MDPI

Article

Geo-Marketing Segmentation with Deep Learning

Oussama Benbrahim Ansari (ORCID)

Triagon Academy Munich, School of Business and Law, Steinheilstrasse 5, 85737 Ismaning, Germany;
Oussama.BenbrahimAnsari@edu.triagon-academy.com

Abstract: Spatial clustering is a fundamental instrument in modern geo-marketing. The complexity of handling of high-dimensional and geo-referenced data in the context of distribution networks imposes important challenges for marketers to catch the right customer segments with useful pattern similarities. The increasing availability of the geo-referenced data also places more pressure on the existing geo-marketing methods and makes it more difficult to detect hidden or non-linear relationships between the variables. In recent years, artificial neural networks have been established in different disciplines such as engineering, medical diagnosis, or finance, to solve complex problems due to their high performance and accuracy. The purpose of this paper is to perform a market segmentation by using unsupervised deep learning with self-organizing maps in the B2B industrial automation market across the United States. The results of this study demonstrate a high clustering performance (4×4 neurons) as well as a significant dimensionality reduction by using self-organizing maps. The high level of visualization of the maps out of the initially unorganized data set allows a comprehensive interpretation of the different clusters and patterns across space. The centroids of the clusters have been identified as footprints for assigning new marketing channels to ensure a better market coverage.

Keywords: spatial clustering; market segmentation; artificial neural networks; deep learning; self-organizing maps; channel management

---

1. Introduction

Marketing literature considers the segmentation, targeting, and positioning (STP) as key pillars of all marketing strategies [1–3]. The purpose of market segmentation is to identify relatively homogeneous groups of consumers with similar consumption patterns. A market segment has four components: (1) it must be identifiable, (2) it must be economically reachable, (3) it is more homogeneous in its characteristics than the market as whole, and (4) it is large enough to be profitable [4]. Customer demand should be considered as a basis for determining the channel structure. The targeting decision, when applied to channel design, entails a choice of whom not to pursue just as much as what segment to pursue. Targeting a channel segment means choosing to focus on that segment, with the goal of achieving significant sales and profits from selling to it [5]. Segmenting a market is not free. There are costs of performing the research, fielding surveys, and focus groups, designing multiple packages, and designing multiple advertisements and communication messages. Inadequate segmentation and clustering problems could lead to missing the strategic marketing opportunity or not cashing in on the rewards of a tactical campaign [6]. From a practical perspective, a lack of well-defined clusters and clear responsibilities often lead to cannibalizing and overlapping in the sales territories. This can jeopardize the performance of the marketing channels and can often create severe channel conflicts.

The customer data usually contain geographic information which should be considered for clustering the data to create the customer segments and thus to decide where the marketing channels need to be located. Clustering of geo-referenced (or spatial) data has become more popular in geo-marketing, however, the traditional clustering methods reveal

various limitations in view of the increased requirements for accuracy and predictability. The more accurately and homogenous the customer segments are conducted, the more successful differentiated targeting through the appropriate channels will be.

Collectively, geospatial data available from several sources have grown into petabytes and increase by terabytes in size every day [7]. The increase in the sources of data and their acquisition have been exponential as compared to the development of processing systems which can process the data in real time [8]. Earlier, storage of data was costly, and there was an absence of technology which could efficiently process the data. Now, the storage costs have become cheaper, and the availability of technology to transform Big Data has become real [9]. The volume of geospatial data at the global scale (e.g., at the petabyte-scale) exceeds the capacity of traditional computing technologies and analytical tools designed for the desktop era. The velocity of data acquisition (e.g., terabytes of satellite images a day and tens of thousands of geotagged tweets a minute) pushes the limits of traditional data storage and computing techniques [10].

Most commercially available GIS provides extended functionality to store, manipulate, and visualize geo-referenced data, but rely on the user's ability for exploratory data analysis. This approach is not feasible with regards to the large amount and high dimensionality of geographic data and demands for integrated data mining technology [11]. The high dimensionality of a dataset can cause serious problems for most analysis methods. One typical problem to address is that it is unlikely for all variables to interrelate meaningfully. Most analysis methods limit or compress the potential hypothesis space by assuming a simple form of pattern, which can be configured with several parameters. For example, a regression analysis assumes a form of pattern (normally a linear form) and uses data to configure its parameters (e.g., coefficients) in relation to this form. However, the number of possible patterns, which can be of various forms, is practically infinite in a multivariate spatial dataset. Patterns can be linear or non-linear, spatial, or non-spatial, with different configurations [12]. In the real world, we encounter problems that exhibit non-linear correlations among the actions. They require complex and nonlinear models. To overcome this problem, a new technique called deep learning was proposed. This technique introduces non-linearity into the network [13]. The applicability of deep learning in business is a research field that intends to research different models that learn patterns from data in a supervised or unsupervised manner. The field presents high interest both for practitioners, as for researchers. Deep learning relates to artificial neural networks with multiple hidden layers, convolutional neural networks, recurrent neural networks, self-organizing maps, Boltzmann machine, and auto encoders [14].

This paper investigates deep learning approaches to spatial clustering and identifies the unsupervised self-organizing maps (SOM) to perform an optimized customer segmentation which should provide a comprehensive visualization of the different customer groups with similar patterns by considering their geo-location. Although some previous research papers applied the SOM algorithm to solve complex clustering tasks, its big advantage in a comprehensive visualization and knowledge acquisition in the geo-marketing has not been fully exploited. A key contribution of this research is the integration of a deep learning unsupervised approach into the channel management environment to enhance the visualization of the homogenous clusters and enable a deep learning approach for the customer segmentation problem. The empirical results of this research demonstrate the importance of using geo-marketing intelligence and visualization in the strategic decision making of manufacturers within the B2B industrial automation market in the US.

2. Theoretical Framework

2.1. Industrial Market Segmentation

B2B markets are considerably more challenging than consumers markets and demand specific skills from marketers. Buyers, with a responsibility to their company and specialist product knowledge, are more demanding than the average consumer [15]. Buyers of B2B products and services often need to deliver a return on investment for their purchase,

highlighting the more complex nature of B2B purchases [16]. In the industrial markets, often the same industrial products have multiple applications; likewise, several different products can be used in the same application. Customers differ greatly, and it is hard to discern which differences are important and which are trivial for developing a marketing strategy [17]. Segmenting industrial markets is different and more challenging because of greater complexity in buying processes, buying criteria, and the complexity of industrial products and services themselves [18].

Industrial market segmentation is a decision process that enables a firm to effectively allocate marketing resources to achieve business objectives. The decision process seeks to implement the major tenets of the marketing concept—to define an offering (products and services) that meets the needs of target buyers, while recognizing the behaviors of competitors and other stakeholders that define the market. While there are several decisions to be made in the process of segmentation, they revolve around the identification of groups of potential organizational buying centers that within each group are similar in response to a marketing program, and between-groups are different in their response [19]. The decision to use B2B market segmentation has three critical components: (1) the market uncertainties faced from the outcome of a situation analysis; (2) the importance of the marketing decisions contemplated; and (3) the organization's readiness to embrace segmentation [20].

The goal for every industrial market segmentation scheme is to identify the most significant differences among current and potential customers that will influence their purchase decisions or buying behavior, while keeping the scheme as simple as possible. This will allow the industrial marketers to differentiate their prices, programs, or solutions for maximum competitive advantage [21]. Marketers should evaluate a myriad of descriptive characteristics when identifying and selecting business market segments. The critical issues can be understood by analyzing geodemographic attributes or firmographics. These segmentation bases provide important decision-oriented insights about high-tech, industrial, and service markets [22]. Industrial market segmentation is currently primarily based on geographics and demographics [23]. However, this leaves industrial suppliers unsatisfied, for segmentation of the market into homogeneous groups regarding buying behavior has proved to be very difficult based on these criteria [21]. Industrial marketers can hardly be blamed for feeling that segmentation is very difficult for them. Not only has little been written on the subject as it affects industrial markets, but such analysis is also more complex than for consumer markets. The problem is to identify the best variables for segmenting industrial markets [17]. According to Webster, segmentation variables are customer characteristics that relate to some important difference in customer response to marketing effort [24]. The selection of segmentation variables should be based on such conditions as measurability, substantiality, accessibility, and actionability [25].

Johnson and Flodhammer [26] investigated how market segmentation in industrial Swedish firms is applied. They proposed a model for identifying variables to measure and group industrial customers in meaningful segments. In their study, they addressed two major problems that should be considered in industrial market strategy: (1) Is there a need for segmentation? If so, what conditions should be met? (2) How to identify segmentation variables—useful and relevant to evaluating industrial markets?

The main conditions are summarized as follows: Market segmentation is appropriate for industrial firms when one of the following conditions is available; the more heterogenous the product assortment, the greater the need for market segmentation; the more the customers differ on buying strategy, the greater the need for segmentation; the more heterogenous the market, the more the need for segmentation; the more the environment changes, the greater the search for expansion into market opportunities, the greater the reason for segmentation concepts. With regards to problem 2, five major criteria are proposed for identifying market segmentation variables—technological, economic, market, competition, and organizational characteristics. Finally, in their study, they suggest a model with segmentation variables suited to industrial markets. These variables include product/process, application (field use), branch (SIC), market size, customer location, buying

process, buying process, buying center, previous relations of seller to buyer, and end-user (environment).

