# A predictive modeling based approach for treatment assistance and diagnosis prediction

**Adriana Mihaela Coroiu**

Faculty of Mathematics and Computer Science, Babeş Bolyai University

adrianac@cs.ubbcluj.ro

**PhD Research Report.** The amount of available data increases in an exponential way by the day. Being able to make sense of all the data will represent an essential skill for people working in a data science industry. Moreover, the type of these data is an important issue [**?**]. Therefore the extraction of relevant information, the discovery of relations between data and the ability to generalize to new data represent a continuous challenge.

Data analysis is an impressive area of concern for domains such as education, healthcare, economics, biology, history or agriculture [**?**, **?**]. Particularly, the purpose of this work is related to medicine and psychology. Machine learning advantages are being investigated in order to improve a treatment, a diagnosis of a patient.

This report, presenting a work in progress, discusses an approach to a relevant supervised learning method: classification. Various aspects are considered, as preprocessing of the input data; selection of the model applied to the data; evaluation of the model; improving the performance of a model, selection of the most relevant features to be included in the model and the generalization to new data [**?**, **?**]. The computed metrics for performance evaluation of a model are also highlighted.

The model evaluation, the model improvements and feature selection ultimately lead to building models able to generalize to new data with a high value of accuracy. All these represent an added value in fields such us medicine and psychology, where a physician or a psychologist may use pattern and information as input in the treatment of a patient.

## References

[1] Bradley A. P., *The use of the area under the ROC curve in the evaluation of machine learning algorithms*, Pattern recognition 30.7 1145-1159, 1997

[2] Coroiu A. M., Găceanu R. D., Pop H. F., *Ordinal data analysis using agglomerative clustering algorithms*, In Knowledge Engineering Principles and Techniques, Book of Abstracts (Cluj-Napoca, Romania, July 2 2015), M. Frentiu, H. F. Pop, and S. Motogna, Eds., Babes-Bolyai University, (2015), 71-74.

[3] Dursun D., Walker G., Kadam A., *Predicting breast cancer survivability: a comparison of three data mining methods*, Artificial intelligence in medicine 34.2 113-127, 2005

[4] Gunawardana A., Shani G., *A survey of accuracy evaluation metrics of recommendation tasks*, The Journal of Machine Learning Research, 10, 2009.

[5] Witten I. H., Eibe F., *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2005.