Article

Geographical and Economic Factors Affecting the Spatial Distribution of Micro, Small, and Medium Enterprises: An Empirical Study of The Kujawsko-Pomorskie Region in Poland

Agnieszka Chłoń-Domińczak [1] **, Anna Fiedukowicz** [2,*] **and Robert Olszewski** [2]

[1] Institute of Statistics and Demography, SGH Warsaw School of Economics, al. Niepodległości 162, 02-554 Warsaw, Poland; achlon@sgh.waw.pl

[2] Department of Geodesy and Cartography, Warsaw University of Technology, Pl. Politechniki 1, 00-661 Warsaw, Poland; robert.olszewski@pw.edu.pl

* Correspondence: anna.fiedukowicz@pw.edu.pl; Tel.: +48-22-234-55-90

check for
updates

Abstract: Micro, small, and medium-sized enterprises (MSMEs) are an essential part of economies at the national, regional, and local levels. Understanding the determinants of the development of this sector is interesting not only for researchers but also for local governments to support the development of this sector. This paper analyses micro, small, and medium enterprises at the *gmina* (local) level in one region, the Kujawsko-Pomorskie voivodship (NUTS2) in Poland. The authors use multivariate linear regression, spatial econometrics, and classification trees to model the influence of different factors on the number of enterprises relative to population size. The authors found that the most crucial factor in all cases, independently of the method used, is the local government's revenue from personal income tax per capita. This finding, together with the lack of significance of variables related to the distance to technological parks or economic zones, indicates that the enterprises in the region produce mainly for local consumption and lack innovativeness. The authors also examined the influence of spatial context on the number of enterprises. The most important factor seems to be the percentage of built-up areas, but there are also others, depending on the model type; again, this confirms the local character of the activity of micro, small, and medium enterprises in the region. Variables representing the spatial context can explain the relative number of enterprises with coefficient of determination (R^2) between 0.30 and 0.45, which shows that this context played a relatively significant role in the development of the MSME sector in the region. On the other hand, the econometric models (that include the neighborhood) are only significant (improving R2) for medium enterprises, which means that medium enterprises expand their activity beyond the local range.

Keywords: micro; small and medium-sized enterprises; small and medium-sized enterprises; local development; administrative data; spatial distribution; spatial context

1. Introduction

Micro, small, and medium-sized enterprises (MSMEs) are an essential part of economies at the national, regional, and local levels. They provide jobs, create investment opportunities, and develop economic potential required for development. Therefore, the literature has researched the understanding of the determinants of the development of this sector. As emphasized by [1], a region facilitates the formation and transmission of social capital as well as geographic platform from

which social capital can be accessed. Geographical location also affects spatial disparity in economic activity [2].

The theoretical developments in the new economic geography aim at increasing the understanding of spatial perspectives and economic growth. Within its framework, it is expected that existing institutional infrastructure measured as industry density can generate spillovers that contribute to entrepreneurship development. Moreover, the variation in the firm formation is higher across regions than in time [3]. Boschma and Frenken [4] described how evolutionary economics may contribute to a new and more dynamic understanding of the location of an industry. They pay attention to the role of geography and firm location behavior as being price signals (neoclassical) and place-specific institutions as conditioning the range of possible (location) behaviors and potential locations, but not determining the actual (location) behavior and locational outcomes. This is due to several factors. First, the geographical concentration of industrial activities can generate agglomeration economies fostering start-ups and innovation and, possibly, the birth of a related industry in the region. Second, it increases the level of competition and makes the exits of firms raise the average fitness of routines. Third, the spatial concentration of firms can also affect the opportunities of collective action. Place-specific institutions (the institutional view) only condition the range of possible behaviors and potential locations of firms, but the actual behavior and location is largely determined by organizational routines acquired in the past. Changing these routines is possible with innovation and relocation. In the article, the authors analyze the spatial distribution of enterprises and their determinants. Such analysis contributes to understanding the characteristics of the MSME sector in the region, which can support recommending directions for further development of the enterprises, including their innovativeness.

Furthermore, [5] claims that territories can only be called relevant and meaningful units when the idea of routines and competences can be transferred from the organizational to the regional level. The region has become an entity on its own, providing intangible and non-tradable assets based on a unique knowledge and institutional base, which is not accessible for non-local firms. In the article, the authors investigate the structure at the most local municipality level, which can provide more insight into the intra-regional variation in enterprise structure. The snapshot of the existing distribution of the MSMEs and to what extent it is affected by the spatial and economic context can inform us about the factors that were underlying the establishment of such structure in the past, which, given the assumed path dependency, also affects current developments.

An important explanation of the spatial distribution of the MSMEs are localization economies that arise from a spatial clustering of economic activities in either the same sector or related industries. As shown by [6,7], localization economies, including agglomeration and urbanization effects, may be decisive in explaining the emerging spatial pattern of industries. In particular, the clustering of small firms, related firms, and supplier/customer linkage can help to construct a favorable environment for manufacturing entrepreneurship. Furthermore, [8] argue that spatial clustering is at the core of economic geography, highlighting the impact of proximity and distance, institutions, and local milieus on economic processes. Spatial clusters induce variation, observability, and comparability, while at the same time allowing increased differentiation without discouraging knowledge exchange. They show that the spatial attributes of interactive learning and innovation process could be a fruitful point of departure not only when analyzing spatial agglomerations, but also when it comes to re-invigorating research into economic geography.

Clusters emerge due to the benefits of proximity. Proximity is discussed by [9], who analyses five dimensions of proximity (i.e., cognitive, organizational, social, institutional, and geographical). It is claimed that geographical proximity combined with some level of cognitive proximity is sufficient for interactive learning to take place. Geographical proximity may also stimulate the formation and evolution of institutions, such as norms and habits that may affect interactive learning and innovation. Too little proximity leads to the lack of spatial externalities. Yet, too much geographical proximity may lead to a lack of geographical openness. This can particularly apply to micro and small companies, which have fewer resources to reach out beyond the local context of activity. Overall,

geographical proximity may facilitate interorganizational learning, but it is neither a necessary nor a sufficient condition. Understanding the diverse patterns of geographical proximity at the local level is therefore an important factor contributing to the identification of possible solutions aimed at enhancing innovation.

The degree of urbanization affects localization economies. Spatial agglomerations create localization advantages in terms of spillovers and cooperation between firms. The benefit of urbanization comes, among others, from lower transport costs and proximity to suppliers and customers [10–12]. Access to customers can be proxied by the degree of urbanization. One of the important characteristics related to urbanization is access to transportation infrastructure, which has been found to have a significant effect on urban growth [13].

The firm size also matters. Companies with different sizes have different characteristics, both with respect to sectoral structure but also the ability to innovate. Studies show that there is a negative relationship between firm size and employees' tendency to become entrepreneurs (e.g., [14–16]). Ref. [17] shows that the occurrence of the small firm effect is closely related to low entry costs in industries typically clustered by small firms. Not only do the industry entry conditions of small parent firms have a strong and positive impact on the likelihood of employees' entrepreneurial entry, but entrepreneurs spawned by small firms show a strong preference for starting a business in the exact same industry as their previous employer. Additionally, evidence in the literature shows that there is a link between firm size and innovative performance [18–20]. Finally, smaller firms face more obstacles than larger firms with financing, taxes and regulations, inflation, or anti-competitive practices. For these obstacles, small firms have the biggest problems, followed by medium-sized and large firms [21]. Moreover, [22] analysis by firm size shows that industrial structure and corporate organization affect the benefits that arise from clustering within a given industry. This finding is strongest with regard to the size distribution of establishments; own-industry at small establishments presents a much greater attraction to potential new arrivals than does a comparable level of own–industry employment at larger establishments.

Multiple factors, both at the regional and local levels, affect the creation process of new firms, their survival, and their innovative capacity. The evidence in the literature indicates that these include, among others, population growth, a high percentage share in the population of people with high managerial and vocational skills, urban concentration, household wealth, and demand [23]. Factors that are conducive to the development of small and medium-sized enterprises (SMEs) differ locally. Evidence from Finland shows that urban and rural communities provide different environments for enterprise development, particularly with regards to human capital, access to technology, the number of cluster enterprises, and the intensity of communal cooperation [23]. However, in their analysis of SMEs' development in the UK, [24] conclude that business growth is possible under different territorial conditions, including different levels of competition and market demand between regions and differences in the occupational and skill structure of the labor market. Many SMEs in peripheral regions may actively work to develop strategies to overcome these constraints. Therefore, an initial locational disadvantage may ultimately benefit rather than inhibit a company's growth and performance. The growth of indigenous companies in peripheral regions results from more active and autonomous roles of the entrepreneurs as well as local governments in creating an entrepreneurship-friendly environment. In turn, this can contribute to the economic and social development of the local areas [25,26]. These considerations can also be placed in the context of the role of the MSMEs in the economic transition in Central and Eastern Europe (CEE). In recent years, there are three major spatial developments that can be observed in the CEE countries [27]. First, there are increasing differences in development between urban core regions that develop better than peripheral rural regions. Second, there are strong trends towards polarization between their main metropolitan area (usually the national capital) and the rest of the country. Third, there is an east-west gradient, with the western parts performing better than eastern regions, which is also confirmed by [28]. As shown by [29], there is a risk that spatial development further concentrates in a smaller number of (metropolitan) regions,

whereas more and more other regions might be affected by the processes of peripheralization. Ref. [30] show that the regional inequalities in CEE countries are strong and persistent. Weaker CEE regions typically lost the greatest part of their industrial base which, being in capital-intensive sectors, was more exposed to international competition; as a result they are more exposed to recession shocks, which gains importance given the recent COVID-19 pandemic situation. An analysis of the non-core regions in CEE by [31] shows that, after the financial crisis, the increasing number of SMEs, along with substantial R&D outlays and the development of human capital, were important stimuli for development. Analyses at the CEE intrastate level conducted by [32] show that while there was general catching-up with the EU–15 average by the state economies, there was also growing economic diversification between regions in the studied countries (internal divergence). Analysis of [33] indicates that, following the Williamson's curve, disparities between regions are lower in the early stages of development, peak in the middle-income stages, and diminish again as a country becomes wealthy. Development policies must not focus extensively on the country as a whole but have to take into account the preferences and possibilities of their peripheral regions as well. To that end, [34] show that while considering business stimulation policies, both the quantity and the quality of the new firm start-ups should be taken into account.

The regional determinants of the growth of SMEs have also been investigated in Poland. According to [35], the human capital development, wages, unemployment, economic activity of the population, and disposable incomes are essential for the development of SMEs. The study of enterprises in southern Poland indicates that access to financial instruments at the regional level is an essential barrier to enterprise development [36]. Simultaneously, IT infrastructure, closeness to markets, suppliers, and cooperators are the most critical factors stimulating such development.

In this article, the authors focus on analyzing the local determinants of the spatial distribution of micro, medium, and small enterprises in the Kujawsko-Pomorskie region in Poland. This region for many years has lagged behind the economic development in the rest of the country. Its innovative capacity remains one of the lowest, both when taking into account human capital as well as investment in research and development [37]. Given these challenges, this article was developed as a part of action research activities that aim at co-learning leading to developing knowledge and understanding necessary to design effective policies aiming at boosting innovative MSME growth in the region [38] under the umbrella of the "REGIOGMINA" project, led by the regional authorities. As discussed by [39], the choice of incubation strategy depends on the characteristics of local areas. Therefore, providing knowledge on the differences and similarities between different municipalities is an important contribution to develop the regional strategy.

