ISSN 1870-4069

Machine Learning for the Retail Trade Behavior Analysis in Mexico

Patricia Soto-Vázquez, Guillermo Molero-Castillo,
Everardo Bárcenas, Rocío Aldeco-Pérez

Universidad Nacional Autónoma de México,
Facultad de Ingeniería,
Mexico

soto.holden@gmail.com, gmoleroca@fi-b.unam.mx,
{ebarcenas, raldeco}@unam.mx

Abstract. In recent years, machine-learning methods have gained prominence when used as a tool for data analysis in different areas such as the economy. This article presents the result of the data analysis of the retail trade in Mexico and the industrial branches that comprise it. For the data analysis, unsupervised learning was used, specifically clustering based on the K-means, through which it was possible to organize clusters with information on the characteristics of the different industrial branches of the retail trade. As a result, it was possible to identify that this sector, the source of many jobs, subsists in the midst of an aggressive, demanding market, with insufficient access to update its technology and complex administrative procedures.

Keywords: Machine learning, clustering, k-means, economics, retail trade.

1 Introduction

The use of information as a source of knowledge is not only limited to understand what the data represents, but also generates value in any area it is applied, which is essential in the context of the actual, globalized world. Thus, because of this constant and necessary search to make total utilization of knowledge, Artificial Intelligence, applied in Industry 4.0, benefits the society in the process of digital transformation, mainly driven by machine learning and deep learning [1].

Knowledge in economics is built on the analysis of information derived from the economic activity of a country [2]. From the categorization of economic activity, economic sectors emerge as pillars of growth and development [3] [4]. Therefore, within the economic context in which Mexico develops, a knowledge-based economy is necessary, with machine learning being an important pillar in the push towards the fourth industrial revolution, which has led to the application of data analysis algorithms in order to build an efficient and quality economic system.

In this sense, given this growing need for data analysis in the economic field, the use of machine learning, as a support tool for advanced data analytics, is important [5] [6], since it offers a wide variety of algorithms, among which are supervised, unsupervised, deep, reinforcement, and mixed, currently achieving an important position in response to the extensive digitization and storage of data. On the other side, current machine learning is not only changing the way a product is produced, marketed and sold, but is also part of the study of economic growth, determined by the increase in productivity and income of a country [7, 8].

This study establishes the basis for analyzing economic development, measured based on improvements in the living conditions of the population. There is no doubt that the development of new products and companies, with unprecedented levels of automation and robotization, can transversally transform the economy and the labor market. Thus, to understand how trade is directly associated with people quality of life, it is worthwhile to focus efforts on the analysis of one of the economic sectors that have a predominant impact on the Mexican economy, this is, retail trade; which is the economic activity defined by the individual sale of goods and services directly to final consumers [9].

This activity (by its nature) is a component of the supply chain in view to its model focused on the sale between the company and the consumer. This type of trade is a fundamental sector in Mexico, since, in terms of gross domestic product (GDP), tertiary activities had an annual percentage structure of 60%, corresponding to 2020, within which 9.2% corresponds to trade retail [10, 11]. In addition to the above, misinformation on the demeanor of economic activity not only leads people and companies to mismanage their business but encourages disinterest in establishing measures or laws that benefit the retail trade [12].

Consequently, this paper aims to show insight about retail trade in Mexico, which represents a field of opportunity, through an analysis, based on machine learning, since, moreover to its considerable percentage share with respect to GDP, it also concentrates a large population that finds, in this sector, a source of employment. The document is organized as follows, Section 2 presents the antecedents of economics as social science, some of the contributions of machine learning in the economy, discuss its applications, the use of algorithms and related work are also presented. Section 3 describes the method established as a proposed solution. Section 4 presents the results obtained, based on an example of application, and Section 5 summarizes some conclusions and future work.

2 Background

2.1 Retail Trade

Retail trade is defined by economic units, within which are several establishments that are under the control of a proprietary entity, permanently established and delimited by fixed facilities [10, 13]. Furthermore, these economic units are located at different geographical levels. For example, country, state, municipality, and locality, where they perform the task of enabling activities of buying and selling merchandise, or providing services, regardless of whether they have mercantile purposes [6] [15]. This group includes micro-businesses and small and medium-sized enterprises (SMEs) [14]. At present, the SME sector is one of the most vulnerable since, like any business, they require correct management of their financial income.

