**Realizado por:**

Juan Camilo Restrepo Velez  
William Leonardo Andrade Collazos  
Wilder Valencia Ocampo

**PROYECTO MINERÍA DE DATOS EN R Y PYTHON**

Seleccionar una base de datos en <https://www.datos.gov.co/> y aplicar los 3 análisis descriptivos vistos en clase (preferiblemente la misma base de datos analizada en la práctica anterior). Después de realizar todos los pasos, responder las siguientes preguntas en un informe, incluyendo un pantallazo de los modelos en Knime:

**Bank Marketing**

Los datos están relacionados con campañas de marketing directo de una institución bancaria portuguesa. Las campañas de marketing se basaron en llamadas telefónicas. A menudo, se requería más de un contacto con el mismo cliente, para poder acceder a si el producto (depósito bancario a plazo) sería ('sí') o no ('no') suscrito.

**Información de atributos**

Información Bancaria de los clientes

Age - Edad

Job - Trabajo: tipo de trabajo

Marital - Estado civil: estado civil

Education - Educación: Nivel educativo

Default - Incumplimiento: ¿tiene el crédito en mora?

Housing - Vivienda: ¿tiene un préstamo de vivienda?

Loan - Préstamo: ¿tiene préstamo personal?

Relacionado con la última llamada de la actual campaña

Contact - Contacto: tipo de comunicación

Month - Mes: último mes de contacto del año

DayofWeek - Día de la semana: último día de contacto de la semana

Duration - Duración: duración del último contacto, en segundos (numérico). Nota importante: este atributo afecta en gran medida al objetivo de salida (por ejemplo, si la duración = 0, entonces y = "no"). Sin embargo, no se conoce la duración antes de una llamada se realiza. Además, después del final de la llamada se conoce obviamente y. Por lo tanto, esta entrada sólo debe incluirse a efectos de referencia y debe descartarse si se pretende tener un modelo predictivo realista.

Otros

Campaign - Campaña: número de contactos realizados durante esta campaña y para este cliente

Pdays - pDías: número de días que pasaron después de que el cliente fue contactado por última vez en una campaña anterior. Nota, 999 significa que el cliente no fue contactado anteriormente

Previous - Anterior: número de contactos realizados antes de esta campaña y para este cliente

Poutcome: resultado de la anterior campaña de marketing

Atributos del contexto social y económico

Emp.var.rate - Tasa de variación del empleo - indicador trimestral

Cons.price.idx: Índice de Precios al Consumidor - Indicador mensual; el Índice de Precios al Consumidor o IPC mide los cambios en los precios pagados por los consumidores por una cesta de bienes y servicios cada mes.

Cons.conf.idx: Índice de confianza del consumidor - Indicador mensual; En Portugal, el índice de confianza del consumidor se basa en entrevistas con los consumidores sobre sus percepciones de la situación económica actual y futura del país y sus tendencias de compra. Se estima utilizando la diferencia entre la proporción de respuestas de evaluación positivas y las respuestas de evaluación negativas, pero no incluye la proporción de respuestas neutras

Euribor3m: euribor 3 meses - Euribor es la abreviatura de Euro Interbank Offered Rate. es un índice de referencia publicado diariamente que indica el tipo de interés promedio al que un gran número de bancos europeos dicen concederse préstamos a corto plazo entre ellos para prestárselo a terceros.

Nr.employed - Número de empleados: Número de empleados - Indicador trimestral; Número de personas empleadas para el trimestre.

y - ¿el cliente ha suscrito un depósito a plazo? (Variable objetivo)

\*Tomado de <https://www.kaggle.com/henriqueyamahata/bank-marketing>

**PREPARACIÓN DE DATOS**

Ya se realizo debido a que se tomaron los datos utilizados en “Practica Análisis Predictivo”

**Con los datos seleccionados en las prácticas anteriores, crear dos modelos de minería de datos con las siguientes características**

**Modelo en R: explicar objetivo del modelo, configuración de la técnica y resultado obtenido. Indicar qué conocimiento nuevo se obtuvo**

**Modelo en Python: explicar objetivo del modelo, configuración de la técnica y resultado obtenido. Indicar qué conocimiento nuevo se obtuvo**

**Observación: se deben aplicar técnicas diferentes**