Driving innovation inside and outside UNHCR

@rebecamorenog
Innovation Officer - Data Scientist
UNHCR Innovation Service
In 2017, an average of 44,400 people being forced to flee their homes every day bringing a total of 68.5 million forcibly displaced people in need of international protection. 25.4 of them have crossed a border, becoming refugees.
UNHCR has been innovating for years, However. . . .
...there were no formal structures to foster, capture & reward innovation
UNHCR Innovation Service provides spaces & structures for experimentation to occur within UNHCR
How can we provide solutions that cater to the well-being and dignity of the forcibly displaced?
Innovation ≠ Technology

Innovation = new ways of doing things

Product
Process
Policy
Paradigm
We respond to field challenges that are demand-driven, not supply-driven.

We search in-house, and externally for solutions that exist.

We collaborate to identify the best possible solution.

We test and refine those solutions.

We collaborate with stakeholders and decide whether or not to scale.
Data Innovation in UNHCR
Why is data innovation important for UNHCR?

Make evidence-based decisions
Better coordination between partners for improving quality of services
Accountability and transparency
What is data innovation in UNHCR?

Application of data science techniques, including -but not exclusively - the use artificial intelligence for analysing non-traditional data sets for advocacy, preparedness and operational response.
What does that actually mean?
Topic #1: Xenophobia and Hate Speech
Detecting xenophobia and hate speech towards refugees, with UN Global Pulse

Number of tweets analysed: 3,433,800 (Nice) + 297,506,445 (Munich) = 300,940,245

Geography: Worldwide

Hypothesis: advocacy-related

“Some people link the 2016-17 terrorist incidents in Europe with refugees”

Data Source: Twitter

Berlin truck terrorist yet another Muslim refugee. Merkel's policy is now a national security threat. Resign now. youtube.com/watch?v=rGsFGv...

VS.

What is indefensible is the creation, funding and arming of terrorist groups that led to the refugee crisis in the first place

Challenges

Technical

- Veracity
- Representativeness

Cultural:

- machine-trainers: native speakers with human rights background
- sarcasm/slang
Topic #2: 
Humanitarian Crisis Early Warning
Monitoring Violent Conflict Incidents & Displacement

DRC/Angola

Social media posts in Kasai Region about violent conflict-related incidents vs. Weekly arrivals UNHCR Angola

Challenges

Technical
- Connectivity
- Representativeness
- Literacy rates

Data Source: Twitter
Topic #3: Forced Displacement Movement (Population Flow)
We became exiled in other countries. We took on the name ‘refugees’. After we were once the children of pride and jasmine [in reference the the Jasmine-scented streets Syria is known for] Oh Syria, I swear on the life of those who created you. To you we will return.

Challenges
Technical
• Connectivity

Technical + Cultural
• Representativeness = not preferred Comms Channel
Monitoring intention of fleeing + asylum applications

Venezuela

Timeline: Asylum Applications vs. Relevant Posts

Data Sources:
Google Trends + Twitter

Related queries: political asylum (100), asylum in Spain (25), political asylum in the U.S. (15), political asylum in Spain (15), how to request political asylum (10)
Quantifying and validating official mixed-migration data, **with QCRI**

**QCRI data on 2018 movement**
- Venezolanos en Colombia: 980,000
- Venezolanos no Brasil: 81,000

Source: OCHA (2018) - 935,000

Source: RAMV Colombia (2018) - 819,034
**Project Jetson: Predicting Population Flow**

Algorithm example (to be read by the machine)

```
Bay_CurrentRegion = max(max(delay(Awdal_CurrentRegion, 1), delay(Bakool_CurrentRegion, 1), delay(Banadir_CurrentRegion, 1), delay(Bari_CurrentRegion, 1), delay(Bay_CurrentRegion, 1)), delay(Nugaal_CurrentRegion, 1), delay(Sanaag_CurrentRegion, 1), delay(Sool_CurrentRegion, 1), delay(Togdheer_CurrentRegion, 1), delay(Woqooyi_Galbeed_CurrentRegion, 1))
```

Model produced by machine:

```
Bay_CurrentRegion = max(max(delay(Awdal_CurrentRegion, 1), delay(Bakool_CurrentRegion, 1), delay(Banadir_CurrentRegion, 1), delay(Bari_CurrentRegion, 1), delay(Bay_CurrentRegion, 1)), delay(Nugaal_CurrentRegion, 1), delay(Sanaag_CurrentRegion, 1), delay(Sool_CurrentRegion, 1), delay(Togdheer_CurrentRegion, 1), delay(Woqooyi_Galbeed_CurrentRegion, 1))
```

**Challenges:**

- **Technical**
  - Computational power vs. data protection
  - Open source vs. time-series ML software
  - Open Data
  - Prediction time - data availability

**Jetson predictions**

**Predicting October**
- Time prior: 1 week
- Jetson: 9,062
- Actual Arrivals: 10,003

**Predicting November**
- Time prior: 2 weeks
- Jetson: 8,928
- Actual Arrivals: 7,750

**Predicting December**
- Time prior: 3 weeks
- Jetson: 9,468
- Actual Arrivals: 9,881
April 2018: 18 regions + Dollo Ado, Ethiopia

Prediction time: 3 to 4 weeks prior, 13 predicted
$R^2 = 0.70$ to 0.95, depending on region/model
among other projects and ideas that don’t require A.I....
Thank you!

unhcr.org/innovation

innovation@unhcr.org