Leveraging digital health and machine learning models to forecast adverse maternal outcome in low resource settings. An Experience from Geita Tanzania

Isaac Lyatuu, → Researcher, Data Scientist, PHIT
Presentation at 2024 IDM Annual Symposium
Seattle, USA

Source: Google Images

Source: Google Images

Rationale

In low- and middle-income countries, the persistently high maternal mortality ratio (MMR) remains a significant concern, representing tragic losses of both mothers and infants. Many of these deaths result from preventable complications

Rationale

In low- and middle-income countries, the persistently high maternal mortality ratio (MMR) remains a significant concern, representing tragic losses of both mothers and infants. Many of these deaths result from preventable complications

Rationale

In low- and middle-income countries, the persistently high maternal mortality ratio (MMR) remains a significant concern, representing tragic losses of both mothers and infants. Many of these deaths result from preventable complications

Rationale

In low- and middle-income countries, the persistently high maternal mortality ratio (MMR) remains a significant concern, representing tragic losses of both mothers and infants. Many of these deaths result from preventable complications

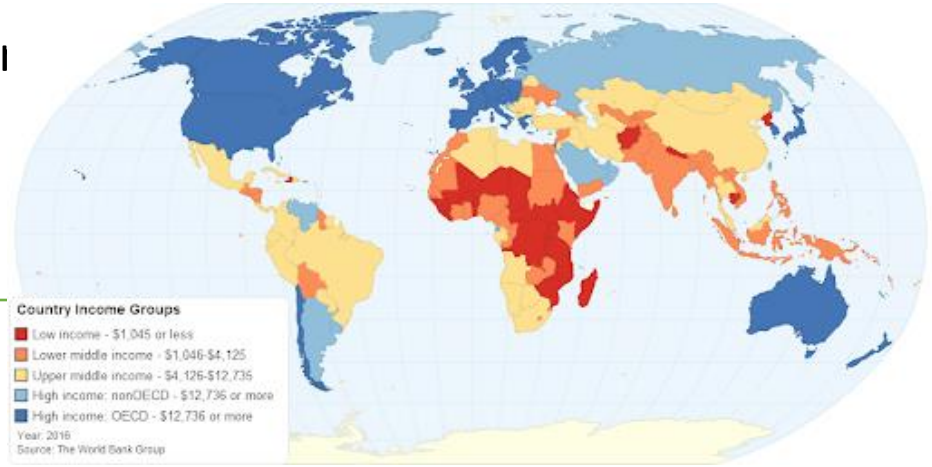
Typical characteristics includes

→ limited power (electricity),

→ Limited infrastructure (internet, roads or

→ Limited expertise (doctors, nurses),

→ Dispersed or scattered population



source: chartbin.com

Rationale

In low- and middle-income countries, the persistently high maternal mortality ratio (MMR) remains a significant concern, representing tragic losses of both mothers and infants. Many of these deaths result from preventable complications