Plank demonstrated in his review of the industrial segmentation literature, that there are three approaches for selecting segmentation bases: (1) unordered segmentation notions (a single segmentation dimension is chosen with no specific rationale for how it was selected), (2) two-step notions (such as the macro-micro-segmentation), or (3) a multistep approach (such as nested approach) [27].

Wind and Cardoza indicated that, while the concept of marketing strategy differentiation is widely accepted among industrial firms, there is little evidence to suggest that firms do follow a conscious segmentation strategy to plan or control their marketing activities. They proposed a two-stage approach to industrial segmentation that consists of macrosegments and microsegments. They explained that macrosegments consist of key organizational characteristics such as size of the buying firm, SIC (Standard Industrial Classification) category, geographic location, and usage factors. The second stage involves dividing those macrosegments into microsegments, based on characteristics of decision-making units [28].

Bonoma and Shapiro [29] proposed general guidelines for segmenting industrial markets following a nested approach. Specifically, they distinguished five general categories of segmentation variables:

(1) Demographics—industry, company size, and customer location.
(2) Operating variables—technology, user status, and customer capabilities.
(3) Purchasing approaches—purchasing function organization, power structures, buyer-seller relationships, and purchase policies/criteria.
(4) Situational factors—urgency of order fulfillment, product application, and size of order.
(5) Buyers' personal characteristics—buyer–seller similarity, attitudes toward risk, and buyer motivation/perceptions.

The "Nested Approach" is still applicable to industrial markets today in the twenty-first century. It is still relevant, but changes are taking place [30].

Dibb and Simkin [31] argued that one of the major problems associated with segmentation in B2B markets is a failure of businesses to implement plans. They mentioned that there are three main reasons for this failure:

(1) Infrastructure barriers: (This concern the culture, structure, and resources which can prevent the segmentation process from starting or being completed successfully).
(2) Process barriers: These barriers reflect a lack of experience, guidance, and expertise concerning the way in which segmentation is undertaken and managed.
(3) Implementation barriers: These are practical barriers concerning a move to a new segmentation model.

However, implementation is not the final step, because the company still needs to react to dynamic internal and external changes. Little attention has been given to the strategy and implementation phases. There have been many contributions in the areas of dynamics, even though it is recognized as an important area [32]. According to blocker and Flint, markets and market segments can be unstable over time, and there must be at least a conceptual understanding of this concern, if not the methodological rigor in tracking the existence of a segment structure of the market [33]. Wind and Thomas suggested using both an "interactive research approach" to measure changing responses to marketing stimuli and a "panel survey" to assess the changing segment structures regarding products [19].

When the segmentation is finished and the strategies are created, it is important to realize what segment factors are critical for the company's success, and they should be continually monitored. Some of the critical success factors are likely to be the criteria that were critical in the identification of the segments. Others could be critical assumptions made while developing the strategy, for example, the technology. Other critical success factors could include the customer needs, relationships, technological development, and competitor offerings and moves [32].

Business segmentation gained credibility and acceptance led by research by Bononma and Shapiro and Plank. Despite this progress—more than 30 years—B2B segmentation is often misunderstood or poorly utilized by marketers [34]. With regards to the methods used for the industrial market segmentation, the literature considers a wide variety of techniques. Statistical models have been widely used for customer clustering [35]. Descriptive methods such as logistic regression or cluster-based techniques face serious limitations. Kotras demonstrated in his study how predictive algorithms can increase the performance of customer segmentation [36].

Machine learning and artificial intelligence have clear advantages over traditional statistical methods when: (a) there are a multitude of variables available for analysis, (b) the associations between the variables are uncertain (and likely to be highly complex), (c) the values of each variable are evolving constantly (such as in the case of a GPS), and (d) when understanding correlations between variables are more important that causation. The great strength of machine learning models is in making predictions, especially where an atheoretical prediction will work well. This is the reason that machine learning models are evaluated on criteria such as scalability, real-time implementation, and cross-validated predictive accuracy rather than on internal and external validity and theoretical foundations which are more suited to the traditional models [37]. Artificial intelligence and machine learning in marketing science are currently gaining more importance to leverage predictive segmentation [38,39].

2.2. Artificial Neural Networks

Neural networks learn to do tasks with progressive improvement in performance, by considering examples, generally without task-specific programming. They have found use in applications difficult to express in a traditional computer algorithm using rule-based programming. The original goal of the neural network approach was to solve problems in the same way that a human brain would [40]. An artificial neuron is a basic building block of every artificial neural network. Its design and functionalities are derived from observation of a biological neuron that is the basic building block of biological neural networks. An artificial neuron can be represented (Figure 1) with its inputs, weights, transfer function, bias, and outputs [41]. An artificial processing neuron receives inputs as stimuli from the environment, combines them in a special way to perform a 'net' input, passes that over through a linear threshold gate, and transmits the (output) signal to another neuron or the environment [42].
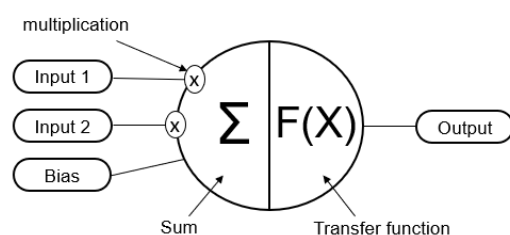

Figure 1. Artificial neuron design.

The artificial neuron model can be seen in its mathematical description:

$$y(k) = F\left(\sum_{i=0}^{m} w_i(k).x_i(k) + b\right) \quad (1)$$

where:

- $x_i(k)$ is input value in discrete time k where i goes from 0 to m,
- $w_i(k)$ is weight value in discrete time k where i goes from 0 to m,
- b is bias,
- F is a transfer function,

- $y_i(k)$ is output value in discrete time k.

The ANNs are also called feedforward neural networks, and recently, deep networks or learning [43]. People began referring to it as "deep" when it started utilizing 3–5 layers a few years ago. Now, networks with more than 200 layers are commonplace [44]. The notion of deep learning refers to an artificial neural network mode that has multiple layers. Studies have shown that a multi-layer artificial neural network is capable of deep learning, namely, is able to model any non-linear relationship in a system [45]. Deep learning is a subset of a more general field of artificial intelligence called machine learning, which is predicated on this idea of learning from example [46] where learning happens through multiple learned layers of neurons [43]. The architecture or topology of the ANN describes the way the artificial neurons are organized in the group and how information flows within the network [47]. According to Haykin [48], there are three different classes of network architectures: single-layer feedforward networks, multilayer feedforward networks, and recurrent networks:

The single-layer feedforward network is the simplest type of neural network. It consists of one output unit and two input units with no hidden layers; thus, it is also known as single-layer perceptron [9]. In a layered neural network, the neurons are organized in layers. One input layer of source nodes projects directly onto an output layer of neurons, but not vice versa. This network is strictly of a feedforward type [49]. The second class of feedforward neural network distinguishes itself by the presence of one or more hidden layers, whose computation nodes are called hidden neurons. The term "hidden" refers to the fact that this part of the neural network is not seen directly from either the input or output of the network output in some useful manner [48]. Among the main networks using multilayer feedforward architectures are the multilayer perceptron (MLP) and the radial basis function (RBF), whose learning algorithms used in their training processes are respectively based on the generalized delta rule and the competitive/delta rule [50]. In the recurrent network, at least one neuron connects with another neuron of the preceding layer and creates a feedback loop. This type of neural network consists of a self-connection between the neurons of the hidden layer. This functionality provides them with a temporary memory. As a result, activation of the following value is forwarded from the lower layer as well as its previous activation value to the hidden layer neuron [50]. Recurrent networks are used to process time varying data, predict future values, classify time series, predict system behavior, and so on [51].

2.3. Learning Algorithms

The performance of a neural network depends to a significant extent on how well it has been trained, and not the adequacy of assumptions concerning the statistical distribution of the data, as is the case with the maximum likelihood classifier [52]. Only after training does the network become operational, i.e., capable of performing the task it was designed and trained to do [53]. Learning or training is one of the most important characteristics of artificial neural networks. The literature on deep learning recognizes two main types of learning algorithms, either supervised or unsupervised.

In supervised learning, the neural network is trained on a training set consisting of vector pairs. One of these vectors is used as input to the network; the other is used as the desired or target input. During the training, the weights of the NN are adjusted in such a way as to minimize the error between the target and the computed output of the network [53]. Though it is biologically implausible, backpropagation learning is the most popular learning rule for performing supervised learning tasks. It is not only used to train feedforward networks such as MLP but also adapted to RNNs [54]. Backpropagation, or back-prop as it is often called, is a more complex way of learning. It reduces the error between the actual output of the network and the desired output by changing the connection weights and biases in such a way that they move slowly toward the correct values [55].

In unsupervised or self-organized learning, there is no external teacher to supervise the learning process. In this type of learning, no specific examples are provided to the network. The desired response is not known, so explicit error information cannot be used to improve network behavior. Since no information is available concerning correctness or incorrectness of responses, the learning process must somehow be accomplished based on observations of responses to inputs that have very little or no knowledge about [56]. The neurons compete to match the input as closely as possible, usually based on Euclidean distance. The neuron closest to the considered input exemplar in the winner taking it all, i.e., adjusting its weight to improve its position and thus move closer to the input [57]. Unsupervised data learning involves pattern recognition without the involvement of a target attribute. That is, all the variables used in the analysis are used as inputs and because of the approach, the techniques are suitable for clustering and association with mining techniques [58].

