

Discovering the Needs of People in the 10/40 Window with Data Science

Harvey Alférez, Ph.D.
Merari González, MSc.

Global Software Lab
School of Engineering and Technology
Montemorelos University



PROBLEM STATEMENT

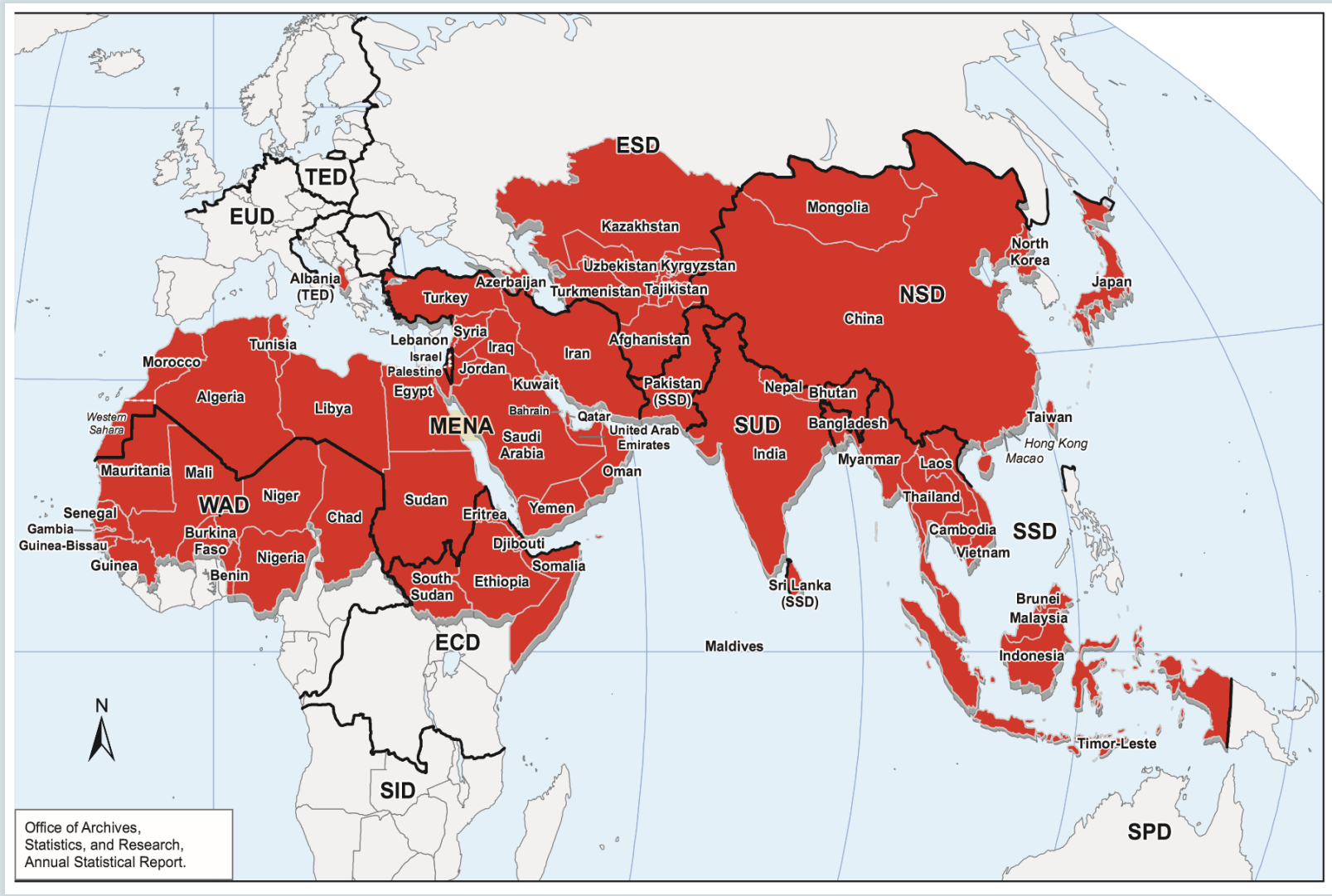


Figure 1. The countries of the 10/40 Window [1]

1. Office of Archives, Statistics, and Research (2017). *2017 Annual Statistical Report*. URL: <http://documents.adventistarchives.org/Statistics/ASR/ASR2017.pdf>

PROBLEM STATEMENT

The 10/40 Window [2]:

- This is where a majority of the world's population lives
- More than 90 percent of the “least-evangelized” people on earth live in the 10/40 Window
- This is where many of the world's most spiritually receptive people live
- The majority of the world's poorest people—some 80 percent—live here

1. Office of Archives, Statistics, and Research (2017). *2017 annual statistical report*. URL: <http://documents.adventistarchives.org/Statistics/ASR/ASR2017.pdf>

2. Gustin, P. (2006). *The 10/40 Window: new misión opportunities*.

URL: http://circle.adventist.org/files/CD2008/CD2/dialogue/articles/12_2_gustin_s.htm

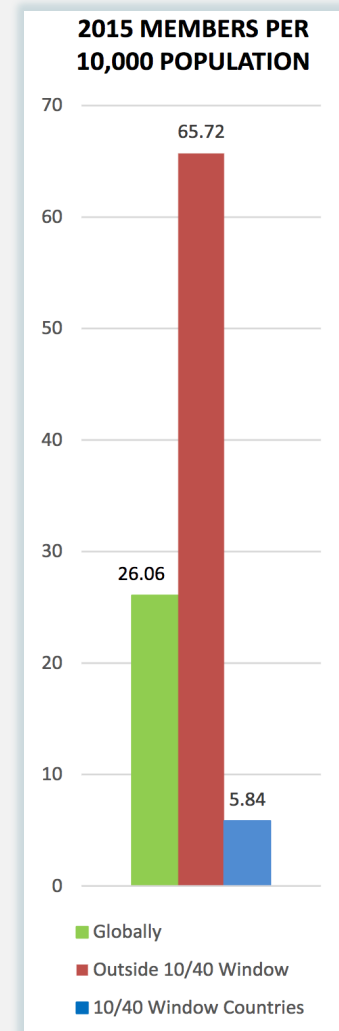


Figure 2. 2015 members per 10,000 population [1]

PROBLEM STATEMENT

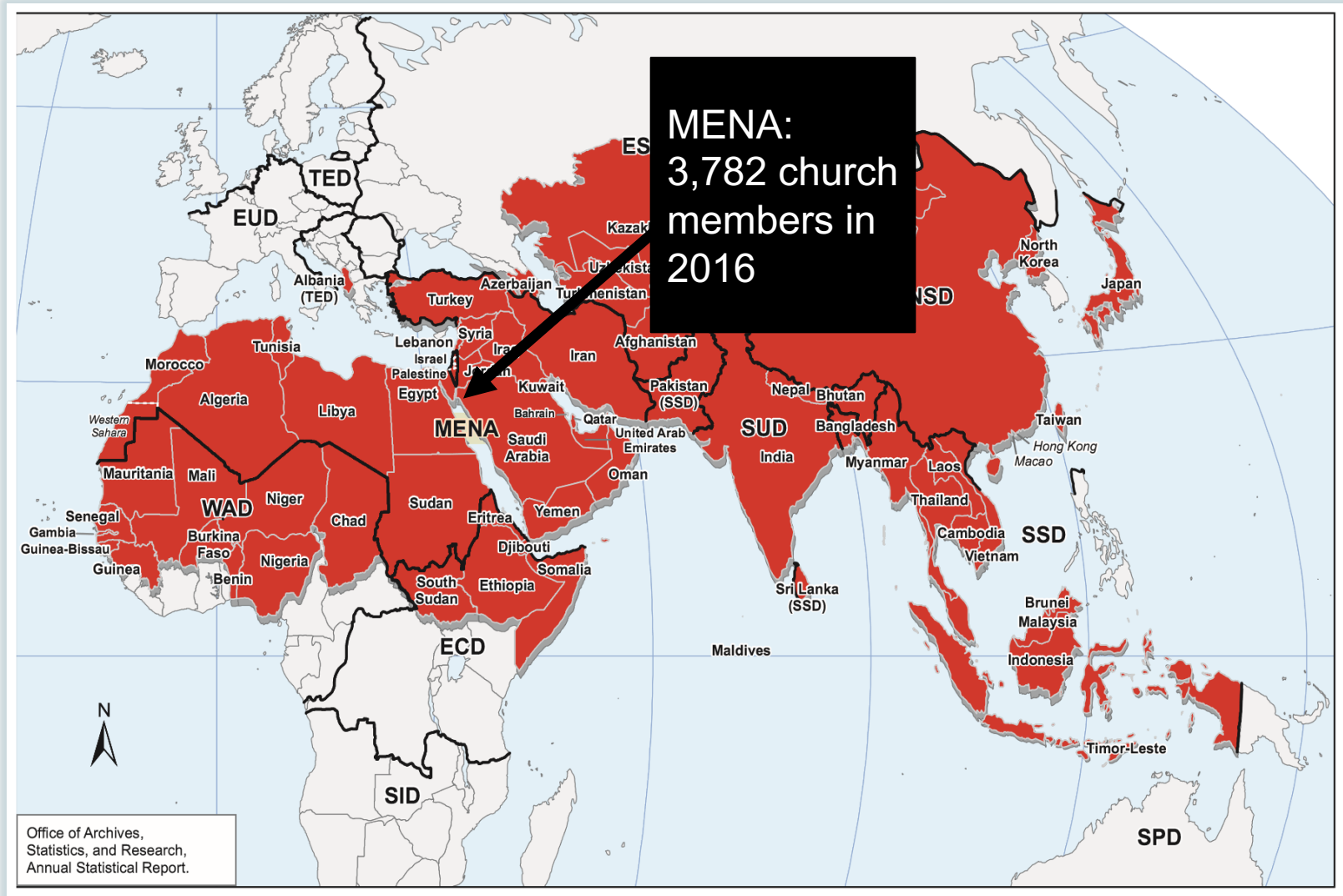


Figure 3. Middle East and North African Union in the 10/40 Window [1]

