

Tutoría Colectiva Inicial

Minería de Datos

Presentación



Prof. Dr. Aurelio López Fernández

aurelio.lopez@professor.universidadviu.com

Formación:

- 2015: Ingeniero Informático (Universidad Pablo de Olavide)
- 2017: Máster Profesorado (MAES) (Universidad Internacional de Valencia)
- 2018: Máster Oficial Ingeniería Informática (Universidad Pablo de Olavide)
- 2022: Doctor cum-laude Ingeniería Informática, especialidad Big Data y Ciencia de datos (Universidad Pablo de Olavide)

Presentación



Prof. Dr. Aurelio López Fernández

aurelio.lopez@professor.universidadviu.com

Experiencia profesional:

- 2023 - Actual: Profesor universitario e investigador en la Università Ca' Foscari, Venecia, Italia.
- 2023 - Actual: Profesor universitario, Universidad Nacional de Asunción, Paraguay
- 2022 - Actual: Profesor universitario, Universidad Internacional de Valencia, España
- 2019 - Actual: Profesor universitario, Universidad Pablo de Olavide, España
- 2015 - Actual: Investigador en el grupo de investigación "Intelligent Data Analysis" de la Universidad Pablo de Olavide.

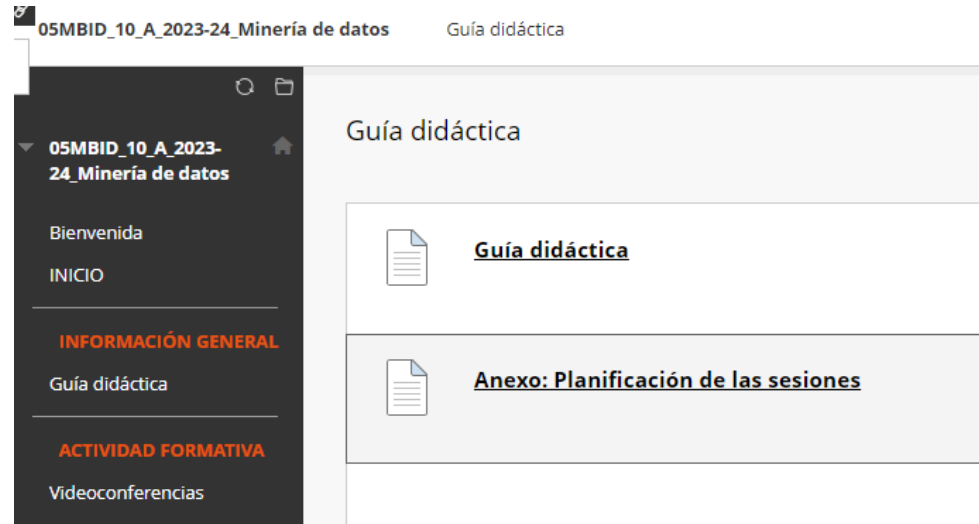
Líneas de investigación: Big Data, High-Performance Computing, Data mining, Bioinformatics y Biomedical & biological computing.

- 2015 - 2016: Profesor universitario, Universidad Nacional del Sur, Buenos Aires, Argentina.

1. Introducción
2. Guía didáctica
3. Calendario
4. Temario.
5. Sistema de evaluación
6. Portafolio
7. Examen
8. Entregas tardías
9. Honestidad
10. Recomendaciones
11. Repaso por el aula

El segundo módulo, tiene como objetivo el desarrollo del alumno en la parte analítica de la ciencia de datos, enfocándose a la parte de tratado, limpieza y entendimiento de datos, aplicación algorítmica y visualización. Este módulo es el de mayor extensión, abarcando un total de 24 ECTS distribuidos en 4 asignaturas de 6 ECTS cada una:

1. Minería de Datos
2. Estadística avanzada
3. Machine Learning
4. Visualización de Datos



CG.1.- Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

CG.2.- Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.

