# UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

**Escuela de Ciencias Matemáticas y Computacionales**

## TÍTULO:

Risk analysis of stocks markets by a merged unsupervised model, time evolution comparison, and optimization.


**Autor:**

Gissela Estefania Pilliza Chicaiza

**Tutor:**

Oscar Chang Tortolero., Ph.D.


Urcuquí, octubre del 2020.

Urcuquí, 27 de octubre de 2020

**SECRETARÍA GENERAL**
**(Vicerrectorado Académico/Cancillería)**
**ESCUELA DE CIENCIAS MATEMÁTICAS Y COMPUTACIONALES**
**CARRERA DE TECNOLOGÍAS DE LA INFORMACIÓN**
**ACTA DE DEFENSA No. UITEY-ITE-2020-00037-AD**

A los 27 días del mes de octubre de 2020, a las 10:00 horas, de manera virtual mediante videoconferencia, y ante el Tribunal Calificador, integrado por los docentes:

| | |
|---|---|
| **Presidente Tribunal de Defensa** | Dr. CUENCA LUCERO, FREDY ENRIQUE , Ph.D. |
| **Miembro No Tutor** | Dr. IZA PAREDES, CRISTHIAN RENE , Ph.D. |
| **Tutor** | Dr. CHANG TORTOLERO, OSCAR GUILLERMO , Ph.D. |

El(la) señor(ita) estudiante **PILLIZA CHICAIZA, GISSELA ESTEFANIA**, con cédula de identidad No. **1723293567**, de la **ESCUELA DE CIENCIAS MATEMÁTICAS Y COMPUTACIONALES**, de la Carrera de **TECNOLOGÍAS DE LA INFORMACIÓN**, aprobada por el Consejo de Educación Superior (CES), mediante Resolución **RPC-SO-43-No.496-2014**, realiza a través de videoconferencia, la sustentación de su trabajo de titulación denominado: **RISK ANALYSIS OF STOCKS MARKETS BY A MERGED UNSUPERVISED MODEL, TIME EVOLUTION COMPARISON, AND OPTIMIZATION.** , previa a la obtención del título de **INGENIERO/A EN TECNOLOGÍAS DE LA INFORMACIÓN**.

El citado trabajo de titulación, fue debidamente aprobado por el(los) docente(s):

| | |
|---|---|
| **Tutor** | Dr. CHANG TORTOLERO, OSCAR GUILLERMO , Ph.D. |

Y recibió las observaciones de los otros miembros del Tribunal Calificador, las mismas que han sido incorporadas por el(la) estudiante.

Previamente cumplidos los requisitos legales y reglamentarios, el trabajo de titulación fue sustentado por el(la) estudiante y examinado por los miembros del Tribunal Calificador. Escuchada la sustentación del trabajo de titulación a través de videoconferencia, que integró la exposición de el(la) estudiante sobre el contenido de la misma y las preguntas formuladas por los miembros del Tribunal, se califica la sustentación del trabajo de titulación con las siguientes calificaciones:

| Tipo | Docente | Calificación |
|---|---|---|
| Miembro Tribunal De Defensa | Dr. IZA PAREDES, CRISTHIAN RENE , Ph.D. | 8,0 |
| Tutor | Dr. CHANG TORTOLERO, OSCAR GUILLERMO , Ph.D. | 10,0 |
| Presidente Tribunal De Defensa | Dr. CUENCA LUCERO, FREDY ENRIQUE , Ph.D. | 9,8 |

Lo que da un promedio de: **9.3 (Nueve punto Tres)**, sobre 10 (diez), equivalente a: **APROBADO**

Para constancia de lo actuado, firman los miembros del Tribunal Calificador, el/la estudiante y el/la secretario ad-hoc.

Certifico que *en cumplimiento del Decreto Ejecutivo 1017 de 16 de marzo de 2020, la defensa de trabajo de titulación (o examen de grado modalidad teórico práctica) se realizó vía virtual, por lo que las firmas de los miembros del Tribunal de Defensa de Grado, constan en forma digital.*

Firmado electrónicamente por:
GISSELA ESTEFANIA PILLIZA

PILLIZA CHICAIZA, GISSELA ESTEFANIA
**Estudiante**

Firmado electrónicamente por:
FREDY ENRIQUE CUENCA LUCERO

Dr. CUENCA LUCERO, FREDY ENRIQUE , Ph.D.
**Presidente Tribunal de Defensa**

Firmado electrónicamente por:
OSCAR GUILLERMO
CHANG TORTOLERO

Dr. CHANG TORTOLERO, OSCAR GUILLERMO , Ph.D.
**Tutor**

CRISTHIAN RENE IZA PAREDES
Digitally signed by CRISTHIAN RENE IZA PAREDES
Date: 2020.11.09 23:13:34 -05'00'

Dr. IZA PAREDES, CRISTHIAN RENE , Ph.D.
**Miembro No Tutor**

TATIANA BEATRIZ TORRES MONTALVAN
Firmado digitalmente por TATIANA BEATRIZ TORRES MONTALVAN
Fecha: 2020.11.09 11:31:05 -05'00'

TORRES MONTALVÁN, TATIANA BEATRIZ
**Secretario Ad-hoc**

# Autoría

Yo, **GISSELA ESTEFANIA PILLIZA CHICAIZA**, con cédula de identidad 1723293567, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de la autora del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, Agosto 2020.

Firmado electrónicamente por:
**GISSELA ESTEFANIA PILLIZA**

Gissela Estefania Pilliza Chicaiza
CI: 1723293567

# Autorización de publicación

Yo, **GISSELA ESTEFANIA PILLIZA CHICAIZA**, con cédula de identidad 1723293567, cedo a la Universidad de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Urcuquí, Agosto 2020.

Firmado electrónicamente por:
**GISSELA ESTEFANIA PILLIZA**

Gissela Estefania Pilliza Chicaiza
CI: 1723293567

# Dedication

*"To my siblings Erika, Fernando, and Domenica who have been the best life partners, this work is proof that you could achieve whatever you purpose in your life. To my loved aunt Sandy which even in the darkest moments gave me her unconditional hand to continue my studies, rest assured that we will continue to achieve our goals. To my life partner Ricardo who struggled by my side unconditionally this challenging road, you almost get here too, keep fighting, I am here for you. To my parents Silvia, Roberto, and to my whole family who has taught me the most valuable life experiences, that brought me till this point. To my friends Nico, Emilia, Javi, Jair, Ari, Juan M., Cris, Erick, Mafer, Emil for the support during this whole road without them it could have been harder thanks for being my family during my time in YT. And to my life friends Carlos, Kira, and Jessi who never have lost the connection and open their home doors to me any time."*

# Acknowledgements

*"In the first place, I want to say thank to my patient and indefatigable thesis advisor and mentor Lorena de los Angeles Guachi PhD. who has heedful in every step of my work and always had been pushing me up to achieve bigger things. Thanks, also to my professor Francisco Ortega Zamorano who provided the necessary AI skills and tools in a prelude of this project. To my partners, Osiris, Winter, and Sergio who played an important role in the initial idea of this project. To Gonzalo Aldana his overseas collaboration, and interest in the project have been constructive. To Elisabeth G. and A. G. for support without interest the goals of any good student at Yachay Tech. To all my Yachay Tech professors without exception, you built every piece of the works of this whole generation, all the achievements and future awards are the result of your exceptional work. And, to all the people who were delighted with the topic somehow, it was a motivation to still working on it."*

# Resumen

El objetivo del presente estudio comparativo fue determinar que algoritmo del tipo mapas auto organizados (SOM) se adapta mejor al agrupar empresas del índice financiero SPLatinAmerica40. Este trabajo buscó un punto de convergencia entre la inteligencia artificial (AI) y la economía, ya que a pesar del aumento del uso de AI, en finanzas no ha aplicado sino en los últimos años. Para cualquier nación, los mercados bursátiles representan un factor potencial para su crecimiento económico, ya que son un motor financiero que genera ingresos a partir del dinero producido por la fuerza industrial de los países. Para comenzar la investigación está enfocada en establecer la mejor arquitectura SOM para el tratamiento del mercado de valores a través de la revisión literaria. Después se procedió a la extracción de datos históricos del índice financiero seleccionado del año 2014 al 2019, usando la plataforma de Yahoo Finance. Luego se realizó un pre-procesamiento de datos mediante un algoritmo de cohesión. La ISOMSP40 es el nombre del método SOM propuesto en este trabajo y utiliza una combinación adecuada de arquitectura hexagonal y función de vecindad basada en la distancia de Manhattan. En adición otros dos métodos, denominados SOM IBEX35 y SOM NYSE, se probaron bajo las mismas condiciones comparando sus métricas y determinando el mejor algoritmo para el conjunto de datos SP Latin America 40. El estudio usó como referencia a las 9 compañías con mayores ganancias en el índice bursátil. El riesgo de inversión se analizó principalmente con las métricas de correlación de densidad de ganancias, área industrial y geográfica detectadas con una tasa de éxito del 80%. También se verificó la correcta agrupación en un análisis de frecuencia de tiempo desarrollado con las 6 principales compañías durante el mismo período de la data. El tiempo de ejecución en el método ISOMSP40 también mejoró en aproximadamente dos cifras significativas teniendo $5,79E01(s)$ como tiempo de ejecución mínimo en contraste con los otros dos modelos. Es así que se logró establecer que el algoritmo ISOMSP40 propuesto muestra un mejor rendimiento para el mercado bursátil al que se apuntó y que los experimentos comparativos de las diferentes métricas demostraron una eficiente adaptación para el índice SPLatinAmerica40 alcanzando así el objetivo principal.

**Palabras clave**: Mapas Auto Organizados, Bolsas de valores, Índices financieros, S&P Latin America 40, IBEX35, NYSE, NASDAQ, riesgo, inversión.

# Abstract

The objective of this comparative study was to determine which algorithm of the self-organizing maps (SOM) type is best suited when grouping companies from the SPLatinAmerica40 financial index. This work sought a point of convergence between artificial intelligence (AI) and economy because AI has only been applied in finance in recent years, despite the AI exponential use increase. For any nation, stock markets represent a potential factor for its economic growth as they are financial engines, which generates income from the money produced by the industrial strength of the countries. First, the investigation was focused on establishing the best SOM architecture for the treatment of stock markets through the literary review. Afterward, the historical data of the selected financial index from 2014 to 2019 were extracted, using the Yahoo Finance platform. Then, pre-processing of data was carried out using a cohesion algorithm. The name of the SOM method proposed in this work is ISOMSP40, and uses a suitable combination of hexagonal architecture and neighborhood function based on Manhattan distance. Two other similar methods were tested under the same conditions to compare their metrics. These measures determined the best algorithm for the SP Latin America 40 data set. The study used as reference the nine companies with the highest profits in the S&PLATAM 40 stock index. There were mainly analyzed the metrics of density by profit, industrial area, and geographic correlation detected with a success rate of 80%. The correct clustering was also verified in a time-frequency analysis developed with the top six companies during the same data period.The execution time in the proposed ISOMSP40 algorithm also improved by two decimal places. The minimum execution time was $5,79E-01(s)$ against the $9,01E+00$ average in the other two models. Thus, it was established that the proposed ISOMSP40 algorithm showed a better performance for the S&P LATAM 40 stock index over two other existing methods.The comparative experiments demonstrated by the metrics an efficient adaptation for the chosen index achieving the main objective of this study.

**Keywords**: Self Organized Maps, Stock Market, Stock Index, S&P Latin America 40, IBEX35, NYSE, NASDAQ, investment risk.

# Contents

# List of Figures

# List of Tables

# Introduction

Beyond doubt, risk analysis for stock exchange markets investment in any region, means a relevant issue when vast amounts of money are circulating and producing around the world. They are directly affected by economic, social, and even political events [4]. The stock exchange market or bursal field is a financial mechanism that allows to the brokers and trades the successful negotiation of different financial instruments such as: bonds, titles, stocks, among others. The risk analysis for this purpose is referred as the process where the incidence of positive and negative episodes on the transactional movements of capitals are evaluated. This analysis avoids significant losses and allows performing the purchases at the right time for the company [5]. Most of those analyses have been treated traditionally by statistical approaches [6].

