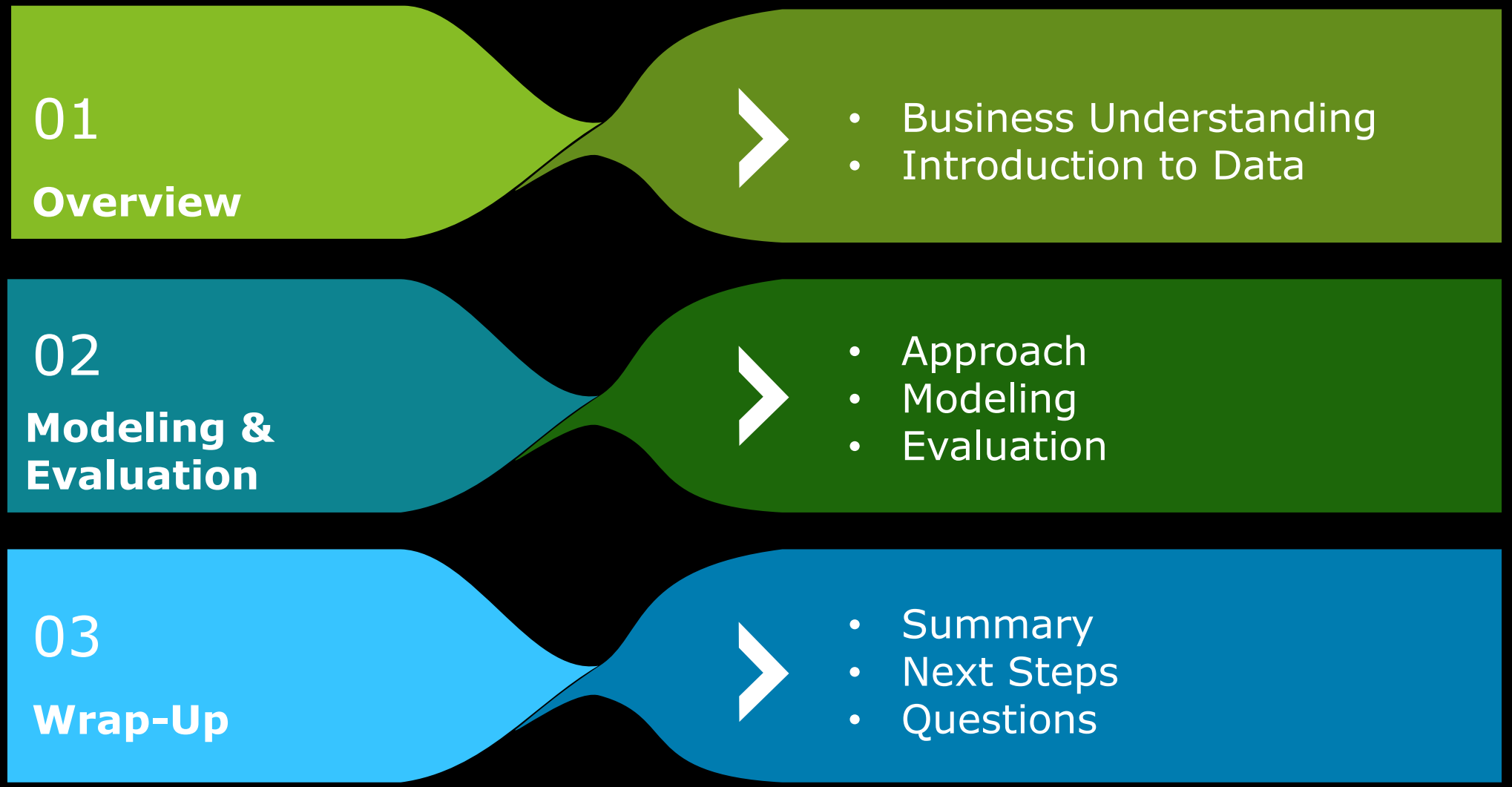
Deloitte.

