Accepted Manuscript

Scaling machine learning for target prediction in drug discovery using Apache Spark

Dries Harnie, Mathijs Saey, Alexander E. Vapirev, Jörg Kurt Wegner, Andrey Gedich, Marvin Steijaert, Hugo Ceulemans, Roel Wuyts, Wolfgang De Meuter

PII: S0167-739X(16)30111-X
DOI: http://dx.doi.org/10.1016/j.future.2016.04.023
Reference: FUTURE 3025

To appear in: *Future Generation Computer Systems*

Received date: 31 July 2015
Revised date: 27 April 2016
Accepted date: 30 April 2016

Please cite this article as: D. Harnie, M. Saey, A.E. Vapirev, J.K. Wegner, A. Gedich, M. Steijaert, H. Ceulemans, R. Wuyts, W. De Meuter, Scaling machine learning for target prediction in drug discovery using Apache Spark, *Future Generation Computer Systems* (2016), http://dx.doi.org/10.1016/j.future.2016.04.023

***Highlights (for review)**

We have used Spark to automatically distribute C++ predictors over a cluster.
Our Spark application allows near-linear speedup and optimal cluster utilization.
The core of the algorithm is easily changed to allow for experimentation.