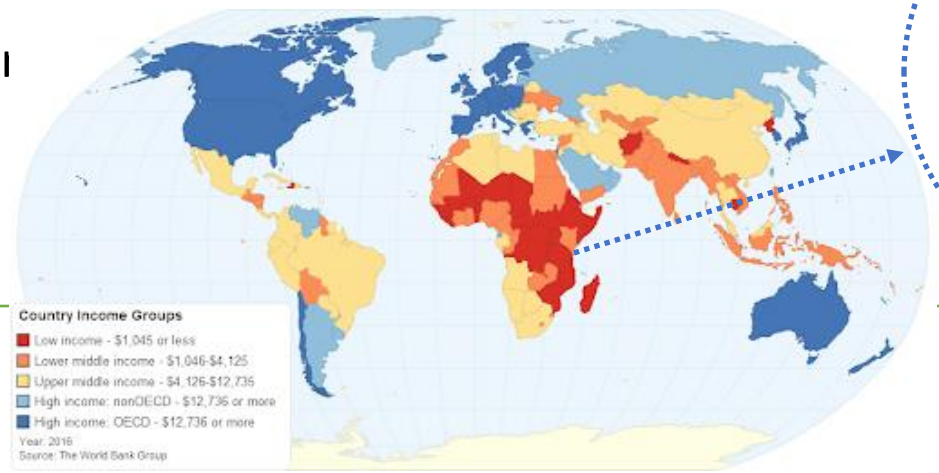
Typical characteristics includes

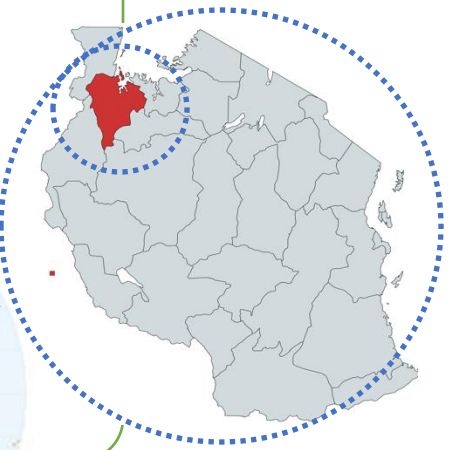
→ limited power (electricity),

→ Limited infrastructure (internet, roads or

→ Limited expertise (doctors, nurses),

→ Dispersed or scattered population





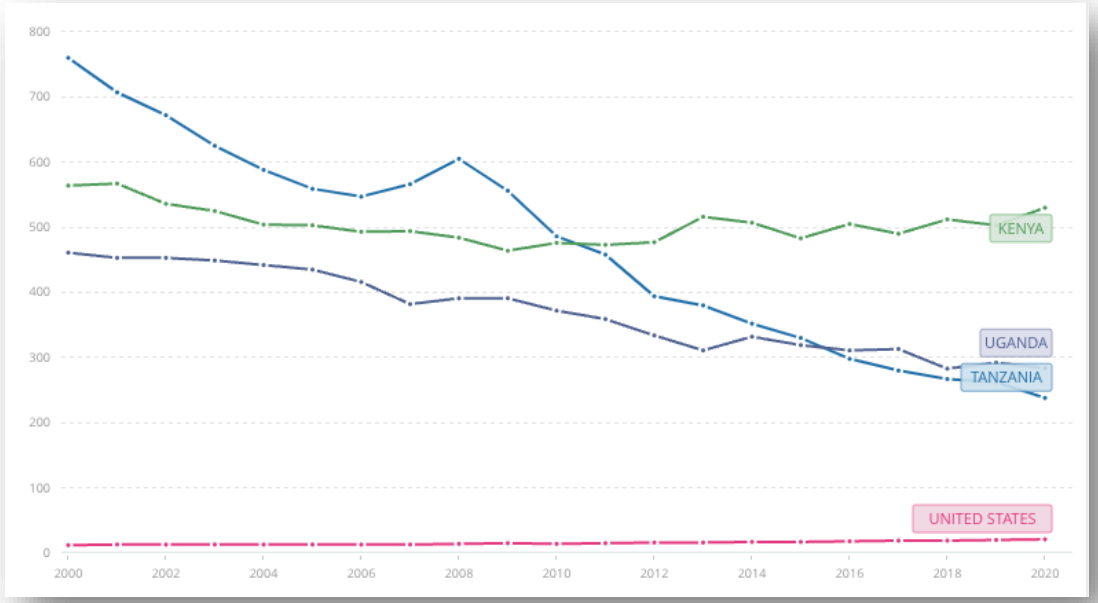
source: chartbin.com

Rationale

In low- and middle-income countries, the persistently high maternal mortality ratio (MMR) remains a significant concern, representing tragic losses of both mothers and infants. Many of these deaths result from preventable complications

Rationale

In low- and middle-income countries, the persistently high maternal mortality ratio (MMR) remains a significant concern, representing tragic losses of both mothers and infants. Many of these deaths result from preventable complications