CG.3.- Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

CG.4.- Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Todas las horas que aparezcan en esta presentación y durante el resto de la asignatura se encuentran en la franja horaria de Europa Central, correspondiente a la hora actual en España:

(GMT +1)

Horario: 18:00 – 20:00h

Sesión	Fecha	Contenido/Tema
SESIÓN 1	17/11/2023	05MBID – TC1 - Tutoría Inicial
SESIÓN 2	21/11/2023	05MBID – VC1 - Tema 1
SESIÓN 3	23/11/2023	05MBID – AG1 - Actividad Guiada 1
SESIÓN 4	28/11/2023	05MBID – VC2 - Tema 2
SESIÓN 5	30/11/2023	05MBID – VC3 - Tema 3
SESIÓN 6	05/12/2023	05MBID – AG2 - Actividad Guiada 2
SESIÓN 7	07/12/2023	05MBID – VC4 - Seminario experto
SESIÓN 8	12/12/2023	05MBID – VC5 - Tema 4
SESIÓN 9	14/12/2023	05MBID – VC6 - Tema 5
SESIÓN 10	19/12/2023	05MBID – VC7 - Tema 6
SESIÓN 11	21/12/2023	05MBID – VC8 - Tema 7
SESIÓN 12	09/01/2024	05MBID – VC9 - Repaso de la asignatura
SESIÓN 13	11/01/2024	05MBID – TC2 - Tutoría Colectiva

- Tema 1: Introducción al proceso KDD.
- Tema 2: Representación gráficas de variables.
- Tema 3: Análisis y transformación de variables.
- Tema 4: Técnicas de agrupamiento.
- Tema 5: Modelos predictivos.
- Tema 6: Comparación, selección y evaluación de modelos.
- Tema 7: Aplicaciones relacionadas con KDD.

Sistema de Evaluación	Ponderación
Portafolio*	60%
Actividad Evaluable 1 (*)	10
Actividad Evaluable 2 (*)	30
Participación en Foro (**)	10
Actividad Evaluativa Continua (**)	10
Sistema de Evaluación	Ponderación
Prueba final*	40 %
<i>La primera parte es un examen tipo test y uno o dos ejercicios prácticos.</i>	

* Es requisito indispensable para superar la asignatura aprobar cada apartado.

** Sólo disponible en primera convocatoria

Sistema de Evaluación	Ponderación
Portafolio*	60%
Actividad Evaluable 1 (*)	10
Actividad Evaluable 2 (*)	30
Participación en Foro (**)	10
Actividad Evaluativa Continua (**)	10
Sistema de Evaluación	Ponderación
Prueba final*	40 %
La primera parte es un examen tipo test y uno o dos ejercicios prácticos.	

Actividad Guiada 1:

Fecha límite de entrega 1ª Convocatoria:
30 de Noviembre de 2023 hasta las 23.59h

Fecha límite de entrega 2ª Convocatoria:
27 de Marzo de 2023 hasta las 23.59h

Ponderación: 10%

¡OBLIGATORIO APROBAR!

En esta actividad se debe hacer una propuesta de un proceso KDD hasta el preprocesamiento y transformación de datos incluido, siguiendo las siguientes pautas:

- **1. Planteamiento de problema y objetivo a alcanzar.** Se debe definir el contexto y las necesidades que llevan a la realización del proceso KDD. Además, se debe describir el problema que se pretende resolver y la necesidad a la que se busca responder. Es por ello por lo que hay que explicar el objetivo del análisis de datos que se espera afrontar en este proyecto.
- **2. Elección de la base de datos y selección de datos útiles.** Hay que seleccionar una base de datos que no provenga de repositorios estándar de KDD (como UCI ML Repository). La base de datos escogida debe respetar el tamaño, formato, contenido y las leyes de protección de datos adecuados. En este punto se debe explicar cómo se han filtrado y seleccionado los datos más relevantes para el problema y objetivo planteados en el anterior punto y hay que justificar la exclusión de los datos que no se consideran útiles.
- **3. Preprocesamiento y transformación de datos.** Se debe describir con detalle que técnicas y métodos se aplicarían para realizar una limpieza y transformación de datos. Además, se debe incluir una discusión sobre por qué y cómo este preprocesamiento y transformación de datos se alinean con el objetivo del proyecto. En esta actividad no es necesaria la implementación de código fuente sobre este proceso.

Este trabajo no solo debe ser visto como una actividad puntual, sino también como una posible base para proyectos futuros en otras asignaturas y para el Trabajo de Fin de Máster (TFM).