1. Office of Archives, Statistics, and Research (2017). *2017 annual statistical report*. URL: <http://documents.adventistarchives.org/Statistics/ASR/ASR2017.pdf>

PROBLEM STATEMENT

Iraq: % net membership growth over the last 10 years: -45% [1]



Figure 4. Middle East and North African Union [3]

1. Office of Archives, Statistics, and Research (2017). *2017 annual statistical report*. URL: <http://documents.adventistarchives.org/Statistics/ASR/ASR2017.pdf>
3. Seventh-day Adventist World Church (n.d.). *Middle East and North Africa*. URL: <https://www.adventist.org/en/world-church/middle-east-and-north-africa/>

PROBLEM STATEMENT

Matthew 24:14: “And this gospel of the kingdom will be preached in the whole world as a testimony to all nations, and then the end will come.” NIV

How?

PROBLEM STATEMENT

“the Savior mingled with men as one who desired their good. He showed His sympathy for them, ministered to their needs, and won their confidence. Then He bade them, ‘Follow Me.’” [4]

4. White, E.G. (1905). *The Ministry of Healing*. Review & Herald, 143.

PROBLEM STATEMENT

- How can we discover those needs?
- Do we have the data to discover those needs?
- What about the places with insufficient official data?

PROBLEM STATEMENT

Iraq is not in the list of countries that offer official open data in the Middle East and North Africa Region [5].

5. Open Data Barometer. (2017). *Middle East and North Africa*. URL: <http://opendatabarometer.org/4thedition/regional-snapshot/middle-east-north-africa/>

OBJECTIVE

Build a **software application** to **discover the needs of people in Iraq**

with **data science** applied to **open big data**
in terms of **refugee crises, humanitarian aid,**
violent protests, artillery fights and **mass**
killings

UNDERPINNINGS OF OUR APPROACH

Data Science: The study of the generalizable extraction of knowledge from data [6]

Big Data: Data sets so large and complex that they become difficult to work with using standard techniques [7]

6. Dhar, V. (2013), Data science and prediction, Commun. ACM , 56 (12), 64-73.

7. Snijders, C., Matzat, U. and Reips, U. D. (2014), 'Big data': big gaps of knowledge in the field of internet science, International Journal of Internet Science 7 (1): 1-5.

UNDERPINNINGS OF OUR APPROACH

Machine Learning: Tries to give computer systems the ability to “learn” with data, without being explicitly programmed [9]

Open Data: Data that can be freely used, re-used and redistributed by anyone [8]

8. Open Knowledge International (n.d.), *What is open data?*. URL: <http://opendatahandbook.org/guide/en/what-is-open-data/>

9. Samuel, A. L. (1959), Some studies in machine learning using the game of checkers, *IBM Journal of Research and Development* 7 (1): 1-5.

UNDERPINNINGS OF OUR APPROACH

The Global Database of Events, Language, and Tone (GDELT) Project: An open platform that monitors the world's broadcast, print, and web news from nearly every corner of every country in over 100 languages: <https://www.gdeltproject.org/>

- Over a quarter-billion event records in over 300 categories covering the entire world from 1979 to present
- The data files use Conflict and Mediation Event Observations (CAMEO) coding for recording events
- AfricaNews, Agence France Presse, Associated Press, Associated Press Online, Associated Press Worldstream, BBC Monitoring, Christian Science Monitor, Facts on File, Foreign Broadcast Information Service, The New York Times, United Press International and The Washington Post

RELATED WORK

- Su, Y., Lan, Z., Lin, Y., Comfort, L. y Joshi, J. (2016, November). **Tracking disaster response and relief efforts following the 2015 Nepal earthquake.** *2nd International Conference on Collaboration and Internet Computing (CIC)*, Pittsburgh, PA, USA.
- Qiao, F., y Wang, H. (2015, October). **Computational approach to detecting and predicting occupy protest events.** *International Conference on Identification, Information, and Knowledge in the Internet of Thing*, Beijing, China.
- Bi, S., Gao, J., Wang, Y. y Cao, Y. (2015). **A contrast of the degree of activity among the three major powers, USA, China, and Russia: Insights from media reports.** Presented at the International Conference on Behavioral, Economic and Socio-Cultural Computing, Nanjing, China.

METHODOLOGY

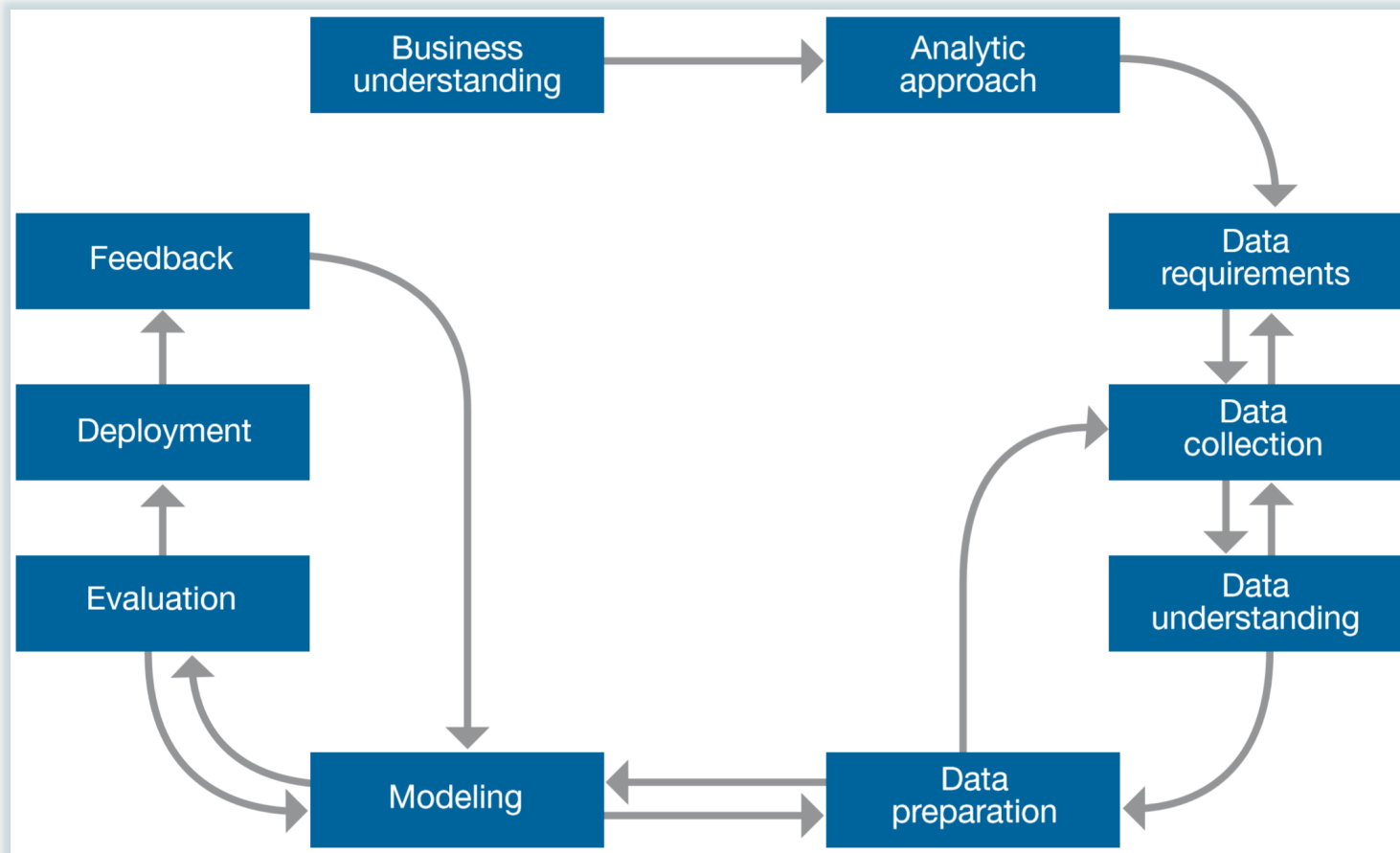


Figure 5. IBM Foundational Methodology for Data Science [10]

10. Rollins, J. B. (2015). *IBM foundational methodology for data science*. URL: <https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IMW14824USEN>

METHODOLOGY

1. Problem Understanding

Lack of official open data in **Iraq**

2. Analytic Approach

Machine learning was applied to understand and analyze the data from **GDELT**

METHODOLOGY

2. Analytic Approach

Table 1. Event description. Information retrieved from [11]

CAMEO Code	Name	Description
REF	Refugees	Refugees: also refers to agencies or multi-national corporations dealing with population migration and relocation issues
073	Humanitarian aid	Extend, provide humanitarian aid, mainly in the form of emergency assistance
145	Protest violently, riots	Protest forcefully, in a potentially destructive manner
194	Fight with artillery and tanks	Attack using artillery, tanks, and rocket fire.
202	Engage in mass killings	Kill a substantial number of people, typically with the intention of ridding a territory of a particular group of people.

