Analysis of Repair Costs of Scholar Buildings Affected by Earthquakes Using Data Mining. Case Study: Earthquakes of 2017 in Mexico

Graciela García-Rueda[1], Rosa M. Valdovinos[1], Jesús Valdés-González[1], Roberto Alejo[2], J. Leonardo González-Ruiz[1(✉)], and José R. Marcial-Romero[1]

[1] Facultad de Ingeniería, UAEM, CU, Cerro de Coatepec, Toluca, Mexico
leon.g.ruiz@gmail.com

[2] Instituto Tecnológico de Toluca, Av. Tecnológico s/n, 52140 Metepec, Mexico

Abstract. Earthquakes are events that cannot be predicted. However, when they occur, devastating consequences are shown in economic, social and structural areas, among others. In this paper, the mining of association rules is carried out in order to estimate the repair cost required by schools affected during the earthquakes of September 7th and 19th, of 2017 in Mexico. For that, we use the public data collected by the Mexican FONDEN.

Keywords: Data Mining · Association Rules · Analysis of the seismic risk

1 Introduction

Throughout history, the structures collapse is a factor that generates the most material and human losses when earthquakes occur. This is mainly due to the use of low quality materials, deficiencies in construction processes and non-compliance with standards, among some other causes [1].

Data Mining (DM) support the diagnosis and analysis of the structures performance. According to the data characteristics collected after an earthquake, the DM techniques most used is the Association Rules (AR) [2]. An example of its application in earthquake is presented by Martínez-Álvarez et al. [3] who apply descriptive techniques for obtaining Quantitative AR (QAR), and uses a regression method (M5P Algorithm) for predict the earthquakes occurrence based on the relationship between Frequency and magnitude in order to observe the earthquakes variation. The QAR obtained showed that for earthquakes of moderate magnitude (between 3.5–4.4) earthquakes occur after short time intervals, while for high magnitudes (earthquakes of magnitude 4.4 to 6.2) the time intervals in relation to the Frequency and Magnitude had a significant decrease.

In similar way Galán Montaño F. [4] uses DM techniques to describe the behavior of earthquakes according to their magnitude too. In their study,

K. M. Figueroa Mora et al. (Eds.): MCPR 2020, LNCS 12088, pp. 45–56, 2020.
https://doi.org/10.1007/978-3-030-49076-8_5

Montaño uses a genetic algorithm capable of finding frequent patterns and obtain behavior models of the time series according to the occurrence of earthquakes. The result shown that before an earthquake of magnitude greater than 4.5 occurs, it is high probability that an earthquake of magnitude 4.4 occur.

On September 7th 2017, an earthquake of magnitude 8.2 was registered in Oaxaca, Mexico, which caused damage to 57,621 homes, 1,988 schools, 102 cultural buildings and 104 public buildings [5]. Few days after, on September 19th 2017 occurs another earthquake of magnitude 7.1 with epicenter in Puebla, which damaged more than 150 thousand homes in Oaxaca, Chiapas, Guerrero, Puebla, Morelos and State of Mexico, with an estimated repair cost up to 38,150 million of Mexican pesos [6].

After these two earthquakes, in the education sector were registered 12,931 schools with damages: 577 will require a total reconstruction, 1,847 a partial reconstruction and the remains with minor damages. The repair costs was estimated around to the 13,650 million of Mexican pesos [7]. Other structures with affectation were historical and culturally valuable buildings, such as the archaeological zone of Chiapa de Corzo, Zocalo of Mexico City, the National Museum of Art, among others, whose repair has an estimated cost of 8000 million of Mexican pesos [8]. In this paper we apply association rule mining for estimate the repair cost required for reconstructing school buildings damaged by the earthquakes of September 7th and 19th, using the data provided by the Fund of Budget Transparency (the Mexican FONDEN).

