# Chapter 8: Improving the model

## Introduction

Contents for this chapter include an introduction and a focus on aspects that improve model acceptability: accuracy, robustness, simplicity and explainability, followed by a summary.

Figure 1: Contents

A green and black circular logo

Description automatically generated

Model development can be a very time-consuming process, varying techniques can be applied and parameters modified. Regardless of the technique used all most should be accurate, robust, simple and

Explainable.

Figure 2: Introduction

A screenshot of a computer

Description automatically generated

What do these things mean in practice? Accuracy relates to model predictions and models should predict new data well.

A decision tree is no exception and the thrust to concentrate observations into the dependent variable categories can lead to some leaf node with all observations concentrated into one of the dependent variable categories.

This is in most cases questionable and should be viewed with caution as to inclusion. Auto-grow with relatively small node size and or lots of variables may suffer this.

Figure 3: Accuracy

A screenshot of a computer

Description automatically generated

Robustness relates to the existence of model segments in the population. For example, a variable that has many splits may not replicate in the population making some of them redundant. To ensure segments are robust they should be of adequate size, a guide being between 1 - 2% of the dataset size but smaller sizes may be considered if they can be justified.

An additional support to ensure robustness is to limit the nodes resulting from any split to be between

4-6, this should ensure segments generalize.

Figure 4: Robustness

A screenshot of a computer

Description automatically generated

A good model should also be simple and explainable. Removing insignificant predictors and

splits that do not further improve predictive accuracy should be top of the list.

The illustration to the right illustrates that additional splits do not improve the model and only

increases complexity.

A comparison of a diagram

Description automatically generated

Model should also be explainable. Results should be relatable and segment characteristics acceptable.

 A complex model will have implications for deployment and may impede getting a model into production unnecessarily.

The elements of accuracy, robustness, simplicity and explainability can be taken into acccount when

building the model and can be further refined during model evaluation and validation.

## Summary

## This lesson focused on model improvement, included an introduction and highlighted aspects that

focus on improving model acceptability: accuracy, Robust, simplicity and explainability