ҮНДЭСНИЙ
СТАТИСТИКИЙН
ХОРОО

Using Big Data
to Measure and estimate indicators
on the Nexus between Environment and Gender

UNESCAP – NSO Joint pilot project outcome

OUTLINE

- PROJECT BACKGROUND
- CONCEPT, DEFINITION OF GER
- METHODOLOGY
- ESTIMATION, OUTCOME
- CHALLENGES, LESSONS LEARNED

ҮНДЭСНИЙ
СТАТИСТИКИЙН
ХОРОО

PROJECT BACKGROUND

PROJECT OVERVIEW

AIM:

According to recommendation issued by statistical commission, to fill the gap and to measure the Nexus between Environment and Gender

INDICATOR:

Share of urban population who living in unsanitary and inadequate housing, by gender

EXPECTED OUTCOME:

1. National guidance accordance with big data utilization
2. Model testing, outcome
3. Project report

ҮНДЭСНИЙ
СТАТИСТИКИЙН
ХОРОО

CONCEPT, DEFINITION OF GER: Can ger be appropriate dwelling?

DO MONGOLIA HAVE THE SLUM DISTRICT?



Over 200 citizens live in Ulanchuluut,

neighbouring area to capital city.

INAPPROPRIATE HOUSING IN MONGOLIA

Is ger districts are slum districts? Whether is ger inappropriate housing?



INAPPROPRIATE HOUSING IN MONGOLIA

Is ger districts are slum districts? Whether is ger inappropriate housing?



CONCEPT, DEFINITION

SDG 11.1 - by 2030, ensure access for all to adequate, safe and affordable housing and basic services, and upgrade slums

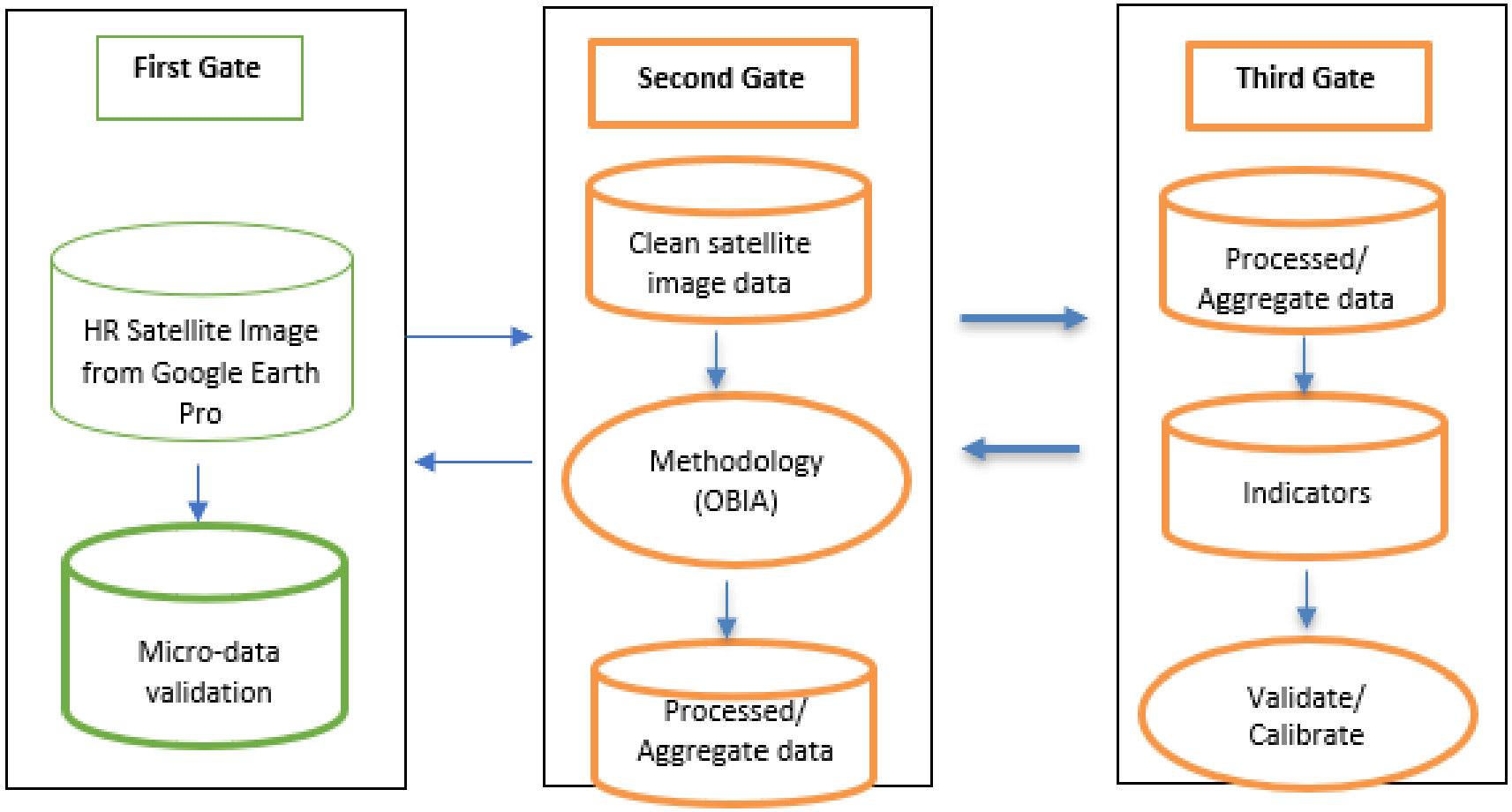
11.1.1- Proportion of urban population living in slums, informal settlements or inadequate housing

