

Prueba de conocimiento analítico

Vicepresidencia Banca Personas & Pymes

Gerencia Analítica Personas & Pymes

El objetivo de la prueba es medir sus capacidades de análisis y desarrollo de modelos predictivos. La idea es que no le dedique más de 15 horas en total, incluyendo el tiempo para **documentar** lo que hizo.

Es posible usar cualquier herramienta que quiera (R, Python, SAS Guide/Miner, SPSS, etc.), y cualquier recurso del internet, pero no se permite consultar directamente con otras personas por ningún medio.

1 Introducción

La Dirección de Estrategia de Clientes Banca Personas & Pymes es un área de gran impacto en la organización dado que son los encargados de entender las necesidades de los clientes y crear propuestas de valor a la medida que satisfagan dichas necesidades. Lo anterior, alineado con los pilares estratégicos del banco es el área responsable, a través de su gestión de generar preferencia por parte de sus clientes, experiencia superior y rentabilidad a sus inversionistas.

Uno de los principales retos que tiene actualmente el área de Estrategia de clientes consiste en **entender** cómo, por medio del uso correcto de información, se puede atender las **necesidades** de financiación (crédito) de los clientes y poder agilizar los procesos internos de crédito, el desarrollo de esta estrategia traerá consigo beneficio tanto para el cliente (quién recibe respuesta a sus requerimientos en menores tiempos) como para el Banco (dado que se crean procesos más ágiles reduciendo el tiempo que invierten los analistas de crédito procesando la solicitud, tiempo que pueden enfocar en nuevas solicitudes).

Dentro de las conversaciones que ha tenido el área de Estrategia de clientes con el área de Estudio de crédito se ha podido evidenciar que el **Ingreso o salario** es una de las variables más complejas de entender dado que requieren mayor grado de verificación (alargando los procesos de estudio) y es la variable que permite identificar ¿cuál es la capacidad de pago? y determinar ¿cuál es el monto de crédito que podría tener el cliente? Es importante, que esta variable sea verificada correctamente dado que un mal cálculo aumentará la exposición del Banco y comprometerá la **rentabilidad**.

Por lo anterior, el área de Estrategia de clientes busca crear un **modelo predictivo** que le permita estimar el ingreso de sus clientes y proactivamente agilizar todos los procesos mencionados.

2 Definición de la población objetivo y variable respuesta

- El área de Estrategia de clientes ha determinado que busca estimar el ingreso para los clientes del segmento **“Personal”** dado que es el segmento masivo que genera mayor operatividad al área de crédito.
- El modelo de estimación tiene como variable de respuesta el **Ingreso o Salario**.

3 Instrucciones importantes

Se le entregarán tres archivos:

- `base_entrenamiento.csv`
- `descrip_variables.xlsx`
- `base_prueba.csv`

Favor seguir las siguientes instrucciones
--

1. El archivo adjunto **`base_entrenamiento.csv`** contiene información demográfica, comportamiento transaccional y tenencia de producto de los clientes del segmento “**Personal**” la información es única por cliente (es decir, que no existe un cliente repetido). La columna “**llave_cod_cliente**” permite identificar al cliente.
2. En la **`base_entrenamiento.csv`** existe una columna llamada “**dem_salario**” dicha variable, es la variable respuesta y corresponde al ingreso de los clientes que tienen el pago de su nómina con el Banco y servirá para entrenar el modelo de estimación.
3. El archivo **`descrip_variables.xlsx`** contiene la descripción detallada del significado de cada una de las columnas de la **`base_entrenamiento.csv`**.
4. El objetivo es que usted desarrolle un modelo analítico que permita, a partir de los datos del archivo **`base_entrenamiento.csv`** predecir cuál es el ingreso del cliente.
5. El archivo adjunto **`base_prueba.csv`** contiene exactamente las mismas columnas del archivo **`base_entrenamiento.csv`**, exceptuando la columna “**dem_salario**”

4 Entregables

Se debe entregar un archivo **`base_prueba_evaluado.csv`** con las columnas de “**llave_cod_cliente**” y “**dem_salario**”, la columna “**dem_salario**” es un valor entero positivo.

- La columna “**dem_salario**” debe contener un valor entero positivo para todos y cada uno de los registros. No se aceptarán valores nulos, NaNs, N/A, N/D, vacíos, o mensajes de texto como, por ejemplo: “datos incompletos”.

Por favor haga todo lo posible por conservar el formato del archivo (csv separado por comas, no otro carácter; el orden de las columnas; la línea de encabezado etc. El orden de las filas no es crítico).
--

- También se debe entregar la implementación de su modelo (archivos de código con comentarios en caso de usar un lenguaje de programación convencional o el archivo de proyecto que incluya documentación, en caso de usar SAS Miner, Azure ML studio, u otra herramienta parecida).
- Un archivo de texto (en `.txt`, `.doc`, `.html`, `.rmd`, `.md`, `pptx` o `.pdf`) que contenga una descripción más o menos detallada del proceso que siguió para generar el modelo (incluyendo exploración, transformaciones de variables, selección de variables, etc.) y luego generar sus predicciones de “**dem_salario**”.

- De manera opcional podrá hacer saber qué otros datos o atributos de los clientes, de su comportamiento frente al banco añadiría idealmente al conjunto de datos, para un modelo predictivo más efectivo. Aquí, tenga en cuenta la factibilidad y el costo de obtener esos datos.

Ahora, es posible que llegue a la conclusión de que no es posible desarrollar un buen modelo predictivo a partir de la información proporcionada o dada la calidad de esta. Si este es el caso, se quiere evaluar el mejor modelo que usted pueda producir y también que dé una sustentación de esa conclusión.

5 Evaluación

La métrica para evaluar el modelo será el **MAPE** calculado a partir del valor real de la variable “**dem_salario**” (información que sólo nosotros conocemos) y el valor predicho por su modelo y consignado en el archivo ***base_prueba_evaluado.csv*** por medio de la columna “**dem_salario**”.

Adicionalmente se evaluará integralmente el informe entregado: el análisis de las variables, su transformación, proceso de selección, y cualquier componente analítico que haya sido útil para la construcción del modelo.

Finalmente usted debe preparar una presentación resumen a partir de la información que suministre en el **punto 3** del literal **4 Entregables** y exponerla en la fecha que se le asigne.

¡Éxitos!