It is important regarding that the data bases obtained during the earthquakes of September 7th and 19th are the first of that type that have been obtained in Mexico. For structural engineering purposes the data bases are not full completed, but the data contained is valuable to do by first time a study of earthquake engineering based on DM. One of the main contribution of the work is to explore the use of DM in earthquake engineering. The rules obtained are valuable because they describe the distribution of earthquake damage costs as a function of some basic structural characteristics of the buildings. These rules are exclusive for the earthquakes of September 7th and 19th in México and for the studied structures. Currently, there are not any parallel study of structural engineering that allows to validate the rules.

2 Association Rules

The AR is a technique for discovering interesting associations or correlations from a transactional data set, where the rows represent the transactions and the column the *items* [9]. Let $I = \{i_1, i_2, ..., i_n\}$ be a set of n attributes or *items* and $D = \{t_1, t_2, ..., t_n\}$ a set of transactions in a data set. $\forall t_i \in D$, $\exists\ Tid$ as an unique identifier where $Tid \subset I$ [10].

An association rule can be defined as an implication of the form $X \Rightarrow Y$, where X is the antecedent and Y is the consequent of the rule. For example the rule $\{Bread, Cheese\} \Rightarrow \{Ham\}$, means that, when *Bread* and *Cheese* occur, *Ham* also occurs. For validate the quality of the rules and the probability that they reflect real relationships, two of the most used metrics for this purpose are [11]: **Support:** Probability P that a set of *items* appears in several transaction, $support(X \Rightarrow Y) = P(X \cup Y)$, and **Confidence:** Fraction of transactions in which X and Y appears, $confidence(X \Rightarrow Y) = \frac{support(X \cup Y)}{support(X)}$.

There are different algorithms for obtaining AR, however in this paper we use the Apriori algorithm and the PSO-GES metaheuristic.

2.1 Apriori Algorithm

The Apriori algorithm is one of the first methods developed for association rules mining and is currently one of the most used. Apriori consider two stages [11]: firstly the algorithm identifies all the frequent *itemsets* and then convert them to an AR (see Algorithm 1).

Algorithm 1. A*priori*

Require: Data set D
MinSupport \\minimum expected value of an item in transactions
$i = 0$ $Full\_Item(C_i)$ \\Includes all size 1 items in C
while $C_i \neq \oslash$ **do**
  for X = element of C_i **do**
    if $MinimumSupport(X) \geq MinimumSupport$ **then**
      $L_i = L_i \cup X$
    end if
    $C_{i+1} = SelectCandidates(L_i)$ \\ Candidates with the MinSupport established
    $i = i + 1$
  end for
end while
return C

2.2 PSO-GES

PSO-GES (see Algorithm 2) is a metaheuristic that generates quality AR with relatively low execution times. The algorithm consists of two stages: in the first stage, the dataset is transformed into a binary matrix and non-frequent items are eliminated. In the second stage, rule mining is carried out through a Guided Exploration Strategy. In which only those items that positively influence the fitness of a rule are added during the particle evolution process. This is done by computing fast estimating, using the summary matrix, values for the support and confidence of the rule represented by a new particle [9].

3 Experimental Set-Up

3.1 Database Description

Dataset used in this work corresponds to infrastructure affected of schools during earthquakes of September 2017 in Mexico (ES17M dataset)[1]. ES17M consists of 19,194 records and 76 features, which was designed by Structuralist Experts in seismic risk analysis [1]. Starting from the total features in ES17M, only the most relevant to seismic risk analysis were chosen, leading 36 of the 76 available features (Table 1). The last four rows (funding source) have six values: cost, total amount of attention, amount exercised and description, which gives a total of 36 items.

