| **Sigla** | **Nombre Asignatura** | **Horas semana** |
| --- | --- | --- |
| **MLY0100** | **MACHINE LEARNING** | *4 h semana 17* |

| **Ítem** | **Puntaje** | **% Ponderación** |
| --- | --- | --- |
| **Competencia Especialidad** | **84 puntos** | **100%** |

**Instrucciones generales:**

A continuación, se presenta el caso a ser utilizado durante el semestre y como examen transversal de la asignatura MLY0100.

**Caso Banco “Monopoly/Dormammu”**

El banco Monopoly lleva muchos años atendiendo a sus clientes en Chile y recientemente ha sido adquirido por un Banco con capitales extranjeros llamado “Dormammu”. Dormammu ha encargado a sus ingenieros hacer estudios sobre los clientes del banco Monopoly, para conocer su comportamiento y detectar patrones. Además, debe analizar cómo será su estrategia para abordar estos nuevos clientes dado el uso que ellos hacen de sus productos financieros.

Los ingenieros del área informática del banco Monopoly han extraído una base de datos según una solicitud recibida por el nuevo dueño del banco y la han compartido con los ingenieros de Dormammu que están en la sede de New York. La base contiene una muestra de clientes con 12 meses de información almacenada en variables mensuales por cada cliente.

Usted es parte del equipo de ingenieros del banco Dormammu, y con sus colegas debe analizar esta base, limpiarla, seleccionar las variables que sirvan para alguna interpretación y entregar la mayor información posible a los dueños de este banco para que puedan conocer a los clientes, preparar una estrategia y abordar a este nuevo mercado.

La base de datos tiene variables asociadas a información del cliente y variables mensuales, es decir una por cada mes, el cliente no se repite y las variables asociadas se van agregando como una columna más. Con esto, la base de datos tiene 574 variables y 51.124 registros. La descripción de las variables se muestra a continuación:

| **Variable** | **Descripción** |
| --- | --- |
| CORRELATIVO | Identificador cliente |
| Región | Región de Residencia |
| Renta | Renta del cliente |
| Sexo | Sexo |
| Subsegmento | Subsegmento |
| Edad | Edad |
| Adicional | Indicador de Tenencia de TC adicionales |
| Antiguedad | Antigüedad del cliente (meses) |
| CambioPin | Indicador del cambio de clave secreta de la tarjeta |
| Consumo | Indicador de Crédito de Consumo |
| Debito | Indicador de Tenencia de TD |
| Ctacte | Indicador de Cuenta Corriente |
| Cuentas | Número de cuentas que tiene el cliente |
| Hipotecario | Indicador de Crédito Hipotecario |
| Internauta | Indicador de cliente Internauta que usa la web del banco |
| Monoproducto | Indicador de si el cliente es solo poseedor de una TC |
| TC | Número de TC que tiene el cliente |
| Dualidad | Indicador de Dualidad (Cliente es dual si tiene 2 o más TC) |
| CUPO\_L1 | Cupo de la tarjeta crédito para compras nacionales |
| CUPO\_MX | Cupo de la tarjeta crédito para compras internacionales |
| CUPO\_L2 | Cupo de la tarjeta crédito para avances en cuotas |
| Col\_T12 | Colocación del cliente en TC en el mes X |
| ColL1TE\_T12 | Revolving del cliente en TC en el mes X |
| EeccInt\_T12 | Monto internacional exigido en el estado de cuenta del cliente en el mes X |
| EeccNac\_T12 | Monto nacional exigido en el estado de cuenta del cliente en el mes X |
| Fac\_T12 | Monto facturado por el cliente en TC en el mes X |
| FacAI\_T12 | Monto facturado en avances internacionales por el cliente en TC en el mes X |
| FacAN\_T12 | Monto facturado en avances nacionales por el cliente en TC en el mes X |
| FacCCOT\_T12 | Monto facturado en compras en cuotas con tasa por el cliente en TC en el mes X |
| FacCCPC\_T12 | Monto facturado en compras en cuotas precio contado por el cliente en TC en el mes X |
| FacCI\_T12 | Monto facturado en compras internacionales por el cliente en TC en el mes X |
| FacCN\_T12 | Monto facturado en compras nacionales por el cliente en TC en el mes X |
| FacCOL\_T12 | Monto facturado en avances en cuotas por el cliente en TC en el mes X |
| FacDebAtm\_T12 | Monto facturado en avances por el cliente en TD en el mes X |
| FacDebCom\_T12 | Monto facturado en compras por el cliente en TD en el mes X |
| FacPAT\_T12 | Monto facturado en PAT por el cliente en TC en el mes X |
| FlgAct\_T12 | Indicador de actividad en el mes X en la TC |
| FlgActAI\_T12 | Indicador de actividad en avances internacionales en el mes X en la TC |
| FlgActAN\_T12 | Indicador de actividad en avances nacionales en el mes X en la TC |
| FlgActCCOT\_T12 | Indicador de actividad en compras nacionales en cuotas con tasa en el mes X en la TC |
| FlgActCCPC\_T12 | Indicador de actividad en compras nacionales en cuotas precio contado en el mes X en la TC |
| FlgActCI\_T12 | Indicador de actividad en compras internacionales en el mes X en la TC |
| FlgActCN\_T12 | Indicador de actividad en compras nacionales en el mes X en la TC |
| FlgActCOL\_T12 | Indicador de actividad en avances en cuotas con tasa en el mes X en la TC |
| FlgActPAT\_T12 | Indicador de actividad en PAT en el mes X en la TC |
| PagoInt\_T12 | Monto de pagos de deuda internacional del cliente en el mes X |
| PagoNac\_T12 | Monto de pagos de deuda nacional del cliente en el mes X |
| Txs\_T12 | Número de transacciones realizados por el cliente en TC en el mes X |
| TxsAI\_T12 | Número de transacciones en avances internacionales realizados por el cliente en TC en el mes X |
| TxsAN\_T12 | Número de transacciones en avances nacionales realizados por el cliente en TC en el mes X |
| TxsCCOT\_T12 | Número de transacciones en compras en cuotas con tasa realizados por el cliente en TC en el mes X |
| TxsCCPC\_T12 | Número de transacciones en compras en cuotas precio contado realizados por el cliente en TC en el mes X |
| TxsCI\_T12 | Número de transacciones en compras internacionales realizados por el cliente en TC en el mes X |
| TxsCN\_T12 | Número de transacciones en compras nacionales realizados por el cliente en TC en el mes X |
| TxsCOL\_T12 | Número de transacciones en avances en cuotas por el cliente en TC en el mes X |
| TxsDebAtm\_T12 | Número de transacciones en avances realizados por el cliente en TD en el mes X |
| TxsDebCom\_T12 | Número de transacciones en compras realizados por el cliente en TD en el mes X |
| TxsPAT\_T12 | Número de transacciones en PAT realizados por el cliente en TC en el mes X |
| UsoL1\_T12 | Monto de deuda en la línea de compras en la TC en el mes X |
| UsoL2\_T12 | Monto de deuda en la línea de avances en cuotas en la TC en el mes X |
| UsoLI\_T12 | Monto de deuda en la línea de compras internacionales en la TC en el mes X |

