International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I10P215

Volume 70 Issue 10, 144-154, October 2022

Original Article

Crime Prediction and Citizen Security Plans using Big Data in Metropolitan Lima

Rosa Perez-Siguas[1], Hernan Matta-Solis[1], Eduardo Matta-Solis[1], Anika Remuzgo Artezano[1], Diana Matta-Solis[1], Segundo Millones-Gomez[2], Roberto Cumpén-Vidaurre[3,4]

[1]*TIC Research Center: eHealth & eEducation, Instituto Peruano de Salud Familiar, Lima-Perú*
[2]*Instituto de Medicina Legal y Ciencias Forenses (IML), Lima-Perú*
[3]*Universidad Nacional Tecnológica de Lima Sur, Lima-Perú*
[4]*Universidad Nacional del Callao, Callao-Perú*

Received: 26 July 2022 Revised: 22 September 2022 Accepted: 06 October 2022 Published: 19 October 2022

Abstract *- Currently, all citizens feel very insecure due to the amount of robbery in the streets because it does not matter if you are alone or accompanied, and they will still rob you. In addition, citizens try not to go out in the streets at very late hours for fear of something happening to them. But the police are taking measures to make the country safer and prevent crime, improving citizen security. The design of the Application will be able to help prevent people who are nearby and to show the police the places that are too dangerous in real time since it will have the interconnection of police cameras so that they can detect assaults or crimes that may occur during that time and will also allow you to put together the strategies that you are going to require. In this work, the SEMMA methodology was used for the development and the Adobe XD tool for the design of the prototype. The results obtained from the investigation will be to know and predict which places there are more robberies to make better decisions.*

Keywords *– Crime Prediction, Citizen security, Big Data, SEMMA, Prototype.*

1. Introduction

Peru is one of the countries most affected by citizen insecurity due to the great amount of violence, assaults and robberies in the streets. This fact is supported by the studies carried out so far. For example, according to a report published by the Peruvian National Police (PNP), an average of approximately 15,348 complaints per month are reported [1]. These include sexual harassment on the street and in public transportation; for this reason, it is estimated that around 30% of women in Lima have been involved in sexual harassment at some point in their lives [2].

Currently, in Peru, according to the INEI, 82.3% of Peruvians were once victims of crime during the 12 months. In places with approximately 20,000 inhabitants, the population living in metropolitan Lima and the constitutional province of Callao had a criminal act that was 22.2%, and in urban population centers, 11.9%. Therefore, 12 out of every 100 inhabitants were victims of theft of money, wallet, or cell phone [1].

There are 2,830 murders, plus 137,000 robberies and 18,582 rapes committed in 2019 that the police received at the national level. Consequently, they were in Lima, Arequipa, Piura, and Lambayeque, according to statistics. In Lima, some 53,112 complaints were filed for theft and 49,924 for assault, and not only do they now steal in person, but also in a computerized way [2]. Although only 9% of the world's population lives in Latin America and the Caribbean, the region registers 1/3 of the world's homicides. It corroborates that one in 3 people in the region considers crime is a more recurrent problem, and more than 50% of the population fears being a victim of a crime [3].

Police forces are essential to prevent crime and improve citizen security. But in general, trust in police forces in the region is very low. It is due to low police quality, related to modest salaries and minimum requirements for entry into the police force. More than 70% of those surveyed affirm that if governments offered better salaries, they could hire better police [3]. Trust plays a fundamental role in determining the demand for public policies. Those who trust the media also support tougher penalties, allocate more resources to the police, and use more traditional policing strategies instead of patrolling at high-insecurity hotspots. [4].

The Big Data in the year 2020, each person produced 1.7 MB of data every second. In a study that they were able to carry out, it can be verified that 90% of corporate companies were put as the most urgent assets in the year 2020. Thanks to analytical analyzes, the use of methodologies such as Artificial intelligence and Machine learning that helps the efficiency of the processes, therefore, must have a place of establishment that helps to predict [5]. Big data not only analyzes large amounts of data but can help process in various ways as many combinations that can be applied in

decisions so that they can have a better result; that is, it allows to enhance and create new strategies to solve some problem [6].

Several years ago, in a crime prediction program based on data analytics, it was possible to reduce 27% of local public robberies and 11% of assaults for a year in California, United States. That was the reason for other countries to replicate the same strategies. In Peru, the forms of management were modernized to combat citizen insecurities. Among them are the use of patrol GPS and the increased resources to have a greater capacity for police solutions. [6].

The design of the Application will be able to help prevent people who are nearby and also to show the police the places that are too dangerous in real time since it will have the interconnection of police cameras so that they can detect assaults or crimes that may occur during that time and will also allow you to put together the strategies that you are going to require.

Big Data has very good results in various parts of the world. For example, in the case of Vancouver, after the application of Big Data at the Police level, it was possible to detect possible crimes within a radius of 150 meters with very good accuracy.

