

Asignatura

Nombre Asignatura	Minería de Datos I
Código	600522038
Créditos ECTS	6.0

Plan/es donde se imparte

Titulación	Doble Grado en ADE y Análisis de Negocios (Plan 2022)
Carácter	OBLIGATORIA
Curso	3

Datos Generales

➤ PROFESORADO

Gabriel Antonio Valverde Castilla, Doctor en Ingeniería Matemática, Estadística e Investigación Operativa, CDO de Celering, Socio Fundador de Datais

➤ CONOCIMIENTOS RECOMENDADOS

El alumno aporta los conocimientos previos de R y Python, así como de análisis exploratorio de datos, técnicas de visualización, gestión de bases de datos, además de los conocimientos aportados en las asignaturas de matemáticas, estadística, econometría métodos de decisión del plan de estudios.

Fundamentos de Estadística y Matemáticas:

Antes de adentrarse en Data Mining, es fundamental que los estudiantes tengan una comprensión sólida de conceptos estadísticos básicos, como la media, la mediana, la desviación estándar y la distribución de probabilidad. También deben estar cómodos con matemáticas, especialmente álgebra y cálculo básico. Estos conocimientos les ayudarán a comprender los algoritmos y las métricas utilizadas en Data Mining.

Programación y Manejo de Datos:

Los estudiantes deben tener experiencia previa en programación, preferiblemente en un lenguaje como Python o R, ya que la implementación de algoritmos de Data Mining a menudo implica escribir código. Además, deben estar familiarizados con la manipulación de datos, incluyendo la carga, limpieza y transformación de conjuntos de datos.

Conceptos de Bases de Datos y SQL:

Tener una comprensión básica de bases de datos y SQL (Structured Query Language) es beneficioso, ya que gran parte de la minería de datos implica trabajar con bases de datos para extraer y analizar información. Los estudiantes deben saber cómo realizar consultas SQL simples para recuperar datos de manera eficiente.

➤ OBJETIVOS TEÓRICOS

Fundamentos del Data Mining No Supervisado: Comprender los principios y teorías detrás de las técnicas de minería de datos no supervisados. Modelos y Algoritmos de Clustering y Reducción de Dimensionalidad: Estudiar en profundidad modelos como K-Means, DBSCAN, PCA, y otros. Evaluación y Selección de Modelos: Aprender a evaluar la efectividad de diferentes modelos de

minería de datos y seleccionar el más adecuado según los objetivos y la naturaleza de los datos.

➤ OBJETIVOS PRÁCTICOS

Los trabajos que los estudiantes van a realizar abarcan distintos aspectos de Data Mining, como Análisis Exploratorio de Datos (EDA), Reducción de Dimensionalidad, Clustering y una tarea global que puede incluir aspectos más amplios de minería de datos, aquí te sugiero algunas recomendaciones específicas previas para cada uno de estos trabajos:

Trabajo de Análisis Exploratorio de Datos (EDA):

Estadísticas Descriptivas: Los estudiantes deben estar familiarizados con estadísticas descriptivas, como media, mediana, desviación estándar y percentiles, para resumir y entender la distribución de los datos.

Visualización de Datos: Deben conocer las herramientas y técnicas para crear gráficos y visualizaciones efectivas que ayuden a comprender los patrones y tendencias en los datos.

Trabajo de Reducción de Dimensionalidad:

Álgebra Lineal: Una comprensión de álgebra lineal, especialmente de conceptos como vectores y matrices, es esencial para comprender los métodos de reducción dimensional como PCA.

Conocimiento de PCA: Deben tener un conocimiento sólido de cómo funciona el Análisis de Componentes Principales (PCA) y cómo aplicarlo para reducir la dimensionalidad de los datos.

Trabajo de Clustering:

Conceptos de Clustering: Los estudiantes deben comprender los conceptos de clustering y cómo funcionan algoritmos como K-Means, DBSCAN y clustering jerárquico.

Evaluación de Clustering: Deben estar familiarizados con las métricas de evaluación de clustering, como la silueta y el índice Davies-Bouldin, para determinar la calidad de los clústeres.

Trabajo Global de Minería de Datos:

Proceso de Minería de Datos: Los estudiantes deben tener una comprensión general del proceso completo de minería de datos, desde la preparación de datos hasta la interpretación de resultados.

