

## Sumilla

Sitio: [manwest.milaulas.com](http://manwest.milaulas.com)

Curso: MACHINE LEARNING CON WEKA

Libro: Sumilla

Imprimido por: Administrador Usuario

Día: jueves, 21 de marzo de 2024, 10:38

## Tabla de contenidos

- 1. Datos generales**
- 2. Destinatarios**
- 3. Descripción**
- 4. Resultados de aprendizaje**
- 5. Contenidos**
- 6. Estrategias didácticas**
- 7. Docencia**
- 8. Bibliografía**

## 1. Datos generales

**Nombre del curso:** Machine learning para la investigación

**Año calendario:** 2024

**Tipo del curso:** Virtual

**Prerrequisitos:** Ninguno

**Unidades:** 8

**Duración:** 8 semanas

**Profesor:** Dr. Manuel Castillo-Cara

## 2. Destinatarios

El curso de Machine learning para la investigación está dirigido a personas que tengan pocos conocimientos de machine learning y quieran adentrarse a este apasionante mundo de dentro del campo de modelado predictivo.

Además, el curso está diseñado para que cualquier estudiante universitario, investigador o tecnólogo que se encuentre realizando o necesite realizar diferentes experimentos a través de grandes conjuntos de datos para poder sintetizarlos en alguna salida predictiva puedan utilizar los muy diferentes recursos de machine learning que nos pone a nuestra disposición el software Weka Workbench.

### 3. Descripción

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 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 desarrollada para cualquier investigador que requiera de estas técnicas pero que no tiene un base previa computacional.

## 4. Resultados de aprendizaje

- Aprender las técnicas de preprocesamiento de datos para machine learning.
- Comprender que es la minería de datos y aplicarla a un conjunto de datos específico.
- Conocer las diferentes posibilidades sobre el pre-análisis y pre-tratamiento de datos para machine learning.
- Comprender y analizar la fase de preprocesamiento en machine learning.
- Comprender y analizar la fase del análisis de datos previos al modelado algorítmico en machine learning.
- Comprender y analizar la fase de modelado algorítmico en machine learning.
- Comprender y analizar la fase de tunning para los diferentes modelos de machine learning.
- Desarrollar y analizar proyectos de machine learning como regresión, clasificación y multiclase.

## 5. Contenidos

<b>Unidad 1: Introducción</b>	<ol style="list-style-type: none"><li>1. Conceptos básicos de machine learning.</li><li>2. Weka Workbench como nuestro entorno de machine learning.</li><li>3. Conclusiones.</li></ol>
<b>Unidad 2: Minería de datos en Weka</b>	<ol style="list-style-type: none"><li>1. Paneles en Weka.</li><li>2. Conociendo nuestros datos en los paneles de Weka.</li><li>3. Conclusiones.</li></ol>
<b>Unidad 3: Pre-análisis y pre-tratamiento de datos</b>	<ol style="list-style-type: none"><li>1. Clasificación de datos en machine learning.</li><li>2. Conjunto de datos para machine learning.</li><li>3. Pre-análisis de datos.</li><li>4. Conclusiones.</li></ol>
<b>Unidad 4: Pre-procesamiento de datos para machine learning</b>	<ol style="list-style-type: none"><li>1. Normalización y estandarización de los datos.</li><li>2. Transformar los datos de machine learning.</li><li>3. Manejar valores perdidos en los datos de machine learning.</li><li>4. Conclusiones.</li></ol>
<b>Unidad 5: Análisis de datos en machine learning</b>	<ol style="list-style-type: none"><li>1. Feature Selection en machine learning.</li><li>2. Uso de algoritmos de machine learning.</li><li>3. Estimar el resultado de los algoritmos.</li><li>4. Estimar una línea base de los resultados.</li><li>5. Conclusiones.</li></ol>
<b>Unidad 6: Fase de modelado en machine learning</b>	<ol style="list-style-type: none"><li>1. Algoritmos de clasificación.</li><li>2. Algoritmos de regresión.</li><li>3. Algoritmos ensamblados.</li><li>4. Conclusiones.</li></ol>
<b>Unidad 7: Fase 'Tuning' en machine learning</b>	<ol style="list-style-type: none"><li>1. Comparar el rendimiento de los algoritmos.</li><li>2. 'Tunear' los parámetros (hiperparámetros) de los algoritmos.</li><li>3. Guardar nuestros modelos y hacer predicciones.</li><li>4. Conclusiones.</li></ol>
<b>Unidad 8: Proyectos en machine learning</b>	<ol style="list-style-type: none"><li>1. Trabajar un proyecto de clasificación multiclase.</li><li>2. Trabajar un proyecto de clasificación binario.</li><li>3. Trabajar un proyecto de regresión.</li><li>4. Conclusiones.</li></ol>

