510
PERSONALIZED HUMANITARIAN
AID THROUGH BIG DATA

✚ AN INITIATIVE OF THE NETHERLANDS RED CROSS

Maarten van der Veen
Marco Velliscig
Jannis Visser

Agenda
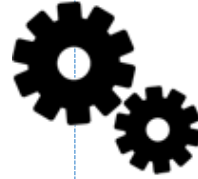
▶ **1. Schedule & Practicalities**

2. Typhoons, the Humanitarian Response and the Priority Index Model

3. The Challenges

Modeling challenge

- 9:05 Start work

- 12:30 – 13:30 Lunch

- 13:30 – 14:00 Communal session if interested.

- ~ In general, you can check one time your intermediate R-squared with us.

- 15:45 Deadline to send prediction results to JannisVisser@redcross.nl. In return, you receive the 'actual damage'-file. Compute your score using the provided R/Python-script.

- 15:55 Deadline to send presentation (ppt) to JannisVisser@redcross.nl

- 16.00 Start presentations (max 5 minutes per group).
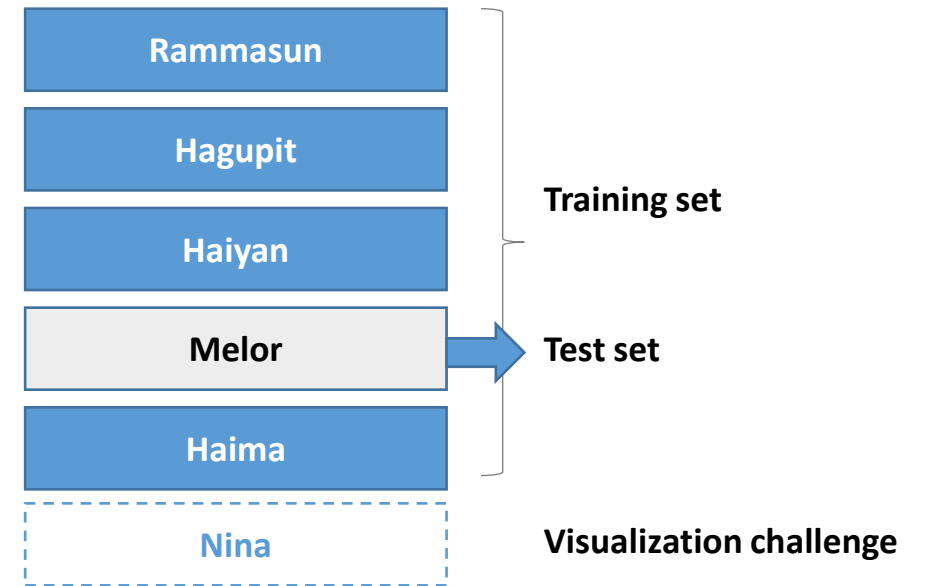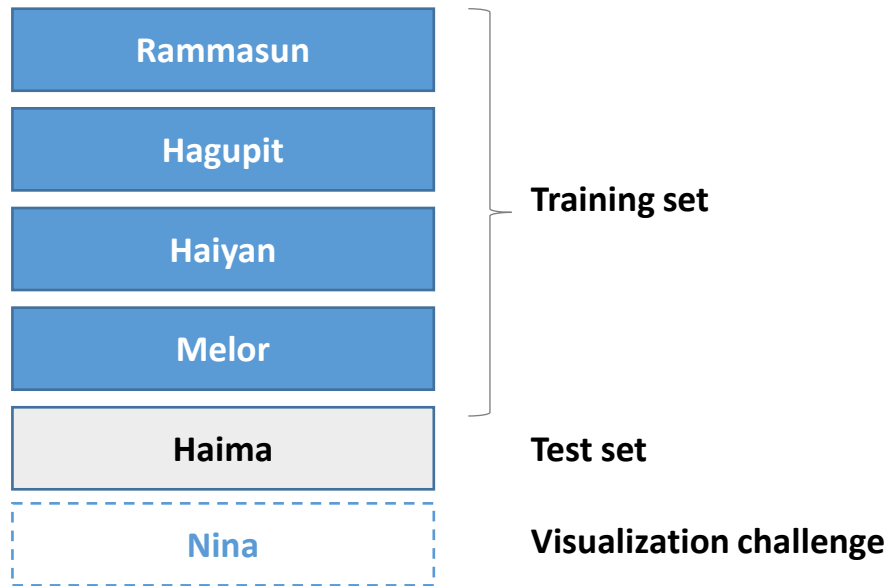
Visualization challenge

- 9:05 – 9:15 Short extra introduction session

- 9:15 Start work

- 12:30 – 13:30 Lunch

- 15:55 Deadline to send presentation to JannisVisser@redcross.nl

- 16:00 Start presentations (after modeling groups).