2.4. Literature Review on Spatial Clustering

The existing literature considers the clustering and segmentation as key activities in geo-marketing. Clustering is the technical process for unsupervised grouping, while segmentation is the application of creating segments of customers or markets. Thus, clustering can be used to segment consumer groups. Clustering helps firms to identify meaningful customer segments, allowing them to target defined groups rather than having to customize for each individual customer [59]. Clustering aims at grouping consumers in a way that consumers in the same segment (called a cluster) are more similar to each other than those in other segments (clusters) [60]. Clustering is a descriptive task that seeks to identify homogeneous groups of objects based on the values of their attributes [61]. A cluster is a collection of data objects with higher similarity within cluster and lower similarity between clusters. The degree of similarity is usually described by the distance between objects. The greater the distance, the smaller the similarity is, vice versa [62]. Due to the complexity and size of the spatial databases, clustering methods should be efficient in high dimensional space, explicit in the consideration of scale, insensitive to a large amount of noise, capable of identifying useful outliers, insensitive to initialization, effective in handling multiple data types, independent to a priori or domain specific knowledge, and able to detect structures of irregular shapes. Conventional clustering algorithms often fail to fulfill these requirements [63].

As shown in Table 1, clustering algorithms can be classified into the following categories: partitional clustering algorithms, hierarchical clustering methods, density-based clustering algorithms, grid-based clustering algorithms. Many of these can be adapted to or are specially tailored for spatial data [64]. General purpose high-dimensional clustering methods mainly deal with non-spatial feature spaces and have very limited power in recognizing spatial patterns that involve neighbors [65].

Table 1. Summary of clustering classifications and the common algorithms used to achieve partitioning, hierarchical, density-based, or grid-based approaches [66].

Clustering Classification	Description	Common Algorithms
Partitioning	Partitions the data into a user specified number of groups. Each point belongs to one group. Does not work well for irregularly shaped clusters.	k-means, k-medoids, clustering large applications-based randomized search (CLARANS), and expectation-maximization (EM) algorithm.
Hierarchical	Decomposes data into a hierarchy of groups, each larger group contains a set of subgroups. Two methods: agglomerative (builds groups from the observation up), or divisive (start with a large group and separate).	Balanced iterative reducing and clustering using hierarchies (BIRCH), Chameleon, Ward's method, nearest neighbor, (dendrograms are used to visualize the hierarchy).
Density-based	Useful for irregularly shaped clusters. Clusters grow based on a threshold for the number of objects in a neighborhood.	DBSCAN, ordering points to identify cluster structure (OPTICS) and density based clustering (DENCLUE), SNN.
Grid-based	Region is divided into a grid of cells, and clustering is performed on the grid structure.	statistical information grid (STING), Wave Cluster, and clustering in quest (CLIQUE).

In geographic segmentation, data are clustered or categorized according to geographic criteria such as nations, states, regions, countries, cities, neighborhoods, or postal codes. However, during the process of segmentation, a serious overlapping issue may occur and leads to an inefficient geospatial analysis. Moreover, geo-marketing is usually active in urban areas and requires clusters to be organized in a three-dimensional (3D) way [67]. For spatial clustering, it is important to be able to identify high-dimensional spatial clusters, which involves both the spatial dimensions and several non-spatial dimensions [12]. Table 2 presents a synthesis of recent studies in the field of spatial clustering.

Table 2. Principle studies on spatial clustering.

Research	Objective	Methods	Reference
Visualizing self-organizing maps with GIS	Demonstrate how SOM can be imported into GIS	Self-organizing maps (SOM), GIS	[68]
Marketing Analytics: Methods, Practice, Implementation, and Links to Other Fields	To provide a practical, implementation-based overview of marketing analytics methodology	Theoretical review	[69]
Local News Online: Aggregators, Geo-Targeting and the Market for Local News	Examines how placement of geo-targeted local news links on Google News affected local news consumption	Empirical analysis is to identify the effect of adding geo-targeted local news links to Google	[70]
Optimal Segmentation in Platform Strategy: Evidence from Geotargeted Advertising	Provides a model to improve ad targeting	Regression discontinuity	[71]
Customer Recognition and Mobile Geo-Targeting	Investigates how combining behavior-based marketing with mobile geo-targeting influences profits and welfare in a competitive environment	Geo targeting	[72]
Competitive Mobile Geo Targeting	Investigates in a competitive setting the consequences of mobile geo targeting, the practice of firms targeting consumers based on their real-time locations.	Mobile targeting, Geo-Targeting analytical models	[73]
3D Geomarketing segmentation: A higher spatial dimension planning perspective	Reduce the overlapping issue during the process of cluster segmentation	Case studies, k-means algorithm	[67]
Application of k-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services	Investigate the utility of k-means algorithm in customer segmentation	MATLAB, k-means algorithm	[74]
Customer Portfolio Analysis Using the Self-Organizing Map	Investigate how the self-organizing map (SOM) can be used for one category of CRM, customer portfolio analysis (CPA)	Customer portfolio analysis (CPA), data-driven market segmentation, self-organizing map (SOM)	[75]
Clustering With GIS: An Attempt to Classify Turkish District Data	Compare different clustering techniques for spatial classification and perform a classification for Turkey's districts	GIS, spatial clustering techniques, SOM algorithm	[65]

Spatial clustering has also long been used as an important process in geographic analysis [65]. Although clustering is an unsupervised learning technique, firms can utilize segments in predictive modeling by creating separate predictive models for each segment [59]. k-means is one of the simplest unsupervised learning algorithms used for

clustering. k-means partitions n observations into k clusters in which each observation belongs to the cluster with the nearest mean. This algorithm aims at minimizing an objective function, in this case, a squared error function [76]. In practice, businesses usually strive to reduce the overlaps and gaps in their market coverage. The k-means algorithm has been widely used in market segmentation. The review of different geo-marketing research revealed that k-means has been popular, especially in clustering applications in combination with GIS [67,74]. Azri argued that segmenting the marketing data could increase time efficiency and sales volume. However, overlapping area among segmented clusters may lead to inefficient data management. In their research, they proposed clustering algorithm, k-means++ to reduce the overlap area. This approach was able to minimize the overlap region during market segmentation [67]. Ezenkwu demonstrated in his paper an application of the k-means algorithm to conduct an efficient customer segmentation. He used a MATLAB implementation of the k-means clustering based on data collected from a mega business retail outfit that has many branches. The result of this study shows that the algorithm has a purity measure of 0.95, indicating 95% accurate segmentation of the customers [74]. Even though the k-means algorithm is widely used in geo-clustering tasks, however, this clustering method has significant disadvantages versus deep-learning algorithms. The main weaknesses of the k-means can be summarized as follows: the number of cluster centers needs to be pre-determined, the algorithm fails for non-linear data sets; applicable only when for numerical data sets and fails for categorical data; unable to handle noisy data and outliers [45,77]. Another clustering algorithm which has been identified in spatial applications is an unsupervised learning approach known as the self-organizing map also called the Kohonen algorithm.

3. Data and Methodology

Self-organizing maps are feedforward, unsupervised neural networks and were developed in the 1980s by Kohonen [78]. The self-organizing maps differ from other neural networks as they apply competitive learning as opposed to error-correction learning (such as backpropagation with gradient descent), and in the sense that they use a neighborhood function to preserve the topological properties of the input space [79]. SOM converts complex, nonlinear statistical relationships between high-dimensional data into simple geometric relationships on a low-dimensional display [80]. The SOM provides a topology-preserving mapping from a high-dimensional input space to a lower-dimensional output space. It is often applied for visualization of high-dimensional data [81]. SOM, also known as the Kohonen network, consists of an input layer, which distributes the inputs to each node in a second layer, the so-called competitive layer. Each of the nodes on this layer acts as an output node. Each neuron in the competitive layer is connected to other neurons in its neighborhood and feedback is restricted to neighbors through these lateral connections. Neurons in the competitive layer have excitatory connections to immediate neighbors and inhibitory connections to more distant neurons. All neurons in the competitive layer receive a mixture of excitatory and inhibitory signals from the input layer neurons and from other competitive layer neurons [82]. The Kohonen layer is a "Winner-takes-all" layer. Thus, for a given input vector, only the Kohonen layer output is 1 whereas all others are 0. No training vector is required to achieve this performance. Hence, the name: self-organizing map layer (SOM-layer) [83].

The self-organizing map differs considerably from the feedforward backpropagation neural network in both how it is trained and how it recalls a pattern. The self-organizing map does not use an activation function or a threshold value. In addition, output from the self-organizing map is not composed of output from several neurons; rather, when a pattern is presented to a self-organizing map, one of the output neurons is selected as the "winner". This "winning" neuron provides the output from the self-organizing map. Often, a "winning" neuron represents a group in the data that is represented to the self-organizing map [84].

