

**SILABO****Módulo: Aprendizaje Supervisado y modelos binarios**

Del (los) programa(s)	:	Diploma Data Scientist
Edición	:	19
Duración	:	15 horas cronológicas 5 sesiones
Horario	:	7:30pm – 10:30pm
Docente	:	Jimmy Salazar Gebol
Correo electrónico	:	Jimsge3@gmail.com

I. SUMILLA

El módulo es de naturaleza teórico y práctico. Le permitirá al participante entrenar y evalúa modelos de clasificación binaria empleando el lenguaje Python y sus librerías asociadas a este propósito, considerando algoritmos como regresión lineal, regresión logística, árboles de decisión y técnicas ensemble. Las temáticas que comprenden este módulo son:

- a) Técnicas de clasificación
- b) Técnicas de clasificación combinadas (Ensembles)
- c) Evaluación de modelos de clasificación
- d) Técnicas de optimización de modelos de clasificación
- e) Fundamentos de Redes neuronales

II. METODOLOGÍA**Metodología DMC: Learning Agile**

La metodología DMC garantiza un aprendizaje significativo mediante la práctica intensiva, combinando tres enfoques:

Secuencia Didáctica Descubre-Explora-Aplica

- **Descubre** 🔎 (Introducción al tema) → Videos, diapositivas y ejemplos.
- **Explora** 🚧 (Práctica guiada) → Ejercicios en clase con apoyo del docente.
- **Aplica** 💾 (Práctica autónoma) → Ejercicios en casa con solucionarios (Puede ser parte de la Evaluación Formativa).

III. HERRAMIENTAS

- Python, TensorFlow, Google Colab.
Provisto por DMC: No.
Costo asociado por alumno: Ninguno

IV. CONTENIDO

Sesión 1

Técnicas de clasificación

- El principio de linealidad en modelamiento de datos.
- El modelo más sencillo de todos: La regresión Lineal.

Casos prácticos.

Caso de Uso: Aplicación de algoritmos de regresión lineal para estimar valores cuantitativos y regresión logística binaria para predecir el objetivo categórico.

Ejemplo: Se solucionará el caso para estimar el ingreso e identificar las principales variables para explicar si una persona abrirá o no una cuenta de ahorro a plazo.

Valor: Identificar las diferencias entre el uso de algoritmos de regresión lineal y logístico según el caso estudiado.

- Taller: Entrenamiento y análisis de un modelo de regresión lineal con Python.
- Regresión Logística y la función sigmoide.
- Taller: Entrenamiento y análisis de un modelo de clasificación binaria basado en regresión logística con Python.

Sesión 2

Técnicas de clasificación

- Árboles de decisión. Algoritmos para su estimación CART, C45.

Casos prácticos.

Caso de Uso: Aplicación de algoritmos de árboles de decisión e identificación de las principales variables que expliquen el objetivo.

Ejemplo: Se solucionará el caso de una entidad bancaria que necesita que propongan un perfil de cliente que abrirá o no una cuenta de ahorro a plazo.

Valor: Crear reglas de perfilamiento a través del árbol de decisión para explicar el objetivo

- Taller: Entrenamiento y análisis de árboles de clasificación binaria basados en CART y C45 con Python.
- Fundamentos de evaluación de modelos de clasificación binaria. La matriz de confusión y errores tipo I y tipo II.
- Taller: Implementación y análisis de matrices de confusión para modelos entrenados.

Sesión 3

Técnicas de clasificación combinadas (Ensembles)

- Técnicas Ensembles. Definición, algoritmos homogéneos y heterogéneos, descripción de Bagging y Boosting.

Casos prácticos.

Caso de Uso: Aplicación de diferentes algoritmos ensembles y elección del mejor modelo de predicción.

Ejemplo: Se solucionará el caso de una entidad bancaria que necesita que prediga si un cliente abrirá o no una cuenta de ahorro a plazo a través de los distintos algoritmos.

Valor: Determinar el mejor modelo de predicción para tener una mejor predicción para la entidad bancaria.

- Taller: Entrenamiento y análisis de un modelo estándar de clasificación binaria basado en RandomForest con Python.
- Taller: Entrenamiento y análisis de un modelo estándar de clasificación binaria basado en AdaBoost con Python.
- Taller: Entrenamiento y análisis de un modelo estándar de clasificación binaria basado en GBM.
- Taller: Entrenamiento y análisis de un modelo estándar de clasificación binaria basado en XGBoost con Python.

Sesión 4

Evaluación de modelos de clasificación

- Métricas de evaluación de modelos: Método de cálculo e interpretación de F1-Score, Sensibilidad y Especificidad.
- Gráficas de evaluación de modelos: Interpretación de curva Precision-Recall y de curva ROC.
- Matriz de confusión y métricas de evaluación de clasificadores multiclas.

Casos prácticos.

Caso de Uso: Aplicación de métricas de machine learning y optimización de algoritmos para comparar y seleccionar el mejor modelo de predicción binario

Ejemplo: Se usarán los diferentes modelos estudiados y se optimizarán a través de Cross-validation para elegir el mejor modelo de predicción.