The number of micro, small, and medium-sized companies per 10,000 people of working age in the Kujawsko-Pomorskie region remains below the national average (Figure 1). In the case of micro and medium-sized companies, the gap between their number and the national average has widened in recent years, while in the case of small companies, it has remained stable. The distribution of the number of companies is asymmetric and the median number of companies is below average in the region. At the same time, there are several gminas, with the number of companies per 10,000 people of working age distinctly above the regional average. As [24] emphasize, the close analysis of such "paradoxical" cases can open up new perspectives for regional and local policies.
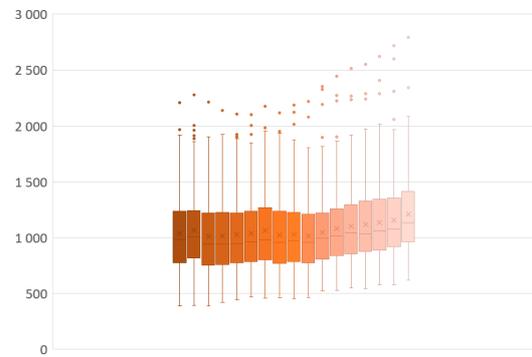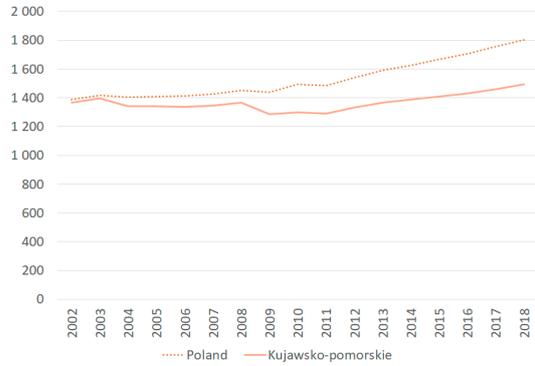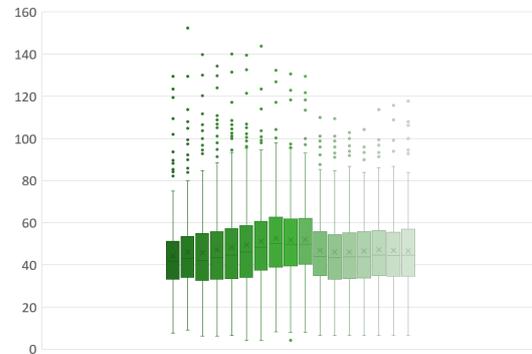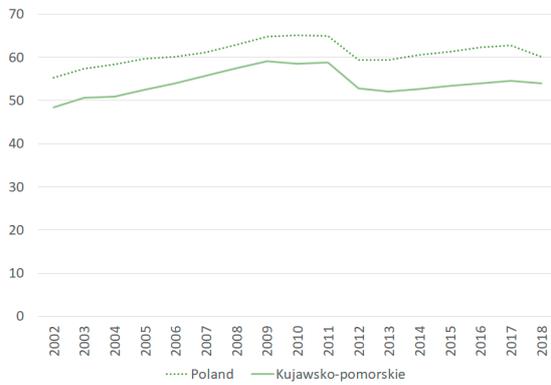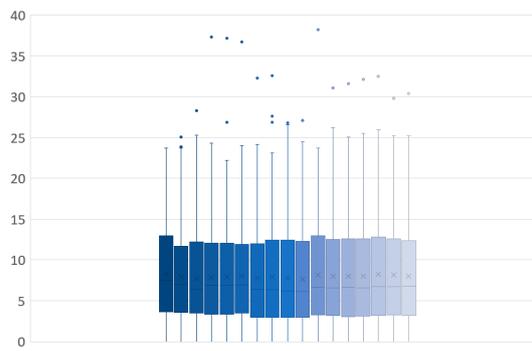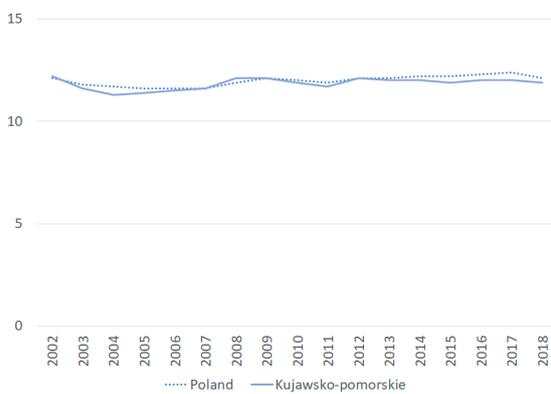
Figure 1. The number of micro, small, and medium-sized companies per 10,000 people of working age in the Kujawsko-Pomorskie region, 2002–2018 (source: Local Data Bank, Statistics Poland).

There are also differences in the sectoral structure of the MSMEs both depending on the size of the companies and at the municipal level. In larger firms, the share of industry is higher. There is also spatial variation—there is a clear gradient between the shares of firms in industry and services, which also depends on locality (Figure 2).

Sectoral structure by enterprise size (NACE[1] sections)

Share of enterprises in services and industry in municipalities

Ratio of enterprises (payers) in services to industry in the municipalities of the Kujawsko-Pomorskie region
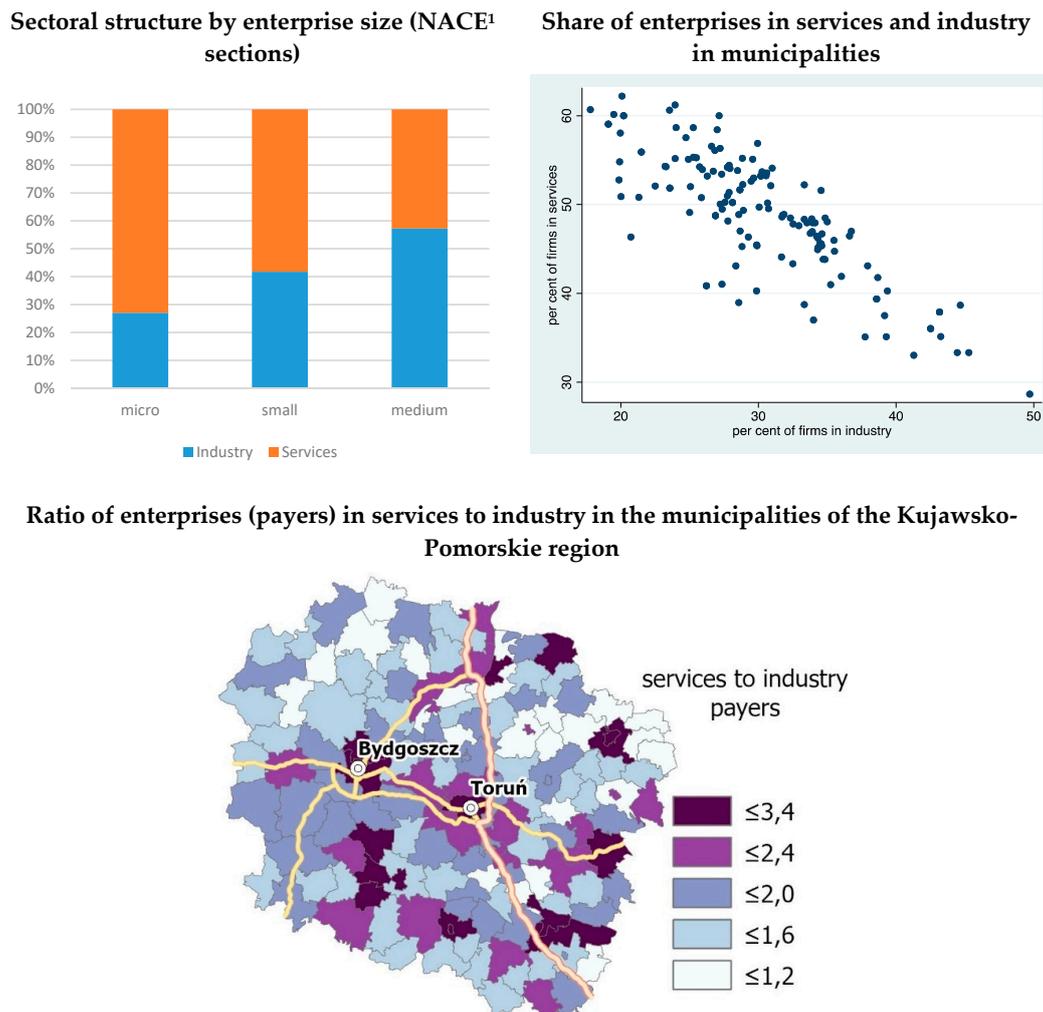
Figure 2. Sectoral distribution of enterprises by company size and municipality (source: own work based on Social Insurance Institution data). [1] Statistical classification of economic activities in the European Community.

The novelty of the authorial approach is threefold. Firstly, the authors focus on the lowest administrative level, gmina (meaning municipality in the urban context or commune in the rural context) as the main actor in creating a friendly environment for conducting business activity [40]. Secondly, the authors use a unique database of local-level enterprises obtained from an administrative register, namely the Social Insurance Institution database. Therefore, the authors use the best possible source of information on Poland's actual activity (and employment) of micro, medium, and small enterprises. Additionally, the authors use gmina-level data from the Local Data Bank of Statistics Poland. Thirdly, the authors also consider spatial factors to explain differences in the MSMEs' development in the Kujawsko-Pomorskie region. The two main research questions are:

- Which factors affect the current distribution of micro, small, and medium enterprises in the region?
- Are spatial factors important in explaining the local MSMEs' development in the region, and can they be successfully included in the analytical models?

In addition to a theoretical understanding of the factors determining the existing structure of micro, small, and medium companies, the authors also identify local areas with higher levels of entrepreneurship than is explained by the analyzed factors.

2. Data and Methods

2.1. Data

The register of Social Insurance Institution is the data source on the number of micro, small, and medium-sized enterprises at the *gmina* level. We define micro enterprises as those that have 1 to 9 employees, small as those with 10 to 49 employees, and medium as those with 50 to 249 employees. Employees are identified as those for whom the enterprises pay social insurance contributions. The information on the number of companies paying social insurance contributions was collected for December 2018 in the "REGIOGMINA" project. The use of the administrative register enables the identification of active companies based on their actual employment; therefore, it provides timely and accurate information on the size of the MSME sector and its structure at the local level. The analytical units are gminas in the Kujawsko-Pomorskie region. This choice forced the selection of explanatory variables that are available in such disaggregation. One should note that most of the statistical information on socio-economic development is provided at the powiat (district)(NUTS 4) or regional (NUTS 2) level.

The response variable adopted in all the methods was the number of micro, small, and medium-sized enterprises (calculated per 10,000 inhabitants of the gmina—Figure 3).
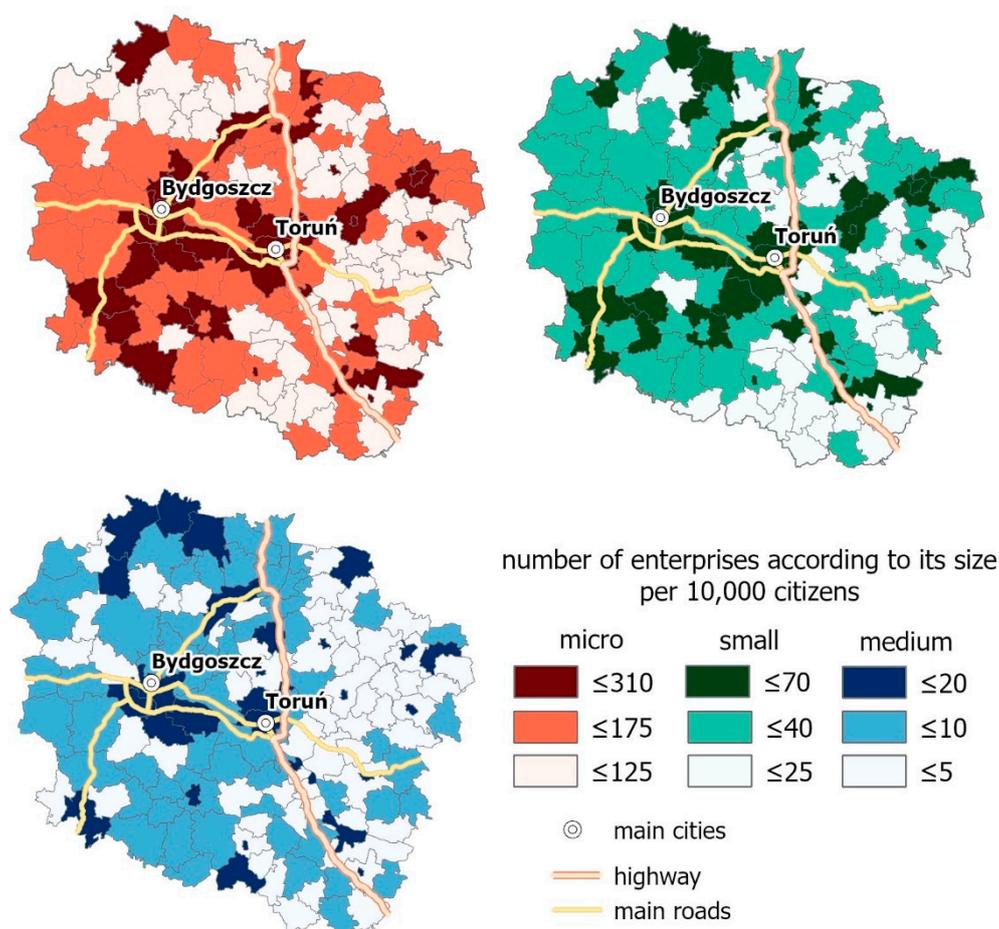

Figure 3. The number of enterprises per 10,000 inhabitants in the gminas of the Kujawsko-Pomorskie region (source: own work based on Social Insurance Institution and Statistics Poland data).

The Social Insurance Institution's registry has provided the number of enterprises. The authors used the registry data, as it includes up-to-date information on the number of people for whom an enterprise paid social security contributions in December 2018. This number differs from the data

provided by Statistics Poland: the enterprises do not regularly update it, so the latter shows a higher number of companies.

The explanatory variables come from the following sources:

- Statistics Poland data, including population and labor market characteristics as well as community wealth, using tax income as a proxy:

  ○　the number of people in gminas, including people in pre-productive, productive, and post-productive ages;
  ○　the number of registered unemployed (total, men, women);
  ○　the number of registered unemployed per 100 inhabitants (derived by the authors);
  ○　gminas' own revenues from personal income taxes (PIT) per capita;
  ○　local budgets' total expenditure per capita.