One further advantage of the SOM is the 2D output grid that can be used for visualization of the results of the SOM. This visualization can enhance the understanding for the underlying dataset [85]. Unlike other neural networks, there is no hidden layer or hidden processing units. As shown in Figure 2, in the SOM, the neurons of the input layer relate to all the neurons of the output layer through synaptic weights. Consequently, the information provided by each neuron of the input layer is transmitted to all the neurons of the output layer. All the neurons of the output layer receive the same set of inputs from the input layer.
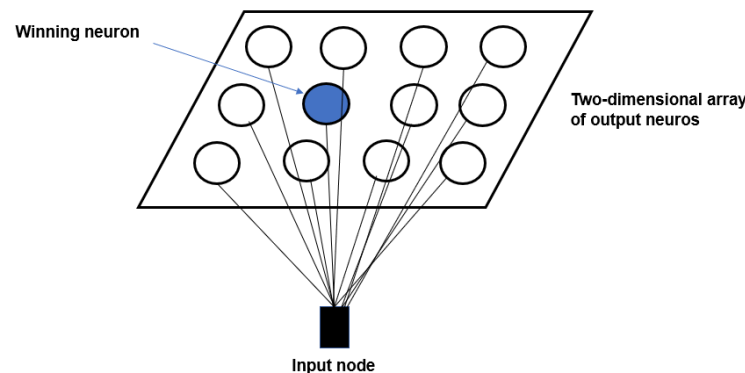

Figure 2. Kohonen model [42].

The SOM is trained by using a combination of neighborhood size, neighborhood update parameters, and a weight-change parameter. The SOM is formed initially with random weights between the input layer neurons and each of the neurons in the SOM. Each neuron in the input layer is connected to every neuron in the SOM. A neighborhood is the region around a given neuron that will be eligible to have the weights adapted. Neurons outside the region defined by the neighborhood do not undergo any weight adjustment. As the training is performed, the neighborhood size is adjusted downward until it surrounds a single neuron. The use of a neighborhood that reduces in size over time allows the SOM to group similar items together [86]. The training utilizes competitive learning. When a training example is fed to the network, its Euclidean distance to all weight vectors is computed. The neuron whose weight vector is most similar to the input is called the best matching unit (BMU). The weights of the BMU and neurons close to it in the SOM lattice are adjusted towards the input vector. The magnitude of the change decreases with time and with distance (within the lattice) from the BMU [79].

The training process of the Kohonen network algorithm consists of the following steps [82]:

1. Initialize network:

   Define $w_{ij}(t)$ $(0 \leq i \leq n-1)$ to be the weight from input i to node j at time t. Initialize weights from n inputs to the nodes to small random values. Set the initial radius of the neighborhood around node j, $N_j(0)$, to be large.

2. Present new input:

   Present input $x_0(t)$, $x_1(t), x_2(t), \ldots, x_{n-1}(t)$, where $x_i(t)$ is the input to node to node i at time t.

3. Calculate distances:

   Compute distances d_j between the input and each output node j using:

   $$d_j = \sum_{i=1}^{n-1} \left(x_i(t) - w_{ij}(t) \right)^2$$

   Select minimum distance:

Designate the output node with minimum d_j to be $j*$.

4. Update weights:

Update weights for node $j*$ and its neighbors, defined by the neighborhood size $N_{j*}(t)$. New weights are:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)\big(x_i(t) - w_{ij}(t)\big)$$

for j in $N_{j*}(t)$, $(0 \leq i \leq n-1)$

The term $\eta(t)$ also called learning rate is a gain term $(0 < \eta(t) < 1)$ that decreases in time, so slowing the weight adaption. The neighborhood $N_{j*}(t)$ decreases in size as time goes on, thus localizing the area of maximum activity.

5. Repeat from step 2.

The data are provided by an international manufacturing company that sells industrial parts to other (B2B) companies. The data set with 2881 observations includes customer data, total sales in 2019, sales for 5 different product lines, industry type, and product category (Table 3). The data include numeric as well as categorical variables. The data set also contains the latitude and longitude of the locations of the customers. Figure 3 shows how the 2881 customers are spread across the United States.

Table 3. Variables description.

| Variable	Description	N° of Observations
Customer Number	Unique customer ID	2881
Latitude	Y-Coordinate of customer locations	2881
Longitude	X-Coordinate of customer locations	2881
Product A-2019	Product A Revenue in 2019	2881
Product B-2019	Product B Revenue in 2019	2881
Product C-2019	Product C Revenue in 2019	2881
Product D-2019	Product D Revenue in 2019	2881
Product E-2019	Product E Revenue in 2019	2881
Ttl. Revenue (2019)	Total Revenue in 2019	2881
Industry_Type	Aerospace, Automotive, Machining	2881
Product_Category	Finished Goods, Repair, Spare Parts	2881
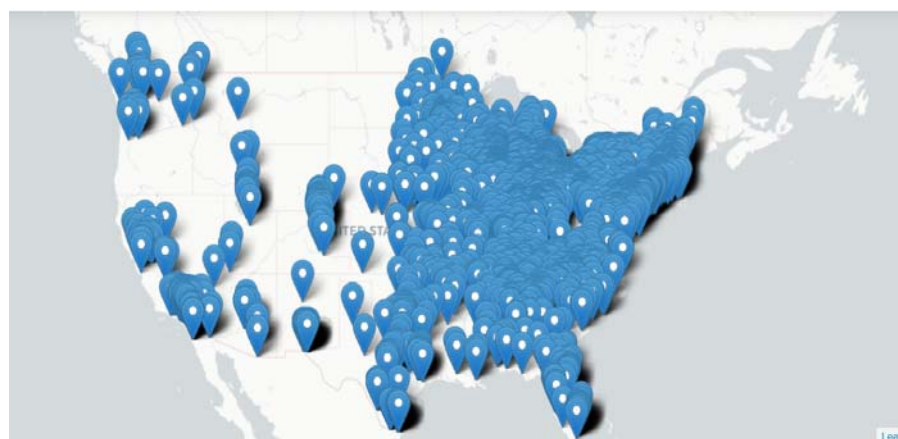
Figure 3. Distribution of customer locations based on their geo-coordinates.

The company wants to perform a customer segmentation to determine which customers with similar patterns should be grouped based on the product category (finished goods, spare parts, repair), and which customers are more service-sensitive and should be supported through its external marketing channels. The strategic objective is to develop a tool to be used by the marketing department to adjust the go-to-market strategy, i.e.,

to determine marketing policies for different customer groups based on their patterns and needs.

This study adopts the Jupyter notebook which is an open-source, browser-based tool acting as a virtual lab notebook to support scientific workflows, coding, data, and visualizations. The python programming environment offers different libraries which support deep learning (Geopy, Scipy), clustering (MiniSom, Kmeans), and spatial analysis and map visualization (Numpy, Pandas, Matplotlib, Seaborn, Folium). The python programming language will be used to perform the unsupervised training and classify the unlabeled data. This study provides a framework which enables the mapping and visualization of performance metrics and results.

The data set contains numeric data as well as categorical data. The latter will be transformed into suitable numeric values. Since the variables are measured at different scales, the data normalization is necessary when working with machine learning algorithms. The goal of data normalization is to convert the values of numeric variables into a common scale by avoiding any distortion of the data or a loss information. The MinMax scaling method will be used prior to the model fitting and training process:

$$y = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Before conducting the unsupervised SOM training, k-means clustering will be performed to benchmark the SOM' performance. By using different performance metrics, this study aims to evaluate the efficiency and accuracy of the unsupervised learning.

4. Results and Discussions

As discussed earlier, k-means algorithm aims to partition the data points into a predefined number of clusters (k) in which each point belongs to the cluster with the nearest mean. It starts by randomly selecting k centroids and assigning the points to the closest cluster, then it updates each centroid with the mean of all points in the cluster. To determine the optimal number (k) of clusters, the elbow method will be applied. This method plots the cluster variance as a function of the number of clusters and selects the k that flats the curve. As presented in Figure 4, the elbow method defines k = 7 as the optimal number of clusters in the data.
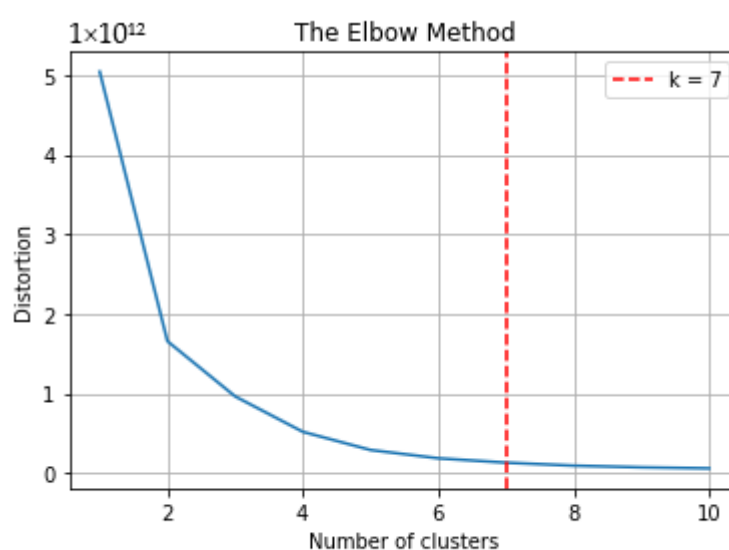

Figure 4. Elbow curve for k-means clustering.

