# Introduction to Statistical Computing in Scala - an Implementation of the K-Nearest Neighbors classifier

Roxana Tesileanu, *Research Assistant, National Institute of Forest Research and Management (INCDS), Romania*

***Abstract*—Statistical computing in ecology evolves at a high speed, mainly because researchers have recognized the advantage of being able to design their algorithms according to their needs. The present paper introduces the implementation in Scala of the k-Nearest Neighbors (kNN) classifier based on Euclidean distances, which can be applied also on small datasets, a situation commonly encountered in ecological research.**

***Index Terms*—machine learning in ecology, k-Nearest Neighbors classification, Scala**

## I. Introduction

One of the drivers of the machine-learning progress is the great amount of data born within and gathered by networked and mobile computing systems which necessitates further processing in order to gain insights into the specific fields from which it originates (Jordan and Mitchell 2015, p. 256). The "Big Data phenomenon" is real but, unfortunately, it can not be generalized to all research areas, especially some areas of ecological research which investigate systems which are by nature data-poor and will probably remain as such unless cost-intensive data collection projects are being proposed and financed. This by no means implies that such areas cannot take advantage of the progress of machine-learning and take the best out of the existing datasets.

## II. Conclusion

## Appendix A
## Source file 1

.....

## Appendix B
## Source file 2

.....

## Acknowledgment

The author would like to thank Jeff Druce and Mike Reposa for very useful learning tips.

**Roxana Tesileanu** My bio.