (1955; 1959), scrutinizes the nature of investments spurred by the export commodity and its inherent capacity to spawn external economies.

The hypothesis posits that the disposable income (Yd) of a municipality/region correlates with two principal variables: the value of exports (X), deriving from the sale of goods and services beyond local confines,[2] and the value of income distribution (Y). As North (1955; 1959) analyzed, sectors capitalizing on natural resources often secure a substantial share of production as profit, with a lesser portion allocated to taxes and an even smaller fraction distributed as workers' wages. The previously posed question contemplates a flexible percentage margin confined to local/regional parameters, acknowledging that profit accrual and significant investment in the sector typically involve importing goods from other national regions or internationally.

Direct and indirect linkages induced by the export sector assume prominence, with notable variances among sectors concerning their capacity for interconnectivity proximate to the primary activity's location. Hirschman's reflections (1958) succinctly distinguish, for example, between the export of raw logs and finished furniture. Meanwhile, North (1955; 1959) contends that the essence of development is encapsulated in the aggregate of export revenue and its potential to stimulate a domestic market—

1 Boisier (2000) states that there is a region in "real terms" and another in "potential terms". In other words, there are regional potentialities that can be fostered with a view to certain development.

2 In this case, export is not necessarily seen as an international action (to another country), but also as an action outside the locality/region.

denoted as 'X + Yd.' The inquiry then shifts to identifying sectors with this bifurcated capacity, partially addressed by Hirschman (1958), who explores sectors/activities conducive to robust linkages.

Further elucidating this matter, North (1955; 1959) differentiates between domestic disposable income (Yd) and regional disposable income (Yrd) through each economic segment's propensity to import (m). Broadly, in a municipality/region with a natural goods-based export economy, there are imports of capital goods (I) and high value-added consumer goods (Ck)—financed by profit capture and elevated wages. Consequently, in such economies, the import propensity is a function of I and Ck (m = I + Ck). Conversely, export goods (X) and goods consumed by workers (Cw) are locally produced, exhibiting a low import propensity. The regional disposable income equation is thus articulated as follows:

$$Yrd = (X + Cw + Gi) - m\ (Ck + I + Cw + T)$$

Where:
X represents the Value of Exports
Cw denotes Goods consumed by workers
Gi is Government spending
m stands for Propensity to Import
Ck symbolizes Capitalist Consumer Goods
I indicates Capital Goods
T refers to Taxes not reimbursed

STRATEGIC APPROACH TO INCREASING REGIONAL DISPOSABLE INCOME

In the context of this discourse, the strategic mechanisms for amplifying disposable regional income, which in effect enhance social benefits through socio-economic diversification, come to the fore. As elucidated, the distinct nature of each economic segment delineates the strategic trajectories. To ascertain the retention of the income generated within the region, the emphasis should be on initiatives that (i) exhibit a low propensity to import, such as consumer goods produced by workers with specialized skills; (ii) possess the potential to transition into exportable products beyond the municipal/region; (iii) demonstrate the capacity for intensive linkage; and (iv) can be feasibly targeted by policies within local/regional planning and development projects.

The theoretical groundwork for these strategies draws from two streams of thought—North and Hirschman—and centers on the capacity of certain economic activities to become key drivers in productive diversification, marked by their external trade and the stimulation of proximate activities within the municipality/region.[3] This foundation also casts light on the theoretical and methodological aspects of where and how to intervene to ensure all-encompassing local development.

Smith (1983), whose work Paiva (2013) examined, discusses the Smithian loop and the prospects of surmounting it. According to Paiva, the wealth of any given locality is intrinsically linked to the ability to enhance labor productivity. Any productive activity that aims to gain new consumer markets needs competitiveness, which, in short, is represented by offering value-added goods and services at the lowest possible cost, i.e. producing more, in less time, with quality and low cost. This effort, in turn, is the result of the division of labor, which is limited by the size of the consumer market for products created with ever-improving productivity rates. Thus, the size of the market is a sine qua non for the expansion of the division of labor. To overcome this loop, the expansion of markets, achieved by increasing labor productivity, brings with it the need to attract new consumers. In other words, exporting outside the locality/region.