- Data from the local policy assessment supporting the entrepreneurship development, obtained from [41].

- Data from the Kujawsko-Pomorskie regional authorities: the number of projects supporting entrepreneurship development funded from the European Regional Development Fund (ERDF) and the amount of funding provided to these projects.

- Spatial data of the Head Office of Geodesy and Cartography from the Topographic Object Database with 1:10,000 Level of Detail regarding, in particular, the degree of urbanization. Following [42,43], we proxy the transportation infrastructure using the road density; we also take into account the characteristics of the landcover. Therefore, the authors include the following:

  ○　land cover: built-up areas (PTZB class, namely land cover: built-up);
  ○　land cover: agricultural areas (PTTR02, namely land cover: grassland and arable farming);
  ○　land cover: orchards (PTUT03, land cover: permanent cultivation);
  ○　transport network: roads (SKDR, namely transport route: roads);
  ○　transport network: railway tracks (SKTR, namely transport network: a rail or tracks);
  ○　location of large cities (from ADMS class, a territorial division unit: a town);
  ○　the geometry of administrative areas (from ADJA class, an administrative division unit).

- Spatial data from other sources (Figure 4):

  ○　Location of large enterprises in the voivodship, obtained using the ranking of the largest companies in Poland. Cooperation with large companies can be an essential lever for increasing the potential of smaller companies, especially in the innovative (technological and organizational) dimension. Large companies need smaller ones because they are more agile and can propose innovative solutions. MSMEs often better know the local markets on which they focus. Specialized MSMEs can meet the diverse and complex needs of large businesses [44].
  ○　Distance to science and technological parks, obtained using the information provided by the website, "Invest in Kujawsko-Pomorskie". Science and technology parks create a base for the commercialization of scientific research, research cooperation, and knowledge transfer, which are vital for the development of MSMEs' innovation and entrepreneurship. These parks offer, among others, management support, training services, venture capital access, intellectual property consultations, and laboratory services [45].
  ○　Distance to areas of the Pomeranian Special Economic Zone, obtained using the information from its website. Special Economic Zones are instruments that support the MSME sector. The zones assure favorable conditions for business activity and foreign investment. Foreign companies operating within the SEZ provide new business standards such as technology, experience in production processes, business contacts, and good practice in training

employees, which are exceedingly significant to the development of the SME sector; they are also their primary source of new technologies [46].

○   Distance to higher education institutions (HEIs) (source: National Court Register). HEIs are important knowledge alliance partners of the SMEs on the regional level; they constitute the source of tacit knowledge for innovative firms [47].

○   Location of the A1 highway exits (source: General Director for National Roads and Motorways): the road network on the local level is vital for economic development at both the local and regional levels, as accessibility is one of the main deciding factors in the location of new businesses [48]. In Poland, as [49] found, the more significant the investment in regional transport infrastructure, including national, regional, and local roads, the more visible the financial and economic outcomes of SMEs.
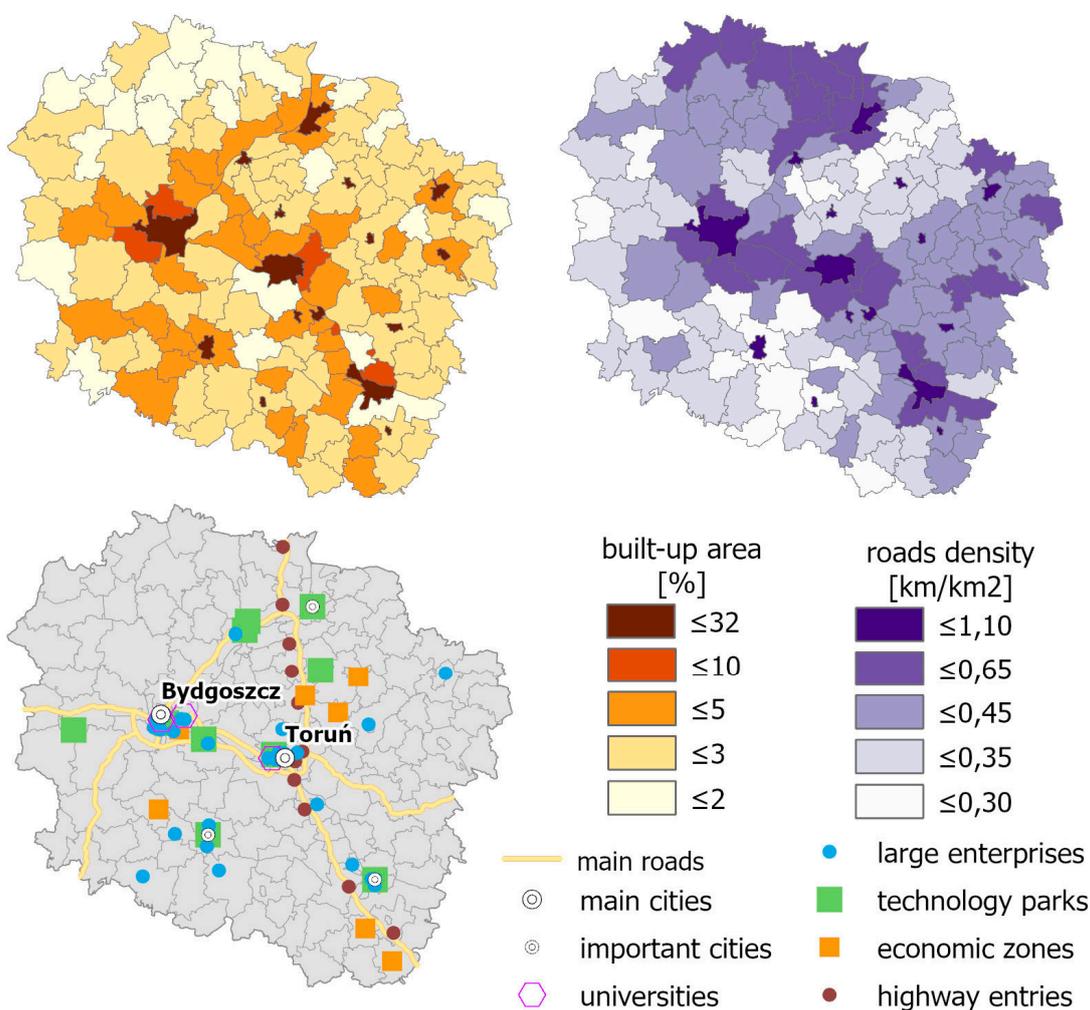

Figure 4. Selected variables derived from spatial data and the location of the more significant structures (source: own work based on data sources listed in text, especially BDOT10k).

Spatial data were not used directly; they were used to enrich the set of data on gminas directly with attributes constituting the formalization of the spatial context and spatial relations of or in the gmina. Consequently, there were additional attributes created, describing the following for each gmina:

- Percentage coverage of the gmina with a built-up area (Figure 3).
- Percentage coverage of the gmina with agricultural areas.

- Percentage coverage of the gmina with orchards.
- Road network density (Figure 3).
- Railway network density.
- Distances from the nearest of the following structures:

  ○ Bydgoszcz or Toruń (largest cities in the region with administrative functions—referred as "main cities" in Figures 2 and 3),
  ○ another large town,
  ○ Bydgoszcz, Toruń or another large town,
  ○ a university,
  ○ a large company,
  ○ a technology park,
  ○ an economic zone,
  ○ a highway entrance/exit,
  ○ key road infrastructure in the voivodship (national roads).

2.2. Methodology

The study of factors influencing the number of MSMEs in a gmina was conducted using several methods:

- studying the correlation of variables,
- multivariate regression,
- models of spatial econometrics,
- classification trees,

First, the authors checked each potential explanatory variable's **correlation** to the number of micro, small, and medium enterprises separately. As in the more populated areas the number of MSMEs is obviously larger, the authors used in their analysis the number of enterprises relative to the population (that is, per 10,000 people).

The next step was to create classical multivariate **regression models** through two methods:

- forward selection—beginning with an empty model and adding further explanatory variables, starting from the one that affects the explained model the most;
- backward elimination—starting with a model with all the variables then removing subsequent variables, starting from the variable with the least significance.

Subsequently, based on the sets of variables defined for the regression, **spatial econometric** models. Such models are also known as geographically weighted regression (GWR) and allow us to include the spatial heterogeneity of the variables in the analysis. In classical regression models, the spatial neighborhood influence is omitted, although it can have significant impact on the explained variable. GWR are particularly used in the modeling of economic indicators (i.e., unemployment rate [50]).

The authors built four GWR models which differ with regards to the type of model and the matrix of weights used. Regarding the type of model:

- a spatial lag model—assumes the influence of explanatory variables of neighbors on the response variable;
- a spatial error model—assumes the relationship between the model error of neighbors.

Regarding the method of determining the matrix of weights (defining the neighborhood) that determines the influence of particular municipalities' values on each other:

- a neighborhood-based matrix—a common border between the gminas indicates that a neighborhood exists;
- a distance-based matrix—the neighboring degree is inversely proportional to the distance between gminas.

In studying correlation and in each of the regression models and spatial econometrics, the authors use a p-value = 0.05 as the significance level.

The next step involved making models that explain the number of micro, small, and medium enterprises per person, in particular, gminas using classification trees (Classification And Regression Tree types, CART) [51,52]. A significant advantage of this type of model is that, unlike linear regression or the spatial econometrics based on it, there are no preliminary assumptions about the linearity of the model. It is also possible to use explanatory variables in the various levels of measurement (variables used in the nominal and ordinal levels were added to the previously used variables).

However, the use of classification trees required the discretization of the response variable. The authors decided on the discretization into three classes (low, medium, high) based on the Jenks natural breaks classification method. While this approach results in a loss of informational content of the data (a "downgrade" of the level of measurement), it does produce higher readability and a more straightforward interpretation of the models created this way. The obtained results can be presented in the form of several logical IF-THEN conditions.

Before determining the trees, the authors checked the validity of the available predictors using the chi-square test to evaluate the results. The five most significant predictors (determined independently for each model), a breakpoint condition based on the Gini index, and a minimum number of nodes of 5 were employed to construct CART.

3. Results

3.1. Correlation

Table 1 shows the correlation results. For any of the types of enterprises, there was no significant correlation between the area of the orchards, the distance from large cities (excluding voivodship cities), the distance to the A1 highway exit, the key road infrastructure in the voivodship, or the expenditure per capita in the gminas.

The strongest correlation with the response variables was the gmina's own revenue from the PIT per capita, which indicates the wealth of the inhabitants of a given gmina. For all enterprises, there are strong and positive correlations between the percentage of built-up area and the density of the road network. Indirectly, both of these parameters show the urbanization level of a given gmina. It is also easy to notice that there were the fewest significant variables found for small enterprises.

However, one should note that the correlation values calculated in this way only indicate the relationship (or the lack of thereof) between pairs of variables, thus providing very general and limited information. One cannot exclude the existence of more complicated relationships, where, e.g., a combination of two variables that have insignificant correlations with a dependent variable will correlate with it significantly. Therefore, in the subsequent part of the study, regression models, spatial econometrics, and classification trees were created using more variables at the same time.

Table 1. Explanatory variables and their correlation with the response variables (the number of enterprises of a given size per 10,000 inhabitants of the gmina). Statistically significant values (p < 0.05) are in bold and with an asterisk (source: own analysis). [1] variables in an ordinal scale, used only in classification trees.