Kenya, 530

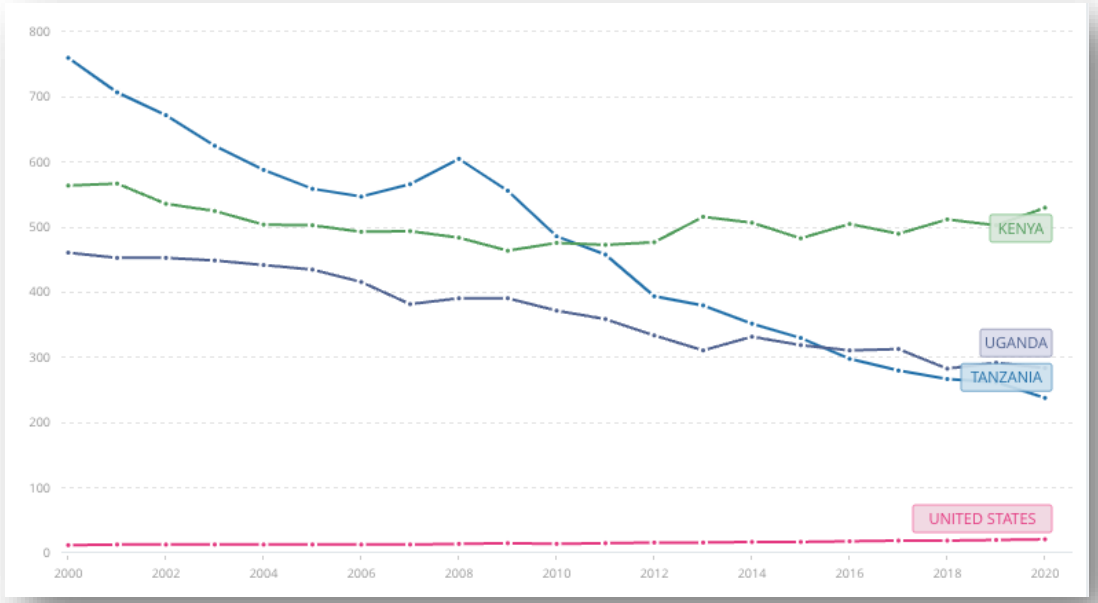
Uganda, 284

Tanzania, 262

source: World Bank Data, https://data.worldbank.org

Rationale

In low- and middle-income countries, the persistently high maternal mortality ratio (MMR) remains a significant concern, representing tragic losses of both mothers and infants. Many of these deaths result from preventable complications



Kenya, 530

Uganda, 284

Tanzania, 262

USA, 20

source: World Bank Data, https://data.worldbank.org

Rationale

In low- and middle-income countries, the persistently high maternal mortality ratio (MMR) remains a significant concern, representing tragic losses of both mothers and infants. Many of these deaths result from preventable complications

Rationale

In low- and middle-income countries, the persistently high maternal mortality ratio (MMR) remains a significant concern, representing tragic losses of both mothers and infants. Many of these deaths result from preventable complications

For example,

No	Complication	Prevention/Management
1	**Hemorrhage**	…
2	**Hypertensive Disorders (**Preeclampsia/Eclampsia) - Preeclampsia is a condition characterized by high blood pressure and damage to other organs, often the liver and kidneys. If untreated, it can develop into eclampsia, leading to seizures, coma, and death	Regular antenatal care, monitoring blood pressure, and early use of medications like magnesium sulfate during delivery can prevent severe outcomes
3	…	
4	…	

Rationale

In low- and middle-income countries, the persistently high maternal mortality ratio (MMR) remains a significant concern, representing tragic losses of both mothers and infants. Many of these deaths result from preventable complications

For example,

No	Complication	Prevention/Management
1	**Hemorrhage**	…
2	**Hypertensive Disorders (**Preeclampsia/Eclampsia) - Preeclampsia is a condition characterized by high blood pressure and damage to other organs, often the liver and kidneys. If untreated, it can develop into eclampsia, leading to seizures, coma, and death	Regular antenatal care, monitoring blood pressure, and early use of medications like magnesium sulfate during delivery can prevent severe outcomes
3	…	
4	…	

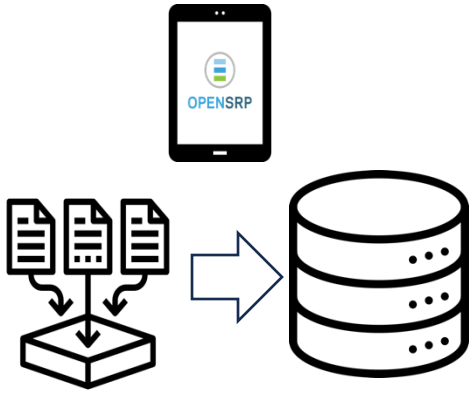
What did we do?

What did we do?

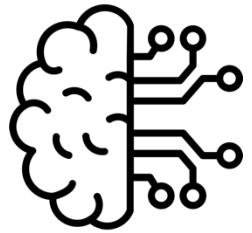
- Joined MoH initiative and deployed Unified Community System (UCS) for data capturing at facility

2. Provide routine support (i.e., education, clinical checks,…)

1. Establish GANC

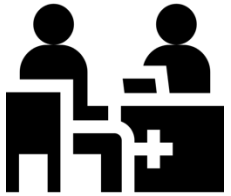


3. Capture the data



4. Predict adverse maternal outcomes before they happens

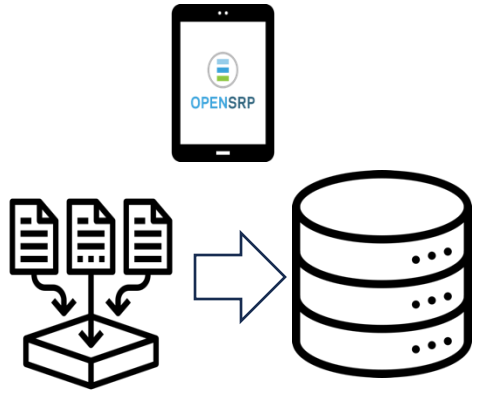
5. Provide feedback to providers



6. Provide opportunities for interventions

What did we do?

