



**Author:** António Aristides Romualdo Carvalho  
**Thesis Advisor:** Manuel Fernandez Veiga  
**PhD Program:** Information and Communications Technology of the University of Vigo

## Motivation Of The Work

Interactive learning systems (ILS) are a set of mathematical procedures for selective sampling and labelling of points, mixing classification by a machine and humans, and as a result classifiers that work in less time and take less effort to implement. As a consequence, these mixed classifiers, which are not completely automated, show ability for adapting to real-time or changing situations, especially in social fields (e.g. smart cities) or behavioral changes in markets.

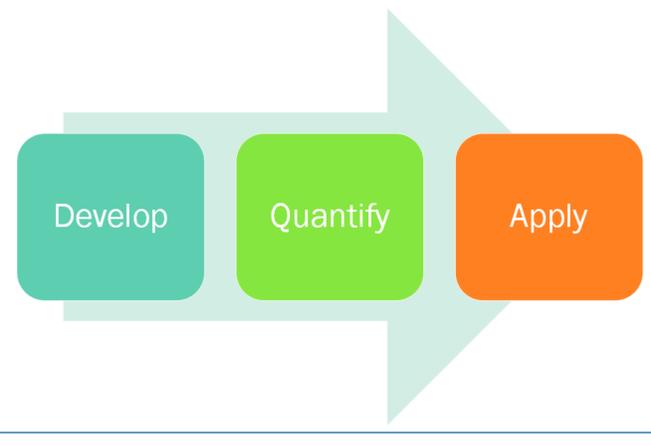
However, due to the biased and limited sampling procedure, interactive statistics learning techniques have to be carefully engineered so that statistical accuracy is preserved.

The motivation is to be able to design efficient interaction statistical algorithms between humans and machines, and quantifying how much these interaction helps in reducing the computational complexity of classical statistical learning applications.



## Thesis Objectives

- Develop new interactive statistical learning algorithms tailored to the retail markets, based on consumer segments (number of children, gross salary, etc), item locations, suppliers, etc (for example to advise products/items based on the number of children of a consumer).
- Quantify the computational complexity of those algorithms and statistical accuracy, for knowing what kind of machine should be used, as well as trying to optimise the algorithms for the computational complexity to be as lower as possible. On the theoretical side, perform an introductory analysis of the interaction step by using information-theoretical tools.
- Apply the techniques to real datasets and implement an algorithmically efficient Interactive Learning System (ILS), so that all the work made in the below objectives can be verified, with real datasets. Datasets will be collected from actual customer service branches, and also from open data repositories.



## Research Plan

- Study retail markets.
- Study interactive learning systems.
- Study R programming language.
- Analyze datasets related with retail markets.



## Results and Discussions

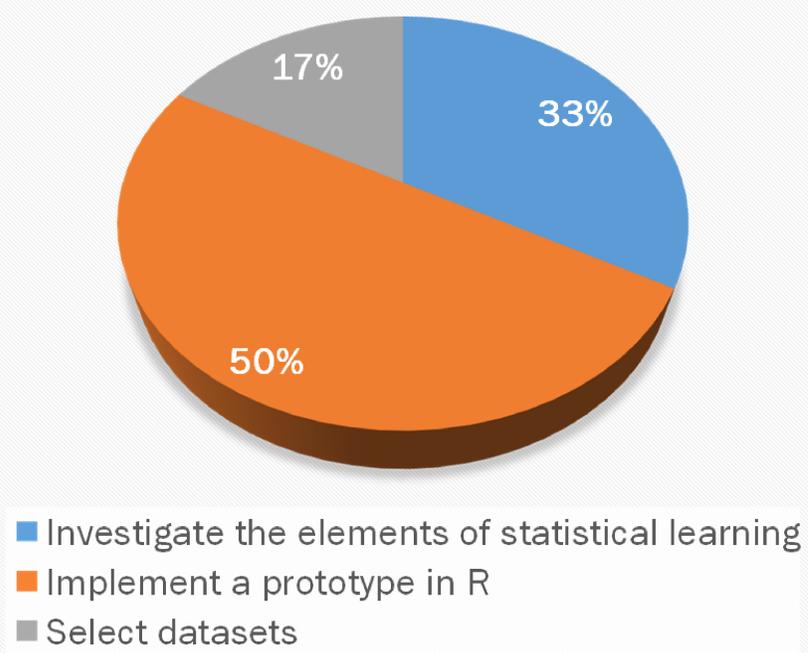
The results and discussions (for now) are in a preliminary phase, to delimit the Programming Language (in this case, it will be R), as well as, the dataset(s) used as a sample for the techniques application.

```
R Console
> source("C:\\under_construction.R")
> under_construction()
[1] "Under Construction!"
>
under_construction.R
under_construction <- function() {
  sprintf("Under Construction!");
}
```

## Next Year Planning

- Investigate the elements of statistical learning (Data Mining, Inference, and Prediction).
- Implement a prototype in R programming language.
- Select datasets that can be used in the prototype tests.

Next Year Planning - Time Distribution



## References

[1] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, Stanford, California: Springer, 2008.

[2] M. Scherer, "00738578.pdf," 1998. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=738578>. [Accessed 01 06 2017].

[3] R. Porter, J. Theiler and D. Hush, "06560028.pdf," Los Alamos National Lab, 2013. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6560028>. [Accessed 02 06 2017].

[4] M. Serasinghe and S. Vasanthapriyan, "07924840.pdf," Sabaragamuwa University of Sri Lanka, 2016. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7924840>. [Accessed 20 05 2017].