Given the number of clusters k = 7 determined above, k-means clustering has been performed by using the python library KMeans. As a result, seven clusters including their centroids are plotted as a scatter plot shown in Figure 5. Since the data points are

geo-referenced, the clusters and their centroids are plotted by using their geographic coordinates. By applying k-means to the data set, it failed to provide useful clusters or distinguish clear patterns or boundaries between the clusters. Traditional clustering algorithms are often used in marketing analytics to perform market segmentation tasks. Based on the pre-defined number of clusters, marketing managers allocate the support capabilities for these clusters. Consequently, an incorrect assessment of the number of clusters can be misleading and thus affecting strategic decision making.
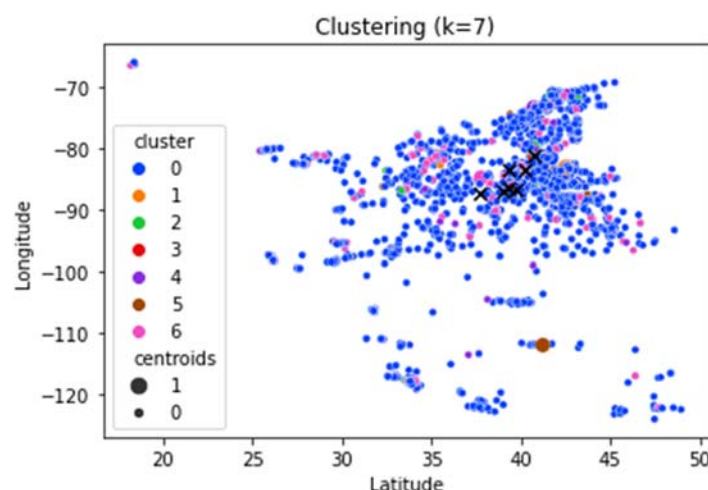

Figure 5. Elbow curve for k-means clustering.

By using the existing python library Minisom, the self-organizing maps (SOM) have been implemented. Minisom accepts two types of training: train_random and train batch. Train_random means that the model will use random samples from the data during training. In train batch mode, the samples are chosen in the given order inside the data set. In this paper, the train_random method was chosen.

Since the SOM algorithm uses unsupervised learning, the labels of the data will be ignored during the training process. Although an unsupervised training will be performed, the label of the target attribute (product category) will be kept only for the SOM visualization of the clusters but is not factored into the training. To determine the size of the SOM, a trial range of options was conducted. Finally, a map with a size 4×4 neurons and 1000 iterations was selected. This grid size gives the feature map enough space to spread out while displaying some overlaps between the classes. In this study, the SOM map uses a hexagonal grid and Gaussian neighborhood function. The other arguments for the Minisom function have been set as follows: Sigma (1.0) which defines the spread of the neighborhood function. The learning rate of 0.5 defines the initial learning rate for the SOM. The learning rate decreases linearly with the training iterations.

For a self-organizing map to be an accurate model, it must preserve the topology and neighborhoods of the input data while also fitting the data [87]. The quality measures are chosen for this study based on their usefulness. The literature recommends two types of quality measures, namely, quantization error and topological error.

The quantization error is a measure to evaluate the resolution of the mapping that can be considered inherent to the process of modeling [78]. The topological error, also known as topographic error, measures the topology preservation and the continuity of the mapping. It is defined by the proportion of all data vectors where the BMU and second BMU are not adjacent units [87]. Both quality measures have been computed during the training. At the end of the training with 1000 iterations, the quantization error reached 0.78. To understand how the training evolves, Figure 6 plots the quantization and topographic error of the SOM at each iteration step.
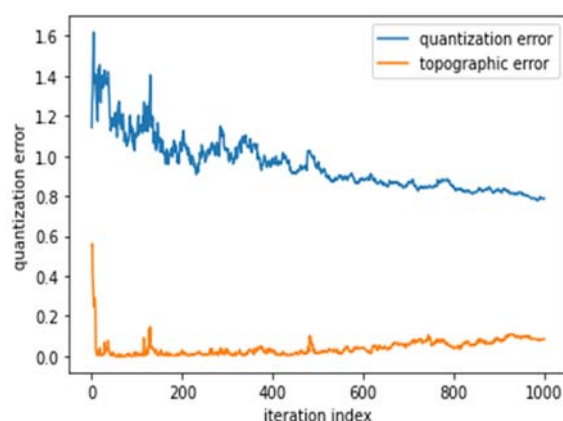
Figure 6. The 4 × 4 SOM with 1000 iterations quantization error and topographic error.

The SOM itself does not explicitly assign data items to clusters, nor does it identify cluster boundaries, as opposed to other clustering methods. Thus, visualizing the mapping created by the SOM is a key factor in supporting the user in the analysis process. A wealth of methods have been developed, mainly to visualize the cluster structures of the data [88].

To visualize the results of the training, the distance map also called U-Matrix (Figure 7) using a pseudo color where the neurons of the maps are displayed as an array of cells and the color represents the (weights) distance from the neighbor neurons.
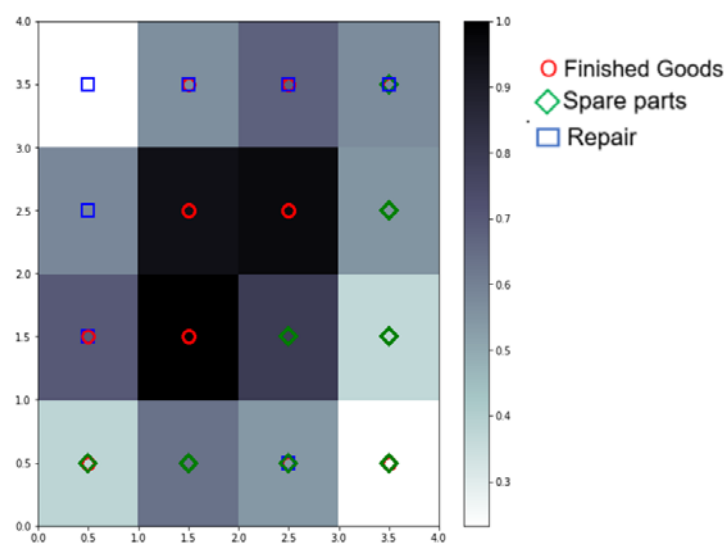

Figure 7. The 4 × 4 SOM with 1000 iterations plot distance with label markers.

The elements with similar patterns are located close to each other. The map representation visualizes the neurons density using colors and markers showing the magnitude. The darker regions show the density and which nodes are close to each other. Light shades mean they are far from one another. The clusters are divided by green (finished goods), red (spare parts), and blue (repair).

The SOM has separated each cluster into topologically distinct areas of the map. Some clusters have been located across different regions in the SOM space. Other clusters fill a distinct region and therefore the SOM has been effective. The U-Matrix also shows that regions with a high density of points are co-habited by data from multiple clusters.

To detect which neurons of the map have been activated more frequently, Figure 8 shows the clusters with different colors and reflects the activation frequencies. The SOM map clearly distinguishes the patterns in the data and visualizes that the customers can be segmented into 16 different clusters. In the bottom part of the map, some green and red clusters overlap in some areas which shows those clusters are similar to each other. In the

center of the map, the clusters are clearly differentiated. In the top side of the map, the clusters related to repair are clearly distinguished, while the top 2 clusters on the right side share similar properties.
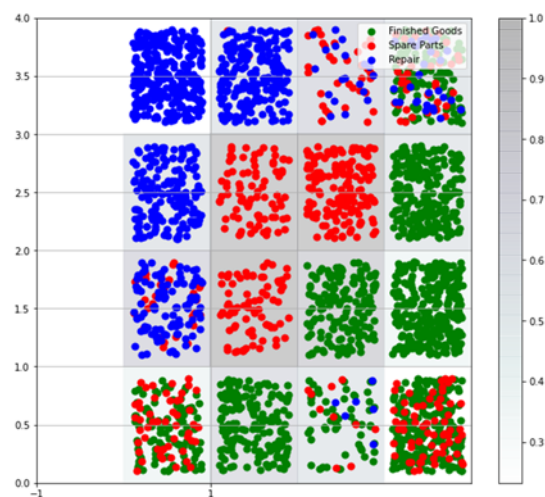

Figure 8. The 4 × 4 SOM with 1000 iterations plot distance and response frequencies.

Figure 9 displays the proportion of samples per class falling in a specific neuron using a pie chart per neuron. In other words, the class pies chart reveals four full clusters with customers related to finished goods, three full clusters with only customers buying spare parts, and three full clusters with repair customers. The remaining six clusters include mixed groups of customers with different patterns.
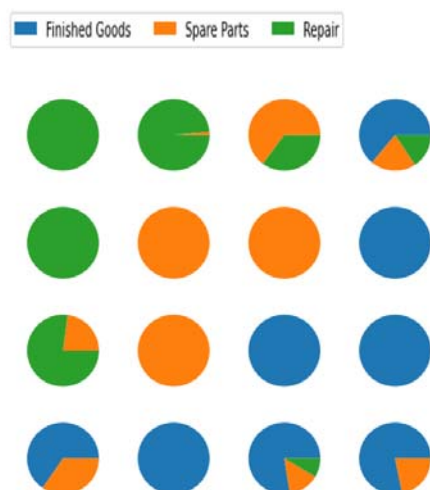

Figure 9. Class pies for 4 × 4 SOM with 1000 iterations.