1. Access to improved water
2. Access to improved sanitation
3. Sufficient living space
4. Structural quality/durability of dwellings
5. Security of tenure

ҮНДЭСНИЙ
СТАТИСТИКИЙН
ХОРОО

METHODOLOGY

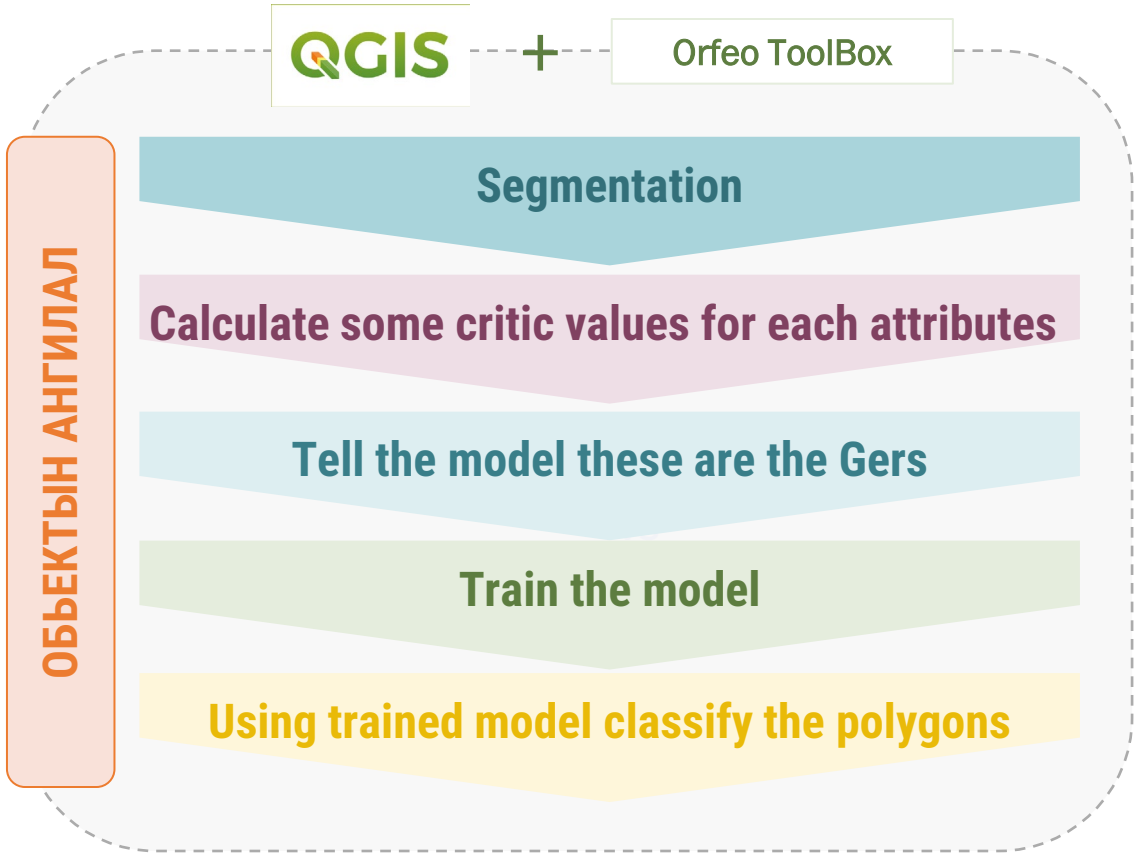
ESTIMATION STEPS/ METHODOLOGY



MODEL FOR GER IDENTIFICATION

On

THE EARTH SURFACE SATELLITE IMAGE



As

AN IMAGE AS WELL AS XLSX FILE

ҮНДЭСНИЙ
СТАТИСТИКИЙН
ХОРОО

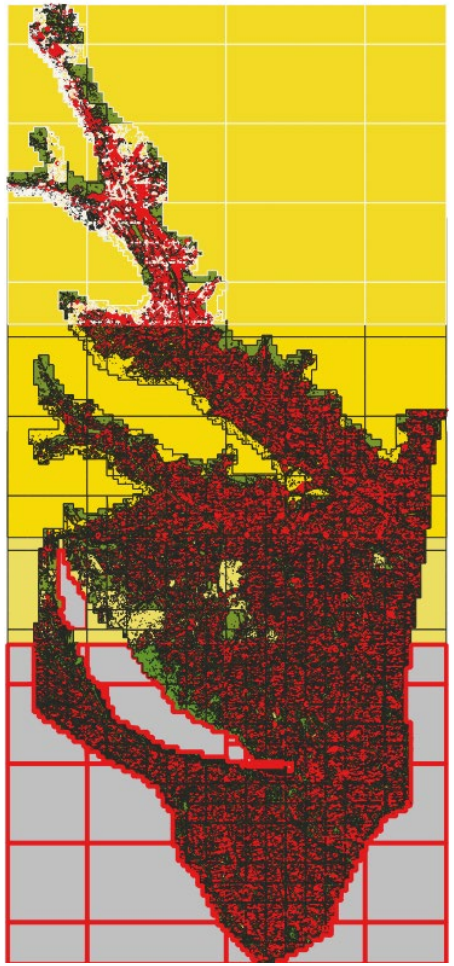
ESTIMATION, OUTCOME

THE MODEL TESTING MADE ON CHINGELTEI DISTRICT.