The article's objective will be to develop a mobile application to predict crimes and put together citizen security plans using Big Data.

The rest of the document is organized as follows; section II will define the Scrum methodology, the Troncoso method and the tools to be used; section III will show the case study; section IV will show the discussions; V will show the results and finally section VI will show the conclusion.

2. Literature Review

According to the author [7], the bibliography is aimed at knowing how the technologies that are being used for Big Data can be tools that are useful for crime prevention. For this reason, several studies have been explained, each with a different approach, which will test algorithms based on Big Data to predict criminal risk. In conclusion, it has been obtained that this type of technology, used respecting all ethical and legal limits, is a more effective tool for predicting the criminal incidence that will occur in a given space-time than the traditionally used methods. It shows that Big Data is a useful and necessary weapon in the fight to achieve optimal situational prevention.

According to the author [8], To face this reality, governments have been working on citizen security programs that seek to develop plans and strategies to fight crime and violence, which will strengthen social cohesion for the protection of the fundamental rights of the citizens. To that end, one of the tools that most contribute to this goal is technology, which plays a very important role since it allows automating processes in the search for operational efficiency and providing relevant information for decision-making to improve results in reducing insecurity rates. What they propose in the investigation is to determine factors that currently do not favor the adoption of new technologies that will be applied in citizen security; therefore, we have a low level of this service.

The author [9] proposes that there are other methods to have security in places; that is, I investigate the percentage of Lima society that uses electronic security products and how much they know about them. Likewise, the performance and satisfaction levels of the main services that Prosegur provides for the home are investigated. Necessary information is collected regarding what technologies and measures could help improve the electronic security system offered by Prosegur for its clients. Likewise, an analysis of the advantages offered by emerging technologies such as Machine Learning (facial and object recognition) focused on home security is carried out.

3. Methodology

3.1. SEMMA Methodology

The SAS Institute developed SEMMA, defined as selecting, exploring, and modeling large volumes of data to discover too many business patterns. The acronym for the process is Sample, Explore, Modify, Model, and Assess.

The SEMMA methodology focuses more on technical aspects, excluding analysis activities and understanding the problems being addressed. It was specially proposed to be able to work with the software of SAS companies. It is organized into nodes based on the different phases that the methodology will be composed of. [10].

Next, we will show the Phases and Activities of the methodology

3.1.1. Sample

It will be allowed to establish the extraction of a significant subset from a statistical point of view of data that will contain the relevant information and, at the same time, is too fast to manipulate

3.1.2. Explore

It will allow us to look for trends and outliers. Etc., to get an idea. In addition to being able to refine the discovery process. Statistics, including factor analysis and segmentation, will be used.

3.1.3. Model

It evaluates the reliability, and you will be able to calculate the efficiency of the models. For this task, it will often be possible to reserve a partition of the data which can be considered so far. If the Model is good, the prediction may be correct

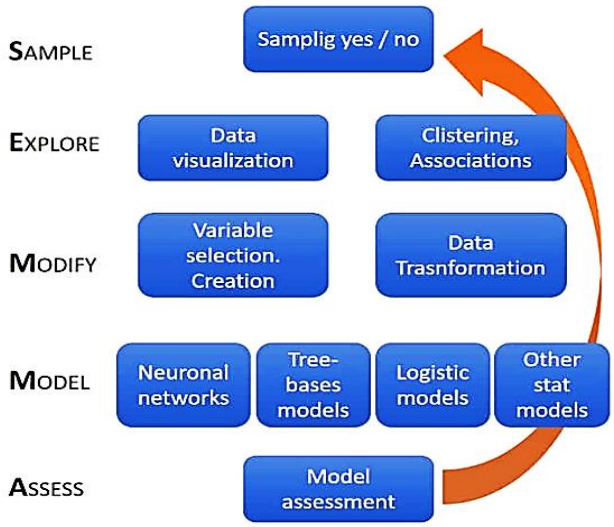


Fig. 1 Phases of the SEMMA methodology

3.2. Prototyping Tools

Adobe XD will be used in web systems designs; it is a graphic editing tool that will aid in developing design interfaces and prototypes of web applications. It allows designers to work much better so that they can navigate with a minimum of errors and less time, and one of the advantages is that it has preview tools that will allow you to see the project when it is finished. Technological tools for the application. It will also be carried out by Balsamiq, an application/service, since it has a native application for OS X and a web version, so we can work from anywhere. And its purpose is none other than to help the development of applications with a tool that facilitates the creation of diagrams [11].