Algorithm 2. PSO-GES

Require: Number of particles N
Number of iterations T
Number of AR that they want to find M
Dataset D
Number of partitions K
Minimum Support $MinSup$
Minimum Confidence $MinConf$
Inertia Factor w
Acceleration constants c_1 and c_2
Ensure: Returns the M best association rules found $GBests$
$Table \leftarrow CreateTable(D, K)$
$POP \leftarrow CreatePopulation(N)$
while t<T **do**
  for $i \leftarrow 0$ a $size(POP)$ **do**
    for $j \leftarrow 0$ a $size(P_i^t)$ **do**
      $v_i^{(t+1)} \leftarrow w * v_{ij}^t + c_i * rand_1 * (PBest_{ij}^t - P_{ij}^t) + c_2 * rand_2 * (GBest_{ij}^t - P_{ij}^t)$
      $P_{ij}^{t+1} \leftarrow UpdatePosition(P_{ij}^{t+1}, j, v_{ij}^{t+1}, Table, MinSup, MinCof)$
    end for
    if $Fitness(P_i^{t+1}) > Fitness(PBest_i)$ **then**
      $PBest_i \leftarrow P_i^{t+1}$
    end if
    if $Fitness(P_i^{t+1}) > Fitness(GBest_i)$ **then**
      $GBest_i \leftarrow P_i^{t+1}$
    end if
  end for
end while
return GBests

3.2 Preprocessing Data

The following procedure was performed in order to properly format the ES17M dataset as an input for the Apriori and PSO-GES algorithms:

1. Standardized data. All data were transformed into categorical data, which are represented for integer numbers from 0–9. For example, cost was categorized in amounts like 1 of (0 to 1000), 2 of (1001 to 10,000), among others.

[1] Budget Transparency Fund, Fuerza México, available from https://www.transparenciapresupuestaria.gob.mx/es/PTP/fuerzamexico.

Strings like “no aplica” were replaced for categorical numbers as 0. Similar string meaning were clustered, for example, “piso”, “pisos”, “superficie”, were included in the same category represented for an integer number.

2. Structural features generation. These were obtained from pdf documents, jpg or png images from official documents. Distance to the epicenter was computed and its relationship with the closeness to geologic faults lines.
3. Discretized and transactional dataset. The dataset was transformed in a binary matrix where 1 value represents the presence of an *item* and 0 value determines the absence of such a *item*; where *items* were obtained on step 1.

Table 1. Summary of the main database characteristics

Attribute	**Description**
Event Date	Date on which the natural disaster occurred
Workplace	Identification Key in the SEP Catalog
Enrollment	Number of students enrolled in the Educational Center
Entity	Name of the federative entity, according to the INEGI catalog of federative entities
Municipality	Name of the municipality, according to the INEGI catalog of municipalities
Location	Name of the location, according to the INEGI catalog of locations
Latitude	Geo-referenced coordinate of the point where the affected infrastructure is located
Longitude	
Type of damage	Classification, according to its severity of the type of damage suffered by the infrastructure
Damage detail	Description of the damages/affectations that the Infrastructure has
URL dictum	Electronic address of the Structural Verification dictum
Costs	Sum of the value reported by source of financing, which are: • Educational Reform Program • Schools at the 100% • Natural Disaster Fund-Immediate Partial Supports • Natural Disaster Fund-Reconstruction • Private insurance • Other budgetary programs and own resources of the Federal Entity

(*continued*)

Table 1. (*continued*)

Attribute	Description
Total amount of attention, Amount exercised	Sum of the reported value and sum of the exerted value of the total amount by funding source after updates to the original cost, which are: • Educational Reform Program • Schools at the 100% • Natural Disaster Fund-Immediate Partial Supports • Natural Disaster Fund-Reconstruction • Private insurance • Other budgetary programs and own resources of the Federal Entity
Description	General description of the works to be carried out by financing, which are: • Educational Reform Program • Schools at the 100% • Natural Disaster Fund-Immediate Partial Supports • Natural Disaster Fund-Reconstruction • Private insurance • Other budgetary programs and own resources of the Federal Entity
Structural description	Type of structure established in the INIFED structure catalog

Final ES17M dataset (transactional dataset) was generated by keeping only the column number where the presence of an *item* to be 1.