These incomes depend, largely, on proper financial management, which many SMEs lack. According to the Development Center for Business Competitiveness, 75% of SMEs close their operations just two years after being created. Moreover, the National Institute of Geography and Statistics (INEGI, by its acronym in Spanish) denotes that the new businesses in Mexico only live on average 7.7 years [16].

Conventionally, the study of economic growth is based on the analysis of indicators such as GDP, thanks to which the significant share of retail trade in the Mexican economy is notable. Although GDP is not a sufficient indicator to determine the economic growth of the country, it is one of the most important, since a rise in this indicator could easily translate into an increase in employment. In addition, the occupation and employment indicators provide relevant information in this area.

2.2 Economic Indicators

Concerning the employed population by size in the economic unit, the employment and occupation indicators establish that 20.1 million people are employed in micro-businesses, 7.5 million in small businesses, and 5.2 million in medium sized businesses.

This information is representative since it has a coverage of 63.9% (84556) of the dwellings in the National Survey of Occupation and Employment (ENOE, by its acronym in Spanish) [17]. Among the organizations that provide information on the activities, economic indicators, and the labor market include:

– The National Survey of Occupation and Employment, which is the primary source of information on the labor market. It provides monthly and quarterly data on the labor force, occupation, labor informality, underemployment, and unemployment [18]. In 2020, they disseminated the occupational characteristics of the population aged 15 years and over, along with demographic and economic variables for the analysis of the labor force.
– The Annual Trade Survey (EAC, by its acronym in Spanish), which provides information on commercial activities and provides a frequent statistical overview that contributes to the decision-making of the different productive sectors of the country [19]. The Annual Trade Survey is based on the Monthly Survey on Commercial Companies (EMEC, by its acronym in Spanish), whose main purpose is to generate statistical information.
– The World Trade Organization (WTO) is the only international organization, of which Mexico is a member, which deals with the rules that govern trade between countries. Its aim is to ensure that commercial exchanges take place in a fluid, predictable and free manner [20]. One of his recent publications was 'Helping MSMES Navigate the Covid-19 Crisis', which explains how SMEs have been affected by the COVID-19 pandemic.

2.3 Clustering Based Machine Learning

Machine learning consists of a set of algorithms for data-driven analysis, which allow establishing models, from the data of examples or experiences, to train machines (computers) and learn from them [8] [21]. Within machine learning, unsupervised methods are algorithms that base their training process on a previously defined, labelless data set. That is, no target or class value is known, either categorical or numeric. Therefore, these methods do not require human intervention [22].

The main applications of unsupervised learning are related to data clustering, where the objective is to find clusters with similar elements, in such a way that the internal elements of a cluster have a high similarity, and are different (dissimilar) with elements of others clusters [23]. There are two main types of clustering algorithms [24]:

1. Hierarchical, which produce a hierarchical organization of the elements that make up the data set, thus enabling different levels of clustering.
2. Partitional, which generate clusters of elements that do not correspond to any type of hierarchical organization. These algorithms are based on the distance between elements.

Partitional algorithms assume a priori knowledge of the number of clusters into which the data set must be divided, that is, they arrive at a division that optimizes a predefined criterion [25]. Among the algorithms that use this type of clustering highlights K-means, whose main idea is to define k centroids (one for each cluster) and then take each element (a record) from the database and place it in the nearest centroid cluster.

The centroid is a point that occupies the middle position in a cluster. The next step is to recalculate the centroid of each cluster and redistribute all elements according to the nearest centroid. The process is repeated until there are no longer changes in the clusters formed [24]. In addition, in K-means the elbow method is used, with different configurations of k, to obtain an approximation to the adequate number of clusters [26].

2.4 Related Work

In recent years, in Mexico, trade is the object of study due to its importance, not only in the national and international economy but also because of its impact on the Mexican population. In this sense, the analysis of retail trade is essential for the benefit of understanding the development and economic growth of a certain region over the years.

