ZINDI

Wazihub Soil Moisture Prediction Challenge

Debrief Session

6 November 2019

INTRODUCTION

AGENDA

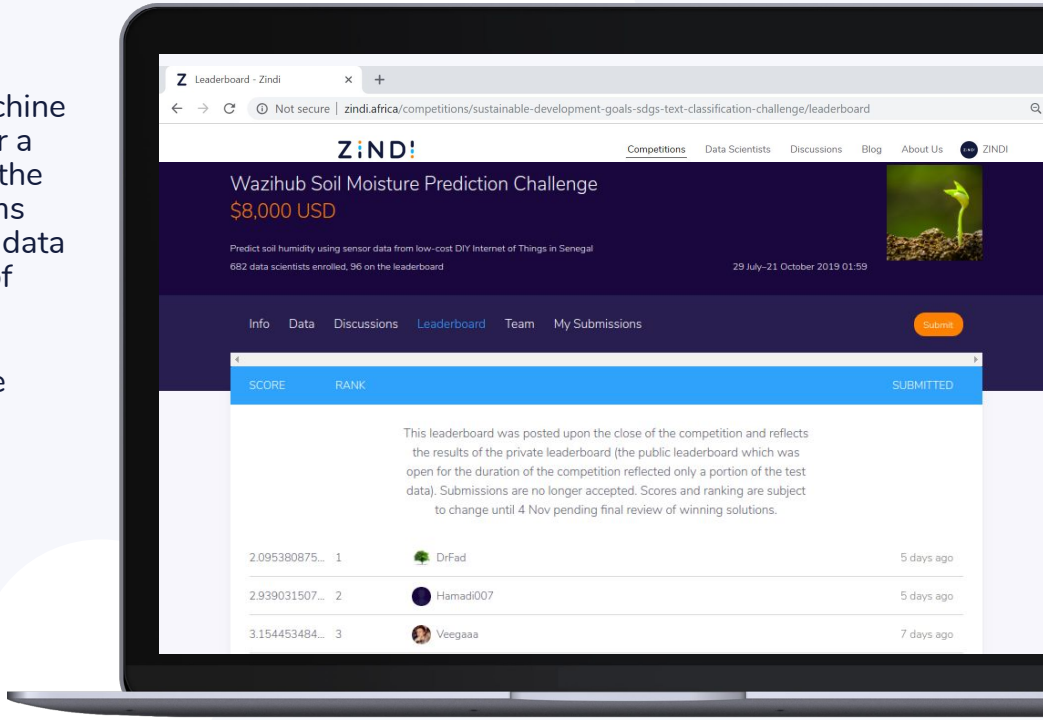
- Quick introduction (Celina)
- Competition Overview (Celina)
- Winning solutions (Amy)
- Open discussion (Wazihub + Zindi)
- Feedback (Wazihub)

COMPETITION OVERVIEW

OBJECTIVE(S) OF THE CHALLENGE

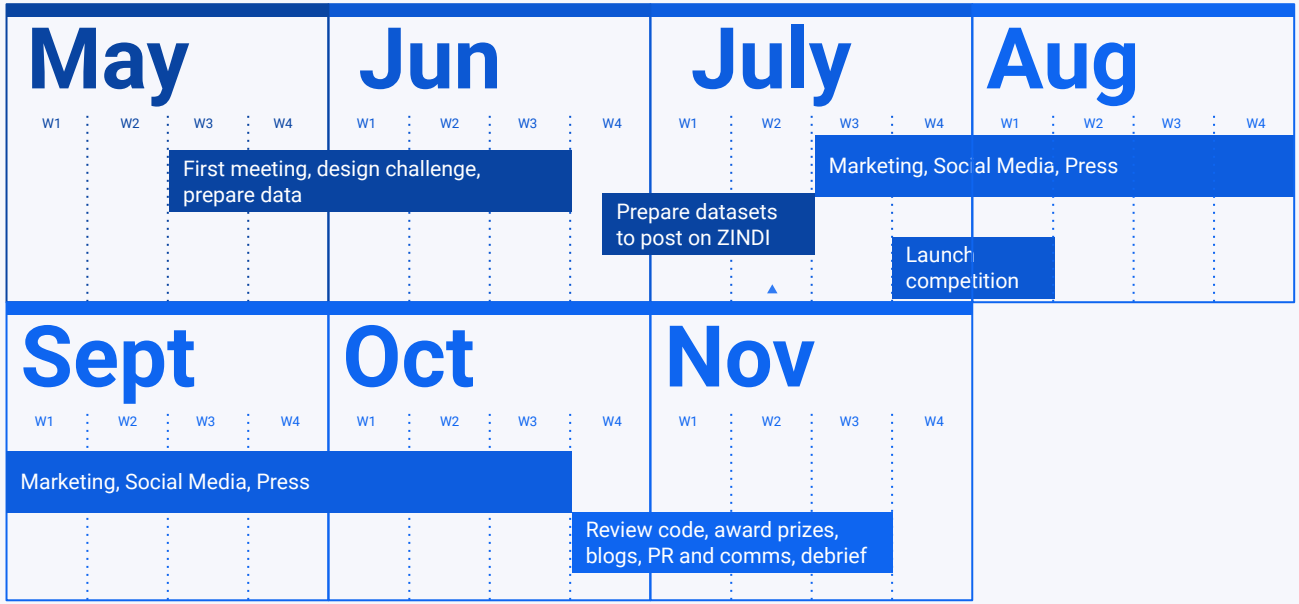
The objective of the competition was to create a machine learning model capable of predicting the humidity for a particular plot in the next few days, using data from the past. A part of the challenge was to design algorithms that are resilient and can be trained with incomplete data (e.g. missing data points) and unclean data (e.g. lot of outliers).