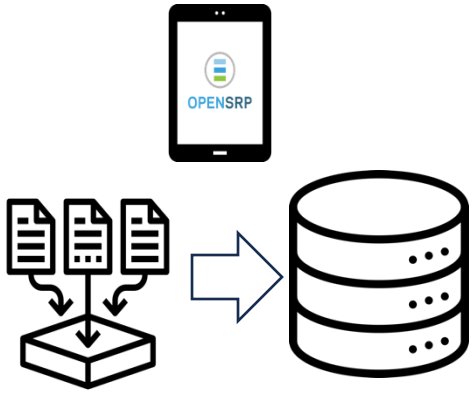
- Joined MoH initiative and deployed Unified Community System (UCS) for data capturing at facility



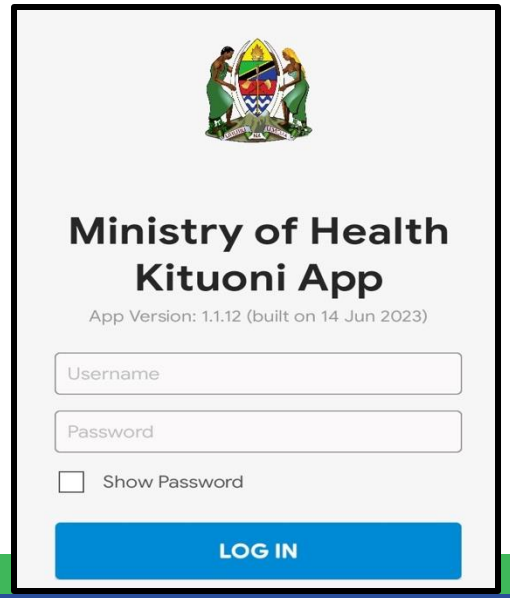
3. Capture the data

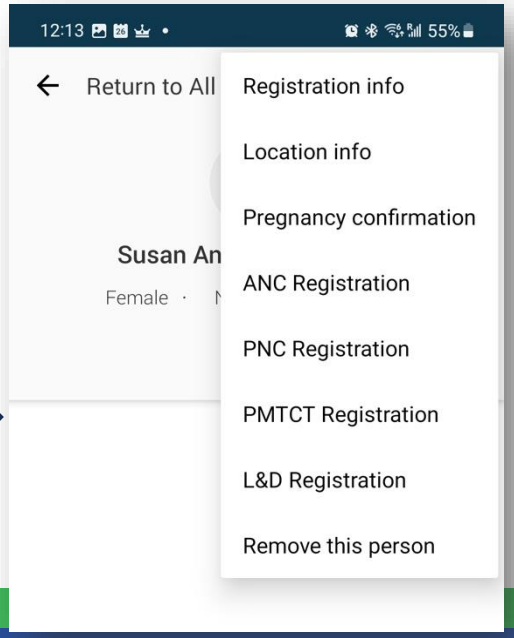
What did we do?

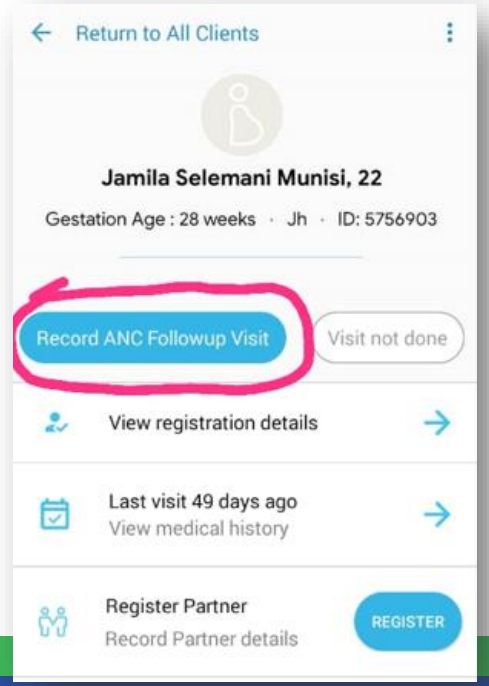
- Joined MoH initiative and deployed Unified Community System (UCS) for data capturing at facility