3.3. Development Tools

The NoSQL database will be used for the project's development, a visual data design tool that integrates software development, database administration, database design, creation and maintenance for the NOSQL database system [8]. For the development of Data Analysis, Weka will be used; it is a software platform for machine learning and data mining (Data Mining) written in Java language [9]. In addition, Python is being used, and it is a simple language to read and write due to its high similarity with human language. In addition, it is an open-source multiplatform language and, therefore, free, which allows you to develop software without limits. Over time, Python has been gaining popularity thanks to its simplicity and extensive possibilities, especially in recent years, since it makes it easier to work with artificial intelligence, big data, machine learning and data science, among many others. [11]. For the development of project management, Team Gantt was used; it is a tool specialized in creating Gantt diagrams for the optimization of the organization of your project. With this method of project management, time and effort are saved, and the scheduled development of your projects is guaranteed.[10]. To develop the project application that shows the interface of the results of the electricity consumption proposed in the prototype, the Android Studio integrated development environment (IDE) will be used, which is designed to allow developers to create mobile applications that can take full advantage of all the resources that a mobile device can offer [11].

4. Case Study

In the research part, it will be explained in such a way that it will be detailed as it is made in the prototype for development, with the SEMMA methodology, since the entire execution will be carried out in such a way that it is accepted in the requirements to have the new prototype.

4.1. Development Stage

This investigation was carried out for the district of Metropolitan Lima in the year 2018-2019; however, the first two years of the country's situation were indeed crossed by different events of a political nature, despite that there was a lot of crime.

4.1.1. Sample

Among the demographic characteristics, the population in metropolitan Lima is 10.8 million, and the geographical area is 2819.26 km2 and a density of about 3,329 inhabitants per km2.

Metropolitan Lima comprises forty-three (43) local governments grouped together to facilitate the secretariat's technical work of supervision and control. In addition, Metropolitan Lima is identified as a Metropolitan region reinforced by the size of the population and its economic diversification. However, it is not immune to the problem of citizen insecurity, such as the persistence of various criminal acts, which are generated by a series of social risk factors that encourage criminal behavior [22]. In Figure 2, the heat map of metropolitan Lima can be shown showing where there is more crime; if it is red, it is more dangerous, and if the color decreases, it is less dangerous.

In the semester of analysis, at the urban national level, the main criminal acts that the population perceives that they may be victims of are: Theft of money, wallet, cell phone and theft of their home Compared to the same semester of the previous year (November 2019 – April 2020), the perception of being a victim of theft of money, wallet, cell phone decreased 3.3 percentage points as shown in Figure 3 [13].

Fig. 2 Integrated System of Statistics on crime and citizen security

Mobile semester	Perception of insecurity	
	Area or neighbourhood without surveillance	Area or neighbourhood with surveillance
Half-year indicators		
Oct 2019-Mar 2020	**89.3**	**87.6**
Nov 2019-Apr 2020	**88.7**	**85.1**
Dec 2019-May 2020	**88.5**	**84.3**
Jan 2020-Jun 2020	**88.0**	**83.4**
Feb 2020-Jul 2020	**86.9**	**82.4**
Mar 2020-Aug 2020	**85.6**	**82.0**
Apr 2020-Sep 2020	**85.5**	**81.8**
May 2020-Oct 2020	**85.8**	**83.7**
Jun 2020-Nov 2020	**86.0**	**84.3**
Jul 2020 - Dec 2020	**86.4**	**84.9**
Aug 2020-Jan 2021	**86.8**	**85.6**
Sep 2020-Feb 2021	**87.4**	**85.7**
Oct 2020-Mar 2021	**87.1**	**86.0**
Nov 2020-Apr 2021	**86.9**	**85.4**
The difference with the previous moving semester (percentage points)		
Oct 2020-Mar 2021/ Nov 2020-Apr 2021	**-0.2**	**-0.6**
The difference with a similar semester of the previous year (percentage points)		
Nov 2019-Apr 2020/ Nov 2020-Apr 2021	**-1.8**	**0.3**

Fig. 3 Citizen Security Statistics

4.1.2. Explore

The final variables used for the investigation will be shown with the respective conceptual and operational definitions as Visualized in Table I, which will be a strategy.

Table 1. Variables

Final Variables				
Variable	Conceptual	Variable	Dimensions	Indicators
Predictive System	Technological tool of autonomous learning that serves for the denunciations of criminal patterns through an algorithm that predicts the possible locations of illicit with time ranges	Independent	Complaints	False accusations
				Police Interventions
				Number of complaints about property crimes
				Time, Place and Zone
				Offender Type
			Predictive Algorithms	Victim Type
			Connectivity with population	PNP 105-Lima Metropolitan Alert
Risk of Crime	Criminal incidence of a patrimonial type (Robbery, Theft, Assault) in Metropolitan Lima	Dependent	Exposition	Number of criminal acts and vulnerable areas
			Fragility	Physical conditions of a community
			Resilience	Adapt and change, Resist and recover
				Assimilate the impact of danger or threat

Table 2. Exponential Smoothing

	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7
Demand	85.8	86	86.4	86.8	87.4	87.1	86.9
Prediction	85.8	85.8	85.85	85.9875	86.190625	86.4929688	86.6447266

It is aimed at the people of Lima Metropolitan:

- Guarantee, maintain and restore the internal order of the jurisdiction
- Provide protection for the community
- Prevent, investigate and combat crime

4.1.3. Modify

The algorithm will proceed with the distribution of police load that will need an objective function based on the crime rate, which will be calculated for the crime information. To obtain the indices, it will estimate the future crimes that will remember the past crimes, that is, the old data up to the most current.