This type of analysis can be achieved based on the observation of similarities, trends, and behaviors, for which machine learning algorithms are useful, some works related to data grouping and partitioning algorithms in economics are:

- Data mining and machine learning in the retail business: developing efficiencies for better customer retention [6] presented an analysis of retail marketing and discussed the application of data mining and machine learning techniques. Within the methodology, the use of K-means as a tool for the identification of incomplete data points stands out.

  This algorithm, together with another one for predicting customer interest and pattern mining techniques, made it possible to identify purchase patterns from user records.
- P2V-MAP: Mapping of market structures for large retail assortments [15], where market structures were analyzed through advances in natural language processing and machine learning. The approach used made it possible to compare data dimensionality reduction techniques, which show a contribution to the market analysis. In addition, the use of machine learning algorithms is proposed to propose solutions in problems related to the structures of the retail market.

- Machine learning for enterprises: Applications, algorithm selection, and challenges [4], in this work the importance of machine learning applied to companies, is exposed, with the aim of promoting their technological development, and thus reducing costs of products and services. In addition, the use of methods and challenges in the application of machine learning algorithms for grouping, classification, and forecasting were analyzed. As well as the increase in the implementation of machine learning tools and algorithms in companies to increase their potential.
- Sustainability of SMEs in the Competition: A Systemic Review on Technological Challenges and SME Performance [27], in this paper the importance of SMEs as an engine of economic development was discussed, described the challenges they face, and reviewed the need for technological progress to drive innovation in the economy and the positive effects it has on production levels and economic growth. In addition, the adoption of information technologies as a means to face competitive challenges in SMEs was exposed.

Due to the growing need to increase research focused on the Mexican economy, based on machine learning, it is important to include, in the solutions, algorithms and varied approaches for understanding trade retail. The purpose is to identify evidence in the form of patterns from the data, with which various informed and thoughtful analyzes can be carried out on the current situation of retail trade and its impact on the national economy.

3 Method

The method defined for the analysis of the behavior of retail trade in Mexico was divided into four stages: a) data acquisition, b) exploratory data analysis, c) selection of variables, and d) algorithmic application.

3.1 Data Acquisition

The analyzed data were obtained from the Annual Trade Survey. This survey is based upon the economic units from the National Statistical Framework of Economic Units (MENUE, by its acronym in Spanish), supplied by the Mexican Business Statistical Registry (RENEM, by its acronym in Spanish) with referenced design variables [28].

The data source corresponds to open data available through the official website of the INEGI[1] , which is made up of data matrices on prime economic indicators of commercial activity by sector, subsector, and branch of economic activity at the national level. The global information is made up of 40 branches of economic activity and is made up of 2134549 businesses in the commercial sector. Of these, 18 branches belong to the wholesale trade (comprising 126933 business) and 22 to the retail trade (comprising 2007616 business), being, the latter, and the object of study in this research work.

Furthermore, the North American Industrial determines the variables used Classification System, which allows us to create clusters systematically, always under

[1] www.inegi.org.mx/app/descarga/ficha.html?tit=110334&ag=0&f=csv

Table 1. Aggregation levels of economic activities by sector, subsector, and industrial branch.

Code	Levels of aggregation
4611	Retail trade of groceries and food products
4612	Retail trade of beverages, ice, and tobacco
4621	Retail trade in self-service stores
4622	Retail trade in department stores
4631	Retail trade of textile products, except apparel
4632	Retail trade of clothing, costume jewelry, and clothing accessories
4633	Retail trade of footwear
4641	Retail trade of health care items
4651	Retail trade of perfumery and jewelry
4652	Retail trade of entertainment articles
4653	Retail trade of stationery, books, magazines, and newspapers
4659	Retail trade of pets, gifts, religious articles, disposables, handicrafts, and other articles for personal use
4661	Retail trade of household furniture and other household goods
4662	Retail trade of furniture, computer equipment and accessories, telephones, and others communication devices
4663	Retail sale of articles for interior decoration
4664	Retail trade of used goods
4671	Retail trade of hardware, plumbing, and glassware
4681	Retail trade of cars and trucks
4682	Retail trade of parts and spare parts for automobiles, vans, and trucks
4683	Retail trade of motorcycles and other motor vehicles
4684	Retail trade of fuels, oils, and lubricating grease
4691	Retail trade exclusively through the Internet, and printed catalogs, television, and similar

the same logic, which helps to avoid controversies and errors of interpretation [29]. Further, within the retail trade, the stratification by a number of workers in each company according to INEGI is defined as [30]: i) micro (up to 10 people), ii) small (11 to 30 people), and iii) medium (31 to 100 people).

