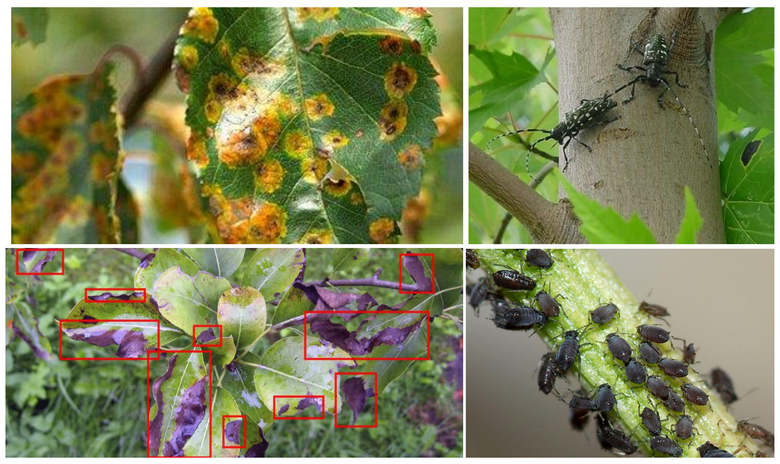
FABI Plant Diseases Diagnostic Clinic Model for Advanced Diagnostic Services

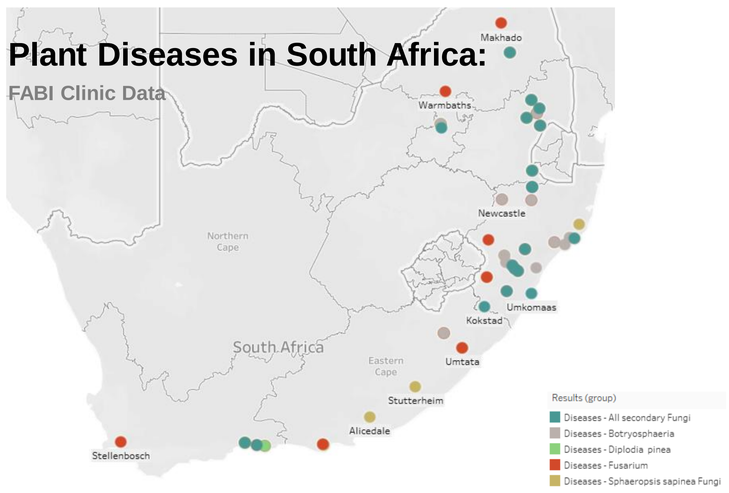
Machine learning model on clinic data about pests and diseases.

MOTIVATION

- Plant pests and diseases cause global damage to forestry plants and crops.
- The distribution of dense forestry plantations makes it difficult for companies to monitor and manage the spread of pests and pathogens



The University of Pretoria Forestry and Agricultural Biotechnology Institute (FABI) Disease Clinic and its partners; Institute of Commercial Forestry Research (ICFR), produced an initiative to tackle the spread of pests and diseases in South Africa and the world at large.



OBJECTIVES

The purpose of this project is to Clean and Integrate FABI and ICFR datasets with GIS dataset and use the data to develop a machine learning model that will perform pre-diagnosis of new samples.

Supervised machine learning models used to:

- determine the likelihood of the damage that a pest or a pathogen can cause to plants
- predict plant pests and diseases to help farmers to make informed decision on pest control mechanism, farming methods and to maintain plant health

Freddie Hove, **Sabelo** Madondo

METHOD & RESULTS

1. Project Scoping & Exploratory Data Analysis

Defining the project scope, objectives and constraints	Data cleaning and integration using data analytics tools	interrogating the data and identifying the occurrence and trends

Data Sources:

- FABI Clinic Data → Plantation Location, Genus Type, Company Name, Sample Type, Diagnosis/Results
- ICFR Sirex Data → Plantation Location, Genus Type, Company Name, Sample Type, Diagnosis/Results
- GIS Data → Plantation Location, Company Name