3. Data requirements

Table 2. Description of the variables used in the query. Information retrieved from [12]

Code	Description	Data Type
ActionGeo_CountryCode	Country code.	String
Actor1Geo_Lat	This is the centroid latitude of the landmark for mapping.	Float
Actor1Geo_Long	Longitud del actor.	Float
Actor1Type1Code	This can be a specific role such as Police Forces, Government, Military, Political Opposition, Rebels, etc., a broad role class such as Education, Elites, Media, Refugees, or organizational classes like Non-Governmental Movement	String
EventCode	Action performed by the actor	String
Year	Year of the event	Integer

METHODOLOGY

4. Data Collection

```
1  Refugiados
2  SELECT Actor1Type1Code,Year, ActionGeo_CountryCode, Actor1Geo_Lat,Actor1Geo_Long, EventCode
3
4  FROM
5  [gdelt-bq:full.events]
6
7  WHERE Actor1Type1Code="REF"
8
9  AND (Year > 2011 AND Year < 2016)
10
11 AND (Actor1Geo_Lat > 29.12 AND Actor1Geo_Lat < 37.29)
12 AND (Actor1Geo_Long > 39.22 AND Actor1Geo_Long < 48.48)
13
14 AND Actor1Geo_Lat IS NOT NULL AND Actor1Geo_Long IS NOT NULL
15
```

Figure 6. Query in BigQuery to get events about refugees in Iraq from 2012 to 2015

4. Data Collection

Table 3. Records returned by BigQuery

CAMEO Code	Description	2010-2016
REF	Refugees	13,476
073	Humanitarian aid	10,414
145	Protest violently, riots	3,068
194	Fight with artillery and tanks	13,247
202	Engage in mass killings	1,822
	Total	42,027

METHODOLOGY

5. Data Understanding

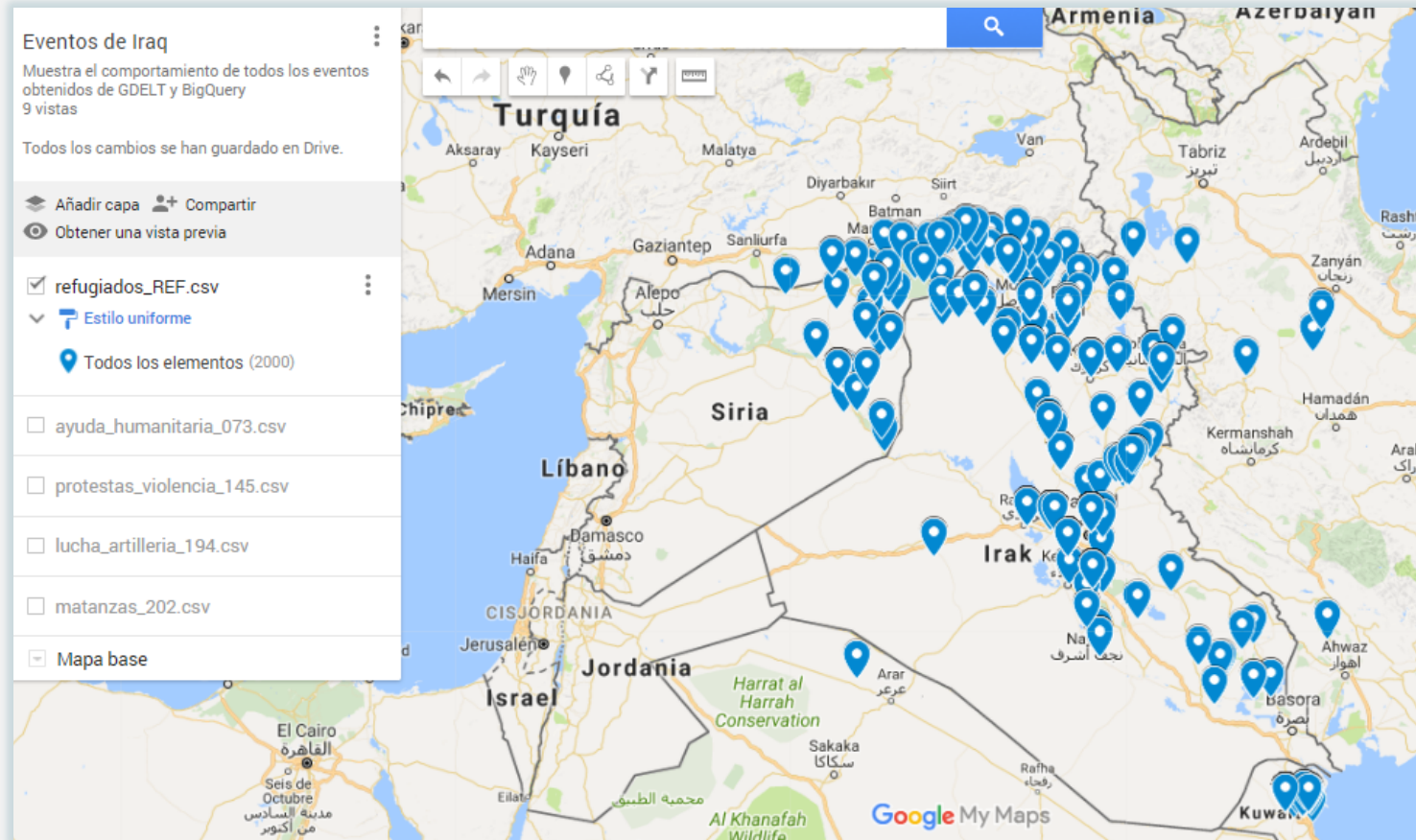


Figure 7. Map for refugee crises generated with GDELT data (2012-2015)

METHODOLOGY

5. Data Understanding

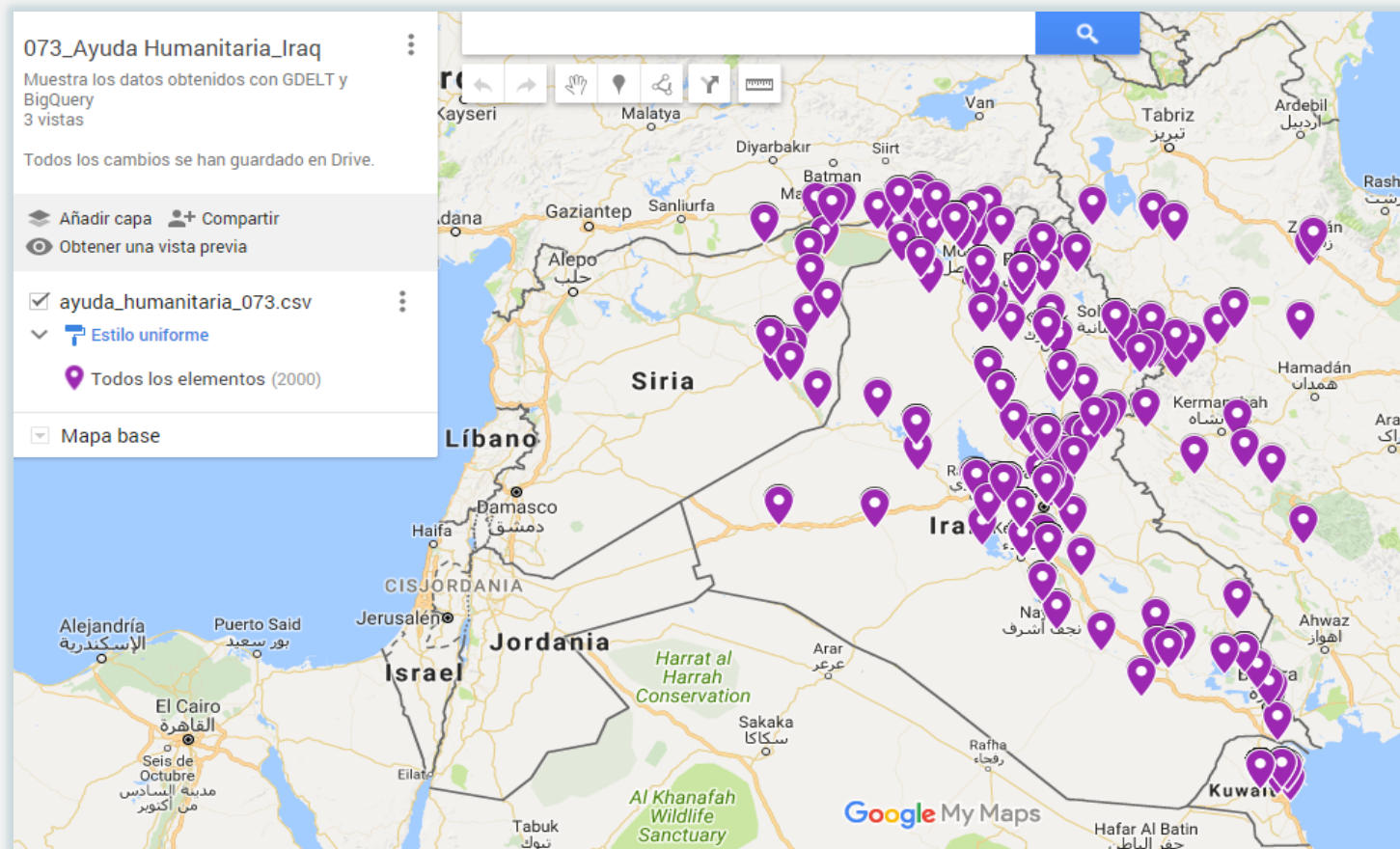


Figure 8. Map for humanitarian aid generated with GDELT data (2012-2015)

