



**UNIVERSIDAD PERUANA  
CAYETANO HEREDIA**

**ESCUELA DE POSTGRADO VICTOR ALZAMORA CASTRO  
UNIDAD DE SERVICIOS ACADÉMICOS**

## **SILABO**

### **I. DATOS GENERALES**

<b>1.1.</b>	Nombre del curso	<b>MACHINE LEARNING PARA INVESTIGACIÓN</b>
<b>1.2.</b>	Código	
<b>1.3.</b>	Año Calendario	2018
<b>1.5.</b>	Créditos	
<b>1.6.</b>	Tipo de curso	Virtual
<b>1.7.</b>	Prerrequisitos	<b>NINGUNO</b>
<b>1.8.</b>	No. Total de Horas	8 UNIDADES
<b>1.10.</b>	Profesor coordinador	Dr. Manuel Castillo cara (mcastillo@uni.edu.pe)

### **II. SUMILLA**

Este curso se enfoca en un subcampo específico de la minería de datos llamado modelado predictivo. Este es el campo de la minería de datos que es el más útil en la industria e investigación siendo estas técnicas las más potente y necesarias que un investigador debe tener en su fase de de desarrollo.

A diferencia de las diferentes formas de estadística, donde los modelos se utilizan para comprender los datos, el modelado predictivo se centra en el desarrollo de modelos que hacen las predicciones más precisas a expensas de explicar el por qué se hacen las predicciones. A diferencia del campo más amplio de minería de datos que podría usarse con datos en cualquier formato, el modelado predictivo se enfoca principalmente en datos tabulares (por ejemplo, tablas de números como una hoja de cálculo).

En este contexto, el curso pretende otorgar a los estudiantes los conceptos básicos e intermedios relacionados al análisis y tratamiento de datos pero llevando este proceso más allá pudiendo aplicar algoritmos basados en aprendizaje, es decir, Machine Learning. Para ello, el curso hará uso de un sistema muy utilizado en cualquier ámbito y línea de investigación como es Weka. Weka es una plataforma de muy sencillo uso que nos permite utilizar todos los conceptos de Minería de datos sin tener que saber programar, es decir, es una plataforma específicamente desarrollara para cualquier investigador que requiera de estas técnicas pero que no tiene un base previa computacional.

### **III. COMPETENCIA**

#### **RESULTADOS DE APRENDIZAJE**

1. Aprender las técnicas de preprocesamiento de datos para machine learning.
2. Comprender que es la minería de datos y aplicarla a un conjunto de datos específico.
3. Conocer las diferentes posibilidades sobre el pre-análisis y pre-tratamiento de datos para machine learning.
4. Comprender y analizar la fase de preprocesamiento en machine learning.

5. Comprender y analizar la fase del análisis de datos previos al modelado algorítmico en machine learning.
6. Comprender y analizar la fase de modelado algorítmico en machine learning.
7. Comprender y analizar la fase de tuning para los diferentes modelos de machine learning.
8. Desarrollar y analizar proyectos de machine learning como regresión, clasificación y multiclase.

#### **IV. CONTENIDOS**

##### **UNIDAD 1. Introducción**

Conceptos básicos de machine learning.  
Weka Workbench como nuestro entorno de machine learning.  
Conclusiones.

##### **UNIDAD 2. Minería de datos en Weka**

Paneles en Weka.  
Conociendo nuestros datos en los paneles de Weka.

##### **UNIDAD 3. Pre-análisis y pre-tratamiento de datos**

Clasificación de datos en machine learning.  
Conjunto de datos para machine learning.  
Pre-análisis de datos.

##### **UNIDAD 4. Pre-procesamiento de datos para machine learning**

Normalización y estandarización de los datos.  
Transformar los datos de machine learning.  
Manejar valores perdidos en los datos de machine learning.

##### **UNIDAD 5. Análisis de datos en machine learning**

Feature Selection en machine learning.  
Uso de algoritmos de machine learning.  
Estimar el resultado de los algoritmos.  
Estimar una línea base de los resultados.

##### **UNIDAD 6. Fase de modelado en machine learning**

Algoritmos de clasificación.  
Algoritmos de regresión.  
Algoritmos ensamblados.

##### **UNIDAD 7. Fase 'Tuning' en machine learning**

Comparar el rendimiento de los algoritmos.  
'Tunear' los parámetros (hiperparámetros) de los algoritmos.  
Guardar nuestros modelos y hacer predicciones.

##### **UNIDAD 8. Proyectos en machine learning**

Trabajar un proyecto de clasificación multiclase.  
Trabajar un proyecto de clasificación binario.  
Trabajar un proyecto de regresión.

#### **V. ESTRATEGIAS DIDÁCTICAS**

**Curso Virtual** el curso es virtual para poder llevarlo a cabo se tiene el siguiente esquema:

**1.- Lectura del material el EVD.** Los participantes debe leer y revisar los contenidos teóricos que se tienen en el EVD, correspondiente a cada unidad.

**2.- Visualización de Videos.** Como ayuda al aprendizaje se tienen videos para cada unidad, los cuales deben ser visualizados por el alumnado.

**3.- Autoaprendizaje.** Se debe resolver el material para refuerzo y aplicación de los contenidos teóricos/prácticos antes de la evaluación.

#### **Software de trabajo**

Se utilizará el software libre Weka Workbench.

#### **Foro de consultas**

Este espacio está destinado para que los estudiantes formulen sus preguntas con respecto a la temática desarrollada y el docente tutor será el responsable de absolver sus interrogantes.

## **VI. EVALUACIÓN**

La evaluación del aprendizaje será un proceso continuo donde los estudiantes tienen la oportunidad de ir trabajando con el software de machine learning practicando lo expuesto en la parte de teoría.

Se tomará en cuenta una evaluación de cada unidad y en la que cada unidad tendrá dos notas:

- Teoría: Tiene un ponderación del 50% de la nota de la unidad. Consistirá en examen tipo test online de los conceptos teóricos.
- Práctica: Tiene una ponderación del 50% de la nota de la unidad. Se pondrán a prueba los conceptos teóricos explicados en los videotutoriales de manera que el alumno trabajará un proyecto real.

Los exámenes se realizarán en 5 de las 8 unidades siendo la la nota final la nota media de estas. Asimismo, los alumnos deberán trabajar un proyecto consolidado con unos datos propios con todas las fases que lleva un problema de machine learning.

Se realizará según la escala de calificación que emplea la Escuela de Posgrado Víctor Alzamora Castro.

### **ESCALA DE NOTAS**

Calificativos y equivalencias en términos literales:

Nota	Nota Literal
17 – 20	A
15– 16	B
13– 14	C
11– 12	D
≤ de 10	E (desaprobado)

- La nota final aprobatoria del curso será de once (11.00). Se expresará en escala vigesimal con dos decimales, no habrá redondeo al valor inmediato superior o inferior.
- La nota de los estudiantes que no rindan la evaluación dentro del plazo establecido para cada unidad será cero (0).

## **VII. BIBLIOGRAFÍA**

1. Ian H. Witten, Eibe Frank, Mark A. Hall and Christopher J. Pal. Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems). 4ta Edición.
2. Cursos MOOCs gratuitos. <https://www.cs.waikato.ac.nz/ml/weka/courses.html>