Se valorará especialmente la originalidad, la exactitud y la lógica en la descripción de la base de datos seleccionada y de cada una de las fases del proceso KDD que se realizaría para lograr los objetivos planteados. El documento final deberá tener una extensión de aproximadamente 4 páginas y seguir la estructura sugerida anteriormente.

Sistema de Evaluación	Ponderación
Portafolio*	60%
Actividad Evaluable 1 (*)	10
Actividad Evaluable 2 (*)	30
Participación en Foro (**)	10
Actividad Evaluativa Continua (**)	10
Sistema de Evaluación	Ponderación
Prueba final*	40 %
La primera parte es un examen tipo test y uno o dos ejercicios prácticos.	

Actividad Guiada 2:

Fecha límite de entrega 1ª Convocatoria:
17 de Enero de 2024 hasta las 23.59h

Fecha límite de entrega 2ª Convocatoria:
27 de Marzo de 2023 hasta las 23.59h

Ponderación: 30%

¡OBLIGATORIO APROBAR!

En esta actividad se debe completar la propuesta del proceso KDD que se comenzó en la Actividad Guiada 1, siguiendo las siguientes pautas:

4. Modelo de Minería de Datos. Se debe indicar de forma precisa cuál será el algoritmo de Minería de Datos más adecuado para resolver el problema que el alumno pretende resolver. Mencionar cuáles adaptaciones en los datos se necesitan para construir el modelo y de qué modo ejecutaría el algoritmo de entrenamiento correspondiente. No es necesario implementarlo, sólo describirlo claramente.

5. Discusión e interpretación de los resultados. Describir cómo se podría evaluar el desempeño del modelo, qué tipo de patrones son los que se obtienen del modelo y cómo se los podría utilizar para tomar decisiones.

En el caso que el estudiante haya entregado la Actividad Guiada 1, el estudiante deberá también corregir los puntos correspondientes a dicha actividad basándose en la retroalimentación (feedback) que el equipo docente suministre. En caso contrario, se recomienda al estudiante que elabore los puntos correspondientes a la AG1 antes de proceder con los puntos de la documentación de esta actividad. En definitiva, en esta tarea deberá **enviarse la documentación del proceso KDD al completo (Actividad Guiada 1 + pautas pertenecientes a la Actividad Guiada 2).**

Por otra parte, se deberá entregar un **Jupyter notebook (archivo .ipynb) que incluya la implementación de todo lo propuesto en la Actividad Guiada 1, junto con los datos necesarios para ejecutarlo (archivo .csv).**

Este trabajo no solo debe ser visto como una actividad puntual, sino también como una posible base para proyectos futuros en otras asignaturas y para el Trabajo de Fin de Máster (TFM).

Se valorará especialmente la originalidad, la exactitud y la coherencia entre el problema/objetivo planteado y el modelo a construir que el alumno haya elegido. El documento final deberá tener una extensión de aproximadamente 2 páginas y seguir la estructura sugerida anteriormente.

Sistema de Evaluación	Ponderación
Portafolio*	60%
Actividad Evaluable 1 (*)	10
Actividad Evaluable 2 (*)	30
Participación en Foro (**)	10
Actividad Evaluativa Continua (**)	10
Sistema de Evaluación	Ponderación
Prueba final*	40 %
La primera parte es un examen tipo test y uno o dos ejercicios prácticos.	

Foro:

Fecha límite de entrega 1ª Convocatoria:
17 de Enero de 2024 hasta las 23.59h

Fecha límite de entrega 2ª Convocatoria:
27 de Marzo de 2023 hasta las 23.59h

Ponderación: 10%

¡SÓLO DISPONIBLE EN 1ª CONVOCATORIA!