Variable	v. Name	Micro	Small	Medium
built-up area [%]	G_Built	**0.44 ***	**0.30 ***	**0.56 ***
agricultural area [%]	G_Agric	**−0.40 ***	**−0.33 ***	**−0.41 ***
orchard area [%]	G_Orch	0.14	0.07	0.13
road density [km/km2]	G_road	**0.49 ***	**0.39 ***	**0.54 ***
railway density [km/km2]	G_railw	**0.22 ***	0.09	**0.32 ***
Bydgoszcz or Toruń dist. [km]	D_BT	**−0.33 ***	**−0.19 ***	−0.15
other important cities dist. [km]	D_Cities	−0.02	0.04	−0.06
B., T. or other cities dist. [km]	D_BTC	−0.33	−0.19	−0.15
large enterprises dist. [km]	D_LE	**−0.35 ***	**−0.25 ***	**−0.13 ***
technology parks dist. [km]	D_TP	**−0.30 ***	−0.13	**−0.19 ***
special economic zones dist. [km]	D_EZ	**−0.20 ***	−0.03	0.00
universities and HEIs dist. [m]	D_Uni	**−0.34 ***	**−0.20 ***	**−0.16 ***
A1 highway exit dist. [km]	D_A1	0.04	0.10	−0.04
main roads dist. [km]	D_mroad	**−0.18 ***	−0.05	−0.08
population	pop	**0.26 ***	0.13	**0.23 ***
population in pre-productive age (0–17)	pop_pre	**0.28 ***	0.14	**0.23 ***
population in productive age (18–59/64)	pop_pro	**0.27 ***	0.13	**0.23 ***
population in post-productive age (60/65 and over)	pop_post	**0.24 ***	0.11	**0.22 ***
unemployed people (total)	unempl	**0.25 ***	0.07	**0.28 ***
unemployed women	une_w	**0.25 ***	0.07	**0.28 ***
unemployed men	une_m	**0.24 ***	0.07	**0.28 ***
unemployed per 100 population	un_p	**−0.40 ***	**−0.41 ***	−0.14
unemployed women per 100 population	uw_p	**−0.44 ***	**−0.42 ***	**−0.21 ***
unemployed men per 100 population	um_p	**−0.32 ***	**−0.36 ***	−0.04
gmina's own revenue from PIT per capita	income	**0.77 ***	**0.63 ***	**0.49 ***
gmina's expenditure per capita	exp	0.01	0.04	−0.01
local policy supporting the enterprise development	supp	**0.26 ***	0.08	**0.24 ***
the number of ERDF-funded projects supporting the enterprise development	proj_n	**0.25 ***	0.12	**0.21 ***
the value of ERDF-funded projects supporting the enterprise development	proj_v	**0.20 ***	0.10	0.16
evaluation by enterprises	eval	**0.28 ***	0.14	**0.19 ***
population class[1]	pop_cl		ordinal data	
[small, medium, high]		≤10,000	≤14,000	>14,000
type of *gmina* [1]	type			
[rural, urban-rural, urban]				

3.2. Regression

3.2.1. Regression with Economic and Spatial Explanatory Variables

The authors created a series of regression models using forward selection and backward elimination methods based on the numeric variables listed in Table 1. Table 2 shows a summary of the results obtained.

In general, models created using the forward selection method typically had a smaller number of variables used in the model (and thus were simpler). Simultaneously, they had only a slightly lower coefficient of determination, and there was no problem with the mutual correlations of the explanatory variables. Instead, there occurred the problem of heteroscedasticity, which indicates that the factors not explained by the model are not random. The models for small enterprises were an exception to this rule, while backward elimination left one with a single explanatory variable.

It was possible to explain the most variability (around 2/3) for the enterprises of the smallest (micro) variability. The models fit slightly worse for small enterprises (around half of the variability), while models for medium-sized enterprises had the lowest coefficient of determination R^2 (improved by using spatial econometrics—see Section 3.3).

A gmina's revenue from personal income tax played a vital role in all the models made for every size of the enterprise; it positively correlated with the number of enterprises. The percentage of the built-up area in a gmina was another frequently occurring variable. As for the other variables, the models most often differ according to the enterprise's size and the way a model has been created.

Table 2. Different variants of created regression models—a summary (source: own analysis).

Parameter/v.Name	Micro		Small		Medium	
	FS	BE	FS	BE	FS	BE
R^2	**0.63**	**0.68**	**0.49**	**0.40**	**0.41**	**0.43**
constant	71.93	71.48	7.90	16.55	1.08	1.39
G_Built	1.52	2.25			0.26	0.27
G_road			20.2			
G_railw		−171.6				
D_BTC						0.27
D_LE					0.05	
D_EZ		−0.36				
D_Uni						−0.25
D_A1		0.28	0.09			
pop_pro		0.004				
pop_post		−0.01				
une_m	−0.03		−0.01			
income from PIT per capita	0.11	0.10	0.03	0.03	0.005	0.0003
standard error	26.3	24.9	8.6	9.2	3.0	3.0
multicollinearity	no	yes	no	no	no	yes
heteroscedasticity	yes	no	yes	no	yes	no
lag significance	no	no	no	no	without D_LE	neigh+dist
error significance	no	no	no	no	no	neigh+dist

The standard errors of the forward selection models ranged from 18% of the mean value for micro-enterprises through 26% for small enterprises to nearly 47% for medium enterprises. Figure 5 presents the number of enterprises per 10,000 inhabitants, projected according to these models and the residuals from regression models as a percentage of the expected value.
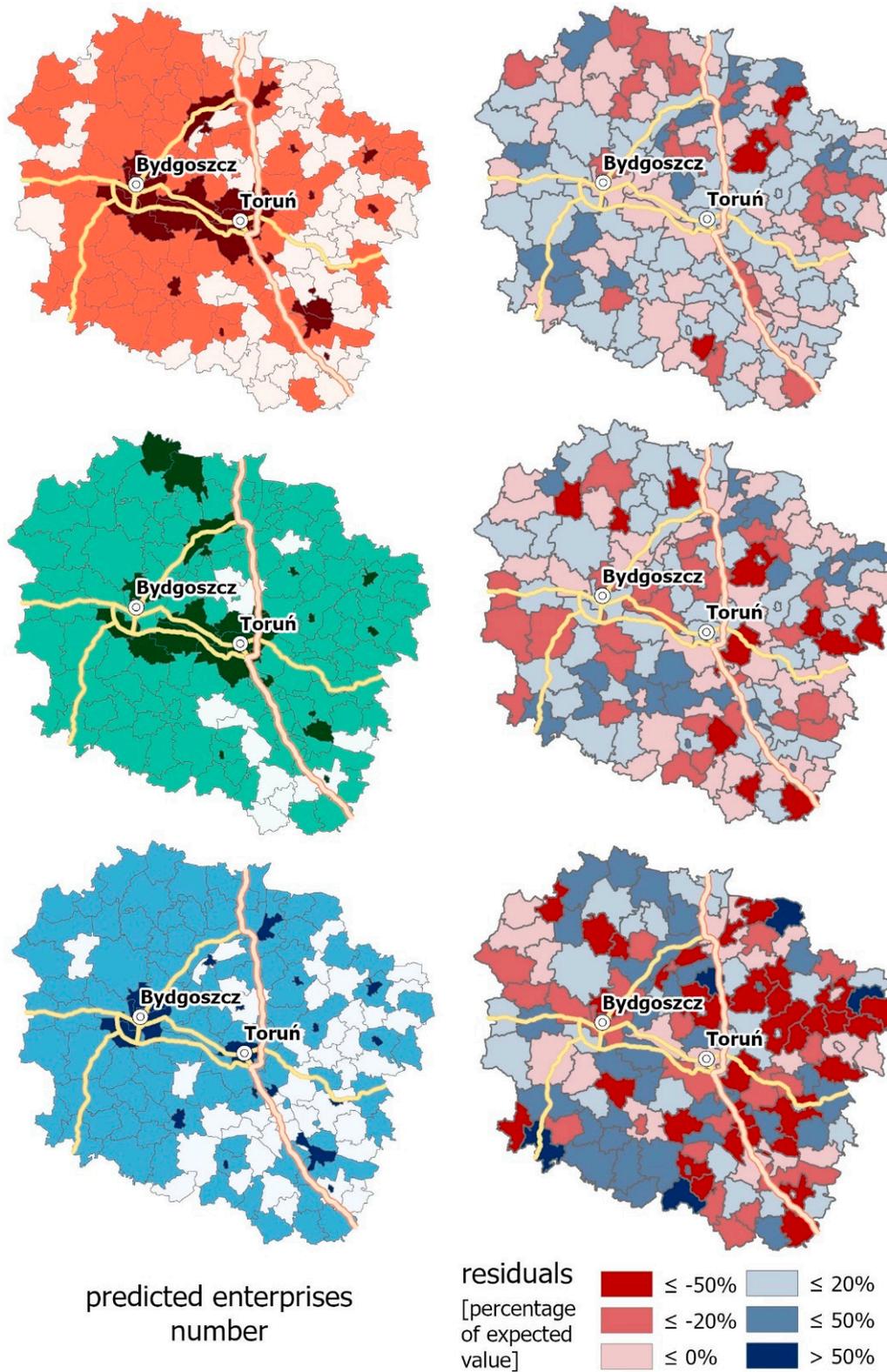
Figure 5. Results of the forward selection models and residuals from regression. The scale and color ranges for the predicted number of enterprises are identical to those in Figure 2 (source: own analysis).

These results identify the local communities, where the relative size of the analyzed group of enterprises is significantly above or below the level explained by the model. The further analysis of the local determinants might indicate which conditions accelerate or hinder entrepreneurship development.

Additionally, the authors performed regression models based solely on spatial variables. As a result, it was possible to check the extent to which the number of enterprises relative to population can be explained without referring to economic data. The authors used the backward elimination method; Table 3 presents the results.

Table 3. Regression models based solely on spatial data—a summary (source: own analysis).

v.Name	Micro	Small	Medium
R^2	**0.45**	**0.31**	**0.39**
constant	99.53	16.87	7.13
G_Built			0.28
G_Agric			−0.035
G_road	155.6	42.6	
G_railw	−174.7	−61.3	
D_BTC		0.79	0.30
D_LE	−0.96	−0.24	
D_TP	−1.23		
D_Uni		−0.82	−0.30
D_A1	0.88	0.17	
standard error	32.4	10.0	3.1
multicollinearity	no	yes	yes
heteroscedasticity	no	no	no
lag significance	no	no	neigh+dist
error significance	no	no	neigh+dist

It is worth noting that, despite the use of a minimal set of output variables, the R^2 coefficient dropped by about 1/3 and remains between 0.31 and 0.45, depending on the enterprise's size. Thus, it is possible to explain this part of the variability by analyzing only the spatial environment of a given gmina without information—e.g., on its own revenue from PIT per capita (although it was the most significant of the correlates). It is also important to note the differences between the models for different sizes of enterprises. For micro and small enterprises, the density of the road network, the distance to the highway interchange (positively), and the density of the rail network (negatively) were relevant. For medium-sized enterprises, it was the percentage of built-up and agricultural areas. The proximity of a large urban center (including the voivodship cities) had a positive impact on the number of small and medium enterprises. Figure 6 shows the values predicted by the models based solely on spatial variables and their residuals as a percentage of the predicted value.

In all the regression models, there were much bigger negative residuals then positive ones. Positive residuals rarely exceeded 50% of the expected number of enterprises (for the models in Figure 5, only for medium enterprises). At the same time, negative residuals (that is, underestimating the number of enterprises per capita) appeared more often and some (single) extreme outliers came off.
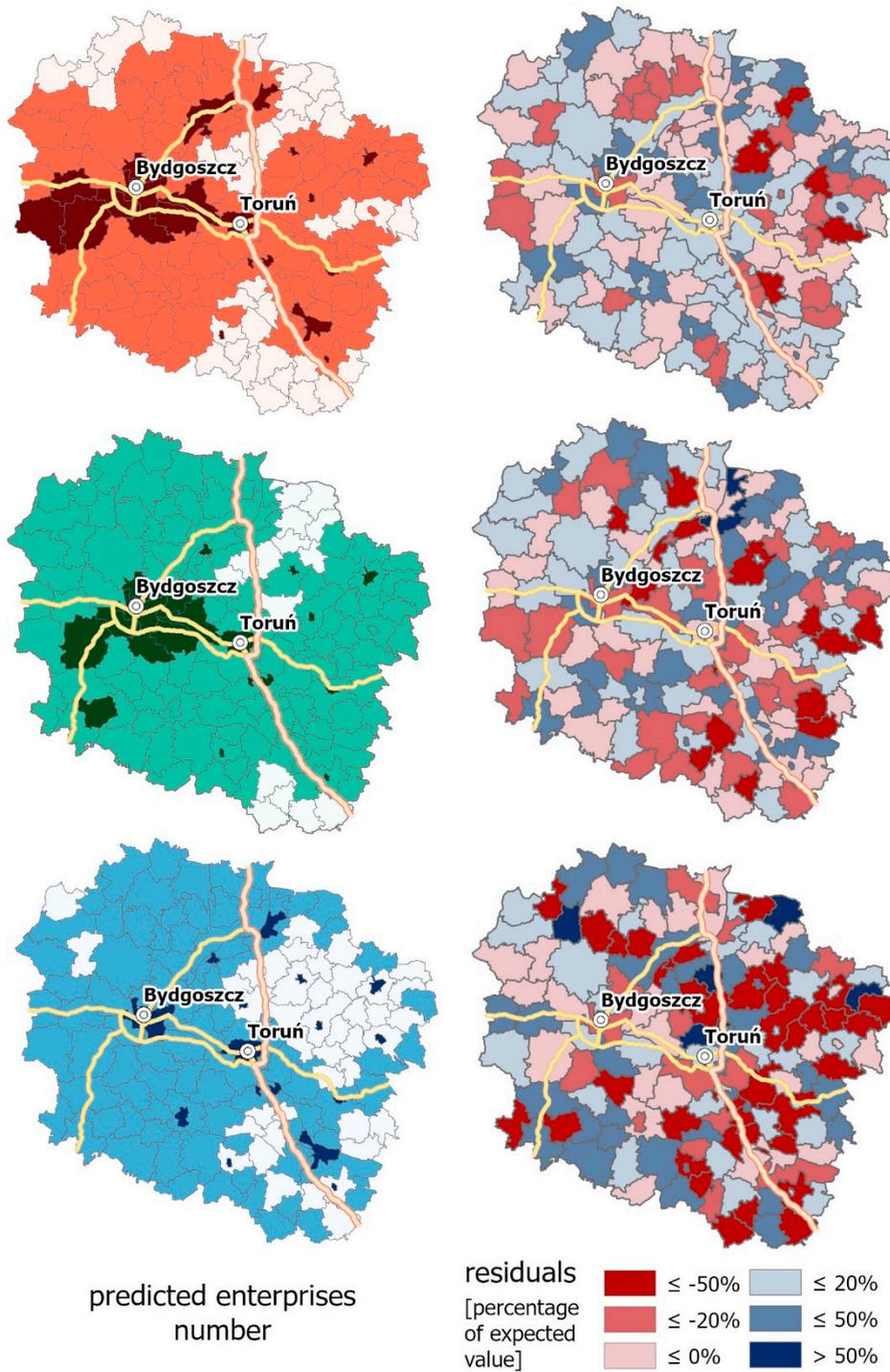
Figure 6. Results of geometric models and the residuals from regression. The scale and color ranges for the expected number of enterprises are identical to those in Figure 2 (source: own analysis).

