

S²C² Statistical Software Creation Circle

SOFTWARE PACKAGE CREATION

Proposal

Principal Scientist: Ernest Fokoue

Date: January 30th, 2021

Title: An R Package for Comprehensive Nearest Neighbors Learning

Abstract/Brief Description:

The nearest neighbors learning paradigm is one of the oldest paradigms in statistical machine learning, and forms the basis of nonparametric statistical machine learning. A number of packages already exist that implement the nearest neighbors learning paradigm, but to the best of my knowledge, none is as comprehensive as the one we intend to build. Among other things, this new package will explore a large number of distances and kernels, a vast number of weighting schemes along with state of the art reporting, plotting, table creations. In at least one of its versions, this package will be made as user-friendly as possible, including the possibility of a corresponding shiny app for educational purpose.

Package name: NearestNeighbors

Language used: R

Existing packages already in existence containing similar functions: `library(class)`; `library(caret)`; `library(kknn)`; `library(tsfknn)`;

Estimated number of collaborators: Minimum 2. Maximum 4.

Estimated time to completion of first version: June 7, 2021

Estimated start date: February 20, 2021

Target users: Statistical Machine Learning and Data Science End users, Professional Practitioners, Researchers, Students and Educators

Collaboration prerequisites:

- Solid demonstrated R programming proficiency
- Statistical Machine Learning Intro or intermediate knowledge
- Passion for coding in R

Collaborators and co-authors:

1. Fokoué, Ernest (Project Leader)
- 2.
- 3.
- 4.
- 5.

Collaborators will be accepted based on credentials, skills, order of expression of interest and commitment

References:

- Max Kuhn (2020). caret: Classification and Regression Training. R package version 6.0-86. <https://CRAN.R-project.org/package=caret>
- Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- Klaus Schliep and Klaus Hechenbichler (2016). kkn: Weighted k-Nearest Neighbors. R package version 1.3.1. <https://CRAN.R-project.org/package=kkn>
- Martinez F, Frias MP, Charte F, Rivera AJ (2019). "Time Series Forecasting with KNN in R: the tsfknn Package." *The R Journal*, 11(2), 229-242.

Contact:

- Ernest Fokoué
ernest.fokoue@gmail.com or
epfeqa@rit.edu