METHODOLOGY

5. Data Understanding

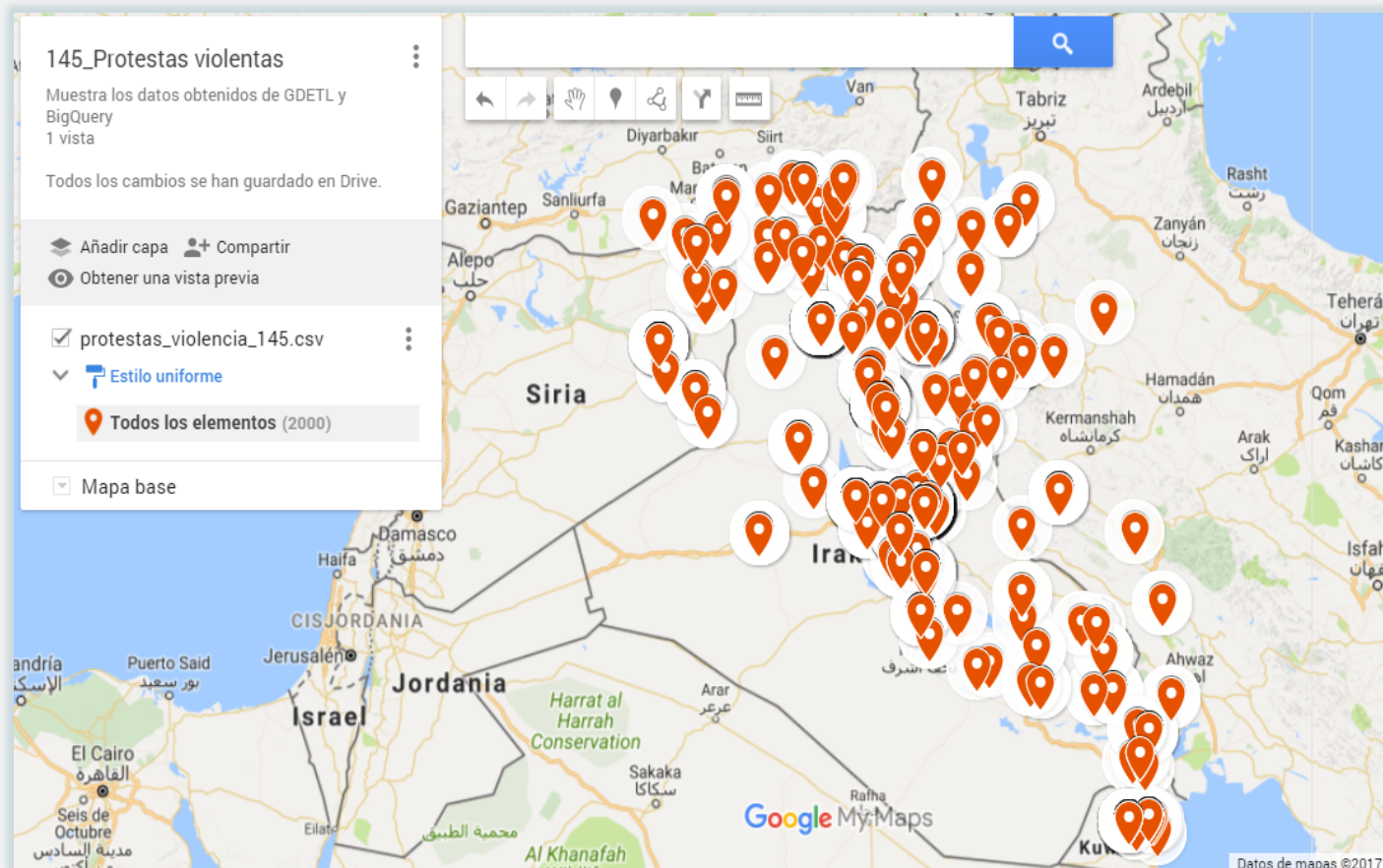


Figure 9. Map for violent protests, riots generated with GDELT data (2012-2015)

5. Data Understanding

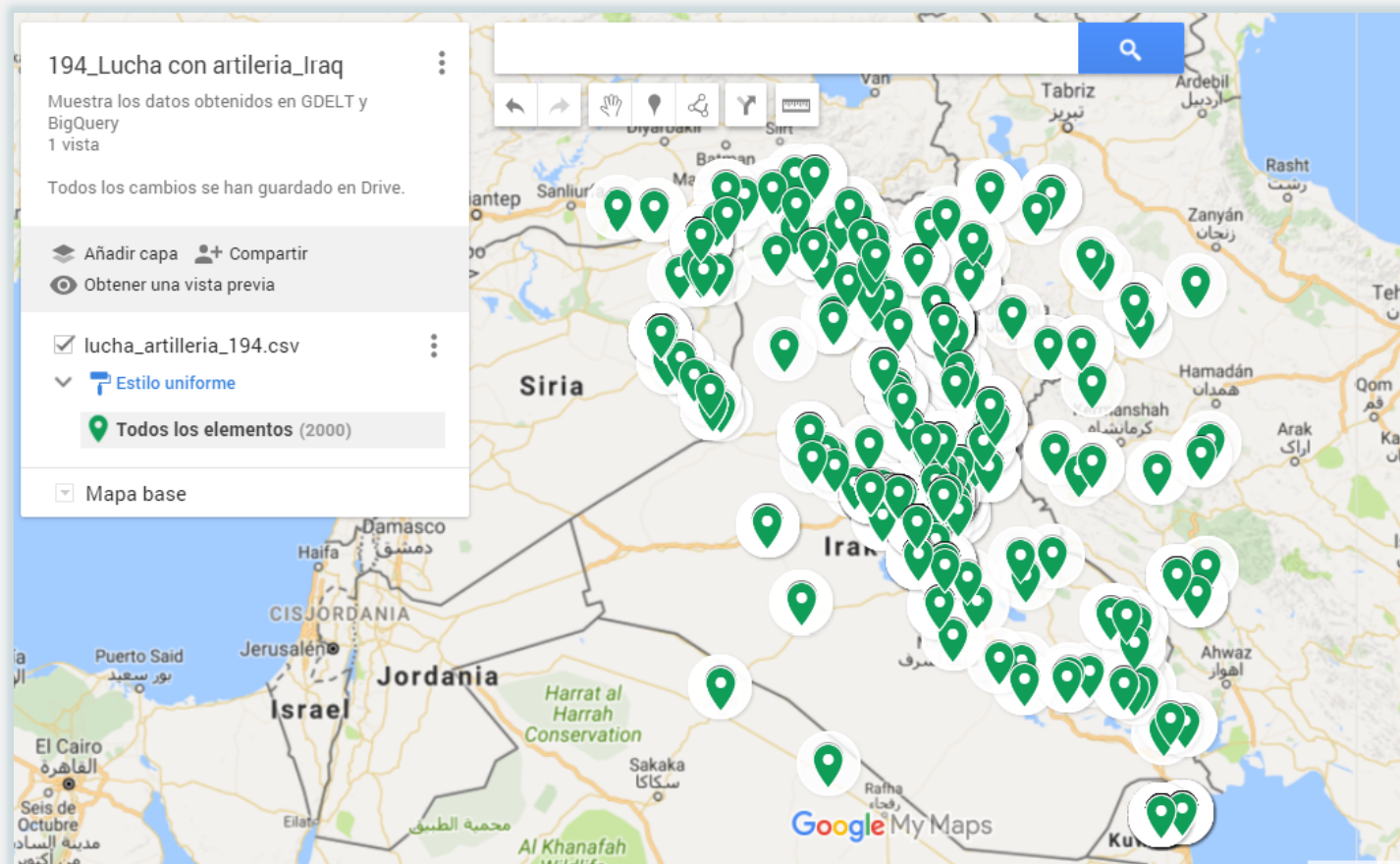


Figure 10. Map for fight with artillery and tanks generated with GDELT data (2012-2015)