The smoothing factor will be used, which is a method used in real life to determine the crimes that are considered or not due to the oldest since it is the most intuitive way [14]. In the tool, when registering a crime, the geographical location is obtained, allowing identification of which block it corresponds to. That is, each block is going to be grouped by each one of the groups of crimes. The algorithm's configuration through the tool will determine how many months ago it was considered for predicting crimes.

This method was performed with the data in Figure 3, using the Smoothing Factor (s) of 0.25, as shown in Table 2. It was calculated recursively since it is defined based on the function:

Forecast (Month 2) =Demand (Month 1) *s + Forecast (Month 1) *(1-s)

4.1.4. Model

After pre-processing the map of Metropolitan Lima and the crime data, it can be integrated with the Python code. It will optimize the distribution of vehicles as well as those of the Police and Serenazgo; when distributing the patrol cars, the criterion will be first to distribute those of the same police and then those of Serenazgo so that it has a better organization. For the algorithm, as mentioned before, Jupyter will be used, which will process the data for the prediction whose objective function will be to minimize the distance between the district police station and the area affected by crimes and maximize that this area has the highest crime rate, that is, that it has the highest probability of some crime occurring as shown in Figure 4.

```
def leer_data():
    data = pd.read_csv('../Data/carpetas-de-investigacion-pgj-de-la-ciudad-de-Lima.csv', sep=';')
    return data
```

```
def transformar_data():
    data.drop(columns=['ao_hechos', 'mes_hechos', 'calle_hechos2', 'geopoint', 'ao_inicio', 'mes_inicio', 'fecha_inicio'], inplac
    data['fecha_hechos'] = pd.to_datetime(data.fecha_hechos, errors='coerce')
    data.longitud.astype('float', inplace=True)
    data.latitud.astype('float', inplace=True)
    data.dropna(inplace=True)
    return data
```

```
def get_clima():
    os.chdir("../Data")
    frequency = 1
    start_date = '02-JAN-2015'
    end_date = '30-AUG-2020'
    api_key = '28f7f02aa28d4afe9dc215223190509'
    location_list = ['peru_city']
    hist_weather_data = retrieve_hist_data(api_key, location_list, start_date, end_date, frequency, location_label = False, expor
    clima = pd.read_csv('../Data/peru_city.csv')
    clima.drop(columns=['maxtempC', 'mintempC', 'totalSnow_cm', 'sunHour', 'uvIndex.1', 'moonrise', 'moonset', 'sunrise', 'sunset
    clima.columns = ['fecha_hechos', 'uv', 'ilu_luna', 'punto_rocio', 'temp_sentir', 'nubosidad', 'humedad', 'precipitacion', 'pr
    return clima
```

```
def get_colonia_delito():
    roma_n = data[(data.alcaldia_hechos == 'Lima') & (data.colonia_hechos == 'ROMA NORTE')]
    roma_n = roma_n[(roma_n.categoria_delito == 'ROBO A TRANSEUNTE EN VÍA PÚBLICA CON Y SIN VIOLENCIA') |
              (roma_n.delito == 'ROBO A TRANSEUNTE DE CELULAR SIN VIOLENCIA') |
              (roma_n.delito == 'ROBO A TRANSEUNTE DE CELULAR CON VIOLENCIA') |
              (roma_n.delito == 'ROBO A TRANSEUNTE SALIENDO DEL BANCO CON VIOLENCIA')]
    roma_n.sort_values(by='fecha_hechos', ascending=True, inplace=True)
    roma_n = pd.merge_asof(roma_n, clima, on='fecha_hechos')
    roma_n['nombre_dia'] = roma_n.fecha_hechos.dt.weekday
    festivos = ['01-01', '01-05', '01-06', '02-05', '02-04', '03-21', '05-01', '05-05', '09-15', '09-16', '10-31', '11-01', '11-0
    años = ['2015', '2016', '2017', '2018', '2019','2020']
    festivo = [i+'-'+x for i in años for x in festivos]
    festivo = pd.DataFrame(festivo, columns=['dia'])
    festivo['dia_festivo'] = 1
    festivo['dia'] = pd.to_datetime(festivo.dia).dt.date
    roma_n['dia'] = roma_n.fecha_hechos.dt.date
    roma_n = pd.merge(roma_n, festivo, how='left', on='dia')
    roma_n.dia_festivo.fillna(0, inplace=True)
    roma_n['dia_festivo'] = roma_n.dia_festivo.astype('int', inplace=True)
    roma_n.drop(columns='dia', inplace=True)
    return roma_n
```

```
def mapear():
    os.chdir("../Images")
    mapa=folium.Map(location=[19.443056, -99.144444], zoom_start=15)
    for index, row in roma_n.iterrows():
        folium.CircleMarker([row['latitud'], row['longitud']], radius=1, fill_color="#3db7e4").add_to(mapa)
    geo = roma_n[['latitud', 'longitud']].as_matrix()
    mapa.add_children(plugins.HeatMap(geo, radius=15))
    mapa.save('mapa.html')
```

```
def preparar_prediccion():
    prueba = roma_n.copy()
    prueba.drop(columns=['fiscalia', 'agencia', 'unidad_investigacion', 'alcaldia_hechos', 'punto_rocio', 'colonia_hechos', 'call
    X = prueba.drop(columns=['fecha_hechos', 'latitud', 'longitud'])
    y = prueba[['fecha_hechos', 'latitud', 'longitud']]
    label = LabelEncoder()
    X.delito = label.fit_transform(X.delito)
    X.categoria_delito = label.fit_transform(X.categoria_delito)
    X.calle_hechos = label.fit_transform(X.calle_hechos)
    X['año'] = y.fecha_hechos.dt.year
    process = StandardScaler()
    X = process.fit_transform(X)
    y['dia'] = y.fecha_hechos.dt.day
    y['mese'] = y.fecha_hechos.dt.month
    y['hora'] = y.fecha_hechos.dt.hour
    y['minuto'] = y.fecha_hechos.dt.minute
    y.drop(columns='fecha_hechos', inplace=True)
    return X, y
```

```
def prediccion():
    etr = ExtraTreesRegressor(n_estimators=750, max_depth=400, random_state=1)
    etr.fit(X, y)
    pre = etr.predict(X)
    prediction = pd.DataFrame(pre, columns=['latitud', 'longitud', 'day', 'month', 'hour', 'minute'])
    prediction[['day', 'month', 'hour', 'minute']] = prediction[['day', 'month', 'hour', 'minute']].round().astype('int')
    prediction['year'] = '2020'
    prediction['fecha_hechos'] = pd.to_datetime(prediction[['year', 'day', 'month', 'hour', 'minute']], errors='coerce')
    prediction.drop(columns=['year', 'day', 'month', 'hour', 'minute'], inplace=True)
    prediction.sort_values(by='fecha_hechos', inplace=True)
    return prediction
```

```
def mapear_prediccion():
    os.chdir("../Images")
    prediction['weight'] = [i for i in range(len(prediction))]
    mapa_final = folium.Map(location=[19.443056, -99.144444], zoom_start=15)
    geo = [[[row['latitud'], row['longitud']] for index, row in prediction[prediction['weight'] == i].iterrows()] for i in range
    index = ['{:%Y-%m-%d %H-%M-%S}'.format(i) for i in prediction.fecha_hechos]
    hm = plugins.HeatMapWithTime(geo, index=index, radius=20, auto_play=True, max_opacity=0.8, name='Robo a transeunte')
    hm.add_to(mapa_final)
    mapa_final.save('mapa_final.html')
```