Modelling Challenge

Improving the Prediction Model

Rammasun	
Hagupit	
Haiyan	Training set
Melor	
Haima	Test set
Nina	Visualization challenge

Rammasun	
Hagupit	
Haiyan	Training set
Melor	Test set
Haima	
Nina	Visualization challenge

AN INITIATIVE OF THE NETHERLANDS RED CROSS

2

Improving the Visualisation and Communication of the Model

Use Cases

- **Philippine Red Cross operations**: interested in priority areas, where to send Early Assessment teams, where to start Early Response operations.

- **UN OCHA** (Office for Coordination of Humanitarian Affairs): more interested in overall picture / damage count. It determines the needed appeal for funding, etc. Also interested in priority areas though.

- **Shelter Cluster:** interested in priority areas, but also in absolute numbers of damaged houses, because they need to start to provide shelter kits and temporary housing and plan rebuilding.

- **General interest in the sector:** The result can also be of interest to not directly involved parties, who are interested in the technology, etc. These might not know everything about typhoons or about the humanitarian setting.

How to deal with various use cases

- It's OK to **choose a use case**, and focus on that, if you have particularly good ideas about a certain use case. Make clear what you're focusing on within your story though.

- However, it's overall OK (and possibly more realistic) to go for a **hybrid mix**, which caters to various use cases. In reality, it might make more sense to send out one product. Possibly by making clear within the storyboard that some sections are more/less interesting to specific audiences.

- In general, keep in mind that all audiences have relatively **low data literacy**. And in all cases, reason from the (various) readers. What context is required, etc?

Data

- http://510.global/hackathon-input-full.zip
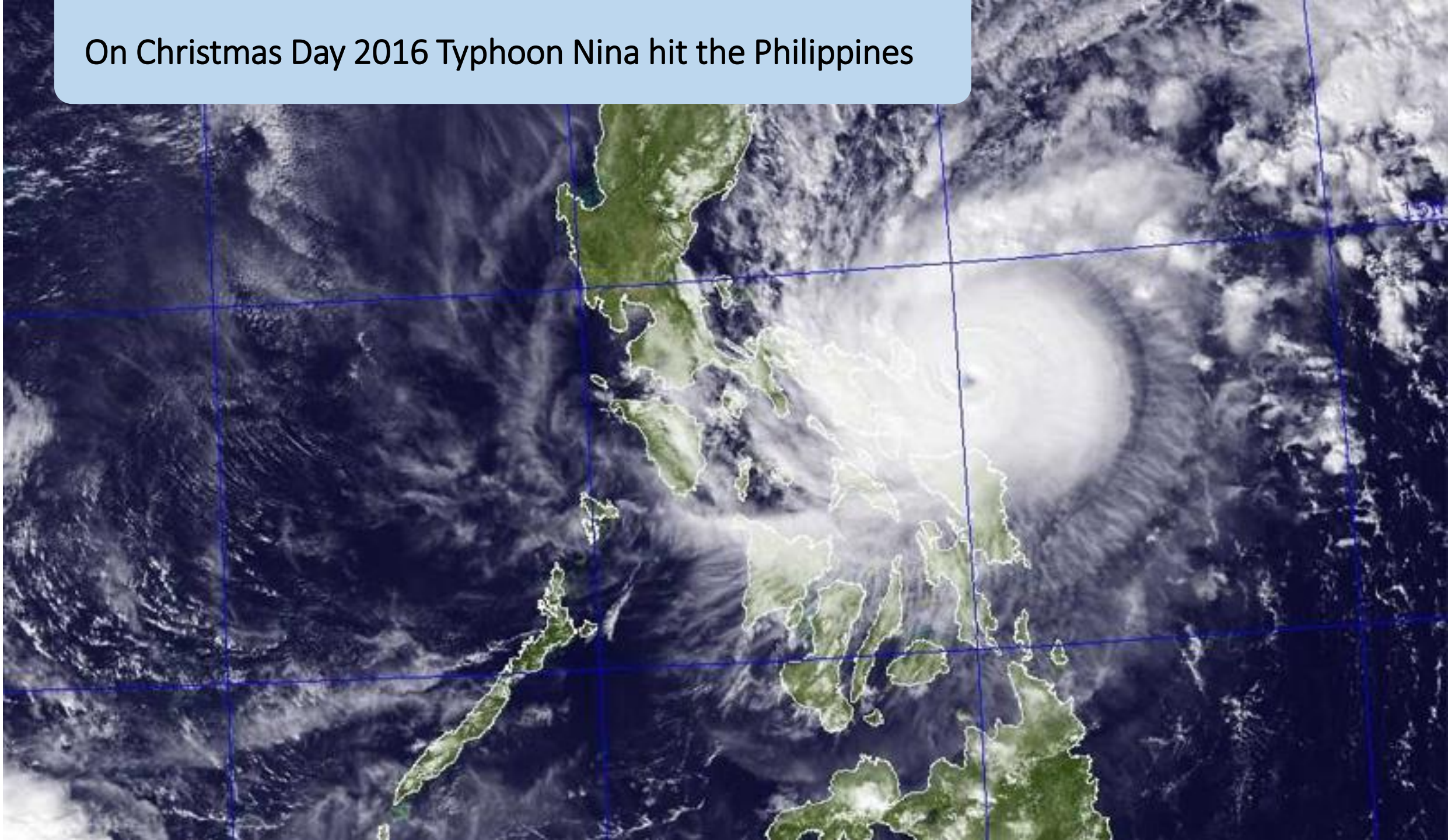- 4 USB-sticks