3.2.2. Spatial Econometrics

For most models, the spatial econometric models (both lag and error) were not significant (regardless of the neighborhood matrix used). The models for medium-sized enterprises were an exception—in most cases, it was possible to improve the model prediction using information about the gmina's neighborhood (already indicated in Tables 2 and 3). In all the cases, models using a matrix of weights based on the distance between gminas worked slightly better (Table 4, Figure 7).

Table 4. Spatial econometrics models—a summary (source: own analysis).

Parameter/ v. Name	All the Attributes				Only Geometrical			
	Lag		Error		Lag		Error	
	Neigh.	Distance	Neigh.	Distance	Neigh.	Distance	Neigh.	Distance
R^2	**0.48**	**0.50**	**0.46**	**0.47**	**0.43**	**0.46**	**0.44**	**0.46**
constant	3.6	5.1	1.71	2.04	10.5	11.3	7.7	7.3
G_Built	0.23	0.24	0.23	0.25	0.23	0.25	0.23	0.26
G_Agric					−0.05	−0.04	−0.04	−0.04
D_BTC	0.37	0.37	0.27	0.28	0.40	0.39	0.32	0.33
D_Uni	−0.35	−0.35	−0.25	−0.26	−0.41	−0.41	−0.33	−0.33
income	0.005	0.005	0.004	0.004				
W_matrix	−0.37	−0.50	-	-	−0.37	−0.50	-	-
LAMBDA	-	-	−0.32	−0.41	-	-	−0.40	−0.51
St.error	2.7							2.8
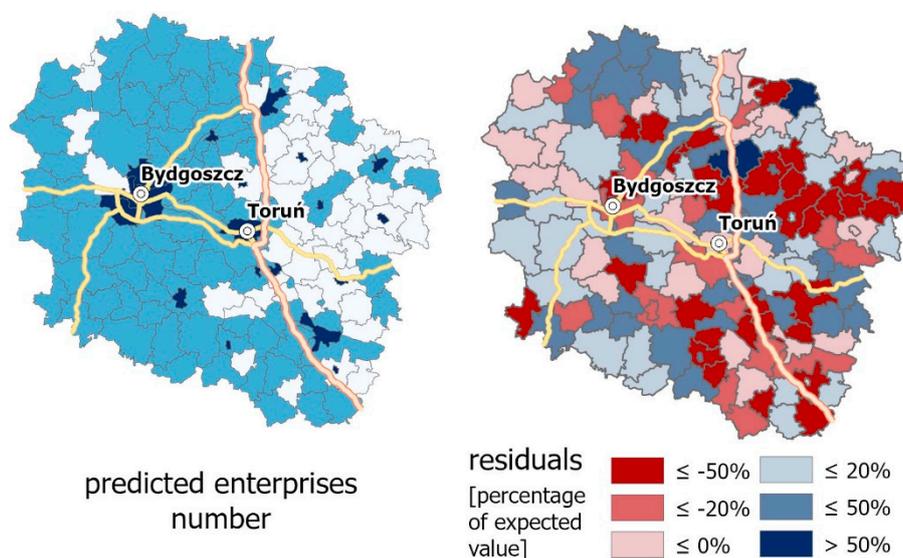
Figure 7. Results of the spatial econometric models and the residuals from regression (lag model with distance matrix) (source: own analysis).

It is worth noting that the use of spatial econometric models made it possible to increase the R2 coefficient for medium-sized enterprises: for complete models—from 0.43 to even 0.50; for models based solely on variables related to the gmina's environment—from 0.39 to 0.46.

What is interesting is the opposite (but comparable, in terms of strength) influence of two factors that would seem very similar: distance from large cities (including Bydgoszcz and Toruń) and distance from universities. It is interesting because the universities are in Bydgoszcz and Toruń. However, using a variable that takes into account the distance to large cities, excluding these two centers (instead of two variables that act oppositely), does not bring the desired effect—the variable D_Cities remains insignificant in all the variants.

3.3. Classification Using Classification and Regression Trees (CART).

The number of enterprises (micro, small, and medium) per 10,000 inhabitants served as response variables in the multivariate regression model. To obtain the comparability of results collected with the CART-type trees and visualization of the multivariate linear regression results, the same range limits were used to divide the data into three classes (Table 5, Figures 3–7). The classes defined in such a way were labeled low, medium, or high.

Table 5. The tripartite division of enterprises number per 10,000 in gminas: classes ranges and cardinality (source: own analysis).

Class	Micro		Small		Medium	
	Number in Gmina	Cardinality	Number in Gmina	Cardinality	Number in Gmina	Cardinality
low	0–125	50	0–25	41	0–5	57
medium	126–175	63	26–40	68	6–10	59
high	176–310	31	41–70	35	11–20	28

When explaining the number of micro enterprises, the analysis of the significance of predictors (Figure 8) indicates that the following are crucial in explaining the dependent variable: the gmina's own revenue from PIT per capita, the distance from large enterprises, the number of inhabitants in a given gmina, the percentage of built-up area, and the density of the road network.
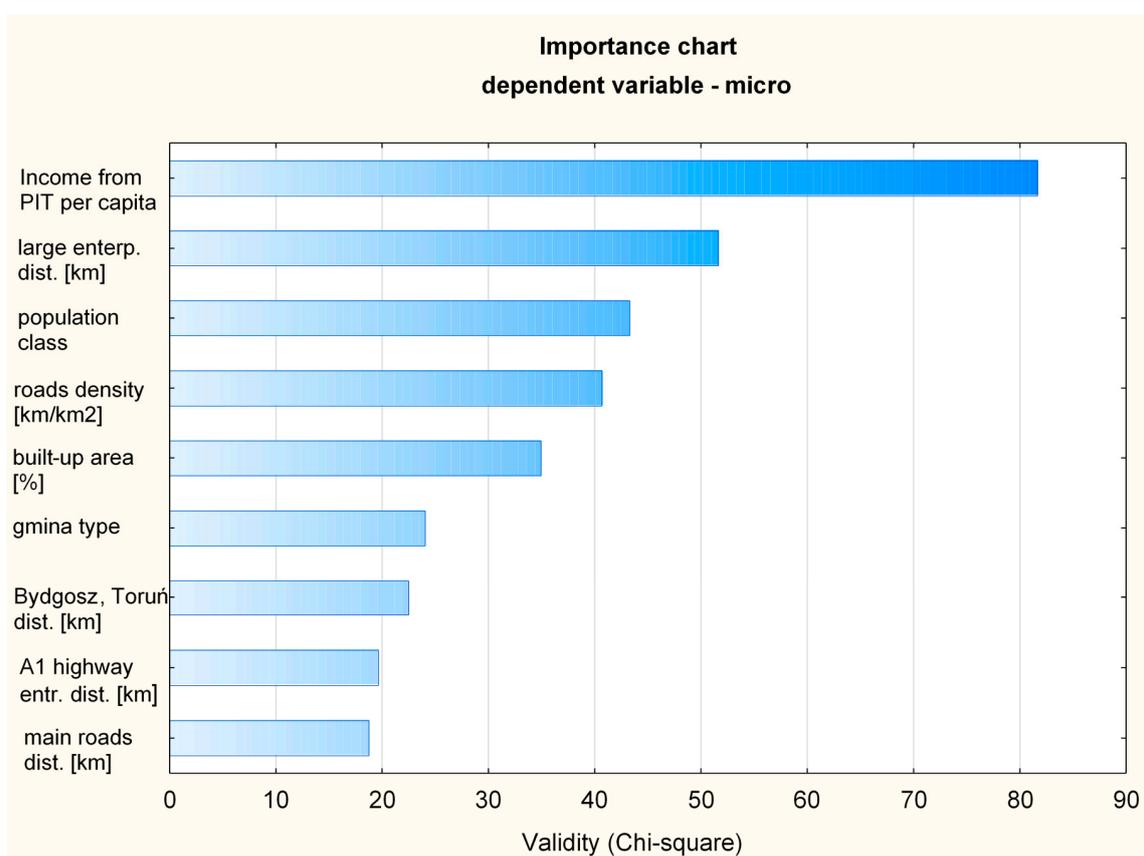
Figure 8. A (chi-square based) significance chart for variables. A dependent variable in number of micro enterprises per 10,000 inhabitants (source: own analysis).

The developed CART model solely uses three of these decision variables, and it is easy to interpret. For instance, the properties of the leaf ID = 6 (Figures 9 and 10) (the preponderance of gminas with a large number of micro-enterprises in relation to the number of inhabitants) means that entrepreneurship (micro) in a given gmina is high if its own revenue from PIT per capita in this gmina is higher than EUR 175.2 (PLN 746.8—we used the medium exchange rate from 2018, which is PLN 1 = EUR 4.2617) and the distance from large enterprises is smaller than 23.6 km. These data confirm the conclusions from the regression analysis; they also indicate the role of large enterprises in entrepreneurship development at the micro-level, which conforms to the conclusions from the literature.
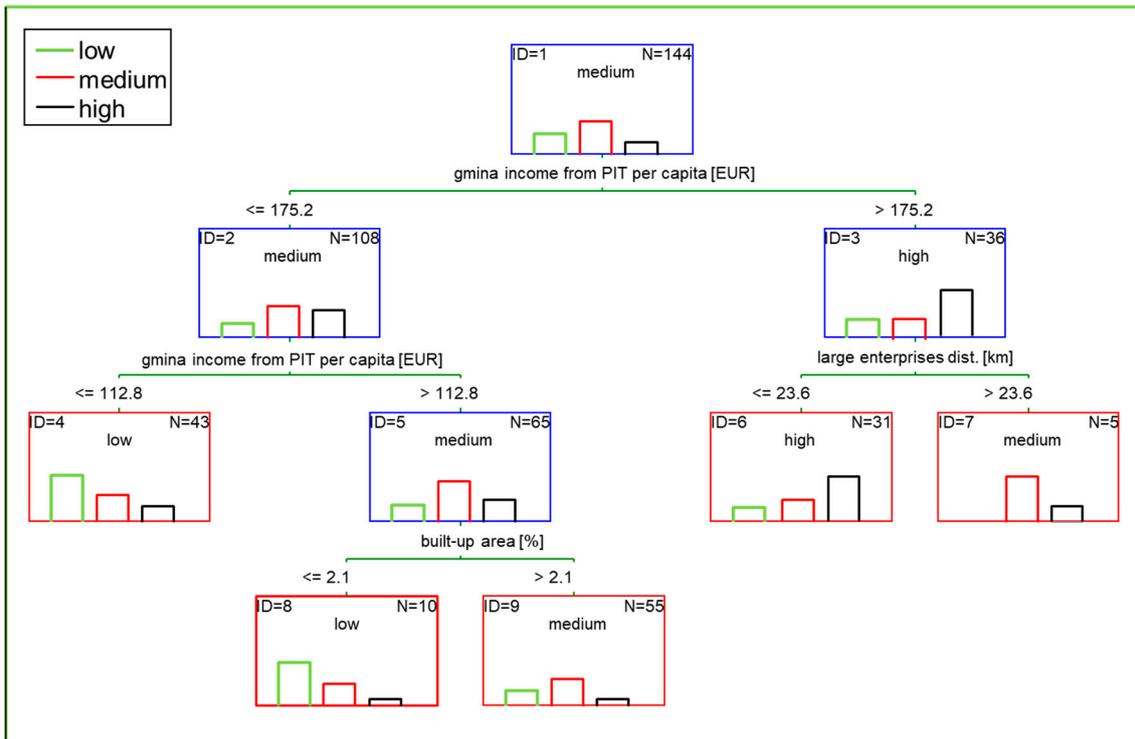
Figure 9. Classification And Regression Tree (CART) for micro enterprises (source: own analysis).
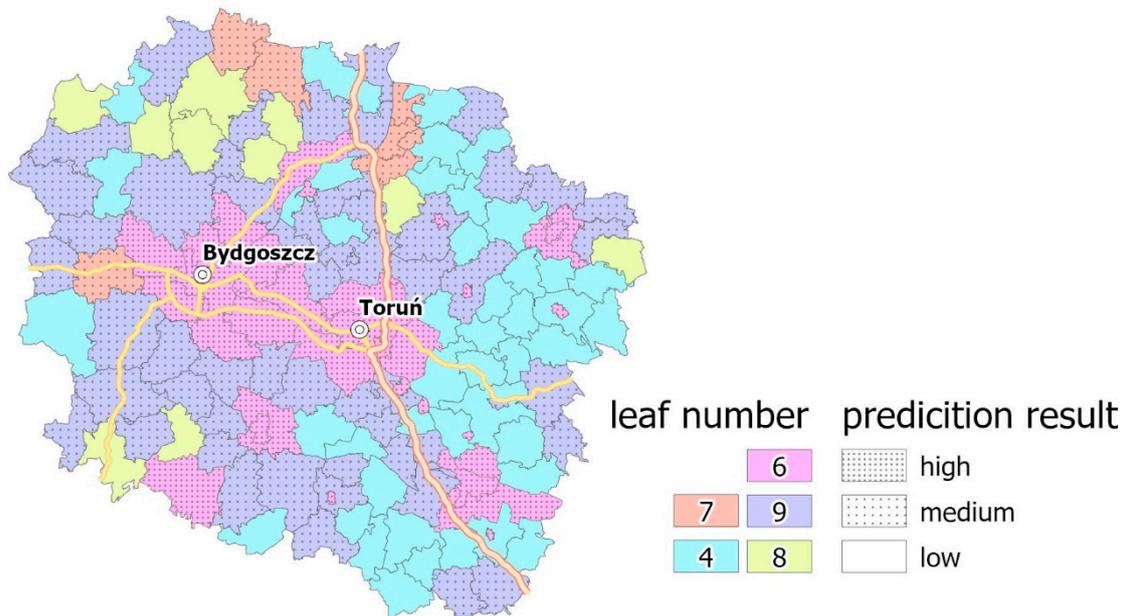

