

---

# Predicting Stellar Rotation Periods Using XGBoost

Nuno Gomes<sup>\*1,2</sup>, Fabio Del Sordo<sup>\*3,4,5</sup>, and Luís Torgo<sup>1,2,6</sup>

<sup>1</sup>Dalhousie University [Halifax] – Canada

<sup>2</sup>Faculdade de Ciências da Universidade do Porto – Portugal

<sup>3</sup>Institute of Space Sciences [Barcelona] – Spain

<sup>4</sup>Institut d'Estudis Espacials de Catalunya – Spain

<sup>5</sup>INAF - Osservatorio Astrofisico di Catania – Italy

<sup>6</sup>Laboratory of Artificial Intelligence and Decision Support – Portugal

## Abstract

The estimation of rotation periods of stars is a key problem in stellar astrophysics. Given the large amount of data available from ground-based and space telescopes, there is nowadays a growing concern to find reliable methods that allow one to quickly and automatically estimate stellar rotation periods accurately and with precision. The work we present is aimed at developing a computationally inexpensive approach to automatically predict thousands of stellar rotation periods. We focused on building a robust supervised machine learning model to predict surface stellar rotation periods from structured data sets, built from the Kepler catalogue of K and M stars. We analysed the set of independent variables extracted from Kepler light curves, investigating the relationships between them and the response. With these variables in hand, we trained regression extreme gradient boosting models that contain different numbers of predictors, and assessed their prediction performance, resorting to several metrics. We obtained models with quality comparable to the state of the art, and we were able to select a minimal set of variables with which we built extreme gradient boosting models without significant loss of performance. We obtained a goodness of fit, as measured by the adjusted coefficient of determination, of about 97%, for a model built with data sets containing stars with rotation periods between 7 and 45 days. Based on the results obtained with this study, we conclude that the best models built with the proposed methodology are competitive with the state of the art approaches, having the advantage of being computational cheap, easy to train, and relying on small sets of predictors.

---

\*Speaker