Agenda

1. Schedule & Practicalities

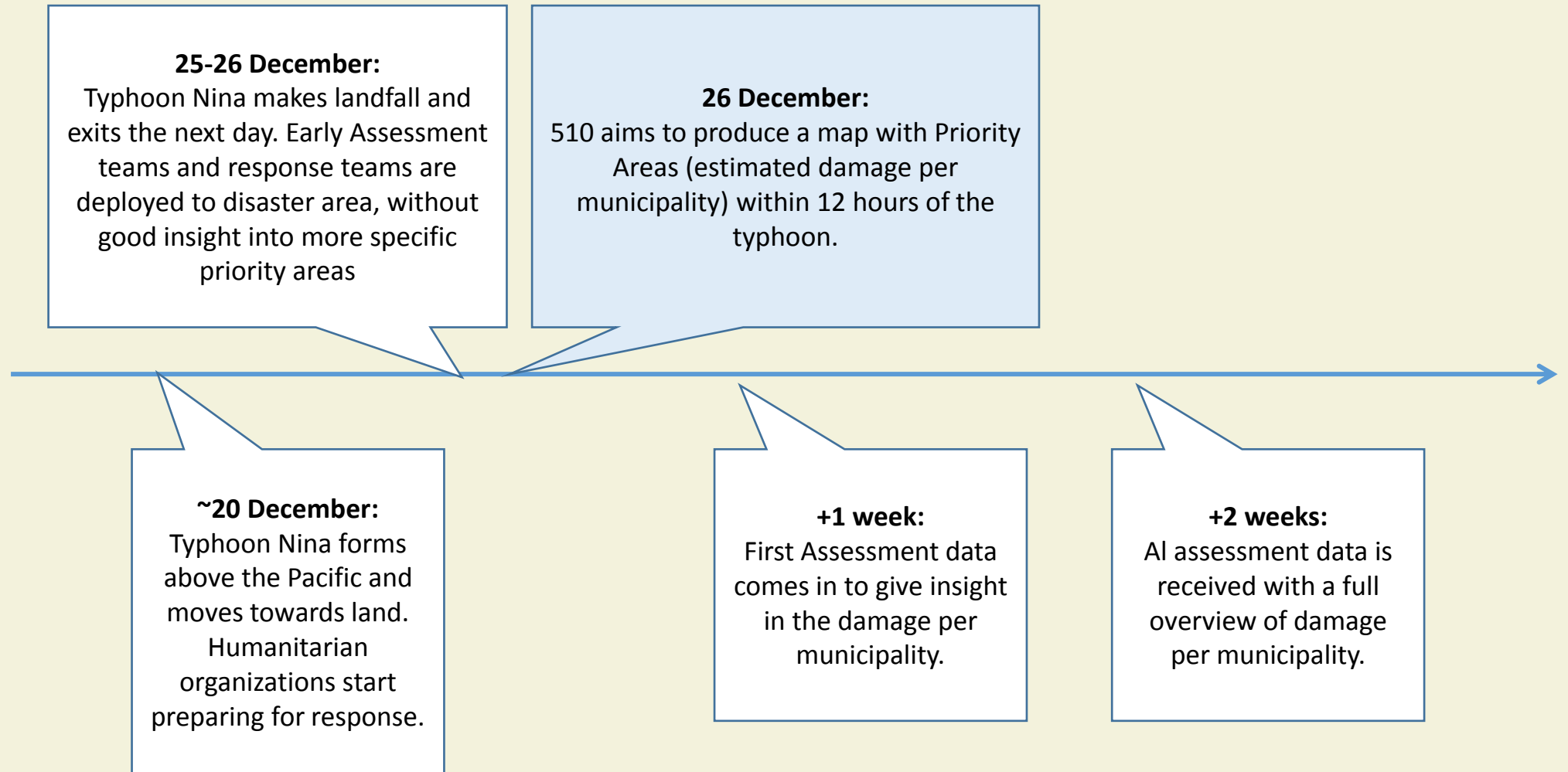2. **Typhoons, the Humanitarian Response and the Priority Index Model**