Eliminar

<input type="checkbox"/>	FORO	DESCRIPCIÓN	PUBLICACIONES TOTALES	PUBLICACIONES NO LEÍDAS	RESPUESTAS PARA MÍ NO LEÍDAS	PARTICIPANTES TOTALES
<input type="checkbox"/>	Foro de debate (Evaluable)	En este foro se debatirá sobre algunas cuestiones planteadas por el profesor/a de la asignatura.	0	0	0	0
<input type="checkbox"/>	Foro de revisión bibliográfica	Foro dedicado a la lectura y revisión de textos de profundización en la materia impartida.	0	0	0	0
<input type="checkbox"/>	Dudas sobre los materiales docentes	En este foro el alumnado puede plantear sus dudas acerca de el material teórico de la asignatura.	0	0	0	0
<input type="checkbox"/>	Dudas sobre las actividades formativas	Foro dedicado a dudas sobre de las actividades guiadas prácticas, las actividades de debate, la evaluación, guía docente o calendarios.	0	0	0	0
<input type="checkbox"/>	Miscelánea	Espacio libre para el alumnado.	0	0	0	0



Eliminar

Para la evaluación de esta sección se permiten publicar temas relacionados con:

- Debate y mejoras sobre ejercicios prácticos realizados en clase.
- Otros ejercicios prácticos que aporten mayor conocimiento sobre el temario de la asignatura.
- Opinión sobre noticias relevantes sobre la materia de minería de datos.
- Elaboración de un estudio sobre técnicas más utilizadas en diferentes tipos de aplicaciones.
- Opiniones y otras cuestiones relevantes sobre el seminario experto que se imparte en la asignatura.

- Se valorará la calidad de las contribuciones y se primará más la cantidad que la calidad.
- Se valorará el número de nuevas aportaciones.
- Responder a temas ya abiertos en el foro por otros compañeros/as será valorado.
- Las publicaciones deben ser originales y resumidas (no más de 150 palabras).
- Se debe evitar el repetir comentarios de otros compañeros/as.
- No forzar la participación ni “eludirla”.

¡IMPORTANTE!

Para la obtención de la calificación en el Foro, todas las contribuciones realizadas en el foro deberán guardarse, con capturas de pantalla, en formato PDF. Este PDF deberá enviarse a una Actividad destinada a ello.

05MBID_10_A_2023-24_Minería de datos Actividades

El modo de edición está: DESACTIVADO

Actividades

AG1 - Actividad guiada 1 (1ª Convocatoria)
Archivos adjuntos: 05MBID_ActividadEvaluableA1.pdf (331,846 KB), Guía.pdf (200,559 KB)

AG2 - Actividad guiada 2 (1ª Convocatoria)
Archivos adjuntos: Guía.pdf (200,559 KB)

Foro (1ª Convocatoria)

Sistema de Evaluación	Ponderación
Portafolio*	60%
Actividad Evaluable 1 (*)	10
Actividad Evaluable 2 (*)	30
Participación en Foro (**)	10
Actividad Evaluativa Continua (**)	10
Sistema de Evaluación	Ponderación
Prueba final*	40 %
La primera parte es un examen tipo test y uno o dos ejercicios prácticos.	

Actividad Evaluativa Continua:

Fecha límite de entrega 1ª Convocatoria:
17 de Enero de 2024 hasta las 23.59h

Fecha límite de entrega 2ª Convocatoria:
27 de Marzo de 2023 hasta las 23.59h

Ponderación: 10%

¡SÓLO DISPONIBLE EN 1ª CONVOCATORIA!

- Una única oportunidad.
- Tipo de examen: Tipo test.
- Número de preguntas: 10.
- Las respuestas erróneas restan.
- Cada pregunta tiene tan sólo una respuesta correcta.
- Basado en el video Recursos y materiales -> 02.Videos de la asignatura -> Video docente 1: José Ángel Olivas Varela



Sistema de Evaluación	Ponderación
Portafolio*	60%
Actividad Evaluable 1 (*)	10
Actividad Evaluable 2 (*)	30
Participación en Foro (**)	10
Actividad Evaluativa Continua (**)	10
Sistema de Evaluación	Ponderación
Prueba final*	40 %
La primera parte es un examen tipo test y uno o dos ejercicios prácticos.	