Sas.planet

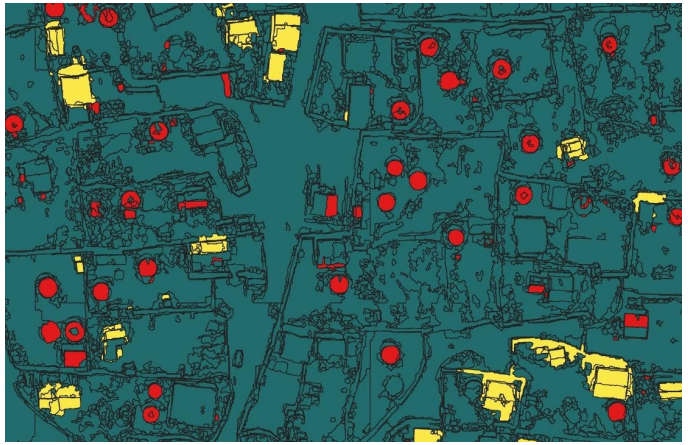


Splatted into **6** pieces



Illustration

The model detected the gers.



Ger

House, buildings

Other

THE MODEL TESTING MADE ON CHINGELTEI DISTRICT.

1. Dark ger and former ger foundation



It is hard to draw the ger shape and to recognize kind of gers because the color not different from the background

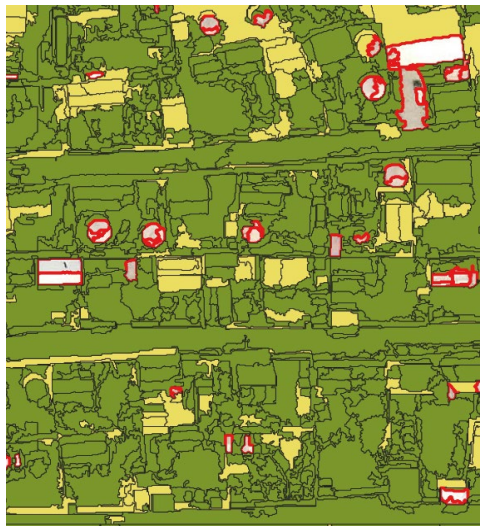
2. Unstructured streets, roads, yards



It might cause the miss recognition.