3. The Challenges

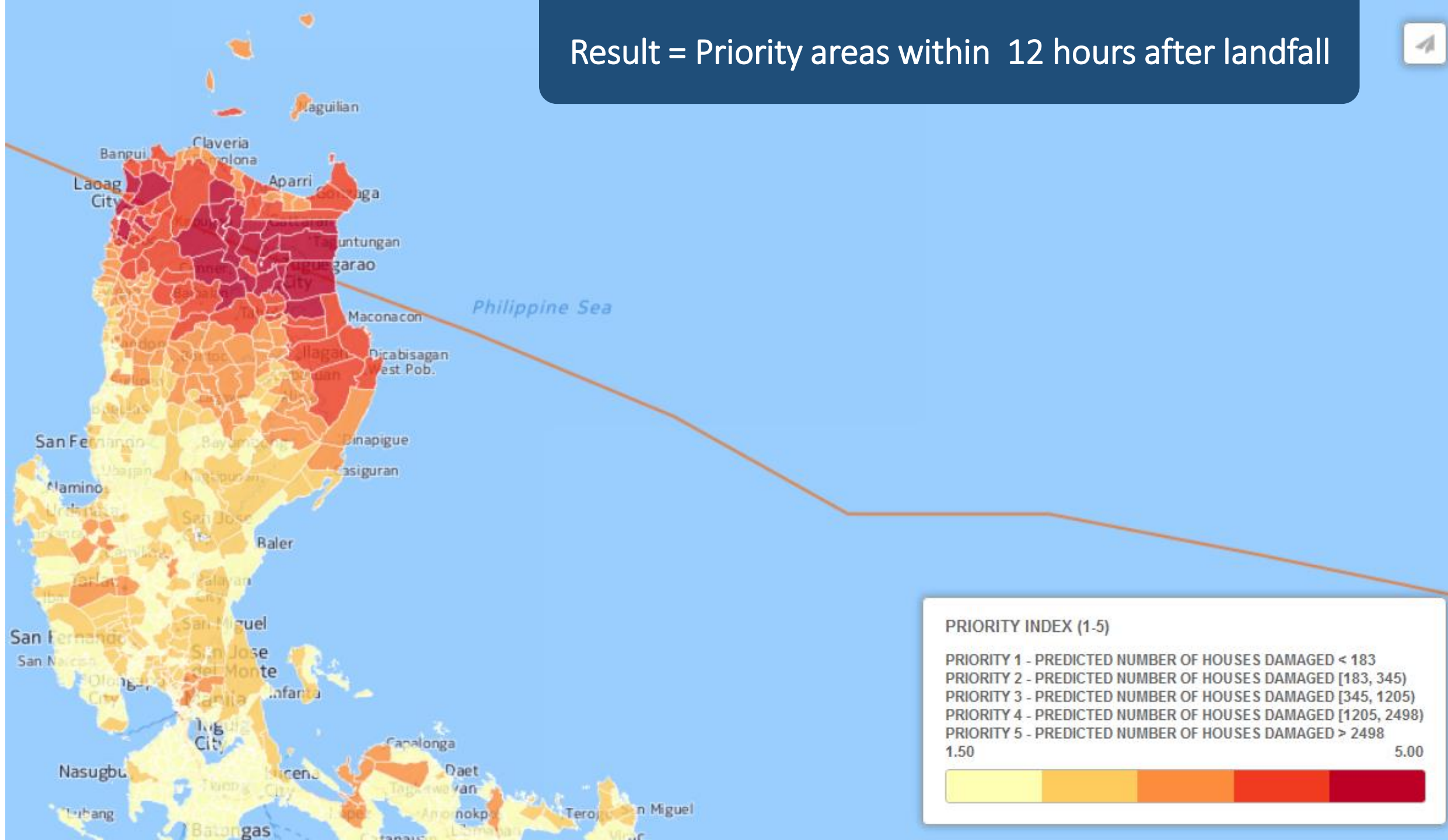On Christmas Day 2016 Typhoon Nina hit the Philippines