Selección de Técnicas: Deben ser capaces de seleccionar las técnicas apropiadas de EDA, reducción de dimensionalidad y clustering según los objetivos del proyecto global.

Comunicación de Resultados: Deben saber cómo presentar de manera efectiva los resultados y conclusiones de su trabajo global, utilizando visualizaciones y explicaciones claras.

➤ COMPETENCIAS QUE SE DESARROLLAN / RESULTADOS DE APRENDIZAJE

- **Competencias Básicas y Generales**

- CB2 Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
- CB3 Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
- CB4 Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
- CG1 Resolver problemas de análisis de negocio en función del contexto y de los factores y variables más relevantes.
- CG2 Analizar y sintetizar la información, hipótesis y variables más importantes de un libro, un tema, un artículo, un caso, etc.

Competencias Específicas

- CE2 Comprender las herramientas estadísticas y econométricas para el análisis de variables económicas y empresariales.
- CE6 Saber manejar herramientas cuantitativas e informáticas para la toma de decisiones
- CE7 Conocer los fundamentos de la estadística multivariante y data mining, así como su aplicación al análisis de negocios.
- CE11 Capacidad para manipular y a modelizar patrones con el uso de la inteligencia espacial a aplicada a negocio. (geolocalización).
- CE19 Adquirir un enfoque de gestión analítico y orientado al dato, ayudar a las compañías en su transformación data-driven. Así como el alumno adquirirá la capacidad para conocer y aplicar los principios fundamentales y técnicas básicas de los sistemas inteligentes y su aplicación práctica

Competencias Transversales

- CT1 Desarrollar el pensamiento crítico propio del espíritu universitario, así como la capacidad de analizar, argumentar e interpretar datos relevantes y complejos para poder integrarlos de manera sólida y solvente en la toma de decisiones.
- CT2 Utilizar con rigor y precisión el lenguaje oral y escrito, siendo capaz de transmitir información a un público tanto especializado como no especializado, teniendo en cuenta los diferentes contextos.

➤ CONTENIDO DEL PROGRAMA

1. Introducción al machine learning
2. Diferencias entre modelos supervisados y no supervisados
3. Aplicaciones de los modelos no supervisados: casos de éxito
4. Preprocesamiento e ingeniería de variables
5. Algoritmos de reducción de dimensión 1. El análisis de componentes principales PCA,
6. Kernel PCA, introducción a autoencoder.
7. Algoritmos de reducción de dimensión 2. T-distributed Stochastic Neighbor Embedding, UMAP, SOM
8. Algoritmos de segmentación 1. Cluster jerárquico
9. Algoritmos de segmentación 2. Cluster de k- medias
10. Algoritmos de segmentación 3. PAM cluster
11. Algoritmos de segmentación 4. Distribution-based clustering, EM
12. Algoritmos de segmentación 5. Density-based clustering
13. Singular value decomposition (SVD)
14. Association rule - Predictive Analytics
15. Criterios de evaluación de los modelos
16. Resumen

➤ **ACTIVIDADES FORMATIVAS**

- **Introducción Teórica:** Comienza con los fundamentos de data mining no supervisado y las métricas de evaluación.
- **Herramientas y Lenguajes:** Python, scikit-learn, colab, github a partir de ejemplos de código construir.
- **Conjuntos de Datos:** desafiantes que tocan distintos casos de uso del ámbito financiero, empresarial.
- **Técnicas de Clustering:** Enseña algoritmos populares y cómo evaluar clústeres.
- **Reducción de Dimensionalidad:** Explora PCA, t-SNE y su uso en visualización.
- **Casos de Estudio y Proyectos:** Ejemplos del mundo real y proyectos prácticos.
- **Evaluación y Retroalimentación:** Realiza evaluaciones y ofrece retroalimentación detallada, de cada práctica.
- **Recursos Adicionales:** Se irán sugiriendo distintos canales informativos.
- **Comunidad y Colaboración:** Buscamos la colaboración entre estudiantes y la participación en comunidades en línea.
- **Actualización Continua:** Mantén el contenido actualizado para reflejar avances en data mining.