3.3 Algorithms Configuration

Python library[2] was used to run the Apriori algorithm, and PSO-GES was executed with the author's proprietary code[3]. The minimum support threshold was 0.00002 for both algorithms and the confidence threshold= 0.75. Additional PSO-GES parameters used in this work were set as it was presented in Ref. [9]: Population of 20 particles, constants c_1 and $c_2 = 2$, inertia $w = 1$, 10 epoch and one transaction per partition ($K = 1$).

[2] efficient-apriori 1.1.0, https://pypi.org/project/efficient-apriori/.
[3] Bernal Baró G. [9].

4 Results

Due to the databases used do not provide enough information to formally conduct a comprehensive and detailed study of the risk and seismic behavior, the analysis was focused on analyze the relationship between the repair costs and type of construction with variables as: epicentral distance, construction type, damage and seismic zone. For the experiments, the algorithms were adapted in their execution, that is to say, the resulting RA should consider: the epicentral distance and type of construction in the antecedent and type of damage, detail of the damage, total repair cost and seismic zone in the consequent. Derived from the execution of the aforementioned algorithms, the rules generated by the PSO-GES algorithm are presented, since the obtained results provided better RA for the seismic analysis.

4.1 Earthquake of September 7th, 2017

From the AR obtained (Table 2) is possible to determine that most of the schools affected by the earthquake are located in the seismic zone C (intermediate zone, where not so frequent earthquakes) and present moderate damage, mainly in the structure and finishes stand out. Bar graph of Figure 1 shows the relation between the repair costs, the epicentral distance and type of construction. In this graph we can seen that the greatest damage occurred in those schools with masonry walls and concrete slab. Similarly, it is generally appreciated that repair costs decrease as schools move away from the epicenter of the earthquake. Thus, at a lower epicentral distance, repair costs are greater than the repair costs of schools that are located at a greater epicental distance.

4.2 Earthquake of September 19th, 2017

Table 3 shows the most important AR obtained with earthquake data from September 19th. The AR obtained show that most of the affected schools are located in seismic zones C and B (intermediate zones, where earthquakes are not so frequently recorded), which presented moderate and severe damage, mainly were structural, in finishes and outdoors. Figure 2 shows a bar graph that relates the distance from the earthquake epicenter to damaged schools with the repair cost and the type of structure. In this case we can see that the greatest damages occurred in schools that are at an epicentral distance of 101 to 150 km. As for the type of construction, the schools most affected were those built based on masonry walls and light technology.

Table 2. Earthquake association rules of September 7th 2017

Distance	Antecedent	Consequent	Distance	Antecedent	Consequent
Distance 101-150 KM	Masonry walls and concrete slab	Moderate damage, Zone C	Distance 201-250 KM	Masonry walls and concrete slab	Minor damage, Zone C
		Structural damage, finishes, electrical installations			Structural damage
		Total repair cost of $32,643,208-$36,723,608			Total repair cost of $28,562,807-$32,643,207
	Steel and concrete slab	Moderate damage, Zone C		Steel and concrete slab	Moderate damage, Zone C
		Structural damage			Equipment damage, outdoors
		Total repair cost of $816,0802-$12,241,202			Total repair cost of $20,402,005-$24,482,405
	Steel and light roof	Minor damage, Zone C		Steel and light roof	Moderate damage, Zone C
		Damage finishes			Structural damage
		Total repair cost of $12,241,203-$16,321,603			Total repair cost of $20,402,005-$24,482,405
	Masonry walls and light roof	Moderate damage, Zone C		Masonry walls and light roof	Moderate damage, Zone C
		Structural damage			Structural damage, outdoors, finishes
		Total repair cost of $24,482,406-$28,562,806			Total repair cost of $24,482,406-$28,562,806
Distance 151-200 KM	Masonry walls and concrete slab	Moderate damage, Zone C	Distance 251-300 KM	Masonry walls and concrete slab	Minor damage, Zone C
		Structural damage, finishes			Outdoors damage, structural
		Total repair cost of $40,804,010-$44,884,410			Total repair cost of $16,321,604-$20,402,004
	Steel and concrete slab	Moderate damage, Zone C		Steel and concrete slab	Moderate damage, Zone C
		Structural damage and damage to hydraulic and sanitary installations			Structural damage and damage to hydraulic and sanitary installations
		Total repair cost of $32,643,208-$36,723,608			Total repair cost of $20,402,005-$24,482,405
	Steel and light roof	Moderate damage, Zone C		Steel and light roof	Moderate damage, Zone C
		Structural damage			Structural damage
		Total repair cost of $32,643,208-$36,723,608			Total repair cost of $20,402,005-$24,482,405
	Masonry walls and light roof	Moderate damage, Zone C		Masonry walls and light roof	Moderate damage, Zone C
		Structural damage, finishes			Structural damage, outdoors, finishes
		Total repair cost of $816,0802-$12,241,202			Total repair cost of $20,402,005-$24,482,405