Classifying Breast Cancer Tumor categories

Capstone Project

MAKING AN IMPACT THAT MATTERS since 1845

Agenda


01
Overview
• Business Understanding
• Introduction to Data
02
Modeling & Evaluation
• Approach
• Modeling
• Evaluation
03
Wrap-Up
• Summary
• Next Steps
• Questions

Business Understanding & Problem

Business Understanding

- Breast cancer is the most common cancer in women, causing over 40,000 deaths every year
- One of the main challenges for radiologists is to diagnose breast cancer
- Early detection of breast cancer is key and can help improve the chances of survival

Business Problem

A hospital wants to build a model that detects if the cancer diagnosed is malign or benignant

Data Understanding

- The dataset has been pulled from Kaggle Website
- It consists of characteristics of the tumors

Data Understanding

Dataset Characteristics

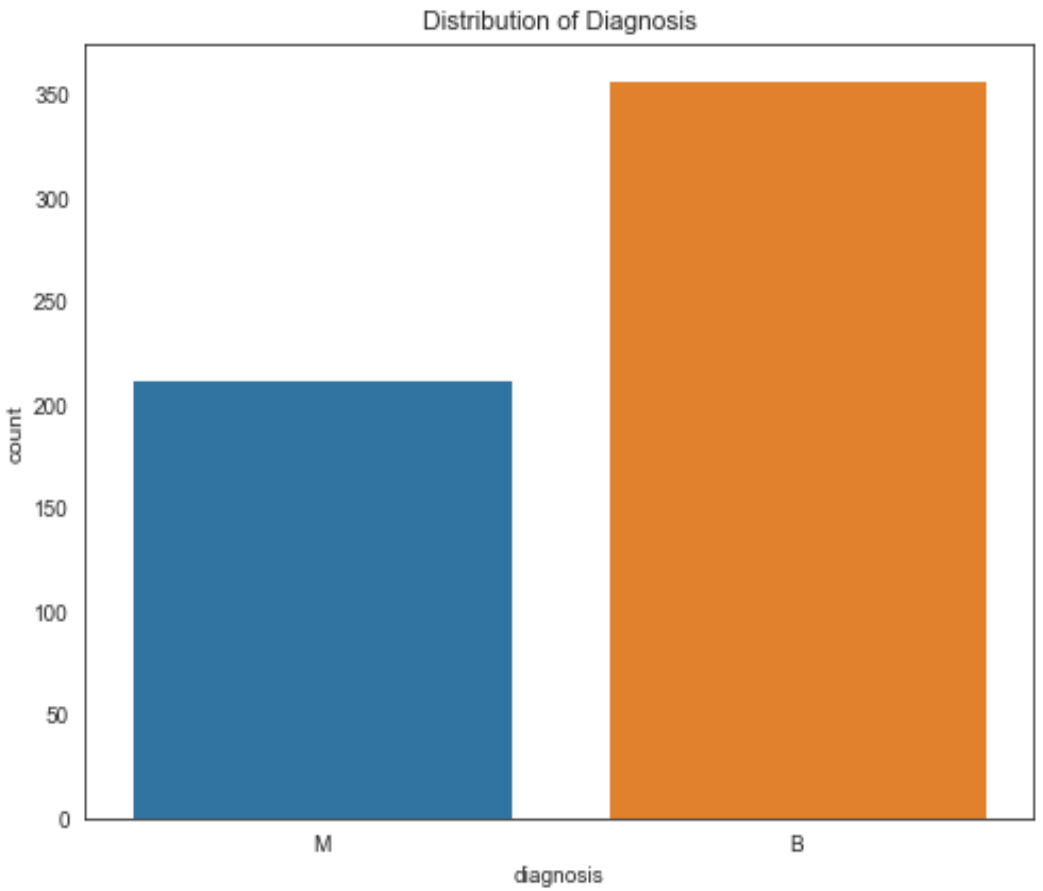
569 Rows

33 Columns

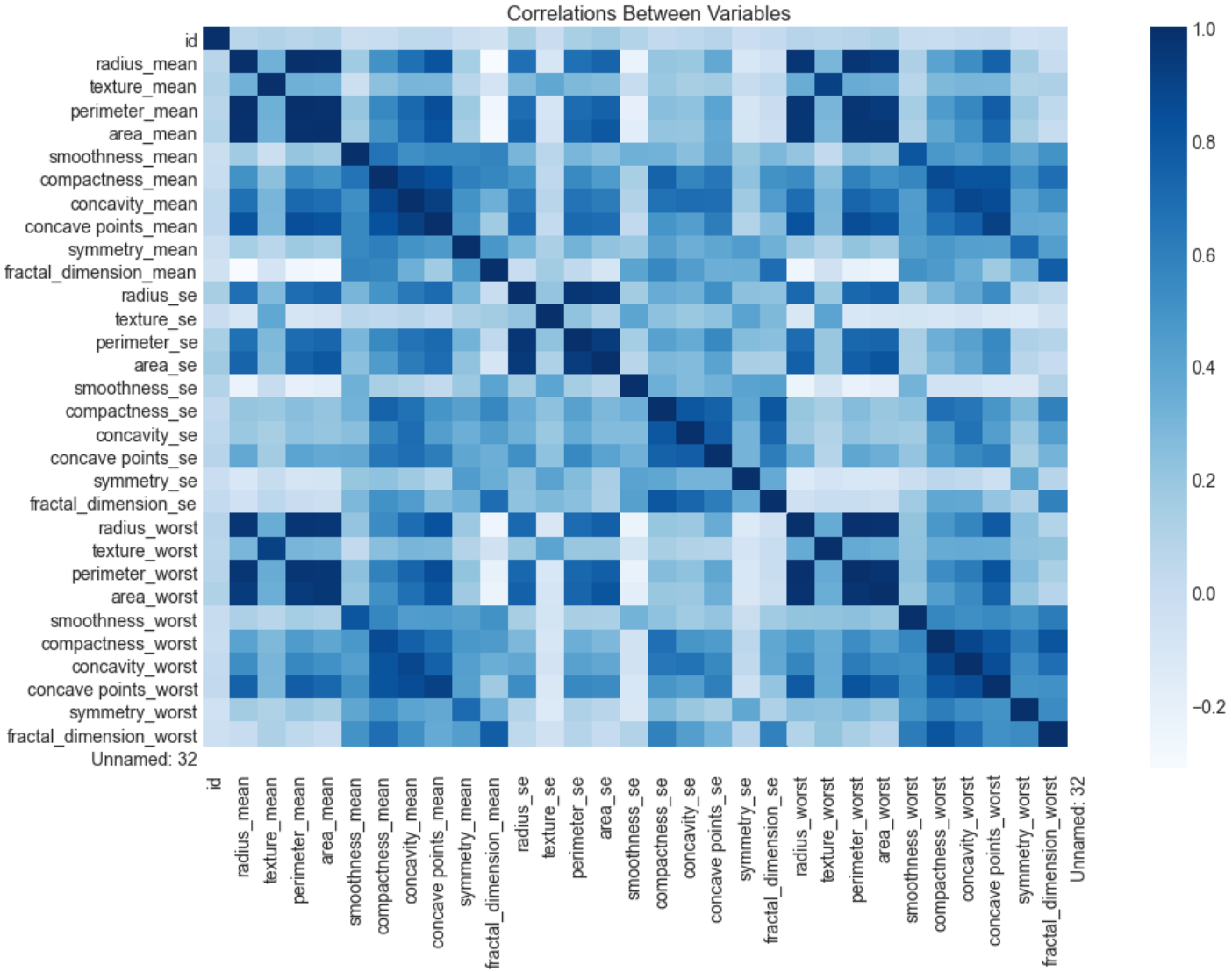
Target Variable

357 Benign

212 Malignant



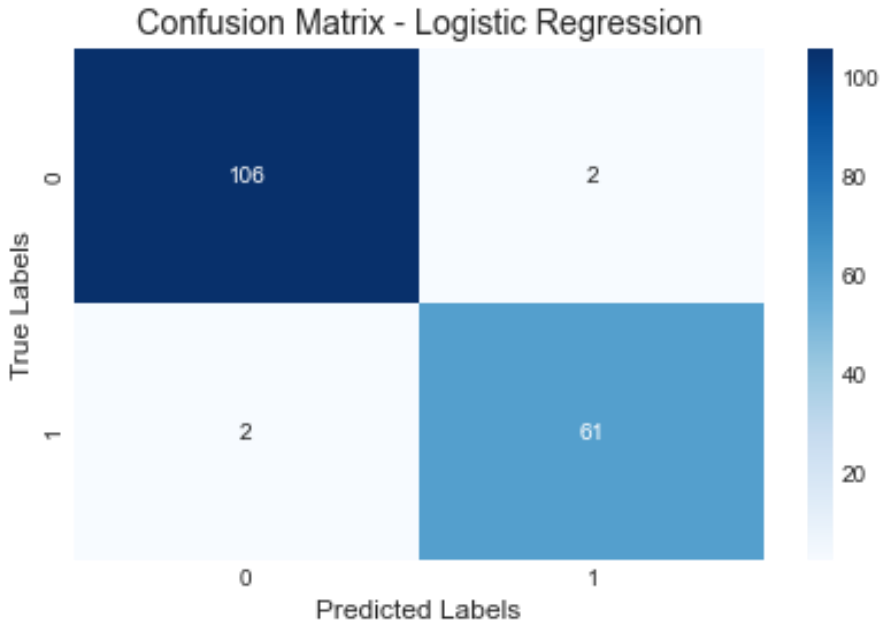
Data Understanding



Multiple columns are highly correlated, causing multicollinearity among independent variables.