3.2 Exploratory Data Analysis

The analysis period was from 2016 to 2019, since in 2020 the final figures were published as part of the 2019 Economic Censuses. In this sense, an exploratory analysis was initially carried out on the data set, which was useful to know the data and understand its main characteristics. Thus, based on data exploration, it was observed that the data structure is made up of 61 variables that represent economic activities.

The data recorded are non-negative numbers, which were grouped into number of establishments, number of people, and thousands of pesos (national currency). Also, there are no null values. It was also observed that there are no out-of-range values. Table 1 shows the levels of aggregation, where the first two digits correspond to the sector (46), the first three digits to the subsector, and the four digits as a whole to the branch of economic activity.

The degree of linear relationship between pairs of variables was also measured. This information was structured in a correlation matrix, finding mostly a weak relationship between the variables. However, those variables that presented a certain relationship were those belonging to the same economic activity. In addition, it was necessary to carry out a selection of variables. This selection allowed focusing the analysis on the significant variables of different economic activities of the retail trade.

3.3 Feature Selection

From each of the 22 retail trade branches, shown in Table 1, a set of 58 variables was obtained, from which categorical variables were filtered, and the year was discarded because it inherently already represents a clustering of data, which is to be avoided, since the objective is to obtain a segmentation that combines all the variables through the measurement of similarities between the available elements. In this sense, the selected variables offer information on the following categories:

1. Stratum of the business: micro, small or medium.
2. Types of establishments: auxiliary or commercial (number of establishments).
3. Dependent and non-dependent personnel: women and men (number of people).

The other selected variables represent different significant activities of the retail trade, such as: a) consumption of goods and services; b) taxes on the activity; c) net sales; and d) fixed assets. These financial activities represent the set of operations that are executed in the supply and demand market, whose path is the acquisition of income and the realization of expenses.

3.4 Algorithm Application

For the cluster analysis, the K-means algorithm was used due to its functionality and efficiency characteristics, where through a method, such as the Elbow, it is possible to define the appropriate number of clusters into which the data vectors should be divided (elements), which make up the data set.

Furthermore, through this algorithm optimization problem are dealt with, in which the elements are distributed in K clusters so that the sum of the internal variances of all of them is as low as possible. Thus, in order to find similarities in the different branches of retail trade, the algorithm was implemented in Python, in such a way that clusters were generated based on the following process of assigning elements and updating centroids:

Start: centroids chosen during each iteration were established randomly for the formation of clusters.

1. Assignment: each data point (vector) was assigned to its nearest centroid.
2. Update: the average of all assigned points in the cluster was calculated to set the new centroid.
3. Repeat: steps 2 and 3 were repeated iteratively until the centroids no longer changed.

The elements were assigned employing the minimum distances, measured through the Euclidean distance, between each element and the centroids, achieving thus, a high intra-cluster similarity and low inter-cluster similarity. The equation of the Euclidean distance is as follows:

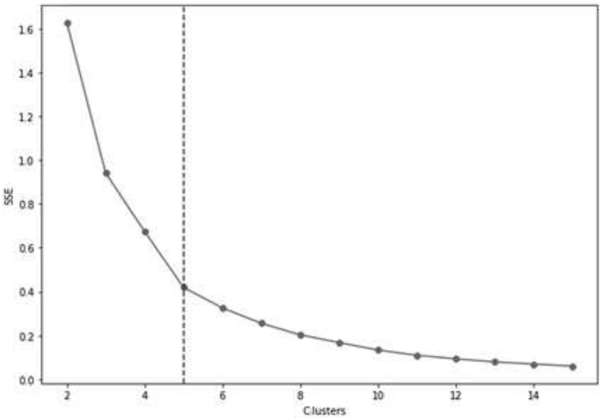

Fig. 1. Elbow method for identifying the adequate number of clusters.

$$\mathrm{dist}(p, c) = \sqrt{\sum_{i=1}^{n} (p_i - c_i)^2}\,, \tag{1}$$

where, p = {p_1, …, p_n} are the elements of the data set and c = {c_1, …, c_k} corresponds to the centroids. Therefore, by virtue of the need for a priori knowledge about the adequate number of clusters, a range of k configurations was established for the implementation of the algorithm.

