Data-Driven Disaster Management in a Smart City

Sandra P. Gonçalves[1], Joao C Ferreira[1(✉)], and Ana Madureira[2]

[1] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, Lisbon, Portugal
{sandra_goncalves,jcafa}@iscte-iul.pt
[2] 2ISRC, Instituto Superior de Engenharia do Porto - Politécnico do Porto, Porto, Portugal
amd@isep.ipp.pt

Abstract. Disasters, both natural and man-made, are extreme and complex events with consequences that translate into a loss of life and/or destruction of properties. The advances in IT and Big Data analysis represent an opportunity for the development of resilient environments once the application of analytical methods allows extracting information from a significant amount of data, optimizing the decision-making processes. This research aims to apply the CRISP-DM methodology to extract information about incidents that occurred in the city of Lisbon with emphasis on occurrences that affected buildings, constituting a tool to assist in the management of the city. Through this research, it was verified that there are temporal and spatial patterns of occurrences that affected the city of Lisbon, with some types of occurrences having a higher incidence in certain periods of the year, such as floods and collapses that occur when there are high levels of precipitation. On the other hand, it was verified that the downtown area of the city is the area most affected by occurrences. Finally, machine learning models were applied to the data and the predictive model Random Forest obtained the best result with an accuracy of 58%.

Keywords: Disaster management · Data mining · Machine learning · Smart city

1 Introductions

Disasters, both natural and man-made, have been occurring more frequently around the world with damaging consequences that are reflected in the loss of human life and material/facilities damage [1]. In fact, in the last ten years, 3 751 natural disasters such as earthquakes, tsunamis, and floods were detected worldwide, representing total damages of $1 658 billion and impacting more than 2 billion people [2]. In this way, it becomes crucial to implement disaster management techniques to minimize the risks associated.

Disaster management can be characterized as a multifaceted process where the primary goals is to avoid, reduce, respond, and recover from disaster impact in the system. Due to the complexity of these events, disaster response involves different organizations such as governmental, public, and private organizations as well as different layers of authority [3]. The involvement of different entities in the disaster management processes

A. L. Martins et al. (Eds.): INTSYS 2021, LNICST 426, pp. 113–132, 2022.
https://doi.org/10.1007/978-3-030-97603-3_9

highlights the need for collaboration and coordination mechanisms since these agencies, to be effective in a disaster situation, need to communicate, coordinate, and collaborate with each other. Some factors may difficult the communication between stakeholders, such as lack of situational awareness or difficulty in adopting technological systems for disaster response since they represent high costs [4].

The increase in population density in cities and the increase in the frequency of disasters in recent years arise the need for cities to provide better services and proper infrastructures to their population. In this context, the concept of Smart City (SC) emerges, considered the ideal solution to overcome the challenges brought by globalization and urbanization [5]. Cities that aim to become a SC use digital and networked technologies to address different types of problems, such as improving the quality of services, becoming more sustainable, growing the local economy, improving the quality of life, and increasing the safety, and security of their inhabitants [6].

In a SC, electronic devices and network infrastructures are incorporated to obtain high-quality services and as cities get the latest network infrastructure, smart devices, and sensors, a substantial amount of data is generated, known as Big Data (BD). This data can contain large amounts of information that can be contextual, spatial, or temporal [7].

In the case of disaster situations, BD plays an important role in disaster management processes since it is possible to apply data mining (DM) and analysis techniques to analyze patterns and predict disasters, allowing the development of appropriate disaster management strategies from the data collected that have occurred in the past [6]. In this way, the application of BD technologies assists agents in the decision-making process, since they enable identifying potential risks and, consequently, the development of appropriate strategies to cope with disaster situations, thus increase the resilience of the SC [2].

This research aims to apply a data-driven approach to extract information about disasters in the context of a SC to contribute to improving the way the city is managed.

The objective is to perform a descriptive and predictive analysis of the data provided by the Lisbon City Hall that contains information regarding incidents that occurred in the city. This analysis is going to be performed using two different data sources: data regarding occurrences registered by firefighters between the years 2011 and 2018 both descriptive and predictive analysis are going to be carried out. The second dataset, where only a descriptive analysis of the data is going to be conducted, comes from the application "*Na Minha Rua Lx*" [8], which is an application for intervention request management in the city of Lisbon. In both cases, the analysis was conducted in two phases, where in a first moment a general analysis of the reported occurrences was carried out and in a second moment the analysis focused on occurrences that affected buildings in the city.

2 State of the Art

Data-driven disaster management is a recent area that has been undergoing an evolution due to the number of works that have been developed [9]. In this sense, a survey and critical appreciation of the literature related to the proposed theme were performed by applying the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology [10] in accordance with the Systematic Literature Review steps proposed by Okoli and Schabram [11].

Accordingly, a systematic search on the topic was conducted in two electronic databases: Scopus [12] and Google Scholar [13], and the main objective was to identify and select research papers related to data-driven disaster management research area. With this in mind, a query was formulated to make the selection of the works carried out in this area. The query is the following: ((“Disaster Management” OR “Incident Management”) AND (“smart city” OR “data analysis” OR “data mining” OR “big data” OR “machine learning”)). Additionally, a ten-year time window was defined (2010–2020), and the research covered areas such as Decision Science, Computer Science, Environmental Science, and Engineering. In terms of document typology, only journal articles, articles, and book chapters were considered. The documents were selected through the abstract and in cases where the information contained in the abstract was not sufficiently complete, the document was consulted in its entirety. The work done in this area covers both natural and man-made disasters.

2.1 Natural Disasters

Natural disasters are events that are characterized by the substantial impact they cause on society, interrupting its normal functioning. Work has been done in this area of data-driven disaster management to provide decision support systems that assist decision-makers in making decisions in a faster and more informed way, that is, based on analytical results. It was with this purpose that Jeong and Kim [14] developed a research where they conducted a statistical analysis of electrical incidents such as fires or failures occurring in Korea caused by climate changes. This study established a relationship between climate change and accidents involving electrical equipment.

Another study [15] conducted in 2017, reflected link between BD systems and disaster management. Big Data Analytics technologies was implemented on a dataset from the National Hydraulic Research Institute of Malaysia in order to analyze the hydroclimate data. The goal was to extract insights on climate change and thus provide information to prepare, mitigate, respond and recover from natural disasters. The application of BD technologies allowed the detection of periods of extreme precipitation and runoff as well as tracing drought episodes.