As indicated above, this strategy should focus on segments that initially add to the available regional income by identifying and fostering production chains, as Marshall (1996) discussed. As production matures due to increased productivity and the accumulation of expertise, a specialty will be generated with an advantage over surrounding municipalities/regions, enabling the product/service to become a new segment to be exported.[4] This proposition, as reflected by North (1955; 1959), repositions the municipality as a provider of goods and services in a new urban and regional hierarchy. As a result, various gears are created around a set of activities that feed back into each other in the form of a production chain.

The movement to boost income by maintaining high demand is one of Keynes' (1985) main contributions to development economics. The author deals with the State's role as an active agent, especially in times of crisis. In fact, the actions of public authorities through strategically defined actions can play an important role in increasing income in market niches that boost regional disposable income.

3 The methods on how to identify these sectors and how they can be taken into account when qualifying public policies are presented in a later section.

4 It is understood as being sold to places outside the municipality/region. It is worth noting that there are two types of exported products/services: those produced in the locality/region and sold abroad, and those produced in the locality/region and sold to consumers outside the locality/region.

State strategies for enhancing regional disposable income can be direct or indirect and approached through demand or supply. These strategies are not exclusive and do not necessarily converge into a single approach for planning public policies. For instance, one approach could be directly targeting the demand of sectors identified as potential stimulants of regional disposable income, such as through municipal public spending.[5] On the supply side, the government acts as a pivotal entity in organizing the flow of goods, information, and people, which directly influences production costs and productivity, as discussed by North (1959) and Smith (1983). Indirectly, public authorities at the municipal/regional level can coordinate stakeholders who may be distant from the production system to enhance production quality and foster innovation-led productivity.

Adding to this narrative is Schumpeter's (1988) discussion on the role of innovation in the production process and economic expansion. He identifies the entry into new markets as one of the key drivers of innovation, thus aligning this strategy with state actions, particularly from the perspective of municipal governance.

Santos' (1978) perspective on socio-spatial formation, as both theory and method, underscores the importance of considering the locality/region when proposing interventions to boost or transform productive activities. This requires acknowledging the interconnectedness of regions and the reliance on their historical, cultural, and economic trajectories, which can lead to regional spillover effects that impact the broader social fabric across spatial, social, and individual dimensions, as noted by Boisier (2000). Moreover, it is posited that productive activities that build a diversified economic base and enhance regional income—beyond what is generated by established export activities—should be promoted. This concept is backed by Paiva's (2013) regional analysis proposal, which distinguishes between two types of economic activities based on their dynamic function. Firstly, propulsive activities are designed to acquire basic income for the local economy by producing goods or services for sale outside the municipality/region, generating a monetary flow that stimulates the local economy. This category also includes activities and services catered to local consumption by individuals from other municipalities and governmental

5 The capacity of each municipality to act must be considered, especially directly. The scale of a direct investment by a municipality with more than one million inhabitants is very different from that of a municipality with less than 50 thousand. However, those that receive royalties from any source for the exploitation of a natural asset are also different, regardless of the number of inhabitants.

operations financed by external taxes.[6]

Secondly, reflex activities serve a different but complementary dynamic function by supplying the local market with retail and fundamental services for the residents.[7] The vigor of these activities is tied to the resource influx from propulsive activities, and they tend to expand in response to this flow. Moreover, mixed activities embody characteristics of both propulsive and reflexive functions. Analyzing each type of dynamic function activity has different strategic objectives. Propulsive activities can help to understand the current level of productive specialization in activities with a real capacity to increase local circulating income. For example, diagnosing just one outstanding activity may indicate a potential satelliteization of the local economy. Thus, its counterpoint would be a diversified, autonomous, and growing local economy with several prominent propulsive sectors — the paths to be promoted by actions with this aim are indicated.

The review of reflex activities seeks to understand the level of evasion of basic income. A place/region with few activities indicates that potential consumers who would heat the local market are consuming in other locations. Furthermore, it may indicate that the purchasing power of families is low and there is no potential for trade, resulting in the generation of income and wealth that does not fuel inclusion.

METHODOLOGICAL PROPOSAL: REGIONAL ANALYSIS AND DATA SCIENCE FOR LOCAL/REGIONAL PUBLIC POLICY

To articulate the "methodological map" proposed herein, an empirical foundation is essential. The method commences with the utilization of microdata from the Ministry of Economy's Annual Social Information Report (Rais), employing the National Classification of Economic Activities 2.0 (CNAE), which categorizes economic activities into 1,357 subclasses. This approach can be tailored to both local (municipal) and regional scales.

