

Finding the nexus between climate change, conflict and forced displacement

UNHCR Innovation

Innovation ≠ Technology

Innovation = new ways of doing things

- Product
- Process
- Policy
- Paradigm

Our process

We collaborate with stakeholders and decide whether or not to scale.



We test and refine those solutions



We respond to field challenges that are demand-driven, not supply-driven.



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

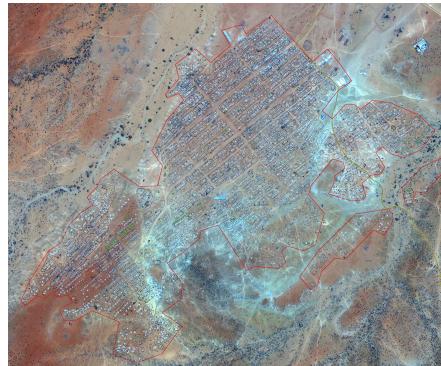


We collaborate to identify the best possible solution.

Institutional memory: 2011 humanitarian emergency (famine)



Dollo Ado, Ethiopia
Dec 31, 2010
Population: **40,479**



Dollo Ado, Ethiopia
July 19, 2011
Population: **115,224**



Dollo Ado, Ethiopia
31 Aug 2018
Population: **219,284**

The Challenge: ask an interesting question(s)

Somalia operation: IDPs

- Where are they moving? = categorical variable
- When are they arriving? = numerical variable
- **How many people are moving? = numerical**

Ethiopia operation: Refugees

- Dollo Ado, Melkadida sub-office: given 2017 conditions (drought/conflict) and with prior institutional memory of displacement from 2011 with similar conditions...
 - **Are we going to receive the same amount of refugees? = numerical**



Why JETSON?

- Help 2 operations (Ethiopia-Dollo and Somalia) **make evidence-based decisions** and **adequately plan/prepare** contingencies
- Set a **precedent** for this type of predictive analytics work in **humanitarian sector** for coordination and compilation of data
- **Open data and model** for other data scientists, computer scientists, programmers can edit, add, and improve upon