Through the application of DM techniques, also in 2017 another research was developed by the authors Briones-Estébanez e Ebecken [16] to identify and analyze the patterns in the occurrence of extensive and intensive events, including floods, river overflows, and landslides, related to precipitation intensity in five cities in Ecuador.

In addition to works developed to analyze disasters from a spatial and temporal perspective, other works have been developed to conduct a quantitative analysis of the damage caused by natural disasters. This is the case of the analysis carried out by the authors Alipour et al. [17]. They present a systematic framework that takes into account the different aspects that explain different types of risk (such as vulnerability and exposure) and apply Machine Learning models to predict the damage caused by flash floods in the Southeast, US.

With a similar approach, Park et al. [18] conducted a study aiming to quantify the possible effects or effectively the damage caused by three types of disasters namely typhoons, heavy rain, and earthquakes on water supply systems in Korea.

The work done in the area of data-driven disaster management is diverse as various techniques are adopted to make information available to decision-makers. In the case of the study carried out by Saha, Shekhar, and Sadhukhan [19], they presented the analytical results in more iterative way by developing a dashboard to predict and identify areas vulnerable to flooding in West Bengal, India, using geographic map visualization.

Other studies [20–23] used a combination of DM and GIS techniques to construct disaster susceptibility maps. The central objective of these studies focuses on the identification and classification of vulnerable areas to natural disasters with the difference that different DM models are used in the different research works.

2.2 Man-Made Disasters

Regarding man-made disasters, Smith et al. [24] developed a research that consisted in the implementation of Big Data technologies for disaster management. They used the statistical tool R, as well as its visualization capabilities, to analyze a dataset regarding fires that occurred in Australia.

Still in the context of fire data analysis, Balahadia et al. [25] applied the K-means clustering algorithm to generate patterns and create clusters of fire events based on the recorded data of fires that occurred in the city of Manila, Philippines. In summary, the goal was to obtain characteristics of fire events that can be used for risk assessment and risk management concerning these types of disasters as well as to assist in the development of prevention measures.

In the study of Asgary et al. [26] an attempt was made to use spatiotemporal methods to analyze the spatial and temporal patterns of fire-related incidents in Toronto, Canada. Insights were extracted by analyzing the relationship between the economic, physical, and environmental aspects of various neighborhoods and the total number of fires that occurred in those neighborhoods.

In the study of Liu et al. [27] was proposed a DM method based on using Bayesian Network to model building fires in urban areas. From the historical records of fires in a city in china between 2014 and 2016, they analyzed the potential fire risk according to building construction characteristics and external influences. Another study aiming to analyze fire patterns was conducted by Lee et al. [28] by applying the Support Vector Machine model to analyze the correlation between building characteristics, occupants, and fire incidents in Sydney.

Finally, in a study developed by Wan, Xu, He, and Wang [29] BD technologies were applied to analyze the distribution and influence factors of harmful gases in the urban underground sewage pipe network of Chongqing city, and explore the impact of smart city developments on harmful gases in the urban underground sewage pipe network.

In short, the literature review allowed to verify that most of the researches developed in this area were in China and it was also found that the research in this field covers natural disasters events as well as man-made disasters and that in the case of natural disasters there is a predominance of analysis of flood incidents and in the case of man-made disasters there is a predominance of the analysis of fire-related incidents.

3 Methodology

The analysis carried out in this research has two distinct focuses that serve the same purpose, i.e., spatial-temporal analysis of occurrences recorded in Lisbon to extract knowledge about the circumstances in which they occur. The Cross-Industry Standard Process for Data Mining (CRISP-DM) [30] methodology was applied separately on both datasets namely, the firefighters' dataset and data extracted from the application "*Na Minha Rua Lx*" to extract insights about disasters that affect the city of Lisbon with emphasis on buildings.

The analysis process based on the CRISP-DM methodology began with the business understanding that allows contextualizing and understanding the scope of the project. In this sense, an assessment of the business problem was made through the analysis of the aspects that characterize the city of Lisbon from different perspectives such as demographic, climatic, and edification aspects. After the business understanding step was completed the next phases consisted of data understanding, data preparation, modeling, and evaluation.

It is important to note that these steps of the CRISP-DM methodology, except for the business understanding, were applied to both datasets, separately, since it was not possible to merge the two datasets. The reason for the impossibility of merging the two datasets is due to the fact that there was no point of interest between the datasets, that is, a column that was common to both datasets and that had the exact same values. The join between the datasets could be made through the parish column, which is common to both, however, the datasets did not have a complete correspondence between the parishes and data could be lost when merging. The impossibility of merging the datasets impacted the prediction process since it was only possible to apply the predictive models on the firefighters' dataset, as it is richer in terms of information.

3.1 Firefighters' Dataset

The firefighter's dataset provided by the Lisbon City Hall, is a CSV file that contains information regarding the occurrences registered by the firefighters. Information covers aspects such as the description of the occurrence, date of the occurrence, location of the occurrence, i.e., latitude, longitude and address, and the human (number of persons) and material resources (number of vehicles) allocated to each occurrence. The dataset contains data from 2011 to 2018 consisting of 135 200 records (rows in the CSV file) and 22 attributes (columns in the CSV file). All the columns are of type "object", and 13 columns have some null values.

During the data preparation it was found that the years 2011 and 2012 have significantly less data than the others and, in order to perform an analysis where all years have representative data, those years where eliminated. Also, in this phase, cleansing techniques were applied that included column format conversion: the selection of relevant features/ attributes for the analysis, where attributes that did not add value to the scope of this research were eliminated. The records with null values were eliminated since it was not possible the replacement of the null values by the mean or median, because they are geographic coordinates, parishes, and descriptions of the occurrences.

This phase also included tasks such as the creation of new attributes from existing attributes and the adding of attributes from external sources such as INE [31] and IPMA [32]. External data contains information that characterizes the city of Lisbon in terms of population and building characteristics. In the last case there is information such as the average age of buildings per parish, the proportion of buildings in need of major repairs or very degraded per parish. Characterizing the city of Lisbon there are meteorological conditions such as average air temperature, relative humidity, average wind speed, and precipitation.

