Scientia Horticulturae 209 (2016) 201–213

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Predictive models in horticulture: A case study with Royal Gala apples

T.M. Logan[a,*], S. McLeod[b], S. Guikema[a]

[a] Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, USA
[b] Mr. Apple New Zealand Ltd., Station Rd, Whakatu 4172, New Zealand

ARTICLE INFO

Article history:
Received 6 April 2016
Received in revised form 20 June 2016
Accepted 21 June 2016
Available online 5 July 2016

Keywords:
Predictive modelling
Statistical learning
Regression tree
Ensembles
Agriculture
Crop forecasting

ABSTRACT

Decision makers in horticulture want to forecast their crop characteristics. Predictions of the crop inform decisions which influence pricing, marketing, logistics, and even consumer satisfaction. This article summarises predictive horticultural models in the literature, and finds confusion exists between predictive and explanatory models. It encourages the use of statistical learning and nonlinear methods for future predictive models. Then it demonstrates how predictive models can be constructed using data for Royal Gala apples from orchards within New Zealand. For the eight years of data available the model has been shown to have a mean predictive error of 6.7%. The best model was an ensemble of a linear model (GLM), a Bayesian additive regression tree (BART), and a boosted classification and regression tree (boosted CART). Statistical learning techniques present substantial opportunity to the horticultural industry and to future attempts to develop more accurate predictive models.

1. Introduction

Predictive models in the apple industry benefit growers, packers, shippers, and buyers (Hughes et al., 1996). For all crops, predictive models can improve the efficacy of the decision making processes and improve the output for the growers. As a result, there has been substantial industry and academic interest in building predictive models for crops from apples, to coffee, to melons (Castro-Tanzi et al., 2014; Hughes et al., 1996; Rad et al., 2015). Statistical and data-mining tools are increasingly powerful, presenting an opportunity for the horticultural field. However, it is important that the difference between explanatory and predictive modelling be understood by the community (Shmueli, 2010). Explanatory models aim to understand the causal relationships between variables, whereas predictive models attempt to find empirical relationships which provide good estimates (Shmueli, 2010). There is significant interest in predictive models for horticultural application, however the validation methods and accuracy metrics used in much of the literature are appropriate for explanatory models, not predictive ones. When creating predictive models, there are two essential elements for validation of predictive accuracy:

1. Validate on a test set, unseen during the training process (cross-validation)
2. Use out-of-sample metrics such as mean square error, rather than in-sample R^2 (Shmueli and Koppius, 2010).

Some publications do successfully validate their predictive models, but if they considered other types of models, their predictive accuracy may have been improved (Table 1). Models such as bagging, boosting, random forests, and Bayesian alternatives have been described as the most significant development in statistical learning (Seni and Elder, 2010); the horticultural industry is poised to take advantage of this opportunity.

To demonstrate predictive modelling, a case study analysing the mean weight of apples from a region in New Zealand is presented. The mean harvest fruit weight provided in that data is at block level, which is a collection of trees within an orchard. The weight influences many important management decisions for the apple industry. On the international market, larger fruit sell for higher prices, and packaging is dependent on the size of apples.

The horticultural industry is vital for New Zealand. In 2014 exports exceeded $4 billion dollars, with more than $500 million coming from apple exports. In dollars per hectare the industry produces approximately 32,000 NZD which is 300% that of NZ's dairy industry. There is room for environmentally and economically sustainable growth in this industry (Cately, 2015). Negotiations with potential buyers begins long before harvest, when little is known about the pertinent crop. Apples are exported to the United King-

* Corresponding author.
E-mail address: tomlogan@umich.edu (T.M. Logan).

http://dx.doi.org/10.1016/j.scienta.2016.06.033
0304-4238/

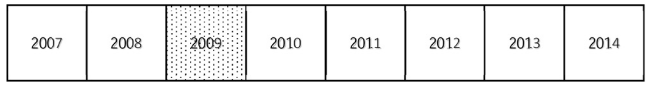

Fig. 1. A data set where the data for 2009 is assigned as the holdout. The models are trained on the remaining years, then tested on 2009.