Fig. 4 Crime Prediction Code

In addition, the mobile application prototype was made with Balsamiq's help so that users can enter through a login and register to the application, as shown in Figure 5 and Figure 6.

For Fig. 7, the module was made where the user can add where he is and where he will go. Depending on it, it will show him the route and see if there is danger or not, as shown in figure 8; that is if it shows the red color is because there was a higher incidence in that place or it will be shaded green if it is less dangerous.

For Fig. 9, you can search for places where the police stations are located, in case you want to resort to making a complaint and in figure 10, you can show the description that there is in the police station, for example, the number of patrol cars or serenade that are in the nearest.

In addition, in Figure 11, you can see in real-time the place of the serenade, where they are, and what license plate they have.

Fig. 12 is a module where you could add if an incident occurred in the place, and you can also access the contact number of the commissioner of the nearest police station.

4.1.5. Assess

Obtaining a minimum of 60% probability of success in the predictions, seeing that more than 70% probability of success of the created Model was obtained, we can say that this evaluation is satisfactory since the established minimum probability was exceeded.

Fig. 5 Login Module

Fig. 6 Registration Module

Fig 7. Search for user location

Fig 8. Search for police stations

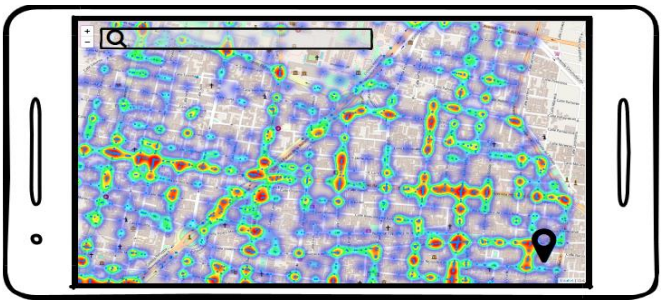


Fig. 9 Places with more danger

Fig 10. Patrol Place

Have a detail of the areas with the most common crimes with their probability of occurrence, the most common crimes for the predictions were obtained according to the frequencies of each one found in the historical data. Additionally, the predictions show the estimated probability of occurrence.

