CU04_MODEL_DEVELOPMENT_01_GAM

June 9, 2023

#

CU04_Optimización de vacunas

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 GAM

1.1.1 Paquetes

[1]:
```
## Paquetes
library(readr)
library(dplyr)
library(mgcv)
library(visibly) # install_github("m-clark/visibly")
library(ggeffects)
```

```
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


Loading required package: nlme


Attaching package: 'nlme'
```

```
The following object is masked from 'package:dplyr':

    collapse


This is mgcv 1.8-42. For overview type 'help("mgcv-package")'.

Loading required package: ggplot2
```

1.1.2 Datos

[2]:
```
data_04_completo <- read_csv("CU_04_05_19_vacunacion_gripe_completo.csv") |>
  mutate(GEOCODIGO = factor(GEOCODIGO),
         scampana = as.numeric(factor(paste0(ano, semana, sep = "-"))))
```

```
Rows: 21736 Columns: 47
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (3): GEOCODIGO, DESBDT, nombre_zona
dbl (44): ano, semana, n_vacunas, n_citas, tmed, prec, velmedia,
presMax, be…

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

1.1.3 Spot checking

Se han probado varios modelos

[3]:
```
# f <- as.formula(paste0("n_vacunas ~  ",
#        paste0("s(", colnames(data_04)[-c(1:2)], collapse = ") + "),
#        ")"))

# f <- as.formula(
#   "n_vacunas ~ s(tmed) + s(presMax) + "
# )


f <- as.formula(
  "n_vacunas ~ GEOCODIGO + s(scampana) + s(tmed) + s(prec) + s(velmedia) +␣
↪s(presMax) + s(benzene) +
    s(co) + s(no) + s(no2) + s(nox) + s(o3) + s(pm10) + s(pm2.5) +
    s(so2) + s(capacidad_zona) + s(prop_riesgo) + s(tasa_riesgo) +
```

```
    s(tasa_mayores) + s(poblacion_mayores) + s(nsec) + s(t3_1) +
    s(t1_1) + s(t2_1) + s(t2_2) + s(t4_1) + s(t4_2) + s(t4_3) +
    s(t5_1) + s(t6_1) + s(t7_1) + s(t8_1) + s(t9_1) + s(t10_1) +
    s(t11_1) + s(t12_1) + s(area) + s(densidad_hab_km)"
)
```

1.1.4 Ajuste del modelo

[4]:
```
mod_04_gam <- data_04_completo |>
  gam(f, data = _, family = poisson(link = log))
```

1.1.5 Guardar modelo para despliegue

[5]:
```
write_rds(mod_04_gam, "mod_04_gam.rds")
```

1.1.6 Generar escenarios

[6]:
```
newdata <- data_04_completo |>
  select(-n_vacunas, -n_citas, -nombre_zona) |>
  filter(ano == 2021 & semana >= 36 | ano == 2022 & semana <= 5) |>
  mutate(scampana = as.numeric(factor(paste0(ano, semana, sep = "-")))) |>
  select(-ano, -semana, -DESBDT)

write_csv(newdata, "NEWDATA.csv")
```

1.1.7 Predicción

[8]:
```
prediction <- predict.gam(mod_04_gam, newdata, se.fit = TRUE, type = "response")

newdata |> bind_cols(data.frame(prediction)) |> write_csv("PREDICTION.csv")
```