



# Protocol (20.01.2026) – Credit Risk Modeling ILAB OeNB

By Anastasia Pryshchepa, Leonid Kuzmishin,  
Tristan Leiter, Martin Yago Tortajada

## Summary

### 1) Univariate Analysis Results *(Slide 5)*

- Check the results for sector (the binary encoded variables). How is the AUC-Score computed here?

### 2) Data Splitting *(slide 7)*

- Rename „Default-Rate“ to „Proportion of Defaults“

### 3) Quantile Transformation *(slide 8)*

- Gaussian Assumption does not matter – has no impact on the feature engineering nor modelling process.
- We only require normality in the response variable, not for the inputs.
- Remove the current motivation and methodology boxes.

### 4) Regularized GLMs | Calibration plots *(slide 14, and others.)*

- Show the calibration plots (average predicted probability of default vs observed, split in deciles) also for the different sectors and categories.
- Mr. Leitner will forward the OeNB framework for showing the discrete folds (instead of the current implementation via deciles).

### 5) NNs *(slide 23)*

- Show the calibration plot instead of the ROC curve.

## Summary

### 6) Outlook (slide 29)

- Add visual representations – Prob. of default for the base learners to understand how shrinkage actually looks like.
- Include more information about the models: feature importance, and others.
- Include partial dependence plots (showing bivariate interactions and the marginal responses one feature at a time); Scatterplots can also be used to show how similar the predictions are. Hexagonal binning.
- Explain what XGBoost „sees“ in contrast to the other models. Find specific exemplary balance sheets to find out what XGBoost catches.
- Try out the „hdo“ package for variational autoencoders (feature compression on the raw positions).
- A previous group achieved a strong result using the hdo package together with GLM modeling.
- Try out GLM with untransformed vs transformed features.
- Look at two balance sheet items and see if the marginal response is the same across models.
- Different gradient boosting methods (CatBoost, LightGBM, etc.) would also be interesting.

### 7) General remarks

- Chart labeling needs to be corrected. Axis elements are not clearly visible when presented via a beamer.