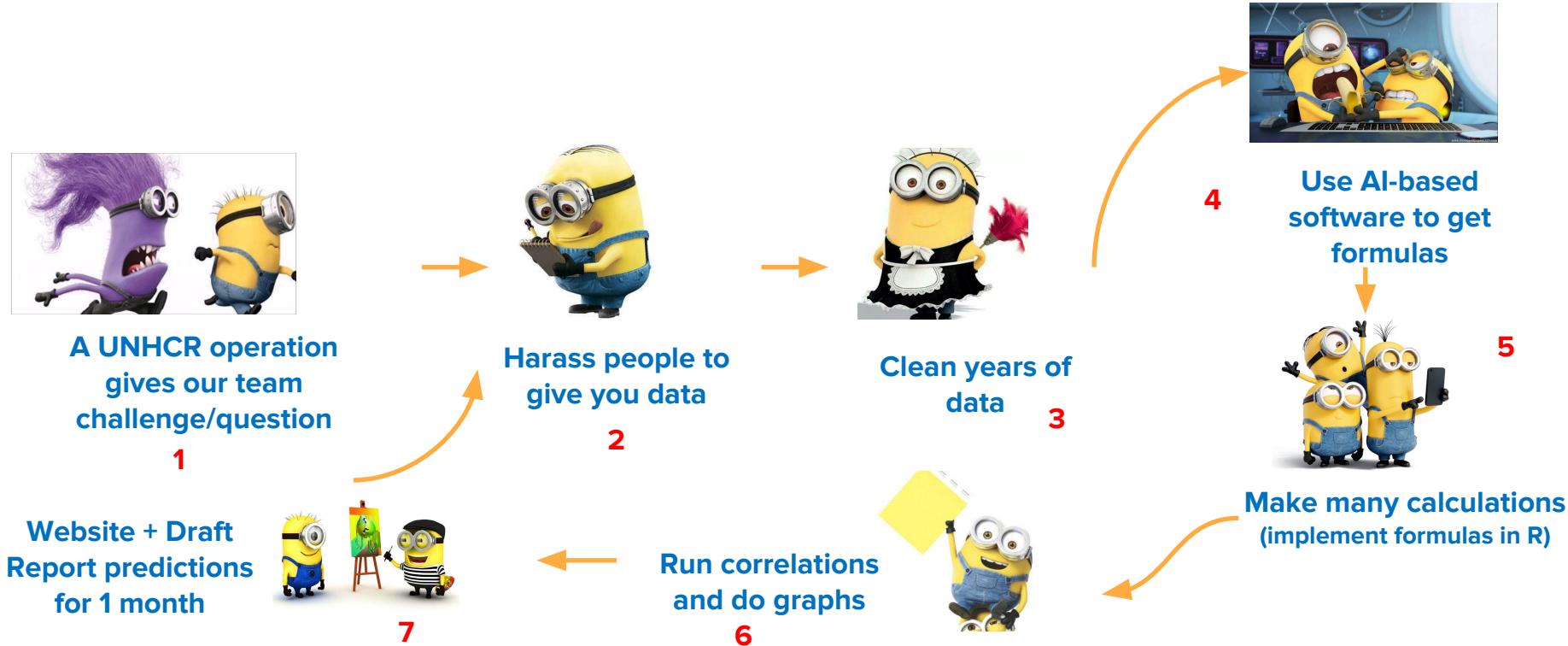
Secondary Data Review

Author/Organization	Title	Topic (Data Science/Stats)
IDMC	<i>Assessing drought displacement risk for Kenyan, Ethiopian and Somali pastoralists</i>	Agent-based model (ABM)
Brunel University London • Groen, D. Bell, D, & Suleimenova, D. (currently working as partners)	<i>A generalized simulation development approach for predicting refugee destinations</i>	Agent-based model (ABM)
Carnegie Mellon/Singapore Management University • Cheng, S., Lin, L. & Carley, K.	<i>An agent-based approach to human migration movement</i>	Agent-based model (ABM)
U.S.A. National Research Council	<i>Model-based approaches to estimating migration flows</i>	Bayesian models Regression models Simulation models
George Mason University • Gulden, T., Harrison, J., Crooks, A.	<i>Modeling Cities and Displacement through an Agent-based Spatial Interaction Model</i>	Agent-based model (ABM)
University of Pennsylvania • Mellers, et.al, Merkle, E. & Tetlock, P.	<i>The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics</i>	Forecasting vs. Intelligence Analysis
Berkeley University • Hsiang, S., Burke, M. & Miguel, E.	<i>Quantifying the Influence of Climate on Human Conflict</i>	Linear Regression

Research Assumptions

Project Phase	Assumption	Way we tested	Result
Challenge Definition	"PoCs in Somalia are fleeing from conflict areas"	<ul style="list-style-type: none"> Desktop research: literature review on historical conflict in Somalia Data on conflict (ACLED) graph trend 	<ul style="list-style-type: none"> True: 7 years of data historical conflict doubled and all literature pointed out on ethnic, inter-tribal and extreme groups conflict
	"PoCs in Somalia are going to places where humanitarian assistance is being provided"	<ul style="list-style-type: none"> Desktop research (data on humanitarian assistance) Semi-structured interviews (phone calls) with partners on their activities in certain regions 	<ul style="list-style-type: none"> Inconclusive: not enough data to conclude assistance is a pull factor. However, partners in their interviews mentioned the fact cash is actually preferred by PoC (qualitative/phone call)