METHODOLOGY

5. Data Understanding

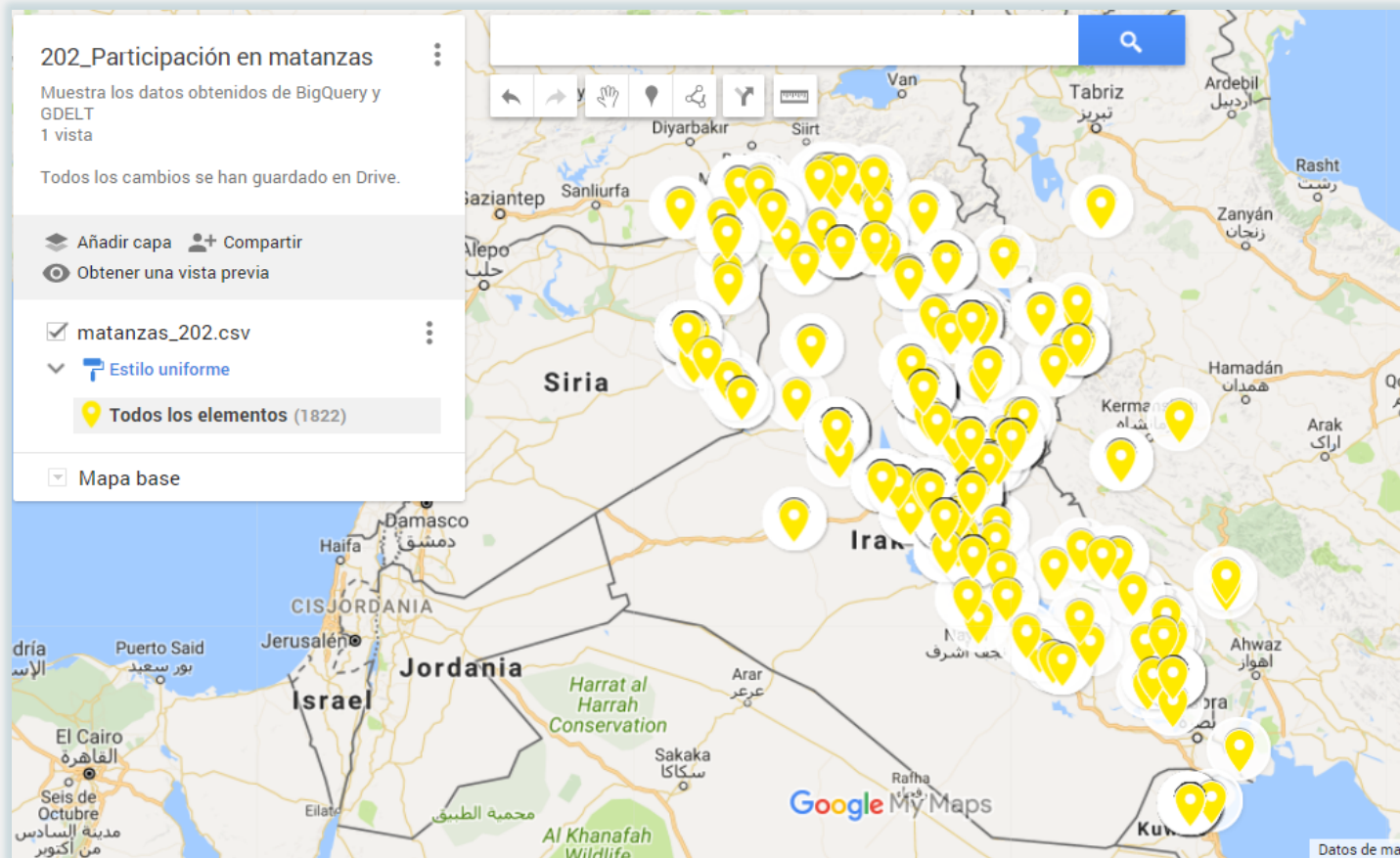


Figure 11. Map for mass killings generated with GDELT data (2012-2015)

METHODOLOGY

6. Data Preparation

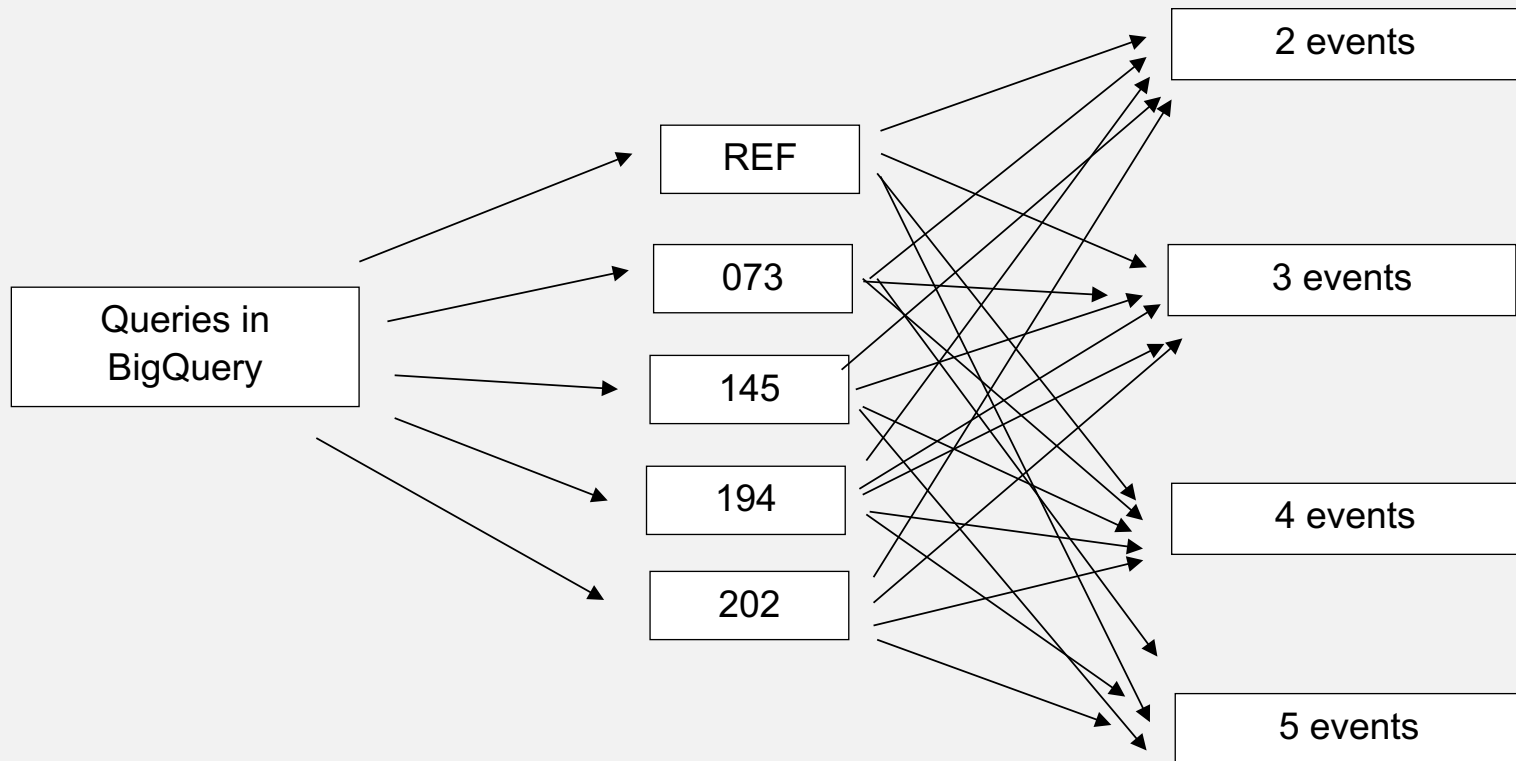


Figure 12. Data preparation

7. Modeling

KNN, Näive Bayes, Decision Trees and Logistic Regression.

- Among the best classifiers [13]

