



**CTIC-UNI**  
BUSINESS SCHOOL

**UNIVERSIDAD NACIONAL DE  
INGENIERÍA**

**CENTRO DE TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIONES**

**I. DATOS GENERALES**

<b>1.1</b>	Nombre del Curso	MACHINE LEARNING CON PYHTON
<b>1.2</b>	Año calendario	2019
<b>1.3</b>	Grupo	“DEJAR EN BLANCO”
<b>1.4</b>	Crédito	2
<b>1.5</b>	Modalidad	Virtual
<b>1.6</b>	Docente	Dr. Manuel Castillo-Cara

**II. RESUMEN**

El curso de “Machine Learning con Python” se centra en un subcampo específico de aprendizaje automático llamado modelado predictivo. Este es el campo del aprendizaje automático que es el más útil en la industria y el cual se utilizar la librería de aprendizaje automático scikit-learn en Python por su gran rendimiento y facilidad en su uso.

A diferencia de las estadísticas, donde los modelos se usan para comprender los datos, el modelado predictivo se enfoca en el desarrollo de modelos que hacen las predicciones más precisas a expensas de explicar por qué se hacen las predicciones.

A diferencia del campo más amplio del aprendizaje automático que podría utilizarse con datos en cualquier formato, el modelado predictivo se centra principalmente en datos tabulares (por ejemplo, tablas de números como en una hoja de cálculo).

**III. RESULTADOS DE APRENDIZAJE**

1. Aplicar técnicas de análisis y visualización de datos en un conjunto de datos complejo para problemas de machine learning.
2. Aplicar técnicas de tratamiento de datos en un conjunto de datos para mejorar la robustez y métrica de salida de los diferentes algoritmos de machine learning.
3. Comprender los diferentes mecanismos y técnicas para aplicar analítica predictiva en problemas de machine learning e interpretar la salida obtenida por los modelos de predicción.
4. Comprender y analizar la fase del análisis de datos previos al modelado algorítmico en machine learning.
5. Realizar modelos algorítmicos robustos con una optimización de sus hiperparámetros para la fase de predicción
6. Desarrollar y analizar proyectos de machine learning como regresión, clasificación y multiclase.
7. Utilizar librerías específicas de Python como scikit-learn para trabajos de Machine Learning

IV. <b>CONTENIDOS</b>	
<b>MÓDULO I</b> Introducción.	<ul style="list-style-type: none"> <li>• Conceptos básicos de machine learning.</li> <li>• Jupyter Notebook como nuestro entorno de machine learning.</li> <li>• Curso rápido de Python.</li> </ul>
<b>MÓDULO II</b> Análisis de datos	<ul style="list-style-type: none"> <li>• Cargar un conjunto de datos.</li> <li>• Estadística descriptiva.</li> <li>• Visualización de datos.</li> <li>• Taller: Trabajo de aplicación de diferentes técnicas analíticas de datos en un conjunto de datos seleccionado por el usuario e interpretar la salida obtenida.</li> <li>• Examen tipo test sobre los contenidos del módulo.</li> </ul>
<b>MODULO III</b> Preprocesamiento de datos	<ul style="list-style-type: none"> <li>• Análisis exploratorio de datos.</li> <li>• Preprocesamiento de datos.</li> <li>• Métodos de remuestreo para estimar la precisión del modelo.</li> <li>• Taller: Trabajo de aplicación de diferentes técnicas de análisis y procesamiento de datos de datos en un conjunto de datos seleccionado por el usuario e interpretar la salida obtenida.</li> <li>• Examen tipo test sobre los contenidos del módulo.</li> </ul>
<b>MÓDULO IV</b> Fase de tratamiento de datos	<ul style="list-style-type: none"> <li>• Evaluación de las métricas.</li> <li>• Feature Selection.</li> <li>• Feature Importance.</li> <li>• Reducción de dimensiones en un dataset.</li> <li>• Taller: Aplicación de diferentes técnicas de tratamiento de datos en un conjunto de datos y verificación de su impacto en las métricas algorítmicas.</li> <li>• Examen tipo test sobre los contenidos del módulo.</li> </ul>
<b>MÓDULO V</b> Fase de modelado	<ul style="list-style-type: none"> <li>• Algoritmos de Machine Learning.</li> <li>• Rendimiento de los algoritmos.</li> <li>• Algoritmos Ensamblados</li> <li>• Taller: Aplicación de diferentes algoritmos de machine learning en un conjunto de datos e interpretar la salida obtenida, así mismo, verificar el algoritmo que tenga mejor comportamiento.</li> <li>• Examen tipo test sobre los contenidos del módulo.</li> </ul>

