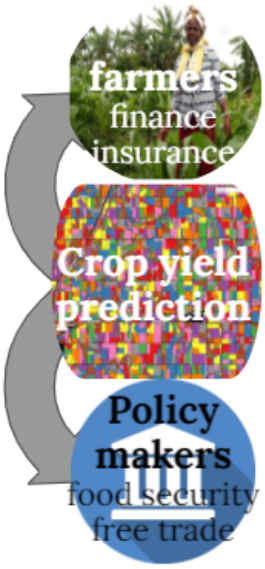
MaizeFinder

Zindi-Lacuna Data Science Competition
Erdős Institute Bootcamp 2021

In the East African countries of Kenya and Tanzania, maize comprises almost a third of daily calorie intake per person, and maize is also the most widely grown crop in Uganda. This staple food is therefore the primary business of **many stakeholders** including individual farmers, traders, processors, and policy-makers. These groups suffer losses due to being unable to predict **maize markets valuing approximately 2 billion USD a year**. With typical year-to-year variability in price of 30%, **a solution to predict maize availability is worth hundreds of millions of USD a year**, in addition to significantly reducing human suffering from famine.



More than half of the region's population directly relies on their agriculture for income or sustenance, but pre-harvest finance for maize farmers is still scarce and expensive. Current sources include wholesale distributors, who typically charge 22% annual interest, or banks, who can charge up to 4% monthly. Similarly, the cost of informal insurance can be very high, up to 25% of income**. Better yield prediction would allow institutions to provide loans and insurance at considerably reduced cost, products massively in demand among small-yield farmers.**

Therefore, maize is a central concern of government policy-makers in East Africa, including trade, taxation, and disaster prevention (food stockpiling). Many of these policy-makers lack confidence in predictions, leading to mismanagement of export bans, insurance, price-setting, or safety nets. The main impact of better policy is to reduce market inefficiencies across the region including reducing transportation costs, the primary cost in selling maize from farm-gate to fork. **Market inefficiencies due to policy uncertainty in Nairobi alone cost more than 10 million USD a year**, on average, since 2009.

To address these issues, we are enabling maize yield prediction using satellite imagery. One key aspect of this issue is being able to scalably identify maize field locations. The Zindi-Lacuna data science competition asks participants to predict the center coordinate (displacement from center of image) of the field for identification and maize yield estimation. Our stakeholders identified **field identification from satellite imagery as the primary obstacle to accurate same-season maize yield forecasting in East Africa.**

Our models were able to process the satellite images and predict corrections to the field centers with a similar accuracy as the original annotation. However, we estimate that a fivefold reduction in loss is necessary to achieve the goals as outlined above, which is considerably better than any posted competition score. We have two proposed extensions to the model: one which would prescreen images for quality and another to combine our predictions using ensemble learning.