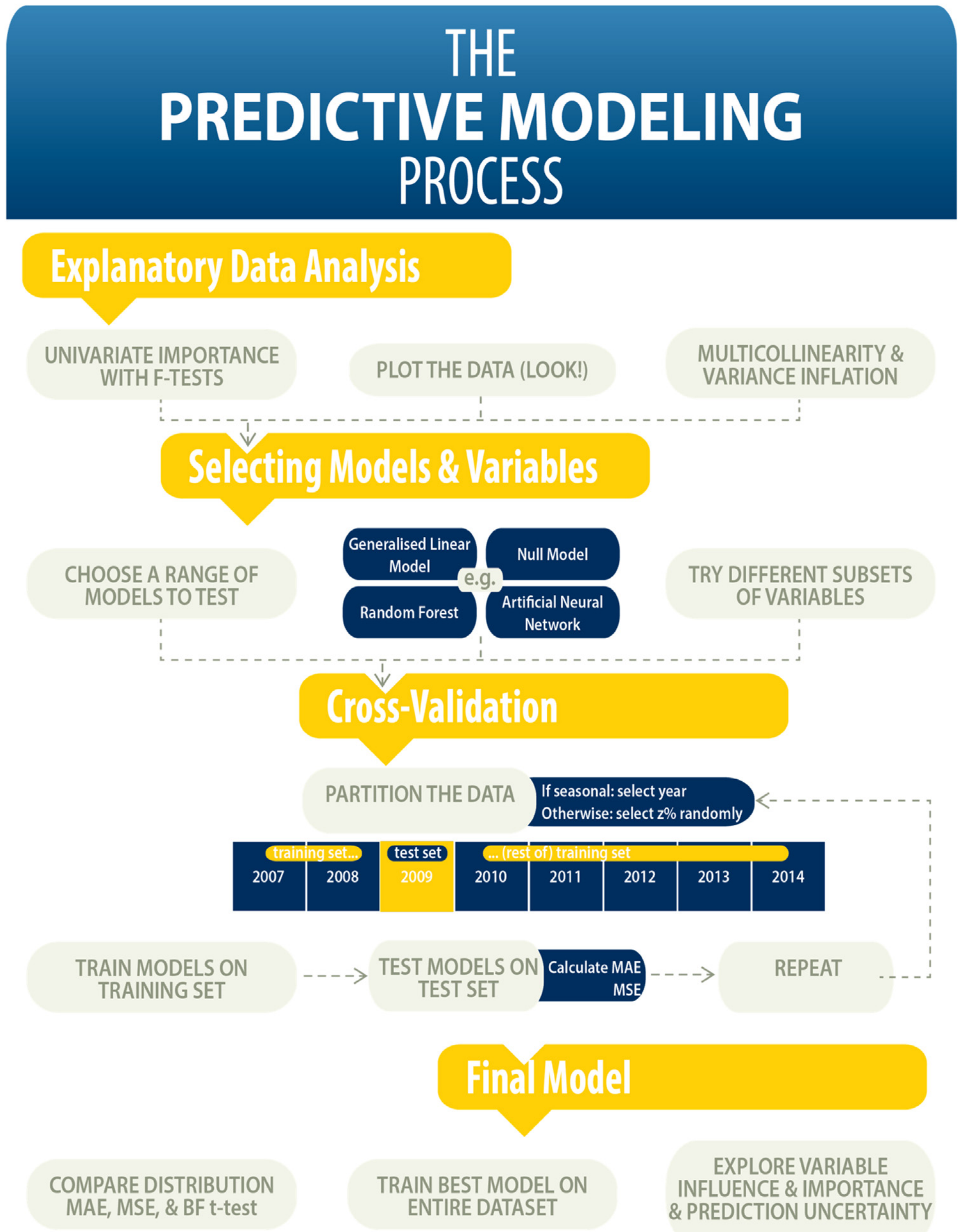

Fig. 2. The process of predictive modelling.

dom, Continental Europe, Asia, and North America (Aitken and Hewett, 2013).

In New Zealand, *Malus domestica* (Royal Gala) is the most common variety, comprising of 27.7% of planted area. Braeburn – the second most widely planted – accounts for a 16.4% of planted area (Lee-Jones, 2014). On a per weight basis, 33% of NZ's apple exports are Royal Gala, making it their most important variety. Braeburn is second, at 22% of exports by weight (Aitken and Hewett, 2013). We focus on Royal Gala in this paper.

2. Models in the literature

A number of predictive models exist in the horticultural literature. When validating a predictive model, it is important to test it on a holdout, and use appropriate error metrics. Table 1 summarises predictive models in the horticultural literature, showing that this is not always done. Also, while many models do a single hold-out, their validation would be more comprehensive if they used k-fold validation. The rationale is explained in Section 3. None of these models compare their results against a mean only model, so their