CompareML – A Web Application for Comparing and Evaluating Machine Learning Algorithms on Datasets

Emmanuel Osimosu

September 2016

Dissertation submitted in partial fulfillment for the degree of

Master of Science in Big Data

Computing Science and Mathematics

University of Stirling

Abstract

The hardest part of solving a machine learning task is often finding the best algorithm for the job. Today, many hundred machine learning algorithms exists with different strengths and weakness, and more are being invented every year. There are different ways to approach machine learning projects. Typically, choosing a machine learning algorithm for a problem relies heavily on human experts. In many cases, the standard procedure is to select a few familiar algorithms which are then implemented, and applied to the data. This is inefficient, and it's too easy to make the wrong choices upfront. CompareML provides a way to quickly and efficiently evaluate several algorithms on a dataset for various measures such as training time, accuracy etc., through a web interface accessible from any modern web browser running on any device, such as a computer, a tablet or a smartphone. There is also a collaborative element as many users can evaluate algorithms on a dataset that has been made public. The dataset would keep track of the most accurate algorithm/values and by which user. This portfolio approach to algorithm selection is more agile, less risky and allows for adapting an algorithm to the real needs of the project.

CompareML is built on top of Weka, a comprehensive suite of Java class libraries that implements many cutting-edge machine learning algorithms. The technology stack consist of Java on the server side, using Spring Boot web framework and a responsive Web front-end on the client side, using Bootstrap and AngularJS. The result is a modern web application where users can upload datasets to evaluate on, choose algorithms to evaluate, choose measures to evaluate for, and perform evaluations. Performance metrics are reported for the evaluations and the algorithms can be ranked according to different measures.