
Trabajo Práctico 2

Predicting Customer Churn at QWE Inc

Métodos Estadísticos Aplicados a los Negocios

Fecha de entrega: 6 de mayo de 2019

1. INSTRUCCIONES

El trabajo puede ser grupal (con un máximo de 4 alumnos). Deberán entregar un informe final **a través del campus virtual** en el cual adjuntarán el Script en R para poder replicar sus resultados.

Tienen tiempo para entregar el TP hasta el lunes 6 de mayo a las 23.55hs. Una entrega tardía implica desaprobar el trabajo práctico.

2. CONSIGNA

Lea detenidamente el caso “Predicting Customer Churn at QWE Inc.” de A. Ovchinnikov (2013). Utilice la base de datos adjunta al caso para contestar las siguientes preguntas:

- (1) **(5 puntos)** Describa la base de datos y obtenga los principales estadísticos descriptivos. Comente.
- (2) **(5 puntos)** En forma aleatoria seleccione el 80% del total de los casos. Dicha submuestra (*training set*) será utilizada para estimar distintos modelos, mientras que el

20% restante (*testing set*) será utilizado para evaluar los pronósticos. Vuelva a calcular los estadísticos descriptivos para cada sub-muestra, ¿hay diferencias significativas? Comente.

- (3) **(10 puntos)** Utilizando el *training set*, estime un modelo lineal de probabilidad (MLP) que identifique cuáles son los factores que condicionan la probabilidad de que un cliente cancele su contrato (incluya la antigüedad del cliente). Interprete.
- (4) **(20 puntos)** Para la misma muestra y utilizando los mismos regresores, estime un modelo probit y otro logit. Interprete a partir de los *odd-ratios* en el caso logit.
- (5) **(10 puntos)** A partir de los modelos estimados en el punto anterior, grafique cómo cambia la probabilidad de cancelar el contrato en función de la antigüedad del cliente. Comente.
- (6) **(10 puntos)** Tanto para el modelo probit como para el logit compute la proporción correctamente estimada. Esto es, se considera que la proporción de clientes que cancelan el contrato es correctamente estimada si $\text{churn}=1$ y la probabilidad estimada del modelo es superior al 50% o si $\text{churn}=0$ y la probabilidad estimada del modelo es inferior a 50%. Comente.
- (7) **(10 puntos)** Usando el *testing set*, calcule el error de pronóstico de cada modelo (MLP, Probit y Logit). Obtenga sus principales estadísticos descriptivos. Comente.
- (8) **(10 puntos)** En función de los errores de pronósticos obtenidos en el punto anterior evalúe los pronósticos calculando el RMSE y MAPE en cada caso. Comente. Recuerde:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2}$$
$$MAPE = \left(\frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{y_i} \right) \times 100$$

- (9) **(10 puntos)** Para evaluar la presencia de sesgos sistemáticos en los pronósticos, regrese los errores de pronósticos en función de una constante y evalúe su significatividad. Comente.
- (10) **¡Vamos por el 10! (10 puntos)** En forma análoga al cálculo de la proporción correctamente estimada de cada modelo que realizó en el punto 6, esta vez trabaje con el *testing set* y compute la proporción de falsos positivos y falsos negativos. Llamamos falsos positivos a los casos en los que el modelo predecía que el cliente iba a cancelar el contrato cuando en realidad no lo hizo, y llamamos falsos negativos a los casos en los que el modelo predecía que el cliente no iba a cancelar el contrato cuando en realidad sí lo hizo. Comente.