IT-BASED MANAGEMENT PROJECT ASSIGNMENT

INTERNIM PRESENTATION

JORDAN TOBIAS 11902127
TRAILOVIC LUKA 11918523
ZIB SEBASTIAN 11907072
SPILYKA ANTAL 11907555

CONTENTS

1. selected binning approach
2. data train/test splitting
3. predictor variables
4. selected generalized linear model for the train sample
5. Gini coefficient for the predicted scores in the test sample
6. Gini coefficient comparison: train vs. test sample

SELECTED BINNING APPROACH

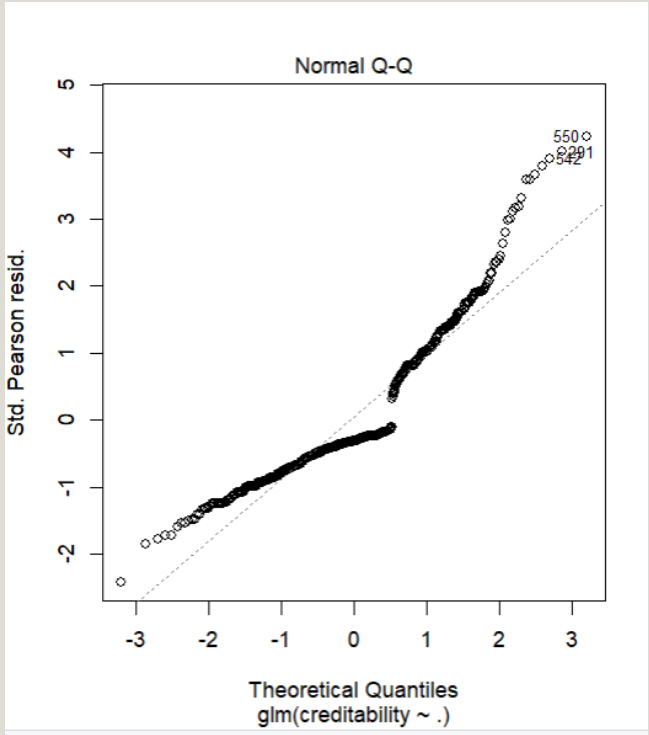
- WOE-Binning
  - Grouping numerical data into bins
  - Calculating the WOE (weight of evidence) for each bin
  - WOE measures relationship between input variable and binary outcome (good/bad)
- GRP-Binning
  - Grouping numerical data into bins
  - Analyze each bin with statistic (mean/median/etc.)
  - GRP identify and create bins that show meaningful difference in the input variables

DATA TRAIN/TEST SPLITTING

- Given in assignment:
  - 75% of data for training
  - 25% of data for testing

SELECTED GENERALIZED LINEAR MODEL FOR THE TRAIN SAMPLE

- We are interested in relation between input and output
- Woe delivers most compact regression models
- => Woe-based logistic regression



PREDICTOR VARIABLES

- Selection of ”informative” variables:
  - Status of existing checking account
  - Duration in month
  - Credit history
  - Age in years
  - Savings account and bonds

variable <chr>	info_value <dbl>
status.of.existing.checking.account	0.6660115
duration.in.month	0.3345035
credit.history	0.2932335
age.in.years	0.2596514
savings.account.and.bonds	0.1960096

5 rows

GINI COEFFICIENT FOR THE PREDICTED SCORES IN THE TEST SAMPLE

- WOE-Binning

```
perf_eva(pred = predProb.list$test,
         label = data_woe.list$test$creditability,
         title = 'test',
         show_plot=c("roc","ks"),
         confusion_matrix = TRUE)
```

```
## $binomial_metric
## $binomial_metric$test
##          MSE      RMSE   LogLoss         R2        KS       AUC      Gini
## 1: 0.1760701 0.4196071 0.5286325 0.1678927 0.4379835 0.7539315 0.507863
```

GINI COEFFICIENT COMPARISON: TRAIN VS. TEST SAMPLE

- Training-Set

```
perf_eva(pred = predProb.list$train,
         label = data_woe.list$train$creditability,
         title = 'train',
         show_plot=c("roc","ks"),
         confusion_matrix = TRUE)
```

```
## $binomial_metric
## $binomial_metric$train
##          MSE      RMSE   LogLoss        R2        KS       AUC      Gini
## 1: 0.1620372 0.4025385 0.4864994 0.2261554 0.4605765 0.7935077 0.5870155
```

- Testing-Set

```
perf_eva(pred = predProb.list$test,
         label = data_woe.list$test$creditability,
         title = 'test',
         show_plot=c("roc","ks"),
         confusion_matrix = TRUE)
```

```
## $binomial_metric
## $binomial_metric$test
##          MSE      RMSE   LogLoss        R2        KS       AUC      Gini
## 1: 0.1760701 0.4196071 0.5286325 0.1678927 0.4379835 0.7539315 0.507863
```