Instance-level explanations for fraud detection (poster)

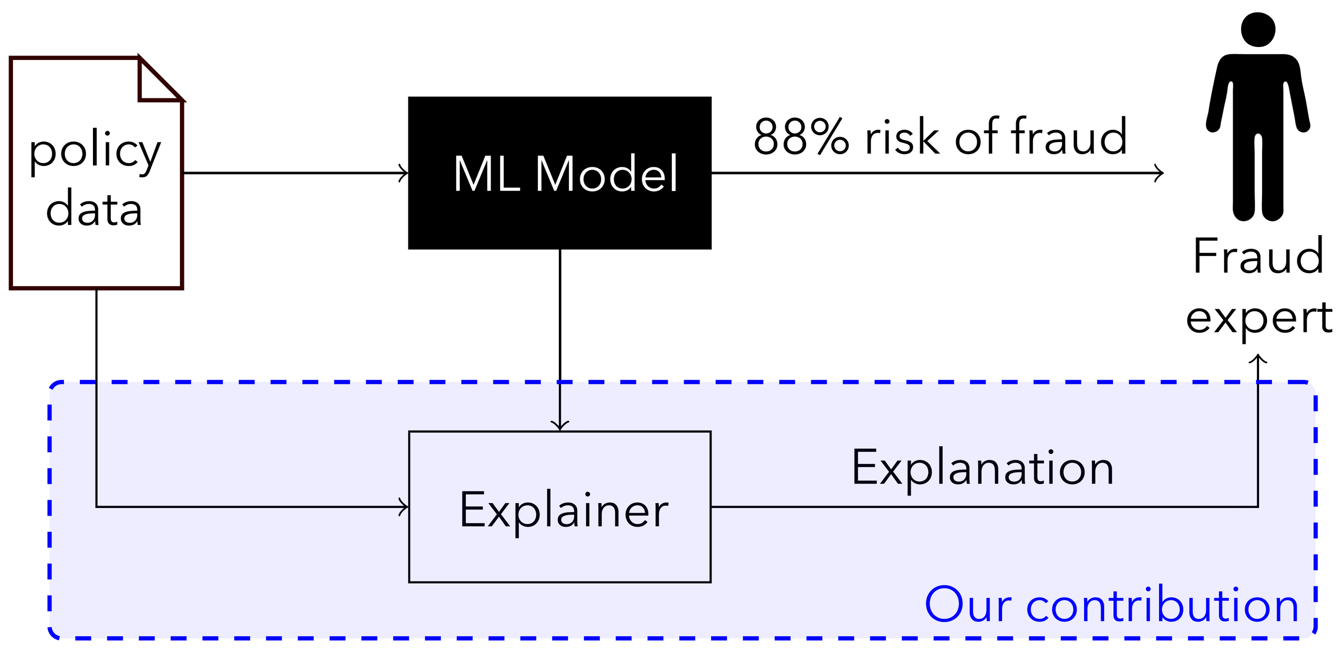
INSTANCE-LEVEL EXPLANATIONS FOR FRAUD DETECTION

Dennis Collaris, Leo M. Vink, Jarke J. van Wijk

TU/e achmea NWO

THE PROBLEM

Fraud detection is a difficult problem that can benefit from predictive modeling. However, the verification of a prediction is challenging; for a single insurance policy, the model only provides a prediction score.



THE SOLUTION

We designed two novel dashboards combining various state-of-the-art explanation techniques.

Feature contribution

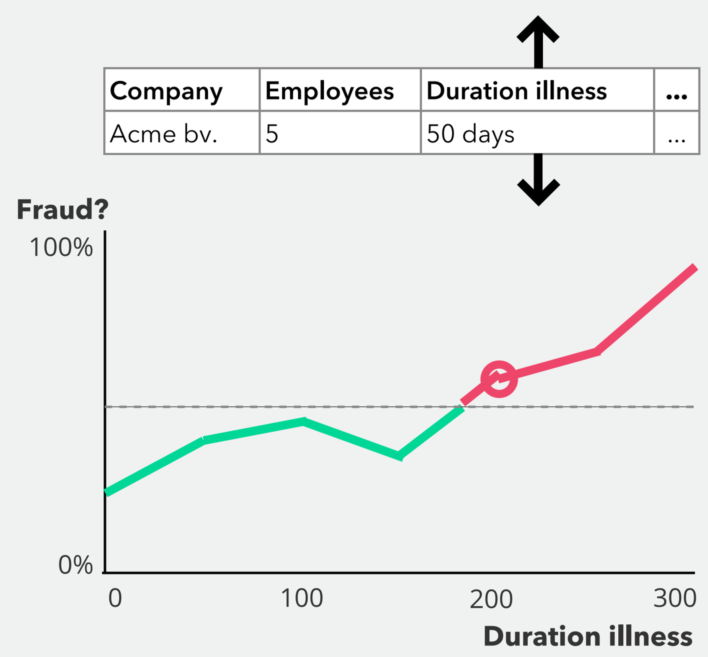
Local increment of feature f:

$$LI_f^c = \begin{cases} Y^c_{mean} - Y^p_{mean}, & \text{Parent splits on feature } f. \\ 0 & \text{Otherwise.} \end{cases}$$

