

Ejemplos de Aplicación - Semana 10

Problema de Asignación de publicidad

Una empresa de telefonía móvil, desea construir un modelo para poder pronosticar la probabilidad de abandono de sus clientes (Churn). Para ello, tomo una muestra de 3000 clientes, que incluye tanto clientes que abandonaron la empresa como otros que no lo hicieron. Para estos 3000 clientes se obtuvieron ciertas variables de comportamiento en el mes anterior.

- *INTL*: si ha contratado el servicio de llamadas internacionales (si/no).
- *VMAIL*: voicemail (si/no).
- *VMAILMSG*: número de mensajes de voicemail
- *DAYS**MIN*: longitud en minutos de las llamadas de día (tarifa 1).
- *DAYS**CALLS*: número de llamadas efectuadas de día (tarifa 1)
- *DAY**CHARGE*: precio total de llamadas efectuadas de día (tarifa 1)
- *EVE**MIN*: longitud en minutos de las llamadas de noche (tarifa 2).
- *EVE**CALLS*: número de llamadas efectuadas de noche (tarifa 2)
- *EVE**CHARGE*: precio total de llamadas efectuadas de noche (tarifa 2)
- *NIGHT**MIN*: longitud en minutos de las llamadas de noche (tarifa 3)
- *NIGHT**CALLS*: número de llamadas efectuadas de noche (tarifa 3)
- *NIGHT**CHARGE*: precio total de las llamadas de noche (tarifa 3)
- *INTL**MIN*: minutos de llamadas internacionales
- *INTL**CALLS*: número de llamadas internacionales
- *INTL**CHARGE*: precio total de llamadas internacionales
- *SV**CALLS*: llamadas al servicio de atención al cliente.
- *CHURN*: abandona (si/no)

Datos: *Ejemplo_Aplicacion_Semana_10.xlsx* Hoja: *Problema1*

Cargamos las librerías

```
# Para importar datos de excel
library(readxl)
# Para graficos más profesionales
```

```
library(ggplot2)
# Librerias
library(dplyr)
```

Cargamos los datos del archivo

```
# IMPORTAR DATOS
datos <- read_excel("Ejemplo_Aplicacion_Semana_10.xlsx",
                    sheet = "Problema1")
```

a) Estime un modelo completo utilizando todas las variables. Analice las variables significativas desde el punto de vista del negocio. Dentro de las variables que resultan significativas hay muchas que tienen que ver con el tipo de servicios que la cuenta utiliza. Sin embargo hay una variable sobre la que la empresa puede tomar acciones correctivas para poder intentar reducir la perdida de clientes. ¿Cuál es? ¿Es significativa?

```
m1 <- glm(Churn ~ ., data = datos, family = binomial)
summary(m1)
```

b) Analice la multicolinealidad del modelo completo. ¿Cree usted que puede haber problemas de colinealidad con las variables que está incluyendo? ¿Indican la métricas (DET y VIFs) problemas de multicolinealidad?

```
mm <- model.matrix(m1)
det(cor(mm[, -1]))
library(car)
vif(m1)
```

c) Corra un modelo utilizando solo las variables que son significativas en el punto anterior. ¿Comparando este nuevo modelo con el anterior, cual es mejor?

d) Genere dos nuevos modelos utilizando los procedimientos de selección hacia adelante y hacia atrás. Compare los 4 modelos obtenidos y seleccione el mejor de acuerdo con el criterio del AIC.

```
modelo_intercepto <-
  glm(formula = Churn ~ 1, data = datos, family = binomial)
modelo_completo <-
  glm(formula = Churn ~ ., data = datos, family = binomial)

sm1 <- step(
  object = modelo_completo,
  scope = formula(modelo_completo),
  direction = "backward",
  trace = 2)

summary(sm1)

sm3 <- step(
  object = modelo_intercepto,
  scope = formula(modelo_completo),
  direction = "forward",
  trace = 2)

summary(sm3)
```

e) Utilizando el modelo del punto c) calcule la probabilidad de que un cliente con plan internacional pero sin vocemail, con 10 llamadas internacionales y sin llamadas al servicio al cliente.

```
predict_dataset <- data.frame(
  IntlPlan = "yes",
  VMailPlan = "no",
  VMailMessage = 0,
  IntlCalls = 10,
  CustServCalls = 0)

predict_dataset
```

```
# Este valor de predicción es de  $Z = b_0 + b_1 x_1 + b_2 x_2 + \dots$ 
predict(m2, newdata = predict_dataset)

# hacemos la conversión para calcular la probabilidad
1/(1+exp(-predict(m2, newdata = predict_dataset)))
```

f) Para el mismo cliente anterior, suponiendo que todas las demás variables permanecen constantes, grafique la probabilidad de Churn en función de la cantidad de llamadas. Compare esta curva para clientes con y sin plan internacional. ¿Cómo podría utilizar esta información para elaborar un plan de retención?

```
predict_dataset1 <- data.frame(
  IntlPlan = "yes",
  VMailPlan = "no",
  VMailMessage = 0,
  IntlCalls = 10,
  CustServCalls = 0:10)
predict_dataset2 <- data.frame(
  IntlPlan = "no",
  VMailPlan = "no",
  VMailMessage = 0,
  IntlCalls = 10,
  CustServCalls = 0:10)

predict_dataset1
predict_dataset2

predict_dataset1$prediccion_con_plan_int <-
  1/(1+exp(-predict(m2, newdata = predict_dataset1)))
predict_dataset2$prediccion_con_plan_int <-
  1/(1+exp(-predict(m2, newdata = predict_dataset2)))

ggplot()+
  geom_line(aes(CustServCalls, prediccion_con_plan_int),
            color = 'red',
            data = predict_dataset1)+
  geom_line(aes(CustServCalls, prediccion_con_plan_int),
```

```

        color = 'blue',
        data = predict_dataset2)+
ylim(0,1)+
theme_bw()

```

g) Agregue un interacción entre las llamadas internacionales y las llamadas al servicio al cliente. Esto le permite que al modelo independizar la relación entre las llamadas a clientes y la probabilidad de churn para ambos tipos de clientes.

```

m5 <- glm(Churn ~ IntlPlan + VMailPlan + VMailMessage + IntlCalls + CustServCalls +
          IntlPlan:CustServCalls,
          data = datos, family = binomial)
summary(m5)

predict_dataset1 <- data.frame(
  IntlPlan = "yes",
  VMailPlan = "no",
  VMailMessage = 0,
  IntlCalls = 10,
  CustServCalls = 0:10)
predict_dataset2 <- data.frame(
  IntlPlan = "no",
  VMailPlan = "no",
  VMailMessage = 0,
  IntlCalls = 10,
  CustServCalls = 0:10)

predict_dataset1
predict_dataset2

predict_dataset1$prediccion_con_plan_int <-
  1/(1+exp(-predict(m5, newdata = predict_dataset1)))
predict_dataset2$prediccion_con_plan_int <-
  1/(1+exp(-predict(m5, newdata = predict_dataset2)))

ggplot()+
  geom_line(aes(CustServCalls, prediccion_con_plan_int),
            color = 'red',
            data = predict_dataset1)+

```

```
geom_line(aes(CustServCalls, prediccion_con_plan_int),  
          color = 'blue',  
          data = predict_dataset2)+  
ylim(0,1)+  
theme_bw()
```