

# Integrating R machine learning algorithms in Stata using rcall 3.0: a tutorial for Stata users and developers

E. F. Haghish

September 18, 2021

# Contents

<b>I</b>	<b>machinelearning package commands</b>	<b>3</b>
<b>1</b>	<b>missforest</b>	<b>4</b>
1.1	Syntax . . . . .	4
<b>2</b>	<b>Description</b>	<b>4</b>
2.1	Example . . . . .	4
2.2	Author . . . . .	4

Part I

machinelearning package commands

Version: 0.0.2 cite: Haghish, E. F. (2021). rcall for all: machine-learning tutorials for Stata users and developers

## 1 missforest

imputing missing data with Random Forest. this command utilizes the **missForest** package from **R** and embeds it in a **Stata** program using **rcall** package (Haghish, 2019). The **missforest** is a powerful command for **EASILY** imputing missing data.

more importantly, the command is also meant to be a tutorial for Stata developers, showing how to embed **R** into **Stata** and how to document **Stata** packages with *Markdown* language, using **markdoc** package. visit the project on github and have a look at the source code!

**Join the Resistance!** Fork this repository on GitHub and contribute to its development or documentation.

### 1.1 Syntax

The Syntax is under development!

## 2 Description

**missforest** is a free and open-source R package for missing data imputation with Random Forest. The current Stata program embeds this package in a **Stata** program using **rcall** package (See Haghish, E. F. (2019) Seamless interactive language interfacing between R and Stata).

The **missforest** Stata program not only meant to provide the capabilities of Random Forest missing data imputation for Stata users, but also meant to be an example programs for developing Stata packages with **R** and **rcall**. In addition, **the documentation of the package is written with markdoc package** (See Haghish, E. F. (2019) Software documentation with markdoc 5.0) using the minimalistic *Markdown* language, which also can serve as a tutorial for the community of Stata developers how to save time on software documentation and make it friendlies and easier!

### 2.1 Example

Here is an example of doing missing data imputation with the *variablewise* option, which provides the OOB error estimation for each imputed variable:

```
. webuse mheart5
. missforest, variablewise
. return list
. matrix list r(oob)
```

### 2.2 Author

**E. F. Haghish**

Department of Psychology

University of Oslo

haghish@uio.no

machinelearning homepage  
Package Updates on Twitter

---

This help file was dynamically produced by MarkDoc Literate Programming package