The proposed revisions should conduct comparative analyses typical of regional analysis since clarity of the dimensions of the indicators can only be achieved when compared with neighboring realities, embedded regions, and the State itself. This study suggests utilizing the

6 All these activities form part of the first function described above, which defines the region's internal income - (X + Gi).

7 Represented in the previous function by Cw.

regional delineations established by the Brazilian Institute of Geography and Statistics (IBGE), which outline intermediate regions, constituent municipalities, and their positions in the urban hierarchy.

Comparative municipalities, which are part of intermediate or adjacent regions, may use as benchmarks for analogy municipalities that hold a superior urban and regional hierarchical position, and in the case of a regional analysis, comparisons with other regions and state and national indicators are considered. Within the Rais microdata framework, three fundamental data points are considered integral to this methodological approach.

The initial data point concerns formal employment across economic segments. This information is proportionally compared to the State to determine if there is a productive specialization in the sectors, as indicated by the Locational Quotient (Lq).[8] North (1955; 1959) posits that an LQ greater than 1 signifies that economic activity is proportionally generating more employment than the state average, suggesting that the activity extends beyond the domestic market, hence classifying it as a propulsive activity. The LQ is calculated using the following equation:

$$Lq_{ij} = \frac{\left[\frac{(E_{ij})}{\sum_{j}(E_{ij})}\right]}{\left[\frac{\sum_{i}(E_{ij})}{\sum_{j}\sum_{i}(E_{ij})}\right]}$$

In which:

Lq_{ij} is the locational quotient of class i in municipality j;

E_{ij} is the employment in class i in municipality j;

$\sum_{j} E_{ij}$ is the employment in all classes in municipality j;

$\sum_{i} E_{ij}$ is the employment in class i in the reference region;

$\sum_{j}\sum_{i} E_{ij}$ is the employment of all classes in the reference region.

8 For more information on the use of LQs for applied regional analysis, please see: Bitencourt and Guimaraes (2012); Mattei and Mattei (2017); and Crocco *et al*. (2006).

The second piece of information based on Rais microdata is the wages paid to formal workers in each of the activities according to their dynamic function - propulsive or reflexive. Being aware of the increase in income for each segment goes beyond the circulating monetary value that each productive sector brings to the local economy in the form of wages for workers. Propulsive segments, given their specialization and productive capacity, tend to pay higher wages than reflex activities as a whole. In addition to regional analysis indicators, this data makes it possible to identify each economic activity according to its dynamic function; it also provides a basis for selecting activities to be promoted with a view to increasing regional disposable income.

Finally, the Rais provides information on establishments and their main characteristics based on the CNAE 2.0 classification. Elements such as the nature and size of the establishment and the number of workers are important in showing how each segment is structured in the municipality. At the same time, this identification makes it possible to understand the size of the companies operating in each dynamic function, since different strategies are needed for micro, small, medium, and large companies.

Parallel to the Rais microdata, the aim is to compile information on individual micro-entrepreneurs (known as MEIs in Brazil) This effort seeks to partially overcome the limitations of the RAIS microdata, which only show indicators for formal employment. In this way, adding MEIs enhances the analysis in terms of the generality of the study, since, in addition to the technical/methodological resource, the number of MEI registrations in Brazil is growing. The LQ should also be sought for eminently rural activities, such as livestock farming and agriculture. However, the informality rate typical of these activities would make it difficult to obtain clear results from the RAIS microdata. As an alternative, we recommend using rural production data from the IBGE Agricultural Census, such as the value of animal production, the number of livestock and the production of temporary and permanent crops.

Additionally, data on international exports from the municipality/region reinforce the assessment of productive specialization. It is assumed that the entry of a certain product into the large and competitive international market is a sign of a high level of productivity. In the same way, significant parts of the inputs and machinery for activities with high productivity are purchased from abroad, so import data makes it possible to identify them. The Ministry of Economy's micro-data on foreign trade provides access to information such as export and import volumes, their respective dollar values, partner countries and the

transformation of municipal/regional integration since 1989.