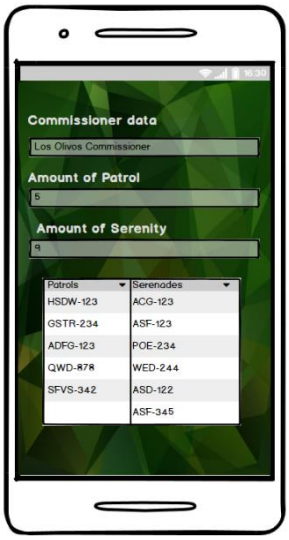


Fig. 11 Serenade details

Fig. 12 Description

5. Results and Discussion

Next, the expected results are shown with respect to the development of the article's research within the case study and the methodology by considering the proposals being developed and the implementation of the control system to carry out later the analysis that is carried out. In comparison with other investigations, of which they present similarities or inequalities in the development of the chosen methodology, considering the phases that intervene in it and in its way of presenting the execution.

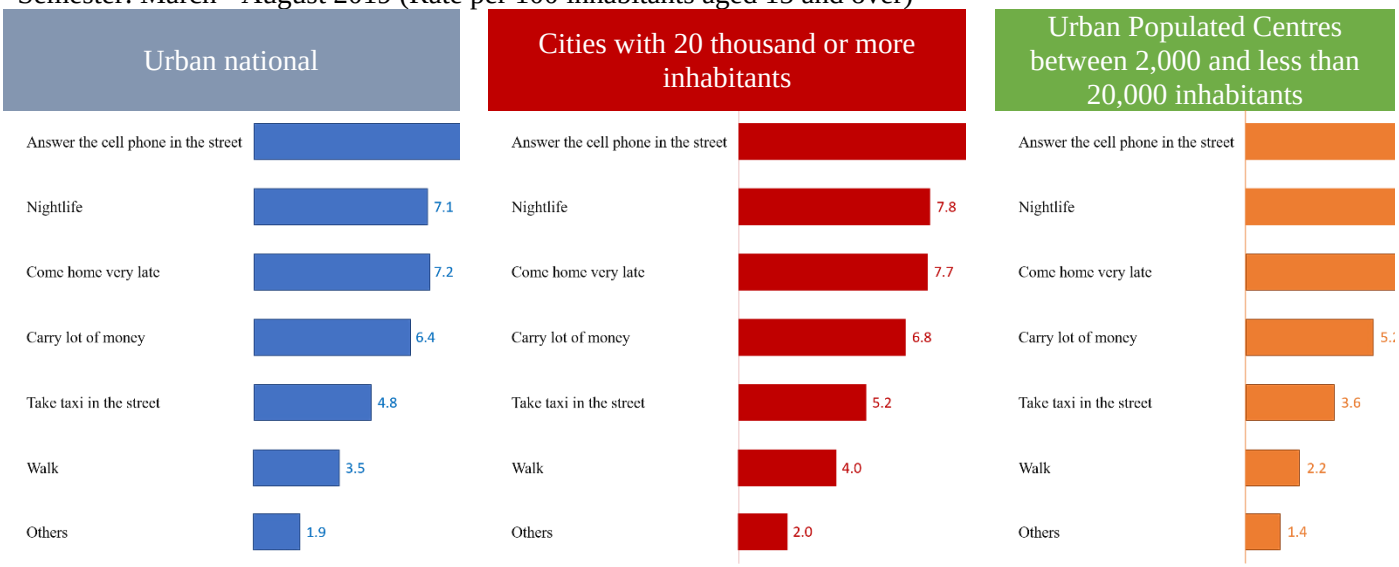
5.1. About the Case Study

One of the many ways crime can affect citizens, in addition to economic losses and the risk to their physical integrity, is by altering their daily activities. These are changes that people introduce in their daily habits, such as stopping going out at night, avoiding walks, stopping answering their cell phones in the street or taking taxis in the street to be less vulnerable to criminals.

23.3% of the inhabitants of the country's cities with 20,000 or more inhabitants claimed to have stopped carrying out any activity in the last 12 months to protect themselves from crime, according to a report by the National Institute of Statistics and Informatics (INEI) that collects information from the period March-August 2019 as shown in figure 13 [19].

But the use of technology by the police decreased in January and December 2019. There was a 16% decrease in the number of criminal events committed in the Cercado de Lima, reported the Municipality of Lima, based on the statistics integrated with the National Police of Peru (PNP), as shown in Figure 14 [20].