Lastly, it was necessary to categorize the types of occurrences that took place in buildings, since they are in a significant amount and a categorization facilitates the visual analysis. The information regarding the types of occurrences is found in the "Occurrence Description" column and this attribute has 25 types of occurrences that were defined by the firefighters' occurrence management system. These 25 types of occurrence were grouped into the following seven categories: Infrastructures – Collapse, Infrastructures – Floods, Infrastructures – Landslide, Fire, Accidents (with equipment or with elevators), Ind. technol. - Gas leak, and Ind. technol. - Suspicious situations (check smoke or check smells).

With the data preparation phase complete, the modeling phase begins. This phase is focused on extracting knowledge that can help decision-makers to manage the city in an efficient way when it comes to disaster situations. The first analysis is focused on understanding the distribution of the data over the years. It was possible to verify that in the period from 2013 to 2018 there was a downward trend in the number of occurrences recorded in the firefighter's occurrence management system, however, this decrease was not linear as there were oscillations over the years. There were 17 176 occurrences registered in the year 2013, 17 607 occurrences in 2014, 16 717 occurrences in 2015, 15 089 occurrences in 2016, 17 582 occurrences in 2017, and 13 368 occurrences in 2018.

Firefighters respond to many different types of occurrences comprising several areas of action. For a better understanding of the activities performed by firefighters, the types of occurrences were analyzed and it was verified that the distribution is not balanced among the nine categories of occurrences recorded in the dataset. There is an over-position of one category, namely the Services category, which represents 45.6% of occurrences recorded in the dataset. This category includes services such as road cleaning services, opening and closing doors, hospital transport, water supply, and prevention services at shows, sports, and patrolling.

The occurrences related to Infrastructures and Communication routes, what includes collapses, floods, landslides, falling trees and structures, and falling electric cables represents, 14.7% of the occurrences recorded in the dataset. While the occurrences related to accidents, what includes railroad accidents, road accidents, and accidents with equipment (elevators, escalators) present a proportion of 10.1%.

The categories that have the smallest representation in the dataset are Activities with 5.9%, Industrial-technological with a proportion of 5.1%, Legal conflicts with 0.5%, and civil protection events that represent 0.004% of the occurrences registered.

After a general analysis of the type of occurrences, the study is focused on the occurrences that took place in the buildings of the city of Lisbon to characterize them spatially and temporally.

As shown in Fig. 1, collapse with 3 742 records and floods with 3 356 records are the types of occurrences that most affect the buildings in the city of Lisbon, followed by occurrences related to suspicious situations that include verification of smells and smoke that count with 3 105 records. Also, with a significant proportion of incidences, but less expressive when compared with the previously mentioned categories, are accidents involving equipment or elevators with a total of 2 399 records, fires with 1 892 records, and gas leaks with 1 259 records.

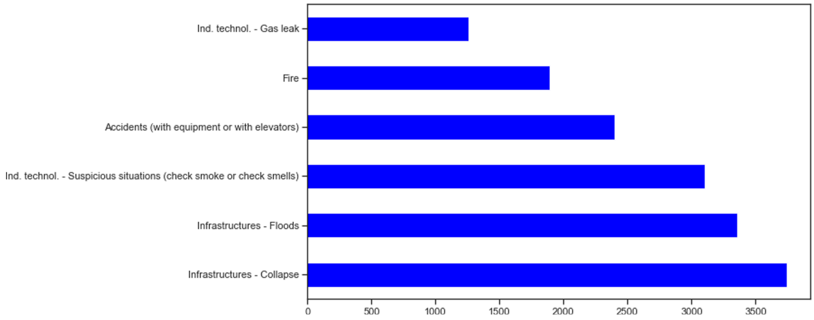

Fig. 1. Distribution of occurrences according to categories

When these occurrences are analyzed over time, i.e., their distribution over the years (Fig. 2), it is verified that there are occurrences that over the years occur in greater proportion, such as collapses, suspicious situations (checking smoke or smells), and accidents with equipment and elevators. The occurrences related to floods had a higher incidence in 2013 and 2014, with a decrease in the following years.

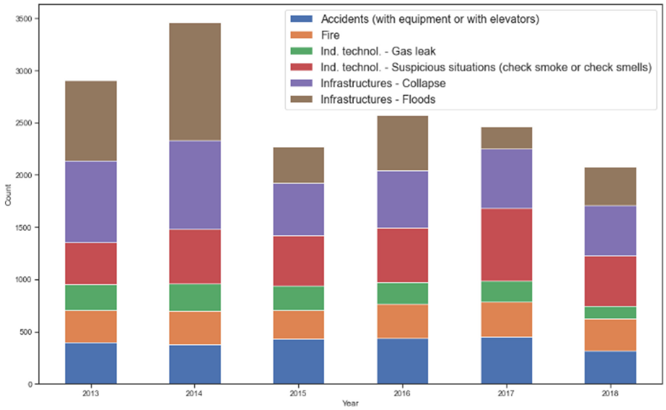

Fig. 2. Distribution of occurrences per year

Focusing the analysis on each occurrence to extract insights about its pattern of occurrence over the 12 months of the year, it is possible to verify that in the case of the occurrences referring to the infrastructure categories, i.e., collapses and floods represented in Fig. 3, that in the case of collapses (A), these occurred more frequently in the autumn and winter months, reaching maximum values (over 400 records) in the months of October and January.

As the spring and summer months approach, the number of records of this type of occurrence decreases, reaching lower values in the summer peak. Regarding floods (B), there is a higher incidence in the winter months with the highest values in the months of October to December, while during the summer months these values are much lower when compared to the winter months. Cases of suspicious situations (C) occur, similar to the types described above, more frequently in the winter months, especially in December. On the other hand, the occurrences related to gas leaks (D) show an oscillation during the months of the year, except for the month of January where values are higher than the other months, exceeding the 140 registered in this month.

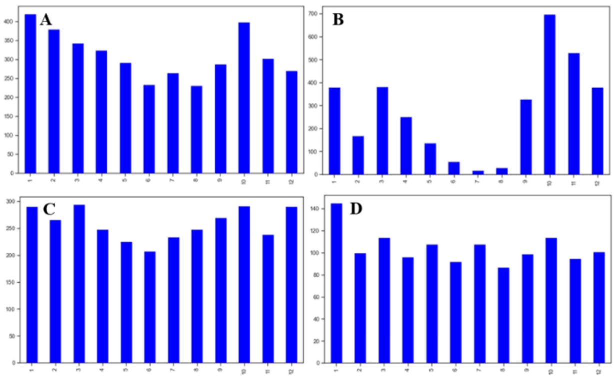

Fig. 3. The bar chart from figure **A** shows the temporal distribution of Collapses, the bar chart from figure **B** shows the temporal distribution of Floods, the bar chart from figure **C** shows the temporal distribution of Suspicious situations (check smoke or check smells), and the bar chart from figure **D** shows the temporal distribution of Gas leaks.