It is important to note that, in contrast with the AR obtained form the earthquake of September 7th, the highest repair costs are not found for the shortest epicentral distances. In this case, the highest costs were presented by intermediate epicentral distances. This is explained because the Mexico City is located just at this epicentral distance (100 to 150 km), so having a large number of schools, the repair costs were higher. From the seismic point of view, this is explained due to the soil characteristics in certain areas of Mexico City, were the seismic waves could be amplified. From a practical point of view, this effect was because the accelerations produced in these areas of Mexico City are comparable to those that occur at sites closer to the earthquake epicenter.

5 Discussion

Historically, the provision of public data related to the damage caused by the earthquakes in Mexico was almost null until 2017. The difficulty of field surveys, data truthfulness, capture and processing times are some of the causes that in the past decades they prevented obtain this information. However, given the advances in technology and the wide use of ICTs, the recording of the damages caused by the earthquakes of September 7th and 19th, 2017 constitutes an invaluable source of information for study the behavior, impact and seismic risk. It is still pending to homologate these databases and rethink the type of data that is collected, in order to being able to make more formal studies about the behavior and seismic risk of the constructions. However, having the databases used in this

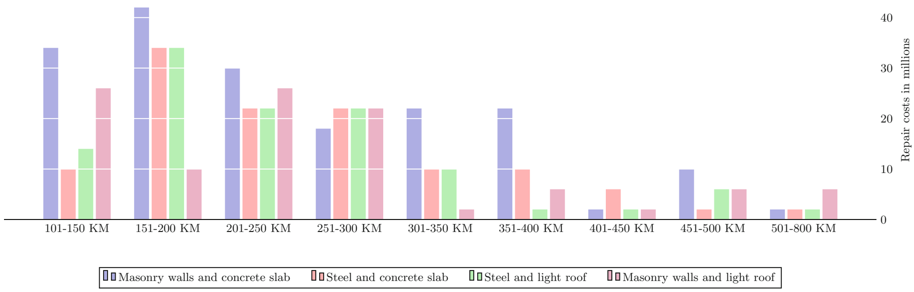

Fig. 1. Values included in the range of estimated costs of damage caused by the earthquake of September 7, 2017, depending on the epicentral distance and the type of construction.

