Project Jetson

Katherine Hoffmann Pham

August 15, 2019

Data Fellow, UN Global Pulse
Ph.D. Candidate in Information Systems, NYU Stern School of Business

Overview

About UN Global Pulse

- UN Global Pulse: "innovation initiative of the UN Secretary-General . . . to accelerate discovery, development and scaled adoption of big data innovation for sustainable development and humanitarian action"

- Model: partner with other UN entities on projects to demonstrate the value of big data

About UN Global Pulse

- Motivation:
  - SDGs: Timely, cost-effective monitoring of 169 targets
  - Appeal of "always-on", "real-time" big data sources
- Sample projects:
  - Measuring Poverty With Machine Roof Counting
  - Making Ugandan Community Radio Machine-Readable

About the Fellows Program

Global Pulse Data Fellows Program:

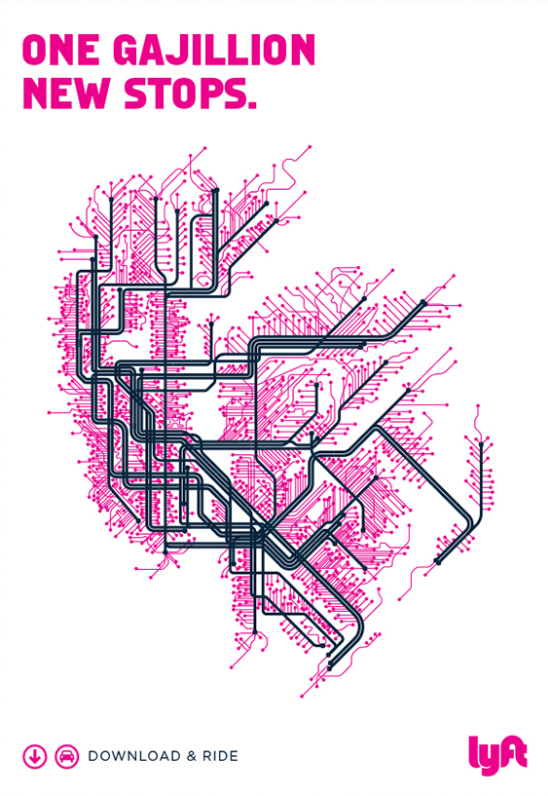
Match PhD students with UN entities for 6-12 months

- 40+ projects applied from within UN system
- 8 projects selected from first cohort
- Example projects:
  - Parsing radio broadcasts for disease surveillance (WHO)
  - Identifying refugee settlements in satellite data (UNOSAT)
  - Exploring the evolution of #metoo movement on social media

About Me

Thesis: Digital Future of Mobility

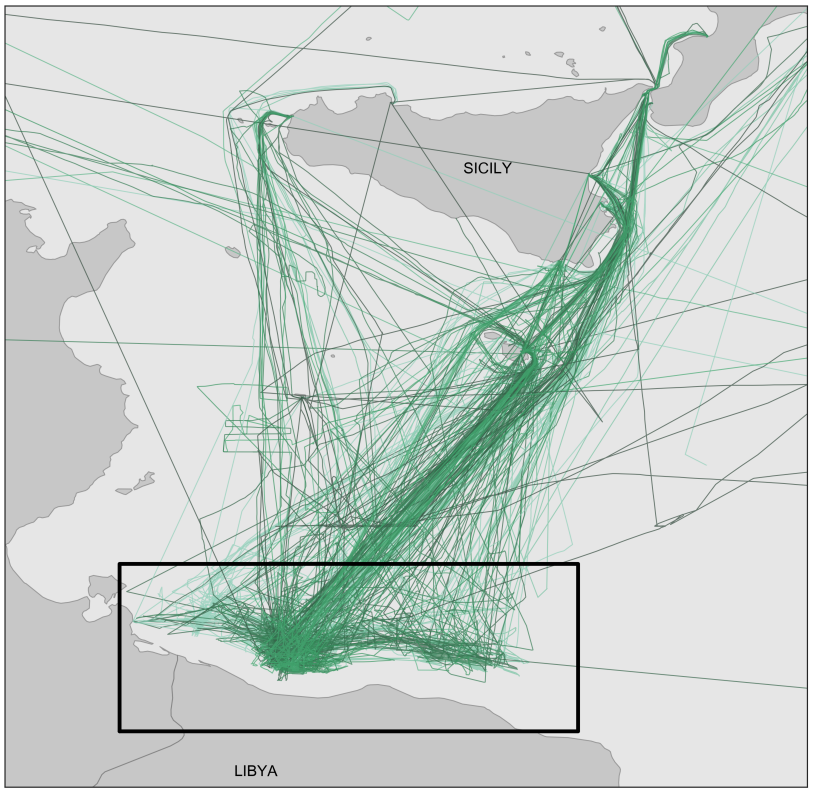
- Impact of Uber/Lyft on public transportation in NYC
- Nudges to improve parking among e-scooter users

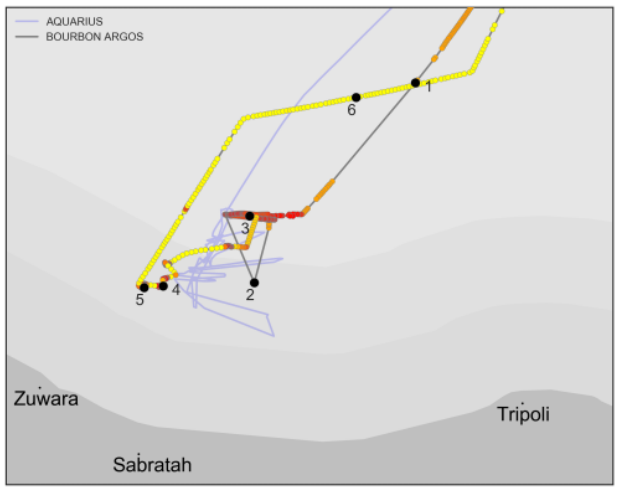


About Me

Projects: Big Data for Social Good

- Migration in the Central Mediterranean using AIS data



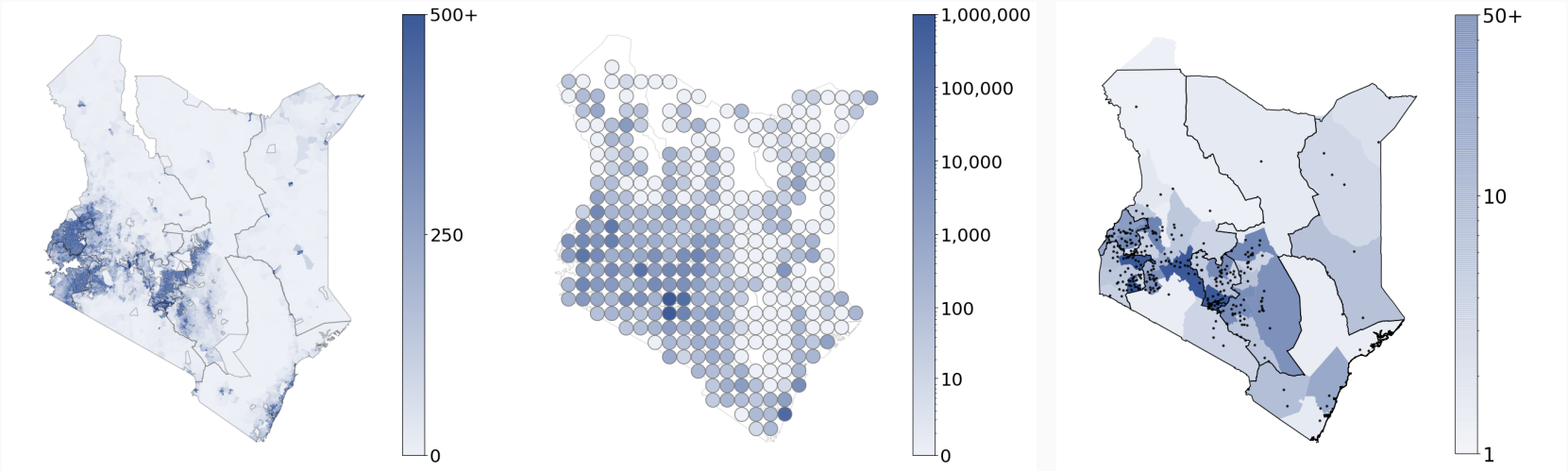


1 10-01 16:18 "@msf_sea: update: the bourbon argos is on its way to the sar zone and should be in position …"
2 10-02 09:25 "@msf_sea: breaking: the #argos has just rescued 26 #people from this tiny speck of a wooden boat …"
3 10-02 18:23 "@msf_sea: update: the #dignity1 and bourbon argos have transferred those rescued today to @savethechildren and @moas_eu …"
4 10-03 11:15 "@msf_sea: then, just as the sun came up the #argos crew rescued this lilting and over crowded boat. followed by 4 more …"
5 10-03 13:27 "@msf_sea: update: the rescues just keep coming on the bourbon #argos. with 7 boats rescued so far the deck is filling fast …"
6 10-03 18:17 "@msf_sea: with 1019 men, women and kids on board there is not a centimetre spare on the #argos …"