3. Objects with whiter roof



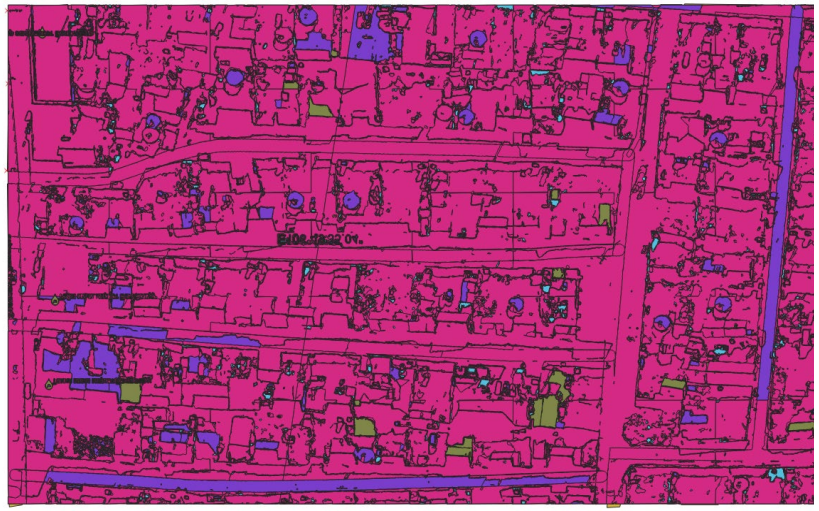


4. Uneven roof surface of ger



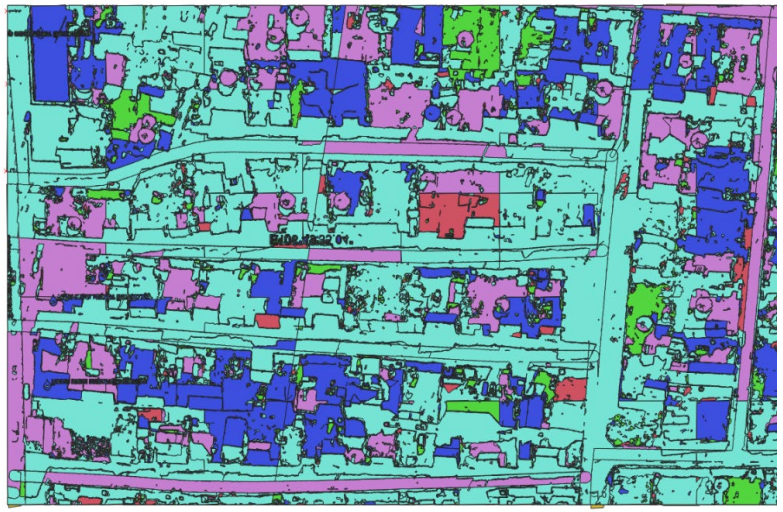
COMPARISONS OF MODEL

RANDOM FORESTS PERFORMED BETTER, AND IT ALSO DETECTED OTHER OBJECTS AS A GER ALTHOUGH NUMBER OF WRONG RECOGNITIONS WAS FEWER COMPARED WITH OTHER TWO MODELS.



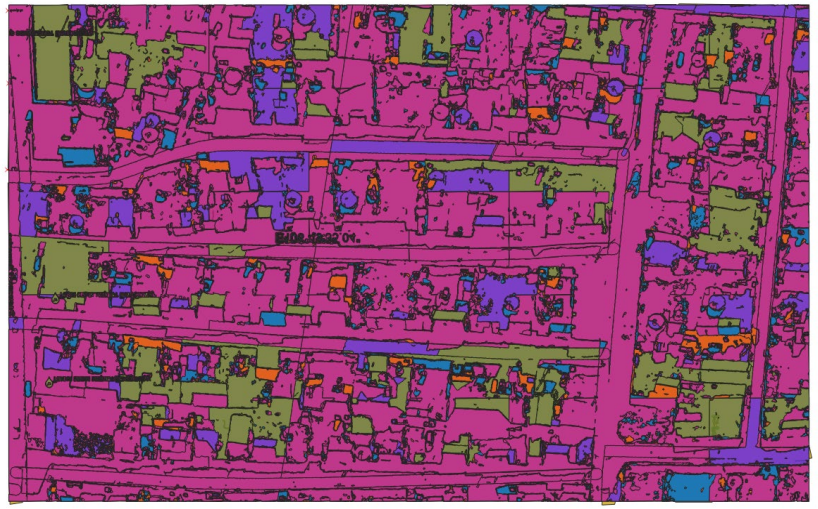
Random forest /70%/

14 out of 20 gers identified.