This range allowed the algorithm to be run iteratively to obtain the clusters. Subsequently, based on the resulting categorization, the sum of the squared error (SSE) between each element of the formed cluster and its closest centroid was calculated. This SSE estimate was for each configuration of k based on the following equation:

$$\mathrm{SSE} = \sum_{k=1}^{k} \mathrm{dist}(p_i, c_i) = \sum_{k=1}^{k} \sum_{p_i \in C_k} (p_i - c_i)^2\,. \tag{2}$$

Since this is a measure of error, the goal of K-means is to try to minimize this value. This measurement of error is used to carry out the elbow method, in which a curve is drawn with the values obtained from SSE to find an inflection point (elbow), through which the optimal number of clusters to be analyzed is established. Figure 1 shows the layout of this curve, in which it is observed that the elbow effect suddenly changes its orientation in k equal to 5.

4 Results

From the results obtained, a differentiated segmentation was observed for the evaluation period (2016-2019). Thus, based on the internal similarities of the clusters and the values of their centroids, Table 2 shows a summary of their most significant characteristics, determined by the aspects in which each cluster stands out for having greater or lesser participation in the retail trade activities; stratum (micro, small and medium-sized business), types of establishments and personnel (dependent and non-dependent); and consumption, taxes, sales and fixed assets.

Cluster 1 is distinguished by having the least amount of money directed to the tax payment levied on commercial activity, and specific taxes on the products sold. Also characterized by the lack of control and tax burden. A case is the businesses that sell through the Internet, which, despite the increase in the consumption of products and services through digital platforms, the payment of their taxes is still low. This causes an impact on tax collection, generating administrative and control weaknesses. According to the economic study carried out by the OECD, tax revenues in this branch continue to be low and fiscal policy has a lower redistributive impact.

This causes an impact on tax collection, generating administrative and control weaknesses for the payment of taxes. Contrary to the previous cluster, Cluster 2 is characterized by allocating the largest amount of resources to consumption, tax payments, sales, and fixed assets. This behavior is because the retail trade, for example, in self-service stores, it stands out for having a higher productivity and more efficient distribution of products. Cluster 3 stands out for being made up of the grocery and food retail trade, as one of the most common industrial branches in Mexican society, where competition is local.

In addition, this type of commerce consists mainly of micro-businesses, since they have fewer barriers thanks to the behavior of consumers, who go to the stores closest to their home to obtain better prices, greater variety, or another benefit. Cluster 4 has differences between the industrial branches that comprise it, since they do not follow the same trend throughout the period analyzed.

This may be due to the fact that these branches have undergone a process of transformation in recent years, for example, the retail trade of hardware stores and glass items has presented significant declines with respect to foreign investment. For its part, Cluster 5 is characterized by having increased participation of small and medium-sized businesses.

This is important because, during 2020, the same trend continued according to the information provided by Data Mexico, there was an in-crease of 15.1% in small businesses and 10.6% in medium businesses, this compared to the previous year (2019). Meanwhile, the micro businesses continued with a fall of 2.45%.

Regarding the participation in the different financial items and activities, it was observed that Cluster 1, unlike Cluster 2, is the one with the most industrial branches and that they have lower participation in all financial activities. This confirms the characteristics mentioned in the background section regarding SMEs and how they survive in the midst of an aggressive and demanding market.

It was also observed that clusters 3 and 5 have greater participation in the three strata (micro, small and medium business), compared to clusters 1, 2, and 4. On the other side, in the conformation of the five clusters, there are industrial branches that have participated in more than one cluster, such as:

1. Retail trade of the furniture, computer equipment, and accessories, telephones, and other communication devices,
2. Retail trade of parts and spare parts for automobiles, vans, and trucks; and

Table 2. Summary of the clusters obtained.

