

# This project demonstrates how XGBoost Regression models can be used to successfully extract feature importance of the *Matebele* Ant raiding process for entomology research, aiding in termite infestation control. Key variables to consider when using *Matabele* Ants as biological control agents are as follows: number of ants in the raid, foraging distance and raid month.

## Using machine learning to identify key features that drive *Matabela* ant raid effectiveness.

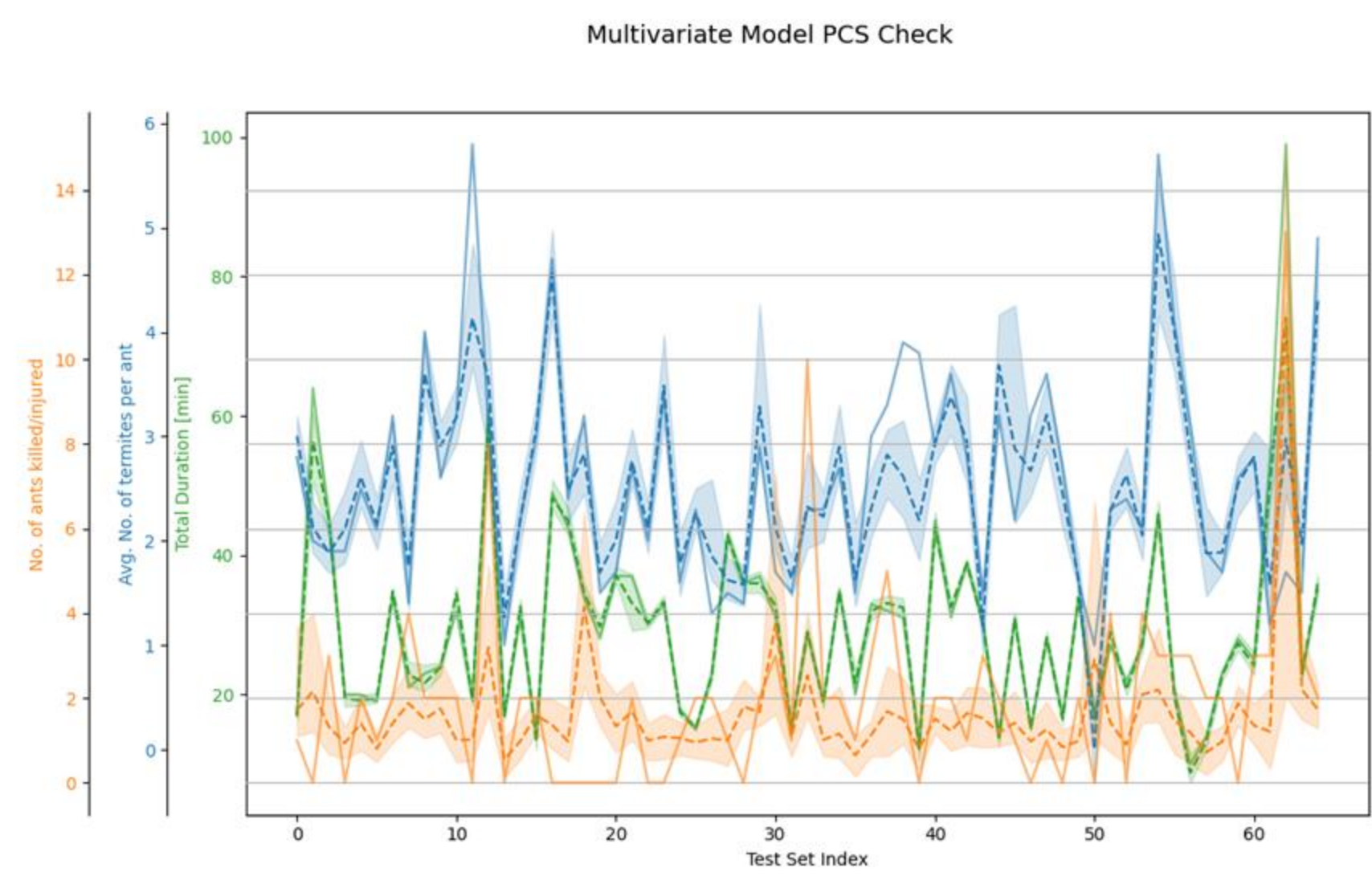