Decision tree /80%/

16 out of 20 gers identified. Many non-ger areas determined as a gers. Therefore, model looks like overfitted.



K nearest neighbors /45%/

9.5 out of 20 gers identified. Many non-ger areas also determined as a ger. Therefore, model did not produce good result.

For big map, it is hard to estimate the accuracy and precision of the estimation.

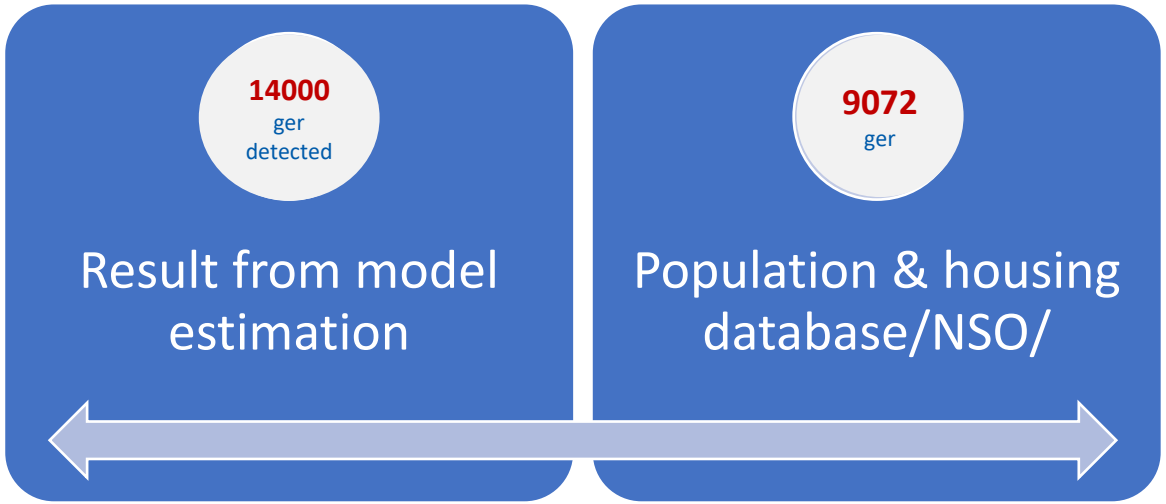
QGIS might has potential to assess the model performance.

RESULTS

Overall population Of Chingeltei: …

Female population Of Chingeltei: …

+



Female population lives in ger Of Chingeltei: …

PROS AND CONS

- Cost efficient and time saving
- Model accuracy can be upgraded
- This model has high potential to make many other estimations.

- Overestimations were a huge
- The result requires a lots of technical data cleaning process and it consumes time
- There is a high error risk while merging the 2 data sources

ҮНДЭСНИЙ
СТАТИСТИКИЙН
ХОРОО

CHALLENGES, LESSONS LEARNED

CHALLENGES, LESSONS LEARNED: IMAGE QUALITY VS QUANTITY

INITIALLY IMAGE PREPARATION WAS PROBLEMATIC, AND WE FOUND ANOTHER WAY TO HANDLE IT.

GOOGLE EARTH PRO



The problem was If we want to download big map, the image will have poor quality. Contrary, If we want to download high-resolution image, then need to zoom in a lot that means the image will capture small area.

Over **100k** pieces of image * Run the model **100k** times

- How to split the map without missing area ?
- How to run model that so many times ?