Cluster	Characteristics
1	**Industrial branch** — Beverages, ice, and tobacco | Textile products, except apparel | Clothing, costume jewelry and clothing accessories | Footwear | Perfumery and jewelry articles | Entertainment articles | Stationery, books, magazines, and newspapers | Pets, gifts, religious articles, disposables, handicrafts and other articles for personal use | Household furniture and other household goods | Furniture, computer equipment and accessories, telephones, and other communication devices (2016 and 2017) | Interior decorating articles | Used goods | Parts and spare parts for automobiles, vans, and trucks (2016 and 2017) | Motorcycles and other motor vehicles | Retail trade exclusively through the Internet, and printed catalogs, television and similar. **Stratum, types of establishments and personnel (dependent and non-dependent)** — The lowest number of small and medium-sized companies. — The lowest number of auxiliary establishments. — The lowest number of dependent personnel (men) and non-dependent personnel (men and women). **Consumption, taxes, sales, and fixed assets** — The lowest consumption of merchandise, materials, raw and auxiliary materials. — The lowest number of taxes levied on the activity and specific to the products. — The lowest net sales of merchandise, manufactured products, services rendered, rental of movable and immovable property. — The lowest purchase and sale of machinery and production equipment, real estate, transportation units and equipment, computer and peripheral equipment, furniture, office equipment, and other fixed assets.
2	**Industrial branch** — Retail trade in department stores. **Stratum, types of establishments and personnel (dependent and non-dependent)** — The greatest participation of non-dependent personnel (men and women). **Consumption, taxes, sales, and fixed assets** — The biggest amount of money for consumption of merchandise, fuels and lubricants, electrical energy, containers, and packaging; payments for the rental of movable and immovable property, personnel non-dependent, advertising, and communication services. — The highest number of specific taxes on products. — The highest net sale of merchandise, manufactured products, income from services rendered, rental of movable and immovable property. — The greatest purchase and sale of machinery and production equipment, real estate, computer and peripheral equipment, furniture, office equipment, and other fixed assets.
3	**Industrial branch** — Groceries and food products | Hardware, plumbing, and glassware products (2018 and 2019). **Stratum, types of establishments and personnel (dependent and non-dependent)** — The greatest number of micro-companies. — The greatest number of auxiliary and commercial establishments. — The greatest number of dependent personnel (men and women). **Consumption, taxes, sales, and fixed assets** — The greatest number of materials consumed for services rendered, raw and auxiliary materials. — The greatest number of taxes levied on the activity. — The lowest consignment and commission income.
4	**Industrial branch** — Retail trade in department stores | Health care items | Hardware, plumbing and glassware products (2016 and 2017) | Cars and trucks | Furniture, computer equipment and accessories, telephones and other communication devices (2018 and 2019) | Parts and spare parts for automobiles, vans, and trucks (2018 and 2019). **Stratum, types of establishments and personnel (dependent and non-dependent)** — The lowest number of small companies. **Consumption, taxes, sales, and fixed assets** — The greatest consignment and commission income.
5	**Industrial branch** — Retail trade of fuels, oils and lubricating grease. **Stratum, types of establishments and personnel (dependent and non-dependent)** — The greatest participation of small and medium-sized companies. — The lowest number of commercial establishments. — The lowest number of dependent personnel (women). **Consumption, taxes, sales, and fixed assets** — The greatest purchase and sale of transportation units and equipment.

3. Retail trade of hardware and glass items. This behavior is due to the years of analysis, from 2016 to 2019, associated with the 22 branches of the retail trade.

Consequently, it is possible to observe that the different trends that prevailed over the years, in the different industrial branches, are a reflection of the economic activity of the country and its industrial sectors. Activities such as imports directly affect merchants' suppliers. On the other hand, foreign investment in different industrial sectors also has a direct impact on retail businesses.

Sectors such as manufacturing industries, food and beverage preparation, wholesale trade, real estate services, construction, mining, agriculture, animal husbandry and exploitation, forestry, fishing, and hunting; as well as generation, transmission, distribution, and commercialization of electrical energy, supply of water and natural gas through pipelines to the final consumer (in this case for use in economic units dedicated to retail trade), have a great impact on the items analyzed for each industrial branch.