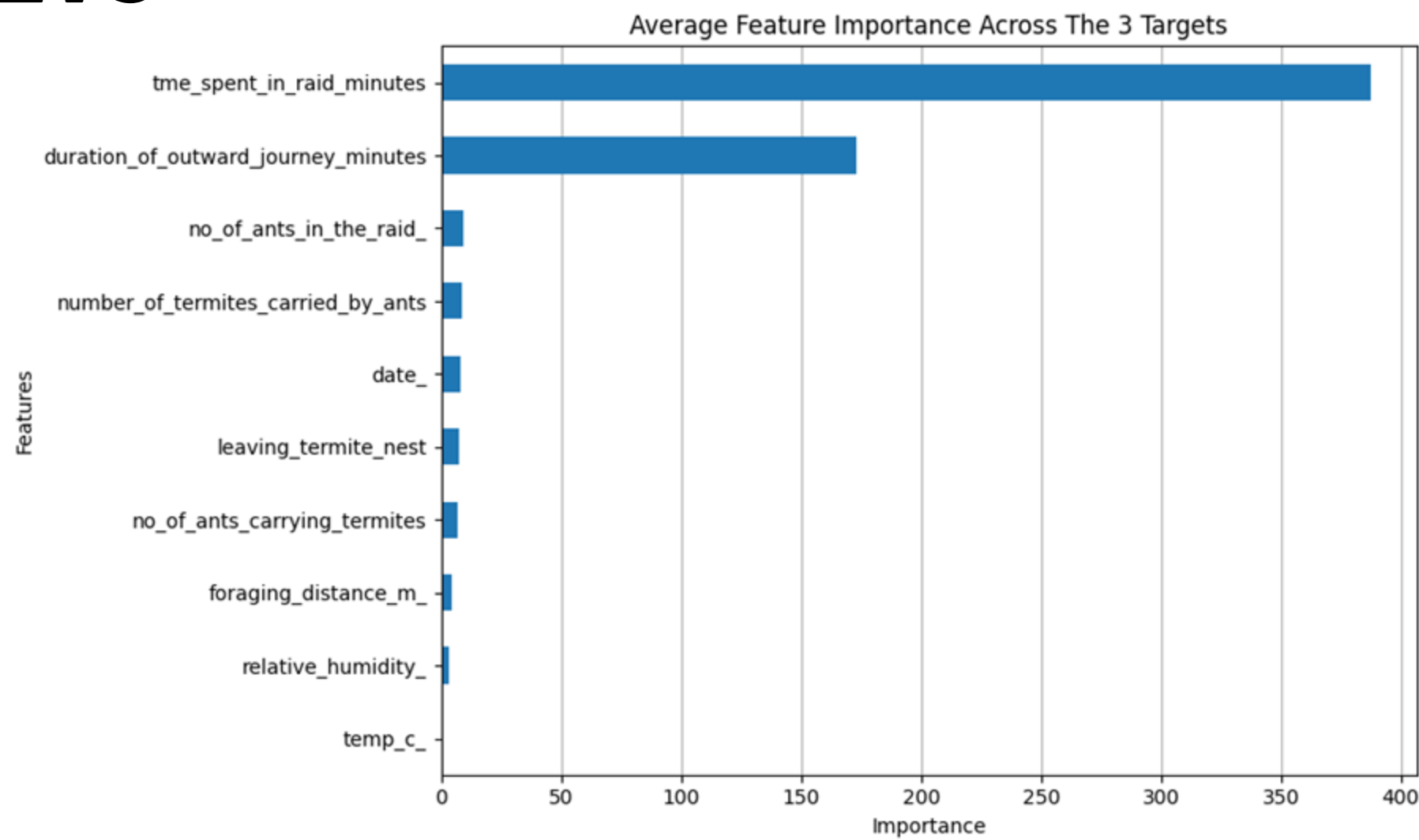
### INTRO

- The world's population is expected to reach 9.1 billion in 2050, increasing the pressure on global food security.
- Ant and termite dynamics have been studied, and few approaches integrate machine learning (ML) to optimize the dynamics for effective termite control.
- We improve Matabele ant raid effectiveness by leveraging ML models to identify key manipulatable features.