Meanwhile, artificial intelligence (AI) involves the creation of intelligent agents starting from the training and learning with large quantities of data generating systems of: prediction, clustering, among others. The artificial neural networks (ANN) and deep learning are branches of this field, and use techniques that allow the approximation of nonlinear systems. They are constantly adapting almost any model efficiently, and in some cases, with better accuracy [6]. During the last two decades in business and finances, just a few studies have applied the AI tools in comparison with other areas such as medicine. The AI studies for medicine according to Scopus, presented an 8:1 relation against the AI other topics studies in 2018. Therefore, the financial AI studies have mainly obtained results in classification tasks [7], stock price forecasting [8], financial distress, bankruptcy [9], and many different applications [6].

The understanding of how different events can affect the stock market tendencies is explored in a work developed in 2018 by Hongping Hu at all [4]. They used the back-propagation ANN technique (BPNN), which is a supervised model that study the prediction of the direction of stock markets. This model was improved in terms of winner function, which was implemented with a sine-cosine algorithm (ISCA), so both joined to form a new structure called ISCA-BPNN [4]. The metric used to test the results of the mentioned work is the radio percentage (%), presenting around $5\% - 10\%$ of improvement over the methods compared without the ISCA-BPNN. An advantage of this implementation was that it didn't imply complex mathematical optimization analysis, but in contrast, it exclusively works with the labeled linear version of the program. At the same time, it struggles with the limits of the descendant gradient present in back-propagation methods by nature.

In Indonesia, Jakarta Composite Index Price is one of the essential financial indicators; for this reason, it was studied in 2018 by F. Fanita and Z. Rustam [10]. The study is focused on predicting the prices of this index using an "Adaptive Neuro-Fuzzy Inference System" (ANIF) in association with Fuzzy Kernel C-Means . This technique belongs to the supervised side of Fuzzy Logic computational intelligent algorithms. The experiments owned target values that allow them to obtain the relative error, which was the determinant to calculate the prediction accuracy metric [10]. The meaningful improvement of this technique was the reduction of the relative error percentage bellow $1\%, 2\%,$, and $3\%$. This accuracy works for the different portions of the training data set on the classification of prediction. Nevertheless, when a large amount of capital is what matters that $1\%, 2\%$, and $3\%$ can mean millions of losses if the prediction gives a bad accurate.

For several social-economical factors, the models developed are not into our regional scope. The underdeveloped countries are economies that have a lot of potential in many industrial sectors, but the bursal sector owns a lack of scientific resources to make investments reliable. Thus, it is necessary to realize studies adapted to the region reality that even there not cover the whole problems, it started to trace the roads to follow. Latin American stock markets and indexes were a potential field almost unexplored. The present work aims the risk analysis by the comparison and implementation of an improved SOM ANN method of the S&P Latin America 40 index denominated improved Self-Organizing Map (ISOMSP40). The comparison was performed with two other existing methods, the SOM IBEX35 and the SOM NYSE. The ISOMSP40 method is going to identify relationships between enterprises in stock exchanges. In this sense, the best parameters and techniques for the adaptability to Latin America stock markets were tested and identified.

In this project, SP Latin America 40 exchange index data are going to be studied using three Self Organized Maps techniques: SOM IBEX35 [2], SOM NYSE [3], and the actual developed improved ISOM SP40. The two previous techniques were selected because of their acceptable results in their correspondent studied stock exchanges. According to the works referred above, for being compared among them and with the improved method. The metrics detected as useful to be compared were the performance time and the density correlations in profit, industrial sector and geography. The data set was extracted from Yahoo Finance [11] open source as the database owns reliable and wide historical information free of charges. The particular feature of this method is the neighbor function used in the clustering. The experimental setup was designed for varying their hyper-parameters and most important the topological distance used in the neighbor function. The experiments have modeled as similar conditions as possible, with five times of repetitions for each different number of maximum iterations 10,50,200, and 500. All of those repetitions measured the accuracy in terms of density correlations and time execution. Firstly, it was evaluated the proposed ISOM SP40 with SP 40 LATAM data set finding that the less number of iterations achieve a better performace. The performance was measured in terms of distribution for the density correlations and time execution reduced in comparison with the other two methods. The second experiment used the SP 40 LATAM data set

to feed the IBEX35 algorithm, but could not find too much accuracy. The clustering of this second experiment had an average time not closer neither to the worst-case nor the best-case. The third experiment processed the NYSE algorithm with the same data set, but it works only for 200 iterations repetitions. The iterations limitations are because of the Matlab tool not allow the modification for a further analysis. This last experiment found that less than 5% percent have a good clustering. The last experiment evidence the correctness of the results in the previous experiments by a time-frequency analysis of some companies sets. This frequency analysis reveals concordance with the results obtained in the first experiment.

The results of this project demonstrate the better performance of the ISOMSP40 algorithm in comparison with the other two studied methods. A good clustering of the SPLatinAmerica40 enterprises allows the analysis of the risk investment in the target. Present an efficient treatment of stock markets data sets, and develop an adapted ISOM SP40 method for the Latin American region reality is the main contribution of this research.

# Chapter 1

# Preliminaries

## 1.1 Problem statement

Although the importance of the stock exchange in the economic development of the nations, artificial intelligence is not necessarily working in economics models. The Latin America region itself has been growing and boosting its markets, and now owns a large database able to feed an ANN or AI models. Unfortunately, it is not the focus of many studies mainly because the region has a reduced budget for research. Besides, the few jobs realized on SOM ANN for world stock markets are unable to fit with stock indexes from Latin American indexes as is evidenced in [2]. Besides, those previous works just use the biggest markets in the world such as NYSE and NASDAQ, thus, leaving apart the idea to rapidly adapt to other regions. For those reasons, this research proposal is focused on introducing an improved SOM clustering model to discover relationships between enterprises for SP Latin America 40.

## 1.2 Justification

The motivation of this work lies in the limited application of artificial intelligence in the economic field since its inception. Recently, SOM techniques have been used in financial applications as financial-market prediction [3], and enterprise clustering [2]. Despite they're demonstrated outperforms, it is challenging to establish a suitable method to obtain a good clustering in Latin America regional indexes. Thus, a new SOM model is introduced for its application to the SP 40 Latin American index. The new approach is evaluated concerning two other SOM models IBEX35, and NYSE, to determine their performance in terms of density correlations of profit, industrial area, geographic, and training time. This work prompts to develop more studies on SOM research, financial and commercial fields as the functions and parameters used are very malleable, and in the economic topic, there is a lot of work to do.

## 1.3 Contribution

The work aims the adaptation of a SOM model to the SP40 LATAM data set through a change in the neighbor function using the Manhattan distance. The distribution in the SOM maps and the time execution was compared to establish that in fact the proposed model is working for the SP40 LATAM data. The model took less execution time in comparison with [2], and [3]. Another contribution is the prepossessing of the data obtained for Yahoo finances. The process of clean the data uses an algorithm to concatenate any amount of files and fix the components to analyze before to enter in any model.

## 1.4 Document organization

This work organizes the whole information in seven chapters labeled as Preliminaries, Objectives, Theoretical Framework, Methodology, Experimental Setup, Results, and Conclusions and Future Work. Each of these chapters contains the following information:

Chapter 1 developed the problem statement, the justification with the problem explained, significant contributions that this study provides.

Chapter 2 points out the general and specifics purposes of this work.

Chapter 3 contains a detailed review of the concepts necessary to follow the work line and a presentation of the relevant architectures to being compared.

Chapter 4 details the methodology followed in the work including the data collection, data prepossessing, and the design of the proposed ISOM SP40.

Chapter 5 depicts the metrics used for the comparison of the methods, the data preparation, and a description of the experiments to develop in the next chapter.

Chapter 6, show the data tables and graphs obtained from each experiment with an analysis focus in the comparison of the three methods.

Chapter 7 includes all the deductions built from the results obtained, and the possible improvements for future works.

# Chapter 2

# Objectives

## 2.1 General Objective

To design and implement an algorithm capable of determining relationships by investment profit, industrial area, and frequency-time evolution among enterprises minimizing investment risk in the SP40 Latin America stock index using different neighbor functions in SOM techniques, and comparing it with other two existing models.

## 2.2 Specific Objectives

The next specific objectives will be followed to achieve the primary goal.

- To collect fixed data able of the enterprises in the SP40 Latin America stock index, to fit with it all the experimental setup needed.
- To perform prepossessing operations to start with cleaner data able to fit in many models.
- To determine the appropriate architecture among the reviewed methods for SOM networks suitable for cluster unlabeled data, in this case, the profit analysis of each company.
- To design and implement an improved SOM network model taking into account remarkable characteristics from explored methods.
- To compare the performance of the proposed approach finding enterprise relationships in terms of topological distance, density correlations, and execution time.

# Chapter 3

# Theoretical Framework

Within the AI increasing area, several sub-branches have been born, being applied almost in any science and industry field [6]. Thus, the problematic modeling of the fluctuating behavior of the markets makes necessary the use of complex and integral forecasting and prediction tools as artificial intelligence tools. This section explains the basics about the work main issues, and analyze some of the existing clustering AI methods. Specifically the AI methods applied in the finances field to find the best features. This section also performs the final architecture selection for the implementation and comparison of this study. Despite there are several methods implemented in this area before, just the techniques which have shown results kindred to the project objectives were reviewed.

## 3.1 Definitions

In order to get a better understanding of the study, it is essential to review some key concepts of AI to let the reader lay in context. Aso the reader can get the most approximated idea of what each thing is for work purposes. However, these fundamental concepts just introduce the main idea of the big topics involved in this research. Thus it does not imply a profound explanation as artificial intelligence is in constant growth.

### 3.1.1 Artificial Intelligence

Despite the several definitions and misconceptions of what Artificial intelligence is, the accurate idea comes from the performing of human tasks, biological brain processes. It simulates the human intelligence system in the computational world [12]. Humanity has not ended to study and understand the brain in biology at this moment. With the knowledge and current technology available, the computer sciences are not ready to completely reproduce, not even a part of the human brain.
The intelligence is defined as the ability to understanding and learning things [13]. Hence, what is real about AI is that in most cases, AI systems are capable of understanding data sets, learn from it, make decisions, and accomplish determined tasks. The increasing wide spectrum of AI has made it difficult for the academy to establish a concise classification for its branches according to the literature. Several of these approaches converge in many

points, but at the same time have the discordance enough to do not being in the same line. To make a capsule idea of the different sources reviewed in this work, the artificial intelligence can be assimilated as is described in Figure.3.1.

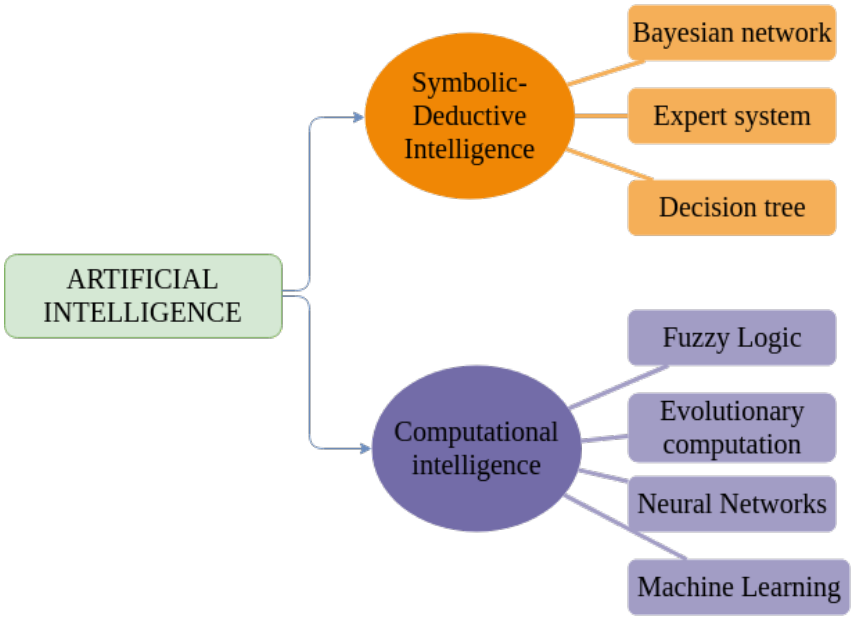


Figure 3.1: Artificial Intelligence Approaches Classification (AI)

For this work purpose, there is pointed out specifically to the Neural Networks approach, which belongs to the computational intelligence bigger branch and is going to be deeply described in the following section. Besides, another absolute fact is that machine features overcome the capabilities of humans in terms of storage, memory, and processing speed; the last characteristic implies that many current human tasks would be candidates to be replaced by AI automation [14]. The dynamic role that the AI plays nowadays is applied in almost any computational and industrial field, or at least it could be applied.