Modeling & Evaluation

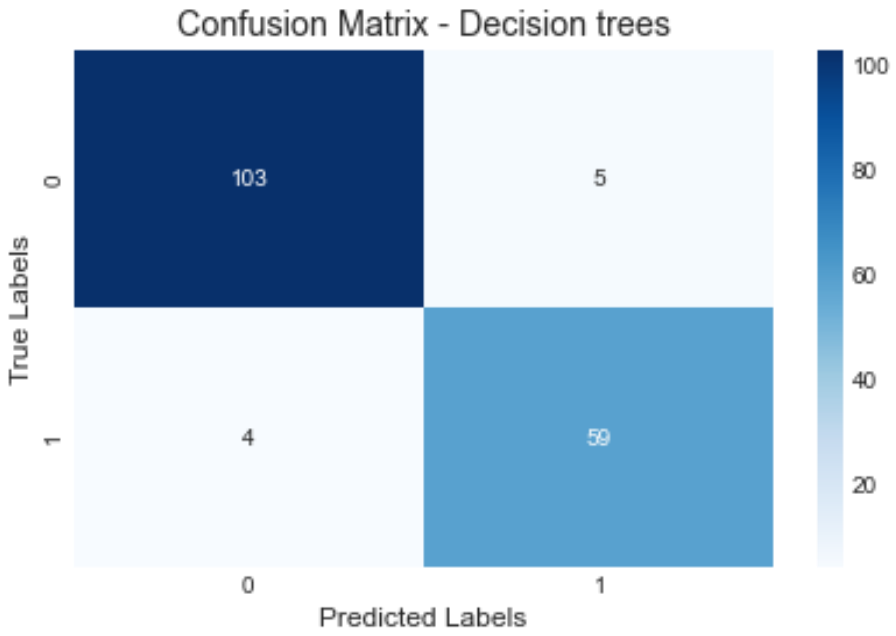
Logistic regression



Labels	Recall score
0 – Benign	0.98
1 - Malignant	**0.97**

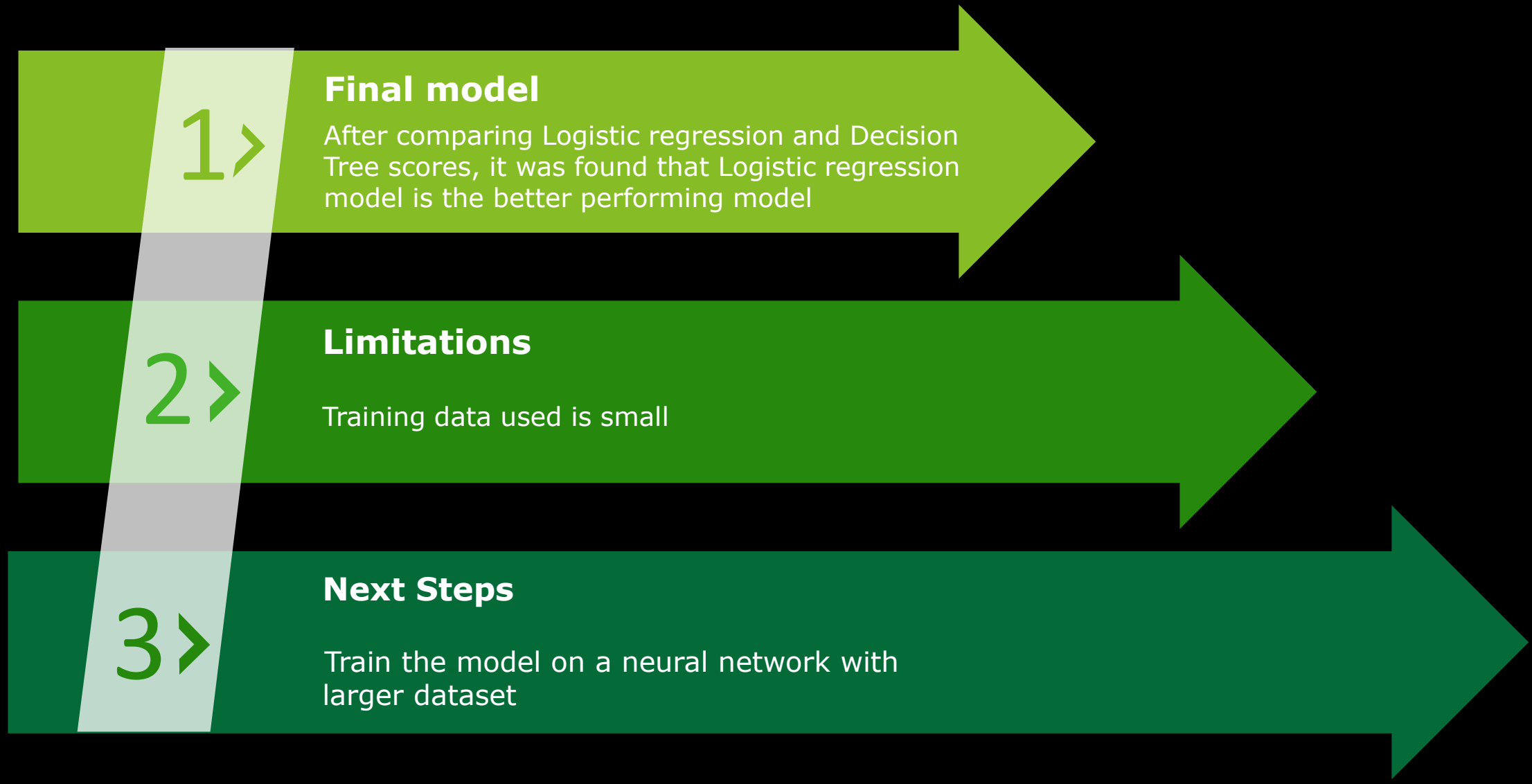
Modeling & Evaluation

Decision Tree – After Optimization



Labels	Recall score
0 – Benign	0.95
1 - Malignant	**0.94**

Conclusion


1
Final model
After comparing Logistic regression and Decision Tree scores, it was found that Logistic regression model is the better performing model
2
Limitations
Training data used is small
3
Next Steps
Train the model on a neural network with larger dataset