About Me

Projects: Big Data for Social Good

- Facebook advertising data for demographic estimates in Kenya



About Me

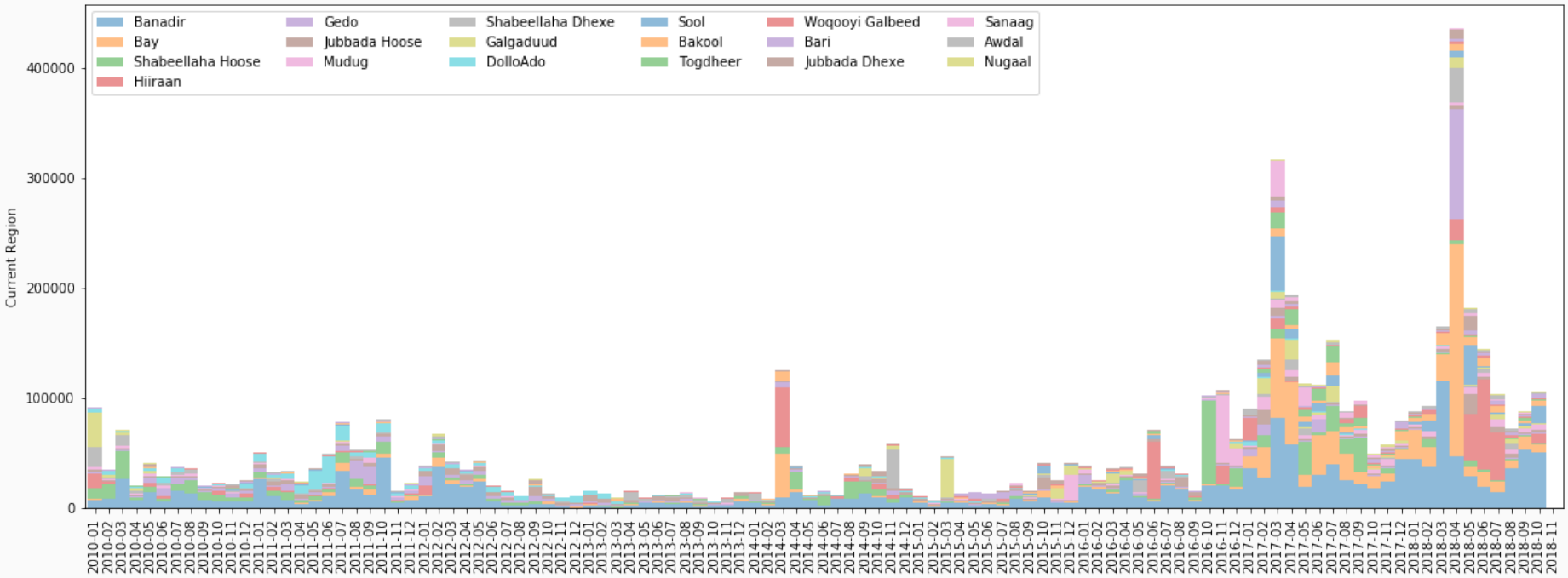
Training: BIGSSS Summer School on Migration

- How do people model movement?
  - Agent-Based Models (rules + simulation)
  - Gravity Models (flows between two regions)
- What is the focus of current research?
  - Host community based (integration, political orientation)
  - Theoretical (climate change, resilience, and agency)
  - Global models (country-level flows)

Project Jetson

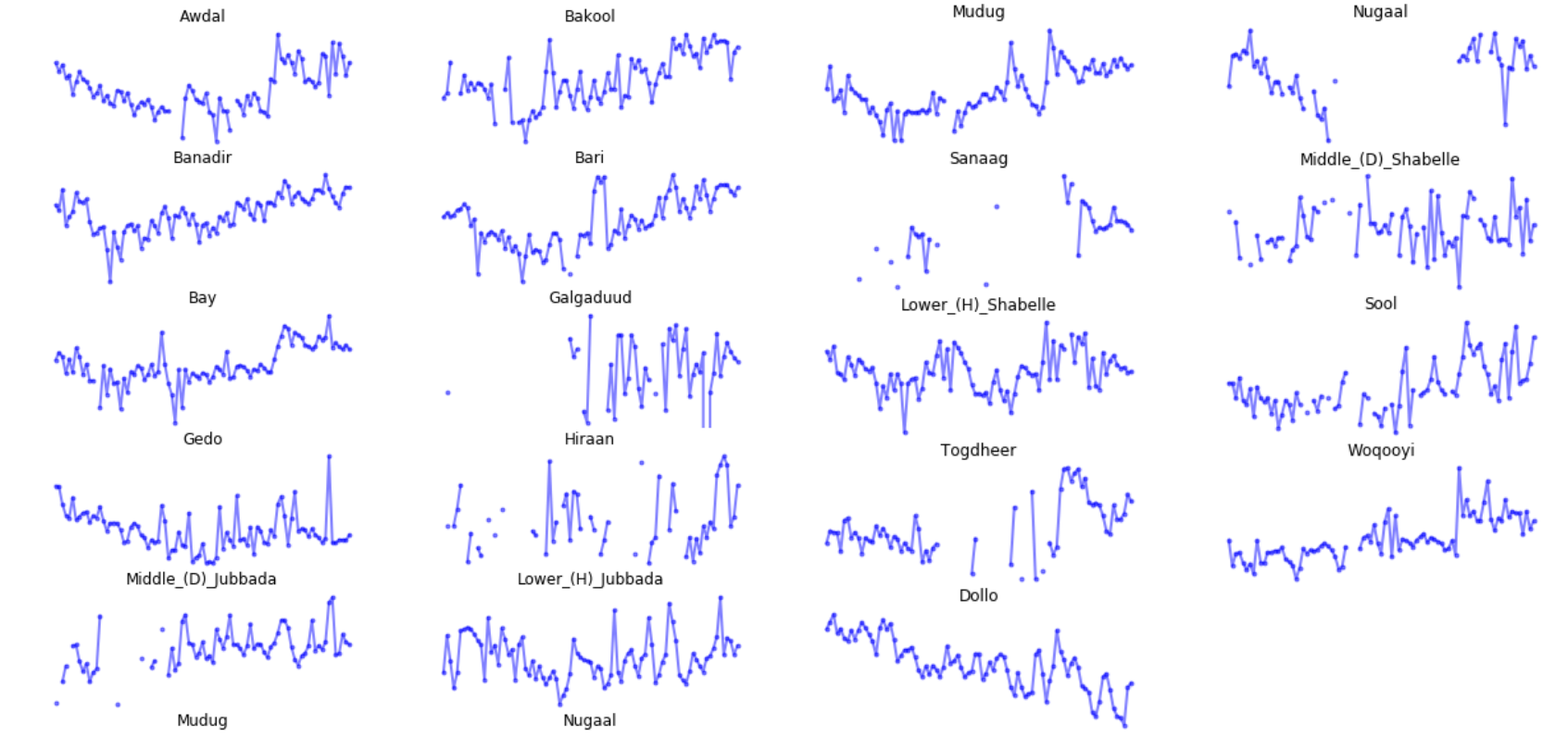
Research Questions

Can we use artificial intelligence to forecast the monthly volume of IDP/refugee arrivals by region in Somalia?



Research Questions

Can we use artificial intelligence to forecast the monthly volume of IDP/refugee arrivals by region in Somalia?