3. Capture the data



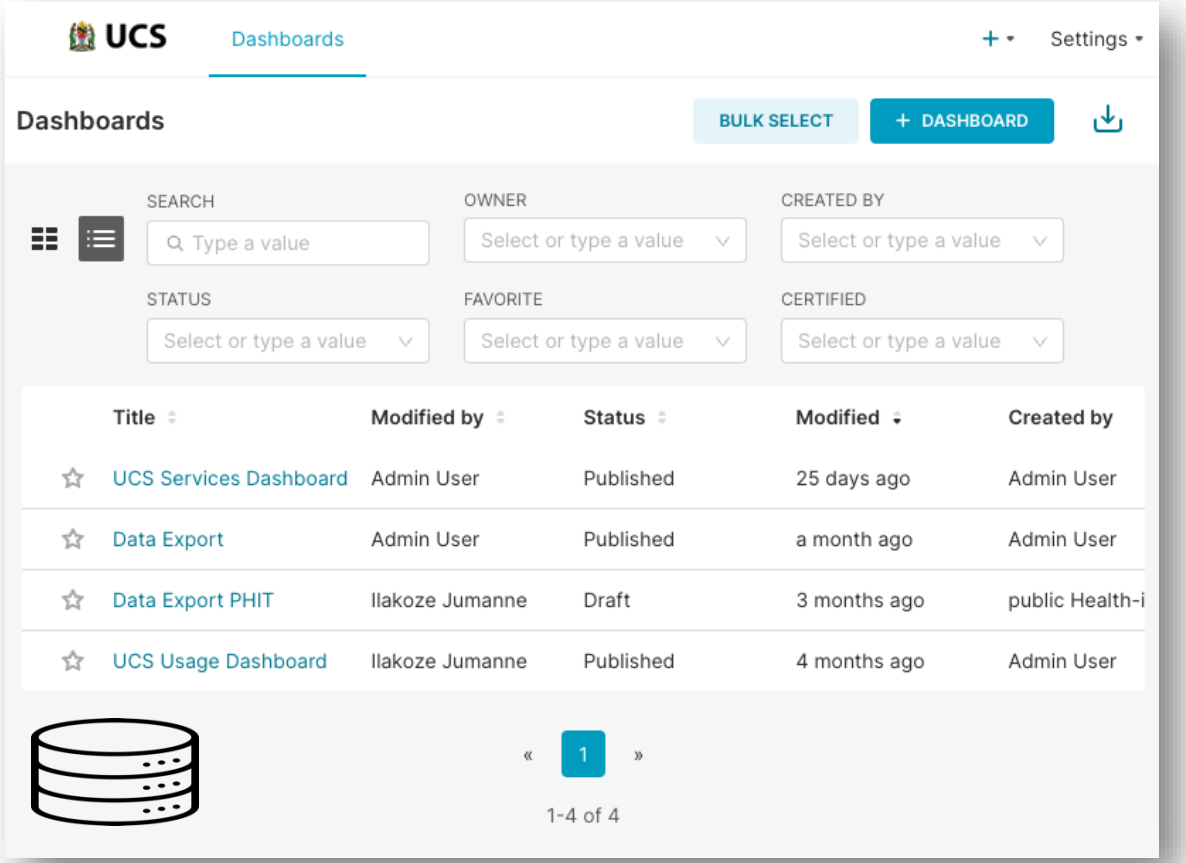




What did we get?

What did we get?

