Informatics Institute of Technology
In Collaboration With

University of Westminster, UK

AutoMLR: Automated Machine Learning in R

A dissertation by
Mr. Thiloshon Nagarajah

Supervised by
Mr. Guhanathan Poravi

Submitted in partial fulfillment of the requirements for the
BEng (Hons) Software Engineering Degree Department of Computing

May 2019

Abstract

Machine learning provides machines ability to learn about data and perform tasks without explicitly programming. It is one of the most important component in artificial intelligence and helps to effectively solve many complex real world problems. Due to the advanced theories and prior knowledge required to even begin machine learning, only experienced programmers are able to develop machine learning systems. If domain experts who are not familiar with programming are given opportunities to develop these systems, it would pave way for technological revolutions in their respective domains.

AutoML systems are built with the aim of automating machine learning and abstracting the complexities involved in developing machine learning models. This dissertation is a result of the project to build an autoML system. The developed system, *AutoMLR,* works by making use of the knowledge of past trainings and several rules based components. The promising combinations and settings of algorithms, preprocessing, evaluations and hyperparameters are recommended by the system. Several such settings are converted to machine learning models, trained, evaluated and finally, the models with best performance are presented to the user.

AutoMLR provides a novel design and algorithm to be used in autoML systems. The decision making mechanisms used throughout the system and the model selection algorithms are novel results produced by this dissertation. Compared to the existing systems, it performs equally or better in small and medium datasets and performs slightly poor in extremely complex datasets. This system is first of its kind in R language and no other existing systems offer as much functionality and automation as *AutoMLR*. It works with both regression and classification tasks and supports around 30 different algorithms. It is available as an R package and interactions with user is facilitated through web graphical interface that eliminates the need for the user to learn any new languages.

Keywords: AutoML, Machine learning, Meta learning, Hyperparameter, Algorithm selection, R.

Thiloshon Nagarajah – 2015298