Machine Learning in Economics
(458657)

term paper

Early Warning System of Fiscal Stress

**comparing the performance of traditional logistic regression versus
a random forest algorithm**

**Department of Economics
University of Bern**

Spring Semester 2022

submitted by Bela Koch

Contents

1 Introduction

test ob zitierung funktioniert Jarmulska (2020).

2 Literature Review

3 Model Describtion

3.1 Performance Metrics

3.2 Logit Model

Hastie et al. (2009)

$$\hat{\beta}^{lasso} = \underset{\beta}{\arg\min} \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \quad \text{subject to} \quad \sum_{j=1}^{p}|\beta_j| \leq t \tag{1}$$

Lagrangian form

$$\hat{\beta}^{lasso} = \underset{\beta}{\arg\min}\left\{\frac{1}{2}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{2}$$

3.3 Random Forest

Gini index

$$g(w) = \sum_{k \neq j} p_{wk}p_{wj} = \sum_{k} p_{wk}(1 - p_{wk}) \tag{3}$$

4 Data Describtion

4.1 Dependent Variable

4.2 Explanatory Variables

5 Empirical results

5.1 Performance

5.2 Interpretability

5.2.1 Variable Importance

5.2.2 Shapley Values

5.2.3 Partial dependence plots

5.2.4 Accumulated local effects plots

6 Conclusion

7   References

Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Vol. 2. Springer.

Jarmulska, Barbara. 2020. "Random Forest Versus Logit Models: Which Offers Better Early Warning of Fiscal Stress?" *ECB Working Paper Series No 2408 / May 2020.*