Contribution of feature f in decision rule R:

$$FC^f_{i,t} = \sum_{N \in R_{i,t}} LI^N_f$$

Partial dependence



Local rule extraction

Synthetic pruning data set, uniform samples from an n-ball:

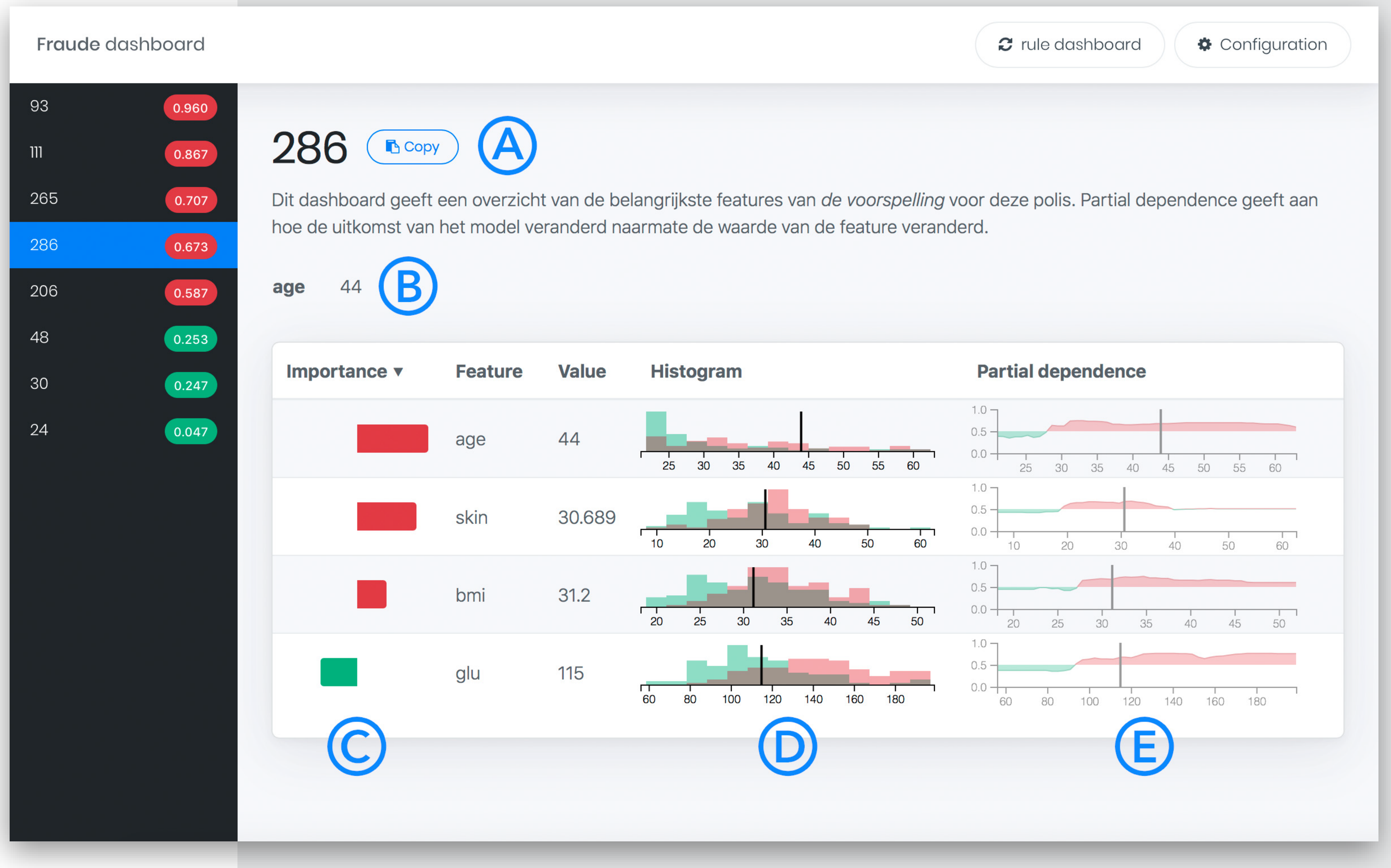
$$n\text{-ball uniform distribution} = \frac{Y * U^{\frac{1}{n}}}{\|Y\|}$$

with $Y \sim N(0,1)$ and $U \sim U(0,1)$

All decision rules applicable to instance i are extracted and pruned.