Research Questions

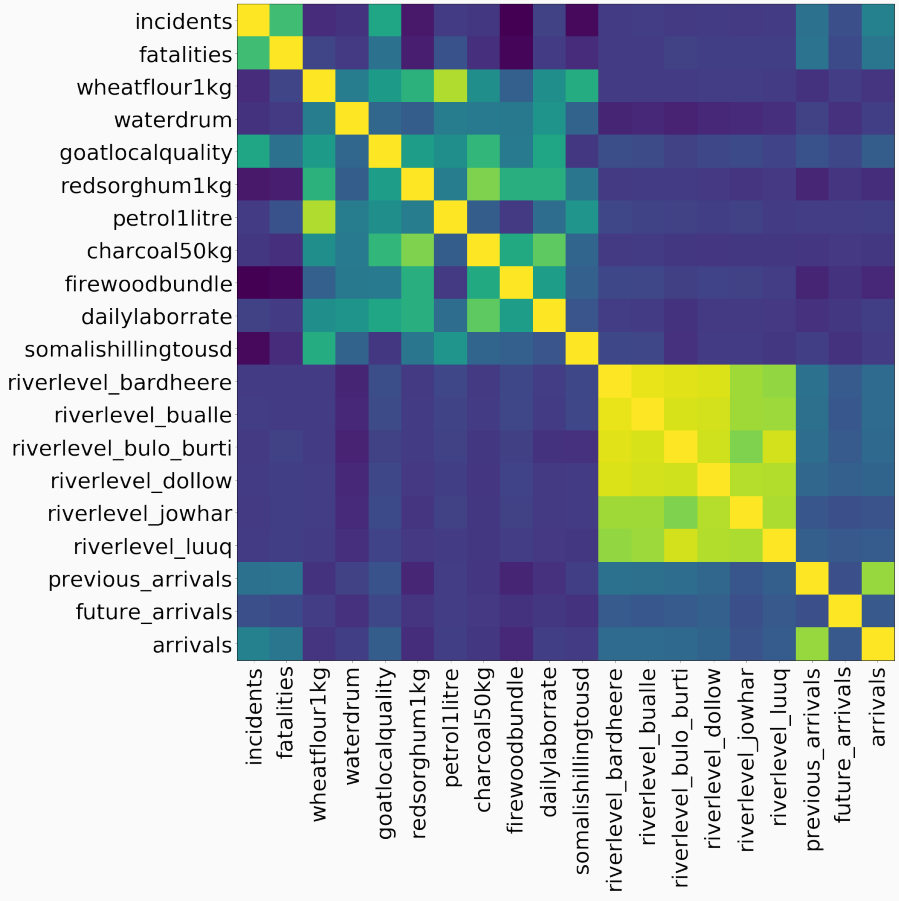
Can we use artificial intelligence to forecast the monthly volume of IDP/refugee arrivals by region in Somalia?

- How far in advance:
  1 months? 3 months?
- How does contextual data
  improve predictions?
- How best to model this
  problem: arrivals? flows?
  changes?

Research Questions

Can we use artificial intelligence to forecast the monthly volume of IDP/refugee arrivals by region in Somalia?

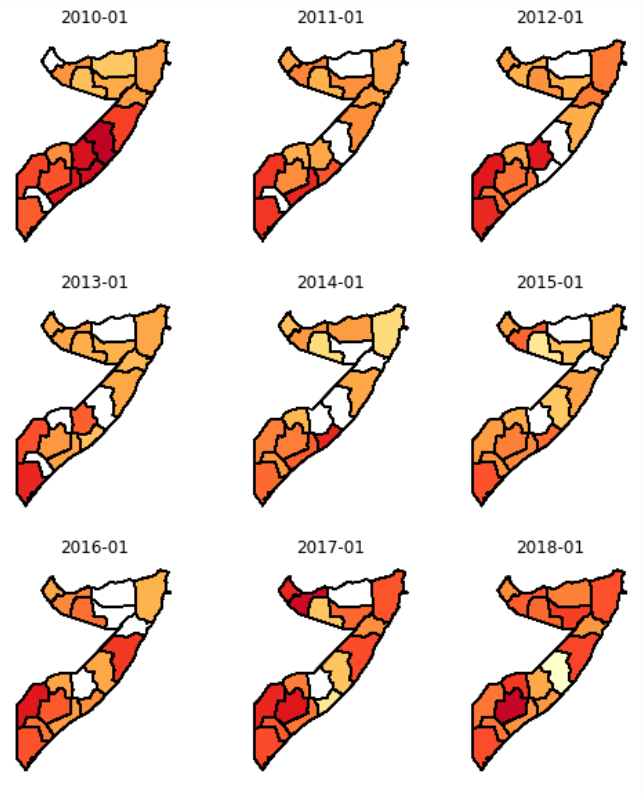
- How far in advance:
  1 months? 3 months?
- How does contextual data improve predictions?
- How best to model this problem:
  - arrivals?
  - flows?
  - changes?



Research Questions

Can we use artificial intelligence to forecast the monthly volume of IDP/refugee arrivals by region in Somalia?

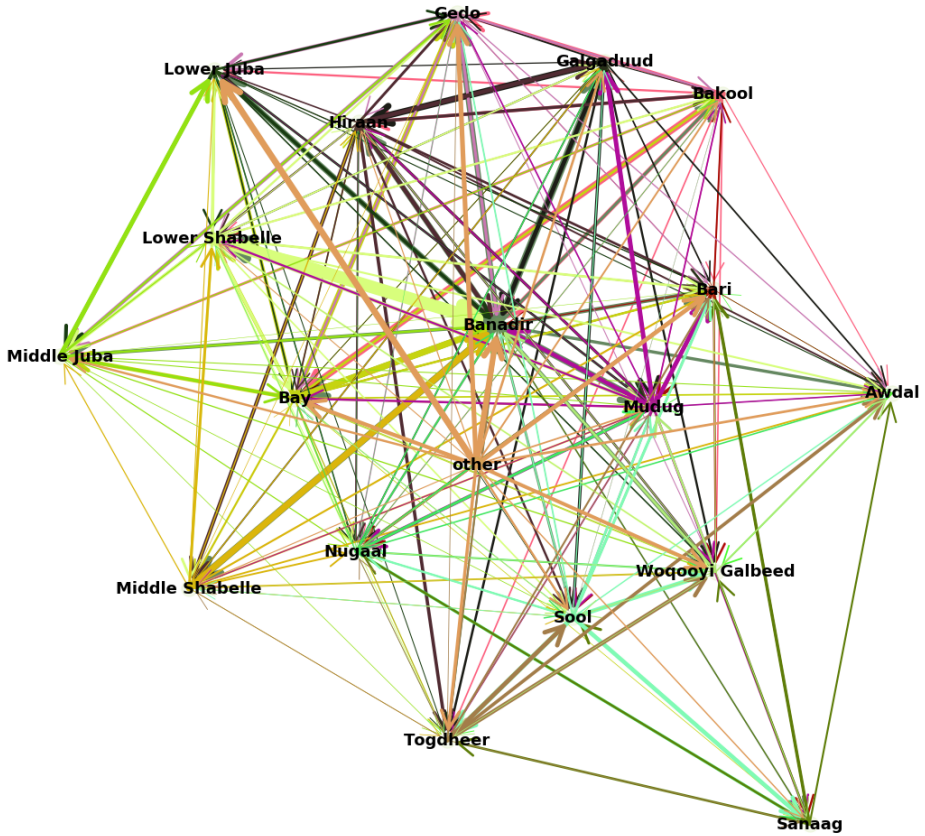
- How far in advance:
  1 months? 3 months?
- How does contextual data improve predictions?
- How best to model this problem: arrivals? flows? changes?



Research Questions

Can we use artificial intelligence to forecast the monthly volume of IDP/refugee arrivals by region in Somalia?

- How far in advance:
  1 months? 3 months?
- How does contextual data improve predictions?
- How best to model this problem: arrivals? flows? changes?

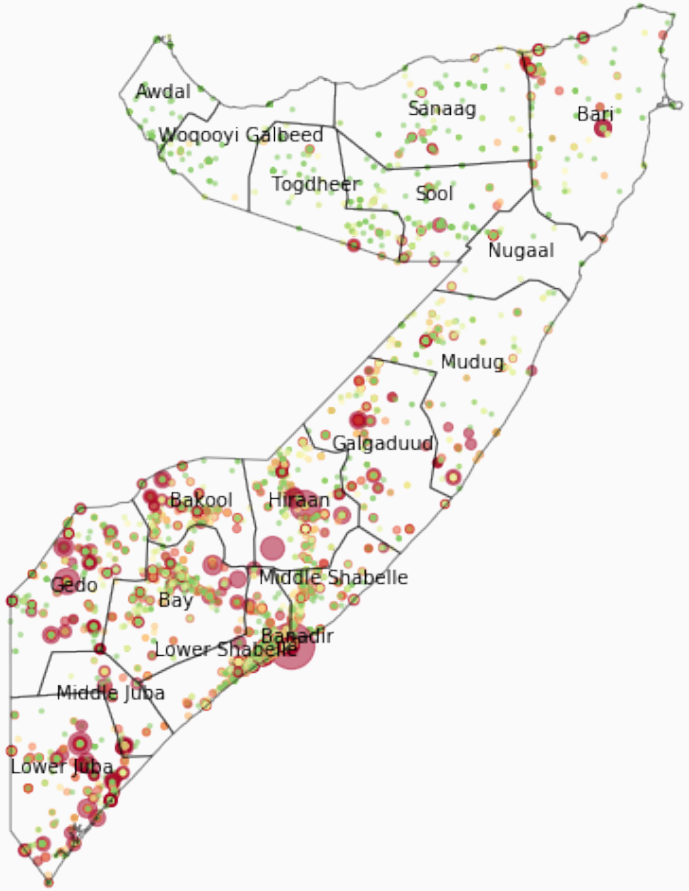


Experiment 1

- Data extracts are gathered and compiled monthly
  Can this be automated?
- $18 + 1$ regions $\times$ 23 models per region
  Can we implement a single model across all regions?
- Goal is to predict arrivals one month in advance
  Can we predict 3 months in advance?
- Models implemented in Eureqa (equations)
  Can we explore alternative tools?
- Each month, 3 winners are chosen based on previous month
  Can this be automated?
- Models are evaluated using several characteristics
  Can we identify the most important metric for the team?