### 3.1.2 Artificial Neural Networks (ANN)

Inside the AI family tree, the Neural Networks approach is focused more than others in developing models that follow the principles of the human cognitive processes [12]. As in the human brain, the primary processing unit is the neuron or soma that in an initial stage owns dendrites that can be compared with the input data ($x$). The synapses interactions are replicated such as weights ($w$) which are generated with operations and instructions to interconnect the layers. Then, they form a complete neural network that result an output ($y$) that is equivalent to the biological axon. The significant highlight of these intelligent systems is that, as in biology, these structures respond to signals and stimulation patterns, which finally allows obtaining significant deductions [13]. The flowsheet of any ANN begins according to the selected data set as it defines the architecture selection as to how it is capable of feeding the initial layer. Each data signal corresponds

to a neuron in the initial layer, so in this layer, there are going to be as many neurons as rows of data set. Then, the first connections were performed towards the second layer being combined with the weights, learning rate, and other hyper-parameters implied in the ANN performance. Later, the structure of each network defines the road that the neurons follow. Also, how they achieve the expected output in the final layer as we can see in Figure.3.2.

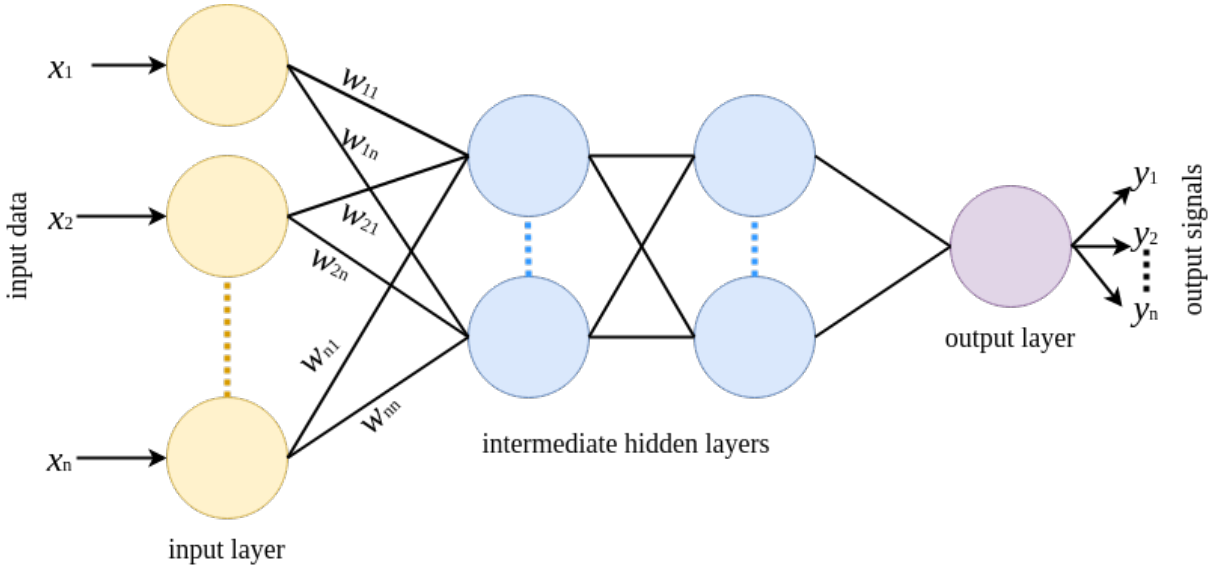


Figure 3.2: Artificial Neural Network Classic Architecture

The nature of the different data sets and the types of interconnections define the division of ANN learning paradigms as follows:

- **Supervised:** This learning paradigm, also known as active learning, is based in a conducted training by its labeled data set. These methods pursue the minimization of the difference (mean square error) enclosed by the output signals and the target signals, which work as the supervisor [13].

- **Unsupervised:** This approach is founded in the understanding of the behavior of its data set. Whatever the data set nature is, there are not specific expected results, references, or targets to follow. Mostly of the data sets which own fluctuating patterns over several variables opt for this type of learning because of its adaptability. It can detect intricate patterns and place them in relevant categories [13].

- **Reinforcement:** The animal-learning is the main inspiration of this paradigm as it is focused on the capability of resolve practical situations [15]. This ability is achieved under the condition to maximize the expected value of a determined function. In other words, is analogous to receive rewards each time that the desired approximation is obtained [16].

### 3.1.3 Self Organized Maps

Self Organized Maps technique belongs to the unsupervised methods inside the ANN class. In 1981 the professor Teuvo Kohonen[1] proposed this model motivated with the idea of "abstract feature maps found in the biological central nervous systems". These organizations of selective responses to the stimulus are denominated brain maps, which is the idea that allows the self-organized map technique to discover patterns in multidimensional data analysis. This model is presented as an alternative for data sets that can be linearly unmodeled. Also, the calibration as input works even there is a huge amount of data or not.

The different versions of the SOM use the concepts of similarity and distance to orient their output. In this work, the distance means one of the essential features for analysis [1]. The distance is measured topologically by a chosen function, a widely used distance in SOM models is the Euclidean distance. The map geometry where the network is distributed is another parameter that can be fixed to achieve the best performance of the model. Commonly it starts as a triangle or a square, but it can be modeled even as a hexagon.

The base structure of a SOM comes from the concept illustrated in Figure.3.2. In this structure the interconnection from layer to layer is defined by a winner model. Significant patterns are detected and begin to be located on the map according to the geometry established and distance found by the network among each node. Therefore, the number and the distribution of the outputs depend on the number of relevant clusters catch by the system ,as is described in Figure.3.3.

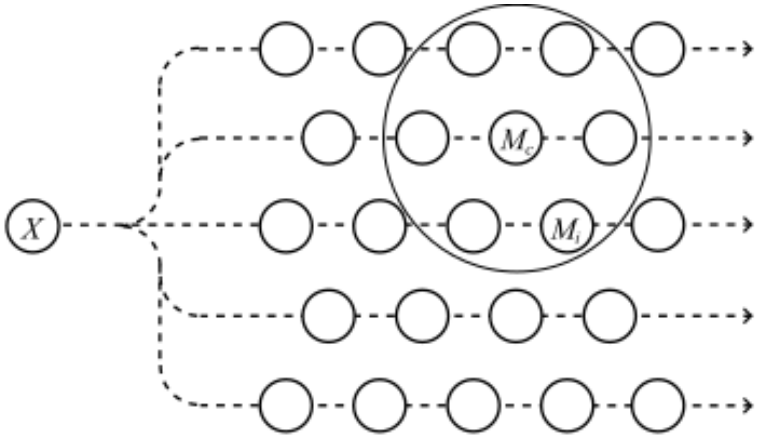


Figure 3.3: Illustration of a self-organizing map. An input data item $X$ is broadcast to a set of models $Mi$, of which $Mc$ matches best with $X$ . All models that lie in the neighborhood (larger circle) of $Mc$ in the grid match better with $X$ than with the rest. *Source:[1]*

## 3.2 Relevant Architectures

According to the literature review, there are many AI tools that can be used in finances [6]. Nevertheless, the selection criteria in this work pursue the main goal of the study described in chapter 2. The first criterion was the orientation of the project, then as it was desired to find relationships in the behavior of the enterprises, a clustering approach was needed. Secondly, it was important to understand the anatomy of the data set available to train the network and how it interacts with the purpose. In this way, for the extensive unlabeled data set in hand, an unsupervised method was required. Under these two first conditions, the SOM maps technique fitted ideally. Afterward, there were not too many works developed in the Stock Markets area with AI unsupervised methods, and specifically with SOM technique, just a few [17][3]. This section describes the two relevant related works for the comparison part, this time selected by the closest hexagonal geometry architecture criteria and similarly among their data sets.

### 3.2.1 SOM IBEX35 - LATIBEX

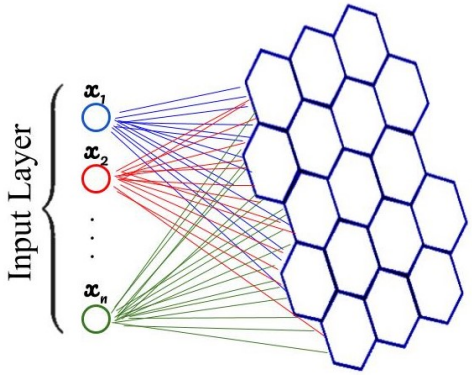


Figure 3.4: SOM Architecture 2D hexagonal IBEX35. *Source:[2]*

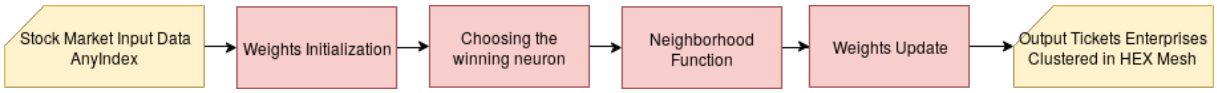


Figure 3.5: SOM IBEX35 Flow Chart

The motivation of this work was the application of the potential of the SOM method to perform an analysis of investment risk. The work published in 2018 [2] used SOM hexagonal network to cluster the companies in the IBEX-35 stock index. The architecture described is distributed in a 2-dimensional network with all its neurons directly connected only with its neighboring neuron. In parallel all of them are connected in different ways, containing a number of $n \times m$.

The input data IBEX-35 were selected for the study as it is part of a robust financial market with a high transactional movement "The Madrid Stock Exchange". This index

owns more than 20 companies with enough historical information to be studied. The time period of the extraction was a whole year from April 2017 to April 2018, and the data source was the Yahoo Finance website[11].

An hexagonal network was implemented making that each neurons connects to six other neurons, such as is illustrated in Figure3.4. The activation of neurons originates when a neuron weight vector is the closest to the input as each input neuron has a related weight vector. The topological distance is determined by the euclidean function and points to the closeness that a neighborhood function is setting. These features show that if the neuron is activated, then the neuron is learning[2]. Besides, another data set associated with Latin America was analysed, but it does not showed the same accuracy in profits fluctuation by the time. The last fact explanation was that the index does not have the high transactional movement needed to feed the network. In summary, this approach was following the process the depicted in Figure 3.5, demonstrating an efficient clustering of the companies. Nevertheless, it was not adaptable to indexes with fewer transactions.

### 3.2.2 SOM NYSE - NASDAQ

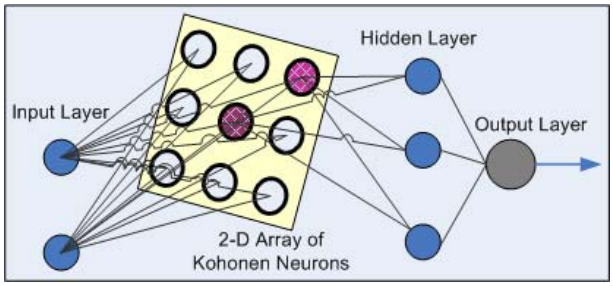


Figure 3.6: SOM Architecture 2D hexagonal NYSE-NASDAQ *Source:[3]*

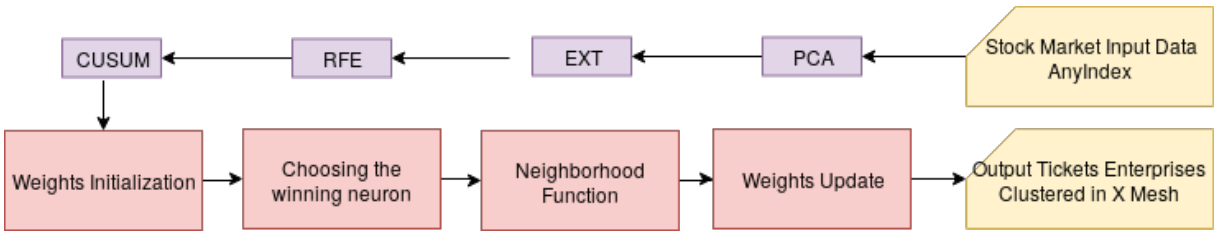


Figure 3.7: SOM NYSE-NASDAQ Flow Chart

In this method proposed in 2015[3], its SOM principles followed the scheme of the SOM described in Figure 3.5. The main difference here was on its previous data prepocessing steps and different geometries applied to the experimental part. They choose the NYSE and the NASDAQ stock indexes as they host the biggest and influential companies in the world. Particularly there were selected only companies with at least nine years of available data, and those data had to be adequate statistical. In this manner, about 6500 companies in a time of 10 years were included in the tests.