	"PoCs movement is also affected by external factors (e.g. drought/floods)"	<ul style="list-style-type: none"> Focus groups with PoCs Interviews with different operational partners 	<ul style="list-style-type: none"> True AND found out because of focus groups we were missing a variable: goats/market prices
[Solution] Ideation	"A machine/computer program could help us predict how many people are going to be arriving/moving to a particular region"	<ul style="list-style-type: none"> Desktop research: academic papers on machine learning Read blogs on other agencies trying to do similar work (predictive analytics) 	<ul style="list-style-type: none"> True: academic research and financial sector research on predictive analytics by analyzing historical data and getting accurate results.
Experimentation	"The machine can predict PoC arrivals with 75-100% of accuracy"	<ul style="list-style-type: none"> Build the computer program and ran it - literally - 65 times, until it provided a good result. 	<ul style="list-style-type: none"> False: after running it for 65 times, the machine can provide the arrivals on a 60% to 110% accuracy (difference between actual arrivals vs. machine predicted)
Implementation	"The predictions will help Somalia operation to do better planning/preparedness" (e.g. food distribution, shelter planning)"	<ul style="list-style-type: none"> Calls with Somalia operation on a monthly basis 	<ul style="list-style-type: none"> False: the operation uses it to fact-check if the arrivals are actually true then they allocate resources (send protection partners to specific areas)
Scaling	"The predictions can help the humanitarian sector to share their data and do more regions or predict other variables"	<ul style="list-style-type: none"> Workshop with partners in Kenya (TBC) 	Working on it :) you're part of it