Lastly, the distribution of accidents involving equipment or elevators and fires is shown in Fig. 4.

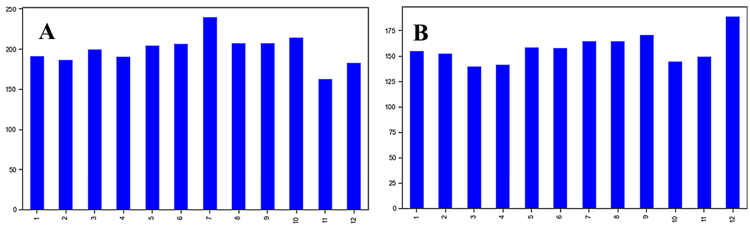

Fig. 4. Temporal distribution of the occurrences. The bar chart from figure **A** shows the temporal distribution of accidents with equipment or elevators and the bar chart from figure **B** shows the temporal distribution of Fires.

Regarding accidents with equipment or elevators (A), this type of occurrence presents an incidence with similar values throughout the months except for the month of July where there is an increase and the month of November where there is a decrease. In the case of fires (B), these occur mainly in the last month of the year, and these observations may be due to the fireplaces and candles that are used in greater density at this time of year.

The first analysis showed that there are types of occurrences that have a higher incidence in certain seasons of the year, such as collapses, floods, suspicious situations (checking for smoke or smells) occurring with a higher incidence in the winter/spring months, the influence of weather conditions on the incidence of different types of events affecting the city of Lisbon has been verified.

With this in mind, the influence of precipitation on the different types of occurrences data was analyzed through four distinct periods, namely when it does not rain, when the rain is low, when the rain is moderate, and when the rain is heavy. The creation of these four levels allows the precipitation to be classified in qualitative terms. For this purpose, an interquartile approach was adopted and from the interquartile ranges it was possible to build 4 datasets with the four precipitation levels previously mentioned.

From the analysis of occurrences according to the four precipitation levels, it was possible to conclude that there are two types of occurrences, namely floods and collapses that increase when precipitation levels increase. In the case of floods, the increase in incidence depending on precipitation levels is outstanding, since in cases where the precipitation was zero its incidence was 4.02%, in situations of low precipitation it was 19.23%, in situations of moderate precipitation it was 47.98%, and finally in situations of heavy precipitation it was 75.81%.

Shifting the focus to an analysis of occurrences from a spatial perspective to verify how occurrences are distributed throughout the city of Lisbon, heatmaps were created for the six types of occurrences that most affect buildings in the city of Lisbon. Figure 5 shows the spatial distribution of collapses and floods.

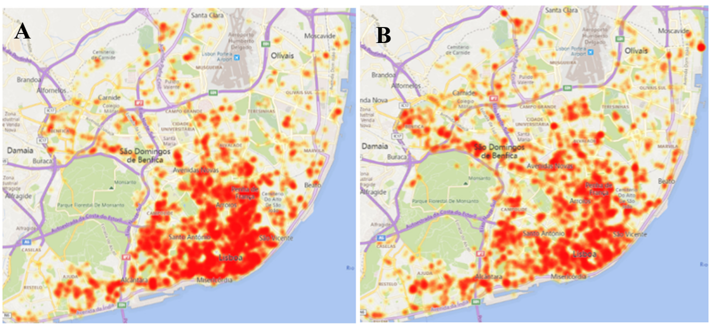

Fig. 5. Spatial distribution of the occurrences. Figure **A** shows the spatial distribution of Collapses and figure **B** shows spatial distribution of Floods,

From the heatmaps presented above it is possible to infer that the occurrences related to collapses, which is the type of events that most affects the city of Lisbon, have a higher concentration of points in the central zone of the city, which means that collapses affect mainly parishes in the central area of the city, such as *Arroios, Santo António, São Vicente, Misericórdia, Campolide, Avenidas Novas, Penha de França*, and areas of the Historical Center of Lisbon. The occurrences referring to floods, similarly to collapses, have a higher concentration in the downtown area of the city with the difference that this type of occurrence also happens with high incidence in the northwestern part of the city, namely the parishes of *Benfica* and *São Domingos de Benfica.*

Figure 6 refer to the geospatial distribution of occurrence regarding Suspicious situations, Gas Leaks, Accidents with equipment or elevators, and Fires.

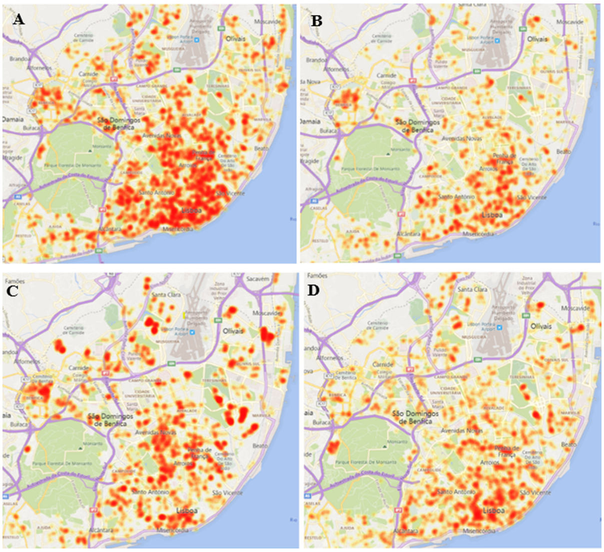

Fig. 6. Spatial distribution of the occurrences. Figure **A** shows the spatial distribution of Suspicious situations (check smoke or check smells), figure **B** shows the spatial distribution of Gas leaks, figure **C** shows the spatial distribution of accidents with equipment or elevators, and figure **D** shows the spatial distribution of Fires.