13. Wu, X. et al. (2008). Top 10 algorithms in data mining. Knowledge and Information Systems, 14(1), 1-37

EVALUATION

8. Evaluation

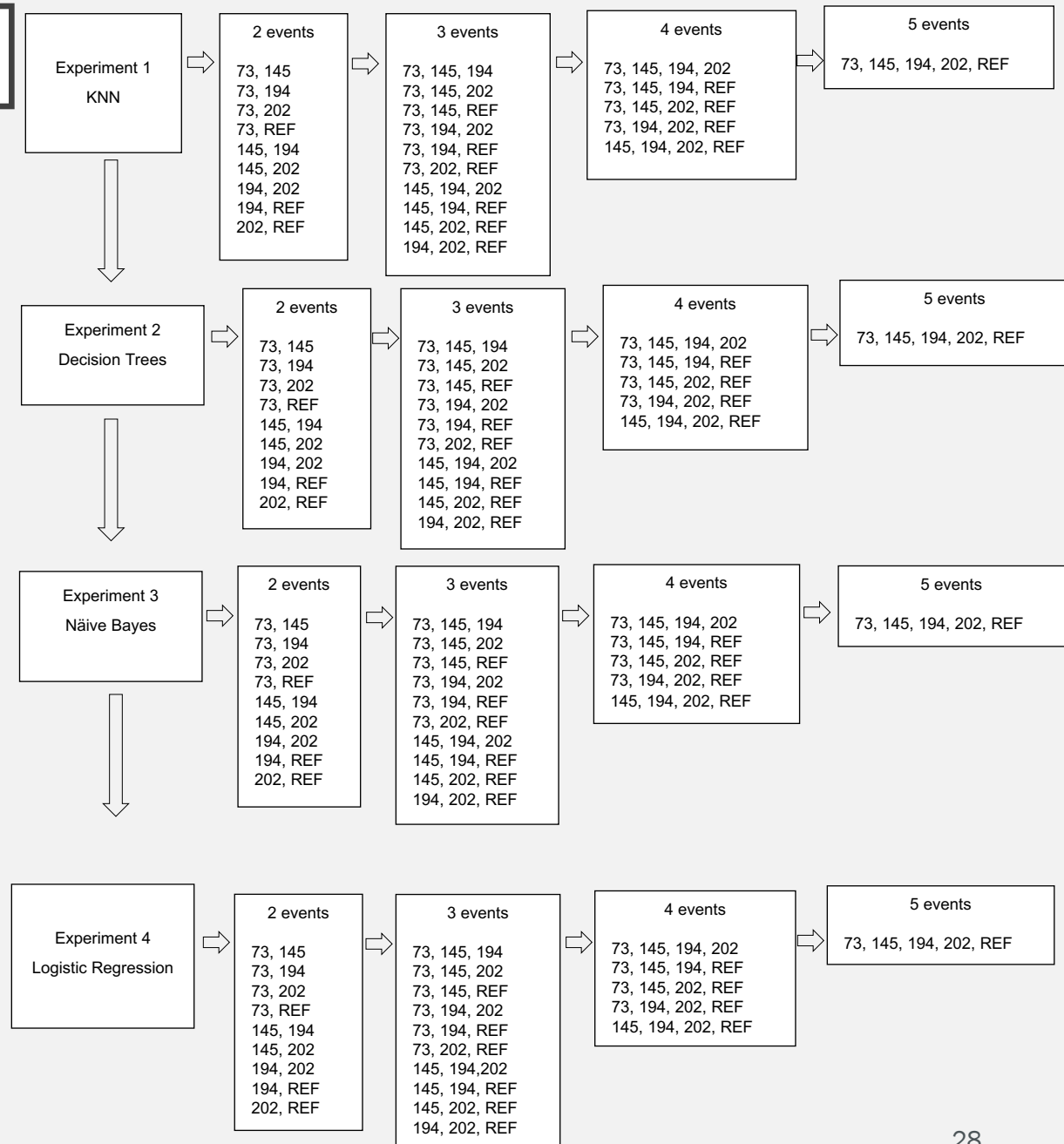


Figure 13. Workflow of the experiments

METHODOLOGY

8. Evaluation

Table 4. Results with Naïve Bayes

Accuracy: 0.8510

Event	Precision	Recall	F1-Score
Humanitarian aid	0.85	1.00	0.92
Engage in mass killings	0.00	0.00	0.00

Table 5. Results with Logistic Regression

Accuracy: 0.8896

Event	Precision	Recall	F1-Score
Fight with artillery and tanks	0.89	1.00	0.94
Engage in mass killings	0.00	0.00	0.00

METHODOLOGY

8. Evaluation

Table 6. Results with KNN

Accuracy: 0.7545

Event	Precision	Recall	F1-Score
Refugees	0.75	0.75	0.75
Fight with artillery and tanks	0.74	0.74	0.74

Table 7. Results with Decision Trees

Accuracy: 0.7629

Event	Precision	Recall	F1-Score
Refugees	0.76	0.76	0.76
Fight with artillery and tanks	0.74	0.75	0.75

