**A PRIORITY INDEX FOR HUMANI-TARIAN AID AFTER A TYPHOON**

**The Priority Index is a data driven solution to predict damage to houses after a (super) typhoon or hurricane. We use data and machine learning techniques to identify priority areas for humanitarian aid. Organizations like the Red Cross  and Red Crescent National Societies, governments or UN OCHA can use these results to better understand the impact of a natural disaster and to mobilize humanitarian response faster.**

(NOTE concerning DataMission Hackathon: this is a stripped version of a blogpost on the subject, which is temporarily taken down. Information on importance of features and R2-scores is filtered.)

Timely information can save lives. Aid organisations must recognize that accurate, timely information is a form of disaster response in its own right

*Niskala, Secretary – General of the*[*IFRC (2003-2008)*](https://webmail.wur.nl/owa/#_ENREF_31)

**1.1 Humanitarian aid improves with accurate and timely information**

When a natural disaster strikes, the local government, NGOs and Red Cross and Red Crescent National Societies quickly need information on the damage (affected population, casualties, road blocks, flood extent, damaged houses) in the areas that were hit by the disaster. The information that is presented to decision-makers in the wake of a disaster needs to be accurate, appropriate, timely and valid.

One of the challenges with disaster response is scarcity of resources: not each affected family can be helped. Therefore it is essential to identify priority areas , by assessing damage and finding vulnerable people that are affected the most. Currently damage assessments and identification of the most vulnerable is a time consuming process, which can takes weeks to complete, due to logistics, safety constraints, or workload.

Assessment teams need to go into the affected area and interview people affected and review damage to houses. Due to time constraints, or limited information sharing, there is a risk that decisions on priority areas are not based on complete and accurate information, and thereby also organizational or political preferences could be taken into account, as well as influence by the media on areas that receive more media attention than other areas.

During a study in the Philippines, 60% of 32 interviewed decision makers (government, NGO’s and UN) have indicated  that a faster, more complete and more objective analysis of priority areas (Priority Index) could be useful to identify areas with high damage and number of people affected. Thereby supporting decision makers to prioritize and distribute aid efforts and reach the most vulnerable people in the worst affected areas more efficiently.

**1.2 Building the priority index for typhoons**

Our aim is  to develop a methodology to identify high priority areas for humanitarian response, based on (open) secondary data of affected areas, combined with disaster impact data (such as windspeeds and rainfall) and by learning from past disasters. It is important that we invest in data preparedness, so that these pre-crisis secondary datasets are available and up-to-date.

Applied research on this objective is ongoing for Typhoons (Philippines), Earthquakes (Nepal) and Floods (Malawi). Our objective is to develop machine learning methodologies that can be applied to different countries, using local data, and with minor modifications reach a fast and sufficiently accurate damage prediction. In this blog we describe initial results for the Philippines during Typhoon Haima on October 19th 2016.

**1.3 Data**

Data used for the prediction model includes country wide base line data (administrative boundaries, population, poverty, house wall and roof types), Geographical features per municipality (ruggedness,  slope, coastline length, distance to coast), combined with impact data (wind speed, rainfall, typhoon path), and uses a number of specific features created from these data.

Official counts by DSWD and NDRRCM (Philippine government) on damaged houses are used to validate the model. For this we used data from four past typhoons: Haiyan, Melor, Hagupit and Rammasun. All data was aggregated to the municipality level. Unfortunately barangay level damage counts are not available in the datasets published by the government. All information per municipality was integrated using the PCODE-system, which assigns a unique identifier to each administrative area in the Philippines. To ease this task, an efficient PCoder-program was developed.

3 second sustained windspeed

*Explained: damage counts. Damage counts, people affected counts and casualties are collected by the barangay council and reported by the barangay captain to the local government units (LGU). The LGU’s report the municipality aggregated data to the regional office of civil defence (OCD), who then reports it to the national OCD. Casualty counts are collected on a personal level and are double checked, accurate and valid. Damage counts are collected without a centralized methodology, and therefor methods of data collection can differ. People affected counts is a very wide term, and there is no centralized definition of when a person is affected. Therefor the counts can differ widely between municipalities. An analysis from Typhoon Haiyan shows that many of the people affected reports are estimates where either 0%, 50% or 100% of the total population in that area is reported as being affected. For a learning algorithm this distribution cannot be used to provide a reliable prediction*

**1.4 The prediction model**

In the risk management domain probabilistic models are being developed for determining the likelihood of losses from a disaster (usually economic loss). It creates impact scenario’s that can be used by decision makers to mitigate risk. These models however are not developed to predict impact on people during a recent disaster. Our approach is not to develop sophisticated hydrologic, seismic, or windspeed models, but to use machine learning methods to find the best predictors in existing base line data to predict typhoon impact. Different machine learning methods have been tried (including neural networks). Currently we are using a method called Random Forest Regressor.

Explained: Random Forest Regressor. Its power comes from an interesting strategy of building multiple predictors (decision trees) and averaging their outputs. Each tree is built in a slightly different way, using different subsets of historical data, and randomly selecting different variables during the process of building the trees. This strategy allows to build a model that can handle multidimensional data well and can estimate importance of each input variable. It is a highly configurable method so several experiments were held to select parameters that produce the best results on training data.

**1.5 Testing the model**

We had the unique opportunity to test  the model in the recent Typhoon. We dropped all our other work and got the team to fast track the development of the model and collect and clean the impact data, so that we were able to release a first Priority Index within 24 hours after landfall. More than four days later the first official counts of damage of parts of the affected area were released. The results where shared with humanitarian organizations, government and through social media. We have produced two types of analysis.

Priority areas within 24 hours

Predicted numbers were used to prioritize municipalities on a scale from 1 to 5 (1 with the lowest predicted number of damaged houses, 5 for the highest predicted number of damaged houses). The map and data were shared in the humanitarian community and reviewed by a few organizations such as UN OCHA and the Shelter Cluster.

Absolute damage to fill gaps in government counts

We used the model to complete gaps in the official counts of DSWD and NDRRCM. For this we included the official counts in the model and ran the model again to predict the gaps in the official data. This map was used by the Shelter Cluster to get a better overview of total house damage in the affected areas.

**1.6 Conclusions**

From our work so far we can conclude that when data preparedness is done right, and disaster impact data collected structurally after an event, then **it is possible to use machine learning techniques to build reliable damage predictions**.

Although damage predictions by using data are not perfect, they are far **more transparent** than other prioritization methods, because the underlying data, assumptions and methodologies are shared openly.