Suspicious situations present a higher concentration in the central zone of the city of Lisbon, namely the parishes of *Arroios, Santo António, São Vicente, Misericórdia, Campolide, Avenidas Novas, Penha de França* and areas in the Historical Center of Lisbon. On the other hand, situations concerning gas leaks (B) are more concentrated in the Lisbon Historical Center area and the districts of *Penha de França*, *Arroios*, and *Benfica.*

In terms of accidents involving equipment or elevators (C), this type of occurrence, unlike the types of occurrences already analyzed, does not present a higher concentration in a single Lisbon area, but instead affects the entire Lisbon city area with a similar proportion. On the other hand, although fires (D) are a type of occurrence that in general is registered in the entire Lisbon area, their concentration is slightly higher in the Lisbon historic center area.

Since there is a concentration of occurrences in a specific area of the city of Lisbon, it was sought to deepen the knowledge about Lisbon by analyzing aspects such as the state of conservation of buildings and the average age of buildings in the different parishes. Through the spatial visualization of the buildings that are degraded or in need of repair and through the visualization of the parishes where the oldest buildings are located, it is possible to establish the association between the spatial concentration of occurrences and the condition of the buildings.

Fig. 7. Figure **A** shows the spatial representation of the proportion of buildings that are degraded or in need of major repairs and figure **B** shows the spatial representation of the average age of the buildings per parish

From the conclusions reiterated from the two heatmaps created (Fig. 7), it is possible to infer that the areas where the older buildings are concentrated and where there is a greater proportion of degraded buildings or with major needs of repair are more affected by the types of occurrences such as collapses, floods, suspicious situations (check smoke or check smells), and gas leaks..

3.2 *Na Minha Rua Lx* Dataset

Regarding the dataset containing the data related to the occurrences recorded on *the Na Minha Rua Lx* application, the first analysis showed that the dataset covers the period from 2017 to 2020 and it is composed of 12 866 records and eight attributes that aggregate information about the occurrences reported in the application such as the date on which the occurrence was reported, the type of occurrence, and the location of the occurrence.

It was verified that there was one duplicate value and zero null values in the entire dataset. Furthermore, six of the eight columns that compose the dataset are of type *object*, except the columns "latitude" and "longitude" that are of type *float*.

The data preparation began with the selection of the relevant attributes and after identifying the necessary attributes to conduct this analysis, data processing techniques were applied to adequate the dataset to the analysis intended to be developed. However, no significant problems were identified, besides one attribute that was not in date format and a duplicate value. Also, in the data preparation phase, new columns were created from an existing column since the information that allows locating the occurrences on a temporal level was extracted from the column "Date-Time". In this way, three new attributes were created from the "Date-Time" attribute: "Year", "Month", and "Hour". Lastly, it was found that the year 2020 has significantly less data when compared to the other years and, to conduct an analysis where all years have representative data, it was necessary to eliminate the year 2020. With all the transformations on the dataset completed, the final dataset has seven columns and 12 865 rows.

With the data preparation phase completed, the modeling phase begins. The analysis of the data from the *Na Minha Rua Lx* application aims to deepen the knowledge about the types of occurrences reported in the application, constituting a tool to help decision-makers to manage the city in an informed and efficient way.

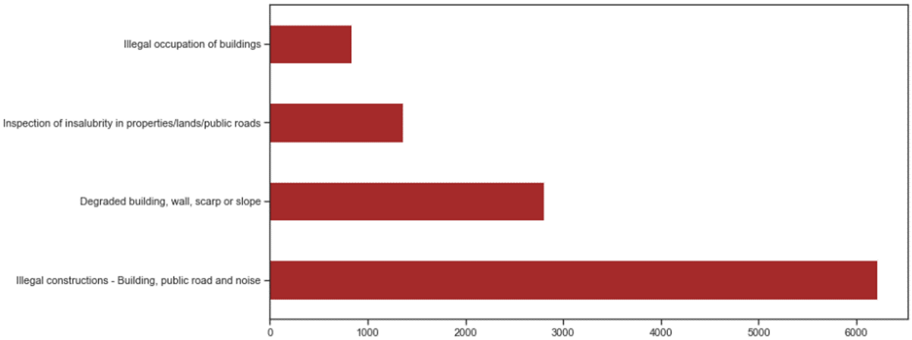

Fig. 8. Types of occurrences reported in the application

From Fig. 8 it is possible to verify that there are four main types of occurrences reported in the application, namely Illegal constructions - building, public roads and noise, Degraded building, wall, scarp or slope, Inspection of insalubrity in properties/lands/public roads, and Illegal occupation of buildings.

In terms of their distribution, it is noted that the occurrences are not equally distributed in the dataset since there is an over-position of the type of occurrence referring to Illegal constructions in relation to the other types of occurrences once, during the period under analysis, 6 217cases of Illegal constructions - Building, public roads, and noise were reported. The other typologies are in lesser proportion with 2 799 cases of Degraded building, wall, scarp or slope, 1 365 cases of Inspection of insalubrity in properties/lands/public roads, and 834 Illegal occupations of buildings.

Analyzing the distribution of these types and occurrences per year, it was concluded that every year there is a large number of reports corresponding to cases of illegal construction, while the other types of occurrences are reported less frequently.

After a general description of the distribution of reports over the years and an analysis of the types of occurrences reported, the analysis focused only on events that took place in the buildings of the city of Lisbon i.e., degraded building, wall, scarp or slope and illegal occupation of buildings.

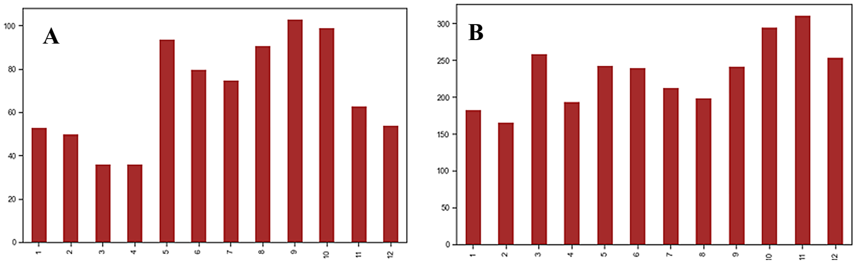

Fig. 9. Temporal distribution of the occurrences. The bar chart from figure **A** shows the temporal distribution of illegal occupation of buildings and the bar chart from figure **B** shows the temporal distribution of degraded buildings, wall, scarp, or slope.