The process (in a nutshell)



The Process



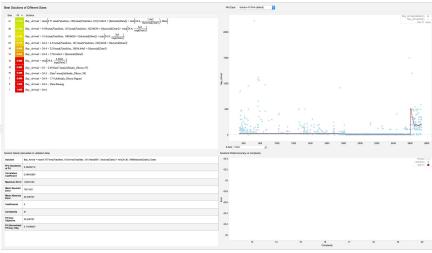
1

A screenshot of a software interface titled "Arrival Data - Overview". It displays a map with several locations marked and a table of data below it. The table includes columns for "Date", "Region", "Arrivals", "Refugees", "Migrants", "IDPs", and "Other". The data shows daily arrivals from various regions.

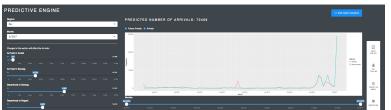
2

A screenshot of a software interface showing a table of arrival data. The table has columns for "Date", "Region", "Arrivals", "Refugees", "Migrants", "IDPs", and "Other". Some rows are highlighted in red or blue, likely indicating specific data points of interest.

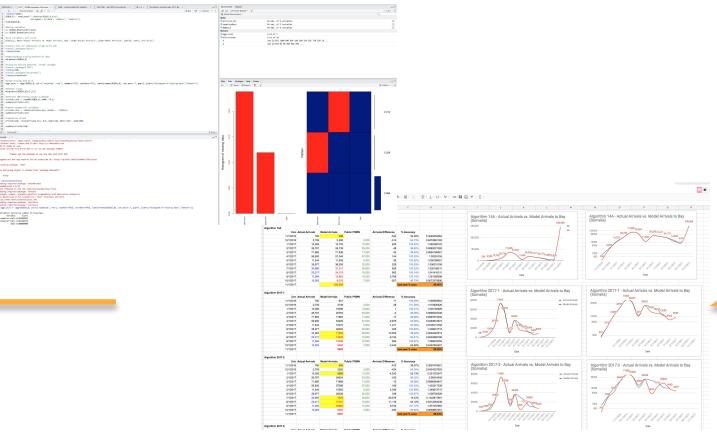
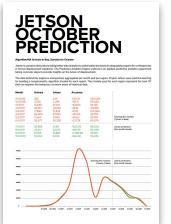
3



4



7



6

```
PREDICTION_CODE
A <- 1.52191998816997*current.long((t-1),"Awad_CurrentRegion"
B <- 0.73695641649494*median(before.long((t-11):(t-1),"Bay_BeforeRegion"))
C <- 0.7689593933939393*mean(before.long((t-15):(t-1),"BeforeRegion"))
D <- future(cars((t-1)),"Arriving_Refugees")*lengths.long((t-15),"Galgeaud_Conflict")
E <- random(worl.long((t-11):(t-1)),"Refugee_Refugees")
XA <- rivers.long((t-10),"Shabelle_Ohexe_Jouharstation_Shabelle_River"]
XB <- 3.61920831646864*rain.long((t-16),"Jubbada_Hoose_rain"]
XC <- rain.long((t-8),"Woqoyi_Gobley_Rain"]
NOTA <- rain.long((t-8),"Jubbada_Hoose_Fatalities"]
NOTB <- not(NOTA)

if(is.na(NOTA)(XD < 1)
else if(NOTA)(XD < 0)
else(XD < 1)

Z <- fatalities.long((t-1),"Jubbada_Hoose_Fatalities"]
K <- less(XA, XB)

if (is.na(XA) || is.na(XB))(Z < 0 )
else if(is.na(XA))(Z < 0)
else if(is.na(XA) || XA < XB)(Z<1)
else (Z < 0)

FIN <- sum(A , B , C , D , E*K , XC*XD , Z , na.rm=TRUE)
```

