CU53_MODEL_DEVELOPMENT_02_XGBOOST

June 13, 2023

#

CU53_impacto de las políticas de inversión en sanidad, infraestructuras y promoción turística en el SPI

1 IV. Model development

En este anexo se incluye el código utilizado durante el desarrollo de los modelos incluidos en el caso de uso.

1.1 Modelo RIDGE

[1]:
```
Sys.setlocale(category = "LC_ALL", locale = "es_ES.UTF-8")
```

'es_ES.UTF-8/es_ES.UTF-8/es_ES.UTF-8/C/es_ES.UTF-8/C'

1.1.1 Paquetes

[2]:
```
## Modelos CU 53
library(glmnet)
library(dplyr)
library(readr)
```

```
Loading required package: Matrix

Loaded glmnet 4.1-6


Attaching package: 'dplyr'


The following objects are masked from 'package:stats':

    filter, lag


The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

1.1.2 Datos

[3]:
```
inversiones <- read_csv("CU_53_05_05_inversiones_cm.csv")

spi <- read_csv("CU_53_05_02_01_spi.csv")
spi_meta <- read_csv("CU_53_05_02_02_spi_metadata.csv")

escenario_spi <- read_csv("ESCENARIO_INVERSIONES_PAISES.csv")
```

```
Rows: 24 Columns: 3
  Column specification

Delimiter: ","
chr (1): grupo
dbl (2): anyo, inversion

 Use `spec()` to retrieve the full column specification for this
data.
 Specify the column types or set `show_col_types = FALSE` to quiet
this message.
Rows: 2364 Columns: 81
  Column specification

Delimiter: ","
chr  (3): country, spicountrycode, status
dbl (78): rank_score_spi, spiyear, score_spi, score_bhn, score_fow,
score_op…

 Use `spec()` to retrieve the full column specification for this
data.
 Specify the column types or set `show_col_types = FALSE` to quiet
this message.
Rows: 81 Columns: 8
  Column specification

Delimiter: ","
chr (8): id_var, name_var, role, Dimension, Component, Unit of
measurement, …

 Use `spec()` to retrieve the full column specification for this
data.
 Specify the column types or set `show_col_types = FALSE` to quiet
this message.
Rows: 2352 Columns: 5
  Column specification
```

```
Delimiter: ","
chr (1): spicountrycode
dbl (4): spiyear, inv_inf, inv_tur, inv_san

 Use `spec()` to retrieve the full column specification for this
data.
 Specify the column types or set `show_col_types = FALSE` to quiet
this message.
```

1.2 Modelo

[4]:
```r
dfmodel <- spi |>
  filter(spicountrycode != "WWW") |>
  filter(!is.na(score_spi)) |>
  select(spicountrycode, spiyear, score_spi) |>
    inner_join(escenario_spi)
x <- dfmodel |> select(-c(spicountrycode, score_spi) ) |> as.matrix()
y <- dfmodel |> select(score_spi)  |> as.matrix()

## Modelo
cv_model <- cv.glmnet(x, y, alpha = 0)

#find optimal lambda value that minimizes test MSE
best_lambda <- cv_model$lambda.min

## MOSTRAR EN INTERFAZ
best_lambda

# plot(cv_model)

best_model <- glmnet(x, y, alpha = 0, lambda = best_lambda)
cc <- coef(best_model)


## MOSTRAR EN INTERFAZ
cc
```

```
Joining with `by = join_by(spicountrycode, spiyear)`
```

1.48045312738535

```
5 x 1 sparse Matrix of class "dgCMatrix"
                      s0
(Intercept) 29.708657618
spiyear     -0.003411833
inv_inf      0.801822794
inv_tur     14.972654971
inv_san     19.117756691
```

1.3 Predicción

[5]:
```
## Predicción escenario región

escenario_region <- read_csv("ESCENARIO_INVERSIONES_REGION.csv")

prediccion <- predict(best_model, escenario_region |> makeX())

## Representar series: escenario + predicción

## guardar modelo

write_rds(best_model, "modelo_reg.rds")
```

```
Rows: 4 Columns: 4
  Column specification

Delimiter: ","
dbl (4): spiyear, inv_inf, inv_san, inv_tur

 Use `spec()` to retrieve the full column specification for this
data.
 Specify the column types or set `show_col_types = FALSE` to quiet
this message.
```