Figure 10. Visualization of CART model prediction for micro enterprises together with the leaf number assigned to gminas, improving the interpretation coming from the model (source: own analysis).

Figure 11 shows the results of the classification carried out by the CART decision tree. Figure 10 indicates also the leaf number in the CART tree used to classify particular gminas. The spatial distribution of the CART results is similar to that in the source expert classification (see Figure 3). However, one should emphasize that the added value of using the decision tree is the assigning of particular gminas to the CART leaf identifier, which enables the explanation of the classification results (Figure 10). The hachure (Figure 11) indicates incorrectly classified gminas; there are 43 such gminas,

representing 29.9%. This error arises from including only three explanatory variables in the decision tree; within individual decision classes, it assumes the values as in Table 6.
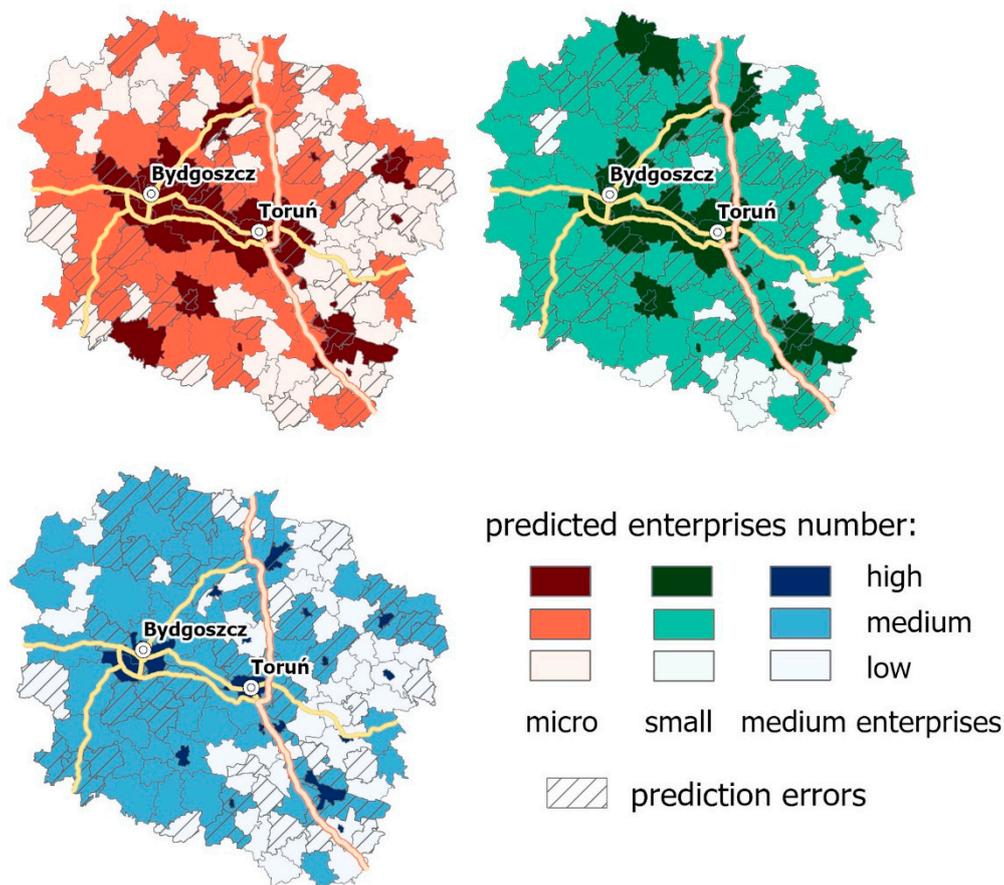
Figure 11. The results of CART classification and the prediction errors (source: own analysis).

Table 6. The incorrectly classified gminas—number and percentage of the class. (Source: own analysis).

Class	Gminas Classified Incorrectly					
	Micro		Small		Medium	
	Number	Percentage	Number	Percentage	Number	Percentage
low	12	24%	20	49%	16	27%
medium	23	36%	11	31%	14	50%
high	8	26%	15	22%	19	33%

A similar model using CART decision trees was developed for small enterprises (Figures 12 and 13). In this model, as in the classification of micro enterprises, the dominant explanatory variable is the gmina's own revenue from PIT per capita. The second independent variable indicates the significance of distances from large companies located in the Kujawsko-Pomorskie region with which MSMEs may cooperate. In this model, the CART tree was created using only two independent variables; thus, the resulting model is even more straightforward and interpretable than the model for the micro-enterprises. (Figure 13)
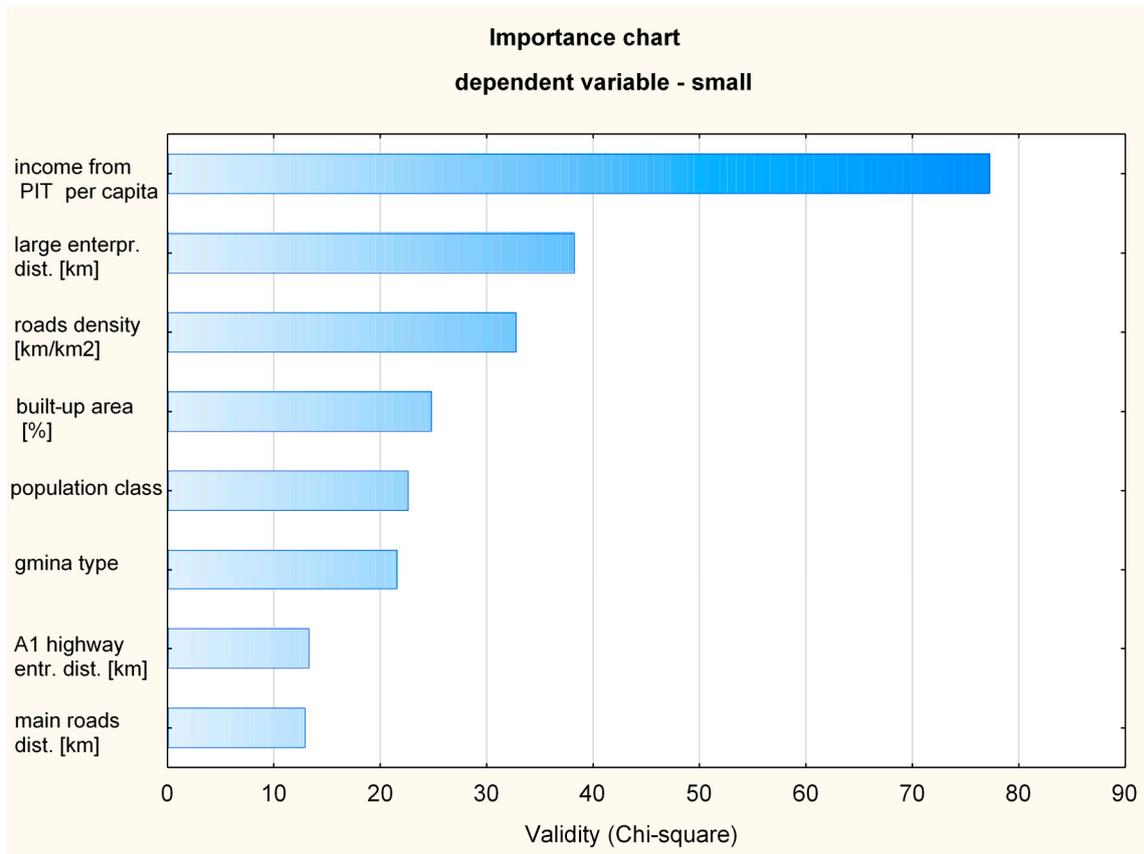
Figure 12. A (chi-square based) significance chart for variables. The dependent variable is the number of small enterprises per 10,000 inhabitants (source: own analysis).
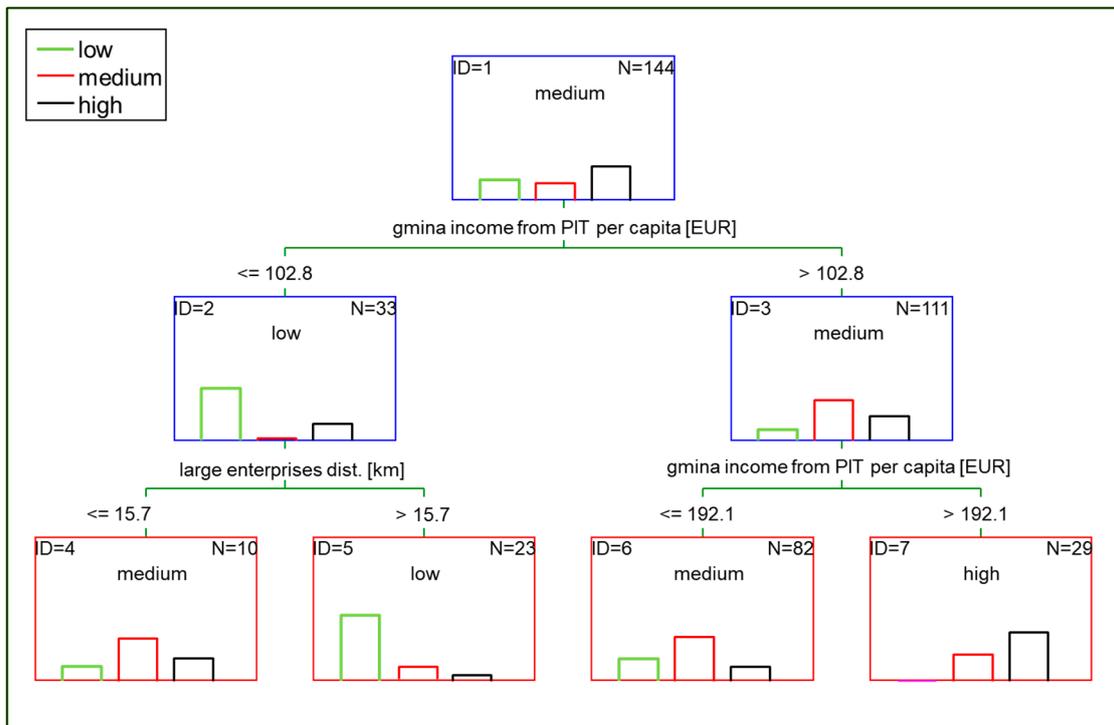

Figure 13. CART tree for small enterprises (source: own analysis).

The obtained results of the classification carried out by the complete CART decision tree indicate that:

- entrepreneurship in a given gmina is high if the revenue from personal income tax per person in this unit is higher than EUR 192.1 (leaf 7);
- entrepreneurship in a given gmina is average if the own revenue from PIT per capita in this unit is lower than EUR 102.8 and the distance from large companies is smaller than 15.7 km (leaf 4), or the revenue is higher than EUR 102.8 (although lower than EUR 182.1) (leaf 6);
- entrepreneurship in a given gmina is low if the own revenue from PIT per capita in this unit is lower than EUR 103.0, and the distance from large companies is greater than 15.7 km (leaf 5).