Data Sources

Organization - Unit	Topic	Data type	Frequency	Original Format, Privacy
ACLED	Conflict	<ul style="list-style-type: none"> Violent incidents Fatalities 	Weekly	CSV, public API
FAO - SWALIM	Climate	<ul style="list-style-type: none"> River levels/discharge Rain 	Daily (sensor data)	Web table, public <ul style="list-style-type: none"> We built parser to extract
FAO- FSNAU	Market Prices	<ul style="list-style-type: none"> Local Goat price Water drum price 	Monthly	CSV, public
WMO- ICPAC	Weather	<ul style="list-style-type: none"> Regional Weather Forecast 	Dekadal Monthly Seasonal	PDF report, public
PRMN - Somalia	Population	<ul style="list-style-type: none"> IDP movement 	Daily	CSV, internal <ul style="list-style-type: none"> We built a parser to aggregate
UNHCR Ethiopia	Population	<ul style="list-style-type: none"> Registration data 	Daily	CSV, internal <ul style="list-style-type: none"> We built a parser to aggregate

Methodology & Methods

Methods:

I. Data wrangling

- Dplyr (R), time-series format

II. Data imputation:

- multivariate imputation by chained equations (MICE)

III. Statistical modelling

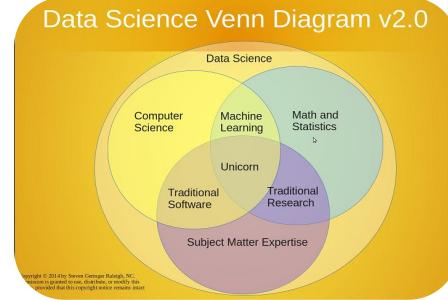
Model selection - 1st phase:

- R²
- Mean Absolute Error (MAE)

Model selection - 2nd phase:

- Simple Correlation (complete.obs, last 20 months only)
- Simple correlation (complete.obs, all 7 years)
- Bayesian information criterion (BIC)
- Akaike Information criterion (AIC)

Overall research methodology:
data science process



IV. Computer Science/Machine Learning

Model selection - 2nd phase:

- Accuracy computer computation (ACC) or percentage computer computation (PCC)

Algorithm transformation

- Math package (python)
- Equation package (python)
- String package (python)

Data Visualization

- R-shiny (app)

In more detail ...

Challenge

What can machine learning do to solve this **challenge**?

How do we deal with lack of data?

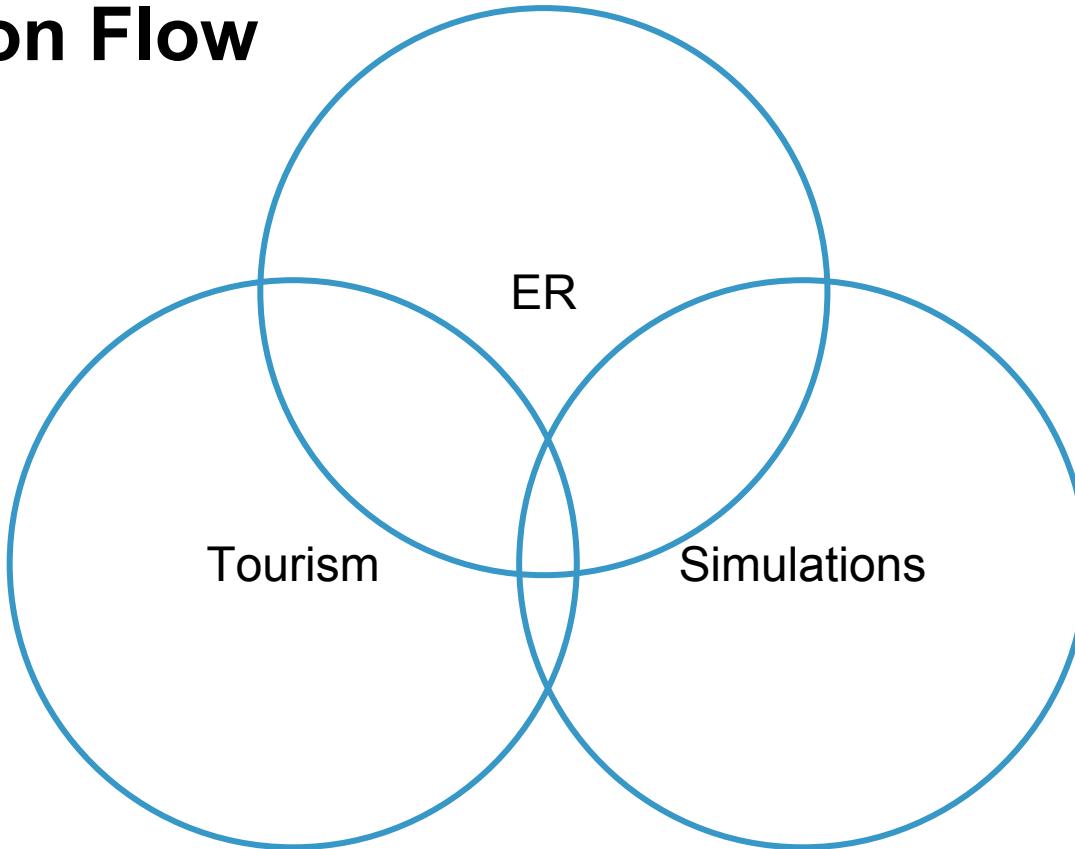
Can we predict at least for one month in advance?