By integrating the geo-coordinates of the data points during the training process, the SOM clusters have been be plotted according to their geographic distribution. Figure 10 exhibits the distribution of 16 SOM clusters as a scatterplot by using their latitudes and longitudes. Likewise, each cluster centroid has been assigned a pair of coordinates in the scatterplot. Figure 11 visualizes the geo-location of the clusters and their centroids across the US map which displays multiple high-density regions such as along the east-coast around Boston-New York and towards the south around North-Carolina-Florida. A high concentration of the clusters can be also seen in the eastern states such as Illinois, Indiana, Michigan, and Ohio. The visualization of cluster centroids in the geo-clustering map can be used for assigning the location of the future marketing channels.
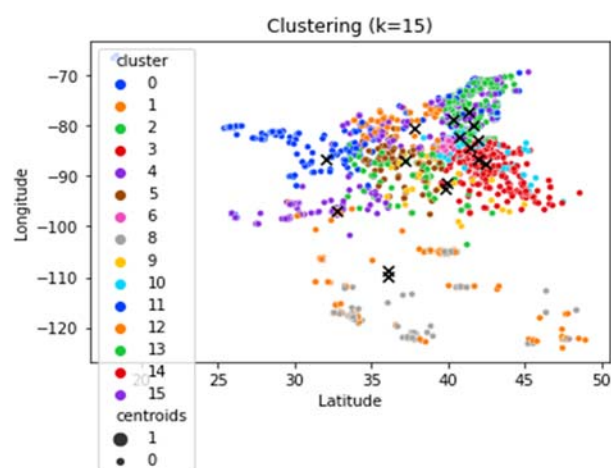
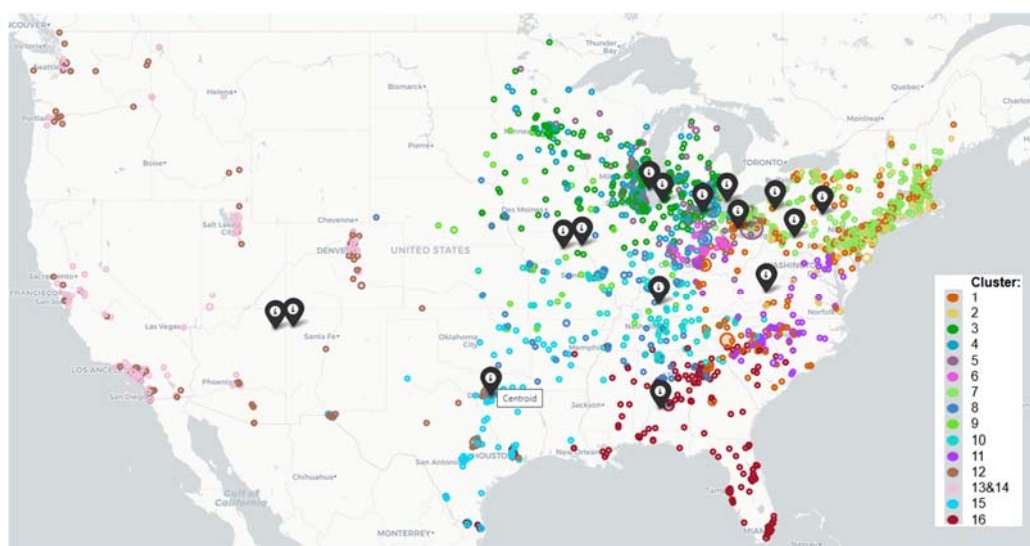**Figure 10.** Scatterplot for SOM 4 × 4 clustering with 1000 iterations.

Figure 11. Geo-clustering map by using 4 × 4 SOM with 1000 iterations.

The geo-clustering by the SOM enables a deeper understanding of how the clusters are spread out through space. Table 4 highlights how the clusters stretch over the different states. Due to a high industrial concentration in Michigan, Wisconsin, Illinois, Indian, Ohio, Pennsylvania, and New York, clusters C1 to C9 intersect with each other at least in one of these regions. C10 covers Missouri, Virginia, Tennessee, and overlaps with C11 in Kentucky. C12 extends over a relatively wide territory including Texas, Louisiana, New Mexico, Arizona, and Colorado. C13 and C14 include Arizona, Utah, San Francisco, and Nevada. C15 covers Arkansas, Oklahoma, and Tennessee. C16 covers mainly Florida but also extends with a lower density over Georgia, Alabama, Louisiana, Mississippi, and Tennessee. Most of the clusters cover several states which requires from future marketing channels a wide market coverage to extend the reach of their product offering. Overlapping clusters especially those across the main industrial manufacturing belt between Wisconsin and New York can be expected to increase the intensity of competition among the marketing channels. In this context, marketing channels are requested to introduce more differentiation. Offering extra services can help them to avoid price conflicts.

Table 4. Summary of spatial distribution of SOM clusters.

Clusters	Centers/Geo-Coordinates	Spatial Distribution
1	40.360841; −78.890987	Ohio, Pennsylvania, New York, Massachusetts, Connecticut, Maryland
2	41.379610; −77.218788	Pennsylvania, New York, Massachusetts
3	41.950033; −82.890716	Ontario, Michigan, Illinois, Indiana, Wisconsin
4	42.442861; −87.512295	Michigan, Wisconsin, Illinois, Ohio, Indiana
5	41.934342; −86.685673	Michigan, Indiana, Wisconsin
6	41.486035; −84.331184	Ohio, Indiana
7	40.740315; −82.233762	Ohio
8	41.604746; −80.024125	Pennsylvania; Iowa, Ohio, Indiana
9	39.968397; −91.513968	Missouri, Illinois, Iowa, South Dakota
10	39.830666; −92.607506	Kentucky, Missouri, Virginia, Tennessee
11	37.242167; −86.926336	Kentucky, Virginia, Maryland, Pennsylvania, Ohio
12	37.824302; −80.553858	Texas, Louisiana, New Mexico, Arizona, Colorado
13&14	36.126735; −108.637728	Arizona, Utah, San Francisco, Nevada
15	32.760898; −96.915516	Texas, Arkansas, Oklahoma, Tennessee
16	32.102329; −86.820285	Florida, Georgia, Alabama, Louisiana, Mississippi, Tennessee

A more precise understanding of the spatial cluster coverage implies the analysis of the relationships between the other segmentation attributes. The results of the SOM customer segmentation are exhibited in Table 5, which shows the sizes and the distinguishing characteristics of each cluster. The size of the clusters is based on the number of customers and the total sales generated in each cluster. C9, C12, and C16 represent the largest clusters. C1, C4, and C15 have a medium size. C2, C3, C5, C6, C7 and C8 are the smallest ones. All clusters contain customers from all over the industry types (automotive, aerospace, and machining), however, the SOM presents the product category as the main differentiating attribute. The clusters C4, C5, C9, C10, C13, and C14 consist of customers demanding finished goods only. These clusters extend over geographical territories with a high OEM density. The clusters C7, C11, C15, and C16 include only customers for repair services. For instance, C16 covers predominantly Florida which includes a low concentration of manufacturing companies but a large base of end users with a need for repair services. C8 and C12 reflect customers requiring spare parts. The remaining clusters combine two or three different product categories.

Table 5. Cluster characteristics according to SOM segmentation.

Clusters	Number of Customers	% in Total	Total Sales 2019	Industry Type			Product Category		
			USD	Automotive	Aerospace	Machining	Finished Goods	Spare Parts	Repair
1	202	7%	1,677,728	104	31	67	88	44	70
2	141	5%	367,479	63	30	48	111	0	30
3	133	5%	961,058	62	28	43	133	0	0
4	143	5%	1,143,886	75	30	38	0	92	51
5	104	4%	849,350	42	19	43	104	0	0
6	136	5%	618,963	76	24	36	119	0	17
7	44	2%	244,528	26	7	11	0	0	44
8	111	4%	677,657	55	20	36	0	111	0
9	418	15%	3,368,942	216	78	124	418	0	0
10	65	2%	418,376	29	17	19	65	0	0
11	186	7%	2,308,062	93	35	58	0	0	186
12	270	10%	3,893,410	137	60	73	0	270	0
13 & 14	330	12%	2,567,582	170	62	98	330	0	0
15	129	5%	1,380,176	52	29	48	0	0	129
16	294	11%	3,011,615	154	62	78	0	0	294

In this study, the SOM clustered the customer groups based on geographic, demographic, market-related, and product-related variables. Most of the clusters consist of customers who are homogeneous in terms of their specific requirements and benefits sought, thus allowing marketing channels to deploy targeted marketing campaigns that promote the products and services to each customer segment. Understanding the differentiating attributes of each segment in terms of needs and expectations enables marketers to derive appropriate channel selection criteria. All the sixteen clusters consist of cus-

tomers who belong to the different industry sectors, namely, automotive, aerospace, and machining. Hence, future marketing channels should possess market knowledge and direct access to these industry segments. Palmatier [5] considered the customer demand as a basis for determining the channel structure. As mentioned earlier, the demand type is captured in the data through the product category variable which has been presented by the SOM as a key differentiating attribute during the segmentation process. The promotion of finished goods requires technical competencies such as product sizing and configuration. Furthermore, the marketing channels need to offer physical selling to better support the customers on selecting the right products. Clusters with spare parts customers require from marketing channels inventory centers close to customer locations to ensure prompt deliveries of a product that is intended to be used as a replacement for installation on a machine that is in service. Moreover, e-commerce capabilities are beneficial in this product category to shorten time to market. In the repair segment, for instance C15 and C16, marketing channels need to have a qualified repair capacity as well as a customer hotline. Repair centers located close to the customers will help shorten travel distance and reduce downtime. Heterogenous clusters (C1, C2, C4, C6) with more than one product category require from the marketing channels matching the different criteria discussed above to fulfill the specific customer requirements of each segment.