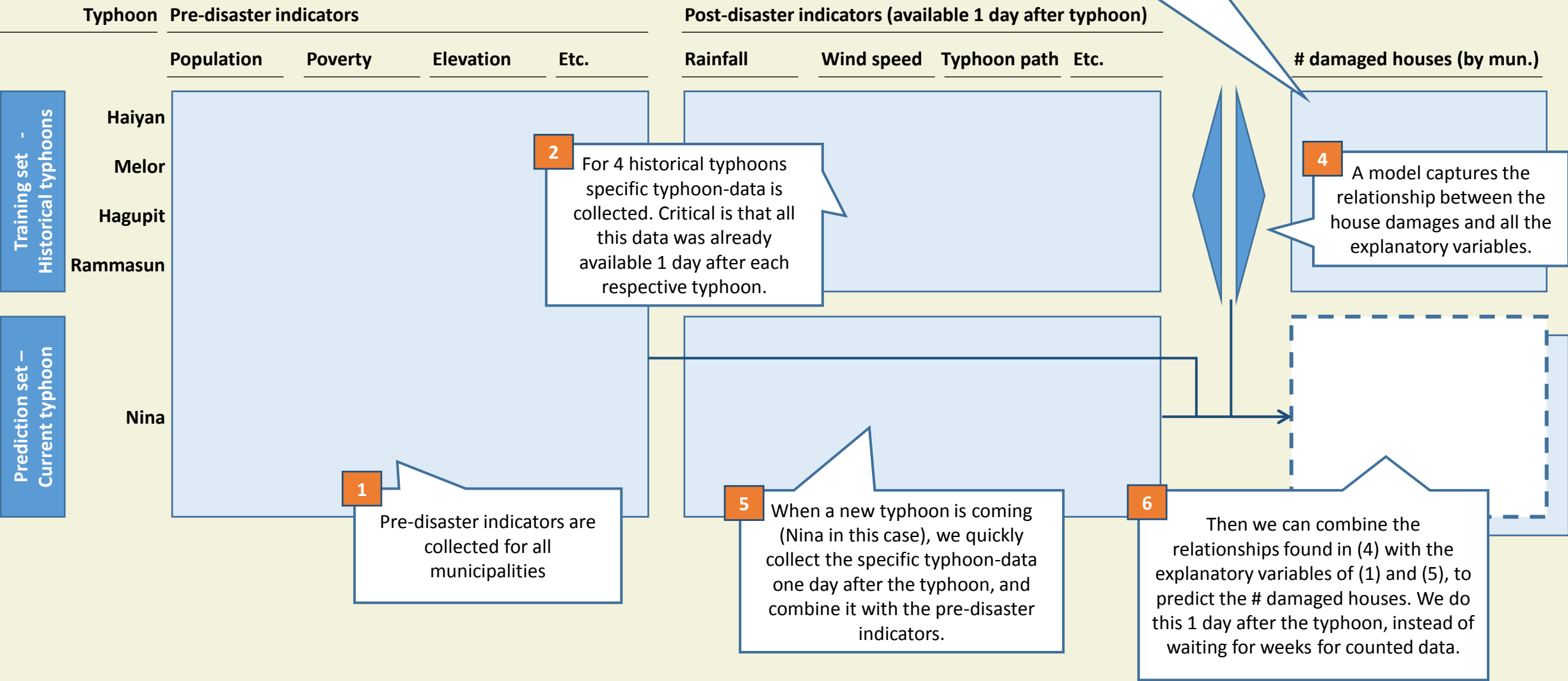# The timeline of typhoon and humanitarian response

25-26 December:
Typhoon Nina makes landfall and exits the next day. Early Assessment teams and response teams are deployed to disaster area, without good insight into more specific priority areas

26 December:
510 aims to produce a map with Priority Areas (estimated damage per municipality) within 12 hours of the typhoon.

~20 December:
Typhoon Nina forms above the Pacific and moves towards land. Humanitarian organizations start preparing for response.

+1 week:
First Assessment data comes in to give insight in the damage per municipality.

+2 weeks:
AI assessment data is received with a full overview of damage per municipality.

Result = Priority areas within 12 hours after landfall

PRIORITY INDEX (1-5)

PRIORITY 1 - PREDICTED NUMBER OF HOUSES DAMAGED < 183
PRIORITY 2 - PREDICTED NUMBER OF HOUSES DAMAGED [183, 345)
PRIORITY 3 - PREDICTED NUMBER OF HOUSES DAMAGED [345, 1205)
PRIORITY 4 - PREDICTED NUMBER OF HOUSES DAMAGED [1205, 2498)
PRIORITY 5 - PREDICTED NUMBER OF HOUSES DAMAGED > 2498

1.50   5.00

AN INITIATIVE OF THE NETHERLANDS RED CROSS

Agenda

1. Schedule & Practicalities

2. Typhoons, the Humanitarian Response and the Priority Index Model

3. **The Challenges**

Two challenges in this hackathon

1 **Improving the Prediction Model**

- Write a machine learning algorithm to predict the damaged houses per municipality of a new typhoon
- Use the supplied data from 4 historical typhoons
- Feel free to come up with additional data
- At the end of the day predictions per team will be compared to actuals, and the team with the highest R-squared wins.