Can we detect push and pull factors that drive internal displacement?

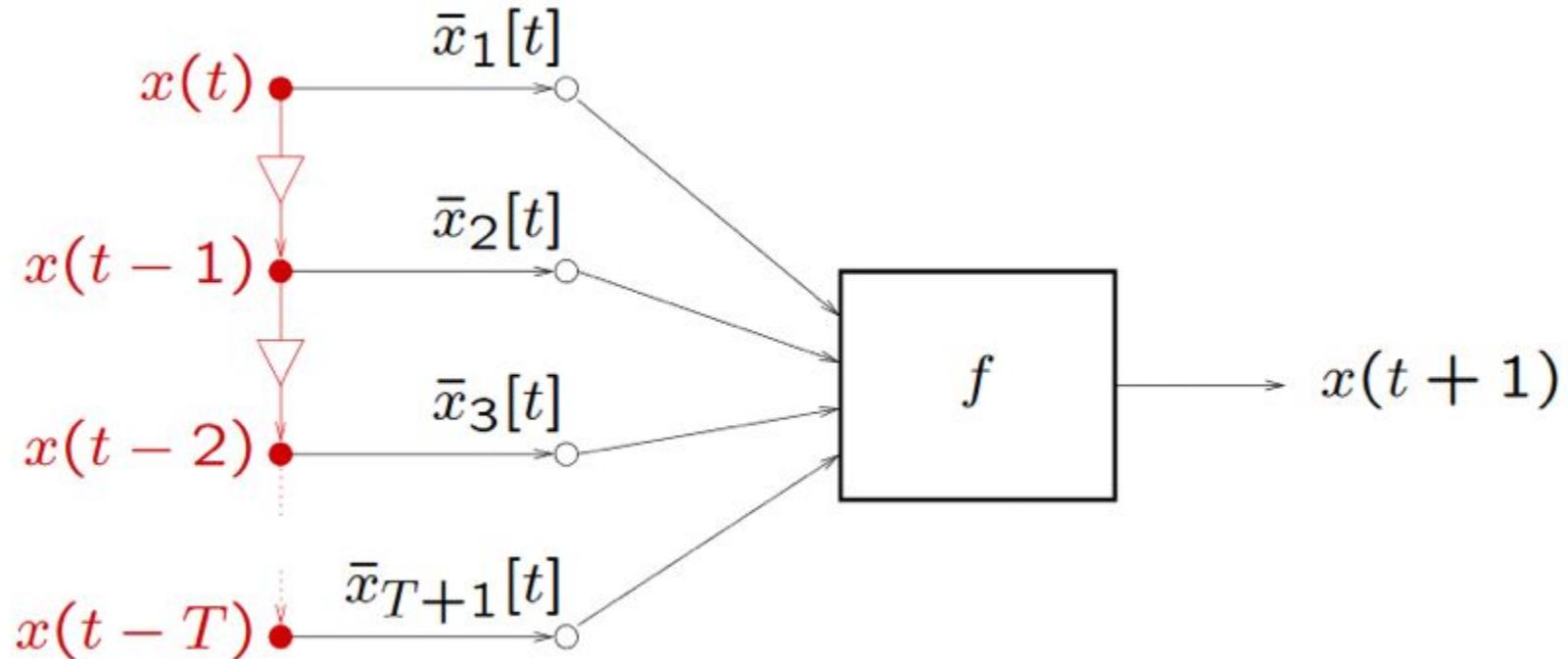
Research and Theories

- Disassociated theories - cognitive process, economic giving receiving region
- Geographical theories – distance, cost of transportation
- Post–phenomenon theories – cannot assist for prediction

Population Flow



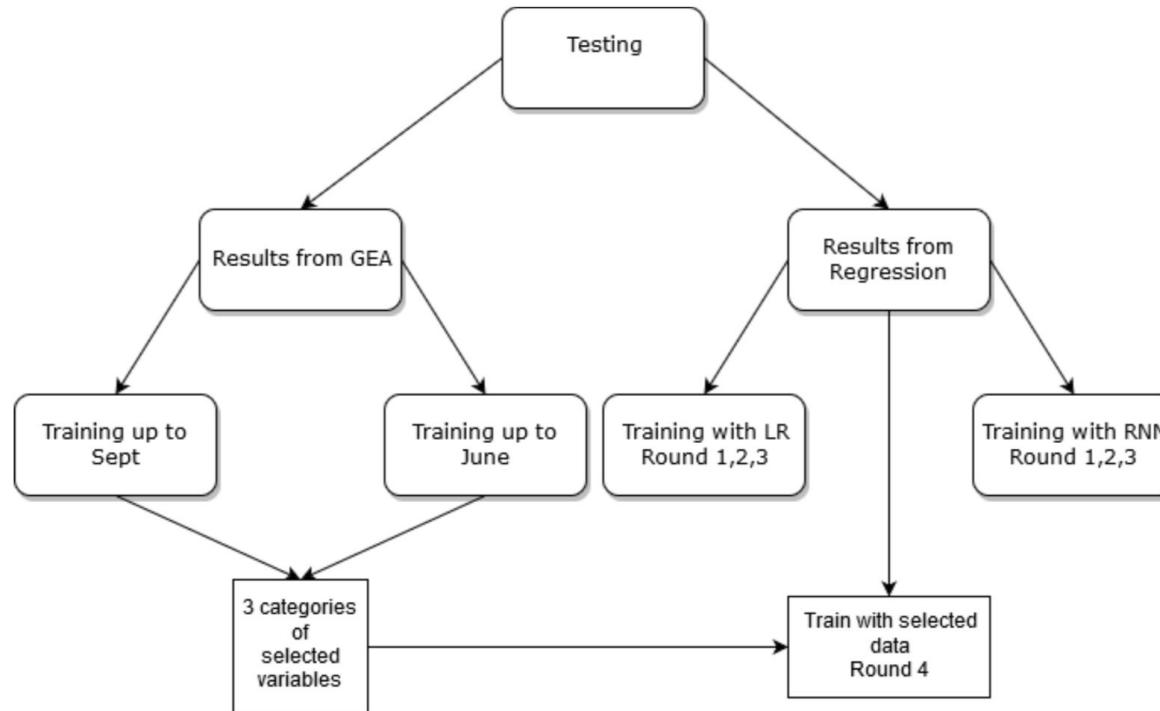
Multivariate time series forecasting



Machine Learning on Time Series Forecasting

- Short - Long Term Memory
Recurrent Neural Networks
- Genetic and Evolutionary
Algorithms

Structure of testing



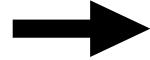
Results for Banadir

DATE	10/17	11/17	12/17	01/18	02/18	03/18	04/18
Accuracy	123%	112%	112%	98%	94%	39%	156%