- Data & Collaborations

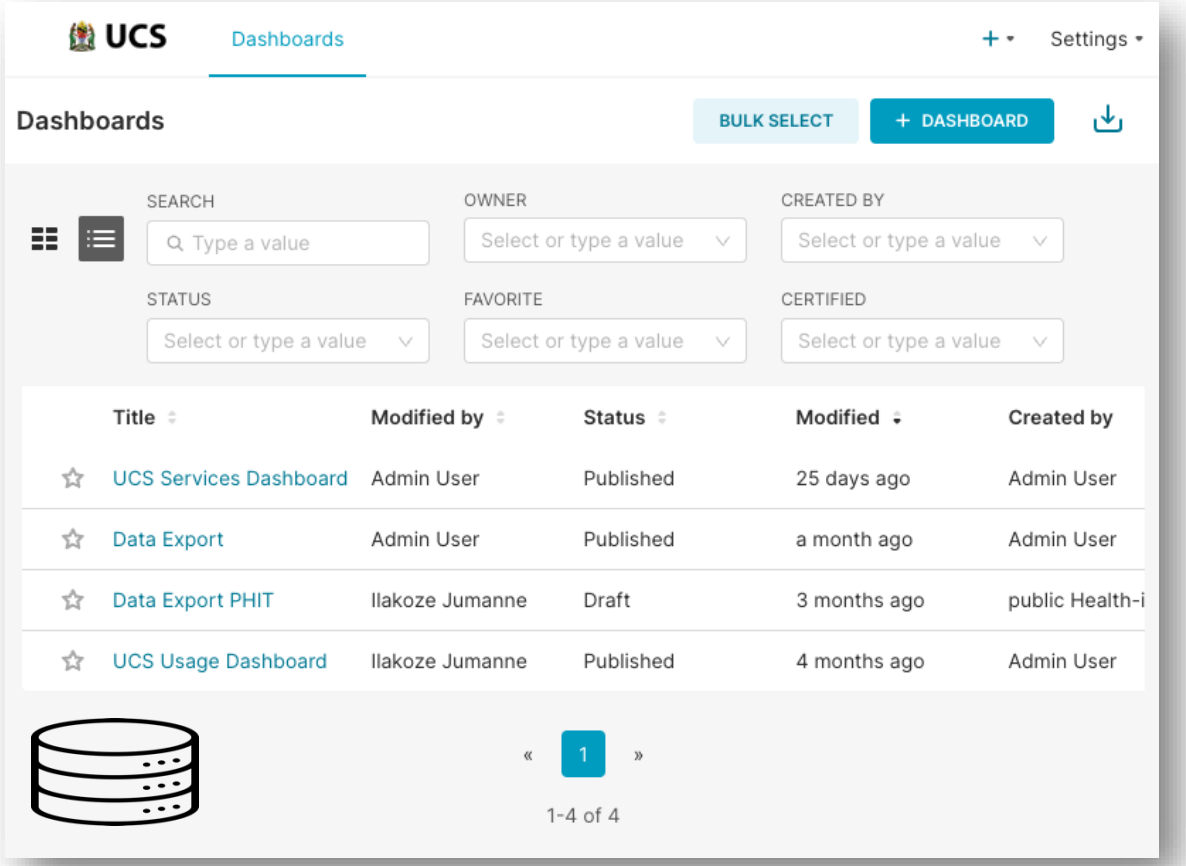


- By March 2024
  - A total of **337,027** ANC Visits have been recorded
  - Multiple visits (longitudinal/repeated observations)
  - Visits ranges from **2020** to **2024**
  - Majority with single visit

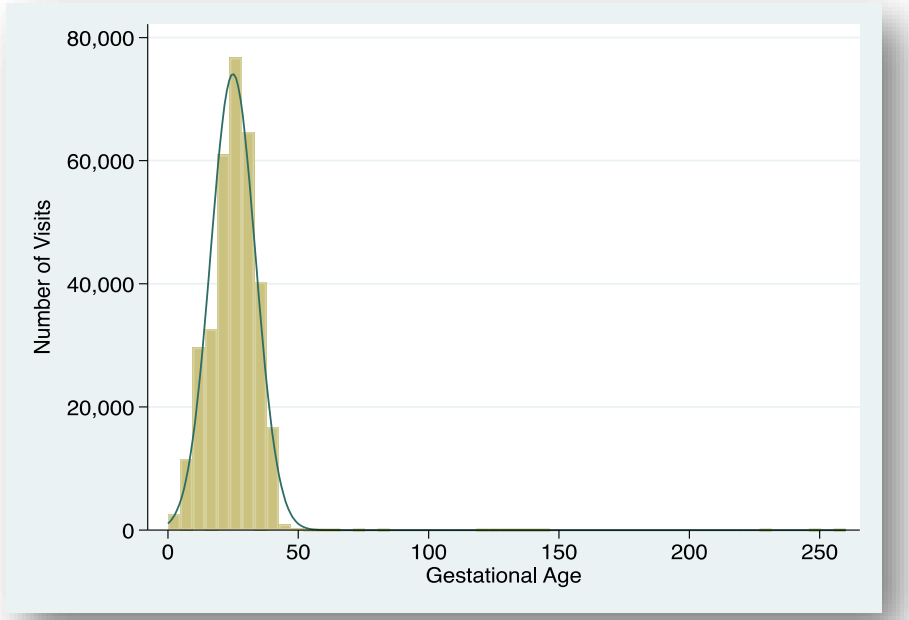
r	Freq.	Percent	Cum.
1	182,313	54.09	54.09
2	104,886	31.12	85.22
3	33,487	9.94	95.15
4	11,606	3.44	98.60
5	3,494	1.04	99.63
6	915	0.27	99.90
7	202	0.06	99.96
8	49	0.01	99.98
9	17	0.01	99.98
10	3	0.00	99.98
11	1	0.00	99.98
99	54	0.02	100.00
Total	337,027	100.00	

What did we get?

- Data & Collaborations



- By March 2024
  - A total of **337,027** ANC Visits have been recorded
  - Multiple visits (longitudinal/repeated observations)
  - Visits ranges from **2020** to **2024**
  - Majority with single visit
  - Average gestational age = 24W with about 80,000 cumulative visits



What did we get?