## 6. Estrategias didácticas

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

1. Lectura del material del campus virtual. Los participantes debe leer y revisar los contenidos teóricos que se tienen en el campus virtual, correspondiente a cada unidad.
2. Visualización de Vídeos. 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.



## 7. Docencia

Manuel Castillo Cara es ingeniero en informática, y doctor en Tecnologías Informáticas Avanzadas por la Universidad de Castilla-La Mancha (UCLM) en 2018 (con la calificación de sobresaliente Cum Laude). Respecto a la trayectoria académica, entre 2014 y 2020, tuvo una trayectoria como profesor investigador en Perú, principalmente en la Universidad Nacional de Ingeniería (UNI) y la Universidad de Lima (ULima). Posteriormente, entre 2021 y 2023 consigue un contrato postdoctoral Recualificación - María Zambrano en la Universidad Politécnica de Madrid (UPM) para el retorno de doctores a España. Desde 2023 a día de hoy se encuentra en la Universidad Nacional de Educación a Distancia (UNED), actualmente como PDI en categoría de Profesor Ayudante Doctor. Actualmente cuenta con la acreditación "Profesor Titular" por parte de la ANECA.

Durante su etapa docente ha participado en más de 15 de investigación de los cuales 7 de ellos son de convocatorias competitivas nacionales. De estos proyectos, ha sido Investigador Principal en 2 de ellos. Así mismo, ha sido miembro de equipo en un proyecto de desarrollo social que consistía en desarrollar destrezas tecnológicas a niños de zonas deprimidas de Perú, habiendo tenido más de 2000 alumnos entre 8 y 15 años en total. Los cursos eran en el ámbito de Robótica educativa, programación con Scratch, entre otros. Por otro lado ha obtenido reconocimientos docentes como profesor con mejor puntuación en las encuestas académicas en la UNI en 4 ocasiones. Sus líneas de investigación se encuentran enfocados en Inteligencia Artificial y arquitecturas de computación dentro de las áreas temáticas de redes de sensores (WSN), plataformas y arquitecturas distribuidas como Fog Computing, análisis/tratamiento de datos, reconocimiento de patrones y Visión por Computador. Así mismo, es desarrollador principal de la librería TINTOlib para el desarrollo de las Redes Neuronales Híbridas e implementó el método para convertir datos tabulares en imágenes sintéticas llamado TINTO. Ha impartido varios seminarios sobre IA y el uso de TINTOlib en Perú y España.

Se resumen a continuación sus indicadores generales de calidad de la producción científica.

- 17 artículos en revista de impacto indexados (5 en Q1 y 9 en Q2)
- 7 congresos internacionales (4 en Scopus)
- 2 competencias en robótica IEEE logrando el tercer puesto
- 1 patente modelo de utilidad
- 2 registros de derechos de autor de software
- 1 libro académico
- Impacto WOS (JCR): Citas totales: 139 / Publicaciones Q1: 5 / Índice H: 7
- Impacto Scopus: Citas totales: 237 / Publicaciones Q1: 11 / Índice H: 8
- Impacto Scholar: Citas totales: 389 / Índice H: 9
- 4 Trabajos Fin de Máster
- 6 Trabajos Fin de Grado

## 8. Bibliografía

- 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.
- Cursos MOOCs gratuitos. <https://www.cs.waikato.ac.nz/ml/weka/courses.html>