Data Sources

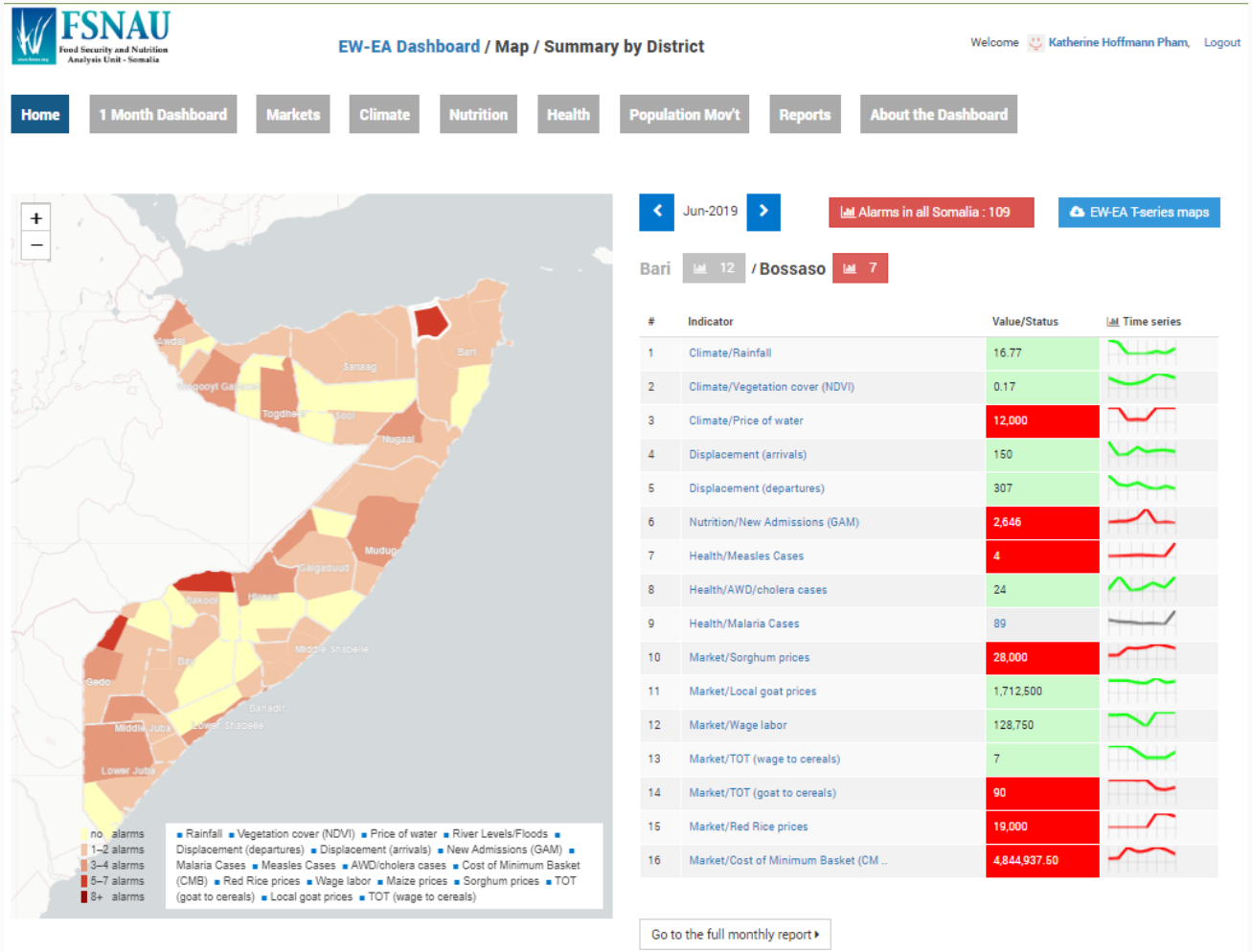
ACLED: conflict incidents and fatalities



Data Sources

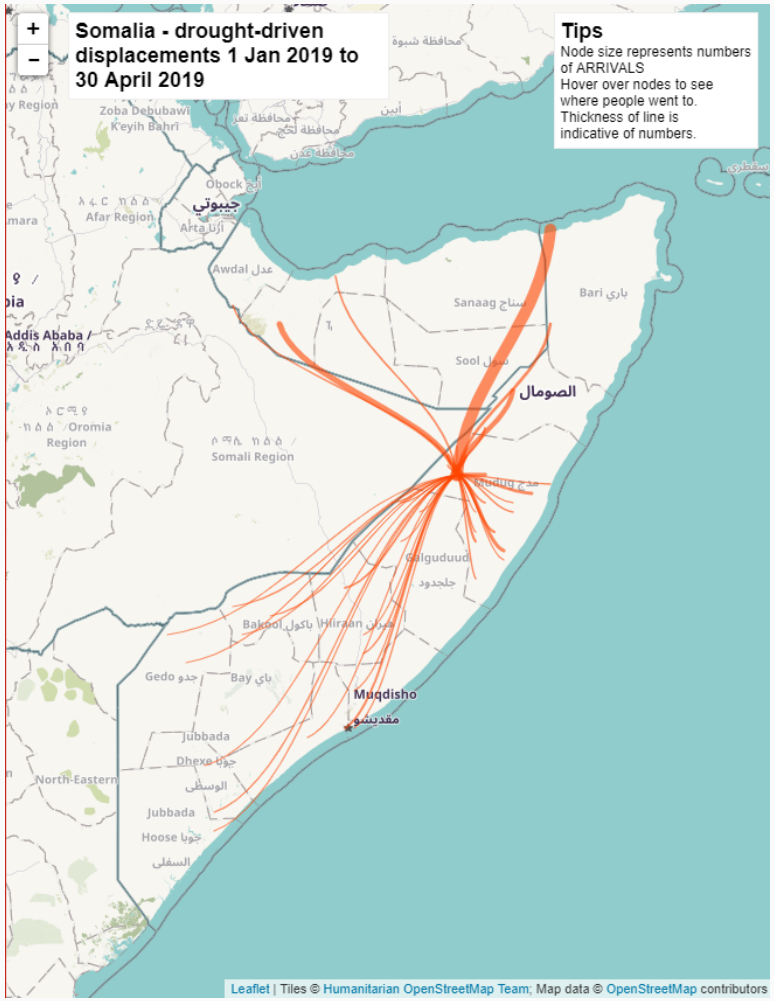
FSNAU Early Warning Early Action Dashboard:

