

## » Mortalidade das Empresas

Um modelo preditivo

«

Alexandre Cunha  
Rui Gouveia  
Sandra Lagarto  
Vasco Cordeiro  
Vera Dias



 21-04-2023

»

1

## Mortalidade das Empresas

- Como obter uma classificação mensal da morte das empresas?
- Uma empresa que pertença à população de empresas ativas do mês  $n$  e não pertença às populações de empresas ativas dos 24 meses imediatamente seguintes é considerada morta no mês  $n$ .

2



## Mortalidade das Empresas

População de empresas ativas – grupo de empresas que, em determinado mês, entregou informação numa das duas seguintes fontes administrativas:

- Autoridade Tributária - portal E-fatura
- Segurança Social - declarações mensais de remunerações (DMRSS)



3



## Mortalidade das Empresas

Restrições de âmbito

- Utilização de informação proveniente de fontes de dados administrativos
- Período: janeiro 2016 a fevereiro 2018
- Restrições de âmbito:
  - Setores Institucionais 11 e 12
  - Sociedades



4



## Mortalidade das Empresas

### Restrições ao Volume de Dados

- Ajustamento do modelo de classificação
  - CAE Secção G – Comércio por grosso e a retalho, reparação de veículos automóveis e motociclos
- Previsão
  - CAE Secção F – Construção



5



## Mortalidade das Empresas

| Variáveis E Fatura             |  | Variáveis DMRSS                                 |  |
|--------------------------------|--|---|--|
| Número de identificação fiscal |  | Número de identificação fiscal                  |  |
| Código de atividade económica  |  | Ano   |  |
| Ano                            |  | Mês   |  |
| Mês                            |  | Valor das remunerações por natureza             |  |
| NUTS III                       |  | Número de trabalhadores remunerados por Empresa |  |
| Volume de Negócios             |  |   |  |

| Rácios Financeiros               |   |  |                            |
|----------------------------------|---|--|----------------------------|
| Rendibilidade do ativo           | Produtividade aparente do trabalho            | Peso dos gastos com o pessoal no VABpm | Rendibilidade das vendas   |
| Rendibilidade do capital próprio | Produtividade do trabalho ajustada ao salário | Taxa de margem bruta de exploração     | Rotação do capital próprio |
| Endividamento                    | Gastos com o pessoal per capita               | Rendimento operacional das vendas      | VVN per capita             |
| Autonomia financeira             | Taxa de valor acrescentado bruto              | Debt to equity ratio                   | Taxa de investimento       |
| Rotação do ativo                 | Peso do EBE no VABpm                          | Solvabilidade                          |                            |



6



# Mortalidade das Empresas

## Modelação

- Classification trees

```
tree_mod <- decision_tree(cost_complexity = tune(), min_n = 2, tree_depth = 30) %>% set_engine("rpart") %>% set_mode("classification")
```

- Logistic regression

```
glm_mod <- logistic_reg(penalty = tune(), mixture = tune()) %>% set_engine("glmnet") %>% set_mode("classification")
```

- K Nearest Neighbors

```
knn_mod <- nearest_neighbor(neighbors = tune(), weight_func = 'gaussian') %>% set_engine("kknn") %>% set_mode("classification")
```

- Random forests

```
forest_mod <- rand_forest(trees = tune()) %>% set_engine("ranger") %>% set_mode("classification")
```

- Gradient boosted machines

```
boost_mod <- boost_tree(trees = tune()) %>% set_engine("xgboost") %>% set_mode("classification")
```

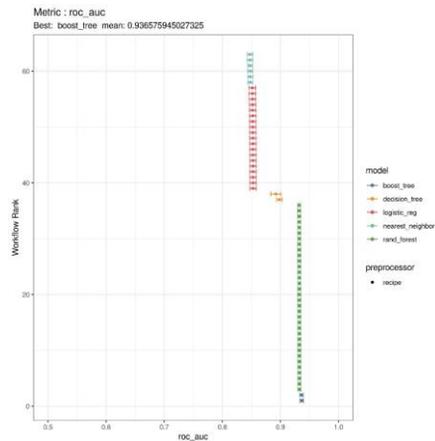
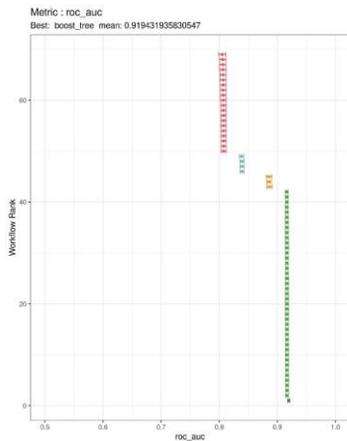


7



# Mortalidade das Empresas

## Modelo sem Rácios vs Modelo com Cinco Rácios



8



# Mortalidade das Empresas

## Gradient Boosted Machines Results

| Métrica     | 5 Rácios | 10 Rácios | Threshold 0.5 |  | Real  |              |                   |
|-------------|----------|-----------|---------------|--|-------|--------------|-------------------|
| Precision   | 2,2      | 1,7       |               |  | Morte | Vida         |                   |
| PR-AUC      | 10,6     | 11,0      |               |  | Morte | 353 (0.2 %)  | 19.613 (11.05 %)  |
| Specificity | 85,2     | 88,9      | Previsto      |  | Vida  | 39 (0.02 %)  | 157.459 (88.73 %) |
| Accuracy    | 85,3     | 88,9      |               |  |       |              |                   |
| Sensibility | 88,3     | 90,1      | Threshold 0.8 |  | Real  |              |                   |
| ROC-AUC     | 93,8     | 96,1      |               |  | Morte | Vida         |                   |
|             |          |           |               |  | Morte | 341 (0.19 %) | 13.827 (7.79 %)   |
|             |          |           | Previsto      |  | Vida  | 51 (0.03 %)  | 163.245 (91.99 %) |



9



# Mortalidade das Empresas

```

== Workflow ==
Preprocessor: Recipe
Model: boost_tree()

--- Preprocessor ---
5 Recipe Steps

• step_rm()
• step_normalize()
• step_dummy()
• step_corr()
• step_downsample()

--- Model ---
Boosted Tree Model Specification (classification)

Main Arguments:
trees = tune()
tree_depth = tune()
learn_rate = tune()

Computational engine: xgboost

Preprocessor1_Model41

trees: 1980
tree_depth: 12
learn_rate: 0.0263

VIP – Variáveis mais importantes
Número de Trabalhadores
Gastos com Pessoal per capita
Soma do Valor Total
Soma do Valor Total lag 1
VVN per capita
Produtividade aparente do trabalho
Rotação do Ativo
Média de Dias Trabalhados

```



10

## Mortalidade das Empresas

### Dificuldades

- Volume de dados vs capacidades computacionais
- Reduzida precisão dos modelos

### Desenvolvimentos Futuros

- Introdução de novas variáveis preditivas
- Análise exploratória após calibração de scores