EINDHOVEN UNIVERSITY OF TECHNOLOGY

The nutcracker framework for ensemble interpretability

Citation for published version (APA):
Zeev Ben Mordehay, O., Duivesteijn, W., & Pechenizkiy, M. (2017). *The nutcracker framework for ensemble interpretability*. Poster session presented at 16th International Symposium on Intelligent Data Analysis (IDA 2017), October 26-28, 2017, London, UK, London, United Kingdom.

Document status and date:
Published: 01/10/2017

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 15. Jun. 2020

Technische Universiteit Eindhoven University of Technology

Where innovation starts

IDA 2017 - London

The *Nutcracker* Framework for Ensemble Interpretability

Oren Zeev-Ben-Mordehai
o.zeev.ben.mordehay@tue.nl

Advisors: **Wouter Duivesteijn & Mykola Pechenizkiy**
Project: KnowYourCustomer
Predictive Analytics (Rabobank)

Introduction The basic principles behind ensembles (e.g. Random Forest [1], AdaBoost [2]) are simple. But we're still in trouble when attempting to explain the logic taken. Where does the problem lie? The reason that ensembles are effective is that the base estimators "work together" and compensate each for the others' shortcomings.

The *Nutcracker* Framework Given a trained ensemble and the relevant training / test dataset, construct prediction matrix, M, cases (rows) against predictions (columns). Bi-cluster M to a given number of R x C bi-clusters [3].
Now, investigate performance per bi-cluster (R x C). Identify feature importance per base estimators group (C). Describe each of the R cases subgroups in terms of features and values. We use Exceptional Model Mining [4] for that task.

Performance of the ensemble against the dataset compared to performance of base estimator groups against subgroups of cases, adds transparency.

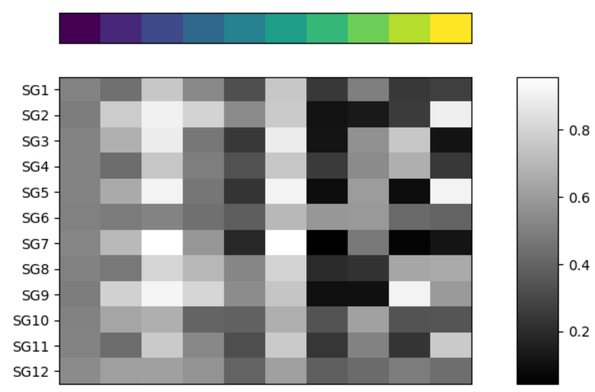

Figure 1 R x C bi-clusters. Each bi-cluster reports its mean accuracy (brighter is better) .

https://doodlemeister.files.wordpress.com/2009/02/nuts1.jpg

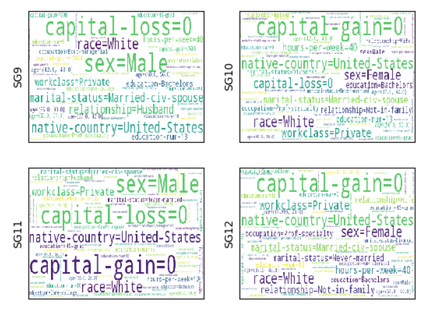

Figure 2 An attempt to name the subgroups using word-clouds

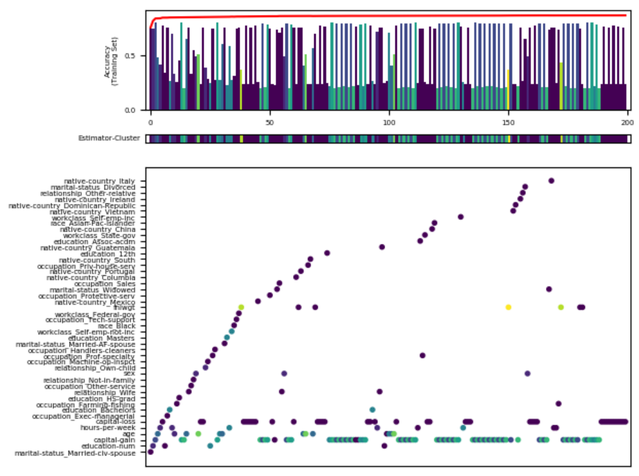

Figure 3 AdaBoost ensemble trained on the Adults Dataset (200 decision stumps) . Stage-wise, new features are explored, and then old ones are revisited.

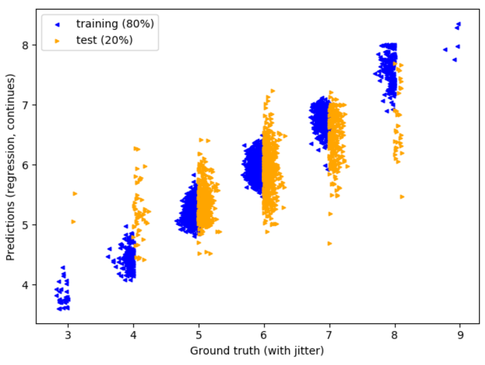

Figure 4 Confusion for RandomForest trained on the Wine dataset.

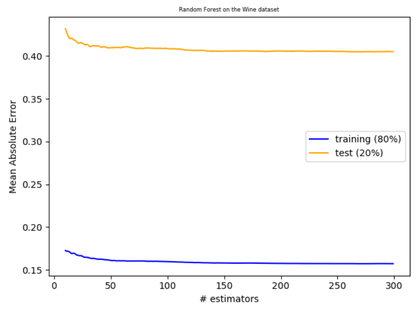

Figure 5 Additional decision trees fail to bring test performance close to training one.