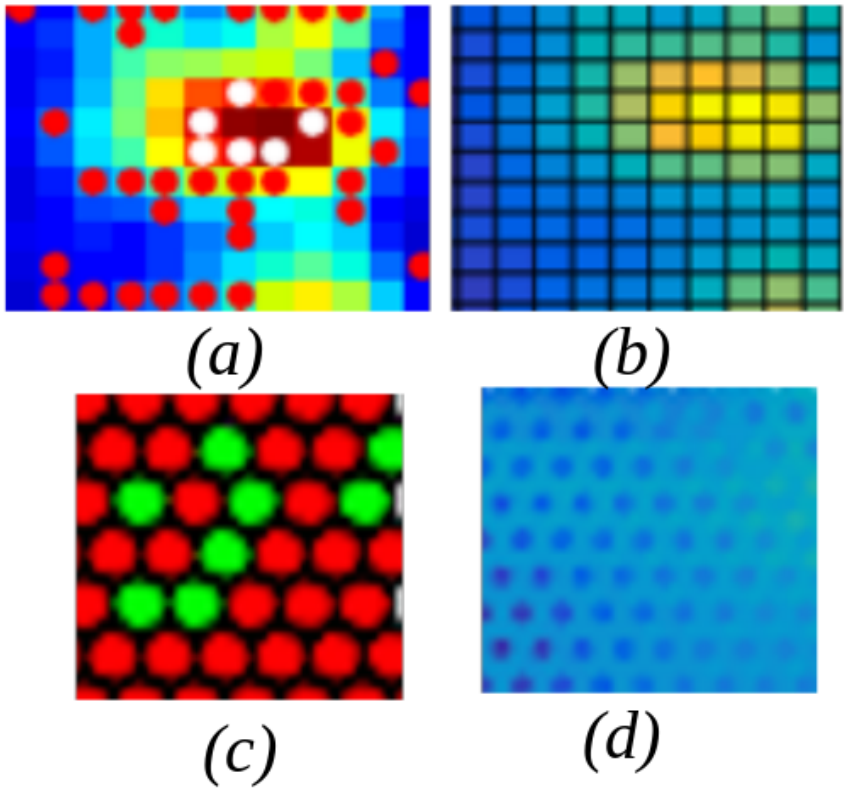


Figure 3.8: *(a)*The cluster after the FWC transformation. Notice the closeness of the top ten investments. *(b)* The raw FWC, showing the single cluster more clearly. *(c)* The purely majority voting section (of the same cluster). These green points are strongly voted as "good investments". *(d)* The UMAT plot. Notice that homogeneous light blue consistency, suggesting close nodes.*Source:[3]*

Illustrated in Figure 3.7, it first performs a Principal Component Analysis (PCA). In this analysis they made choices without the necessity of labels trying to reduce the components because at the starting point, the companies owned a list of 125 features. Secondly, it uses the Recursive Feature Extraction (RFE) that picks and replace characteristics making different combination until finding the best possible combination. Then, Extra Trees Classifiers (EXT) technique is applied as a variation of the random forest features selection. Just before entering in the SOM, the data is identified by the Cumulative Sum (CUSUM) method that is a labeling technique of data for enterprises precisely described in the source.

Finally, the pre-processed data are entered into the SOM network finding successfully good approximations for the top ten enterprises of investment. These findings are represented by color patterns and density distribution ruled by the euclidean distance. This time the vast selected data set make ably the adaptation of multiple prepossessing techniques that make more manageable the use of AI itself. The results of this study just talked about the profit probability but did not resolve to sell tendencies.

# Chapter 4

# Methodology

The established methodology for designing a SOM approach for S&P Latin America 40 index is depicted in this chapter. It starts with a preliminary exploration mainly to identify the most relevant SOM models and their distinctive characteristic that influence to performance results in financial fields. The next stages are named as Data Collection, Data Pre-processing, Design ISOM Approach, Train Models, and Obtained Results.

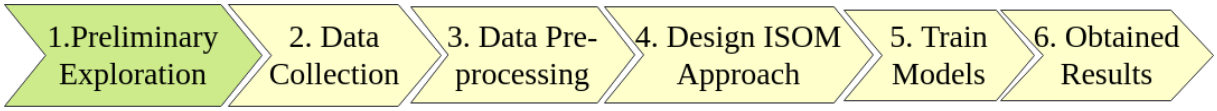


Figure 4.1: Methodology Flow Chart

## 4.1 Data collection

### 4.1.1 Data Source

The data generated for the transactional movement of the companies and capitals are somehow classified because of the market dynamics; thus usually obtaining determined historical information from stock markets official pages, the users must pay considerable quantities of money. Fortunately, since 1997 Yahoo finances have collected integral information of world stock markets being considered nowadays as the number one online platform for stock market values.

Yahoo finances site offers to the user the stock markets and indexes historical information of at least 20 years ago of bursal market completely open adjusted to its self policies of data usage and reproduction. Yahoo Finances website obtains its information from Global Market Intelligence, Commodity Systems, Inc. (CSI), among other trustworthy companies, thus making the data collected quite reliable. Nowadays, almost more than 50% of Yahoo Finance data providers are delivering the information in real-time. Just a few are not in this percentage but they are updated in a maximum time of 25 minutes [11]. For these reasons, several studies have used the Yahoo information as is referred to in the previous chapter. This work is not the exception, this project also uses this source.

The steps to obtain each company's historical information started with the identification of the unique ticket in the corresponding stock market website. Then, once the ticket was entered into the Yahoo Finances browser, there is unfold all the information in tabs, there had to be selected the "Historical Data" tab. There the final time parameters have to be set to deliver the companies data daily, weekly, and monthly.

### 4.1.2 SP 40 Latin America data set

In the stock markets field, there is important to note that a stock market is not the same than a stock market index. While the first one is a group of companies listed in the same stock market . The second measures the performance of a group of enterprises that could be in different stock markets. From the series of S&P Dow Jones well-known indexes, there is the S&P Latin America 40, which concentrates the 70% of the capitals in Latin America. As the work main principle is the adaptability to the region, then the daily data from 5 years (April of 2014 to April 2019) were extracted to be analyzed. All of them downloaded in .csv format compatible with readers or processors of spread-sheets.

## 4.2 Data pre-possessing

From the Yahoo Finances platform, the historical data collected were "dirty", containing columns of data that are no going to be used independently in the study; for this reason, the data need to be sieved.

- Each file collected contains six indicators of stock market daily prices, one more useful than others to diversify the input parameters for the network. Then the parameter that would be selected for this study are:

    1. **Stock Price**: In this column, there are the actual values of the prices for a single stock indicating the amount of currency of the stock that costs that company piece [2].
    2. **Open Price**: The values in this column are equal to the last maximum that a price can achieve in a day. There it is used in the variance calculation as the starting point for the day before. [2].
    3. **Close Price**: The values in this column are equal to the last minimum that a price can achieve in a day. There it is used in the variance calculation as the ending point for the current day. [2].
    4. **Price Variance**: The calculation of this value derived a new column that is equal to the difference between the open and close values. This calculation point the daily variance for the company and allows to determine episodes of profit o losses in a period. [2].

- Then, there was made an automatic concatenation of the files mentioned before in section 4.1.1; in other words from the .csv documents downloaded from the Yahoo

finances were be concatenated in a single folder by each data set group. Then, an algorithm implemented in python is used to have the data compiled in a single document.

- As a result, an only new .cvs document was generated representing the average percentages of changes between open and close values in each frequency called Variation Percentage for all the companies.

## 4.3 Proposed ISOM Approach

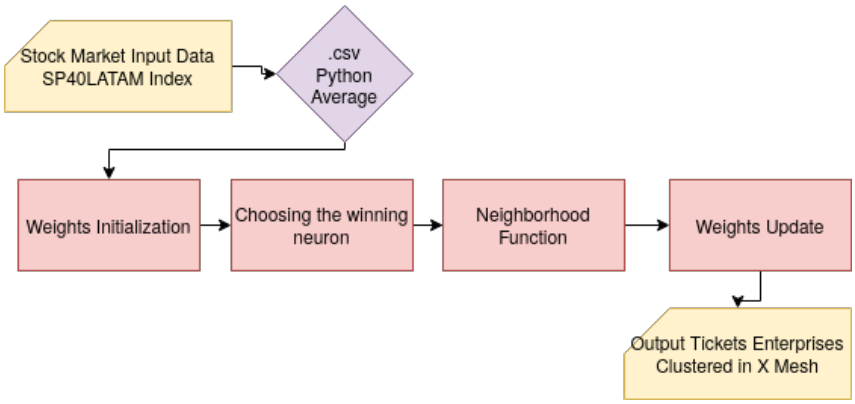


Figure 4.2: Improved SOM Flow Chart

For the model to be implemented, it is expected that follows the hexagonal structure of the IBEX35 approach developed in previous risk analysis paper [2]. But, the hyper-parameters are fixed for a different test, and in the neighbor mechanism it is modified whit the Manhattan distance. Then, it would use the equations 4.1,4.2, 4.3, and 4.4. In the weights update step according to the mathematical formula improvement, it could be applied building a topological structure. This structure can represent the original surface of the system adapting the mesh organization [18]. Meanwhile, there is the automation in the initialization scheme of data the data set through the "phyton prepossessing algorithm" to be entered after in Matlab. This program is going to be the IDE used in the whole project because the dynamic handling of the modules to build the project. The general flowchart of the proposed approach is depicted in Figure 4.2.

### Proposed SOM Pseudocode

*Weights Initialization*:

$$w(0) = random([a, b], [n, m, o]) \tag{4.1}$$

*Choosing the winning neuron*:

$$G(x(p)) = argmin_{\forall i}\{||x(p) - w_i(p)||\}\forall i = 1, 2, 3, ..., n \times m \tag{4.2}$$

*Neighborhood Function*:

$$\Lambda(P_j, P_k) = Sum(Abs((P_j - P_k))) \tag{4.3}$$

*Weights Update*:

$$w_i(p+1) = w_i(p) + \eta(p)\Lambda(P_r, P_i)(x(p) - w_i(p)) \tag{4.4}$$

**Main Differences regarding to IBEX35**

The model proposed where written in base to the code of the IBEX35, which at the experimental moment seems to be the best SOM choice. Nevertheless, after the parameter calibration and much more after the study of improvements of SOM metrics, the critical improvement is in the Neighborhood Function.

- + The hyper-parameters involved in SOM the number of working layers and learning rate were the correct since the beginning, so they stayed fixed in the experiments, not duplicating the work .
- + Then the max number of iterations was found as an important role in the clustering distribution; for this reason, this was set as dynamic for the tests. This hyperparameter allows a better visualization of the time metrics and repeated verification of the results.
- + The euclidean distance was the protagonist for the most of the SOM methods, and especially in both studies selected for comparison. Thus, it was changed by the Manhattan Distance, which, according to the literature, showed features such as time reduction, operations simplicity, and almost the same accuracy.
- **-** That almost the same accuracy represents a risk in the correctness of the results.
- **-** The method is accurately aligned to the SOMIBEX35 routine, but for the NYSE some parameters can not be modified. Thus, there were fewer samples to be compared with the NYSE method.

### 4.3.1 Training Models and Resources

- The data was prepossessed in the same way described in the previous section; for the three cases, they were introduced equally in the three algorithms IBEX35, NYSE-NASDAQ, and SPLatam40 Figure. 5.1 and are described deeper in Experimental Setup Section 5.1.
- Each algorithm uses its own routine for training and cluster as the schemes explained before in the methodology section.
- The whole process was run in Matlab IDE, including the prepossessing and training and with an available in a Laptop Dell core i5 from 4th generation and Ubuntu operative system.

# Chapter 5

# Experimental Setup

In this Chapter, the experimental configuration is described giving parameter details and model configuration used in each experiment.