Certainly, it is important to analyze the market in which the retail trade operates from the point of view of the consumer and how its decisions affect the retailer. The results obtained provide the basis for understanding the power that the consumer has within commerce in Mexico. For example, today supermarkets have displaced retailers, causing consumers to prefer supermarkets instead of small businesses.

This is due to the ability of large companies to offer lower prices and offer a huge variety of products. These characteristics have led some retailers out of business. To avoid business closures, the participation of retail businesses in tax regimes should be promoted.

For example, the Ministry of Finance and Public Credit, together with the Tax Administration Service, must continue with the incorporation of SMEs to the Tax Incorporation Regime (RIF), to obtain benefits, such as discounts on income tax (ISR), deduction of payments, issuance of electronic invoicing, social security, financing or credits. However, pro-retail policies for many entrepreneurs are partial solutions, causing them to fail or survive in the midst of an aggressive and demanding market.

5 Conclusions

The retail trade has experienced risks and difficulties over the years that impede its development and growth. Since this type of trade is an important part of the economic and social stability of a country, it is essential to understand its behavior, which, hand in hand with specialized machine learning algorithms, is possible to do so. The importance of data in economics is useful for the development of social impact studies that provide relevant information. Under this idea, working with open data from the retail trade represented a significant challenge, since key variables were identified for obtaining useful results on the population dedicated to the retail trade.

In line with the above, it is essential to highlight the importance of retail trade within the Mexican economy, considering that it is the sector to which many SMEs are dedicated and, in terms of GDP, it is the source of numerous jobs. However, it is also struggling to survive in the market. These SMEs subsist in the midst of the vulnerability of their businesses since they lack financial management, which does not allow them to develop within a demanding market. They also do not have the resources to update their technology. From the foregoing, the need for policies to benefit workers dedicated to the retail trade without requiring complex administrative procedures is derived.

The incorporation of retail businesses into tax regimes is a solution that seeks the protection and well-being of workers in any eventuality. However, it is a complex activity to manage for most workers, so focusing efforts to promote support programs for micro, small and medium-sized businesses is essential for their growth and economic development. On the other side, in order to identify strengths, weaknesses, opportunities, threats and even risks that the retail trade faces, the use of machine learning algorithms becomes essential, since they provide a way to the resolution of specific problems. For this reason, the application of the K-means clustering algorithm allowed a better understanding of the behavior of the retail trade.

Undoubtedly, making use of clustering algorithms proved to be a powerful tool for observing the dynamism over the years of the trade sector, thus confirming the importance of applying machine learning as a support tool in data analytics in the field of economics and, with this, to know the growth economic of Mexico. It was observed that the retail trade has experienced risks and difficulties that impede its development and growth, therefore, measures must be established to guarantee support plans, regardless of their size; provide guidance on the shortage of skilled labor; advise on subsistence in the sector after the crisis due to the COVID-19 pandemic; and diversify sales channels, especially by helping small physical retailers sell online.

As future work, it is intended to make a new analysis with updated information to enrich the results obtained. This may be important due to the behavior of the COVID-19 pandemic and its impact on the retail trade, which, according to sources consulted, registered a significant drop, compared to months prior to said pandemic. In addition, it would be relevant to focus efforts on the visualization of results through visual resources such as graphics, maps or a graphical interface aimed at interested users and thus contribute to decision-making, or information analysis in this area.

References

1. Molero-Castillo G., Maldonado-Hernández G., Mezura-Godoy C., Benítez-Guerrero E.: Interactive system for the analysis of academic achievement at the upper-middle education in Mexico. Computación y Sistemas, vol. 22, no. 1, pp. 223–233 (2018) doi: 10.13053/CyS-22-1-2773
2. Montuschi, L.: Datos, información y conocimiento. De la sociedad de la información a la sociedad del conocimiento. Universidad del CEMA, vol. 192, no. 6, pp. 2–32 (2001)
3. Piedras, E.: Industrias y patrimonio cultural en el desarrollo económico de México. Cuicuilco, vol. 13, no. 38, pp. 29–46 (2006)
4. Lee, I., Shin, Y. J.: Machine learning for enterprises: Applications, algorithm selection, and challenges. Business Horizons, vol. 63, no. 2, pp. 157–170 (2020)
5. Hansen, S.: Aplicación del aprendizaje automático al análisis económico y la formulación de políticas. Papeles de economía española, vol. 157, pp. 216–234 (2018)
6. Kumar, M. R., Venkatesh, J., Rahman, A. M.: Data mining and machine learning in retail business: Developing efficiencies for better customer retention. Journal of Ambient Intelligence and Humanized Computing, pp. 1–13 (2021) doi: 10.1007/ s12652-020-02711-7
7. Quiroga-Persivale, G.: ¿Qué es la inteligencia artificial y cómo se aplica en los negocios? (2018)