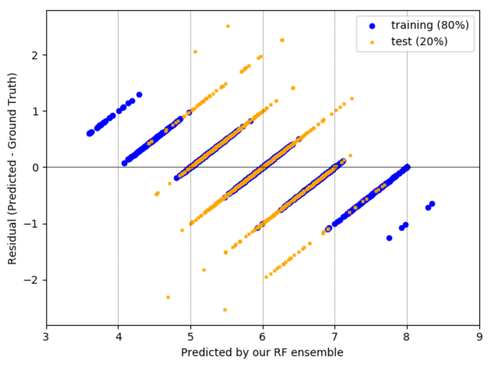

Figure 6 Residuals

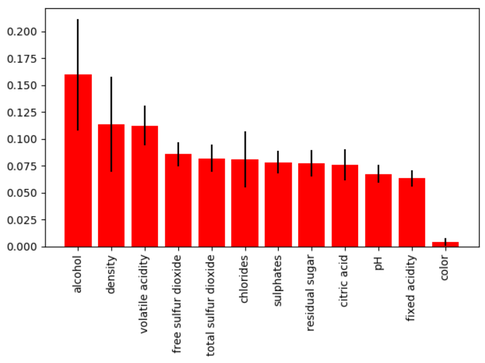

Figure 7 Feature Importance

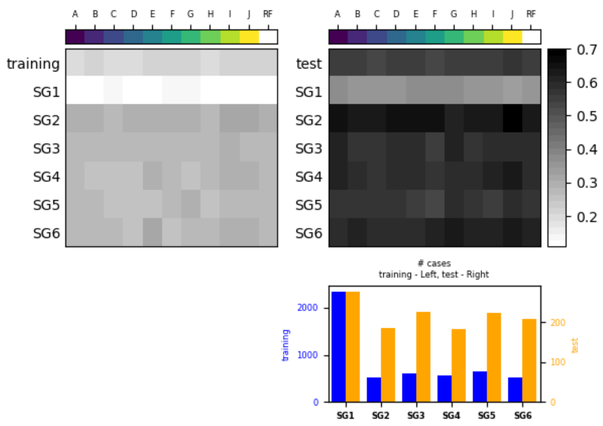

Figure 8 Mean Absolute Error (MAE) per bi-cluster. Brighter is better.

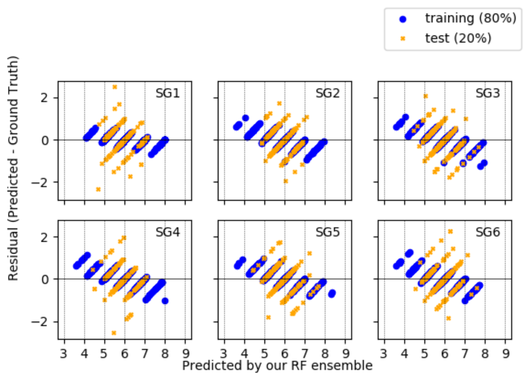

Figure 9 Residuals by subgroup

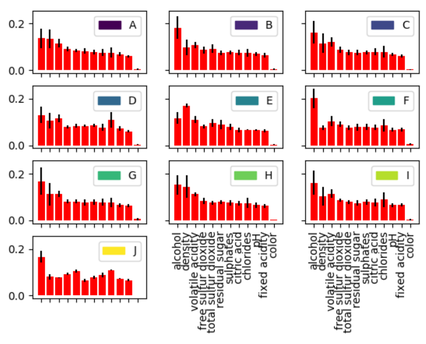

Figure 10 Per base estimator group feature importance.

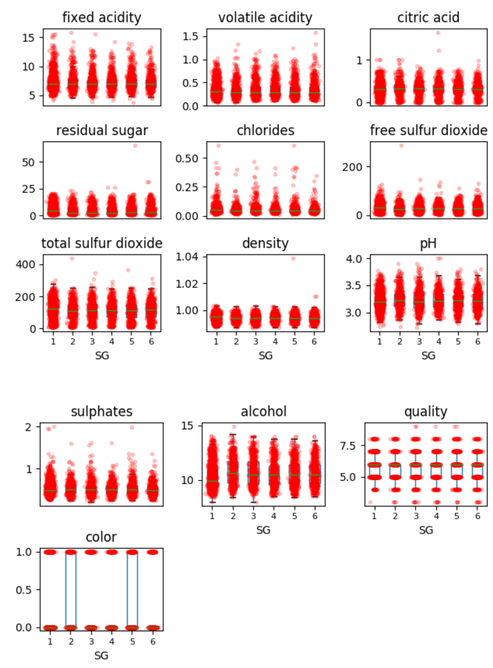

Figure 11 Values distribution per subgroup

Table 1 Descriptions for the subgroups

SG1	alcohol > 9.97 and residual sugar <= 19.23 and total sulfur dioxide <= 192.0
SG2	alcohol <= 9.97 and chlorides <= 0.18 and quality != 4
SG3	alcohol > 9.97 and alcohol <= 12.93 and sulphates <= 0.73
SG4	alcohol > 9.97 and total sulfur dioxide <= 192.0 and fixed acidity <= 10.71
SG5	quality != 6 and density <= 0.99 and color != 1
SG6	alcohol > 9.97 and volatile acidity <= 0.72 and pH > 3.09

http://www.colonialsoldier.com/136/279/

References
[1] Tin Kam Ho. (1995). Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition, 1, 278–282. https://doi.org/10.1109/ICDAR.1995.598994
[2] Y. Freund, R.E. Schapire (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1):119–139.
[3] Kluger, Y., Basri, R., Chang, J. T., & Gerstein, M. (2003). Spectral biclustering of microarray data: Coclustering genes and conditions. Genome Research. https://doi.org/10.1101/gr.648603
[4] D. Leman, A. Feelders, A.J. Knobbe (2008). Exceptional Model Mining. Proc. ECML/PKDD (2), pp. 1–16.

Rabobank