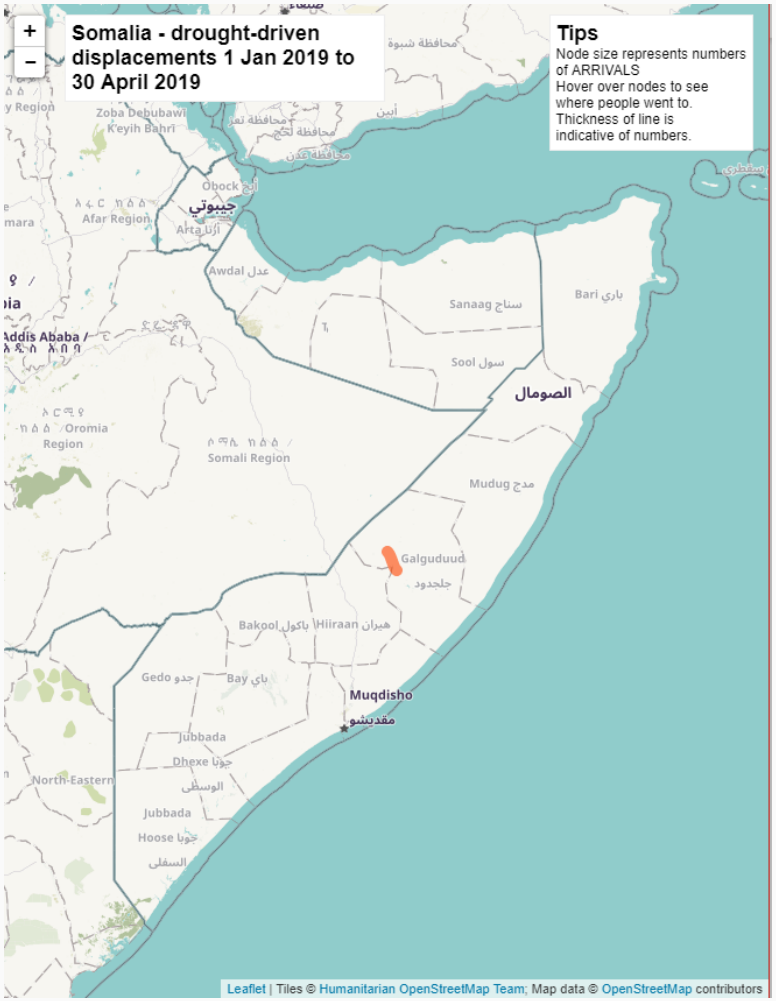
prices; climate; health



Data Sources

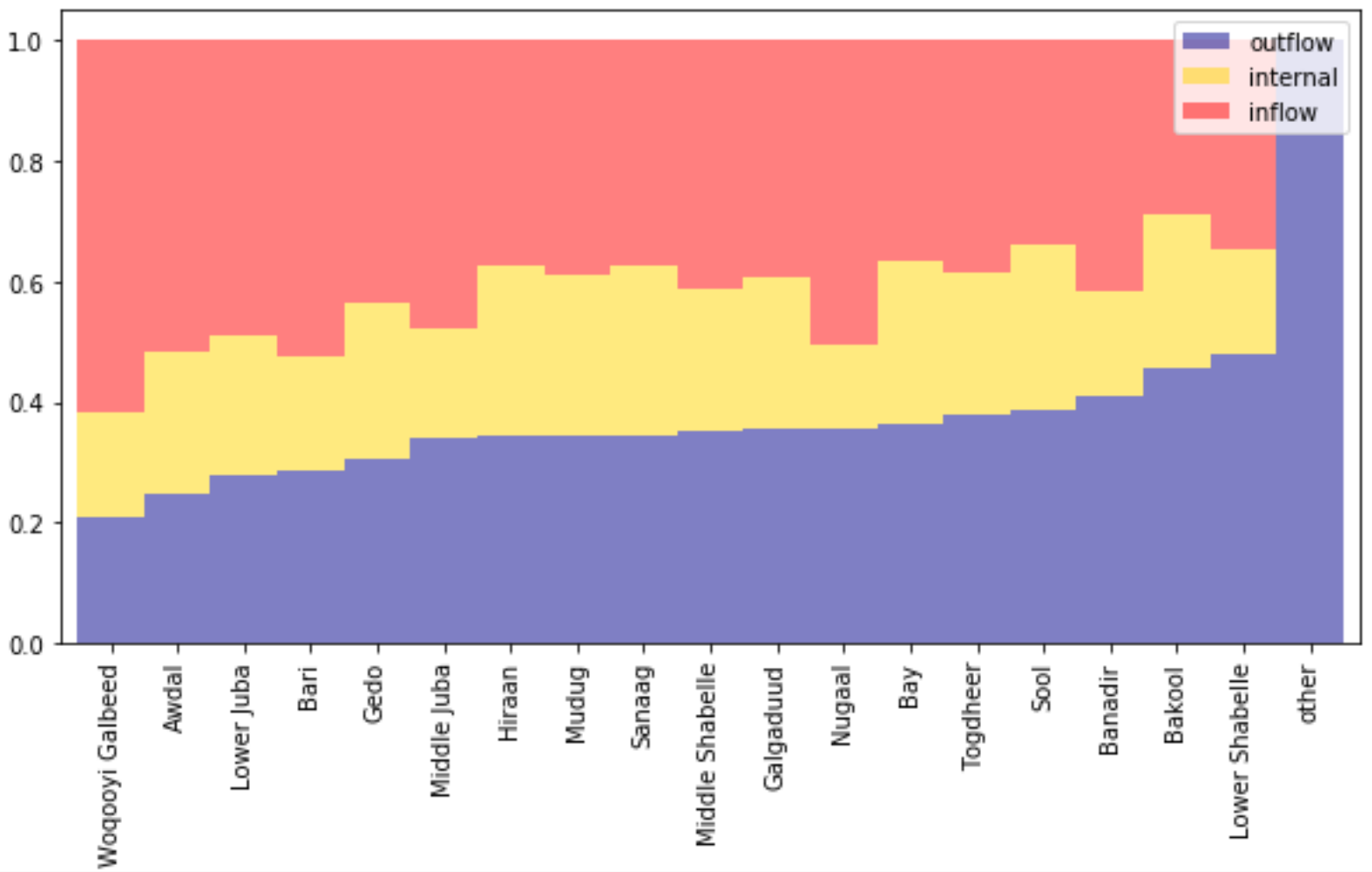
PRMN: current, previous, or future location





Data Sources

PRMN: current, previous, or future location



Problem Setup

1. Model output: # monthly arrivals in region
2. Model input:
    - Lagged data for the region
      (e.g. previous 3,4,5,6,12 months)
    - Lagged data for other regions
      (e.g. previous 3,4,5,6,12 months)
    - Indicator for missing values
    - Indicator for region and month of year
    - Linear time trend

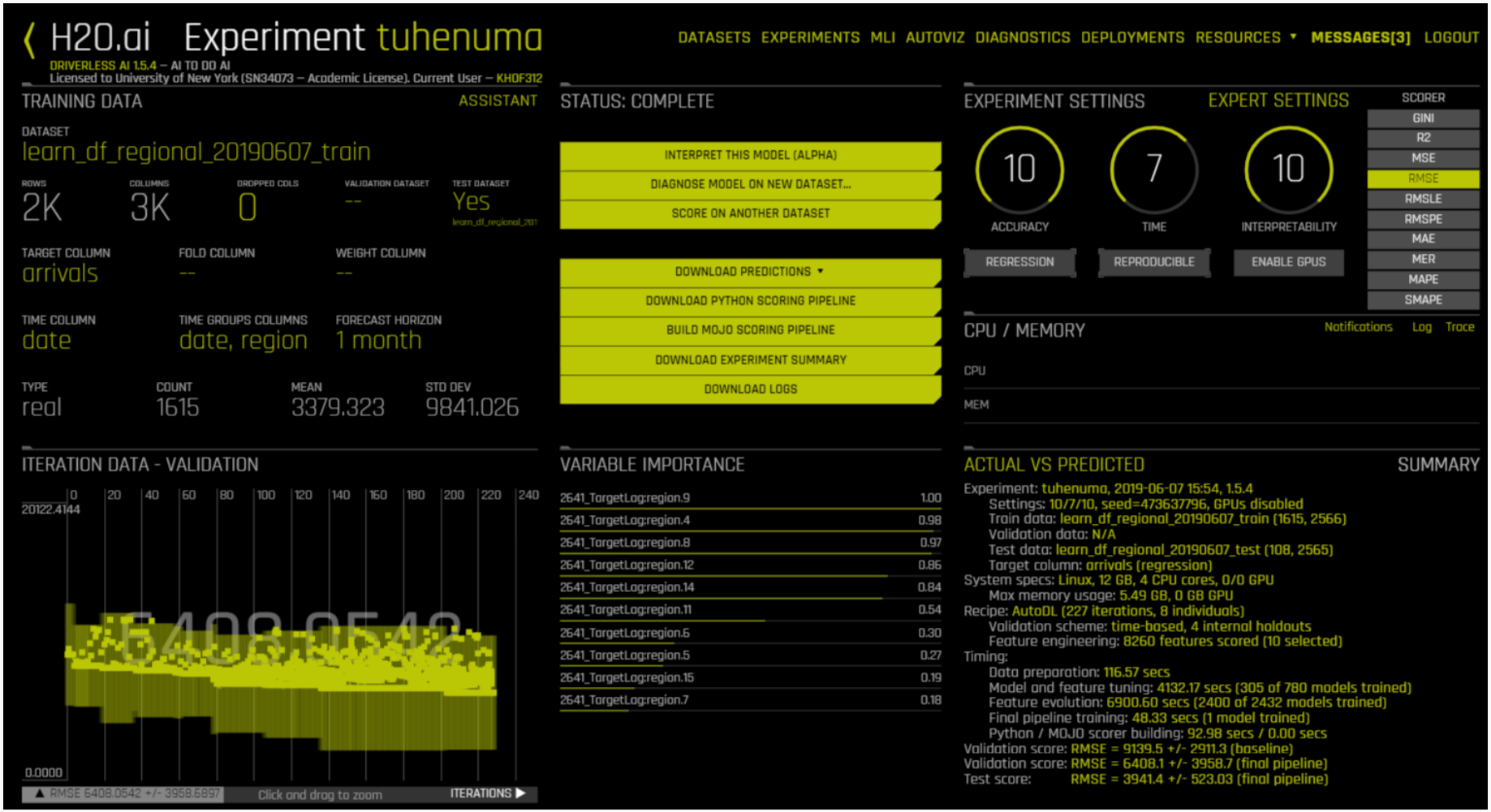
Tools tested

Python (SciKitLearn)

```
# ...

# Specify model
m = DecisionTreeRegressor(min_samples_leaf=5, max_depth=5, criterion='mse')

# Fit model
m.fit(X_train, y_train)

# Make predictions
y_fit  = m.predict(X_train)
y_pred = m.predict(X_test)

# ...
```