### METHODS

- Data.** Employed Matabele ant behavioral data during raids from Mpala Research Center in Kenya.
- Process.** Explored feature importance on various models to identify key features that drive ant raid effectiveness.
- Model.** XGBoost model was used to predict our 3 success metrics namely, raid duration, number of ants injured/killed and average number of termite per ant.
- Deployment.** Use Streamlit to build an interactive user interface where users can simulate real-life scenarios to observe how they affect the 3 predicted variables.

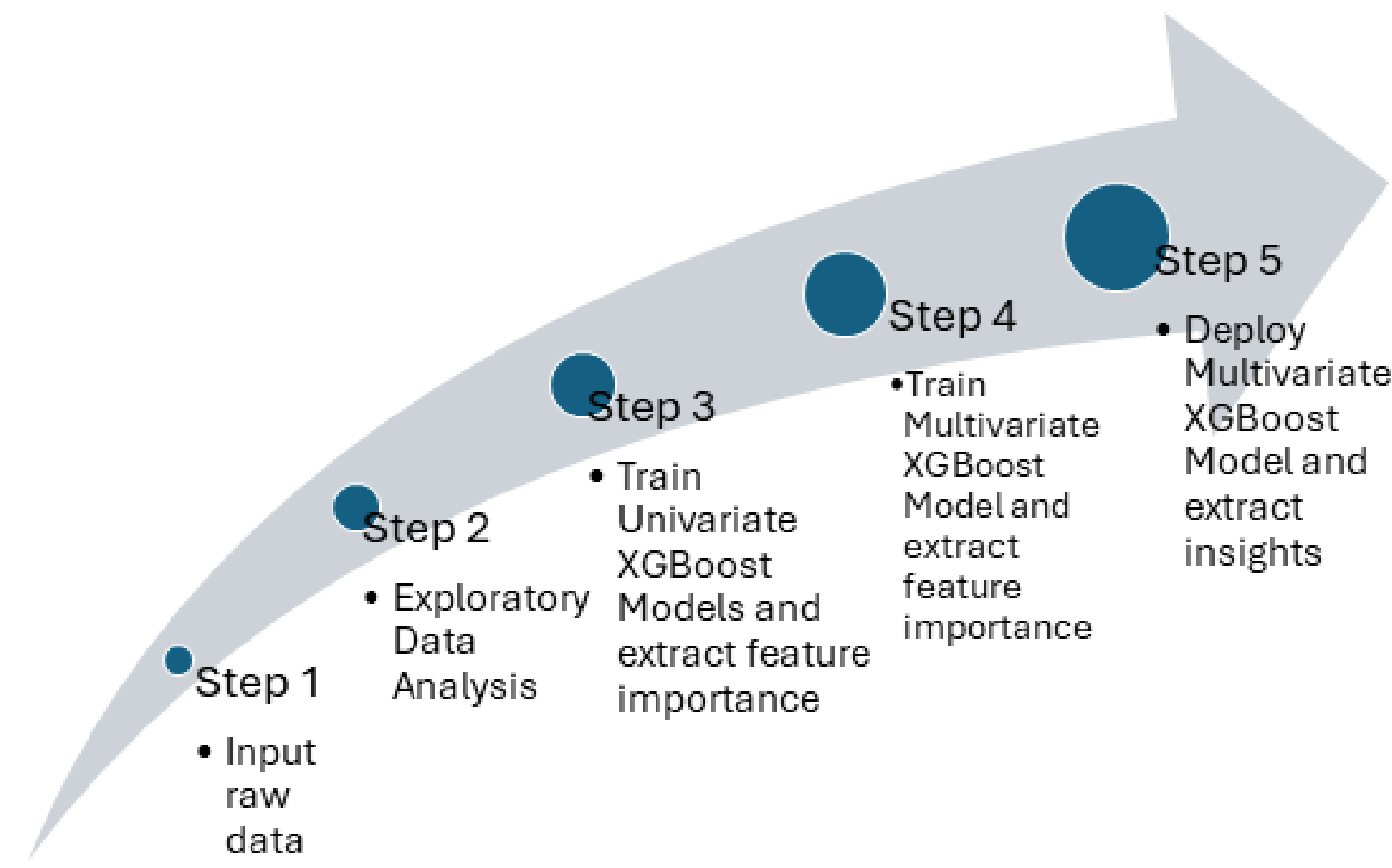
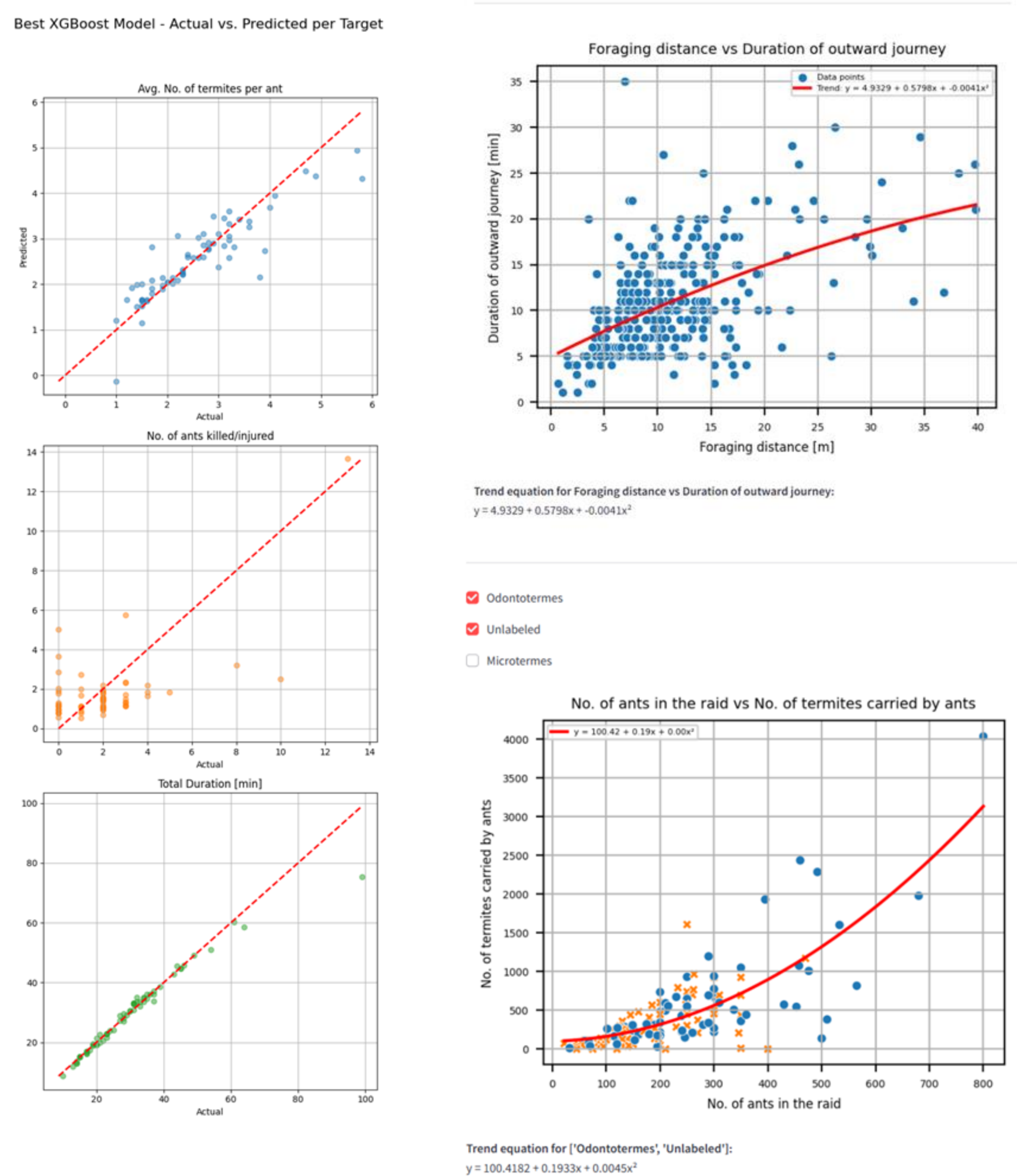
### RESULTS