The first test: Bay Region

Algorithm example (to be read by the machine)

```
Bay_CurrentRegion = max(max(delay(Awdal_CurrentRegion, 1),  
0.00012717740817059*delay(Bakool_WaterDrumPrice,  
1)*delay(Awdal_CurrentRegion, 1) +  
log(delay(Jubbada_Hoose_rain,  
1)) *snn(delay(Awdal_CurrentRegion, 1), 11) +  
2.11983998917739e-6*delay(Mudug_BeforeRegion,  
1)*delay(Shabeellaha_Dhexe_BeforeRegion,  
1)*sma(delay(Nugaal_FutureRegion, 1), 10) +  
delay(Sool_BeforeRegion, 15)) -  
delay(Jubbada_Hoose_FutureRegion, 1),  
0.805982905429227*delay(Awdal_Conflict,  
1)*2*delay(Sool_Fatalities, 1)*delay(Bari_FutureRegion,  
1))
```



Model produced by machine:

```
Bay_CurrentRegion =  
max(max(delay(Awdal_CurrentRegion, 1),  
0.00012717740817059*delay(Bakool_WaterDrumPrice,  
1)*delay(Awdal_CurrentRegion, 1) +  
log(delay(Jubbada_Hoose_rain,  
1)) *snn(delay(Awdal_CurrentRegion, 1), 11) +  
2.11983998917739e-6*delay(Mudug_BeforeRegion,  
1)*delay(Shabeellaha_Dhexe_BeforeRegion,  
1)*sma(delay(Nugaal_FutureRegion, 1), 10) +  
delay(Sool_BeforeRegion, 15)) -  
delay(Jubbada_Hoose_FutureRegion, 1),  
0.805982905429227*delay(Awdal_Conflict,  
1)*2*delay(Sool_Fatalities, 1)*delay(Bari_FutureRegion,  
1))
```