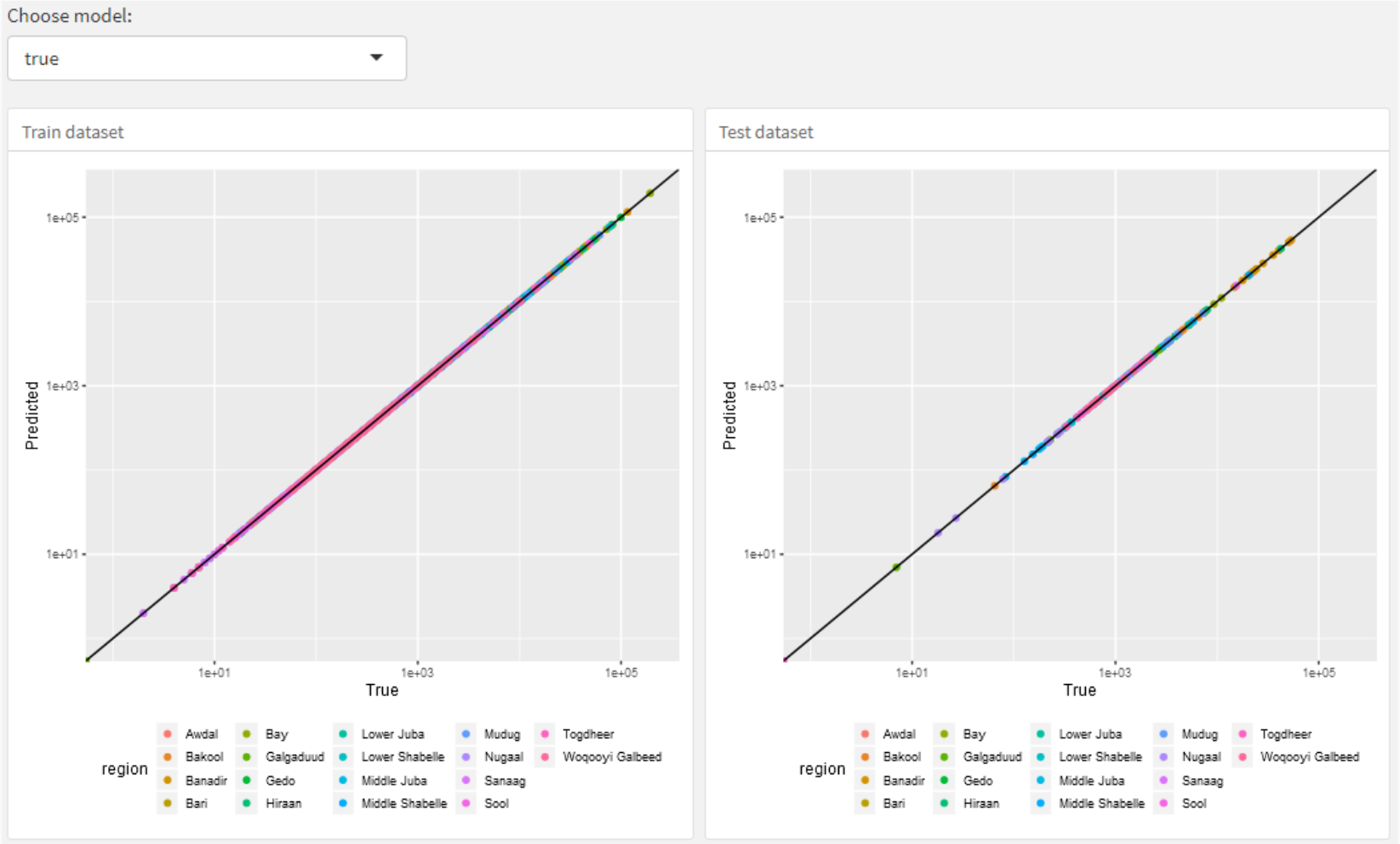
Tools tested

H2o.ai



Models tested

Ground truth:



Models tested

Baseline: Last observation carried forward (1 month)

Models tested

Baseline: Last observation carried forward (3 months)