5. Implications

This study yields several implications for research and practice. First, little attention has been paid in the literature to study geo-segmentation in B2B markets. While the existing literature focused more on B2C markets, this research on B2B segmentation provides insightful implications for predictive segmentation. Moreover, previous studies using unsupervised learning for customer segmentation have not placed emphasis on the cluster centroids and their usefulness to optimize the market coverage within the different clusters.

This paper also provides several practical implications. Using geo-marketing methods in conjunction with unsupervised deep learning allows the visualization of the spatial cluster distribution and detection of interdependencies between the different segmentation variables when making marketing decisions. The results of this study have demonstrated how the specific patterns of the clusters can be used to improve B2B channel management strategies. Without a deep understanding of how customers are segmented, firms often lack market focus and fail to allocate their limited resources efficiently. A differentiated cluster strategy should support creating clear roles and functions among the marketing channel members for each segment. Assigning dedicated channels for the different customer clusters and defining clear responsibilities will help minimize the channel conflicts and channel cannibalization. The model adopted in this paper demonstrates how the SOM can deploy the cluster centroids to localize where the future marketing channels can be ideally established. Moreover, the centroids can be used for an optimal allocation of the inventory locations and service centers. Well defined cluster territories and boundaries will help overcoming multichannel challenges such as overlaps and white spots in the market coverage.

Due to the increasing dimensionality of customer data, descriptive segmentation methods face serious limitations. User interventions on selecting of number of clusters or manual handling of the segmentation variables often lead to a bias in results and a loss in cluster efficiency. The adoption of deep learning methods into B2B marketing opens new horizons for marketers. In the era of big data, predictive analytics and location intelligence should be considered as an integral part of the modern geo-marketing. The results of this study provide insightful outputs for the marketing managers to improve their current clustering approach. Deep learning has become attractive to classify unlabeled data, detect non-linear and hidden relationships, and propose efficient clusters. Although self-organized learning does not involve an external teacher to supervise the learning process, the model requires technical capabilities such as programming and data analytics.

This requires a change in the skillset of the marketing managers to better leverage the power of artificial intelligence.

6. Conclusions

This research deals with a capability of unsupervised deep learning to perform geospatial clustering in the B2B marketing context. While the k-mean algorithm could not find any useful clusters out of the data set, the unsupervised SOM has proven its capability of detecting both spatially homogeneous and spatially heterogeneous regions across the US. As shown in this study, after the integration of the geographical coordinates in the SOM, combined with various non-spatial variables, the model performs well and produces well-defined clusters, which will allow the marketing channels to identify which products and services they should offer and which benefits they should promote.

This research has some limitations that should be highlighted to provide avenues for future research. For the scope of this research and the implementation of this study, only the data of the year 2019 have been considered. This paper suggests future research to investigate the historical trends including temporal changes of the patterns inside the clusters. In this context, it would be also useful to investigate how the clusters and their centroids shift their positions over time.

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

References

1. De Sarbo, W.; Blanchard, S.; Selin, A. *A New Spatial Classification Methodology for Simultaneous Segmentation, Targeting, and Positioning (STP Analysis) for Marketing Research*; Malhotra, N.K., Ed.; Review of Marketing Research; Emerald Group: Bingley, UK, 2009; Volume 5.
2. Fill, C. *Simply Marketing Communications*; Pearson Education: London, UK, 2006.
3. Havaldar, K. *Industrial Marketing: Text and Cases*; McGraw-Hill Education: New York, NY, USA, 2005.
4. Ghauri, P.; Cateora, P. *International Marketing*, 2nd ed.; Mc Graw Hill Education: New York, NY, USA, 2006.
5. Palmatier, R.; Louis, S.; El-Ansary, A. *Marketing Channel Strategy: An Omni-Channel Approach*; Routledge: Abingdon, UK, 2016.
6. Gupta, D. *Tourism Marketing*; Pearson Education India: Delhi, India, 2011.
7. Eldawy, A.; Niu, L.; Haynes, D.; Su, Z. Large scale analytics of vector + raster big spatial data. In Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Redondo Beach, CA, USA, 7–10 November 2017.
8. Alkathiri, M.; Jhummarwala, A.; Potdar, M. Multi-dimensional geospatial data mining in a distributed environment using MapReduce. *J. Big Data* **2019**, *6*, 82. [CrossRef]
9. Kaur, D.; Gill, N. *Artificial Intelligence and Deep Learning for Decision Makers: A Growth Hacker's Guide to Cutting Edge Technologies*; BPB Publications: Delhi, India, 2019.
10. Li, Z.; Gui, Z.; Hofer, B.; Li, Y.; Scheider, S.; Shekhar, S. Geospatial Information Processing Technologies. In *Manual of Digital Earth*; Guo, H., Goodchild, M.F., Annoni, A., Eds.; Springer: Singapore, 2020; pp. 191–227. [CrossRef]
11. Giannotti, F.; Pedreschi, D. *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*; Springer: Berlin/Heidelberg, Germany, 2008.
12. Diansheng, G.; Gahegan, M.; MacEachren, A.; Zhou, B. Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach. *Cartogr. Geogr. Inf. Sci.* **2005**, *32*, 113–132. [CrossRef]
13. Singh, N.; Ahuja, P. *Fundamentals of Deep Learning and Computer Vision: A Complete Guide to Become an Expert in Deep Learning and Computer Vision*; BPB Publications: Delhi, India, 2020.
14. Pedro, F.; Marques, G. *Handbook of Research on Big Data Clustering and Machine Learning*; IGI Global: Hershey, PA, USA, 2019.
15. Zimmerman, A.; Blythe, J. *Business to Business Marketing Management: A Global Perspective*; Routledge: Abingdon, UK, 2013.
16. Laasch, O. *Principles of Management: Practicing Ethics, Responsibility Sustainability*; SAGE: New York, NY, USA, 2013.
17. Shapiro, B.; Bonoma, T. *How to Segment Industrial Markets*; Harvard Business Review: Cambridge, MA, USA, 1984; pp. 104–110.
18. Raghavendra, G.; Hemanth, Y. *Marketing Management*; Wizard Publisher: Delhi, India, 2021.
19. Wind, Y.; Thomas, R. Segmenting Industrial Markets. *Adv. Bus. Mark. Purch.* **1994**, *6*, 59–82.
20. Lilien, G.; Grewal, R. *Handbook on Business to Business Marketing*; Edward Elgar Publishing: Camberley, UK, 2012.
21. Verhallen, T.; Frambach, R.; Prabhu, J. Strategy-Based Segmentation of Industrial Markets—An Investigation of the Market Orientation of Dutch Companies. *Ind. Mark. Manag.* **1998**, 27. [CrossRef]
22. Weinstein, A. *Handbook of Market Segmentation: Strategic Targeting for Business and Technology Firms*; Psychology Press: Abingdon, UK, 2004.

