Slope, Altitude and planted Area are most useful class separation features. Logistic regression model outperforms the decision tree model.

Beating The Beetle

Paddington Chiguvare, Masana Khosa

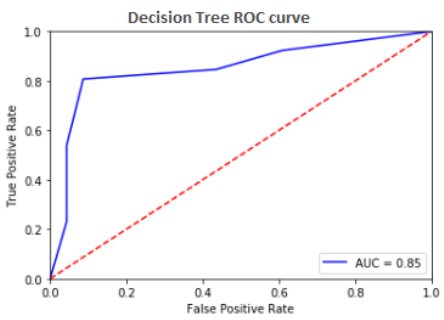
1 Intro

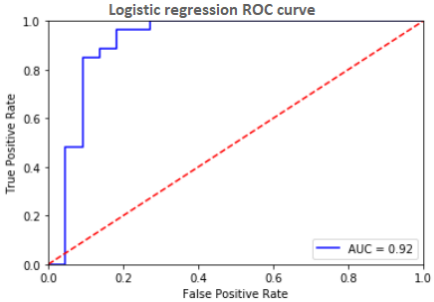
- Eucalyptus plantations in South Africa have been infected by Eucalyptus Beetles.
- A dataset containing comprehensive information about infected and healthy compartments was provided.
- The aim is to develop classification models to classify compartments as healthy or unhealthy and predict sites likely to be threatened.

2 Methods

- Logistic regression and decision tree classification models were developed.
- logistic regression model is given by: $\log\left(\frac{(p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$
- To ensure stability, cross-validation was used to train the models. The models were tested using the testing set.
- PCS framework was used to test reliability, reproducibility and transparency.
- Area Under ROC curves were used to measure performance.

3 Results





- Logistic regression model has a better performance.

Extra figures

Performance comparison

Cross-Table for Decision Tree Model

	Predicted Positive	Predicted Negative
Actual Positive	16	6
Actual Negative	6	21

Cross-Table for Logistic Regression Model

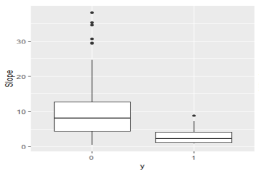
	Predicted Positive	Predicted Negative
Actual Positive	16	6
Actual Negative	0	27

Performance Comparison of the models

	Precision	Recall	Accuracy
Decision Tree	0.75	0.75	0.88
Logistic regression	0.91	0.86	0.75

Useful features

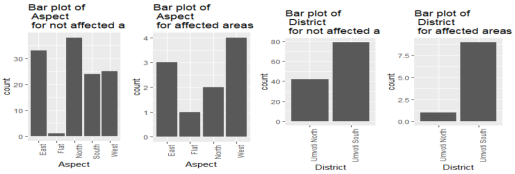
- Slope and Altitude are numerical features useful in class separation since their box plot notches do not overlap.



Effect of hill slope gradient on class separation

Effect of altitude on class separation

- Aspect and District are categorical features useful in class separation because they have different categorical distributions.

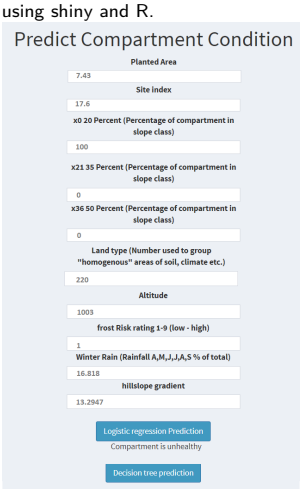


Effect of aspect on class separation

Effect of District on class separation

Deployment of Models

- The two classification models were deployed using shiny and R.



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
UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Capstone Project - MIT 808

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