9. Deployment

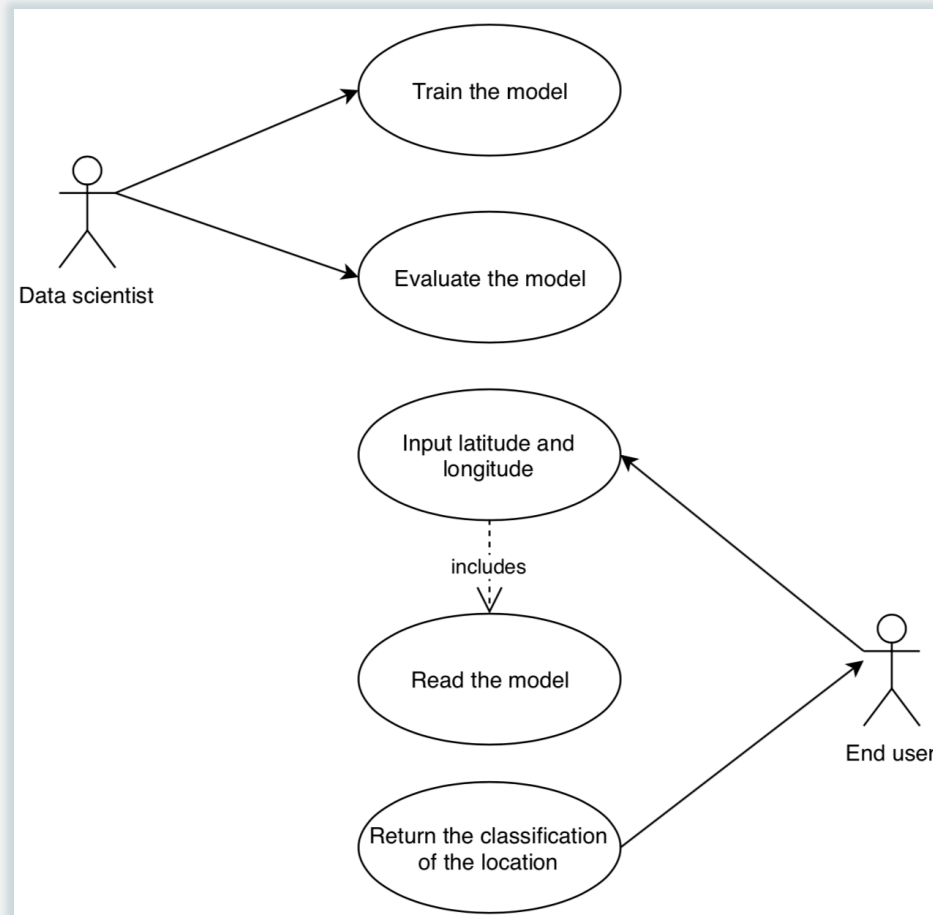


Figure 14. Use case diagram

RESULTS

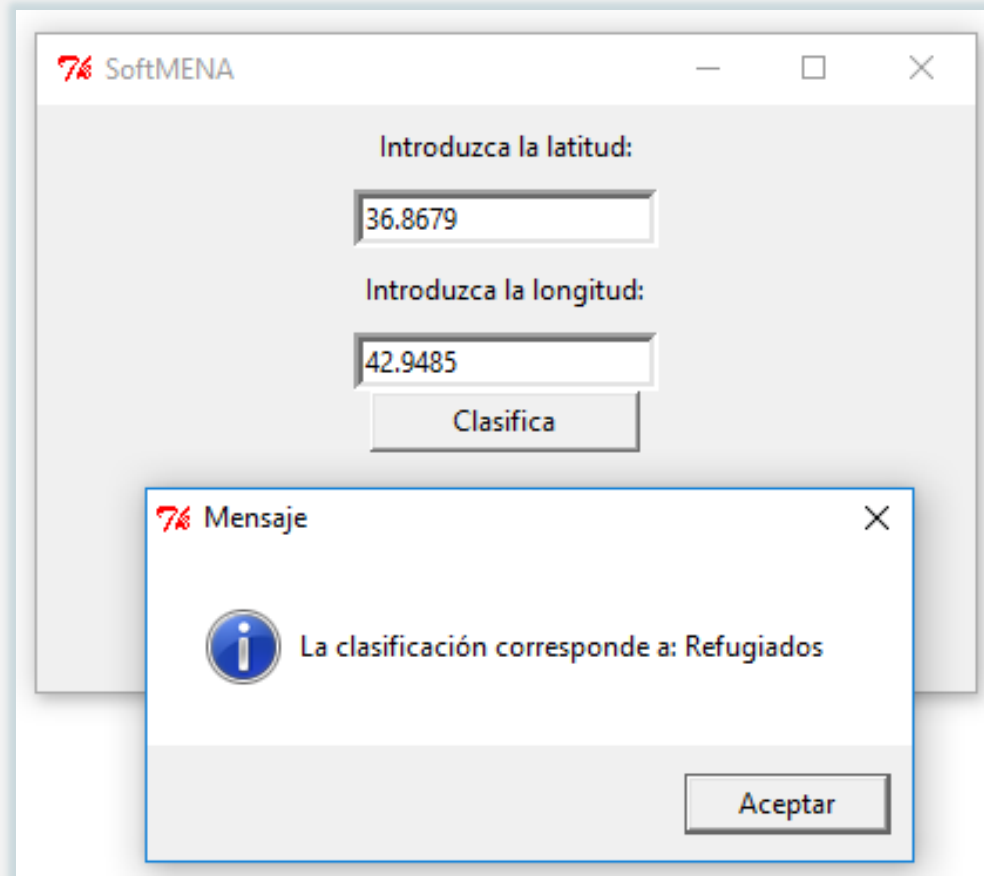


Figure 15. Classification of Duhok, Iraq

<https://vimeo.com/268047670>

RESULTS

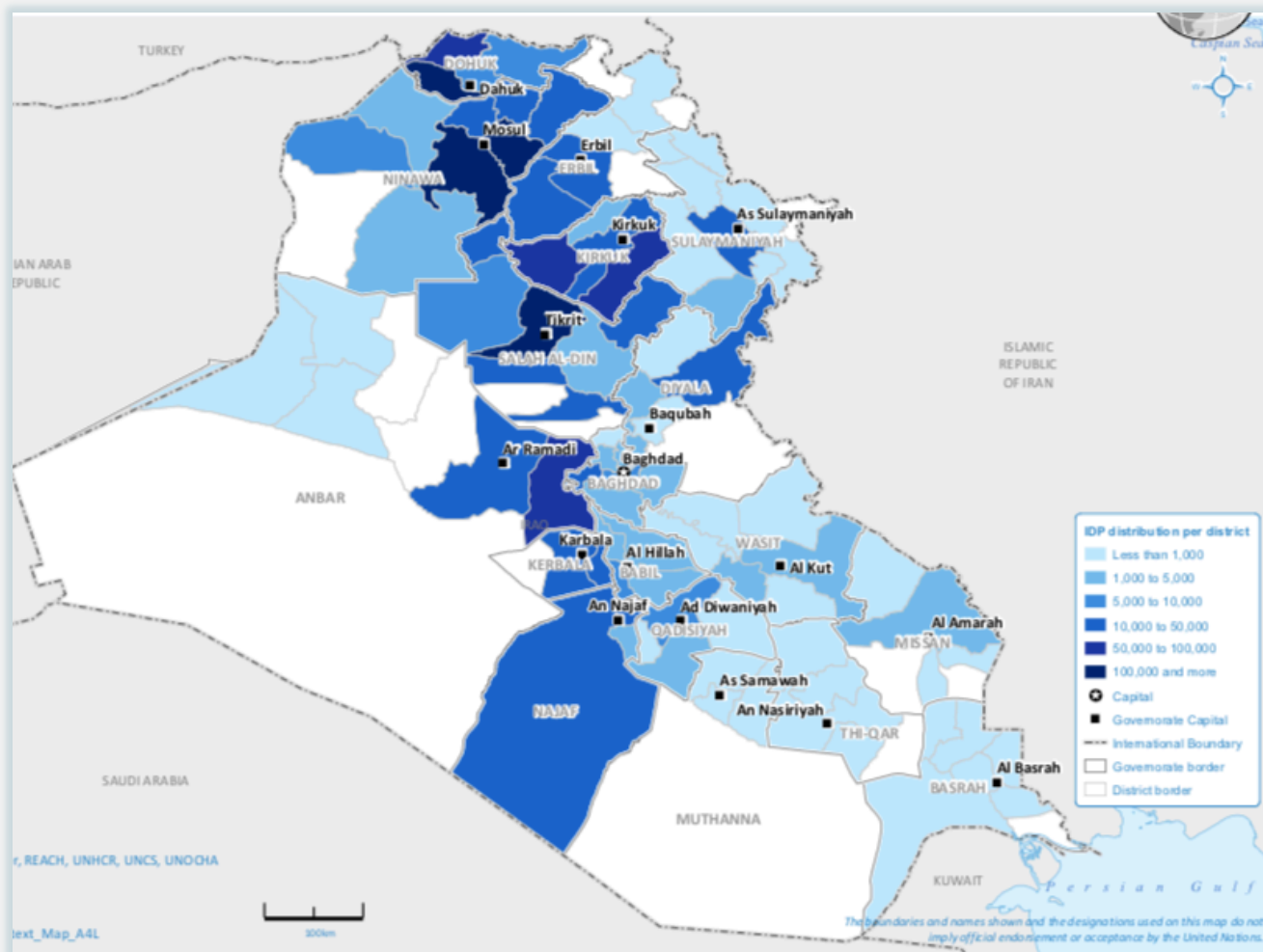


Figure 16. UNHCR populations in temporary settlements by district [14]

14. UN High Commissioner for Refugees (UNHCR) (2017). *Iraq: CCCM - IDP populations in temporary settlements by district*. URL: <http://www.refworld.org/docid/5977453c4.html>

RESULTS

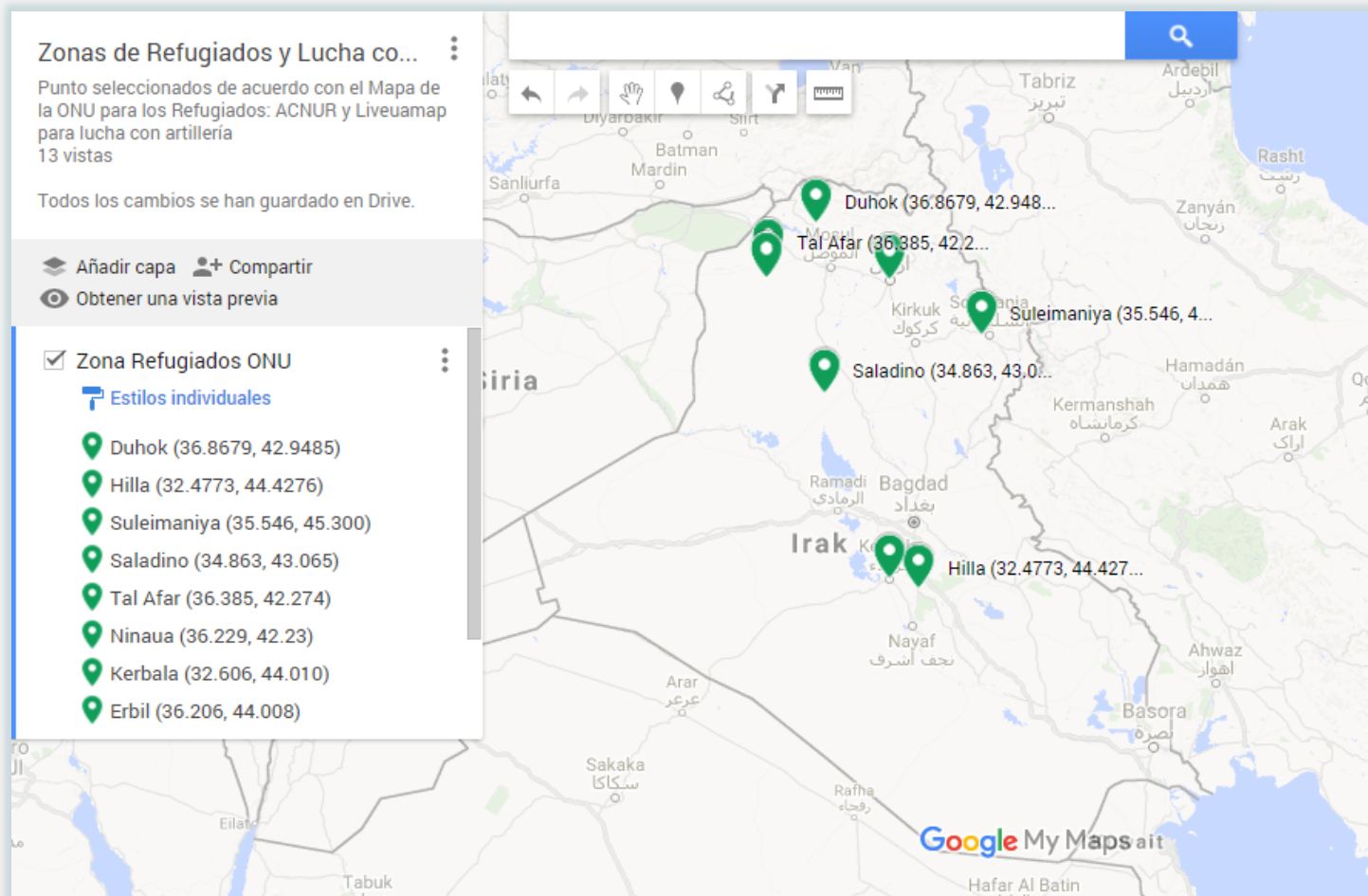


Figure 17. Eight refugee areas with data from the UNHCR, 2017

RESULTS

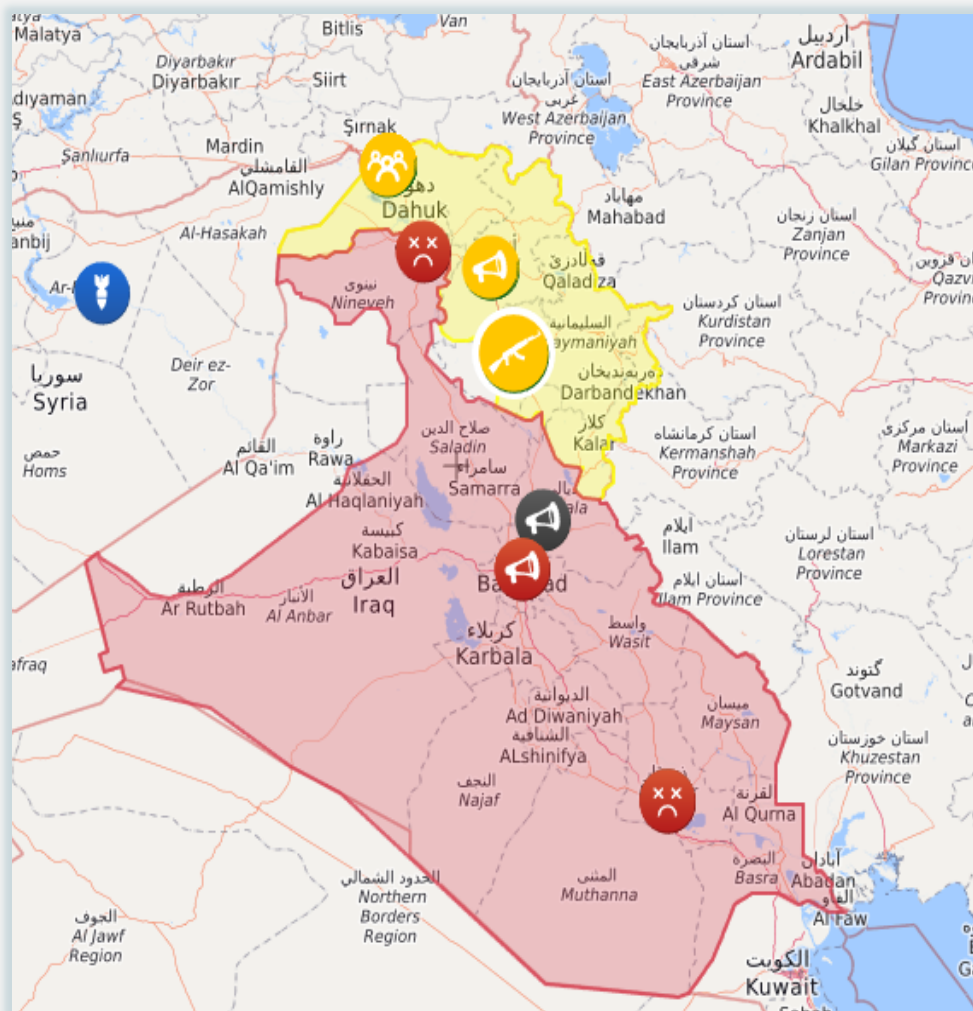


Figure 18. Map of artillery fight events [15]