With the purpose of deepening the knowledge about these two types of occurrence, their temporal distribution was analyzed (Fig. 9) and for cases of illegal occupation of buildings are reported in greater expression between the months of September and October, with emphasis on the month of May where there is an increase in this type of occurrence.

On the other hands, cases of degraded buildings, wall, scarp, or slope, there is an increase in reports in the last four months of the year, i.e., from September to October, and then an increase again in March and May.

Shifting the focus to an analysis of occurrences from a spatial perspective, heatmaps were created that present the geospatial distribution of the above-mentioned types of occurrences with the goal of verifying how these occurrences are distributed throughout the city.

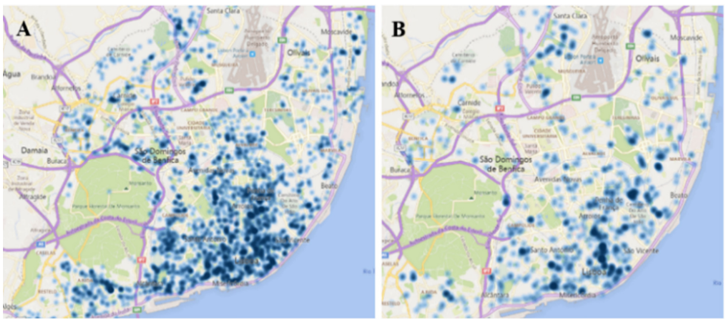

Fig. 10. Spatial distribution of the occurrences. Figure **A** shows the spatial distribution of degraded buildings, wall, scarp, or slope and figure **B** shows the spatial distribution of illegal occupation of buildings.

From the heatmaps presented in Fig. 10, it is possible to verify that in both cases these events are registered with a higher incidence of the downtown area of the city. Cases regarding Degraded building, wall, scarp or slope have higher incidences in the following parishes: *Penha de França*, *Arrois*, *Avenidas Novas*, *Misericórdia*, *Santo António, São Vicente, Ajuda, São Domindos de Benfica* and *Campolide*. On the other hand, the occurrences related to Illegal occupation of buildings have a higher incidence in the parishes of *Arroios*, *São Vicente*, *Santo António*, *Penha de França*, *Misericórdia*, and the historical downtown area.

3.3 Prediction Process

In this phase, predictive models were applied to predict disasters. Since this is a classification problem and the attribute that is intended to be predicted is a categorical attribute (“Occurrence Type”), supervised classification algorithms [68] were applied and then compared to determine the most efficient classification algorithm for this case. The following predictive models were applied: Random Forest, Decision Tree Classifier, Support Vector Machine, Gaussian Naive Bayes, and Logistic Regression.

Before proceeding to the application of the predictive models it was necessary to make a feature selection, i.e., the selection of relevant attributes for the construction of the predictive models, and the feature selection was conducted using the correlation matrix presented in Fig. 11.

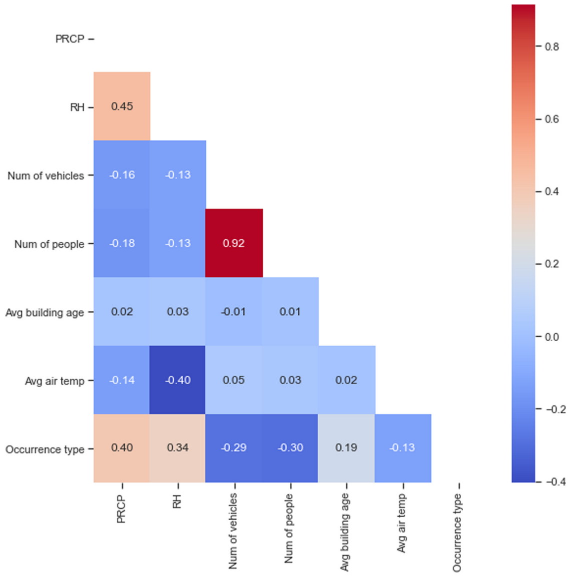

Fig. 11. Correlation matrix

Considering the data presented in the correlation matrix, only the attributes that have greater correlation with the attribute Occurrence Type were selected, since it is the target attribute (independent attribute). Thus, the following attributes were selected as dependent or explanatory attributes (dependent attributes): “PRCP” (precipitation), “RH” (relative humidity), “Num of people”, “Num of vehicles”, “Avg building age”, Avg air temp”, “Resident pop”, and Avg WS” (average wind speed).

To train the predictive models, it was necessary to split the data to allow not only the training but also the testing of the models. The training set is used to find the relationship between dependent and independent attributes, while the test set evaluates the model's performance. In numerical terms, the division was made so that 70% of the data was for training and the remaining 30% of the data was for model testing.

After defining the X and Y attributes and dividing the data for testing and training, the algorithms were applied, and the predictive results were analyzed. It is important to emphasize that all algorithms were applied twice, wherein a first moment they were applied without hyperparameters and in a second moment, with the purpose of improving the performance of these algorithms, functions that present the best set of hyperparameters were applied and the algorithms were reapplied considering the hyperparameters. The graph in Fig. 12 shows the summary of the predictive results for each model after tuning.

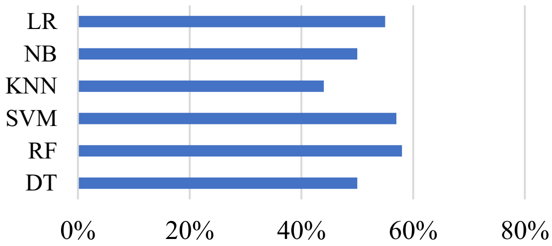

Fig. 12. Summary of the predictive results for each model

The prediction results were not satisfactory (best result 58%), and these results can be explained by the fact that there was no strong correlation between the attributes and also by the fact that the independent attribute (the one predicted) had six possible values: Infrastructures - collapses, Infrastructures - floods, Ind. Technol. -suspicious situations (check smoke or check smells), Ind. - Gas leak, Fires, and Accidents (with equipment or with elevators).

4 Evaluation

The work developed in this research was evaluated through a satisfaction questionnaire. The evaluations were carried out by 3 experts and, regarding the characterization of the panel of evaluators, they hold degrees in territorial engineering, geology and architecture and the average experience of the evaluators is 26 years.