2 **Improving the Visualisation and Communication of the Model**

Come up with a way to communicate the results to relevant stakeholders in the Philippines in an engaging and accurate manner. Consider:
- Storytelling: engage stakeholders with a convincing story.
- Visually pleasing and informative format and visualisations. (map, dashboard , storyboard, …)
- Accuracy is key. Get across the benefits and potential errors or biases of the model and results. What does it really mean?
- Output will be presented and judged at the end of the day by all the teams to determine a winner.

See next slides for more details

Modelling Challenge

Improving the Prediction Model

1. You are supplied with data on 4 historical typhoons, with for each typhoon one data-row per municipality that was hit by the typhoon.
2. Per municipality you have data in the following categories
   a. Actual damage (variables to predict): Damaged Houses separated in Completely and Partially damaged.
   b. Event-specific (this is post-disaster data, such as windspeed, rainfall and distance to typhoon path)
   c. Geographic (elevation, slope, coastal length, etc.)
   d. Exposure (population and density)
   e. Socio-economic (poverty and wall- and rooftype materials)
3. Also added is one new typhoon, which has exactly the same columns, but for which the actual damage variables are empty. This is your task to predict.
4. The variable to predict is total number of damaged houses. But it is very much encouraged to split this in two separate models, for Completely damaged houses and Partially damaged houses, as the underlying processes may work very differently.
5. Write a machine learning algorithm in Python or R to make these predictions. Where possible, add explanatory comments to your scripts.
6. You are also free to add any other data sources or transformations that you think are relevant. (NOTE: for event-specific data, this can only be a data source that is available very quickly after a typhoon, otherwise it cannot be used in our model.)
7. At the end of the data provide us with your predictions per municipality. Then we will compare this with the actuals to compute the R^2. The team with the highest R^2 wins.

Included files on USB
- Datamatrix described above
- Description of all supplied variables (excel-list + folder of images with visualizations of most variables)
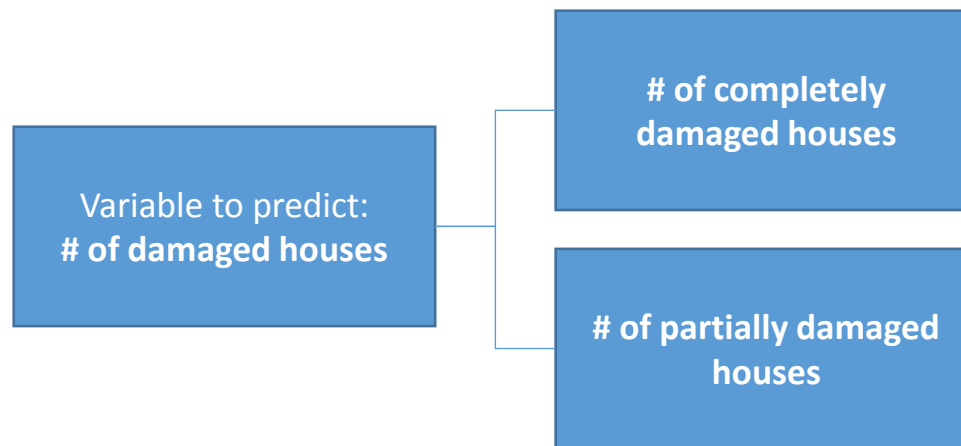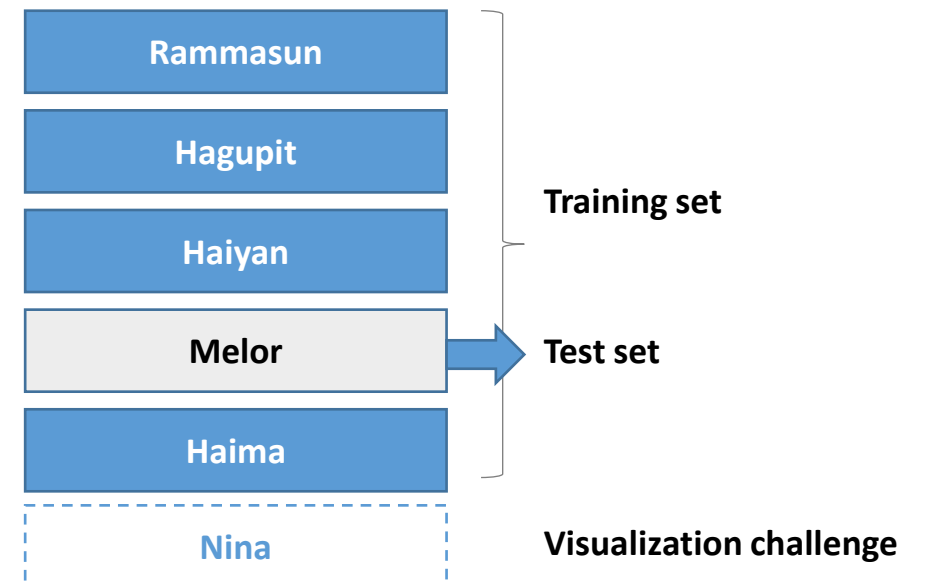
Modelling Challenge

1

Improving the Prediction Model

Prediction objective

Variable to predict:
# of damaged houses

# of completely damaged houses

# of partially damaged houses

Typhoons

Rammasun
Hagupit
Haiyan

Training set

Melor

Test set

Haima

Nina

Visualization challenge

- You can make any transformation you want to the prediction variable, but at the end you have to transform back to exact # of damaged houses. Turning it into a classification problem is therefore not likely.