This resulting model will enable farmers to anticipate water needs and prepare their irrigation schedules.



COMPETITION OVERVIEW

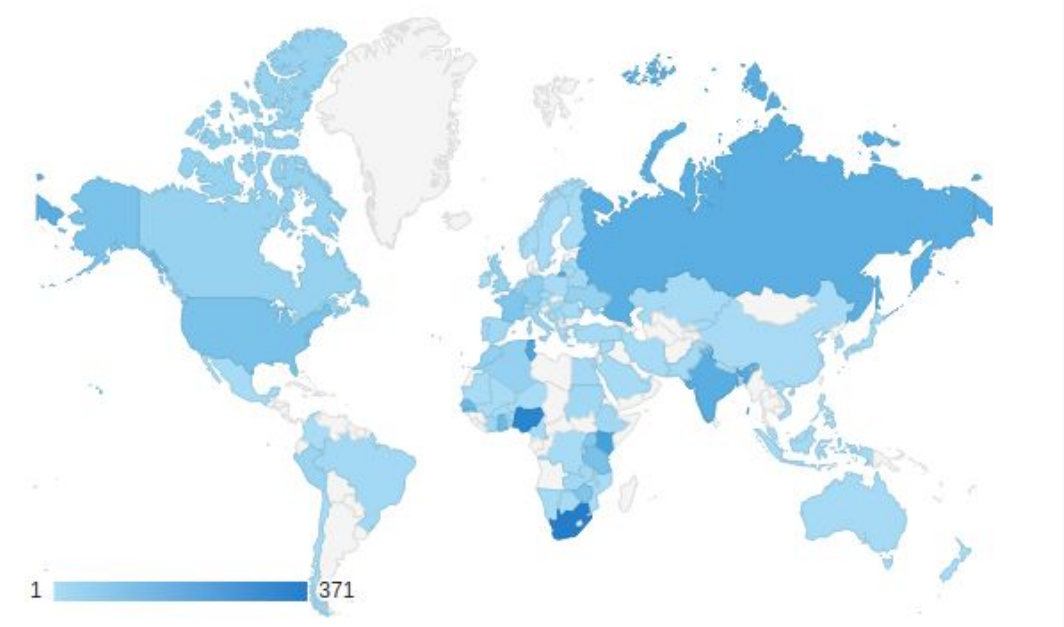
TIMELINE



COMPETITION OVERVIEW

CHALLENGE WEBPAGE ENGAGEMENT

- 677 data scientists enrolled
- 6,443 total sessions across 96 countries



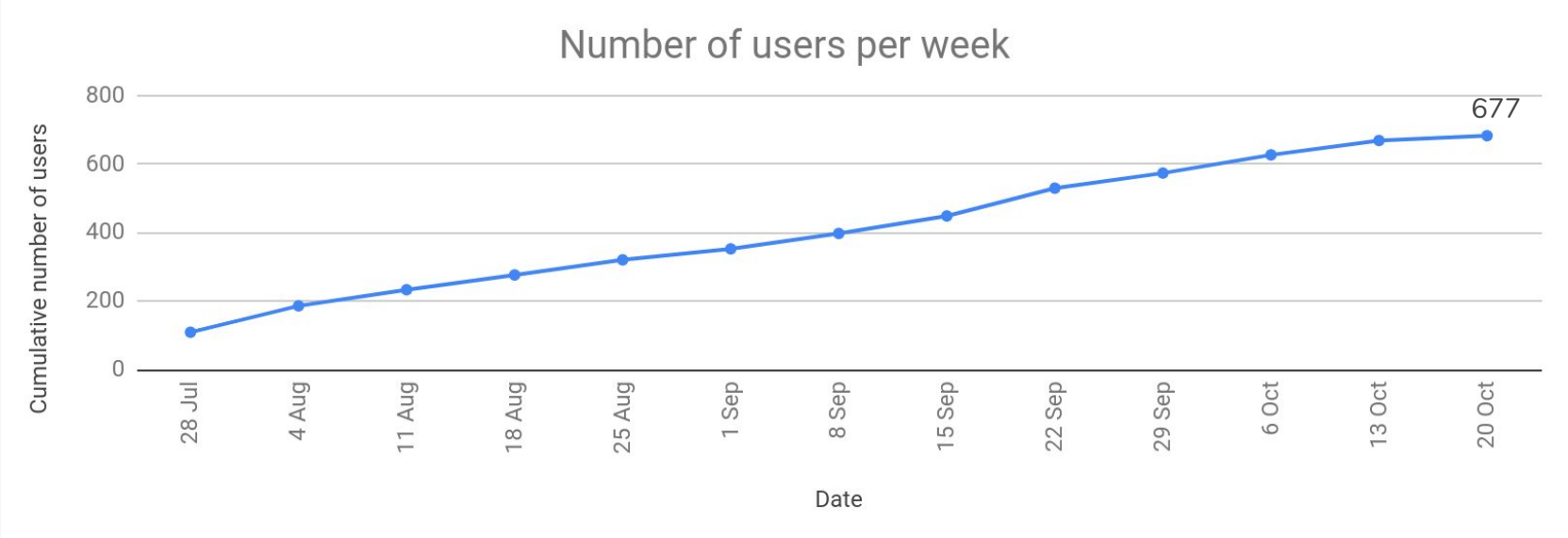
COMPETITION OVERVIEW

MARKETING AND COMMUNICATION

- Targeted activity on Facebook, Instagram, LinkedIn & Twitter
- Winners blog on Zindi
- Wazihub Medium blog
- ITweb press release
- TechCrunch press release