- Data Quality

```
data1.notnull().sum()/len(data1)*100
```

```
client_id              100.000000
ganc                   100.000000
gest_age                99.986055
blood_for_glucose       23.940218
glucose_in_urine        70.199420
protein_in_urine        70.198827
blood_group             57.584704
syphilis                55.499411
rh_factor               27.805487
visit_number           100.000000
height                  99.923745
weight                  50.419403
bmi                     50.418216
systolic                50.419106
diastolic               50.419403
temperature             50.419106
fundal_height           37.124919
fetal_heart_rate        37.123139
dtype: float64
```

Good, but not very useful

- Good for management purposes

Good,
but dig deeper

```
counts = data1['protein_in_urine'].value_counts(dropna = False)
percs = data1['protein_in_urine'].value_counts(dropna = False, \
                normalize=True).mul(100).round(1).astype(str) + '%'
pd.concat([counts,percs], axis=1, keys=['count', 'percentage'])
```

	count	percentage
negative	173065	51.4%
NaN	100438	29.8%
test_not_conducted	62799	18.6%
positive	725	0.2%

Key variables

ML Model

ML Model

Predictors definition

```
# Collapse: Change data from longitudinal to cross-section
data2 = data1.sort_values(['client_id','visit_number'],ascending = [False, True])
data3 = data2.replace('[null]', np.nan)
data4 = data3.groupby('client_id',as_index = False
                      ).agg(
                              {
                                  'ganc':'last',
                                  'gest_age':'max',
                                  'glucose_in_urine':'last',
                                  'protein_in_urine':'last',
                                  'blood_group':'last','syphilis':'last',
                                  'visit_number':'last',
                                  'blood_for_glucose':'mean',
                                  'height':'mean',
                                  'weight':'mean',
                                  'bmi':'mean',
                                  'systolic':'mean',
                                  'diastolic':'mean',
                                  'temperature':'mean'
                                }
                              )
print('The shape of data before collapsing:', data2.shape)
print('The shape of data after collapsing:', data4.shape)
print('Maximum number of visit per client:')
data4['visit_number'].value_counts().sort_index(ascending=True)
```

```
The shape of data before collapsing: (337027, 18)
The shape of data after collapsing: (187438, 15)
Maximum number of visit per client:
1     80773
2     72448
3     22368
4      8247
5      2609
6       732
7       157
8        33
9        14
10        2
11        1
99       54
Name: visit_number, dtype: int64
```

ML Model

Outcome definition

Hypertensive Disorders During Pregnancy ➔ Sys >= 140 | Dia. >= 90

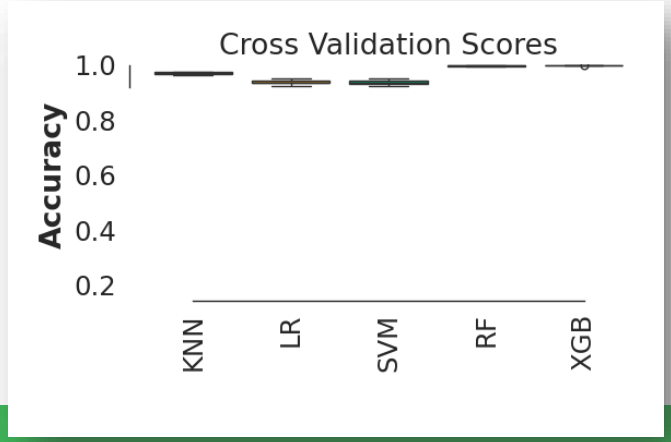
Label	Count
Data with outcome condition (Risk)	1,894
Data without outcome condition (No Risk)	48,680
Total	**50,574**

Count
1,894
1,894
3,788

Balance

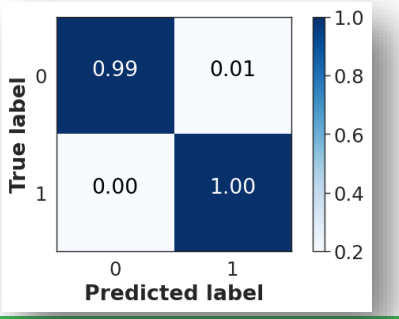
Final Set
379
3409
3,788

10% Test / 90% Train



KNN: 97.07%
LR: 93.78%
SVM: 93.72%
RF: 99.74
XGB: 99.94

→ Train 90%
Test 10% →



Model Use Cases

Model Use Cases → New dataset (N=120,232)

Detect Risk of HDDP using **arithmetic methods**
(Sys. BP >= 140, Dia. Bp >= 90)
Results

- → Risk HDDP = **1,725**,
- → No Risk HDDP = **118,507**

Model Use Cases → New dataset (N=120,232)

Detect Risk of HDDP using **arithmetic methods (Sys. BP >= 140, Dia. Bp >= 90)**

Results

- → Risk HDDP = **1,725**,
- → No Risk of HDDP = **118,507**

Predict Risk of HDDP using XGB **Trained Model**

Results

- → Risk of HDDP = **12,603**,
- → No Risk of HDDP = **107,629**

Model Use Cases → New dataset (N=120,232)

