

HACKATHON N°01 INCABANK

DESCRIPCIÓN DEL BANCO:

INCABANK es un banco local de Perú, con una trayectoria en el mercado de más de 30 años. Sus clientes se distribuyen en banca mayorista (grandes empresas y corporaciones) y banca minorista (personas y pequeñas empresas), siendo la banca minorista el segmento que más contribuye a la utilidad del negocio (por encima del 60%).

El número de clientes de banca minorista de INCABANK ha aumentado de forma moderada pero constante desde su creación. Sin embargo, teniendo en cuenta el crecimiento del país en los últimos años, se ha detectado un gran aumento de la demanda en tarjetas de crédito entre los clientes del banco.

ROL DEL ÁREA DE METODOLOGÍA Y MODELAMIENTO DE GESTIÓN DEL RIESGO:

El área de Metodología y Modelamiento de Gestión del Riesgo es un área que da soporte al proceso de crédito del banco. En concreto, desarrolla herramientas que consisten en modelos estadísticos que permiten discriminar los niveles de riesgo de crédito, estimar la probabilidad de incumplimiento de las operaciones de créditos, así como estimar los niveles de ingreso de los sujetos de créditos: personas o negocios.

SU ROL:

Como **Subgerente de Metodología y Modelamiento de Gestión del Riesgo**, debe ser capaz de trabajar con la información que tiene sobre los clientes, para poder ayudar al área de crédito del banco a minimizar sus riesgos, identificando qué clientes son susceptibles de retrasarse en sus pagos o incumplirlos.

En estos momentos, el banco cuenta con una base de datos acerca de los clientes que poseen una tarjeta de crédito. En función a estos datos, su área confecciona modelos estadísticos de predicción que permiten que el banco tome decisiones sobre a quién otorga créditos.

Puede ver esta información en el archivo de Excel. Esta base de datos incluye variables relevantes para poder evaluar qué probabilidad de mora tiene cada cliente.

PROGRAMA DE ESPECIALIZACIÓN EN “MACHINE LEARNING FOR DATA SCIENCE”

Las variables incluidas en el archivo son las siguientes:

Variable	Descripción
ID	Identificador
MORA60	Indicador mora>60 días en el transcurso de un año luego de haberle otorgado el crédito (0: No tuvo mora 60, 1: Sí tuvo mora 60).
VAR01	Número de días de atraso máximo durante los 12 meses antes de la aprobación del crédito.
VAR02	Situación de la casa (alquilada, familiar, propia, otras)
VAR03	Edad
VAR04	Fecha ingreso al empleo.
VAR05	Máxima antigüedad con Tarjeta de crédito en el Sistema Financiero (en meses).
VAR06	Número de meses con algún producto Pasivo (ahorro) durante los 12 meses antes de la aprobación del crédito.
VAR07	Ingreso mensual.
VAR08	Monto de Línea de crédito Tarjeta de Crédito en el Sistema Financiero.
VAR09	Monto de Saldo deudor promedio total en el Sistema Financiero.
VAR10	Score con el que fue aprobada la Tarjeta de crédito en el Banco.
VAR11	Número de veces con Abono de Pago de Haberes en el Banco durante los 12 meses antes de la aprobación del crédito.
VAR12	Zona geográfica.
VAR13	Máxima clasificación de Riesgos SBS durante los 12 meses antes de la aprobación del crédito (0: Normal, 1: CPP, 2: Deficiente, 3: Dudoso, 4: Pérdida).
VAR14	Nivel educativo.

RETO GRUPAL:

Teniendo en cuenta su rol y los datos que se le han facilitado, debe realizar una presentación en donde responda a las siguientes cuestiones:

- Identifique los clientes que tienen una alta probabilidad de mora.
- Justifique su propuesta con datos, utilizando las herramientas que considere oportunas (tablas, gráficos, etc.).
- Identifique las variables más relevantes para este análisis.

ENTREGA:

Debe realizar una presentación en Power Point (5 slides como máximo) respondiendo a las preguntas que figuran en el apartado “El reto grupal” y enviar el score dentro del tiempo establecido dentro del tiempo estipulado.

HORARIO Y CONSIDERACIONES:

- ✓ La entrega del score final es hasta culminar la clase.
- ✓ Los grupos de trabajo serán seleccionados aleatoriamente con los participantes del trabajo, bajo la supervisión de los docentes del curso.
- ✓ A partir de ahí comienza la ronda de exposiciones con la presencia de la Coordinadora de la Maestría en Ciencia de Datos.
- ✓ El score final sobre el cuál se calificará será el “**Accuracy**” o **Precisión Global del Modelo**.
- ✓ Los modelos a desarrollar para el score final los puede hacer en el lenguaje R o Python de acuerdo a su preferencia.
- ✓ Cada grupo tendrá una carpeta compartida en el drive donde subirá su presentación en PPT y sus archivos de score en formato “**csv delimitado por comas**” o **Excel** con las columnas ID y Predicción (Binario: 0 y 1).
- ✓ El formato del archivo score debe ser por ejemplo: **grupo1_score001** y así progresivamente.
- ✓ Tendrá solo “10 minutos” para ejecutar su exposición y responder las preguntas del panel.
- ✓ La nota del programa será en estricta jerarquía de los resultados obtenidos y de la creatividad mostrada en el transcurso del análisis.

$$\text{Nota Final} = 0.70 * \text{Score del Modelo} + 0.30 * \text{Presentación de Resultados}$$



PROGRAMA DE ESPECIALIZACIÓN EN “MACHINE LEARNING FOR DATA SCIENCE”

CODIFICACIÓN DE VARIABLES:

NIVEL:

EDUCACION BASICA	1
PROFESIONAL	2
SIN EDUCACION	3
SUPERIOR	4
TECNICO	5

ZONA:

CAPITAL	1
CENTRO	2
NORTE	3
ORIENTE	4
SUR	5

TIPO VIVIENDA:

ALQUILADA	1
FAMILIAR	2
OTRAS	3
PROPIA	4