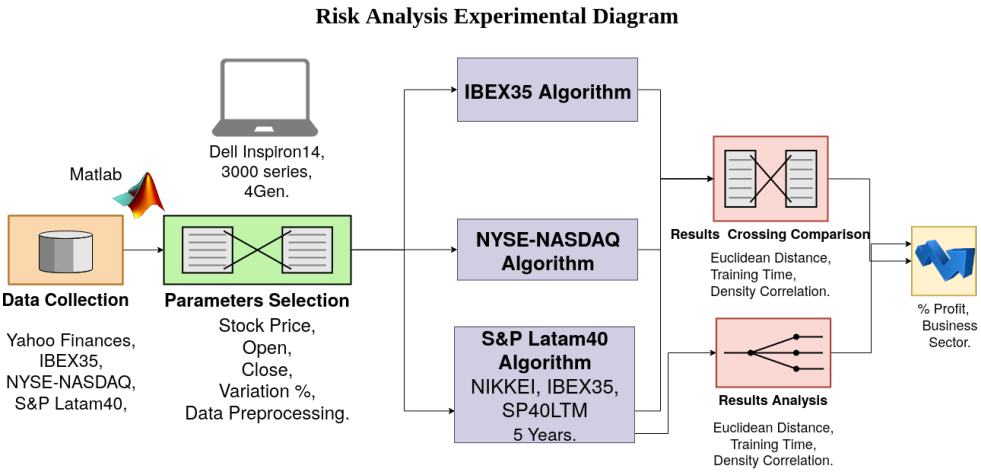


Figure 5.1: Experimental Setup Scheme

## 5.1 Performance metrics

For the present work, the metrics that allow measure the results in a standard manner to compare the three methods refereed before are: Topological Distance, Training time, and Density Correlations.

### 5.1.1 Topological Distance

The topological distance is the metric that measure the distance among the nodes of data of each company. It can impact significantly in the accuracy expected in the density correlations described below.

### Euclidean Distance

The SOM algorithms major clustering tools are based on distances, the one of the most wide used distances is the Euclidean distance and it could be applied to the three selected models. It is a quantitative metric and in this project was calculated as follow [2]:

$$d(P_j, P_k) = \sqrt{(P_{j1} - P_{k1})^2 + (P_{j2} - P_{k2})^2} \tag{5.1}$$

$\forall k = 1, 2, 3, ..., M$.
,where:
$P_j$ is the treated company,
$P_k$ is the k-st company to get the distance, and
$M$ is the total number of companies.

### Density Correlations

The density correlations are qualitative metrics which are identified as areas of the graphics generated by the algorithms in each case, as is shown in the enterprises relationship graph Figure. 5.2 .Therefore, codes could be adapted to analyze the correlation areas by:

- General Company Tickets: distribution of the companies with more profit in the SOM structure.
- Geographical Distribution: distribution into the SOM of companies which belong to the same country.
- Industrial Areas: distribution of companies associated in same business line such as banking, energy, food, chemicals, oil, and many others.

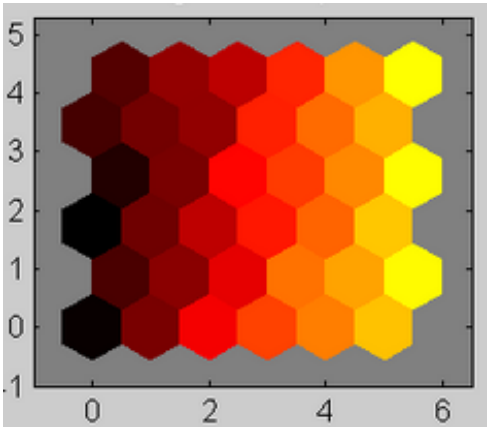


Figure 5.2: SOM Density Correlation Illustration

### 5.1.2 Training Time

This quantitative metric is referred as the time that the Matlab program is going to take to measure the model construction with the data set. The execution time program was quantified by the tic-toc function integrated into Matlab that, according to its documentation, it measures the time until the program completes the last operation.

- From the three algorithms, the same result features were extracted under the metrics founded before.
- After the three algorithms were run, the Topological distance for density correlations, Training Time was crossed to detect the differences between the three methods.
- At the same time, the metrics were be widely analyzed to determine if the method is efficient in the adaptation to the SP40 LATAM data set.

## 5.2 Data Preparation

After the download process from Yahoo Finance site the data was processed and concatenated by an Phyton algorithm following the next steps:

Phyton Preprocesing Pseudocode

```
months=include all the year months;
files=(load all the Yahoo files);
for files in file do :
Principal Components Analysis and Selection

for files in file do:
Join to the main file

for file do:
Main file name and attributes assignation
```

## 5.3 Experiments ISOM SP40

In this section the relevant proofs realized in order to verify the efficiency of the algorithms is explained and conceptualized in order to get clear the

- **Experiment 1:SP40LATAM data set in ISOM SP40 Algorithm**

  In this experiment, the data-set described in section **??** was entered into the proposed method with Manhattan Distance as neighborhood function. Following the

steps represented in Figure. 4.2. The experiment consisted of 5 times the verification for each maximum number of iterations. The number of iterations was extracted for the experiments developed in [2] and [3]. In this way, the maximum number of iterations of all the experiments were 500 as in [2]. Then followed by the 200 of the [3] set as default in MATLAB. Later by the proportions, a 50 times iteration included. And ten iterations were tested because it is the minimum number.

- **Experiment 2: SP40LATAM data set in IBEX35 Algorithm**

  This experiment followed the same scheme as in the previous one, but with the IBEX35 algorithm keeps the original Euclidean distance as in the 2018 study[2]. Also, it varied in the presentation of the graphic results because it was adapted to the proposed model. Always with the goal of obtaining similar images of both experiments for comparison.

- **Experiment 3: SP40LATAM data set in NYSE Algorithm**
  This experiment used the same data set of the two past experiments, but it uses the Matlab AI tools to replicate the experimental conditions of [3]. In this case the iterations can not be fixed then just the five times the routine was performed.

- **Experiment 4: SP40LATAM data set Top Ten companies Analysis**

  This experiment consisted of the selection of the top ten profit enterprises into the SP40 Latin American index extracted from its official web-page. Thus, as a posterior step after the results visualization of the three previous experiments, three sets of six companies were selected to perform a time-frequency peaks comparison. The first set was the top six of the companies. The second set was the outline data consider into the top. And the last set was the not related companies neither in the business sector nor in the profit.

The experimental conditions for all these algorithms were the same in terms of hardware and software. And for all of them the hyper-parameters of the learning rate was equal to 0.9 and layers $FC = [1010]$ stayed fixed.

# Chapter 6

# Results

The results of each experiment are presented to determine the points of interest of the comparisons of ISOM SP40, SOM IBEX35, and SOM NYSE. Also, the analysis of the facts that can be deduced from the newly available information was explored in detail. There is a total of four experiments, and their results are represented by different sources of analysis.

## 6.1 Results Analysis Resources

In the first three subsections in general terms, a table is provided with three columns: the number of iterations, accuracy of the top 8 enter prices, and the execution time in seconds. The content of that columns is over understood, except for the top 8 accuracy, it is about how every clustering test has its graphic result in the correspondent SOM; thus it was affirmative if those eight companies were near or negative if they not. The Yes/No case represents a particular distribution in which the companies, even all the top 8 enterprises, were no close enough; the SOM distribution represents a relation between them in subsections. The second source is the SOM graph generated by the models and analyzed thoroughly bellow. Moreover, for the last experiment, three graphs are provided to an overall analysis trough the time. A complete overview of the source codes used in all the experiments is attached in Appendix Section.

Table 6.1: SP40 Latin America Top 9 Strongest Companies

| ID | Company name | Ticker symbol | Industry | Country |
|---|---|---|---|---|
| 31 | Itaú Unibanco | NYSE: ITUB | Banking Brazil | Brazil |
| 39 | Vale | NYSE: VALE.P | Mining Brazil | Brazil |
| 03 | Banco Bradesco | NYSE: BBD | Banking Brazil | Brazil |
| 34 | Petrobras | NYSE: PBR.A | Oil Brazil | Brazil |
| 06 | Banco do Brasil | B3: BBAS3 | Banking Brazil | Brazil |
| 17 | Companhia de Bebidas das Americas (AmBev) | NYSE: ABEV | Beverages Brazil | Brazil |
| 02 | América Móvil | BMV: AMX L | Telecommunications Mexico | Mexico |
| 27 | Fomento Económico Mexicano (FEMSA) | BMV: FEMSA UBD | Beverages Mexico | Mexico |
| 32 | Itaúsa Investimentos Itau | B3: ITSA4 | Banking Brazil | Brazil |

## 6.2 Experiment 1

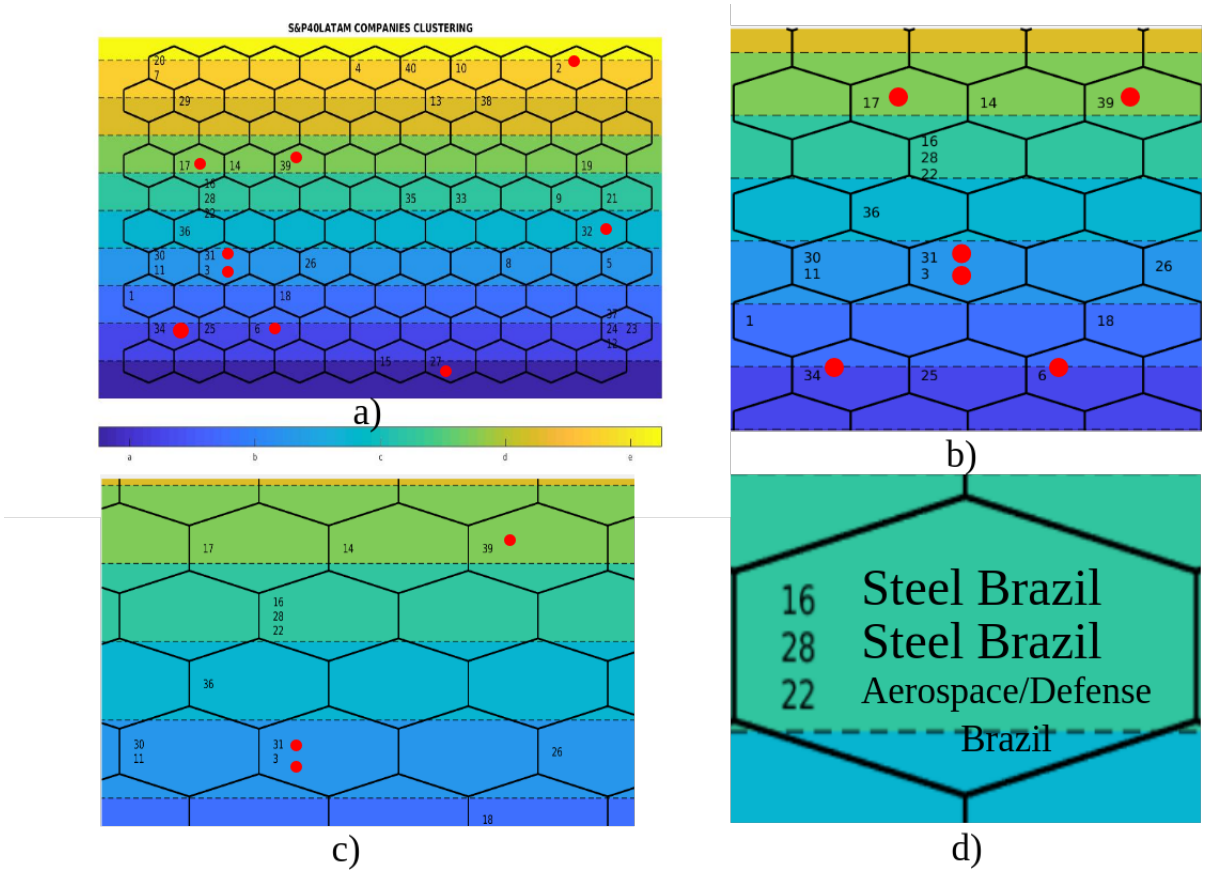


Figure 6.1: Experiment 1 - Clustering and Density Results for ISOM SP40 Algorithm

Table 6.2: Experiment 1 - Execution Time Comparison and Qualitative Results for ISOM SP40 Algorithm

| Experiment 1. S&P40 Data set in ISOM SP40 Algorithm | | |
|---|---|---|
| N Iterations | Accuracy Top8 | Time(s) |
| 10 | Yes | 5,79E-01 |
| 10 | Yes/No | 5,95E-01 |
| 10 | Yes/No | 5,77E-01 |
| 10 | Yes | 5,71E-01 |
| 10 | Yes | 5,79E-01 |
| 50 | No | 2,91E+00 |
| 50 | Yes/No | 2,94E+00 |
| 50 | Yes/No | 2,95E+00 |
| 50 | Yes/No | 2,96E+00 |
| 50 | No | 3,05E+00 |
| 200 | No | 1,23E+01 |
| 200 | Yes/No | 1,17E+01 |
| 200 | No | 1,83E+01 |
| 200 | No | 1,78E+01 |
| 200 | No | 1,52E+01 |
| 500 | No | 3,25E+01 |
| 500 | Yes/No | 3,12E+01 |
| 500 | Yes/No | 3,39E+01 |
| 500 | Yes/No | 3,32E+01 |
| 500 | No | 3,14E+01 |

This initial experiment was developed with two objectives: first, to calibrate the parameters correctly in the proposed the model, and to measure the metrics for being compared. Thus, after the configuration of settings specified before in the experimental setup chapter, the table 6.2,and graph 6.1 were obtained.