In addition to productive specialization itself—in which the LQ plays a leading role—four other complementary indicators are used for specific reasons. The Herfindahl-Hirschman Index (HHI) is used to determine the concentration of a particular sector in the municipality (i.e., the LQ indicates whether there is specialization). The HHI, on the other hand, tells us whether this sector is marked by a power of attraction due to specialization or, conversely, a low power of attraction. This indicator refers to the level of competition between sectors, configuring them in a gradual way, ranging from perfect competition to monopoly. As previously discussed, this methodology seeks to verify sectors with the potential to generate products and services for export which, at the same time, do not tend towards concentration, given their lower capacity to impact on regional disposable income. The HHI is calculated as follows:

$$HHI = \frac{P_{ij}}{P_{it}} - \frac{P_{tj}}{P_{tt}}$$

In which:
P_{ij} is the sector i of municipality j;
P_{it} is the total for sector i;
P_{tj} is the total for municipality j;
P_{tt} is the total reference region.

Another indicator associated with specialization issues is the location coefficient. It varies between zero and one, making it possible to identify the dispersion between the various sectors of economic activity and classifying those with a greater or lesser tendency towards spatial concentration. The classification is made into three groups: the best dispersed, close to zero; those in between; and those with highly concentrated sectors, with a result close to one. The location coefficient is determined by the following formula:

$$CL = \frac{\left(j_{ei} - \sum_{i} j_{ei}\right)}{2}$$

In which:
j_{ei} is the employment in i in municipality e;
$\sum_{i} j_{ei}$ is the employment of all classes in the reference region.

The specialization coefficient indicator is also used to check the diversification and specialization of production activities. This also varies between zero and one and checks whether the municipality/region has a similar production structure to the reference region. If the result is close to one (i.e., with a different productive structure to the reference region), the productive sector will be considered specialized, as it has a high presence in parallel to that present in the reference region. The equation used is as follows:

$$C\,E_i = \sum_i \frac{\left(i_{ej} - \sum_j i_{ej}\right)}{2}$$

In which:

$\sum_i$ is the employment of all classes in the reference region;

i_{ej} is the participation of sector e in municipality j;

$\sum_j i_{ej}$ is the percentage share of sector i in the reference region.

To evaluate the geographical association between different sectors, the geographical association coefficient is used. This index, which also ranges from zero to one, indicates the degree to which locational patterns are geographically associated. This analysis, combined with the HHI, elucidates the complementarity within the productive structure, particularly for sectors that rely on others for efficient performance, thus shedding light on propulsive and reflexive activities. The geographical association coefficient is formulated as follows:

$$CAG_{ik} = \sum_j \frac{\left(j_{ei} - j_{ek}\right)}{2}$$

In which:

$\sum_i$ is the employment of all classes in the reference region;

J_{ei} is the participation of sector i in municipality e;

J_{ei} is the participation of sector k in municipality e.

By obtaining the results of the five indicators described, it is possible to ascertain, with a high degree of precision, the productive structure of the place/region under study, as well as glimpse the segments that are conducive to strategic public policy interventions for productive diversification. Based on the application of these regional analysis techniques and the theoretical assumptions presented above, a description and diagnosis of the productive structure will be presented, as well as the first situational indications of the retention and distribution of available regional income. The result provides indications of the sectors of activity with potential for inducing development. In summary, the five indicators will make it possible to see the following aspects (Table 1):

Table 1 | Interpretative summary of the selected indicators.

Indicator	Summary of results
Location quotient	Weak/medium/significant localization
Herfindahl-Hirschman Index	Perfect competition/monopolization
Location coefficient	Dispersion/concentration
Coefficient of specialization	Diversification/specialization
Coefficient of geographical association	Weak/medium/significant association

Source: Produced by the authors.

The interpretations derived from the indicators reinvigorate Keynes' (1985) theories, which position the State—represented by municipalities in this context—as a catalyst for development. This is operationalized through two strategic approaches: direct/indirect actions and demand/supply, supported by two databases. The municipality's Transparency Portal is the primary database, which grants public access to all municipal commitments. This resource is utilized to categorize commitments by the type of activity or service corresponding to the strategic segments identified in the previous analysis. It also allows for tracking financial outlays based on the location of the contracted products or services.