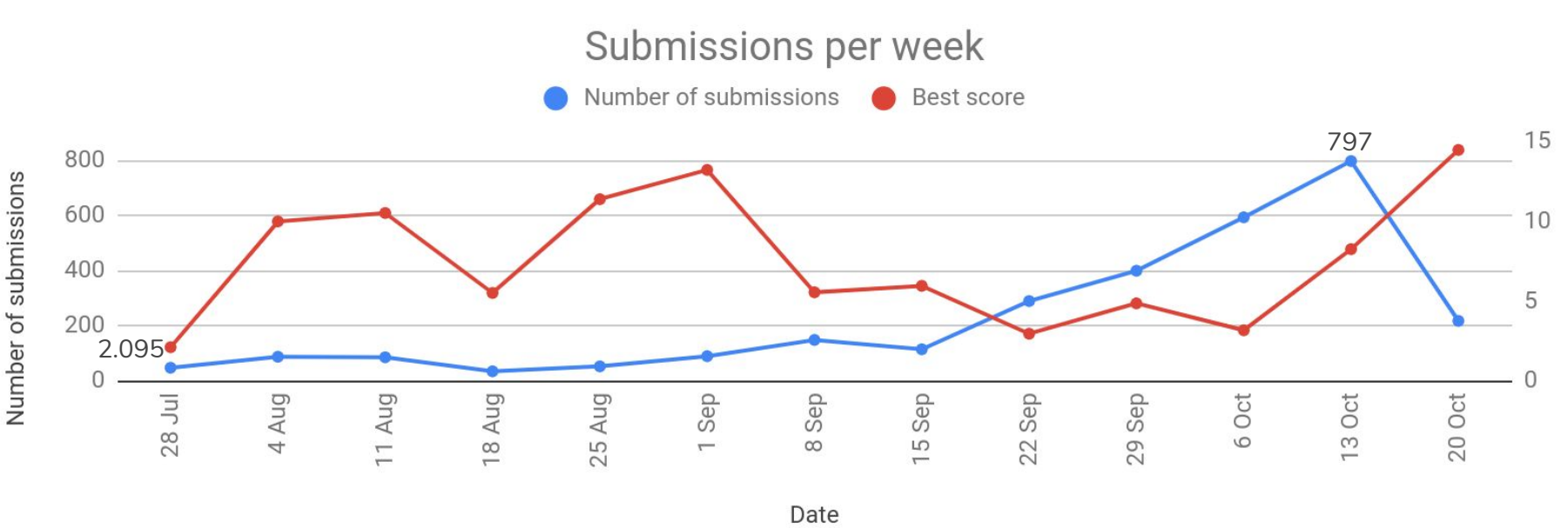
WINNING SOLUTIONS

DATA SCIENCE COMMUNITY ENGAGEMENT



COMPETITION OVERVIEW

DATA SCIENCE COMMUNITY ENGAGEMENT



WINNING SOLUTIONS

TOP 6 PARTICIPANTS' PROFILES

1st Olayinka Fadahunsi

From Nigeria

He works as a Data Scientist with Stanbic IBTC Bank. He enjoys creating solutions and he is a role model in the Zindi community.

2nd Hamadi Chihaoui

From Tunisia

Works as a machine learning engineer for Vneuron

3rd Ageev Alex

From Russia

Graduate from Higher School of Economics. Ageev is currently working as a Senior Data Scientist at a bank.

WINNING SOLUTIONS

TOP 6 PARTICIPANTS' PROFILES

4th Jasseur Abidi

From Tunisia

Works as a machine learning engineer for Vneuron

5th daoudi

From Tunisia

Joined Zindi for this challenge

6th Sertac_Ozker

From South Africa

Has worked for more than 10 years in Turkey's leading banks in data warehouse, business intelligence, reporting and data analytics positions.

WINNING SOLUTIONS

1ST PLACE SOLUTION

- Programming language - R
- Step 1: Data wrangling - extract, manipulate, transform and fill missing values.
- Step 2: Build 5 different models
- Step 3: Ensemble different models to create a single powerful model
- Soil humidity changed with irrigation ON/OFF and also when coefficient cultural, evapotranspiration rate and evapotranspiration reference deviate from normal values. Modeled these changes.
- Assumed a linear growth or decline in soil humidity. Future work could explore non-linear relationships to improve the model score.
- Used the irrigation schedule to enhance model.

WINNING SOLUTIONS

2ND AND 4TH PLACE PROBLEM-SOLVING APPROACHES

- Solution 2
  - Programming language - Python
  - Conventional forecasting solution.
  - Used temperature and pressure as features. Filled missing values with the value 24 hours prior.
  - XGboost regressor

- Solution 4
  - Programming language - Python