Rate of population aged 15 and over in the urban area, who stopped carrying out any activity to protect themselves from crime, in the last twelve months, by type of activity, according to field of study
Semester: March - August 2019 (Rate per 100 inhabitants aged 15 and over)



1/ Includes: Avoid carrying a wallet, stop wearing jewellery, avoid taking a motorcycle taxi, constantly changing routes, avoid carrying a cell phone when you go out, etc.

Source: National Institute of Statistics and Informatics - National Survey of Budget Programs 2019 (preliminary information).

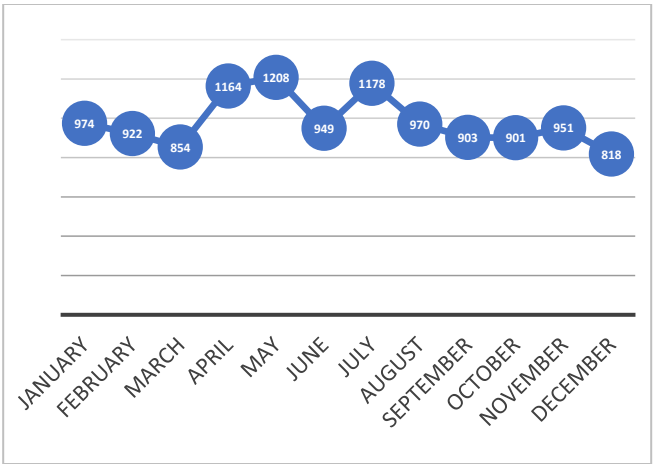


Fig. 14 Statistics of the improvement of delinquency

5.2. About the methodology

Several methodologies guide the data mining process; they intend to facilitate the execution of new projects with similar characteristics, optimize their planning and management, reduce their complexity, and allow them to be better monitored.

CRISP-DM can be integrated with a specific project management methodology that complements administrative and technical tasks; it is also freely distributed, at no cost, unlike SEMMA (SAS, 2009b). CRISP-DM defines a structure for data mining projects and guides their execution. It consists of a reference model and a user guide (Chapman et al., 2000). The reference model gives an overview of the life cycle of a data mining project and contains the phases with their objectives, the tasks and the relationships between them, and the step-by-step instructions that must be carried out. The phases defined by the reference model are business understanding, data analysis, data preparation, modeling, evaluation and deployment. Each of these phases (level 1) is made up of generic tasks (level 2), which are divided into specific tasks (level 3). Finally, at level 4, there is the process instance, which describes the specific activities to perform in a data mining project. The user guide provides detailed tips, hints for each phase, and each operation within a phase, and exemplifies how to do a data mining project. This user guide is an excellent option for developers with little experience developing this type of project. [21].

Table 3. Difference between Traditional Methodology and Agile Methodology

CRISP-DM	SEMMA
It is oriented toward the development of the MD process	Oriented to business goals
It starts with analyzing the data.	It starts by analyzing the business objectives
Arrival at SAS products	Open and free methodology
Free Access	Oriented to a project management methodology
	Maturity and clear and simplified phases

Next, the differences between SEMA with CRISP-DM will be shown as presented in table III

6. Conclusion

In conclusion, the IOT prototype was designed to prevent robberies in young areas of North Lima. With the help of tools such as Marvel App and Balsamiq, each interface and module of the application was successfully designed to show its operation and application during a robbery. On the other hand, the main functionalities were implemented following the guidelines of the user stories.

In addition, thanks to the Scrum methodology, it was possible to structure the information in a way that is understandable to all readers; in the same way, following the methodology, an order in the development of the project activities is achieved, assigning a role to each team member, thus strengthening the harmonious development of the activities.

Future work

The experience analyzed on the issue has shown that in order to combat crime and counteract delinquency, if the active participation of citizens is not counted on, the actions that are undertaken fail. Therefore, public policies must not only be framed by existing legal provisions and regulations, but to achieve more efficient and effective results, it is necessary to lead the entire system towards the construction of a collective and intelligent consciousness integrated into a single unit between the administration and the social group that leads to understanding as theirs all the problems that are circumscribed, from those that are evidenced in their original source due to the little coexistence that is observed between members of the household, neighborhood, neighborhood, town and city, to understanding the social conflict of each context and urban environment. It is to articulate and integrate the values of life and citizenship with urban spaces in a sustainable and friendly way.

The system can generate a positive impact in reducing exposure to danger through the information/recommendation provided via alerts, as well as the feeling of security by viewing the integrated patrol.