For this, the secondary database is used; it is constructed from the CNPJs (corporate tax numbers for Brazilian companies) of contractors, as listed on the Transparency Portal. This information is cross-referenced with data from the Federal Revenue Service, detailing businesses' location and primary and secondary economic activities as per CNAE classification. The objective at this juncture is to create a matrix that identifies potential commitments made outside the local or regional context. When combined with indirect actions, this data may guide and support strategic

sectors that generate internal regional income and promote diversification and socio-economic specialization within the municipality or region.

Furthermore, within the scope of municipal governments' ability to enhance their development, when activities with the potential to increase available regional income are identified, the State can act as a facilitator between the productive segments and the scientific and technological competency centers present in the (ICTs), and such facilitation aims to merge the needs of the public and private sectors with academic offerings. Research prospecting can be executed via the CNPq Directory of Research Groups (DGP-CNPq), which is organized to include a range of information: group leader, field of expertise, associated ICT, location, contact details (email and phone), summary of activities, research lines, networks of cooperation, human resources, partnering institutions, and available equipment. With this data, public administrators can swiftly pinpoint and leverage scientific and technological expertise conducive to the municipality or region's advancement, fostering growth and innovation that bolsters labor productivity, aligning with the principles set forth by Smith (1983) and Schumpeter (1988).

The final resources to be utilized in applying this method are the Brazilian Digital Library of Theses and Dissertations and the CAPES Periodicals Platform. These repositories are broadly recognized as principal sources of scientific output from Brazilian and international scholars. For the review of socio-spatial development in specific municipalities or regions, these databases offer a wealth of existing research on various Brazilian regions, providing valuable insights and data.

All the techniques and databases presented so far are commonly used in regional analyses focusing on the productive structure, as pointed out at the beginning of this section. They are invariably used to obtain interpretations and syntheses that are highly representative of the reality being analyzed. However, even though they deal with relatively similar aspects and complementary databases, the techniques for analyzing these indicators are mostly applied singularly and with joint interpretations made by "adding up" the individual summaries. Although this practice has important scientific relevance, there are possibilities for improving it. It is in this gap that Data Science enters the method on screen.

Data Science is introduced as a fundamental element of this methodological proposal, interpreting the array of indicators collectively instead of in isolation. This interdisciplinary field, which has permeated social and human sciences, deals with the systematic study of large-scale data encompassing volume,

diversity, veracity, and velocity, allowing for multifaceted problem analysis and the generation of valuable, solution-oriented outcomes, as detailed by Marquesone (2016).

Figure 1 encapsulates the Data Science analytical process. The initial stages—descriptive and diagnostic—are well-established in both the scientific community and public and private management. In recent years, with the intensification of data production and operational capacities for storing and processing it, there has been an increasing demand for analysis aimed at predicting what will happen. To do this, historically established conditions are taken as a basis and then what can be achieved, with the aim of altering undesirable results, such as through public policies.

Figure 1 | Synthesizing the analysis through Data Science.