The first idea that can be extracted for Tab. 6.2 have a relation with the execution training time, it was found a significant improvement while the number of iterations was reduced. The referent number of iterations toke form the previous studied tends to overfit the model. In the majority of the cases, it did no establish a uniform relation of the top nine enterprises in 6.1. The best approximation was $0,579 seconds$, and it represents almost two decimal places of difference with the worst-case founded in NYSE SOM method analyzed later. In further cases, those two decimals are essential to take advantage of the hardware resources in studies with much more amount of data.

At the same time, on the 6.1 the red points represented the top nine enterprises listed in Tab. 6.2. Here the *a*) part of the figure presents a complete overview of the SOM distribution. The verification points are well accurate clustered as at least the top 6 enterprises in part *c*) and *b*) of the graph are inside a 3 range of neighborhood. That means that from each hexagon corner there is maxi mun 3 corners of separation. Indeed, an explanation can be deduced with the result of the fourth experiments for the three important points outside of the profit area. It could be related to their position in the top 9 list.

The enterprises in the profit area that were not in the top 9 list, this time were zoom in the part *d*) of the graph. In this way it is verified this behavior with the other two density correlation metrics. Also, there was quick to detect that they belong to the same industrial area the metallurgic, and more much consistent in the same geographical location Brazil.

In the first moment, the background colors in all graphs expected to detect concentration areas of profit or looses. Even though, at the end with the experimental repetitions, just the 50% of them correspond to a correct association which that percentage it is not considered relevant. The situation for these data set and algorithm is that with a lower size of iteration they always find good approximations. On the other hand, its worst fact is that two of five in the repetitions in these experiments gave relationships somehow ambiguous.

Table 6.3: Experiment 2 - Execution Time Comparison and Qualitative Results for IBEX35 Algorithm

| Experiment 2. S&P40 Data set in IBEX35 Algorithm | | |
|---|---|---|
| N Iterations | Accuracy Top8 | Time(s) |
| 10 | No | 6,90E+00 |
| 10 | Yes/No | 6,35E+00 |
| 10 | No | 6,40E+00 |
| 10 | No | 6,15E+00 |
| 10 | No | 6,06E+00 |
| 50 | No | 3,08E+00 |
| 50 | No | 3,78E+00 |
| 50 | No | 4,92E+00 |
| 50 | No | 3,01E+00 |
| 50 | No | 3,01E+00 |
| 200 | No | 3,00E+05 |
| 200 | Yes/No | 1,25E+01 |
| 200 | Yes | 1,21E+01 |
| 200 | No | 1,19E+01 |
| 200 | Yes/No | 1,19E+01 |
| 500 | No | 2,87E+01 |
| 500 | Yes | 2,99+01 |
| 500 | No | 2,89E+01 |
| 500 | No | 3,23E+01 |
| 500 | No | 2,94E+01 |

## 6.3 Experiment 2

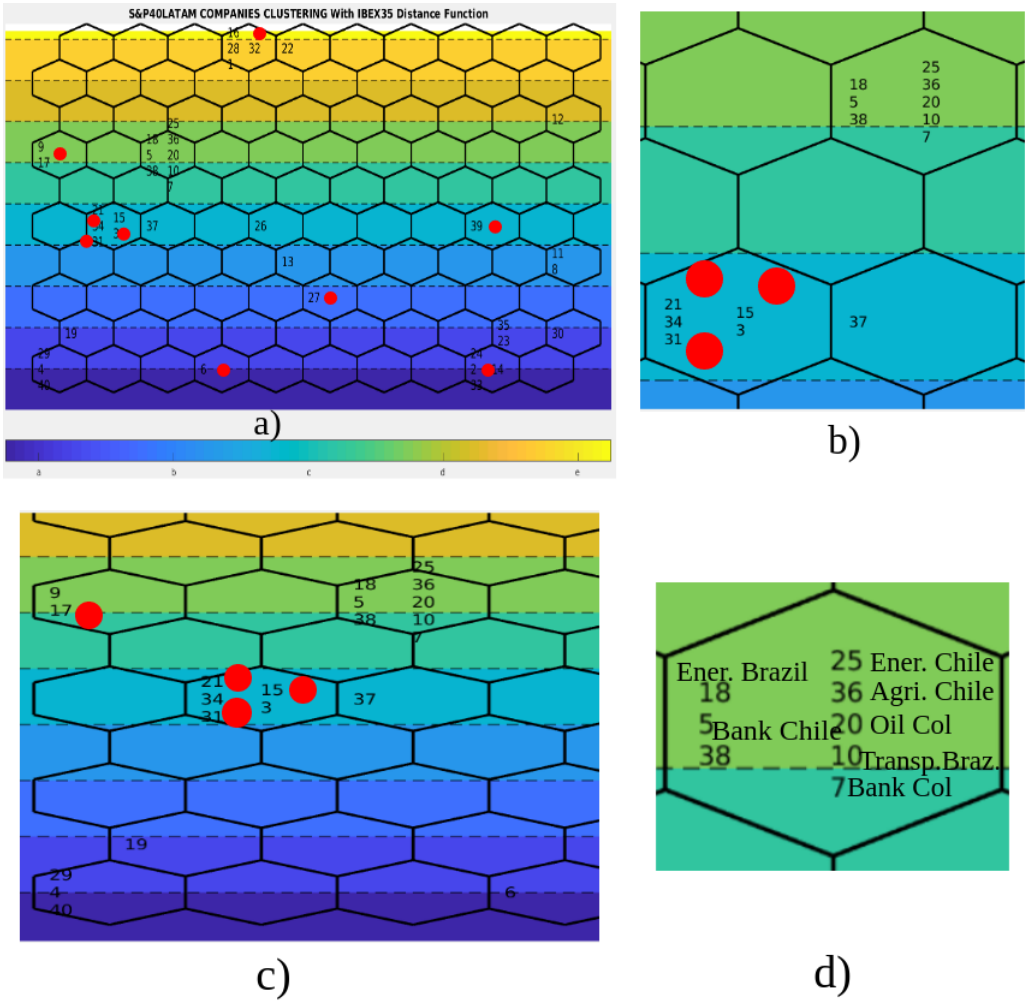


Figure 6.2: Experiment 2 - Clustering and Density Results for IBEX35 Algorithm

This experiment follows the same experimental line that the previous experiment, the table 6.3 and the figures 6.2 represent the same characteristics. The different method applied IBEX35 give us a different perspective for the analysis. Talking about the Table 6.3, the number of iterations of the five last repetitions were the same as in the 2018 study [2]. However, they do not have the same accuracy for this data set. This result is very concordant with the analysis performed in that time because an additional data set related to a Latin American index was studied without success. However, in the rest of the iterations, there is a negative behavior in the results as it does not find relations in the companies 16 times of 20. The execution time is another notable factor, in the best case for ten iterations it was $6,06E+00(s)$. This time it is six times bigger than the time execution in the proposed method ISOM SP40 $5,79E-01(s)$. The result is demonstrating that the machine operations modified by the Euclidean and the Manhattan distances are crucial because their impact in the time execution. They could mean be a big deal.

Table 6.4: Experiment 3 - Execution Time Comparison and Qualitative Results for NYSE Algorithm

| Experiment 3. S&P40 Data set in NYSE Algorithm | | |
|---|---|---|
| N Iterations | Accuracy Top8 | Time(s) |
| NA | NA | NA |
| NA | NA | NA |
| NA | NA | NA |
| NA | NA | NA |
| NA | NA | NA |
| NA | NA | NA |
| NA | NA | NA |
| NA | NA | NA |
| NA | NA | NA |
| NA | NA | NA |
| 200 | No | 9.23E+00 |
| 200 | No | 9.10E+00 |
| 200 | No | 8,91E+00 |
| 200 | No | 8,95E+00 |
| 200 | Yes/No | 9,10E+00 |
| NA | NA | NA |
| NA | NA | NA |
| NA | NA | NA |
| NA | NA | NA |
| NA | NA | NA |

For the graph 6.1, we can observe in the section *a*) that the distribution across the map is not grouping more than three companies correctly. They are separated with more than three corners of the neighborhood. Once again, the color distribution does not show a correlation with the group in more than 70% of the proofs. And going more in-depth in the cell amplified in *d*), there are concentrated eight companies that do not belong to the top ten and are besides the important concentration group illustrated in *b*). These companies were identified, and even there are a couple of energy companies, and banking its density correlations (business sector and geographical area) not allow a real connection between these companies. Then, the important issue of this method, although its low accuracy, is the adaptable and modular architecture. Architecture that could work as a base for developing new models for the different world regions.

## 6.4 Experiment 3

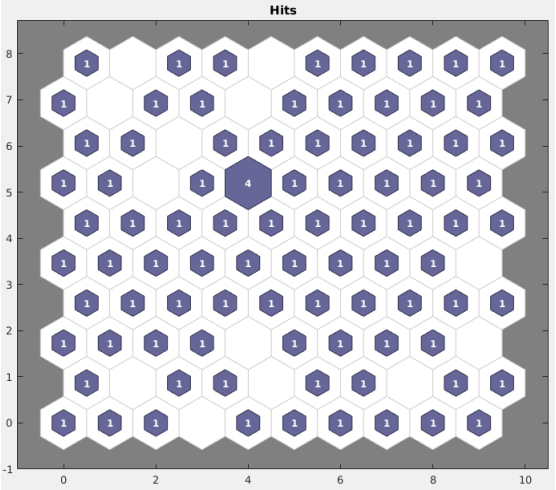


Figure 6.3: Experiment 3 - Clustering Results for NYSE Algorithm

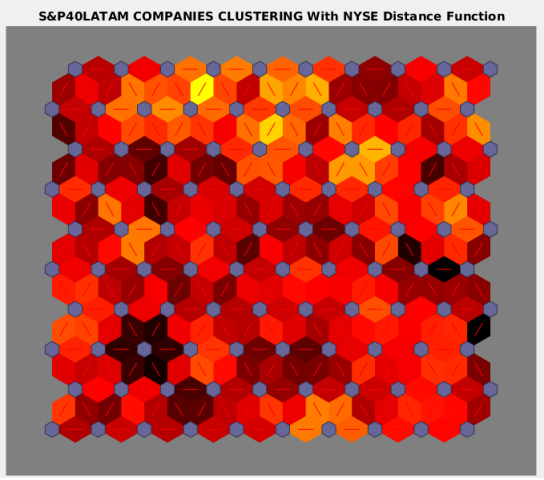


Figure 6.4: Experiment 3 - Density Results for NYSE Algorithm

The architecture of this experiment is explained the work developed in 2015 [3]. The documentation of the SOM architecture is similar to the steps described in the methodology section. Figure. 6.3 represents a layer of 10 by 10 neurons, and the number contained in the hexagons is the number of inputs grouped by the neuron. Then, from near 100 neurons, just classified a single company and just one cell achieve to get four together. In the same way in the Figure. 6.4, there is the representation of the "SOM Neighbor Distances", which also uses the hexagons. Later, the colors represent areas in which the profits are grouped, thus the darkest areas correspond to the major profits. In the distribution, the lightest is not completely clear, but as long as the profits decrease, the

color too. Thereby, all the density correlations can not be well visualized because of the graphical distribution design proper of matlab tool.

Although the data set was easily processed, there were some parameters that cannot be fixed. In this case, the number of iterations 200 that was one of the dynamical features in the experiments. Because of this, the rest of the experiments were also performed with this number to get a comparative section. The Matlab tool for AI measured the training time automatically, and in the best case, it has $9,10E+00(s)$; thus, it is inferior to the same number of iterations in the two previous experiments, which have $1,170E+01(s)$ and $1,190E+01(s)$ in the best 200 iterations case. For those reasons, the final statement for this method is that the data can not be easily adapted to new data sets. Essentially because the other functions cannot be manipulated to get improvements by the Matlab tools as they are not open. One point on favor is time training enhancement, but if the accuracy is not working, it goes to a second place of relevance.