Modelling Challenge

1

Improving the Prediction Model

Scoring

$$R^2 = 1 - SSR/SST$$

- Scoring scripts in R / Python

- After sending in your prediction results on Saturday, you will receive the file with actual damage data in return.

- Run one of the scripts to calculate your R-squared, which you can present in your presentation. (We will also check it though with our own script.)

- During the day you are allowed one time to ask us to calculated the R-squared of your current model.

Presentation

- Present your R-squared value, as well as your approach, newly found/created features, important features, etc.

- Make sure to also include any good ideas, which might not have found their way fully into the R-squared.

Modelling Challenge

1

Improving the Prediction Model

Modelling

- Use any software and modeling technique you want.

- It might be very fruitful to think about additional explanatory variables

  - Finding data online (Note: typhoon-specific data should be available very quickly after typhoon.)

  - Creating new geo-features (make use of GIS-experts)

- Provided files
  1. Training set: 4 typhoons >> X + Y variables
  2. Test set: Haima typhoon >> X variables only
  3. Output predictions file
  4. Variable descriptions

2

Improving the Visualisation and Communication of the Model

1. You are presented with a datamatrix which contains both predicted and actual damage for 4 historical typhoon, as well as all the underlying predictive variables. You are also presented with the predicted damage for a new typhoon.
2. Your challenge is to present the predicted data in an optimal manner to stakeholders in the Philippines.
3. Think about
   a. Storytelling: build up the story from high-level to detailed. The context: what should the reader know to understand everything? How can you cater to differing backgrounds/knowledge of stakeholders?
   b. What is the predicted measure: number of houses, or for example priority classes 1 to 5? Also, think about how to incorporate the difference between partly and completely damaged houses in your results.
   c. How do you visualize the output?
4. Most importantly we ask you to think about how to communicate the accuracy of the model (what are good measures, how to visualize/communicate them). Specifically, this splits in two tasks:
   a. On day 1 after the typhoon, we want to inform users about the accuracy; this must be based obviously on historical typhoons/predictions.
   b. On day 14 after the typhoon, data on actual damages is coming in. Think about how to add a section to your story that incorporates this data, including an analysis of prediction errors. Where do you over- and underestimate?
5. Output can be in any visual format. Powerpoint, Illustrator, etc. It is also possible to write a html/bootstrap-storyboard, which can be scrolled through in a browser.
6. Teams are judged at the end of the day by all other teams on 3 criteria, after shortly presenting their output.
   a. **Visualisation/Simplicity**: How simple are the results and message communicated?
   b. **Storytelling**: How engaging is the storyline?
   c. **Accuracy**: How clearly is the accuracy communicated