References

[1] F. A. P. Quiroz, "INEI: Only 15.5% of the Victims of a Criminal Act File a Complaint," [Online]. Available: https://elperuano.pe/noticia/123704-inei-solo-el-155-de-las-victimas-de-un-hecho-delictivo-hacen-la-denuncia

[2] Anonymous, "A Total of 2,803 Murders, More than 137 Thousand Robberies and 18,582 Sexual Violations were Committed in 2019," [Online]. Available: https://peru21.pe/peru/peru-un-total-de-2803-asesinatos-mas-de-137-mil-robos-y-18582l-violaciones-sexuales-se-cometieron-en-el-2019-noticia/?ref=p21r

[3] S. Perez, "Fight Crime and Trust to Fight Crime You Have to Increase Trust," [Online]. Available: https://blogs.iadb.org/seguridad-ciudadana/es/para-combatir-el-delito-hay-que-incrementar-la-confianza/

[4] F. G. Cafferata and C. Scartascini, "How to Fight Crime in Latin America," pp. 68, 2021.

[5] "Big Data and Business Intelligence are Emerging as the Main Engines of Digital Transformation," [Online]. Available: https://directivosygerentes.es/innovacion/como-abordar-un-proyecto-data-intelligence

[6] N. Marquez, "How to Fight Crime with Big Data?," [Online]. Available: https://www.tuataratech.com/2016/08/como-combatir-el-crimen-con-big-data.html

[7] M. Á. J. Hernández, "Final Degree Project Big Data as a Prevention Tool," 2021.

[8] M. J. Baldárrago Torreblanca, A. L. Durán Huamantinco, O. I. Lévano Muñante, and D. M. Silva Cumpa, "Factors that Prevent the Adoption of Emerging Technologies to Improve Citizen Security," 2017.

[9] Z. Milagros and H. Julian, "Implementation of an Electronic Security Management System with Machine Learning Aimed at Prosegur Peru for Security Management in Homes in Metropolitan Lima," *Technological University of Peru*, pp. 1–47, 2020, [Online]. Available:https://repositorio.utp.edu.pe/bitstream/handle/20.500.12867/2842/ZulyHuaman_TrabajodeInvestigacion_Bachiller_2020.pdf?sequence=1&isAllowed=y.

[10] I. Juan, M. Moine, D. S. Gordillo, D. Ana, and S. Haedo, "Data Mining Projects," *Comparative Analysis of Methodologies for Data Mining Project Management*, pp. 931–938, 2011.

[11] Anonymous, "Python: What it is and Why you Should Learn to Use it," 2020. [Online]. Available: https://www.becas-santander.com/es/blog/python-que-es.html

[12] Satvik Shukla, Hari Purnapatre, Gitalee Jadhav, Rasika Gohokar, "Improving Women's Safety by Accelerating Spatio Temporal Crime Prediction of Hotspots," *International Journal of Computer Trends and Technology*, vol. 68, no. 12, pp. 34-39, 2020. Crossref, 10.14445/22312803/IJCTT-V68I12P108

[13] Anonymous, "What is Weka and what does it have to do with Big Data?," [Online]. Available: https://agenciab12.pe/noticia/que-es-weka-que-tiene-que-ver-big-data

[14] Anonymous, "TeamGantt: Project Management and Creation of Gantt Charts," [Online]. Available: https://www.appvizer.es/organizacion-planificacion/gestion-proyectos/teamgantt

[15] Anonymous, "Android Studio," [Online]. Available: https://www.ecured.cu/Android_Studio

[16] R. D. E. S. Ciudadana and D. E. L. Metropolitana, "Proposal for a Regional Action Plan for Citizen Security in Metropolitan Lima," 2022. [Online]. Available: https://www.munlima.gob.pe/images/PROPUESTA-PLAN-DE-ACCION-REGIONAL-DE-SEGURIDAD-CIUDADANA-2022.pdf

[17] National Institute of Statistics and Informatics, "Citizen Security Statistics," 2021. [Online]. Available: https://www.inei.gob.pe/biblioteca-virtual/boletines/estadisticas-de-seguridad-ciudadana/1/

[18] V. Gutierrez, "Police Load Distribution System through Crime Prediction," 2018.

[19] Anonymous, "Crime Alters the Activities of Almost a Quarter of the Inhabitants of the Cities of Peru," [Online]. Available: https://gestion.pe/peru/la-delincuencia-altera-actividades-de-casi-la-cuarta-parte-de-los-habitantes-de-las-ciudades-del-peru-noticia/

[20] Anonymous, "Lima: Criminal Acts in the Cercado were Reduced 16% in 2019." [Online]. Available: https://andina.pe/agencia/noticia-lima-actos-delictivos-el-cercado-se-redujeron-16-2019-785102.aspx

[21] C. Cobos, J. Zuñiga, J. Guarin, E. León, and M. Mendoza, "CMIN - A CRISP-DM-based Case Tool for Supporting Data Mining Projects | CMIN - Case Tool based on CRISP-DM to Support Data Mining Projects," 2010.

[22] Anonymous, "What is MySQL Workbench? Concept and Characteristics Title," [Online]. Available: http://xtmiporti.blogspot.com/2016/04/que-es-my-sql-workbench-concepto-y.html