### DISCUSSION

- The XGBoost model scored 70% accuracy.
- We found the number of ants in the raid, foraging distance and raid month to be key features.
- Resampling with replacement used to do PCS check.
  - Shaded regions represents the 5th to 95th percentile of predicted values across 100 bootstrapped models; 90% perturbation for each predicted test set sample
  - Predictability - Median and perturbation interval track the actual values well, thus prediction variation is acceptable
  - Computability - usable model is deployed with standard computational resources
  - Stability - Shaded regions are not too wide indicating good stability

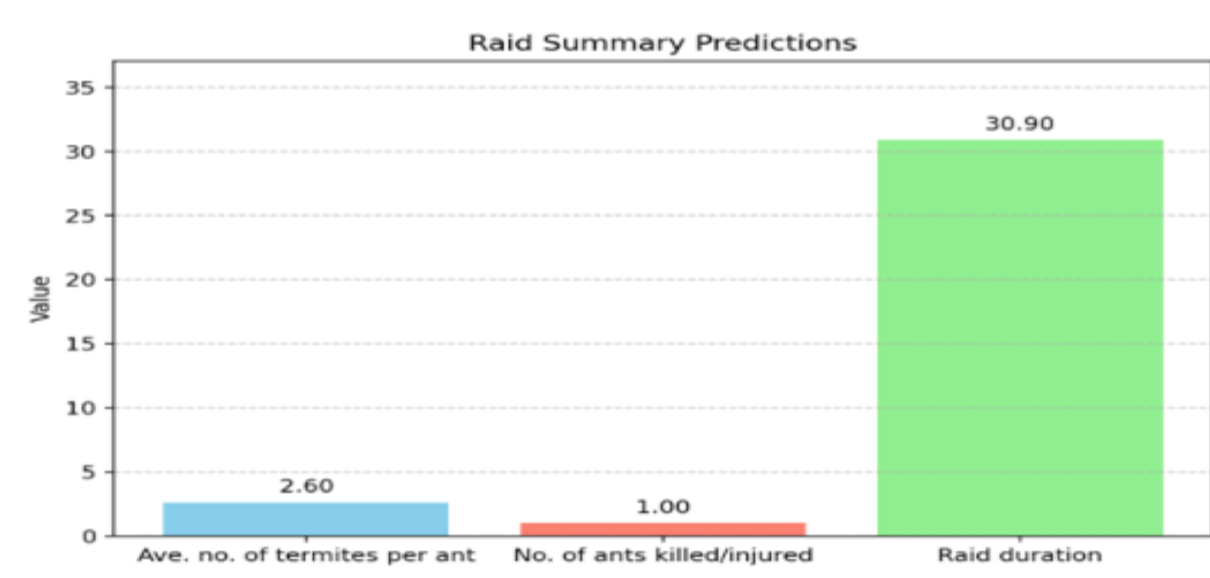
Mzwakhe Didishe, Realeboga Monare



#### Adjust variables:



#### Prediction:



Department of Computer Science

Faculty of Engineering,  
Built Environment and  
Information Technology  
Fakulteit Ingenieurswese, Bou-omgewing en  
Inligtingtegnologie / Lefapha la Boetšenere,  
Tikologo ya Kago le Theknolotši ya Tshedimošo

Capstone Project - MIT 808

Course Coordinators:  
Dr. Vukosi Marivate (vukosi.marivate@cs.up.ac.za)  
Abiodun Modupe (abiodun.modupe@cs.up.ac.za)