<b>MÓDULO VI</b> Fase de optimización y forecasting	<ul style="list-style-type: none"> <li>• Pipelines</li> <li>• Procesamiento de datos avanzado</li> <li>• Configuración de hiperparámetros</li> <li>• Guardado e integración del modelo</li> <li>• Taller: Una vez seleccionados los algoritmos candidatos a modelo realizar una optimización de estos a través de la configuración de sus hiperparámetros.</li> <li>• Examen tipo test sobre los contenidos del módulo.</li> </ul>
<b>MÓDULO VII</b> Proyectos de machine learning	<ul style="list-style-type: none"> <li>• Trabajar un proyecto de clasificación multiclase</li> <li>• Trabajar un proyecto de regresión.</li> <li>• Trabajar un proyecto de clasificación binaria.</li> <li>• Proyecto: Realizar un proyecto completo analizando todas las fases estudiadas en los diferentes módulos.</li> </ul>
<b>MÓDULO VIII</b> Aprendizaje No supervisado	<ul style="list-style-type: none"> <li>• Trabajar diferentes algoritmos ULAs</li> <li>• Trabajar proyectos de ULAs</li> <li>• Trabajar con PCA</li> <li>• Trabajar con ULAs en imágenes</li> </ul>

<b>V. ESTRATEGIAS DIDÁCTICAS</b>
<p>El curso se desarrolla considerando actividades no presenciales a través de la plataforma virtual, para lo cual se utilizan las siguientes estrategias didácticas:</p> <p><b>Actividad virtual</b></p> <ul style="list-style-type: none"> <li>• Sesiones de videoconferencias</li> <li>• Análisis de casos</li> <li>• Foros de discusión</li> <li>• Trabajos parciales de los módulos en foros</li> <li>• Examen tipo test.</li> <li>• Proyecto final del curso</li> <li>• Lecturas comentadas y</li> <li>• Búsquedas de información científica.</li> </ul>

<b>VI. EVALUACIÓN</b>
<p>La evaluación del aprendizaje será un proceso continuo donde los estudiantes tienen la oportunidad de ir trabajando con el lenguaje de programación Python practicando lo expuesto en la parte de teoría.</p> <p>Se tomará en cuenta una evaluación desde el módulo 3 al módulo 6 y en la que cada unidad tendrá dos notas:</p> <ul style="list-style-type: none"> <li>• Teoría: Tiene un ponderación del 30% de la nota final (el total de todas la unidades). Consistirá en examen tipo test online de los conceptos teóricos.</li> <li>• Taller: Tiene una ponderación del 20% de la nota final (el total de todas la unidades). Se pondrán a prueba los conceptos teóricos explicados en los videotutoriales de manera que el alumno trabajará un proyecto de cada unidad.</li> </ul>

Así mismo se realizará un:

- Proyecto: Tiene una ponderación del 50% de la nota final. Se realizará un proyecto final del curso el cual deberá poner a prueba todos los contenidos vistos en los diferentes módulos.

Para aprobar el curso se deberá obtener una calificación aprobatoria producto de la participación en las actividades virtuales y presenciales, y en cada uno de los resultados establecidos en el curso.

Escala de notas:

Nota	Nota Literal
17.00 – 20.00	A
15.00 – 16.99	B
13.00 – 14.99	C
11.00 – 12.99	D
< de 11.00	E (desaprobado)

## VII. CERTIFICACIÓN

La Universidad Cayetano Heredia como entidad Certificadora que oferta Actividades de Educación Continua, al finalizar el Curso podrá otorgar los siguientes tipos de certificación:

### 1. Certificado

- Para profesionales y bachilleres que hayan aprobado satisfactoriamente las actividades, cumplan con la participación mínima requerida y hayan aprobado con nota 11.00 como mínimo.

### 2. Constancia de participación

- Para estudiantes de pregrado que hayan aprobado satisfactoriamente las actividades, cumplan con la participación mínima requerida y hayan aprobado con nota 11.00 como mínimo.

## VIII. BIBLIOGRAFÍA

R in a Nutshell: A Desktop Quick Reference de Joseph Adler

Applied Predictive Modeling de Max Kuhn and Kjell Johnson

Machine Learning with R de Brett Lantz

Practical Data Science with R by Zumel Mount

An Introduction to Statistical Learning: with Applications in R de Gareth James

## IX. REQUISITOS

- Conocimientos intermedios de programación.
- Conocimientos básicos de estadística inferencial y cálculo de probabilidades.