**Formato Entregas**

Las entregas tanto parciales como finales, consisten en un Jupyter Notebook con formato de informe que detalle cada una de las etapas de la metodología CRISP-DM y los procedimientos realizados para el proyecto de Machine Learning.

Escribir y hacer uso de markdown o cuadros de texto que expresen y detallen los descubrimientos de patrones en la data, considerando análisis sobre aspectos del negocio, de los datos y del contexto del enunciado.

Hacer uso de componentes estadísticos y matemáticos para describir la información, los datos y su distribución, con el objetivo de tener conocimiento de estos a través de un análisis exploratorio.

Realizar transformaciones y preprocesamiento a los datos, utilizando las mejores prácticas y combinando las mejores técnicas con el objetivo de tener buenos indicadores y resultados en las predicciones y/o segmentaciones realizadas.

PARTE 1: **Análisis Exploratorio y** **Preprocesamiento**

**INSTRUCCIÓN:**

Entrega avance de Proyecto a nivel de Análisis Exploratorio inicial y Preprocesamiento de los datos. Se deben analizar las entidades representadas en el conjunto de datos (tipos de datos, cantidad de datos, estadísticos, distribuciones, etc), detectar distribuciones para los campos que consideren relevantes, además de analizar la existencia de anomalías (valores atípicos) y de missing values. Proponer y ejecutar limpieza de los mismos, incluyendo las transformaciones que consideren necesarias.

Entrega parcial 1: Semana 6

PARTE 2: **Aprendizaje Supervisado**

**INSTRUCCIÓN**

Entrega avance del Proyecto: se debe presentar un análisis aplicado de algoritmos de aprendizaje supervisado considerando la columna target, seleccionando aquel modelo que presente mejores resultados en predicción y mayor ajuste con los datos de entrenamiento, testeado en un set de pruebas o de validación y que su selección sea respaldada y justificada a través de sus métricas de desempeño.

Entrega parcial 2: Semana 12

PARTE 3: **Aprendizaje no supervisado**

**INSTRUCCIÓN**

Entrega avance de Proyecto: se debe presentar un análisis aplicado usando algoritmos de aprendizaje no supervisado, seleccionando aquel modelo que presente mejores resultados con los datos de entrenamiento y que su selección sea respaldada y justificada a través de sus métricas de desempeño.

Entrega parcial 3: Semana 17

Examen Transversal: **Conclusiones Finales y presentación del ET**

**INSTRUCCIÓN**

Entrega Final con las correcciones y feedbacks del Proyecto, a nivel del uso de Machine Learning en conocimiento y descubrimiento de patrones extraídos para uso de predicciones, presenta los resultados de los análisis seleccionando el mejor modelo. Realiza presentación de los resultados obtenidos de una manera visual, complementando con conocimiento que aporte a las decisiones del negocio.