Models tested

Linear regression: $\approx$ try to fit a line to the data

Models tested

Decision trees: ≈ break the dataset into groups

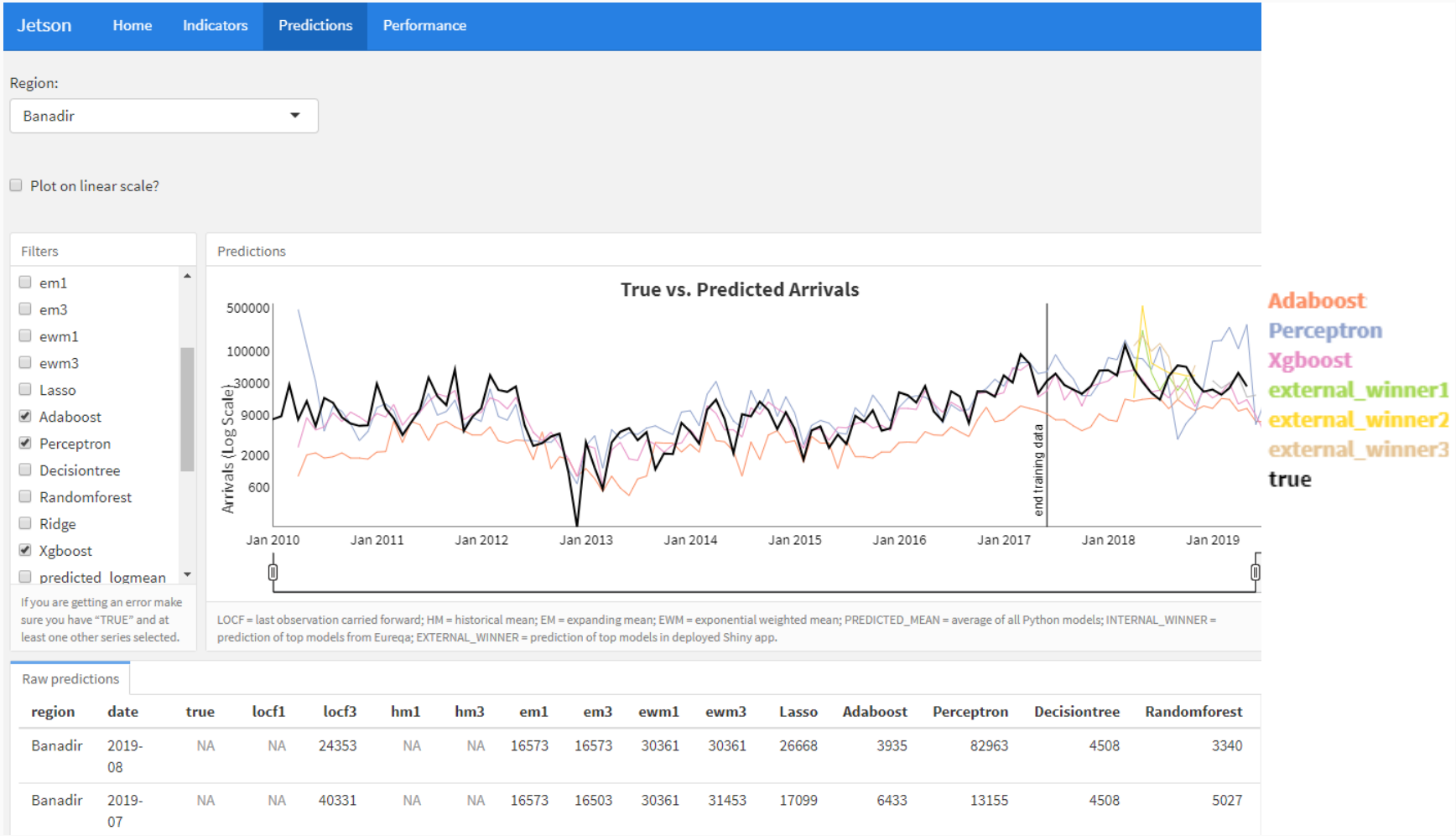
Models tested

Boosting: combine many models in a smart way

Models tested

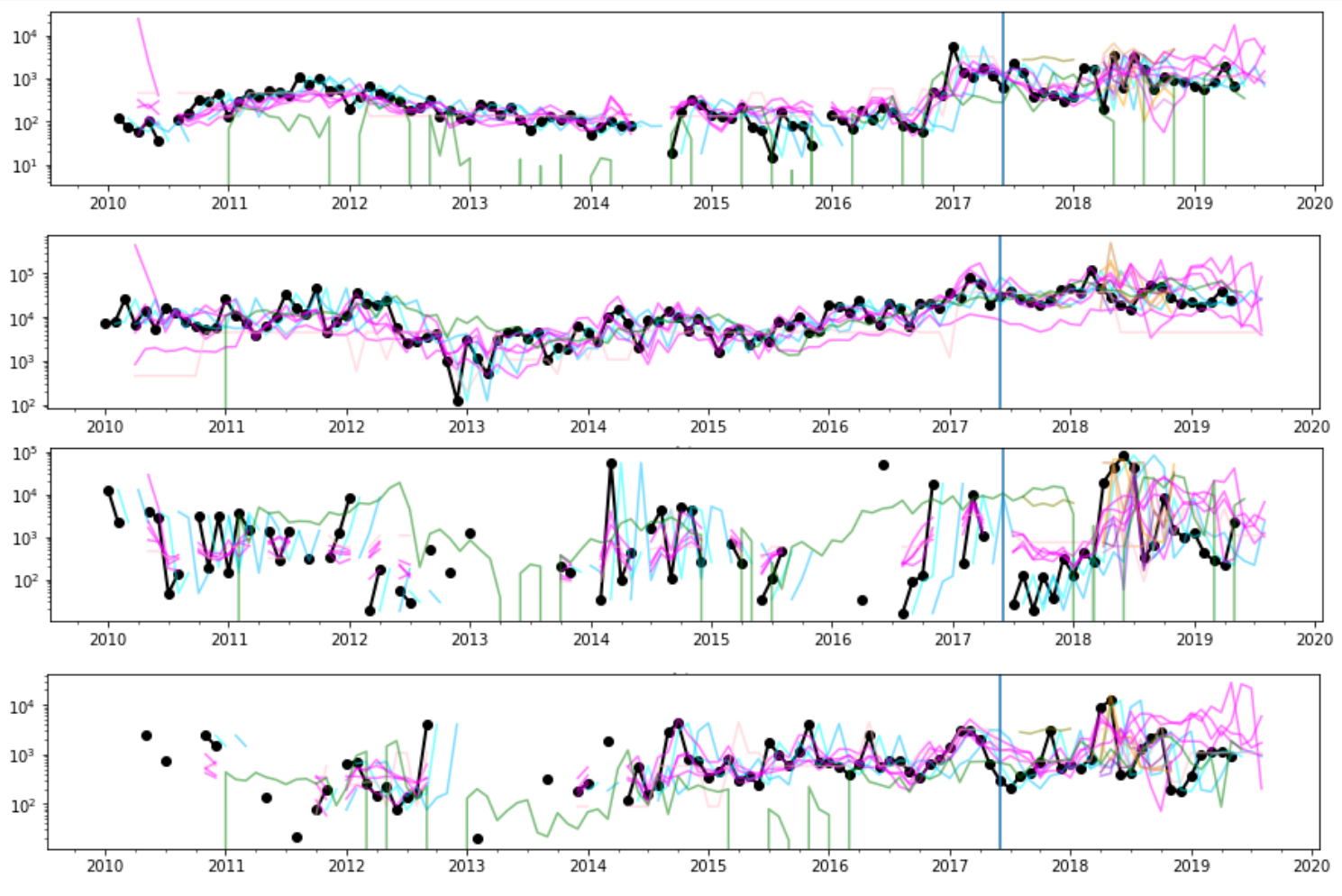
Perceptron: fit a very flexible model

Model Comparison



region	date	true	locf1	locf3	hm1	hm3	em1	em3	ewm1	ewm3	Lasso	Adaboost	Perceptron	Decisiontree	Randomforest
Banadir	2019-08	NA	NA	24353	NA	NA	16573	16573	30361	30361	26668	3935	82963	4508	3340
Banadir	2019-07	NA	NA	40331	NA	NA	16573	16503	30361	31453	17099	6433	13155	4508	5027

Model Comparison



Performance Metrics

1. Percentage of computer prediction accuracy (ACC): $\frac{\text{predicted value}}{\text{true value}}$

   penalizes the PERCENT by which we are off

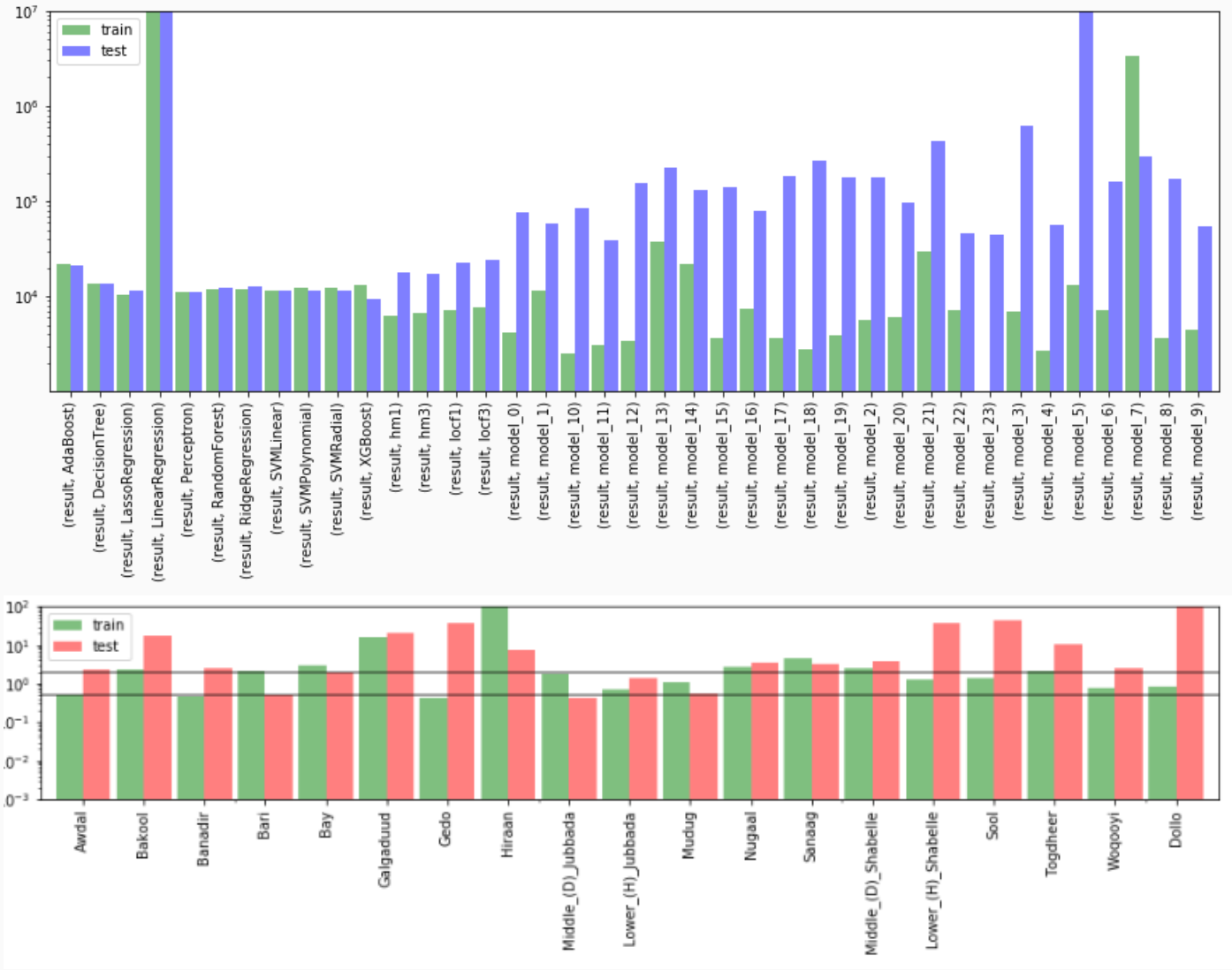
2. Mean squared error (MSE): $(\text{true} - \text{predicted})^2$