Figure 11 shows the results of the classification carried out by the CART decision tree for small enterprises. Thanks to the visualization of the results of spatial distribution, it is possible to compare and evaluate the classification results while pointing to the highest number of small enterprises predicted by the model in the central part of the region, between Bydgoszcz and Toruń.

The number of incorrectly classified gminas is 46, representing 31.9% (error distribution among classes—Table 6). The prediction results indicate a significant overestimation of the "medium" value gminas in relation to the source expert classification (92 gminas in relation to 68) and the underflow in relation to "low" values gminas number (23 in relation to 41). It is worth emphasizing that the CART classifier does not generate the so-called "gross" errors. None of the gminas defined as "low" in the expert classification were classified as "high" or vice versa.

Finally, the model classifying gminas in relation to the number of medium-sized enterprises per 10,000 inhabitants was estimated. In this case, apart from the own revenue from PIT per capita, the key to explaining the dependent variable is the road density; built-up density; and the type of gmina—urban, urban-rural, or rural. (Figure 14)
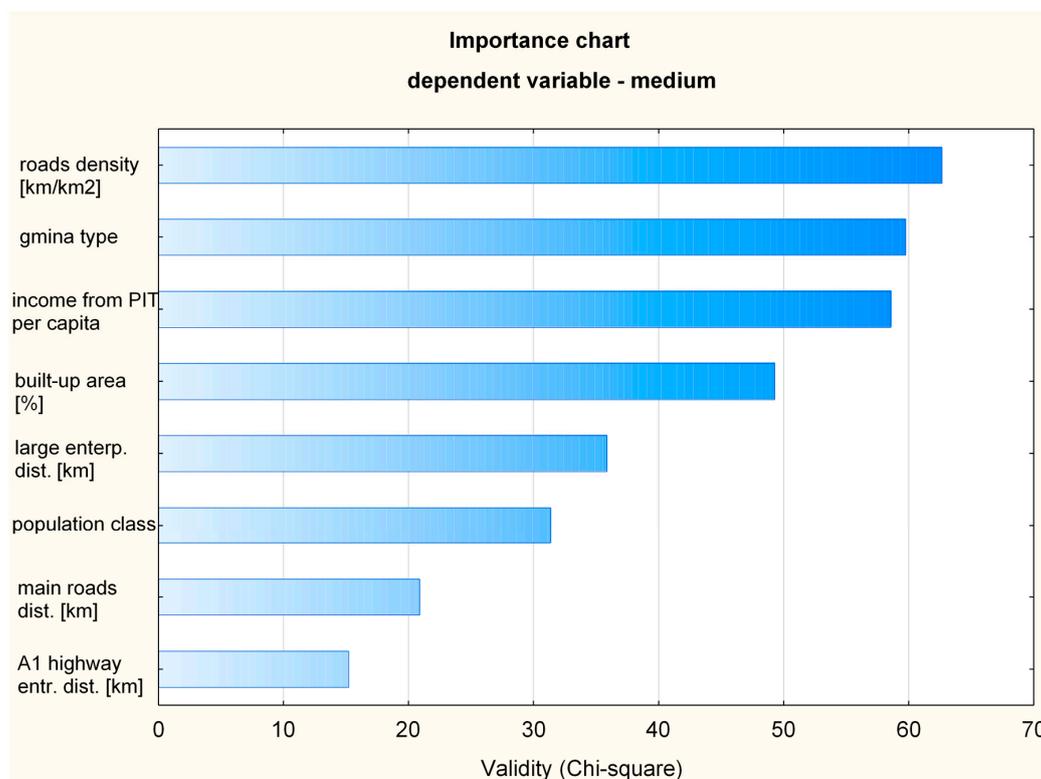

Figure 14. A (chi-square based) significance chart for variables. The dependent variable is the number of medium enterprises per 10,000 inhabitants (source: own analysis).

The model results indicate that a low revenue from the PIT per capita translates into a low number of medium-sized enterprises in the gmina. On the other hand, a high revenue from PIT per capita (over EUR 112.8 per person) and the gmina's high level of urbanization (measured by the percentage of the built-up area above 7.2%) means a higher number of enterprises. (Figure 15)
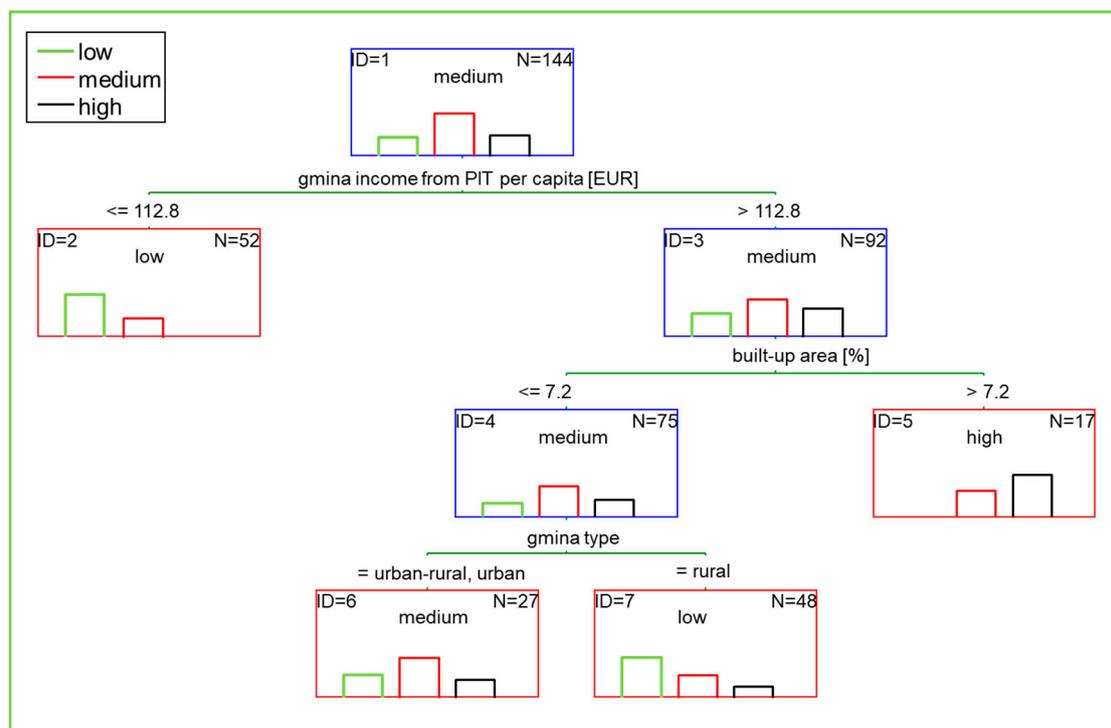

Figure 15. CART tree for medium enterprises (source: own analysis).

The obtained results indicate a significant underestimation by the CART decision tree of the number of gminas with a high level of entrepreneurship development in medium-sized enterprises (17 compared to the 28 obtained in the expert classification) and an excess of gminas classified at the "medium" level (75 compared to 59). One should principally focus on the CART model's underestimation of entrepreneurship development in the central area of the voivodship near Bydgoszcz and the northern region (see Figure 11). The CART model based on a more significant number of explanatory variables—one that takes into account the road network density, distance from large companies in this voivodship, and the number of inhabitants of individual gminas—yields an almost 60% improvement in the quality of classification at the expense of doubling the complexity of the decision model.

4. Discussion

Out of a significant number of generalized regression models, the authors used three models—i.e., a multivariate linear regression model, spatially weighted regression models (spatial econometric models), and CART nonlinear classification trees. The use of diverse regression models, both parametric and non-parametric, would enable obtaining more varied results. The purpose of the conducted research is not to assess the adequacy or effectiveness of different models for generalized regression, but to analyze the development of MSME entrepreneurship in the Kujawsko-Pomorskie region. The obtained results indicate that it is possible to develop a relatively reliable model for estimating the level of entrepreneurship development, explaining even 2/3 of the variability in the phenomenon, using simple models of multivariate linear regression and several explanatory variables (for micro-enterprises). In some models, the spatial distribution of the phenomenon, through the spatial weight matrix, enables

a several-percent increase in the value of the coefficient of determination, R^2. It is possible to obtain similar results using non-parametric models—e.g., CART nonlinear regression models. This article is limited only to the use of such classification models, thus facilitating the creation of conceptually simple models that explain the spatial differentiation of entrepreneurship development in a voivodship. Thanks to CART classification trees, it was not only possible to extract the decision variables critical for understanding, but also to explicitly formulate the decision rules.

The conducted analyses show a significant correlation between the level of entrepreneurship development, measured by the number of enterprises per 10,000 inhabitants, and the gminas' own revenues from the PIT per capita.

Gmina's own revenue from the PIT per capita is a proxy for households' incomes at the local level and reflects the potential consumer demand. This result can indicate that the micro, small, and medium enterprises in the region belong to a relatively low hierarchy cluster focused on producing for local consumption, following the grouping proposed by [53]. They divide SMEs in developing countries into three groups. At the lowest tier are small companies that produce for local consumption. The medium-tier companies are better endowed (in capital and skills) and can generate an investible surplus and produce, either directly or on contract, for the domestic and, often, export markets. The third tier includes technically innovative firms that maintain high quality, capable of entering export markets and aspiring to grow. What also confirms the relatively low tier of development of this sector in the analyzed region is the lack of significance of variables related to the distance to technological parks or higher education institutions. Furthermore, the CART classification indicated that the distance to large companies is significant to the density of the micro and small companies at the local level, while it is not significant to medium enterprises. Thus, this confirms that local factors, such as the consumption needs and cooperation between these companies and large enterprises, determine the activity of micro and small enterprises.

Moreover, the authors observed a substantial significance of variables that essentially formalize the spatial context, such as, e.g., the percentage of built-up areas in gminas, the density of the road network, or distances to some structures. The presence of such variables in both models shows the importance of the spatial context to entrepreneurship development. This is essential because economic analyses usually do not include these variables in such a comprehensive manner. However, this study shows that it is necessary to take into account an area's spatial characteristics when determining the possibilities of entrepreneurship development. Therefore, one should consider the data enrichment of the standard data used in economics with information on the spatial context as an essential step in the construction of econometric models, especially since the models based solely on spatial variables used in this study explain between 30 and 45 percent of the variability in the number of enterprises.

Another way of considering the spatial nature of data is by using spatial econometrics instead of classical regression models. These models prove efficient when dealing with the autocorrelation of the spatial distribution of the residuals from the classical regression models (thus, the residuals are not randomly distributed). In this research, the only enterprises for which this phenomenon occurred were medium-sized enterprises, and spatial econometric models were constructed for them. This may be because the business activities of larger enterprises cover a larger area. Thus, their surroundings in neighboring gminas have a more significant impact on the businesses, as per the classification proposed by [53]; medium-sized enterprises belong to the group with a range beyond the local market. In all of the cases, it was better to use a matrix of weights based on the distances between gminas and not on their neighborhood, which confirms that space and distance are essential, not administrative divisions. This confirms analyses of the labor market areas in Poland that form beyond the administrative borders [40,54]. The use of spatial econometric models in place of classical regression models for medium-sized enterprises increased the R^2 coefficient and improved the model's quality of explaining the variable.

5. Conclusions

The results presented in the paper indicate that spatial development largely determines the local development of micro, small, and medium enterprises. The number of companies at the local level is strongly correlated to the population size. Therefore, the analysis focused on the relative number of companies compared to the population size.

The authors' analysis indicates that the type of entrepreneurship observed in the Kujawsko-Pomorskie region mainly focuses on meeting the demand of local consumers and large companies, as there are more enterprises in those localities where the personal income taxes paid are higher, but also in those with a smaller distance to large enterprises. Furthermore, the results show that the existence of technological parks or special economic zones does not have a significant role in the MSMEs' sector's current structure in the Kujawsko-Pomorskie region. This may indicate that micro, small, and medium companies focus on local consumption. The further development of this sector requires providing the companies with the capital and skills needed to produce for broader markets (both domestic and foreign) and, ultimately, towards a high level of technical innovation with a high potential for growth and international competition. Thus, regional policies should focus on providing access to financial instruments and investing in the current and future workforce's qualifications, including vocational or higher education that recognizes innovative skills.

The authors have also confirmed that the spatial context matters; models that include only geographical variables explain a large share of the variance related to the development of micro, small, and medium enterprises. In the case of micro-enterprises, it correlates with local factors, while in the case of medium-sized enterprises a more comprehensive (geographical) context is essential.

While the presented models do not indicate causality, they can direct the further monitoring of the MSME sector. The results show that there is a group of communities that enjoy a much higher or much lower enterprise development than is explained by the proposed models. These gminas should be further analyzed (using qualitative methods) to identify the factors that stimulate or hinder entrepreneurship development. This may inform the framing of regional policies focused on the development of the MSMEs in the Kujawsko-Pomorskie voivodship.