The questionnaire aimed to evaluate five criteria, namely effectiveness, consistency with the organization's objectives, utility, level of detail, clarity, applicability, transferability, and impact. These criteria have four evaluation possibilities: NA (Not Achieved), PA (Partially Achieved), LA (Largely Achieved), and TA (Totally Achieved). The result is shown in Table 1.

The evaluation assigned to the criteria were PA and LA with the exception for the Clarity criterion which was assigned a TA classification by Evaluator 1. The reason why the criteria were classified as PA and LA is due to the fact that the stockholders responsible for this project initially expected this analysis to be conducted at the level of the buildings with a higher level of detail relating the state of conservation of each building with the events that occurred, but it was later concluded that the data did not allow such analysis and therefore the possible analysis would be from a high-level perspective.

Table 1. Results of the evaluation

Project evaluation				
Criterion	**Objective statement**	**#Eva1**	**#Eva2**	**#Eva3**
Efficacy	The research effectively informs about incidents that affect buildings in the city of Lisbon	LA	LA	LA
Consistency with the organization's objectives	The results achieved are aligned with the objectives set and correspond to the needs identified, providing relevant insights to the decision-makers	PA	PA	PA
Utility	Through the research, useful insights were extracted for the decision support process	LA	LA	LA
Detail level	The proposed solution provides the necessary level of detail to assist in the decision support process	LA	PA	PA
Clarity	The research provides clear and easy-to-understand information from the elaborate graphics	TA	LA	LA
Applicability	The solution has practical applicability in the field of disaster management and civil protection	PA	PA	PA
Transferability	The results gathered in this research can be applied to other contexts or areas	PA	LA	LA
Impact	The results achieved positively impact the way disaster situations are managed, thereby increasing the city's resilience	PA	PA	PA

5 Discussion

The analyses performed in this study allowed to verify that disaster management is a multifaceted field which encompasses several aspects. The spatial aspect permitted the analysis of the areas of the city most affected by the occurrences. This information allows the allocation of the intervention teams and placing the resources near the risk locations.

The temporal aspect allowed the verification of the incidence of the different types of occurrences throughout the year. Finally, the analysis of the human and material resources allocated to each type of occurrence allowed to identify the type of event that requires greater resources.

In this way, all this information enables the decision-makers to strategically allocate resources in order to respond to occurrences in a timely manner, since the reduction in response time reduces the impact of the events on the community.

This study allowed to conclude the historical center of the city of Lisbon is the most affected area of the city. Although the other areas are affected by the occurrences, it is the historical center area that has older and more degraded buildings, meaning they are more exposed to risk.

On the other hand, the application of the Random Forest algorithm is an asset in disaster management as it enables prediction of occurrences and helps decision makers to be better prepared to cope with the incidents.

In short, historical data on disasters that have occurred in the past is an important tool for disaster management in the present and in the future since the effective analysis of this data allows not only the extraction of knowledge about the patterns of occurrence, but also their prediction.

This research was based on data from a single city, so the conclusions are only for the city of Lisbon. However, this process can be replicated for other cities, taking into account their specific characteristics.

6 Conclusion

The research carried out shows that the evolution in the IT area has positively impacted the disaster management area since when city management policies are grounded on analytical results resulting from the application of DB technologies to the large amounts of data that have begun to be stored, it allows increasing the authorities' capabilities to cope with disaster situations.

The spatial-temporal analysis conducted in this research is important to understand the types of occurrences to which the city of Lisbon is vulnerable. The extraction of knowledge regarding the patterns of occurrence of these events is useful in the various stages of disaster management as it allows generating an overview of easy understanding among the various stakeholders, which allows the development of appropriate strategies for the various phases of disaster management.

In short, through the application of DM techniques to the firefighters' dataset, it was possible to conclude that the buildings in the city of Lisbon are affected by six types of events namely collapses, floods, suspicious situations (check smoke or check smells), gas leak, fires, accidents (with equipment or with elevators). It was also verified that there is a temporal pattern with regard to these occurrences since in some cases there is a greater predominance of certain occurrences at certain times of the year. In terms of the distribution of the occurrences, it was concluded that the historic center of the city is, in general, the area most affected by these types of occurrences and it is in this area where are concentrated the degraded buildings or with a great need for repair and also the older buildings.

On the other hand, it was verified that in the dataset with data from the application "*Na Minha Rua Lx*", the data reported in the application are not of the same type as the data registered in the firefighters' occurrence management system, since the occurrences involving buildings are Illegal occupation of buildings, Degraded building, wall, scarp, or slope. Furthermore, it was verified that the occurrences of both the firefighters and the application cover the same areas and that in both cases there is a predominance of occurrences in the historic center area of the city.

Additionally, it was expected that the events reported in the *Na Minha Lx application* could complement the firefighters' data, but results showed that the events reported in the application, despite covering the same areas, do not have events of the same extent and weight as the events recorded in the LFBR management system.

Finally, predictive algorithms were applied to the firefighters' data, however the results obtained were not satisfactory and the explanation for these results is due to the poor richness of the data since there were not very strong correlations between the dependent variables and the independent variable had six possible values.

Thus, the fact that it was not possible to join the datasets due to their characteristics, the fact that the data were not rich and consequently did not allow a more detailed analysis (as intended by the stakeholders responsible for this project) and achieve better predictive results, were the main limitations identified during this research.

References

1. Wellington, J.J., Ramesh, P.: Role of Internet of Things in disaster management. In: 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1–4 (2017). https://doi.org/10.1109/ICIIECS.2017.8275928
2. Yang, C., Su, G., Chen, J.: Using big data to enhance crisis response and disaster resilience for a smart city. In: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), pp. 504–507 (2017). https://doi.org/10.1109/ICBDA.2017.8078684
3. Zagorecki, A., Johnson, D., Ristvej, J.: Data mining and machine learning in the context of disaster and crisis management. Int. J. Emerg. Manag. **9**, 351–365 (2013). https://doi.org/10.1504/IJEM.2013.059879
4. (PDF) Crowdsourcing Disaster Response. https://www.researchgate.net/publication/268448750_Crowdsourcing_Disaster_Response. Accessed 26 Aug 2021
5. Shah, S.A., Seker, D.Z., Rathore, M.M., Hameed, S., Yahia, S.B., Draheim, D.: Towards disaster resilient smart cities: can internet of things and big data analytics be the game changers? IEEE Access **7**, 91885–91903 (2019). https://doi.org/10.1109/ACCESS.2019.2928233
6. Chaudhari, S., Bhagat, A., Tarbani, N., Pund, M.: Dynamic notifications in smart cities for disaster management. In: Computational Intelligence in Data Mining, Singapore, pp. 177–1902019). https://doi.org/10.1007/978-981-10-8055-5_17
7. Shah, S.A., Seker, D.Z., Hameed, S., Draheim, D.: The rising role of big data analytics and IoT in disaster management: recent advances, taxonomy and prospects. IEEE Access **7**, 54595–54614 (2019). https://doi.org/10.1109/ACCESS.2019.2913340
8. Na Minha Rua LX - Lisboa Inteligente. https://lisboainteligente.cm-lisboa.pt/lxi-iniciativas/na-minha-rua-lx/. Accessed 10 Aug 2021
9. Li, T., et al.: Data-driven techniques in disaster information management. ACM Comput. Surv. **50**(1), 1–45 (2018). https://doi.org/10.1145/3017678
10. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G.: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Int. J. Surg. **8**(5), 336–341 (2010). https://doi.org/10.1016/j.ijsu.2010.02.007