Valor: Tomar la mejor decisión entre los modelos de predicción para asegurar un alto performance del modelo.

Técnicas de optimización de modelos de clasificación

- Sobreajuste y subajuste de modelos de clasificación. Definición, métodos para su identificación.
- Taller: Tratamiento de datos desbalanceados. Impacto en el modelo.
- Taller: Muestreo basado en Cross-Validation. Impacto en el modelo.

Sesión 5

Fundamentos de Redes neuronales

- Principio de No linealidad en modelamiento de datos.
- De la neurona biológica a la neurona artificial: Arquitectura y componentes del perceptrón simple.
- Del perceptrón a la red neuronal. Arquitecturas (“The neural network Zoo”), y aplicaciones.

Casos prácticos.

Caso de Uso: Aplicación de algoritmos de redes neuronales y conocer los principales hiperparámetros a usar.

Ejemplo: Se solucionará el caso de predecir que una persona pueda abrir una cuenta de ahorro a plazo en una entidad bancaria a través de modelos con redes neuronales.

Valor: Identificar las bondades de usar modelos con redes neuronales.

- Taller: Implementación de un modelo de regresión basado en perceptrón simple con TensorFlow. Caso conversión de unidades.
- Taller: Implementación de un modelo de clasificación binaria construido con TensorFlow y Python.

V. SISTEMA DE EVALUACIÓN

La calificación de este MÓDULO se realizará considerando lo siguiente:

CRITERIO	PESO	FECHA DE EVALUACIÓN
Evaluación Final 100% 1. Evaluación Final : Entregable de Caso de Uso.	100%	SESION 5 Inicio [10/02/2026] Fin [17/02/2026]

Leyenda:

Fecha de inicio : Sesión de clase en vivo , donde se brinda las indicaciones del examen y el docente sube toda la información (data , formatos)en el aula virtual referente a la evaluación
 Fecha fin : Fecha máxima que los alumnos envían el entregable o finaliza el criterio de evaluación

Calificación de Diploma Total:

$$\text{Nota de Diploma} = \text{Promedio de Módulos } (N1, N2, \dots, N10) * 40\% + N \text{ Proyecto} * 60\%$$

Leyenda:

N : Nota

EVALUACIÓN FINAL

CORRESPONDIENTE A:

Programa	Diploma Data Scientist Edición 19
Módulo	Aprendizaje Supervisado y modelos Binarios
Objetivo	El módulo le permitirá al participante demostrar los conocimientos adquiridos durante las cinco sesiones tanto en lo teórico como práctico a través del desarrollo de un modelo analítico que ayude a predecir la probabilidad que una persona adquiera una tarjeta de crédito de una entidad bancaria.

CONSIDERACIONES GENERALES :

Modalidad	Individual
Medio y plazo de entrega	Aula virtual DMC, hasta el 17/02/2026 a las 23:59.
Archivos y herramientas	<ul style="list-style-type: none"> • El Caso a desarrollar se encontrará en el aula virtual de DMC junto con el dataset "db_venta_tc.csv"
Producto académico (entregables esperados).	El alumno deberá desarrollar un modelo analítico usando el dataset "db_venta_tc.csv" donde encontrará las variables predictoras y el target (adquiere la tarjeta de crédito), en este caso se le pedirá que use las herramientas estudiadas de las cinco sesiones, se le pedirá que realice análisis y conclusiones de los resultados y entregue la Notebook Final con su nombre.

INSTRUCCIONES PARA REALIZAR LA ACTIVIDAD

- 1) Utilizar el material de las cinco sesiones estudiadas, es importante para que sepan como iniciar la solución.
- 2) Se recomienda escuchar las grabaciones en caso les quedó alguna duda o no recuerden.
- 3) Es importante realizar el análisis de los resultados y culminar con las conclusiones (será muy esencial en la evaluación).

I. FUENTES DE INFORMACIÓN COMPLEMENTARIA

Las siguientes son fuentes de información que el docente le sugiere revisar para que pueda complementar los contenidos revisados en clase:

- Duda, R. O.; Hart, P. E.; Stork, D. G. (2001). Pattern Classification (2.a ed.). USA: John Wiley and Sons, Inc.
- Frank, A.; Asuncion, A. (2010). UCI Machine Learning Repository (Disponible online: <http://archive.ics.uci.edu/ml>). USA: University of California, School of Information and Computer Science.
- Python Software Foundation (2025). The Python Tutorial. Recuperado de <https://docs.python.org/3/tutorial/index.html>
- NumFOCUS Inc. (2024). Pandas documentation. Recuperado de <https://pandas.pydata.org/pandas-docs/stable/index.html>

- “Three ways to detect outliers” (Blog): <http://colingorrie.github.io/outlier-detection.html>
- Anomaly detection with Local Outlier Factor (LOF) http://scikit-learn.org/stable/auto_examples/neighbors/plot_lof.html
- How to Identify Outliers in your Data (Blog): <https://machinelearningmastery.com/how-to-identify-outliers-in-your-data/>