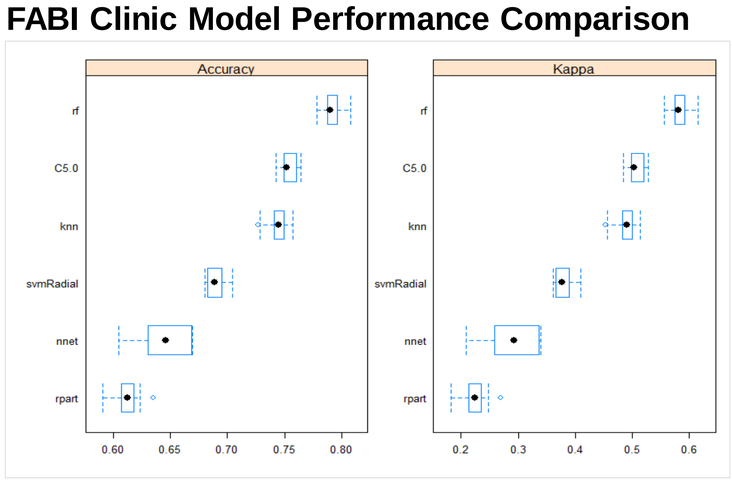
2. Management & Modelling

Construction of machine learning models	Evaluation of the model accuracy and stability	Model selection and validation

Best Model = Random Forest

	Accuracy	Kappa
Sirex Model	0.95	0.98
FABI Clinic Model	0.80	0.58

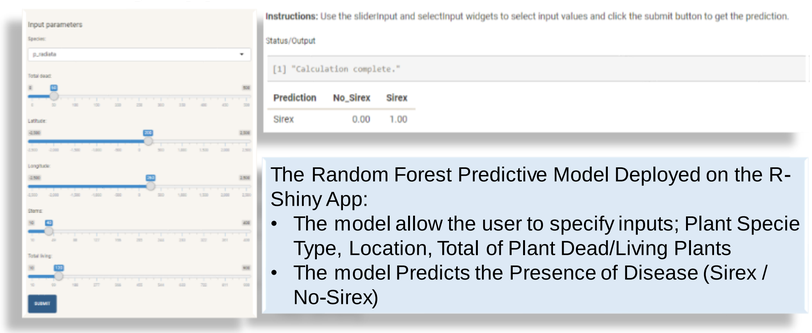
FABI Clinic Model Performance Comparison



3. Visualisation & Deployment

Visualise trends, outliers and patterns	Visualise predictive model results	Deploy in R-Studio and R-Shiny App

R-Shiny App:

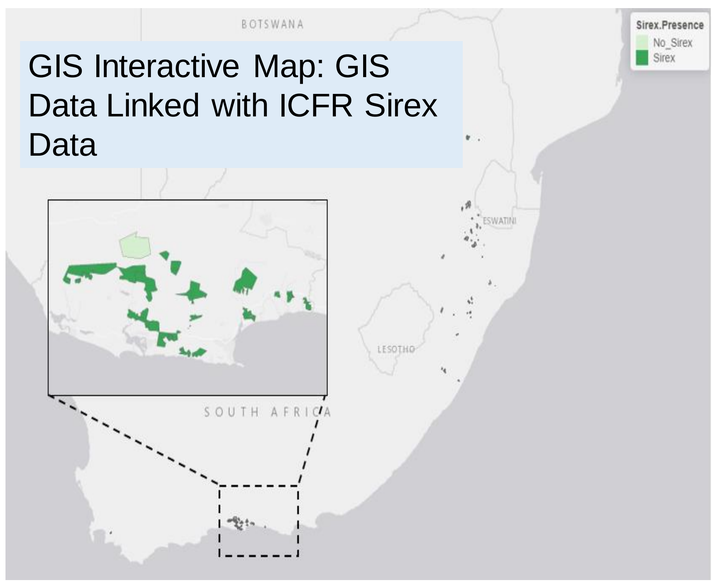


The Random Forest Predictive Model Deployed on the R-Shiny App:

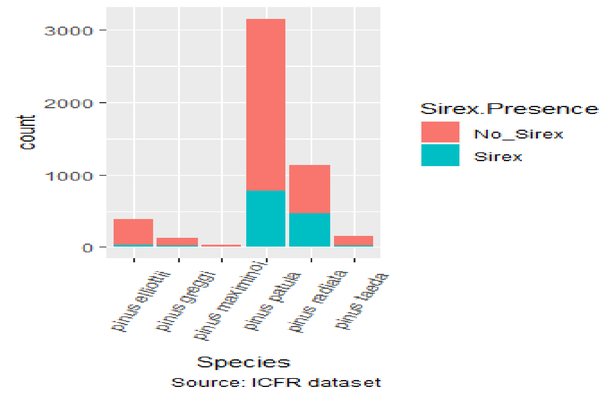
- The model allow the user to specify inputs; Plant Specie Type, Location, Total of Plant Dead/Living Plants
- The model Predicts the Presence of Disease (Sirex / No-Sirex)

VISUALISATION

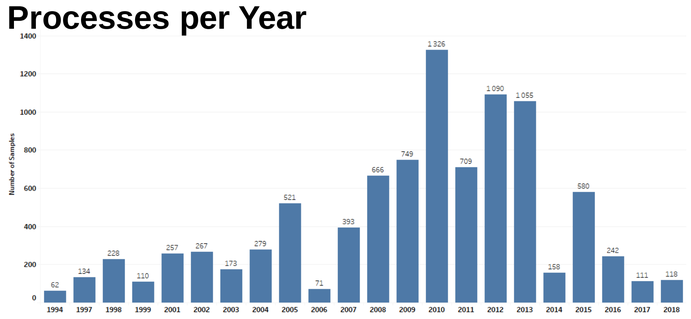
Plantation Location



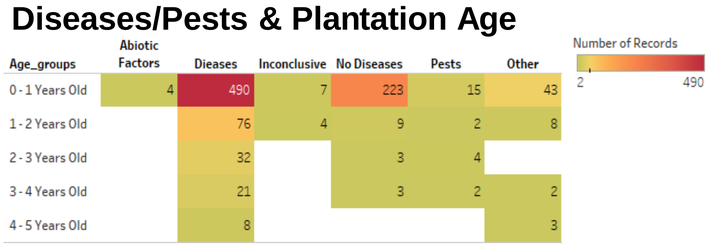
Sirex Presence by Species



FABI Clinic Data: Number of Samples Processes per Year

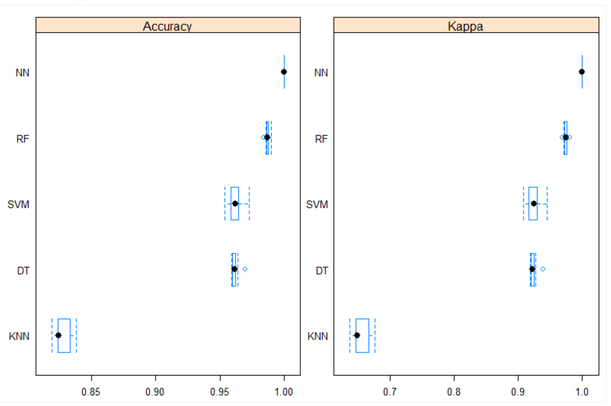


FABI Clinic Data: Presence of Diseases/Pests & Plantation Age



Age_groups	Abiotic Factors	Dieases	Inconclusive	No Diseases	Pests	Other
0 - 1 Years Old	4	490	7	223	15	43
1 - 2 Years Old		76	4	9	2	8
2 - 3 Years Old		32		3	4	
3 - 4 Years Old		21		3	2	2
4 - 5 Years Old		8				3

ICFR Sirex Model Performance Comparison



Take a picture to learn more about FABI