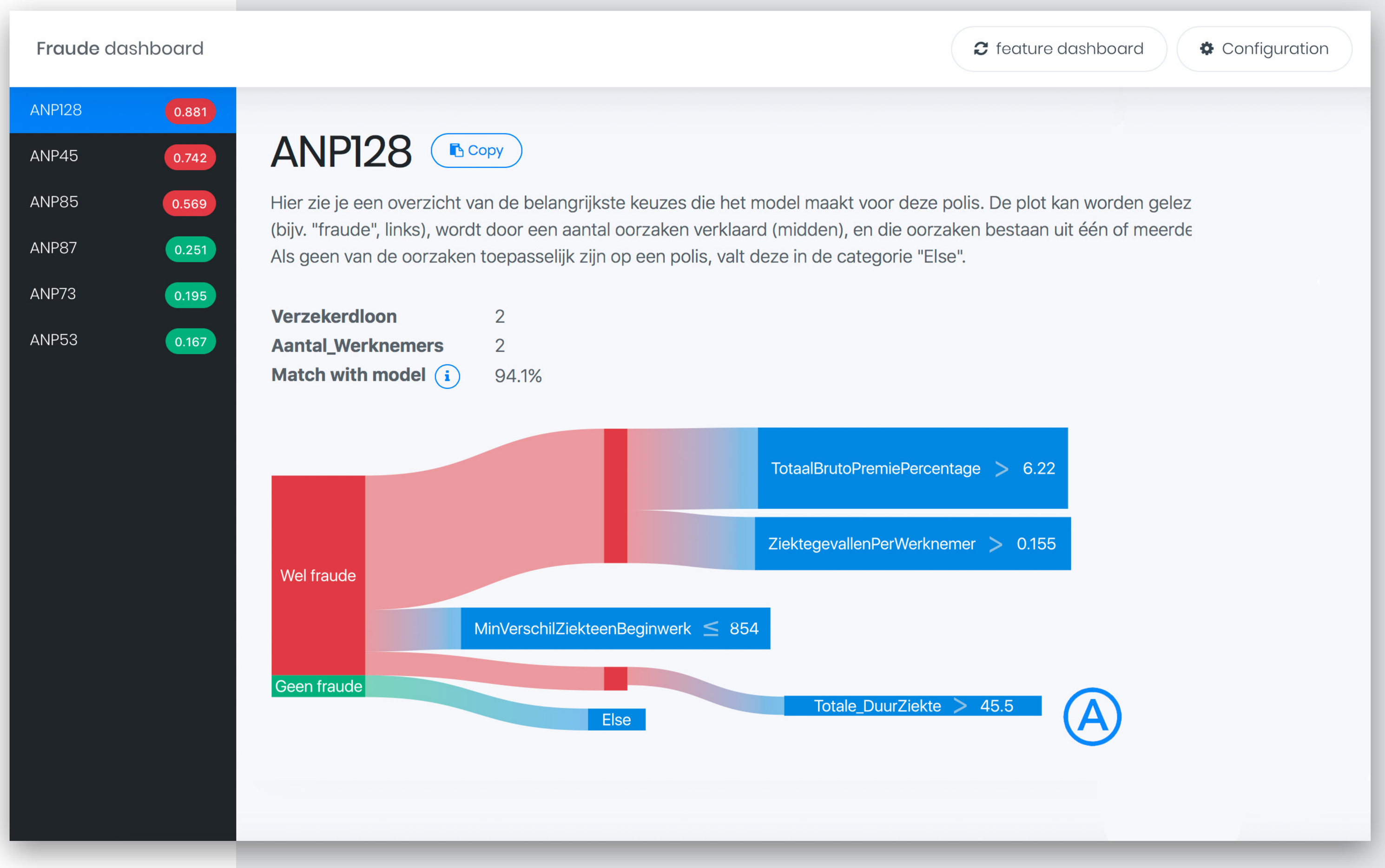
A Regularized Random Forest is trained on binary matrix of applicability of rules on the pruning dataset. Feature importance of that forest constitutes a metric of importance of individual decision rules.

For the example on the right, 1.300.000 rules are reduced to only 4, while still retaining 94.1% of the local fidelity of the reference model.



FEATURE DASHBOARD

This dashboard shows bar charts (A) expressing the contribution of a feature to the prediction. Additionally, partial dependence plots (B) show the impact of changing the feature value indicated with a vertical line) on the prediction.



RULE DASHBOARD

This dashboard shows a flow diagram (A) representation of locally extracted rules. Four rules are shown, and rule and feature importance is encoded by the width of the edges.

Check out the paper:
arxiv.org/abs/1806.07129

Dennis Collaris
d.a.c.collaris@tue.nl

Commit2Data