## 6.5 Experiment 4

In this experiment, we corroborate the accuracy of the results found in the first experiment with iSOM SP40 proposed method with a time events comparison for selected companies. Also, the behavior correspondent to the density correlation of the business sector is better explained with the time period frequency graphs.

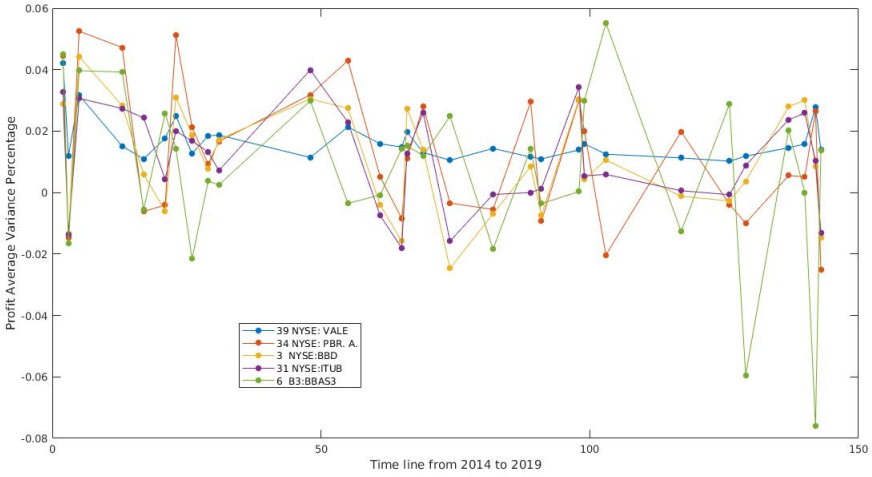


Figure 6.5: Experiment 4 - Time Frequency Analysis For Top Ten Companies in SP40LATAM index

Figure. 6.5, shows the tendency lines comparison between the top 6 enterprises with very satisfactory results as they manifest similar incidences. In other words, the stock price behavior of the companies used as a reference in the clustering accuracy has a similar variation percentage in the events through the time. The accuracy established in experiment one is affirmed as the positive and negative peaks make them related

companies.Hence, it could be affirmed that those enterprises that are near this group with at least three hexagon corners of separations have a similar profit by the time.

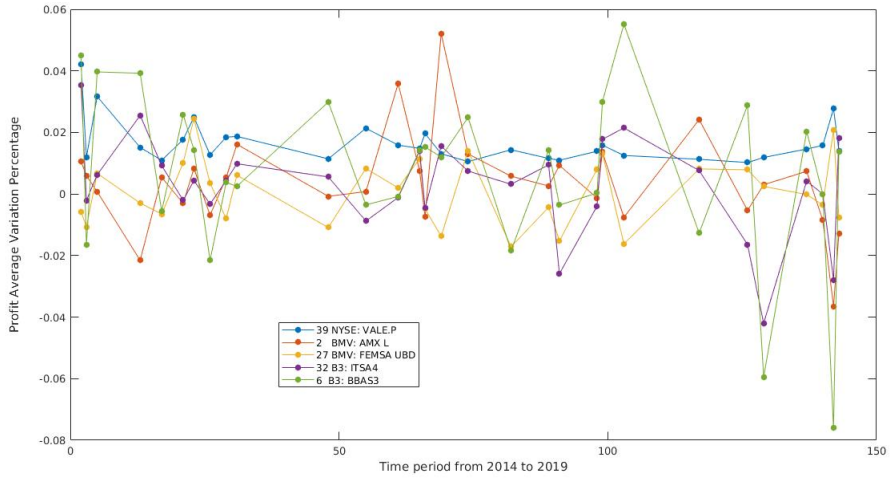


Figure 6.6: Experiment 4 - Time Frequency Analysis For Outline Companies into Top Ten Companies in SP40LATAM index but not near in SOM

Figure. 6.6 allows to visualize the comparison between the top 9 enterprises outside the SOM map profit area showing tendency lines not so related. That means that the SOM makes a good interpretation of the relations among the enterprises. Sometimes this can give the idea that the network is just making an aleatory selection of the top enterprises, but it is not the case. Besides the outline position of this data in the SOM map, they occupied the last three positions in List. 6.1. According to the bibliographical resources reviewed, this top ten last companies are in constant change because of the money fluctuation. Also it is influenced by the entries and the exits of new companies to the index SP40 LATAM, which have more concordance with the Figure. 6.1 *a*) and with Figure. 6.6.

The results of the last part of the experiment are plotted in the Figure. 6.7, here the comparison among non-related companies neither in density correlation of geography, industrial area or profit is performed by time period. They are spread in opposite limit corners or far areas in the SOM map of the proposed method. In this Figure.6.7, it is notable the nonuniform coincidence in the peaks of the variation during the five years. Thus, it once again demonstrates the reliability of the proposed model.

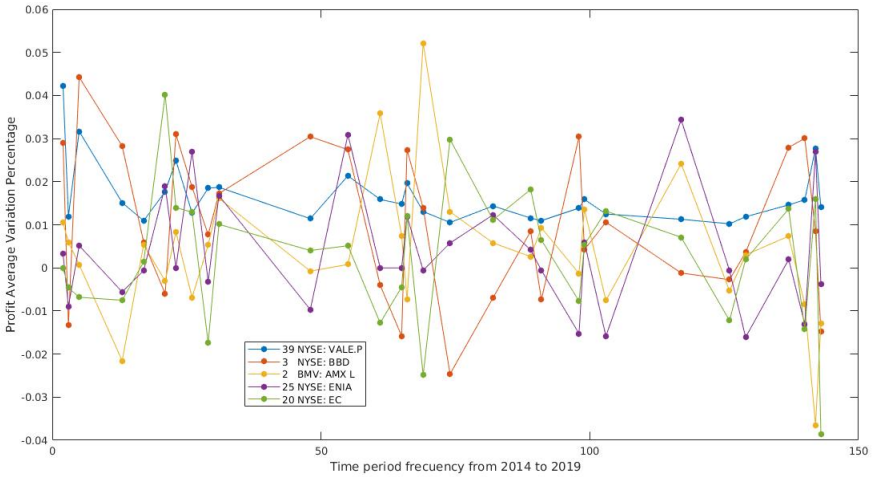


Figure 6.7: Experiment 4 - Time Frequency Analysis for nonrelated Companies in SP40LATAM neither in SOM position nor in Business Sector

## 6.6 Overall results

The concatenation of the results obtained after the implementation of the three methods is available in Tables. 6.2,6.3, and 6.4. It allows to determine that the actual proposed method is superior in terms of accuracy of density correlations with the other two methods. It presents an 80% of well clustering performance against a 20% of IBEX35, and a 5% in NYSE methods. At the same time, under the ten times iteration condition, the proposed model evidence the minimum time in execution over the three methods. Both deductions make clear that the developed model is improved to work with the SP40 Latin American Index.
In addition, a Github repository containing the information of the whole work and the final results are attached in this link.

# Chapter 7

# Conclusions and Future work

This chapter closes the whole study presenting the finding of the research according to the objectives proposed. Also here are displayed the possible improvements and further studies that can be developed from this work.

## Conclusions

Broadly speaking, the implementation of an improved SOM algorithm for the SP40 Latin America data set was achieved successfully in terms of reduction of time execution, and accuracy of the density correlations. The relationships among the top nine companies of the index were corroborated in the SOM map distributions and a final time-frequency analysis. Being precise in technical terms and specific metrics, it is concluded that:

1. The prepossessing of the data sets were improved with practical tools such as Python libraries panda and numpy allowing the automatizing of data treatment and concatenation. This works with different formats and is compatible with multiple programming languages and development environments.

2. The hexagonal architecture of the SOM has demonstrated a good performance in the studies, specifically in the financial field for stock market prediction and clustering. Also, the proposed algorithm implemented used hexagonal structure for being compared with the performance of the architectures in [2] and [3] studies. Thus, the ISOMSP40 showed positive results in its performance.

3. The design of the SOM architecture comes along with the improvement of the topological distance used in the neighborhood function. This work demonstrated that besides the classic Euclidean distance, there is the Manhattan Distance, which reduces the machine operations. This topological change did not affect the accuracy of the densities correlations demonstrating its adaptability to the SP40 Latin America index market.

4. The performance of the three methods with the SP40 LATAM data set was compared, showing that the adequate algorithm for those companies is the proposed

ISOMSP40 proposed algorithm. This statement is corroborated with the metrics selected for the comparison execution time, and the different density correlations. Thus, the higher accuracy corresponds to the proposed method with 80% overall in all the experiments against the 10%, and 5% of the other two methods.

5. The execution time was reduced in almost two significant decimals having $5,79E-01(s)$ as the minimum time in the experiments with ten iterations and well-sorting distribution in SOM. In contrast, the IBEX35 method, which even with the ten iterations it achieve at least $6,15E+00(s)$ time of execution.

6. The density correlations were pointed out by the Figure. 6.1 and verified by Figure. 6.5, with the top nine enterprises and the time-frequency analysis among the companies ITUB, VALE, BBD, PBRA, the profit analysis is done. Then, the geographic distribution and business sector were also verified in the amplified cell, which group SID, ERJ, and GGB three metallurgic enterprises from Brazil.

7. This work boosts the automation of the data pre-processing of any set extracted from reliable sites such as Yahoo Finances. These data can be handled by any structure that admits the python libraries formats, which are almost all the existent to the date. Secondly, the SOM adaptation to the SP40LATAM data set was achieved through the calibration of the hyper-parameter of the iterations. Finally, the modification of the topological distance formula by the Manhattan equation has obtained well-sorted clusters and reducing the time of execution.

# Future Works

As future works, first with the whole information generated and the interpretations of the study, these can be transformed into a more customer-oriented tool. This tool should own a graphical interface in which any user with non-programming skills can easily set the parameters. It should be very visual and owns interactive indicators such as set combinations of the different data sets and methods for being presented with a click. If the implementation and use of these tools become popular, the risk of investment would be minimized.

Finally, from the mathematical side, there are several topological possibilities that can be analyzed and implemented to still improving the SOM models. As it was shown, just a few of these mathematical functions are used in neighbors functions, which impacts a lot in accuracy measures. In addition, the improvement of geometric structures inside self-organized maps can also be combined in order to get new models that can fit different scenarios.

# References


[1] T. Kohonen, "Essentials of the self-organizing map," *Neural Networks*, vol. 37, pp. 52 – 65, 2013, twenty-fifth Anniversay Commemorative Issue. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608012002596

[2] G. E. Pilliza, O. A. Román, W. J. Morejón, S. H. Hidalgo, and F. Ortega-Zamorano, "Risk analysis of the stock market by means self-organizing maps model," in *2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM)*, Oct 2018, pp. 1–6.

[3] M. H. Wu, "Financial market prediction," *Preprint submitted to arXiv*, 2015. [Online]. Available: https://pdfs.semanticscholar.org/e043/45fda2bd571209f202fe2abd7e743e981587.pdf

[4] H. Hu, L. Tang, S. Zhang, and H. Wang, "Predicting the direction of stock markets using optimized neural networks with google trends," *Neurocomputing*, vol. 285, pp. 188 – 195, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231218300572

[5] T. E. Times. (2019) Definition of 'investment risk'. [Online]. Available: https://economictimes.indiatimes.com/definition/investment-risk

[6] M. Tkáč and R. Verner, "Artificial neural networks in business: Two decades of research," *Applied Soft Computing*, vol. 38, pp. 788 – 804, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494615006122

[7] C. Chang, Z. Lin, W. Koc, C. Chou, and S. Huang, "Affinity propagation clustering for intelligent portfolio diversification and investment risk reduction," in *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, Nov 2016, pp. 145–150.

[8] A. H. Moghaddam, M. H. Moghaddam, and M. Esfandyari, "Stock market index prediction using artificial neural network," *Journal of Economics, Finance and Administrative Science*, vol. 21, no. 41, pp. 89 – 93, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2077188616300245

[9] K. Kumar and C. Tan, "Artificial intelligence in financial distress prediction."

[10] F. Fanita and Z. Rustam, "Predicting the jakarta composite index price using anfis and classifying prediction result based on relative error by fuzzy kernel c-means,"

*AIP Conference Proceedings*, vol. 2023, no. 1, p. 020206, 2018. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.5064203

[11] Y. Finances. (2018) Exchanges and data providers on yahoo finance. [Online]. Available: https://nz.help.yahoo.com/kb/exchanges-data-providers-yahoo-finance-sln2310.html

[12] M. Nwadiugwu, "Neural network, artificial intelligence and the computational brain," Ph.D. dissertation, 08 2015.