Table 3. Earthquake Association Rules of September 19th 2017

Distance	Antecedent	Consequent
Distance 1-50 KM	Steel and Concrete Slab	Moderate damage, Zone C
		Structural damage
		Total repair cost of $28,562,807-$32,643,207
	Masonry Walls and Light Roofing	Moderate damage, Zone C
		Structural damage
		Total repair cost of $32,643,208-$36,723,608
Distance 51-100 KM	Masonry Walls and concrete slab	Moderate damage, Zone C
		Structural damage
		Total repair cost of $32,643,208-$36,723,608
	Steel and Concrete Slab	Moderate damage, Zone C
		Structural damage
		Total repair cost of $36,723,609-$40,804,009
	Steel and Light Roof	Moderate damage, Zone C
		Structural damage
		Total repair cost of $32,643,208-$36,723,608
	Masonry Walls and Light Roofing	Moderate damage, Zone C
		Structural damage, finishes
		Total repair cost of $32,643,208-$36,723,608
Distance 101-150 KM	Masonry Walls and concrete slab	Severe damage, Zone D
		Structural damage, finishes, outdoor
		Total repair cost of $40,804,010-$44,884,410
	Steel and Concrete Slab	Moderate damage, Zone B
		Structural damage, finishes, outdoor
		Total repair cost of $36,723,609-$40,804,009
	Steel and Light Roof	Severe damage, Zone D
		Structural damage, finishes, electrical, hydraulic and sanitary installations
		Total repair cost of $44,884,411-$48,964,811
	Masonry Walls and Light Roofing	Severe damage, Zone B
		Structural damage
		Total repair cost of $48,964,812-$53,045,212
Distance 151-200 KM	Masonry Walls and concrete slab	Minor damage, Zone B
		Structural damage
		Total repair cost of $32,643,208-$36,723,608
	Steel and Concrete Slab	Minor damage, Zone B
		Structural damage, finishes
		Total repair cost of $24,482,406-$28,562,806
	Steel and Light Roof	Minor damage, Zone B
		Structural damage, electrical installations
		Total repair cost of $24,482,406-$28,562,806
	Masonry Walls and Light Roofing	Moderate damage, Zone B
		Structural damage
		Total repair cost of $28,562,807-$32,643,207
Distance 201-250 KM	Masonry Walls and concrete slab	Minor damage, Zone C
		Structural damage
		Total repair cost of $24,482,406-$28,562,806
	Steel and Concrete Slab	Minor damage, Zone C
		Daño estructural, equipo
		Total repair cost of $24,482,406-$28,562,806
	Steel and Light Roof	Minor damage, Zone C
		Structural damage, outdoor
		Total repair cost of $24,482,406-$28,562,806

work is already a great advance, because commonly other researchers are based on simulations or using only data as: latitude and longitude of the earthquake, magnitude, depth and location. However, this study can be used for different sectors such as health, housing, cultural heritage, among others, provided that the structure of the data is the same as that used in the case of studies.

In this sense, one of the main contributions of the research presented in this paper is the provision of a standardized knowledge base for using machine learning and DM algorithms, making it available to the scientific community for its exploitation and study. The initial results shown in this paper and with the support of an expert in Structural Engineering, was possible to show an overview of the behavior of the repair costs of the schools affected by these earthquakes, depending on the epicentral distance and the type of construction. It this way, it was found that the largest damage was produced by the earthquake of September 19th because the higher reconstruction costs exceed $43'000,000, while the earthquake of September 7th needed costs below this amount. Also was possible to identify that for the earthquake of September 7th (with an epicenter on the Oaxaca coast), at greater epicentral distances the damages were lower compared with the costs recorded for schools located at an epicentral distances closest to the earthquake.

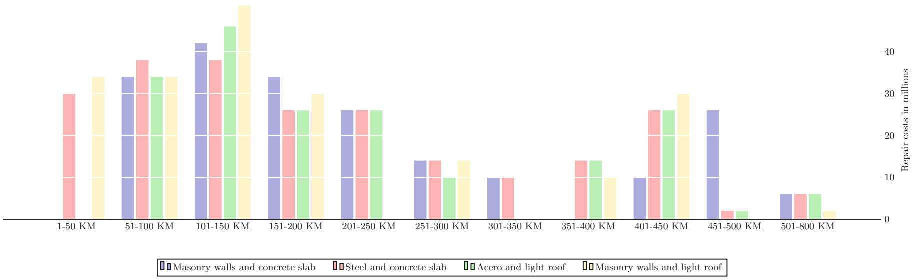

Fig. 2. Values in the range of estimated costs of damage caused by the earthquake of September 19, 2017, depending on the epicentral distance and the type of construction.