➤ **CRITERIOS Y MÉTODOS DE EVALUACIÓN**

Convocatoria Ordinaria 1a Matricula

Los criterios que se aplican son los siguientes:

- Evaluación Continua: 60%

- Asistencia y participación activa: 10%
- Actividades individuales o grupales realizadas dentro o fuera del aula: 40%
- Examen Final: 40% (necesario tener mínimo un 4) para tener en consideración la evaluación continua)

Convocatoria Ordinaria a partir de 2ª Matricula y dispensa académica

El criterio de asistencia y participación se suprime. El alumno será evaluado con los siguientes parámetros en convocatoria ordinaria:

- Evaluación continua (50%): Actividades individuales realizadas dentro o fuera de clase planificadas e informadas a los alumnos con la suficiente antelación.
- Examen final (50%) (necesario tener mínimo un 4 para tener en consideración la evaluación continua)

Convocatoria Extraordinaria

- En caso de tener la evaluación continua aprobada (mayor o igual que 5), se aplicarán los mismos criterios que en la convocatoria ordinaria, manteniéndose la nota ya obtenida en la evaluación continua ordinaria.

- **Evaluación Continua: 50%**
- **Examen Final: 50% (necesario tener mínimo un 4 para tener en consideración la evaluación continua)**

- En caso de haber suspendido la evaluación continua a lo largo del curso, el alumno podrá recuperar la mitad del porcentaje de esta a través de pruebas o trabajos decididos por el profesor. La otra mitad del porcentaje de la evaluación continua no se puede recuperar y ponderará la nota obtenida en la evaluación continua ordinaria.

- **Evaluación Continua: 25%**
- **Evaluación Continua recuperable: 25% (a través de pruebas o trabajos)**
- **Examen Final: 50% (necesario tener mínimo un 4 para tener en consideración la evaluación continua)**

Calificación No presentado

La calificación de "No Presentado", en cada una de las dos convocatorias oficiales, no consume convocatoria.

Cuando el alumno se haya presentado a todas las pruebas de evaluación, o a una parte de las

mismas, siempre que su peso en la nota final supere el 30%, aunque no se presente al examen final, no podrá ser calificado como No presentado en convocatoria ordinaria, siendo la nota final de la asignatura en convocatoria ordinaria la nota de la evaluación continua.

Se entenderá que un alumno se ha presentado a una prueba si asiste a la misma aunque la abandone una vez comenzada la misma.

La condición de "No Presentado" en la convocatoria extraordinaria estará ligada a la no asistencia al examen final.

TRATAMIENTO DE LA REDACCIÓN Y LAS FALTAS DE ORTOGRAFÍA

En la evaluación de los exámenes y trabajos escritos del estudiante se tendrá en cuenta su capacidad de redacción, manifestada en la exposición ordenada de las ideas, el correcto engarce sintáctico, la riqueza léxica y la matización expresiva. Se tendrá además en cuenta la propiedad del vocabulario, la corrección sintáctica, la corrección ortográfica (grafías y tildes), la puntuación apropiada y la adecuada presentación. En el caso de examen, los errores ortográficos sucesivos se penalizarán con un descuento de 0,25 cada uno, hasta un máximo de dos puntos.

» BIBLIOGRAFÍA Y RECURSOS BÁSICOS

"Introduction to Data Mining" por Pang-Ning Tan, Michael Steinbach, y Vipin Kumar. "Data Mining: Practical Machine Learning Tools and Techniques" por Ian H. Witten, Eibe Frank, y Mark A. Hall. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction" por Trevor Hastie, Robert Tibshirani, y Jerome Friedman. "Mining of Massive Datasets" por Jure Leskovec, Anand Rajaraman, y Jeffrey D. Ullman. "Pattern Recognition and Machine Learning" por Christopher M. Bishop.

Recursos bibliográficos de la Universidad Villanueva: <https://biblioteca.villanueva.edu/>

» HORARIO, TUTORÍAS Y CALENDARIO DE EXÁMENES

1.-Los horarios podrán ser consultados en tiempo real en: Horarios

Aproximadamente una sesión por tema.

2.-El calendario de exámenes podrá ser consultado en el siguiente link sobre el general para Grados del curso 2024-25. Calendario Académico

Fechas y Tareas Clave: Iremos concretando en las próximas semana

Primera práctica EDA:

Segunda práctica Reducción Dimensional:

Tercera práctica Clustering:

Práctica Final:

Horarios de atención (tutorías): Jueves de 1830 a 1930

➤ **PUBLICACIÓN Y REVISIÓN DE LA GUÍA DOCENTE**

Esta guía docente se ha elaborado de acuerdo a la memoria verificada de la titulación.