Examen:**Fecha 1ª Convocatoria:**

Franja A: 18 de Enero de 2024 de 12.00h a 13:45h

Franja B: 18 de Enero de 2024 de 20.00h a 21:45h

Fecha 2ª Convocatoria:

Franja A: 28 de Marzo de 2024 de 12.00h a 13:45h

Franja B: 28 de Marzo de 2024 de 20.00h a 21:45h

Ponderación: 40%**¡OBLIGATORIO APROBAR!**

- 20 preguntas de tipo test aproximadamente.
- 1 o 2 preguntas de desarrollo.
- Duración: 1 hora y 30 minutos (+15 minutos extra).
- Tener conectada la webcam y activada la monitorización (SMOWL).
 - Si no lo tenéis activado por favor mandar un email al servicio técnico de la VIU.



- ✓ Obligatorio
- ✓ Realizar pasos indicados, cualquier problema contactar con soporte
- ✓ Hacedlo con tiempo

Primera convocatoria:

En este caso, la entrega se ignora y su calificación sería de **0**. El alumno puede presentar el mismo trabajo, si así lo desea, en segunda convocatoria, en cuyo caso se corregirá de forma normal sin penalización.

Segunda convocatoria:

Se ignorará la entrega y se considerará la actividad como **No presentada**.

- Todos los **trabajos** deben hacerse de manera **individual** e independiente, aunque se permite y fomenta la participación y colaboración.
- Si se utilizan **recursos** externos (libros, páginas web,...) **deben citarse** claramente.
- El **plagio** descalifica.



Repaso por el Aula: Material del profesor

05MBID_10_A_2023-24_Minería de datos Recursos y materiales 03. Materiales del profesor

05MBID_10_A_2023-24_Minería de datos

Bienvenida

INICIO

Calendario

INFORMACIÓN GENERAL

Guía didáctica

ACTIVIDAD FORMATIVA

Videoconferencias

Recursos y materiales

Actividades

Mis calificaciones

Grupos

COMUNICACIÓN

Anuncios

Foro

Enviar e-mail

SERVICIOS

Laboratorios Virtuales

Smowl Proctoring Tool

03. Materiales del profesor

Desarrollar contenido Evaluaciones Herramientas Contenido de cola

Tema 1. Introducción al proceso KDD

Tema 2. Representación gráficas de variables

Tema 3. Análisis y transformación de variables

Tema 4. Técnicas de agrupamiento

Tema 5. Modelos predictivos

Tema 6. Comparación, selección y evaluación de modelos

Tema 7. Aplicaciones relacionadas con KDD

Otros recursos

05MBID_10_A_2023-24_Minería de datos Recursos y materiales 03. Materiales del profesor Otros recursos

Otros recursos

Desarrollar contenido Evaluaciones Herramientas Contenido de colaborador

- [Machine Learning CheatSheet](#)
- [El proceso KDD \(Knowledge Discovery in Databases\)](#)
- [Big Data: Preprocesamiento y calidad de datos](#)
- [Introduction to Statistical Learning Theory](#)
- [CRISP-DM Step-by-step data mining guide](#)
- [Data Warehousing and Knowledge Discovery](#)
- [RankEval: Open Tool for Evaluation of Machine-Learned Ranking](#)
- [Google Colab: Python y Machine Learning en la nube](#)

Dentro del mundo del **Data Science**, existen iniciativas muy interesantes, y una de las que más no puede interesar, además c

05MBID_10_A_2023-24_Minería de datos Recursos y materiales 02. Vídeos de la asignatura

05MBID_10_A_2023-24_Minería de datos

- Bienvenida
- INICIO
- Calendario
- INFORMACIÓN GENERAL
- Guía didáctica
- ACTIVIDAD FORMATIVA
- Videoconferencias
- Recursos y materiales
- Actividades
- Mis calificaciones
- Grupos
- COMUNICACIÓN
- Anuncios
- Foro

02. Vídeos de la asignatura

Desarrollar contenido Evaluaciones Herramientas Contenido de colaborador

Video docente 1: José Ángel Olivas Varela

Habilitado: Seguimiento de estadísticas

05MBID_JAOlivas_consultor.mp4 (51:54)

Minería de datos

José Ángel Olivas Varela

Máster Universitario en Big Data y Ciencia de Datos

Video docente 2: Jupyter Notebook for Data Science

Publishing, Packt. *Jupyter Notebook for Data Science*. PACKT Publishing, 2018, video.alexanderstreet.com/watch/jupyter-notebook-for-data-science. Accessed 17 May 2021.

¿Preguntas?



Gracias