1 region

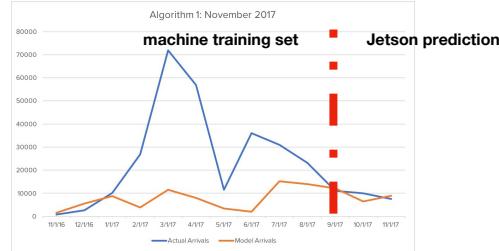


Predicting October

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

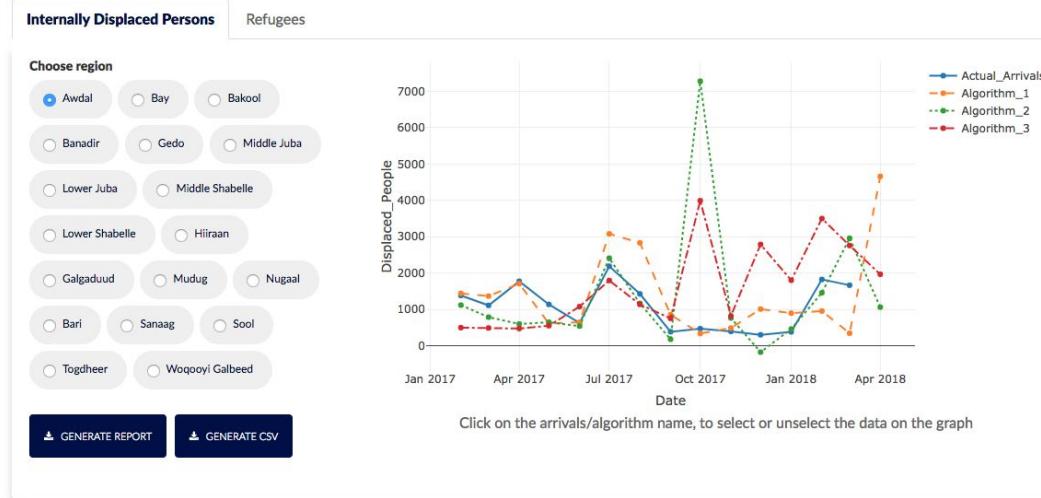
Time prior: 3 weeks

Jetson: **9,468**

Actual Arrivals: **9,881**

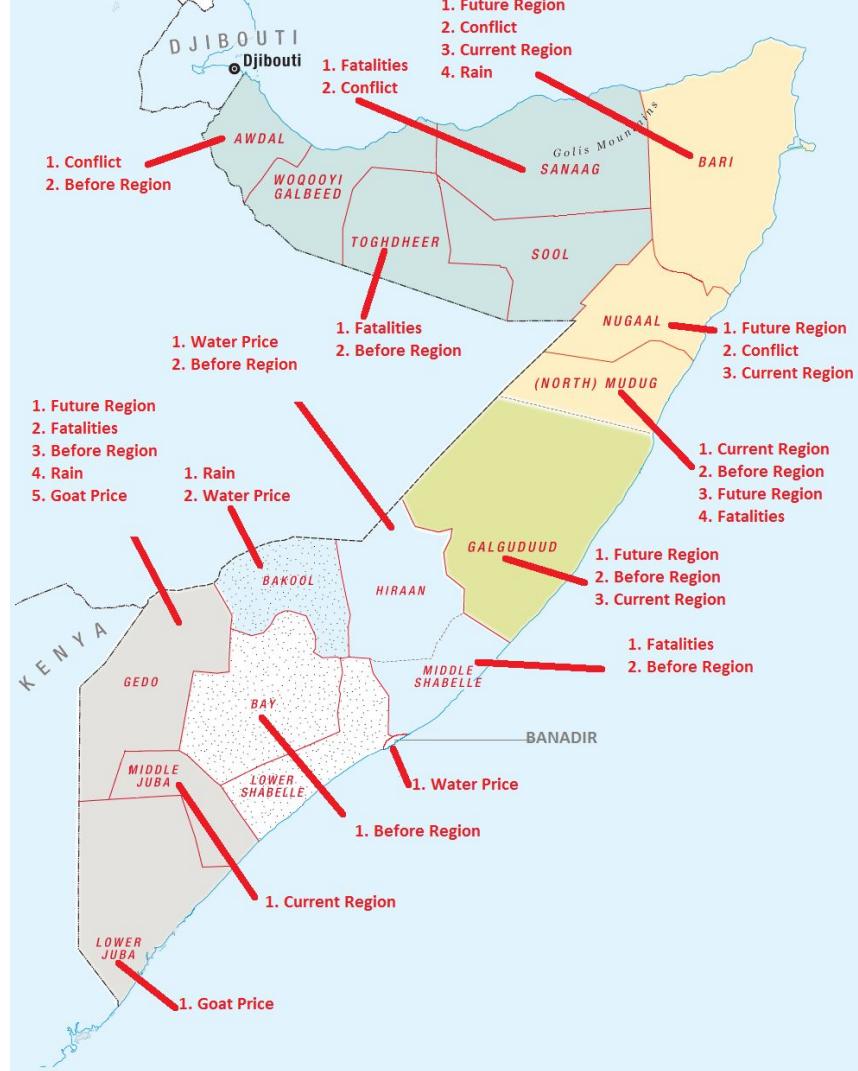
October 2018: 13 regions predicted

(out of 19, including Dollo)
time prior: 3-4 weeks



Most Influencing Factors

- Concatenating the Models
- What increases or decreases the number of arrivals?
- Sensitivity
- The answer is different per region



Main movement ‘predictors’ - sensitivity mapping

Variables Prevalence in 380 Algorithms

PREDICTORS (what variables are influential in these regions displacement, either in their own region or neighboring regions?)						
REGION	Violent Conflict	Fatalities	Anomalies Rain patterns	Anomalies Rivers affectations	Goat prices*	Water drum prices
Awdal		yes	yes			
Bakool	yes	yes		yes		
Banadir	yes	yes	yes	yes	yes	yes
Bari	yes	yes	yes			
Bay	yes	yes		yes		yes
Galgaduud	yes	yes	yes	yes	yes	
Gedo	yes	yes	yes	yes		
Hiiraan	yes	yes	yes			
Middle_Juba	yes	yes				
Lower_Juba	yes	yes	yes	yes	yes	yes
Mudug	yes	yes			yes	
Nugaal	yes	yes		yes	yes	yes
Sanaag	yes			yes		yes
Middle_Shabelle	yes	yes	yes	yes		
Lower_Shabelle	yes	yes	yes	yes		yes
Sool	yes	yes				
Togdheer	yes	yes				yes
Woqooyi_Galbeed	yes	yes		yes		

Somalia is an unstable country, we probably missed a lot of sensitive data while trying to model Internal Displacement . . .

Thank you!



unhcr.org/innovation

innovation@unhcr.org