SAS.PLANET IS ADVANTAGEOUS TO DOWNLOAD SATELLITE IMAGE WITH GEO REFERENCES

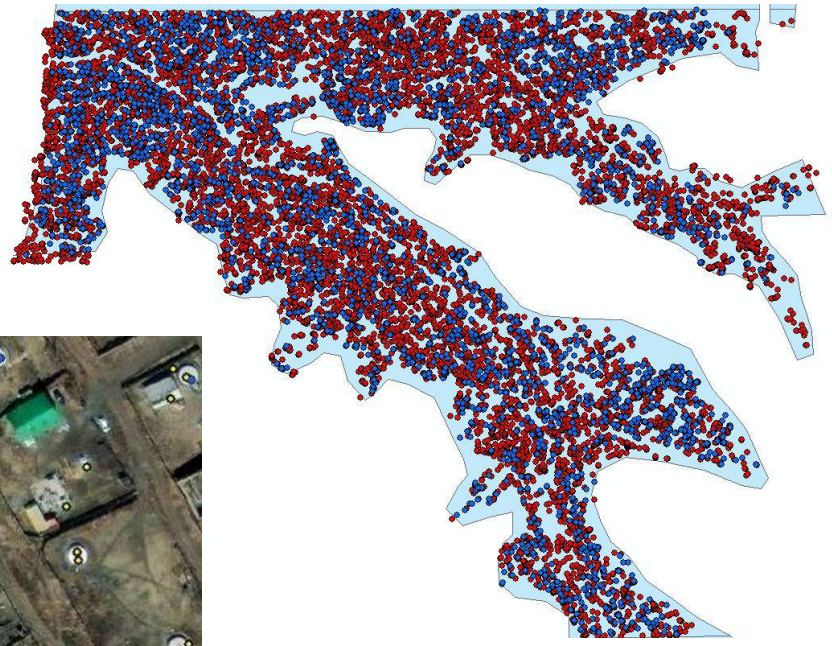
- High resolution satellite imagery
- Downloadable whole map at once
- Save locally, and offline working is possible

CHALLENGES, LESSONS LEARNED: CONSTRAINT OF DATA MERGING

THERE HAVE 2 WAY TO MERGE THE INFORMATION.

Option 1: Spatial join





- Centroid of Detected ger
- Location point of Household from population and household database

Constraints:

- ✓ Coordinates will not match exactly
- ✓ Match within 20m or 30m/
- ✓ Might connect with wrong points/
- ✓ Might create duplications

Option 2: Join using the cadastral parcel map

Detected gers + **Cadastral layer** */From LAMGG/* + Household location



Constraints:

- ✓ Quality and coverage of cadastral parcel map
- ✓ Extra data processing step

CHALLENGES, LESSONS LEARNED: TIME

ESTIMATION OF THE REALISTIC INDICATOR FOR NATIONAL LEVEL IS NOT IMPOSSIBLE, HOWEVER, MORE TIME AND WORK NEEDED.

The most time-consuming steps:

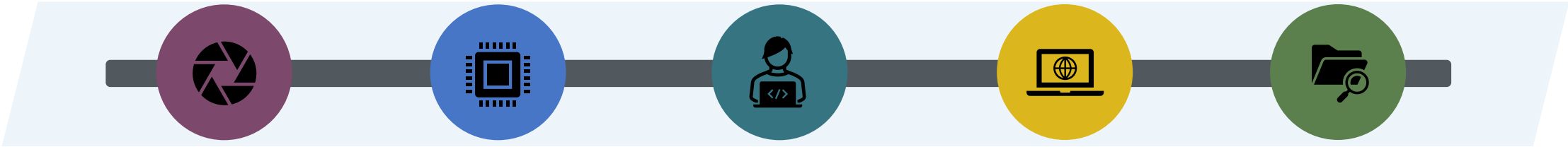
Preparation of input imagery with high resolution

Generate the sample shapes for model

Run Segmentation
/each time at least 2-3 hours/

Result data cleaning
/Needs to delete false detection/

Data cleaning for the joining



THANK YOU FOR YOUR CONSIDERATIONS!

Summary

- ✓ THE MODEL DETECTED GERS SUCCESSFULLY, HOWEVER, IT HAS SEVERAL BIASES.
- ✓ QUALITY OF INPUT IMAGE IS CRUCIAL AND SAS.PLANET IS PREFERABLE.
- ✓ ESTIMATION OF THE REALISTIC INDICATOR FOR NATIONAL LEVEL IS FEASIBLE, HOWEVER, MORE TIME NEEDED.
- ✓ BETTER TO UTILIZE COMBINATION OF SOFTWARES SINCE EACH OF THEM CONVENIENT WITH DIFFERENT FUNCTIONS.

ҮНДЭСНИЙ
СТАТИСТИКИЙН
ХОРОО

АНХААРАЛ ХАНДУУЛСАНД БАЯРЛАЛАА