Handout 2

Plazo de entrega: 16 de Julio de 2018, hasta las 23:59.
Medio de entrega: Enviar link del *notebook* a hl.benavides@uniandes.edu.co y s.ramirez34@uniandes.edu.co

---

Instrucciones

1. El handout debe ser desarrollado en una notebook (R o python).
2. El notebook debe ser publicado por uno de los integrantes del grupo en Github (Python o R) o Rpubs (R). En el caso de Github, no es necesario contribuirlo al repositorio del curso y por lo tanto no se deben hacer pull requests.
3. Como las notebooks de R pueden tener problemas en Github con los gráficos, se recomienda utilizar como opción Rpubs.com. Para aquellos que utilicen Github, la totalidad del contenido del notebook (texto y visualizaciones) debe poder compartirse usando el link en nbviewer (https://nbviewer.jupyter.org).
4. El uso de Rpubs es muy sencillo, sólo requiere registrarse en la página web (https://rpubs.com/) y luego usar la opción *Publish* en Rstudio. Roger Peng tiene un tutorial de 3 minutos que fácilmente pueden seguir (https://www.youtube.com/watch?v=yuQG6078L1U)
5. El contenido del notebook no debe ser modificado después de la fecha límite.
6. Se debe trabajar en grupos de 2 a 4 personas.
7. El link sólo debe enviarlo un integrante del grupo antes del plazo de entrega.
8. Se calificaran también normas básicas de presentación de trabajos.

Ejercicio

Mobile Money (*momo*) platforms are gaining traction across developing markets as a convenient way of sending and receiving money over mobile phones. These systems operate on low-cost feature phones and do not require users to have a bank account. These low barriers-to-entry make mobile money platforms excellent tools for financial inclusion of the poor.

The mobile-based saving and lending product is a joint collaboration between a mobile network operator (MNO) and a regional bank in East Africa. The MNO provides access to the customer through their mobile money platform as well as data on recent customer activity. This data is shared with the banking partner when a customer signs up for the savings and loan product in order to help determine risk. In turn, the bank provided labeled data of customers who have repaid their mobile loans and those who have defaulted.

Available data includes 19 numeric features, 1 categorical feature and 1 categorical label for 50,000 customers of MNO and banking partner services in 2013 and 2014. There are 5 types of *momo* transactions:

- *send* is sending *momo* to another person
- *received* is receiving *momo* from another person
- *deposit* is turning cash into *momo*
- *withdrawal* is turning *momo* into cash
- *bill* is paying a bill through *momo*

Customer use of every type of transaction during the period of study is described by three associated variables:

- *amt* is the total cash amount involved in the associated type of transaction.
- *qty* is the total count of associated type of transaction.
- *dgr* is the number of unique people involved in the associated type of transaction (short for degree).

Three additional customer network use variables are available:

- *airtime* is how much airtime has been used
- *topup* is how much money has been spending on topping up the phone's airtime.
- *days no balance* is a measure of how often the phone has no airtime balance

Demographic self explanatory variables *age* and *gender* are also available. Finally, categorical label *default* indicates whether customer failed to repay their mobile loans.

What can you say about customer credit worthiness from available phone use data? What does your analysis using machine learning tools suggest about product design for financial inclusion of the poor? Your analysis should clearly show that you have:

- Explored the available dataset
- Framed the policy issue as a machine learning problem
- Identified the more relevant machine learning techniques for the problem at hand
- Used an adequate model validation strategy that supports your conclusions

Good luck!

Referencias

James, G. Witten, D. Hastie, T. Tibshirani, R (2013). An introduction to Statistical Learning with Applications in R. Springer New York Inc. Available in: http://www-bcf.usc.edu/ gareth/ISL/