Included files on USB
- Data on modeled and actual damage for 4 historical (training set) typhoons
- Data on predicted damage for a new typhoon Nina
- Data on actual damage for a new typhoon (would be supplied normally 14 days after typhoon)
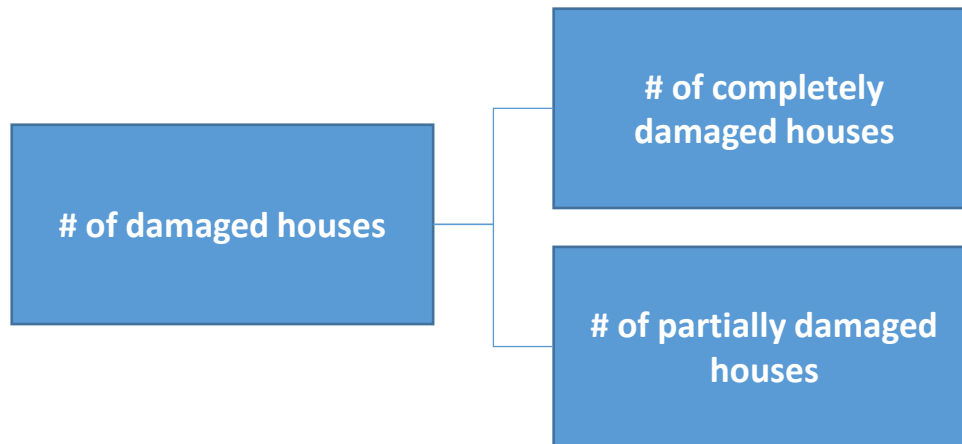
AN INITIATIVE OF THE NETHERLANDS RED CROSS
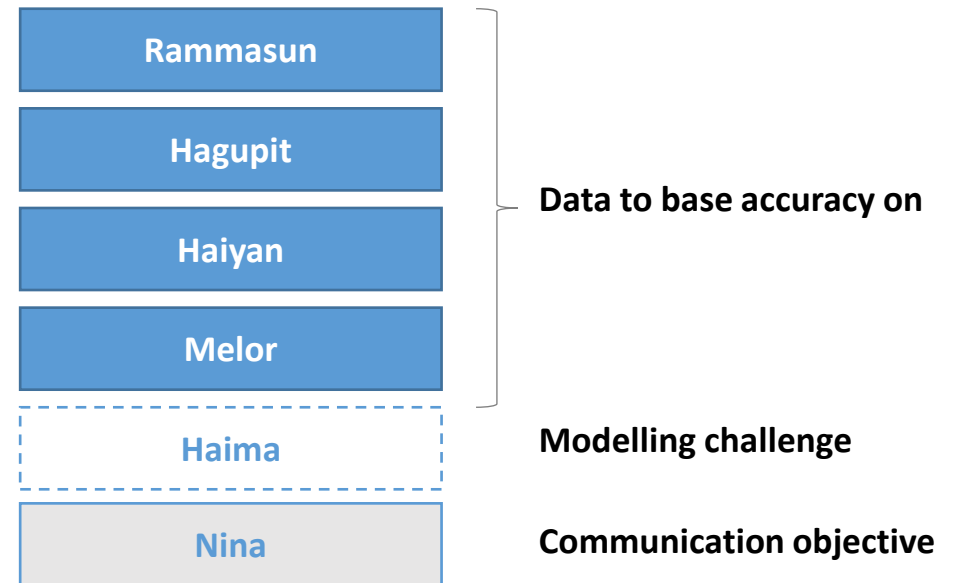
Visualisation Challenge

2

Improving the Visualisation and Communication of the Model

Accuracy

1. Accompany prediction results with accuracy measures

- Use data from historical typhoons
- Use data on underlying features

2. Analyze and communicate prediction results

- You are supplied with actual damage data on Nina, to be compared to the predictions

Storytelling

- Think about how to build up the story top-down

- Think about the context various readers need to know about

Visualisation Challenge

2

Improving the Visualisation and Communication of the Model

Visualization

QGIS ® Trademark

- Use (for example) QGIS to create simple maps (basic instructions included with data)

Expected format & Presentation

- The output format is completely free ..

- .. but is ideally a visual static mockup of the full online storyboard, that you would want the results to be communicated through.

- A working prototype is obviously also fine, but not expected.

- Work in Powerpoint or any other software you like to create your output.

2

Improving the Visualisation and Communication of the Model

Use Cases

- **Philippine Red Cross operations**: interested in priority areas, where to send Early Assessment teams, where to start Early Response operations.

- **UN OCHA** (Office for Coordination of Humanitarian Affairs): more interested in overall picture / damage count. It determines the needed appeal for funding, etc. Also interested in priority areas though.

- **Shelter Cluster:** interested in priority areas, but also in absolute numbers of damaged houses, because they need to start to provide shelter kits and temporary housing and plan rebuilding.

- **General interest in the sector:** The result can also be of interest to not directly involved parties, who are interested in the technology, etc. These might not know everything about typhoons or about the humanitarian setting.

How to deal with various use cases

- It's OK to **choose a use case**, and focus on that, if you have particularly good ideas about a certain use case. Make clear what you're focusing on within your story though.

- However, it's overall OK (and possibly more realistic) to go for a **hybrid mix**, which caters to various use cases. In reality, it might make more sense to send out one product. Possibly by making clear within the storyboard that some sections are more/less interesting to specific audiences.

- In general, keep in mind that all audiences have relatively **low data literacy**. And in all cases, reason from the (various) readers. What context is required, etc?

Data Journalism Inspiration

- http://hno-syria.org/ (shown here)
- https://scroll.lab.nos.nl/een-halte-te-ver
- http://www.thebiggerpicture.online/
- http://www.gapminder.org/

Interested?

- Graduates / Internship opportunities
- Volunteering opportunities working directly on challenges like today
- Opportunities to be included in SIMS Roster for post-disaster Information Management deployments

CONTACT US

Leeghwaterplein 27
2521CV, The Hague, The Netherlands

info@510.global

twitter.com/510global

facebook.com/510global

510
PERSONALIZED HUMANITARIAN
AID THROUGH BIG DATA

AN INITIATIVE OF THE NETHERLANDS RED CROSS