RESULTS

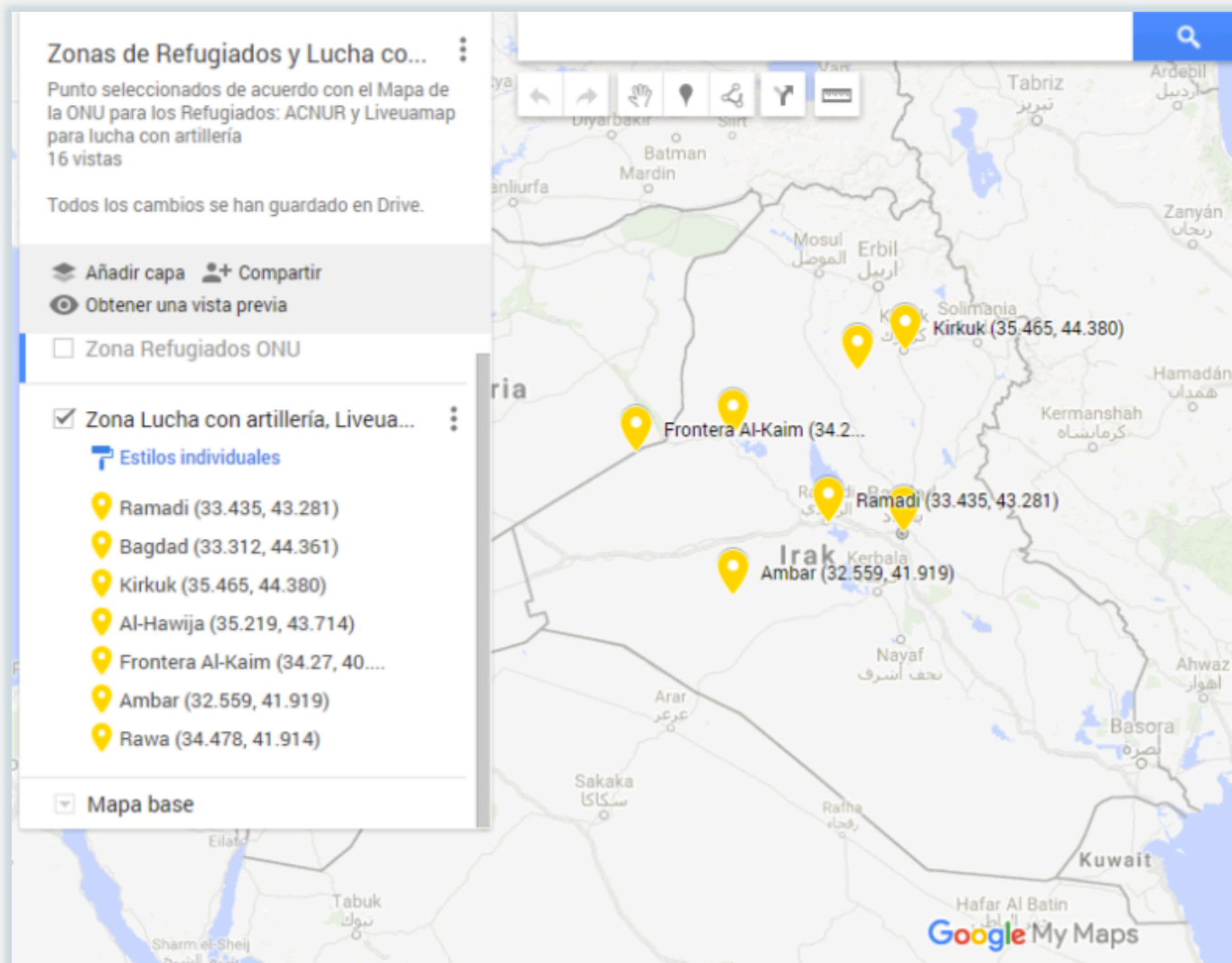


Figure 19. Seven artillery fight areas according to the Live Universal Awareness Map

RESULTS

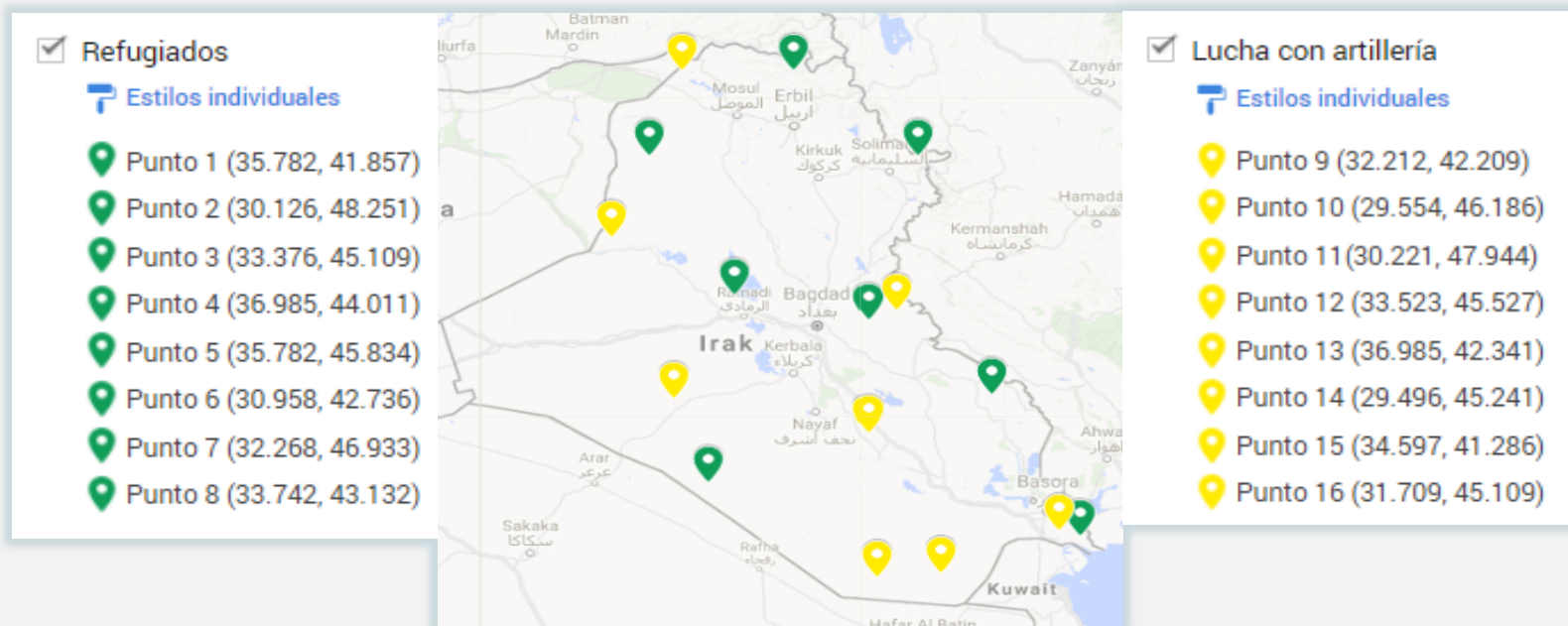


Figure 20. Areas without official data in Iraq classified as areas with needs related to refugees (*in green*) and needs related to artillery fights (*in yellow*)

CONCLUSIONS

- A software was created to automatically discover the needs of people in Iraq, a country in the 10/40 Window
- This software uses a predictive model (i.e., a classifier) generated with machine learning
 - This model was applied to open big data from GDELT
 - It uses the decision trees algorithm to classify two events: refugees and fights with artillery and tanks
 - **Areas without available official data can be classified with an accuracy of 76%!**

FUTURE WORK

1. Waiting for feedback from MENA
2. Carry out additional experiments with other events and other countries in the 10/40 Window
3. Real-time analysis with data from GDELT by means of big data technology (Apache Spark)

Discovering the Needs of People in the 10/40 Window with Data Science

Harvey Alférez, Ph.D.
Merari González, MSc.

Global Software Lab
School of Engineering and Technology
Montemorelos University