  - Inspired by controlled dynamic system models
    - What are the optimal system parameters (performance, safety, economy, etc.)?
    - Can the system be stabilized and, if yes, what are the “best” (cost, performance, etc.) control configurations?
    - What happens if a sensor fails and how can the system’s robustness be increased?
  - Used forward and back propagation. It would be interesting to see how the model performs without the enhancement of the backpropagation.
  - Used LGBMRegressor with 5 k fold validation

WINNING SOLUTIONS

5TH AND 6TH PLACE PROBLEM-SOLVING APPROACHES

- Solution 5
  - Programming language - R
  - Used peaks to back propagate and used that to help create model
  - Found number of hours field was watered in a 12hr, 24hr and 36 hr intervals and used those as features.
  - Used regressor model.

- Solution 6
  - Programming language - Python
  - Trained 2 models, one with irrigation on and the other with irrigation off.
  - Essentially saying, "if the irrigation is on, this field is gaining humidity with x.xx%, if the irrigation is off the field is losing humidity with x.xx%."
  - Used XGBRegressor, trained the model separately for each field and for irrigation on/off cases.
  - Predicted % humidity change for each field. Went to each humidity value for on/off and added this % humidity change. If there was a peak humidity he used this as his new humidity to add his % humidity to.

ZINDI

GITHUB SOLUTIONS

Links to Repos

- These are solutions that were found on Github. They have not been reviewed. They are from Zindi Data Scientist who attempted the challenge but did not win.
- Link1
- Link2
- Link3
- Link4

OPEN DISCUSSION

Implementation Strategy

- Wazihub implementation needs
- Wazihub implementation challenges
- Feedback from Zindi
- Open Discussion

Next steps

- Competition write-up
- Host video

Any feedback?

- How was the experience for Wazihub?
- Next steps for the partnership?