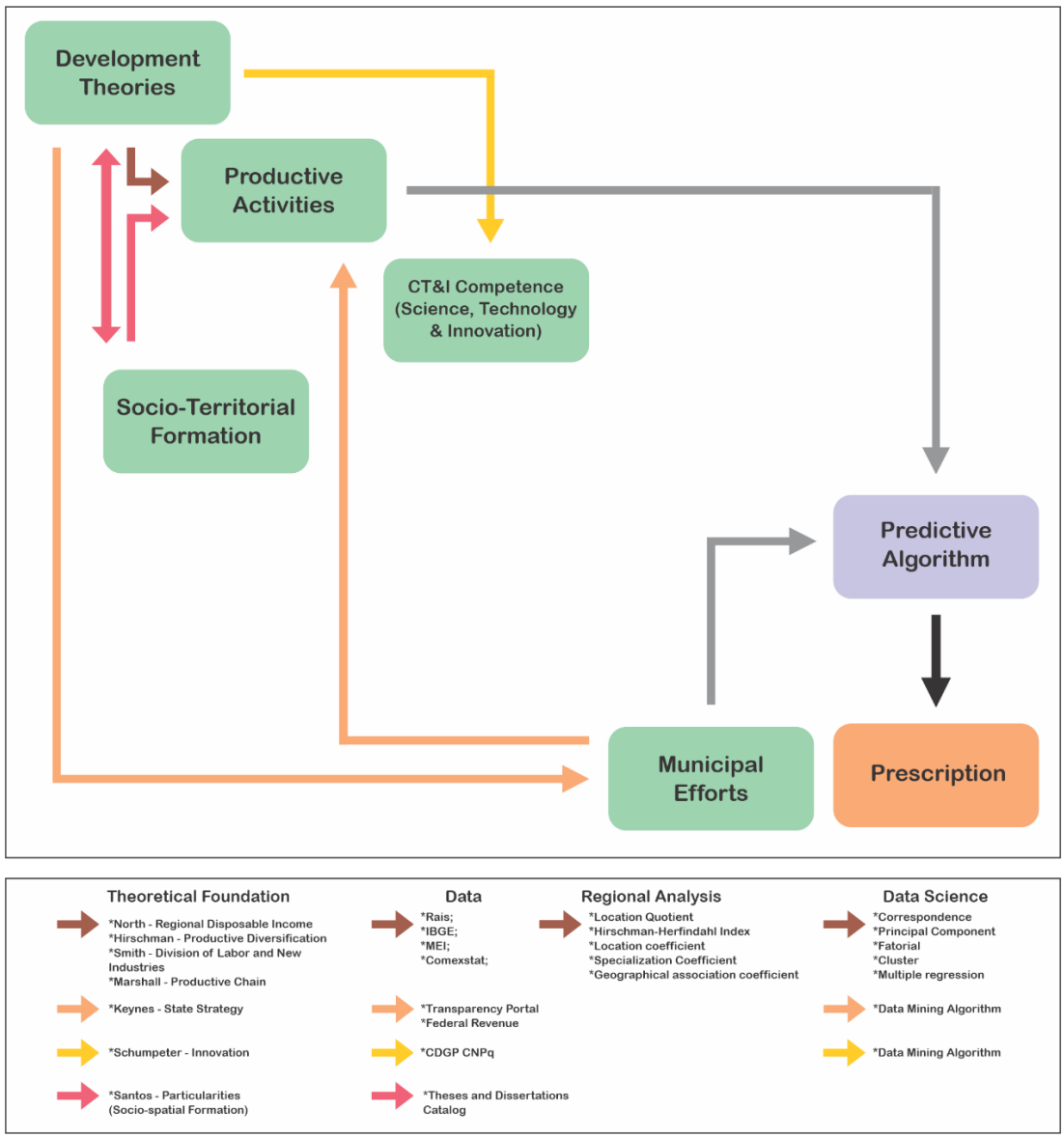
Source: Adapted from Verma *et al.* (2016).

The foundational theory for the proposed methodological framework emphasizes socio-economic diversification as a means to boost regional disposable income, using both direct and indirect public policy interventions. This approach is predicated on developmental theories and regional analysis. Data Science plays a critical role in this context, initially through exploratory analyses utilizing techniques such as correspondence, principal component, multifactorial, cluster, and multiple regression analyses. These techniques are applied to regional analysis indicators to uncover the most effective explanatory and predictive models for the diversification and enlargement of regional income.

The application of the aforementioned techniques, starting with exploratory analysis and progressing to derive meaningful insights, is warranted by the extensive array of both raw and synthesized indicators that are instrumental for an in-depth understanding of the municipality/

region's dynamic functions. The detailed examination of each indicator, such as the CNAE's 1,357 subclasses, necessitates the sophisticated analytical tools provided by Data Science. Figure 2 consolidates the concepts discussed previously, integrating the information outlined and visually representing the methodological pathway.

Figure 2 | Synthesizing the theoretical and methodological proposal.



Source: Produced by the authors.

RESULTS AND DISCUSSION

The methodological approach we have introduced explores the potential of Data Science in enhancing public policies for local and regional development (Boisier, 2000). It provides concrete justifications for utilizing the outlined mechanisms while acknowledging limitations and considerations for future applications. This section delineates these focal points: i) data mining and systematization; ii) application of regional analysis techniques; and iii) strategic prescription.