23. Griffith, R.; Louis, G. Segmenting Industrial Markets. *Ind. Mark. Manag.* **1994**, *23*, 39–46. [CrossRef]
24. Webster, F. *Industrial Marketing Strategy*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 1991.
25. Kotler, P. *Marketing Management*; Prentice Hall International Editions; Upper Saddle River: Hoboken, NJ, USA, 1994.
26. Johnson, G.; Flodhammer, Å. Some factors in industrial market segmentation. *Ind. Mark. Manag.* **1980**, *9*, 201–205. [CrossRef]
27. Plank, R. A Critical Review of Industrial Market Segmentation. *Ind. Mark. Manag.* **1985**, *14*, 79–91. [CrossRef]
28. Wind, Y.; Cardozo, R. Industrial market segmentation. *Ind. Mark. Manag.* **1974**, *3*, 153–165. [CrossRef]
29. Bonoma, T.; Shapiro, B. *Segmenting the Industrial Market*; Lexington Books: Washington, DC, USA, 1983.
30. Weinstein, A. Segmenting technology markets: Applying the nested approach. *Mark. Intell. Plan.* **2011**, *29*, 672–686. [CrossRef]
31. Dibb, S.; Simkin, L. Market Segmentation: Diagnosing and Overcoming the Segmentation Barriers. *Ind. Mark. Manag.* **2001**, *30*, 609–625. [CrossRef]
32. Freytag, P.; Clarke, A. Business to Business Market Segmentation. *Ind. Mark. Manag.* **2001**, *30*, 473–486. [CrossRef]
33. Blocker, C.; Flint, D. Customer Segments as Moving Targets: Integrating Customer Value Dynamism into Segment Instability Logic. *Ind. Mark. Manag.* **2006**, *36*. [CrossRef]
34. Weinstein, A.; Brotspies, H. Rethinking Business Segmentation: A Conceptual Model and Strategic Insights. *J. Strateg. Mark.* **2018**, *27*. [CrossRef]
35. Green, P.E.; Krieger, A.M. Alternative Approaches to Cluster-Based Market Segmentation. *J. Mark. Res. Soc.* **1995**, *37*, 221–239. [CrossRef]
36. Kotras, B. Mass personalization: Predictive marketing algorithms and the reshaping of consumer knowledge. *Big Data Soc.* **2020**. [CrossRef]
37. Syam, N.; Sharma, A. Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Ind. Mark. Manag.* **2018**, *69*. [CrossRef]
38. Huang, M.H.; Rust, R.T. A strategic framework for artificial intelligence in marketing. *J. Acad. Mark. Sci.* **2021**, *49*, 30–50. [CrossRef]
39. Sanjeev, V.; Rohit, S.; Subhamay, D.; Debojit, M. Artificial intelligence in marketing: Systematic review and future research direction. *Int. J. Inf. Manag. Data Insight* **2021**, *1*. [CrossRef]
40. Vasuki, A.; Govindaraju, S. *Deep Neural Networks for Image Classification*; IOS Press: Amsterdam, The Netherlands, 2017. [CrossRef]
41. Suzuki, K. *Artificial Neural Networks: Methodological Advances and Biomedical Applications*; BoD—Books on Demand: Norderstedt, Germany, 2011.
42. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Macmillan: New York, NY, USA, 1994.
43. Gollapudi, S. *Practical Machine Learning*; Packt Publishing Ltd.: Birmingham, UK, 2016.
44. Gulli, A.; Pal, S. *Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and More with TensorFlow 2 and the Keras API*, 2nd ed.; Packt Publishing Ltd.: Birmingham, UK, 2019.
45. Ghavami, P. *Big Data Analytics Methods: Analytics Techniques in Data Mining, Deep Learning and Natural Language Processing*; Walter de Gruyter GmbH: Berlin, Germany, 2019.
46. Buduma, N.; Locascio, N. *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*; O'Reilly Media Inc.: Sebastopol, CA, USA, 2017.
47. Moropoulou, A.; Korres, M.; Georgopoulos, A.; Spyrakos, C.; Mouzakis, C. Transdisciplinary Multispectral Modeling and Cooperation for the Preservation of Cultural Heritage. In Proceedings of the First International Conference, TMM_CH 2018, Athens, Greece, 10–13 October 2018.
48. Haykin, S. *Neural Networks and Learning Machines*, 3rd ed.; Prentice Hall: Hoboken, NJ, USA, 2009.
49. Pereira, B.; Rao, C.; Oliveira, F. *Statistical Learning Using Neural Networks: A Guide for Statisticians and Data Scientists with Python*; CRC Press: Boca Raton, FL, USA, 2020.
50. Silva, I.; Spatti, D.; Flauzino, R.; Liboni, L.; Alves, S. *Artificial Neural Networks: A Practical Course*; Springer: Berlin/Heidelberg, Germany, 2016.
51. Swingler, K. *Applying Neural Networks: A Practical Guide*; Morgan Kaufmann: Burlington, MA, USA, 1996.
52. Mather, P.; Tso, B. *Classification Methods for Remotely Sensed Data*; CRC Press: Boca Raton, FL, USA, 2016.
53. Tzanakou, E. *Supervised and Unsupervised Pattern Recognition: Feature Extraction and Computational Intelligence*; CRC Press: Boca Raton, FL, USA, 2017.
54. Du, K.; Swamy, M. *Neural Networks and Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2019.
55. Grillmeyer, O. *Exploring Computer Science with Scheme*; Springer: Berlin/Heidelberg, Germany, 2013.
56. Rutkowska, D. *Neuro-Fuzzy Architectures and Hybrid Learning*; Springer: Berlin/Heidelberg, Germany, 2001.
57. Zhang, Y. *Machine Learning*; BoD—Books on Demand: Norderstedt, Germany, 2010.
58. Berry, M.; Mohamed, A.; Yap, B. *Supervised and Unsupervised Learning for Data Science*; Springer: Berlin/Heidelberg, Germany, 2019.
59. Vidgen, R.; Kirshner, S.; Tan, F. *Business Analytics: A Management Approach*; Red Globe Press: Bloomsbury, UK, 2019.
60. Dolnicar, S.; Grün, B.; Leisch, F. *Market Segmentation Analysis: Understanding It, Doing It, and Making It Useful*; Springer: Berlin/Heidelberg, Germany, 2018.
61. Ester, M.; Frommelt, A.; Kriegel, H.; Sander, J. Algorithms for characterization and trend detection in spatial databases. In Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 27–31 August 1998; pp. 44–50.

62. Fan, W.; Wu, Z. Advances in Web-Age Information Management. In Proceedings of the 6th International Conference, WAIM 2005, Hangzhou, China, 11–13 October 2005.

63. Leung, Y. *Knowledge Discovery in Spatial Data*; Springer: Berlin/Heidelberg, Germany, 2010.

64. Neethu, C.; Subu, S. Review of Spatial Clustering Methods. *Int. J. Inf. Technol. Infrastruct.* **2013**, *2*, 15–24.

65. Aksoy, E. Clustering with GIS: An Attempt to Classify Turkish District Data. In Proceedings of the XXIII FIG Congress, Munich, Germany, 8–13 October 2006.

66. Lamb, D.; Downs, J.; Reader, S. Space-Time Hierarchical Clustering for Identifying Clusters in Spatiotemporal Point Data. *Int. J. GeoInf.* **2020**, *9*, 85. [CrossRef]

67. Azri, S.; Ujang, U.; Rahman, A.; Anton, F.; Mioc, D. 3D Geomarketing Segmentation: A higher spatial dimension planning perspective. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *42*, 1–7. [CrossRef]

68. Fincke, T.; Lobo, V.; Bação, F. Visualizing Self-Organizing Maps with GIS. Available online: https://www.semanticscholar.org/paper/Visualizing-self-organizing-maps-with-GIS-Fincke-Ba%C3%A7%C3%A3o/71f251e53953b988788e490c13a8d168e2da57fc#references (accessed on 3 April 2021).

69. France, S.L.; Ghose, S. Marketing analytics: Methods, practice, implementation, and links to other fields. *Expert Syst. Appl.* **2019**, *119*, 456–475. [CrossRef]

70. George, L.; Hogendorn, C. Geo-Targeting and the Market for Local News. 2019. Available online: https://ssrn.com/abstract=2357586 (accessed on 12 April 2021).

71. Cowgill, B. *Optimal Segmentation in Platform Strategy: Evidence from Geotargeted Advertising*; Columbia University Working Paper; Columbia University: New York, NY, USA, 2009.

72. Baye, I.; Reiz, T.; Sapi, G. *Customer Recognition and Mobile Geo-Targeting*; DICE Discussion Paper No. 285; DICE: Dusseldorf, Germany, 2018.

73. Chen, Y.; Li, X.; Sun, M. Competitive Mobile Geo-Targeting. *Mark. Sci.* **2017**, *36*, 666–682. [CrossRef]

74. Ezenkwu, C.; Ozuomba, S.; Kalu, C. Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services. *Int. J. Adv. Res. Artif. Intell.* **2015**, *4*. [CrossRef]

75. Holmbom, A.; Eklund, T.; Back, B. Customer Portfolio Analysis using the SOM. *Int. J. Bus. Inf. Syst.* **2011**, *8*. [CrossRef]

76. MacQueen, J. Some Methods for classification and Analysis of Multivariate Observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Davis, CA, USA, 21 June–18 July 1965; Volume 1, pp. 281–297.

77. Huang, D.; Bevilacqua, V.; Premaratne, P. Intelligent Computing Theories and Methodologies. In Proceedings of the 11th International Conference, ICIC 2015, Fuzhou, China, 20–23 August 2015.

78. Kohonen, T. *Self-Organizing Maps*, 3rd ed.; Springer: Berlin, Germany, 2001.

79. Brunton, S.; Kutz, J. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*; Cambridge University Press: Cambridge, UK, 2019.

80. Kohonen, T.; Simula, O.; Kangas, J. Engineering applications of the self-organizing map. *Proc. IEEE* **1996**, *84*, 1358–1384. [CrossRef]

81. Cord, M.; Cunningham, P. *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*; Springer: Berlin/Heidelberg, Germany, 2008.

82. Beale, R.; Jackson, T. *Neural Computing—An Introduction*; CRC Press: Boca Raton, FL, USA, 1990.

83. Graupe, D. *Principles of Artificial Neural Networks: Basic Designs to Deep Learning*, 4th ed.; World Scientific: Singapore, 2019.

84. Heaton, J. *Introduction to Neural Networks with Java*; Heaton Research, Inc.: St. Louis, MO, USA, 2008.

85. Riese, F.; Keller, S.; Hinz, S. Supervised and Semi-Supervised Self-Organizing Maps for Regression and Classification Focusing on Hyperspectral Data. *Remote Sens.* **2019**, *12*, 7. [CrossRef]

86. Priddy, K.; Keller, P. *Artificial Neural Networks: An Introduction*; SPIE Press: Bellingham, WA, USA, 2005.

87. Kiviluoto, K. Topology preservation in self-organizing maps. In Proceedings of the IEEE International Conference on Neural Networks, Washington, DC, USA, 3–6 June 1996; pp. 294–299.

88. Mayer, R.; Taha, A.; Rauber, A. *Visualising Class Distribution on Self-Organising Maps*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4669, pp. 359–368. [CrossRef]