11. Okoli, C., Schabram, K.: A guide to conducting a systematic literature review of information systems research. SSRN Electron. J. (2010). https://doi.org/10.2139/ssrn.1954824
12. Scopus - Document search. https://www.scopus.com/search/form.uri?display=basic&edit.scft=1#basic. Accessed 09 Aug 2021
13. About Google Scholar. https://scholar.google.com/intl/en/scholar/about.html. Accessed 09 Aug 2021
14. Jeong, M.-C., Kim, J.: Prediction and analysis of electrical accidents and risk due to climate change. Int. J. Environmental Res. Pub. Health **16**(16), 2984 (2019). https://doi.org/10.3390/ijerph16162984
15. Abdullah, M.F., Ibrahim, M., Zulkifli, H.: Big data analytics framework for natural disaster management in Malaysia. In: Proceedings of the 2nd International Conference on Internet of Things, Big Data and Security, Porto, Portugal, pp. 406–411 (2017). https://doi.org/10.5220/0006367204060411
16. Briones-Estébanez, K.M., Ebecken, N.F.F.: Occurrence of emergencies and disaster analysis according to precipitation amount. Nat. Hazards **85**(3), 1437–1459 (2016). https://doi.org/10.1007/s11069-016-2635-z
17. Alipour, A., Ahmadalipour, A., Abbaszadeh, P., Moradkhani, H.: Leveraging machine learning for predicting flash flood damage in the Southeast US. Environ. Res. Lett. **15**(2), 024011 (2020). https://doi.org/10.1088/1748-9326/ab6edd
18. Park, J., et al.: Ensemble model development for the prediction of a disaster index in water treatment systems. Water Switz. **12**(11), 1–19 (2020). https://doi.org/10.3390/w12113195
19. Saha, S., Shekhar, S., Sadhukhan, S., Das, P.: An analytics dashboard visualization for flood decision support system. J. Visual. **21**(2), 295–307 (2017). https://doi.org/10.1007/s12650-017-0453-3
20. Célia, R., et al.: Mapping characteristics of at-risk population to disasters in the context of Brazilian early warning system. Int. J. Dis. Risk Reduct. **41**, 101326 (2019). https://doi.org/10.1016/j.ijdrr.2019.101326
21. Lee, S., Lee, S., Lee, M.-J., Jung, H.-S.: Spatial assessment of urban flood susceptibility using data mining and geographic information system (GIS) tools. Sustainability **10**(3), 648 (2018). https://doi.org/10.3390/su10030648
22. Liu, Y., Li, Z., Wei, B., Li, X., Fu, B.: Seismic vulnerability assessment at urban scale using data mining and GIScience technology: application to Urumqi (China). Geomat. Nat. Hazards Risk **10**(1), 958–985 (2019). https://doi.org/10.1080/19475705.2018.1524400
23. Chen, W., Zhang, S., Li, R., Shahabi, H.: Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. Sci. Total Environ. **644**, 1006–1018 (2018). https://doi.org/10.1016/j.scitotenv.2018.06.389
24. Smith, S., et al.: Adoption of data-driven decision making in fire emergency management. In: Presented at the 24th European Conference on Information Systems, ECIS 2016 (2016)
25. Balahadia, F.F., Dadiz, B.G., Ramirez, R.R., Luvett, M., Lalata, J.P., Lagman, A.C.: Application of data mining approach for profiling fire incidents reports of bureau of fire and protection. In: 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), pp. 713–717 (2019). https://doi.org/10.1109/ICCIKE47802.2019.9004420
26. Asgary, A., Ghaffari, A., Levy, J.: Spatial and temporal analyses of structural fire incidents and their causes: a case of Toronto, Canada. Fire Saf. J. **45**(1), 44–57 (2010). https://doi.org/10.1016/j.firesaf.2009.10.002
27. Liu, X., Lu, Y., Xia, Z, Li, F., Zhang, T.: A data mining method for potential fire hazard analysis of urban buildings based on Bayesian network. In: Proceedings of the 2nd International Conference on Intelligent Information Processing, New York, NY, USA (2017). pp. 1–6. https://doi.org/10.1145/3144789.3144811

28. Lee, E.W., Yeoh, G., Cook, M., Lewis, C.: Data mining on fire records of New South Wales, Sydney. Procedia Eng. **71**, 328–332 (2014). https://doi.org/10.1016/j.proeng.2014.04.047
29. Wang, Z., Xu, J., He, X., Wang, Y.: Analysis of spatiotemporal influence patterns of toxic gas monitoring concentrations in an urban drainage network based on IoT and GIS. Pattern Recognit. Lett. **138**, 237–246 (2020). https://doi.org/10.1016/j.patrec.2020.07.022
30. [PDF] Crisp-dm: towards a standard process modell for data mining|Semantic scholar. https://www.semanticscholar.org/paper/Crisp-dm%3A-towards-a-standard-process-modell-for-Wirth-Hipp/48b9293cfd4297f855867ca278f7069abc6a9c24. Accessed 27 Aug 2021
31. Portal do INE. https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_inst_legislacao&xlang=pt. Accessed 04 May 2021
32. IPMA – Serviços. https://www.ipma.pt/pt/produtoseservicos/index.jsp?page=dados.xml. Accessed 04 May 2021