8. Mathur, P.: Overview of machine learning in retail. Machine Learning Applications Using Python. Apress, Berkeley, CA, pp. 147–157 (2019) doi: 10.10 07/978-1-4842-378787
9. Comercio al por menor. http://centro.paot.org.mx/documentos/inegi/comercio menor.pdf
10. INEGI: Producto interno bruto trimestral: Por actividad económica.
11. INEGI: Glosario.
12. Arana, D.: Pymes mexicanas, un panorama para 2018 (2017)
13. Ávila-Lugo, J.: Introducción a la economía. Ed. Plaza y Valdez, México, pp. 390, ISBN: 970-722-256-5 (2007)
14. INEGI: Clasificación para actividades económicas. Encuesta Nacional de Ocupación y Empleo (ENOE), www.inegi.org.mx
15. Gabel, S., Guhl, D., Klapper, D.: P2V-MAP: Mapping market structures for large retail assortments. Journal of Marketing Research, vol. 56, no. 4, pp. 557–580 (2019) doi: /10.1177/00222437198336
16. INADEM: Conflictos en el emprendimiento. www.inadem.gob.mx/conflictos-en-el-emprendimiento
17. ENOE: Resultados del tercer trimestre de 2020. www.inegi.org.mx/contenidos/programas/enoe/15ymas/doc/enoe_n_presentacion_ejecutiva_trim3.pdf
18. INEGI: Encuesta nacional de ocupación y empleo (ENOE), población de 15 años y más de edad. www.inegi.org.mx/programas/enoe/15ymas/
19. INEGI: Encuesta anual del comercio 2019. www.inegi.org.mx/prog ramas/eac/2013/
20. OMC: México y la OMC. Available in: www.wto.org/spanish/thewto_s/countries s/mexico_s.htm
21. Palma Méndez, J. T., Marín Morales, R.: Inteligencia artificial: Métodos, técnicas y aplicaciones. Madrid, MacGraw-Hill, pp. 1022 (2008)
22. Rouhiainen, L.: Inteligencia Artificial. Madrid, Alienta Editorial, https://static0 planetadelibroscom.cdnstatics.com/libros_contenido_extra/40/39308_Inteligencia_artificial.pdf
23. Cambronero C. G., Moreno, I. G.: Algoritmos de aprendizaje: KNN & Kmeans, www.it.uc3m.es/~jvillena/irc/practicas/08-09/06.pdf
24. Pla, D., Pascual, F., Sánchez, S.: Algoritmos de agrupamiento. Métodos Informáticos Avanzados, pp. 164–174 (2007)
25. Soto, A. J., Ponzoni, I., Vazquez, G. E.: Análisis numérico de diferentes criterios de similitud en algoritmos de clustering. Mecánica Computacional, pp. 993–1012 (2006)
26. Pérez, M.: Aplicación de Kmeans y SOM.
27. Prasanna, R., Jayasundara, J., Naradda Gamage, S., Ekanayake, E., Rajapakshe, P., Abeyrathne, G.: Sustainability of SMEs in the Competition: A systemic review on technological challenges and SME performance. Journal of Open Innovation: Technology, Market, and Complexity, vol. 5, no. 4, pp. 1–18 (2019)
28. EAC: Síntesis metodológica: Encuestas Económicas Nacionales
29. SCIAN: NAICS–SCIAN. https://naics-scian.inegi.org.mx/naics_scian/defaul t_e.aspx
30. INEGI: Censos económicos 2019. Micro, pequeña, mediana y gran empresa, www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825198657.pdf