Detect Risk of HDDP using **arithmetic methods (Sys. BP >= 140, Dia. Bp >= 90)**

Results

- → Risk HDDP = **1,725**,
- → No Risk of HDDP = **118,507**

Predict Risk of HDDP using XGB **Trained Model**

Results

- → Risk of HDDP = **12,603**,
- → No Risk of HDDP = **107,629**

Key Discovery

→ The model over detected 10,878 records (12,603 – 1,725) → predictability

→ These 10,878 would not have been captured using convertional routine procedures

Model Use Cases → New dataset (N=120,232)

Detect Risk of HDDP using **arithmetic methods (Sys. BP >= 140, Dia. Bp >= 90)**

Results

- → Risk HDDP = **1,725**,
- → No Risk of HDDP = **118,507**

Predict Risk of HDDP using XGB **Trained Model**

Results

- → Risk of HDDP = **12,603**,
- → No Risk of HDDP = **107,629**

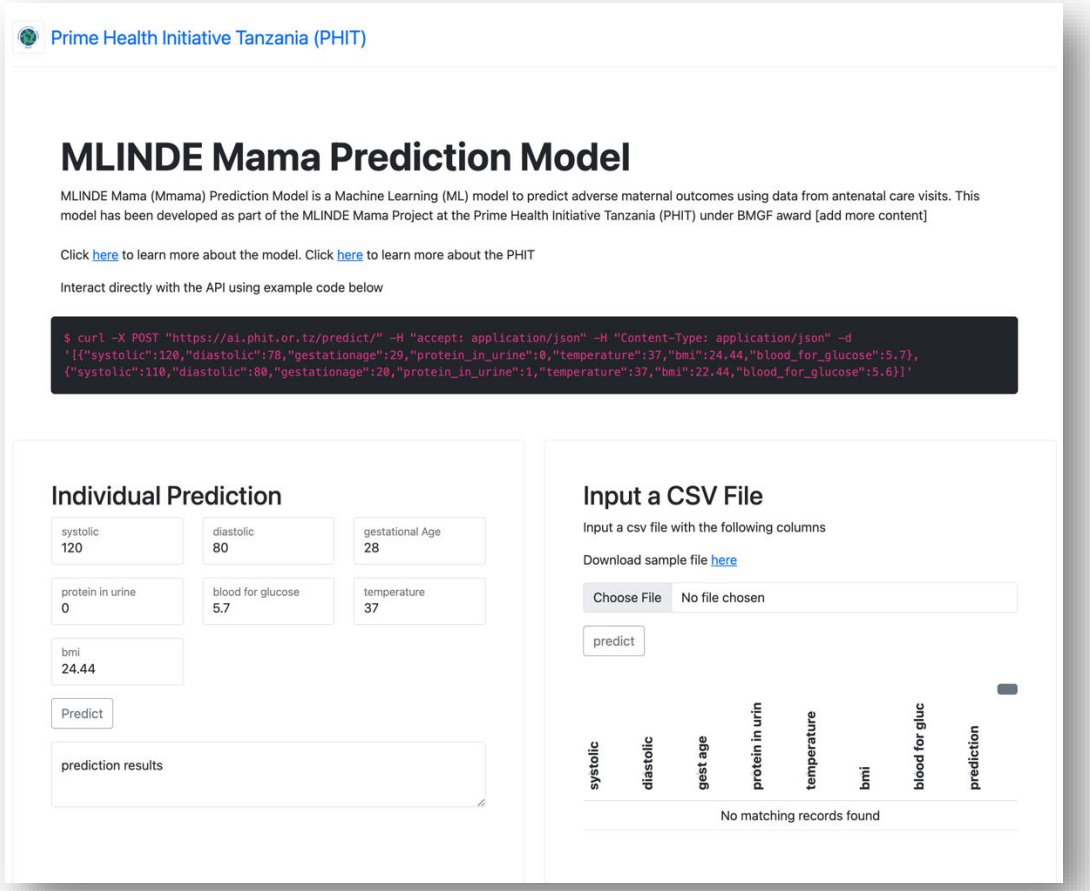
Prediction accuracy: 90.95%
Classification report :

	precision	recall	f1-score	support
No Risk	1.00	0.91	0.95	118507
Risk exist	0.14	1.00	0.24	1725
accuracy			0.91	120232
macro avg	0.57	0.95	0.60	120232
weighted avg	0.99	0.91	0.94	120232

Model Deployment

Model Deployment

https://ai.phit.or.tz



Service Oriented Architecture, (SOA)

- can be consumed by general public
- can be consumed by experts through
  - CSV Upload
  - API Integration

What Next?

- Model Validation with an active follow up
- Model fine-tuning
- Model application in routine services

Acknowledgements

- The Ministry of Health Tanzania
- Geita Regional Administration
- Service providers at facilities & the GANC team
- Gates Foundation for funding support and routine technical discussions

Thank you

Discussions

Additional Slides

Additional Slides

Features importance