One should emphasize that conducting comprehensive analyses into the differentiation of entrepreneurship's spatial distribution in particular gminas of the Kujawsko-Pomorskie region required the integration of descriptive data collected by various institutions, as well as the consideration of spatially localized information. The use of the BDOT 10k Topographic Object Database and other spatial data sources (e.g., the location of special economic zones or technology parks) enabled data enrichment. The enrichment of source tabular data with spatial information on, e.g., the distance to the main cities of the region, highway exits, and large enterprises, enabled obtaining additional explanatory variables in generalized regression models. What is more, determining the distance and neighborhood matrices facilitated the creation of spatial econometric models and a more comprehensive explanation of the spatial variability of the phenomenon through a geographically weighted regression model.

The developed generalized regression models, as well as classification trees, have significant advantages over the frequently used non-parametric models. Multivariate linear regression models are straightforward to interpret and apply; they also allow the recipient to determine the impact of specific factors on the model intuitively. The CART classification trees used in this article enable the automatic extraction of decision rules that explain the hierarchical influence of particular predictors on the value of the dependent variable. The obtained results also show the analytical potential of a unified database that depicts entrepreneurship in the Kujawsko-Pomorskie voivodship, as well as the possibility for further data mining with other types of algorithms for generalized regression. This issue will be the subject of further research.

There are, of course, data limitations that need to be considered. As presented in Figure 1, there are outliers in the observed variables that affect the results, which is unavoidable in such analysis. There are also specific factors that affect the MSME development in particular localities (i.e., municipalities at the region border that are affected by centers localized outside the region), that cannot be explained

by any type of regression model. Furthermore, in the analysis, the authors did not take into account additional characteristics of the MSMEs in the region, such as sectoral structure, which requires further investigation.

Finally, in the article the authors show that there is additional utility of the administrative data to investigate local characteristics of the MSME sector. The presented models can be extended further to other Polish regions but also to the wider context of CEE or European countries (provided that similar data exists). The authors show that the differences that are observed between regions in the CEE countries, as discussed in the literature, are also observed within regions, with more developed regional capitals and lagging behind peripheral municipalities. The strengthening of the MSME sector capacity and innovativeness requires designing policies that take into account such differences and designing MSME development strategies to the specificity of the local environment—both socio-economic and spatial.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Karlsson, C.; Dahlberg, R. Entrepreneurship, Firm Growth and Regional Development in the New Economic Geography: Introduction. *Small Bus. Econ.* **2003**, *21*, 73–76. [CrossRef]
2. Allen, T.; Arkolakis, C. Trate and the Topography of the Spatial Economy. *Q. J. Econ.* **2014**, 1085–1139. [CrossRef]
3. Armington, C.; Acs, Z.J. The determinants of regional variation in new firm formation. *Reg. Stud.* **2002**, *36*, 33–45. [CrossRef]
4. Boschma, R.A.; Frenken, K. Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *J. Econ. Geogr.* **2006**, *6*, 273–302. [CrossRef]
5. Boschma, R.A. Competitiveness of regions from an evolutionary perspective. *Reg. Stud.* **2004**, *38*, 1001–1014. [CrossRef]
6. Boschma, R.A.; Wenting, R. The spatial evolution of the British automobile industry: Does location matter? *Ind. Corp. Chang.* **2007**, *16*, 213–238. [CrossRef]
7. Guo, Q.; He, C.; Li, D. Entrepreneurship in China: The role of localisation and urbanisation economies. *Urban Stud.* **2016**, *53*, 2584–2606. [CrossRef]
8. Malmberg, A.; Maskell, P. The elusive concept of localization economies: Towards a knowledge-based theory of spatial clustering. *Environ. Plan.* **2002**, *34*, 429–449. [CrossRef]
9. Boschma, R.A. Proximity and innovation: A critical assessment. *Reg. Stud.* **2005**, *39*, 61–74. [CrossRef]
10. Krugman, P. *Geography and Trade*; MIT Press: Cambridge, MA, USA, 1992.
11. Aghion, P.; Howitt, P. A model of growth through creative destruction. *Econometrica* **1992**, *60*, 323–351. [CrossRef]
12. Elert, N. What determines entry? Evidence from Sweden. *Ann. Reg. Sci.* **2014**, *53*, 55–92. [CrossRef]

13. Duranton, G.; Turner, M.A. The Fundamental Law of Road Congestion: Evidence from US Cities. *Am. Econ. Rev.* **2011**, *101*, 2616–2652. [CrossRef]

14. Wagner, J. Are Young and Small Firms Hothouses for Nascent Entrepreneurs? Evidence from German Micro Data. *Appl. Econ. Q.* **2004**, *50*, 379–391.

15. Dobrev, S.D.; Barnett, W.P. Organizational Roles and Transition to Entrepreneurship. *Acad. Manag. J.* **2005**, *48*, 433–449. [CrossRef]

16. Gompers, P.; Lerner, J.; Scharfstein, D. Entrepreneurial Spawning: Public Corporations and the Genesis of New Ventures, 1986 to 1999. *J. Financ.* **2005**, *60*, 577–614. [CrossRef]

17. Chen, J. The Small Firm Effect and the Quality of Entrepreneurs. *Acad. Manag. Proc.* **2012**, 11180. [CrossRef]

18. Gomes, C.M.; Kruglianskas, I.; Scherer, F.L. Company size effect in innovative performance. *J. Technol. Manag. Innov.* **2009**, *4*, 14–31. [CrossRef]

19. Martínez-Ros, E.; Labeaga, J.M. The relationship between firm size and innovation activity: A double decision approach. *Int. J. Phytoremediat.* **2002**, *21*, 35–50. [CrossRef]

20. Symeonidis, G. *Innovation, Firm Size and Market Structure: Schumpeterian Hypotheses and Some New Themes*; OECD Economics Department Working Papers; OECD: Paris, France, 1996; Volume 161.

21. Schiffer, M.; Weder, B. Firm Size and the Business Environment: Worldwide Survey Results, The World Bank. 2001. Available online: https://EconPapers.repec.org/RePEc:wbk:wbpubs:13988 (accessed on 17 June 2020).

22. Rosenthal, S.S.; Strange, W.C. Geography, industrial organization, and agglomeration. *Rev. Econ. Stat.* **2003**, *85*, 377–393. [CrossRef]

23. Ritsila, J.J. Regional differences in environments for enterprises. *Entrep. Reg. Dev.* **1999**, *11*, 187–202. [CrossRef]

24. Vaessen, P.; Keeble, D. Growth-oriented SMEs in Unfavourable Regional Environments. *Reg. Stud.* **1995**, *29*, 489–505. [CrossRef]

25. Ribeiro-Soriano, D. Small business and entrepreneurship: Their role in economic and social development. *Entrep. Reg. Dev.* **2017**, *29*, 1–3. [CrossRef]

26. Valliere, D.; Peterson, R. Entrepreneurship and economic growth: Evidence from emerging and developed countries. *Entrep. Reg. Dev.* **2009**, *21*, 459–480. [CrossRef]

27. Komornicki, T.; Czapiewski, K.Ł. *Europa XXI—14. Core and peripheral regions in Central and Eastern Europe*; Polish Academy of Sciences: Warsaw, Poland, 2006; Volume 39. [CrossRef]

28. Vida, G.; Dudás, G. Geographical context of the revealed competitiveness of urbanised areas in Hungary excluding the Budapest agglomeration. *Geogr. Pannonica* **2017**, *21*, 179–190. [CrossRef]

29. Lang, T. Socio-economic and Political Responses to Regional Polarisation and Socio-spatial Peripheralisation in Central and Eastern Europe: A Research Agenda. *Hung. Geogr. Bull.* **2015**, *64*, 171–185. [CrossRef]

30. Psycharis, Y.; Kallioras, D.; Pantazis, P. Regional Inequalities in Central and Eastern European Countries: The Role of Capital Regiona and Metropolitan Areas. In *Economic Development and Financial Markets Latest Research and Policy Insights from Central and Southeastern Europe*; Śliwiński, A., Karasavvoglou, A., Polychronidou, P., Eds.; Springer Nature Switzerland AG: Cham, Switzerland, 2020; pp. 3–20. [CrossRef]

31. Smętkowski, M. The role of exogenous and endogenous factors in the growth of regions in Central and Eastern Europe: The metropolitan/non-metropolitan divide in the pre- and post-crisis era. *Eur. Plan. Stud.* **2018**, *26*, 256–278. [CrossRef]

32. Kisiała, W.; Suszyńska, K. Economic growth and disparities: An empirical analysis for the Central and Eastern European countries. *Equilibrium* **2017**, *12*, 613–631. [CrossRef]

33. Szörfi, B. Development and Regional Disparities—Testing the Williamson Curve Hypothesis in the European Union. *ONB Focus* **2007**, *2*, 100–121.

34. Van Stel, A.; Suddle, K. The impact of new firm formation on regional development in the Netherlands. *Small Bus. Econ.* **2008**, *30*, 31–47. [CrossRef]

35. Łuczka, T.; Przepióra, P. Regional Determinants of Efficiency Growth of Small and Medium-Sized Enterprises. Evidence from Poland. *J. Entrep. Manag. Innov.* **2012**, *8*, 138–154. [CrossRef]

36. Wach, K. Regional barriers and stimuli of SMEs development in southern Poland. *MPRA Pap.* **2007**, 31662. [CrossRef]

37. Statistics Poland. *Science and Technology in 2017*; Statistics Poland: Warszawa, Poland; Szczecin, Poland, 2019. [CrossRef]

38. McGrath, H.; O'Toole, T. Critical issues in research design in action research in an SME development context. *Eur. J. Train. Dev.* **2012**, *36*, 508–526. [CrossRef]

39. Atherton, A.; Hannon, P.D. Localised strategies for supporting incubation: Strategies arising from a case of rural enterprise development. *J. Small Bus. Enterp. Dev.* **2006**, *13*, 48–61. [CrossRef]

40. Statistics Poland. *Labour Market Areas as an Alternative Approach*; Statistics Poland: Warszawa, Poland; Bydgoszcz, Poland, 2019.

41. PwC. *Analiza Sytuacji Społecznej Województwa*; Regiogmina Project: Toruń, Poland, 2020.

42. Burchfield, M.; Overman, H.G.; Puga, D.; Turner, M.A. Causes of sprawl: A portrait from space. *Q. J. Econ.* **2006**, *121*, 587–633. [CrossRef]

43. Lee, Y.S. Government guaranteed small business loans and regional growth. *J. Bus. Ventur.* **2018**, *33*, 70–83. [CrossRef]

44. Cieślik, J. Współpraca Małego i Dużego Biznesu. Uwarunkowania i dobre praktyki. In *Forum Współpracy Małego i Dużego Biznesu*; Warsaw, Poland, 2017; Available online: https://forumbiznesu-zpp.pl/wp-content/uploads/2018/01/kodeks.pdf (accessed on 24 May 2020).

45. Huczek, M. Parki naukowo-technologiczne a rozwój małych i średnich przedsiębiorstw. *Zesz. Nauk. Wyższej Szkoły Humanit. Zarządzanie* **2007**, *2*, 47–60.

46. Moszkowicz, K.; Bembenek, B. Potencjał technoparków jako kluczowy czynnik rozwoju klastrów w Polsce. *Pr. Nauk. Uniw. Ekon. Wrocławiu* **2018**, *538*, 244–261. [CrossRef]

47. Huggins, R.; Johnston, A. Knowledge networks in an uncompetitive region: SME innovation and growth. *Growth Chang.* **2009**, *40*, 227–259. [CrossRef]

48. Peeters, D.; Thisse, J.-F.; Thomas, I. Transportation Networks and the Location of Human Activities. *Geogr. Anal.* **2010**, *30*, 355–371. [CrossRef]

49. Grzyb, U.; Trzepacz, P.Z. Investment in Transport Infrastructure as a Crucial Factor of Entrepreneurship Development in the New Eu Member States—The Polish Case. *Eur. Integr. Stud.* **2010**, *6*, 94–100. [CrossRef]

50. Lewandowska-Gwarda, K. Geographically weighted regression in the analysis of unemployment in Poland. *ISPRS Int. J. Geo Inf.* **2018**, *7*, 17. [CrossRef]

51. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Chapman and Hall: London, UK, 1993.

52. Ripley, B. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, UK, 1996; ISBN 0-521-46086-7.

53. Nadvi, K.; Schmitz, H. Industrial clusters in less developed countries: Review of experiences and research agenda. *IDS Discuss. Pap.* **1993**, 339. Available online: https://www.ids.ac.uk/publications/industrial-clusters-in-less-developed-countries-review-of-experiences-and-research-agenda/ (accessed on 24 May 2020).

54. Ryczkowski, M.; Stopiński, P. Labour market areas in Poland. *Przegląd Stat.* **2018**, *65*, 350–368. [CrossRef]