# Detecting Damaged Regions after Natural Disasters using Mobile Phone Data: The Case of Ecuador


**Galo Castillo-López**
Escuela Superior Politécnica del Litoral
`gadacast@fiec.espol.edu.ec`

**María-Belén Guaranda**
Escuela Superior Politécnica del Litoral
`mguarand@fiec.espol.edu.ec`

**Fabricio Layedra**
Escuela Superior Politécnica del Litoral
`fblayedr@fiec.espol.edu.ec`

**Carmen Vaca**
Escuela Superior Politécnica del Litoral
`cvaca@fiec.espol.edu.ec`


## Abstract


In this work, we use mobile phone activity data to infer the affected zones in the Ecuadorian province of Manabí, after the 2016 earthquake, with epicenter in the same province. We calculate a series of features to train a classifier based on the K-Nearest Neighbors algorithm to detect affected zones with a 75% of precision. We compare our results with official reports published two months after the disaster.


## 1 Introduction and Dataset

Large scale natural disasters involve budgetary problems for governments [1] and prioritizing the allocation of resources requires near real time information about the impact of the hazard in different locations [2]. Such information is not available through sensors or other devices, specially in developing countries that do not have such infrastructure. A rich source of information is the data resulting from mobile phone activity that citizens in affected areas start using as soon as they become available after the disaster. We exploit such data in this work to conduct different analyses in order to identify the affected zones after the earthquake that took place in the Ecuadorian province of Manabí on April 16th, 2016.

Our main dataset consists of anonymous records published by a telecommunication provider operating in Ecuador. This dataset is the same used in [3] and contains 11 million records of SMS (Short Message Service) messages and phone calls, which were produced from April 15th to the 18th (one day before and two days after the earthquake, respectively). In order to explain what the entries of the dataset represent, the following concepts are needed:

1. Event: either a mobile phone call or a SMS. All the events in the dataset started in any city located in the province of Manabí.
2. Event tower (ET): this is the tower the user's device connected to when generating the *event*.
3. Home tower (HT): this is the tower where the user's device has been connected to most of the times, historically, when generating an *event*.

Our second dataset is provided by *SENPLADES*, an Ecuadorian government's entity in charge of the planning of strategies for the development and well-being of the country. This dataset consists of labels of the level of damage that each canton at Manabí suffered due to the earthquake, presented two months after the disaster. From this dataset, we obtain two labels for each canton: highly and moderately damaged. Both datasets contain information about 20 out of the 22 cantons from Manabí.

Table 1: Classifiers precision, recall and F1 score results.

| Classifier | Precision | Recall | F1 Score |
|---|---|---|---|
| Linear SVC | 0.70 | 0.56 | 0.63 |
| K-Nearest Neighbors | 0.75 | 0.67 | 0.71 |
| Logistic Regression | 0.70 | 0.33 | 0.43 |

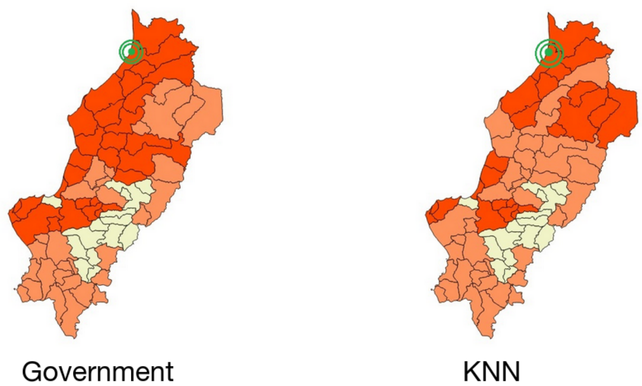


Figure 1: Left: map reported by SENPLADES. Right: map reconstructed with our model. A darker red represents a higher level of damage. The green mark depicts the epicenter of the earthquake.

## 2 Methods

The aim of this work is to automatically label a canton as highly or moderately damaged after a natural disaster using mobile phone activity. We calculate four features to characterize the activity presented in each of the cantons using data produced one day before and one day after the earthquake.

First, we propose *Visitors Diversity Index*, a tuned *Shannon Entropy*, which is a metric that explains the popularity of a place $P_i$ (i.e. canton) in terms of the amount of different places where people, currently located in $P_i$, come from. Second, with the aim of detecting significant changes between the activity produced on April 15th and 17th, we aggregate cell phone activity at canton level and by time spans of one hour. Then we build a pair of 24-dimensional vectors for each canton, comprising hourly activity of both days. We use the Euclidean Distance on each pair of vectors as a metric for calculating the difference among them. Third, we use records from April 17th to build a directed graph $G = (V, E)$, where $V$ is the set of vertexes that represent each tower (ET or HT) and $E$ is the set of weighted edges, where an edge $e_{ij}$ indicates that a group of $k$ clients from a HT $v_i$ made an event in the ET $v_j$. We calculate the weight of an edge based on the distance in kilometers between the towers. Finally, we also use the distance to the earthquake's epicenter (Haversine formula).

Using the previous metrics, we perform three supervised learning algorithms: Linear SVM (SVC), Logistic Regression and K-Nearest Neighbors (KNN). We use Leave-One-Out cross validation in order to evaluate our results, due to the low amount of observations (i.e. 20 cantons).

## 3 Results and Conclusion

Table 1 shows the classification performance for all of our models. We note that KNN presented the best results in all the performance metrics. Although we obtained the same precision with Linear SVC and Logistic Regression, the SVC's recall metric outperformed the Logistic Regression's sensitivity. We could only correctly label as highly damaged 3 and 5, out of 9 highly damaged cantons using Logistic Regression and Linear SVC respectively. In contrast, we correctly identify 6 out 9 highly damaged cantons using K-Nearest Neighbors (KNN). Figure 1 shows a heatmap of the different levels of damage of cantons from Manabí that we were able to reconstruct with the KNN classifier.

# References

[1] Cardona, O.D., Ordaz, M.G., Marulanda, M.C. & Barbat, A.H. (2008) Estimation of probabilistic seismic losses and the public economic resilience an approach for a macroeconomic impact evaluation. *Journal of Earthquake Engineering* **12**(S2), pp. 60—70.

[2] Wilson, R., zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonza-lez, M., Guthrie, S., Chamberlain, H., Brooks, C., Hughes, C., et al. (2016) Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 Nepal earthquake. *PLoS currents* **8**, pp. 1—25.

[3] Andrade, X., Layedra, F., Vaca, C., Cruz, E. (2018) RISC: Quantifying change after natural disasters to estimate infrastructure damage with mobile phone data. *2018 IEEE International Conference on Big Data (Big Data)*, pp. 3383—3391. IEEE.