[13] M. Negnevitsky, Addison-Wesley, Ed., 2005, no. 2.

[14] A. Kaplan and M. Haenlein, "Rulers of the world, unite! the challenges and opportunities of artificial intelligence," *Business Horizons*, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0007681319301260

[15] A. G. Barto and R. S. Sutton, "Chapter 19 - reinforcement learning in artificial intelligence," in *Neural-Network Models of Cognition*, ser. Advances in Psychology, J. W. Donahoe and V. P. Dorsel, Eds. North-Holland, 1997, vol. 121, pp. 358 – 386. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0166411597801057

[16] A. G. E. Collins, "Reinforcement learning: bringing together computation and cognition," *Current Opinion in Behavioral Sciences*, vol. 29, pp. 63 – 68, 2019, sI: 29: Artificial Intelligence (2019). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S235215461830175X

[17] M. O. Afolabi and O. Olude, "Predicting stock prices using a hybrid kohonen self organizing map (som)," in *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, Jan 2007, pp. 48–48.

[18] T. J. Oyana, L. E. Achenie, E. Cuadros-Vargas, P. A. Rivers, and K. E. Scott, "A mathematical improvement of the self-organizing map algorithm," in *Proceedings from the International Conference on Advances in Engineering and Technology*, J. Mwakali and G. Taban-Wani, Eds. Oxford: Elsevier Science Ltd, 2006, pp. 522 – 531. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780080453125500571

# Appendices

# Appendix A

# List of Stock Indexes and Companies

Table A.1: SP40 Latin America Companies Numbers and Tickets

| ID | Company name | Ticker symbol | Industry | Country |
|---|---|---|---|---|
| 01 | ALFA | BMV: ALFA A | Conglomerate Mexico | Mexico |
| 02 | América Móvil | BMV: AMX L | Telecommunications Mexico | Mexico |
| 03 | Banco Bradesco | NYSE: BBD | Banking Brazil | Brazil |
| 04 | Banco Santander Chile | NYSE: BSAC | Banking Chile | Chile |
| 05 | Banco de Chile | BCS: CHILE | Banking Chile | Chile |
| 06 | Banco do Brasil | B3: BBAS3 | Banking Brazil | Brazil |
| 07 | Bancolombia | NYSE: CIB | Banking Colombia | Colombia |
| 08 | BM&F Bovespa | B3: BVMF3 | Stock Exchange Brazil | Brazil |
| 09 | BRF S.A. | NYSE: BRFS | Food processing Brazil | Brazil |
| 10 | CCR S.A. | B3: CCRO3 | Transportation Brazil | Brazil |
| 11 | Cemex | BMV: CEMEX CPO | Cement Mexico | Mexico |
| 12 | Cencosud | BCS: CENCOSUD | Retail Chile | Chile |
| 13 | Cielo S.A. | B3: CIEL3 | Financial services Brazil | Brazil |
| 14 | Compañía de Minas Buenaventura | NYSE: BVN | Mining Peru | Peru |
| 15 | Companhia Energetica de Minas Gerais (CEMIG) | NYSE: CIG | Energy Brazil | Brazil |
| 16 | Companhia Siderúrgica Nacional | NYSE: SID | Steel Brazil | Brazil |

| ID | Company name | Ticker symbol | Industry | Country |
|---|---|---|---|---|
| 17 | Companhia de Bebidas das Americas (AmBev) | NYSE: ABEV | Beverages Brazil | Brazil |
| 18 | CPFL Energia | NYSE: CPL | Energy Brazil | Brazil |
| 19 | Credicorp | NYSE: BAP | Banking Peru | Peru |
| 20 | Ecopetrol | NYSE: EC | Oil Colombia | Colombia |
| 21 | Grupo Elektra | BMV: ELEKTRA * | Retail Mexico | Mexico |
| 22 | Empresa Brasileira de Aeronáutica (Embraer) | NYSE: ERJ | Aerospace/Defense Brazil | Brazil |
| 23 | Empresas CMPC | BCS: CMPC | Paper/Pulp Chile | Chile |
| 24 | Empresas Copec | BCS: COPEC | Energy Chile | Chile |
| 25 | Enel Américas | NYSE: ENIA | Energy Chile | Chile |
| 26 | Enel Generación Chile | NYSE: EOCC | Energy Chile | Chile |
| 27 | Fomento Económico Mexicano (FEMSA) | BMV: FEMSA UBD | Beverages Mexico | Mexico |
| 28 | Gerdau | NYSE: GGB | Steel Brazil | Brazil |
| 29 | Grupo Financiero Banorte | BMV: GFNORTE O | Banking Mexico | Mexico |
| 30 | Grupo Televisa | BMV: TLEVISA CPO | Media Mexico | Mexico |
| 31 | Itaú Unibanco | NYSE: ITUB | Banking Brazil | Brazil |
| 32 | Itaúsa Investimentos Itau | B3: ITSA4 | Banking Brazil | Brazil |
| 33 | LATAM Airlines Group | NYSE: LFL | Airline Chile Chile / Brazil | Brazil |
| 34 | Petrobras | NYSE: PBR.A | Oil Brazil | Brazil |
| 35 | S.A.C.I. Falabella | BCS: FALABELLA | Retail Chile | Chile |
| 36 | Sociedad Química y Minera de Chile | NYSE: SQM | Agricultural Chemicals Chile | Chile |
| 37 | Southern Copper Corp. | NYSE: SCCO | Mining Peru | Peru |
| 38 | Ultrapar Participacoes S.A. | B3: UGPA3 | Energy Brazil | Brazil |
| 39 | Vale | NYSE: VALE.P | Mining Brazil | Brazil |
| 40 | Wal-Mart de México | BMV: WALMEX V | Retail Mexico | Mexico |

# Appendix B

# Project Source Code

This chapter included the main codes and routines implied in the experimental setup, all the auxiliary functions are stored in the Github repository attached at the end of the results section.

## B.1 Prepossessing

Source Code 1

```
#Preprocesing Code

#This code is designed for the preprosing data of the .cvs files obtained from Yahoo finances stock Market

#The download of this libraries is strictly necessary.
import pandas, os, sys, openpyxl
import numpy as np


meses = {'01':'ene.',
'02':'feb.',
'03':'mar.',
'04':'abr.',
'05':'may.',
'06':'jun.',
'07':'jul.',
'08':'ago.',
'09':'sept.',
'10':'oct.',
'11':'nov.',
'12':'dic.'}

#Include exclusively all the stock markets files in a single path.Change './input' for the current path.
archivos = os.listdir(os.path.join(sys.path[0],'./input'))
excel_writer = pandas.ExcelWriter((os.path.join(sys.path[0],'final.xlsx')), engine='xlsxwriter')
print(archivos)

#Starting the loading for concatenation
for archivo in archivos:
    data_frame = pandas.read_csv(os.path.join(sys.path[0],'./input/' + archivo))
    for j in data_frame['Date'].iteritems():
        fecha = j[1].split("-")
        nuevaFecha = fecha[2] + " " + meses[fecha[1]] +" " + fecha[0]
        data_frame.set_value(j[0], 'Date',nuevaFecha)
        #Incluiding and setting the names of all the parameters
    data_frame.rename(columns={'Date':'Fecha', 'Open':'Abrir', 'High':'Mx.', 'Low':'Mn.', 'Close': 'Cierre*', 'Adj
        Close':'Cierre ajus.**', 'Volume':'Volumen'}, inplace=True)
    nombre = archivo.split(".")
    data_frame.to_excel(excel_writer, nombre[0] + "." + nombre[1], index = False)
excel_writer.save()

#All the information will be wirtten in 'final.xlsx'

wb = openpyxl.load_workbook((os.path.join(sys.path[0],'final.xlsx')))
for sheet in wb.worksheets:
```

```
        sheet.insert_rows(1)
        mycell = sheet['A1']
        nombre = sheet.title.split(".")
        mycell.value = nombre[0]
        mycell = sheet['B1']
        mycell.value = nombre[1]
    wb.save((os.path.join(sys.path[0], 'final.xlsx')))
```

## B.2 ISOM SP40

Source Code 3

```
clear;
clc;
close all;

fileName='DataDailyS&P40.mat';
load(fileName);
data1=num2data(num);
txt1=txt;



%Hyperparameters
data=[data1];
data(ismissing(data))=0;
txt=[txt1];
eta0=0.9;
IterMax=500;
FC=[10 10];

Indices=GenerarIndices(FC);
[NumDatos,NumA]=size(data);
W=rand(NumA,FC(1),FC(2));

initime=tic;
for i=1:IterMax
    fprintf('i: %d\n',i);
    ind=randperm(NumDatos);
    for j=1:NumDatos
        eta=eta0*(1-i/IterMax);
        Patron=(data(ind(j),1:NumA))';

        [Gx,Gy]=CalculoGanadora(W,Patron);
        IndGan=[Gx,Gy]';
        %%Aqu se llama la funcion de vecindad
        [Vecindad,Dist]=FuncionVecindadSP40(IndGan,Indices);
        W=IncrementarPesos(W,Patron,Vecindad,eta);
    end

end
finaltime=toc(initime)
Ganadoras=CalculoGanadorasData(W,data);
DibujarWFcHex(FC)
set(gca,'XColor', 'none','YColor','none')
title 'S&P40LATAM COMPANIES CLUSTERING With ISOM SP40 Distance Function'
EscribirTickets(Ganadoras,FC,txt,Indices)
```

## B.3 SOM IBEX35

Source Code 3

```
clear;
clc;
close all;

fileName='DataDailyS&P40.mat';
load(fileName);
data1=num2data(num);
txt1=txt;



%Hyperparameters
data=[data1];
data(ismissing(data))=0;
txt=[txt1];
eta0=0.9;
IterMax=500;
FC=[10 10];

Indices=GenerarIndices(FC);
[NumDatos,NumA]=size(data);
W=rand(NumA,FC(1),FC(2));

initime=tic;
for i=1:IterMax
    fprintf('i: %d\n',i);
    ind=randperm(NumDatos);
    for j=1:NumDatos
        eta=eta0*(1-i/IterMax);
        Patron=(data(ind(j),1:NumA))';

        [Gx,Gy]=CalculoGanadora(W,Patron);
        IndGan=[Gx,Gy]';
        %%Aqu se llama la funcion de vecindad
        [Vecindad,Dist]=FuncionVecindad(IndGan,Indices);
        W=IncrementarPesos(W,Patron,Vecindad,eta);
    end

end
finaltime=toc(initime)
Ganadoras=CalculoGanadorasData(W,data);
DibujarWFcHex(FC)
set(gca,'XColor', 'none','YColor','none')
title 'S&P40LATAM COMPANIES CLUSTERING With IBEX35 Distance Function'
EscribirTickets(Ganadoras,FC,txt,Indices)
```

## B.4 SOM NYSE

Source Code 3

```
% Solve a Clustering Problem with a Self-Organizing Map
% Script generated by NCTOOL
%
% This script assumes these variables are defined:
%
%   simpleclusterInputs - input data.
fileName='DataDailyS&P40.mat';
load(fileName);
data1=num2data(num);

inputs = data1;

% Create a Self-Organizing Map
dimension1 = 10;
dimension2 = 10;
net = selforgmap([dimension1 dimension2]);

% Train the Network
[net,tr] = train(net,inputs);

% Test the Network
```

```
outputs = net(inputs);

% View the Network
%view(net)

% Plots
% Uncomment these lines to enable various plots.
% figure, plotsomtop(net)
%figure, plotsomnc(net)
 figure, plotsomnd(net)
 set(gca,'XColor', 'none','YColor','none')
 title 'S&P40LATAM COMPANIES CLUSTERING With NYSE Distance Function'

% figure, plotsomplanes(net)
% figure, plotsomhits(net,inputs)
% figure, plotsompos(net,inputs)
```

## B.5 Neighbor Function with Mathattan Distance

Source Code 3

```
function [Vecindad,Dist]=FuncionVecindadGiss(IndGan,Indices)
    [~,FCa,FCb]=size(Indices);
    FisGan=Indices(:,IndGan(1),IndGan(2));
    IndGanSOM=repmat(FisGan,1,FCa,FCb);


    Dist=sum(abs(Indices-IndGanSOM));
    Dist=reshape(Dist,FCa,FCb);
    Vecindad=exp(-Dist*2);
```