Regarding data mining and systematization, the data recommended for constructing the analytical matrix include that from the microdata of Rais and MEI. Specifically, this refers to Rais employment relationships and data on establishments. These microdata are categorized by Brazil's mesoregions, meaning researchers, irrespective of the municipalities involved, must initially manage significant numbers of jobs. For instance, there were nearly 24 million active jobs in the Southeast in 2021, detailed through about 60 variables encompassing more than 10,000 possible attributes.

Data Science emerges as a pivotal tool for database exploration and utilization, especially in discerning the dynamic roles of economic activities within a region, as previously mentioned. Utilizing specialized software such as RStudio, researchers can implement the necessary filters to refine their research and effectively utilize standard hardware for research activities.

In municipal data mining, standardizing data retrieval from transparency portals remains challenging due to the lack of uniformity across municipalities. Furthermore, typically, there are no agile mechanisms to obtain a large quantity of information. Hence, the creation of specific computational algorithms to capture the data on municipal commitments and filter out relevant information for the research—such as values, destinations, and the CNPJ of contracted companies—proves to be a feasible element for applying the methodology.

However, linking CNPJ data with the Federal Revenue Service database to verify the main activity and location of the contracted company is only possible, through the revenue platform, by searching one CNPJ at a time. The large number of commitments from a city hall—even of small size—makes this task impossible without the support of an automated data mining mechanism as provided by Data Science. In this regard, this data aids in diagnosis and shows the possibility of redirecting public spending that can induce the maintenance of an effective local/regional demand (Keynes, 1985), corroborating the

theoretical presuppositions of productive diversification.

The use of regional analysis techniques in Data Science aids in uncovering trends, patterns, and hidden correlations. This is beneficial both exploratorily throughout the research and in analyzing local/regional productive sectors, particularly for predictive purposes. The vast quantity of data and variables proposed for use and the operationalization of indicators that reinterpret classic theories is a complex task. Implementing a single statistical tool, such as trend identification that includes regional analysis indicators based on diverse databases, is intricate and often beyond the capabilities of conventional software and hardware used in everyday scientific research. However, employing data architecture and cloud processing capabilities intrinsic to Data Science makes this intersection possible.

Pragmatically, another advantage of informing public policy using this method is the statistical prediction of returns to regional income. This aligns with the theoretical and methodological intentions, facilitating strategic actions within government procurement to pinpoint and foster economic activities with potential linkages to catalyze production chains (Hirschman, 1958; Marshall, 1996). In conclusion, considering the potential of our method, a strategic prescription is paramount in enhancing public policies for local/regional development. Despite the risks to researchers, it is essential to pragmatically suggest alternatives to achieve the desired outcomes. Grounded in theoretical insights, regional analysis mechanisms, and statistical predictions, the formulation of public policies sharpens the allocation of scarce public resources and minimizes the reliance on uninspired or theoretically ungrounded efforts.

FINAL CONSIDERATIONS

The method detailed in this paper is robust, technically oriented, and anchored in a comprehensive theoretical framework and regional analysis, utilizing dependable data from credible sources. This approach aims to provide public authorities - whether at municipal or regional level—with the ability to address specific development issues, with an emphasis on increasing disposable income and improving individual, social and spatial aspects.

The qualification of public policies based on the application of the proposed methodology can be seen in different ways that complement each other. Three are mentioned. The first is strategic action to allocate public efforts and resources. As has been pointed out throughout the article,

there are many ways to intervene in the local/regional reality. The proposal put forward outlines one aspect; therefore, with a translucent initial approach, the endless trajectories of public authority action find concrete alternatives which, when encouraged, can generate local/regional benefits in the medium and long term.

Secondly, the method allows for tangible monitoring and evaluation of outcomes. Employing predictive indicators for application and returns establishes social oversight mechanisms for regional public initiatives, delivering concrete benchmarks for success. Lastly, the application of Data Science in refining public policies for local and regional development is validated by its strengths in thorough analysis, precise forecasting, and resource optimization. Data Science facilitates the mining and organization of extensive data sets, simplifying the discovery of trends, concealed patterns, and correlations essential for a holistic regional examination. Moreover, Data Science's predictive modeling capabilities enable the anticipation of strategic action outcomes, thereby enhancing the policy-making process with data-driven insights.