On the other hand, in the September 19th earthquake data, was observed how the repair costs increased for distances between 100 Km and 150 km. This behavior was congruent with the existing seismological models, that recognize an amplification of the accelerations in certain areas of Mexico City, which is located in this range of epicentral distances. This explains the increase in repair costs. Regarding the type of construction, it was possible to establish that the structures with masonry walls and light roof were the ones that presented the greatest damage in both earthquakes. With respect to the type of damage that occurred most, it correspond to the damage to the structure and finishes of the schools.

6 Conclusions and Future Works

With the results obtained, it is possible to conclude that the MD can be a useful tool to perform a seismic risk analysis, since it was capable to find relations among different variables related with studied earthquakes and structures. These variables were the distance from the earthquake epicenter to the schools, the type of structure (materials), the type of damage and the reparation costs. It is important to emphasize that the quality of the RA was stablished a confidence degree equal or greater than 75%. The study performed here was focused on obtain a description of the damages caused by the earthquakes of September 7th and 9th to the affected schools.

The open lines of study are initially oriented to study a priori the seismic risk of the constructions, that is, before an earthquake occurs, so the prediction is being worked on to determine the cost of a school if an earthquake of a certain magnitude occurs, having certain characteristics, located at a defined epicentral distance. In same way, is our interest to mining Association Rules for different scenarios of seismic risk analysis, taking specific values, either by the type of damage, seismic zone, federative entity, among other parameters of interest. In the same way, it is contemplated to replicate the preprocessing strategy to other types of buildings, such as hospitals and factories where the main challenge is to include the human losses that unfortunately occurred.

Acknowledgments. This work has been partially supported by the 5046/2020CIC UAEM project and the Mexican CONACYT under scholarship [930395].

References

1. Mondragón, F.P.: Estrategias para el modelado y el análisis sísmico de estructuras históricas. Ingeniería sísmica **1**(83), 1–7 (2010)
2. Gupta, S., Mamtora, R.: A survey on association rule mining in market basket analysis. Int. J. Inf. Comput. Technol. **4**(4), 409–414 (2014). ISSN:0974–2239
3. Álvaro, P.M.: Una metaheurística para la extracción de reglas de asociación. Aplicación a Terremotos. Master's thesis, Escuela Técnica Superior de Ingeniería Informática (2012)
4. Galán Montaño, F.J.: Metodología para el análisis de terremotos de gran magnitud. Master's thesis, Universidad de Sevilla (2013)
5. Ortiz, G.: Sismo de 8.2, el más intenso en casi un siglo en méxico, excelsior. https://www.excelsior.com.mx/nacional/2017/09/08/1187054. Accessed 16 Jan 2020
6. Milenio: Los daños por el sismo: reporte del gobierno federal. http://www.milenio.com/negocios/los-danos-por-el-sismo-reporte-del-gobierno-federal. Accessed 16 Jan 2020
7. EFE. Sep: reparar 12 mil 931 escuelas tras sismos costará 13 mil 650 mdp. http://www.eluniversal.com.mx/nacion/sociedad/reparar-12-mil-931-escuelas-tras-sismos-en-mexico-costara-760-millones-de-dolares. Accessed 16 Jan 2020
8. El Universal: Los monumentos dañados por el sismo del 19 de septiembre de 2017. http://www.eluniversal.com.mx/destinos/los-monumentos-danados-por-el-sismo-del-19-de-septiembre-de-2017. Accessed 16 Jan 2020

9. Gretel, B.B.: Diseño de una Metaheurística para el minado de reglas de asociación en bases de datos transaccionales. Master's thesis, Universidad Autónoma del Estado de México, Facultad de Ingeniería (2018)
10. Zyt, J., Klosgen, W., Zytkow, J.: Handbook of Data Mining and Knowledge Discovery. Oxford University Press, New York (2002)
11. Maria, R.Q., Orallo José, J.H., Cesar, F.R.: Introducción a la Minería de Datos. Pearson Prentice Hall (2004)