   penalizes the AMOUNT by which we are off

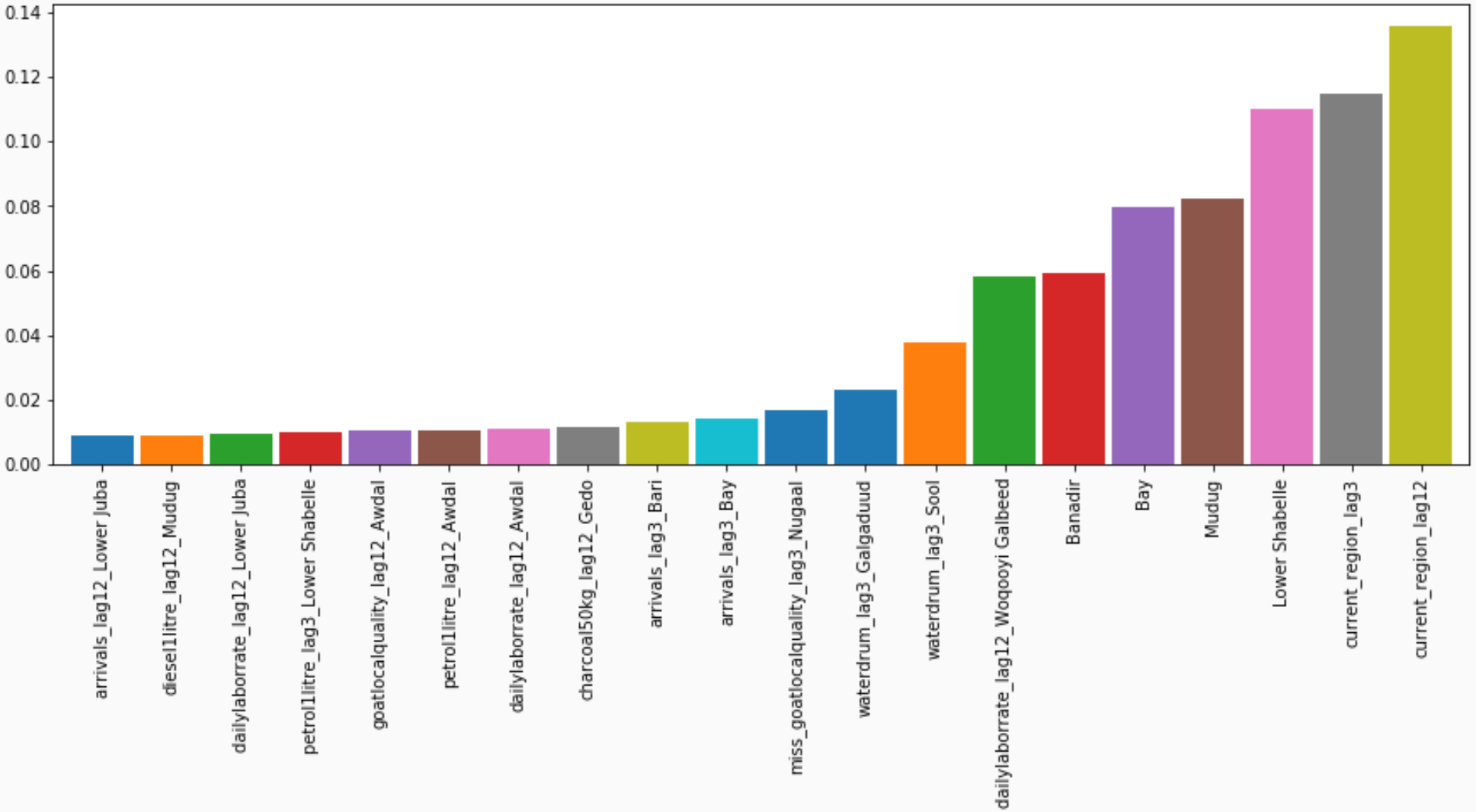
3. Accuracy: $\frac{\text{correctly classified}}{\text{all classified}}$

   penalizes the DIRECTION in which we are off

Performance Metrics



Explainability



Current Priorities

1. Documentation

   Clean, reproducible, and open-source code

2. Automation

   Automatic data downloading and model selection

3. Experimentation pipeline

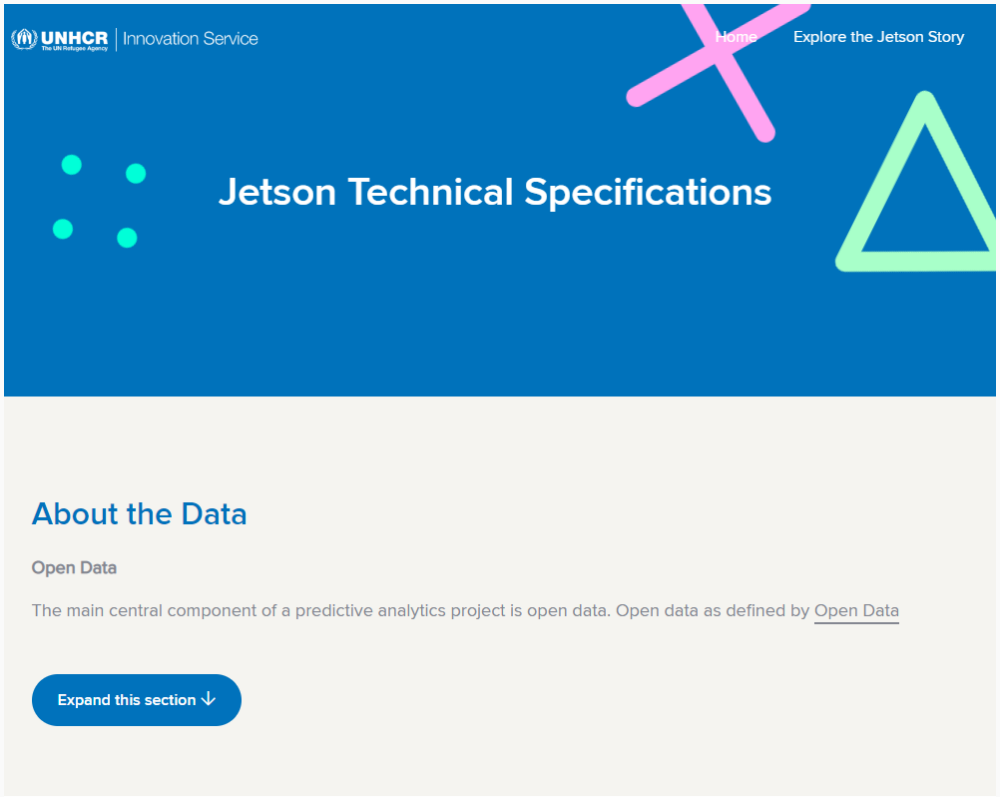
   Tune models with clear goal in mind ;

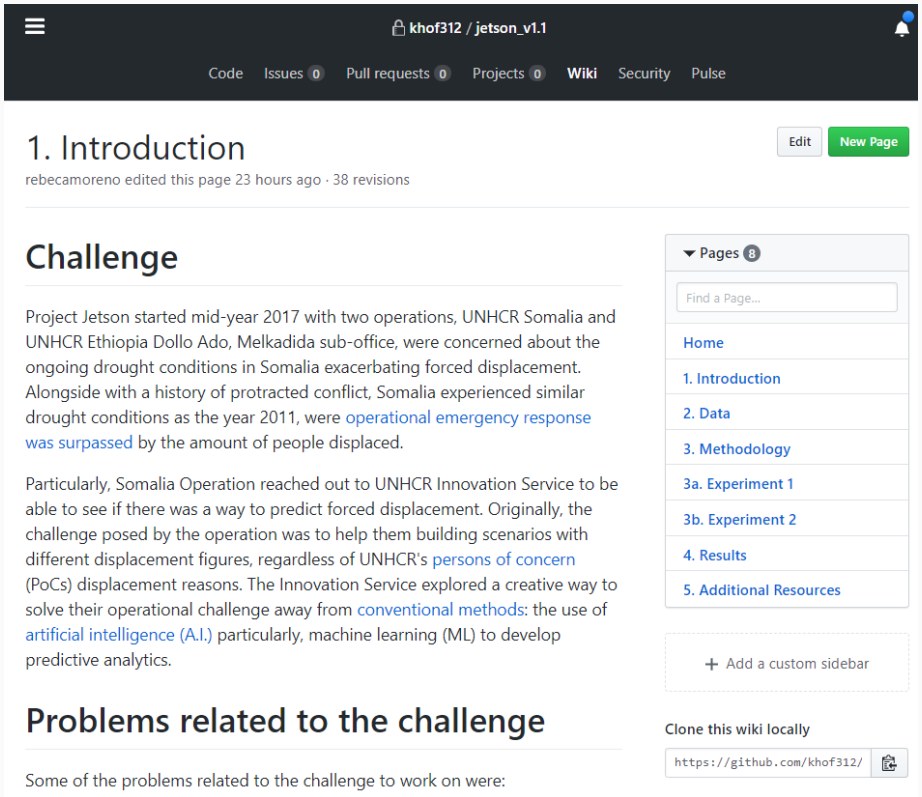
   Test different model settings and compare

4. Explainability

   Visualization and model insights

Questions?


UNHCR | Innovation Service
Home
Explore the Jetson Story
Jetson Technical Specifications
About the Data
Open Data
The main central component of a predictive analytics project is open data. Open data as defined by Open Data
Expand this section


khof312 / jetson_v1.1
Code
Issues 0
Pull requests 0
Projects 0
Wiki
Security
Pulse
1. Introduction
Edit
New Page
rebecamoreno edited this page 23 hours ago · 38 revisions
Challenge
Project Jetson started mid-year 2017 with two operations, UNHCR Somalia and UNHCR Ethiopia Dollo Ado, Melkadida sub-office, were concerned about the ongoing drought conditions in Somalia exacerbating forced displacement. Alongside with a history of protracted conflict, Somalia experienced similar drought conditions as the year 2011, were operational emergency response was surpassed by the amount of people displaced.
Particularly, Somalia Operation reached out to UNHCR Innovation Service to be able to see if there was a way to predict forced displacement. Originally, the challenge posed by the operation was to help them building scenarios with different displacement figures, regardless of UNHCR's persons of concern (PoCs) displacement reasons. The Innovation Service explored a creative way to solve their operational challenge away from conventional methods: the use of artificial intelligence (A.I.) particularly, machine learning (ML) to develop predictive analytics.
Problems related to the challenge
Some of the problems related to the challenge to work on were:
Pages 8
Find a Page...
Home
1. Introduction
2. Data
3. Methodology
3a. Experiment 1
3b. Experiment 2
4. Results
5. Additional Resources
Add a custom sidebar
Clone this wiki locally
https://github.com/khof312/