Nonetheless, it is vital to acknowledge the inherent limitations of this proposal. Its primary focus is on the technical stages of public policy formation, sidelining the initial decision-making, the objectives of the intervention, and the practical enactment—all inherently political elements. The pragmatic implementation of this method encounters significant hurdles, such as the need for skilled personnel and the appropriate technical infrastructure to manage the extensive data compilation, tabulation, and analysis. Additionally, the heterogeneous availability of data across regions precludes a one-size-fits-all approach to algorithm development, necessitating customization.

For future research, we advocate for a closer constructive collaboration between the technical and political dimensions and an examination of case studies that validate the effectiveness of this method in fostering local and regional development over time. This is imperative as the envisioned outcomes of the method's application are projected to materialize in the medium to long term.

ACKNOWLEDGEMENTS

The authors acknowledge the National Council for Scientific and Technological Development – CNPq - Brazil - (CNPq) for the grant to this research project (process: 307567/2022-2).

REFERENCES

ARAUJO, Tania Bacelar de. **Brasil:** territorialidade e políticas públicas: curso de ambientação para analista técnico de políticas sociais. Brasília: ENAP, 2013.

BITENCOURT, Rosimeire Sedrez; GUIMARAES, Lia Buarque de Macedo. Aplicação do coeficiente de Gini locacional e do Quociente locacional como apoio à delimitação geográfica de Sistemas locais de produção: o setor coureiro calçadista doRs. In: **XXXII Encontro nacional de engenharia de produção.** Bento Gonçalves/RS, 2012.

BOISIER, Sergio. Desarrollo (Local): ¿ de qué estamos hablando? In: BECKER, Dinizar Fermiano.; BANDEIRA, Pedro Silveira. (Orgs.) **Desenvolvimento Local-Regional**: Determinantes e desafios contemporâneos, v. 1. Santa Cruz: Edunisc, 2000.

CROCCO, Marco Aurélio**. Metodologia de identificação de aglomerações produtivas locais.** Economia e sociedade brasileiras. Revista nova econ. vol.16 n.2 Belo Horizonte,2006.

HIRSCHMAN, Albert Otto. **The strategy of economic development**. New Haven: Yale University Press, 1958.

KEYNES, John Maynard. **A teoria geral do emprego, do juro e da moeda**. Os Economistas. São Paulo: Nova Cultura, 1985.

MARQUESONE, Rosângela. **Big Data**: Técnicas e Tecnologias para Extração de Valores dos Dados, 1. ed. São Paulo – SP: Editora Casa do Código, 2016.

MARSHALL, Alfred. **Princípios de economia**. São Paulo: Nova Cultural, 1996.

MATTEI, Taise Fátima; MATTEI, Tatiane Salete. **Métodos de Análise Regional**: um estudo de localização. Revista paranaense de desenvolvimento, Curitiba, v.38, n.133, 2017.

NORTH, Douglass. **A agricultura no crescimento econômico**. Journal of Farm Economics, nº 41(5), 1959.

_____. **Teoria da localização e crescimento econômico regional**. Journal of Political Economy, nº63(3), 1955.

PAIVA, Carlos Águedo Nagel. **Fundamentos da análise do planejamento de economias regionais**. Foz do Iguaçú: Editora ParqueItaipú, 2013.

SANTOS, Milton. **Por uma geografia nova.** São Paulo: HUCITEC-EDUSP, 1978.

SCHUMPETER, Joseph Alois. **Teoria do desenvolvimento econômico**: uma investigação sobre lucros, capital, crédito, juro e o ciclo econômico. Traduzido por Maria Sílvia Possas. 3. ed. São Paulo: Nova Cultural, 1988

SMITH, Adam. **A riqueza das Nações**: investigação sobre sua natureza e suas causas. São Paulo: Abril cultural, 1983.

THEIS. Ivo Marcos. **Hic et nunc**: qual concepção de Desenvolvimento quando se trata de Desenvolvimento regional?. Revista Brasileira de Estudos Urbanos e Regionais, v.24, 2022.

VERMA, JaiPrakash *et al*. **Big data analytics**: challengesandapplications for text, audio, video, andsocial media data. InternationalJournalon Soft Computing, Artificial IntelligenceandApplications (IJSCAI), v.5, n.1, 2016.