

Materia: Introducción al Análisis de Datos y Data Mining

1. IDENTIFICACIÓN

Materia: Introducción al Análisis de Datos y Data Mining

Carrera a la que pertenece: 4º año del Ciclo Profesional de la carrera Licenciado en Economía Empresarial.

Plan de Estudios: Plan de Estudios del Cincuentenario (50º).

Carga horaria: 80 horas.

2. EQUIPO DOCENTE Y COORDINADOR

Docentes:

- Ing. (Dra.) María del Carmen Romero (Profesor Asociado)
- Ing. María Jesús Martín (Ayudante Graduado)

Coordinador:

- Ing. (Dra.) María del Carmen Romero (Profesor Asociado)

3. CARGA HORARIA

La asignatura tiene una carga horaria total de 80 horas y una carga horaria semanal de 6 horas. La carga horaria semanal se distribuye en 4 (cuatro) horas presenciales divididas en 2 (dos) bloques de 2 (dos) horas cada uno, y en 2 (dos) horas a desarrollarse a través del entorno virtual de manera asincrónica.

4. MARCO REFERENCIAL

Ubicación de la asignatura en el Plan de Estudios

La asignatura se inserta en el 4º año del Ciclo Profesional de la carrera Licenciado

Programa aprobado por Resolución de Consejo Académico N° 026/2022, de fecha 30 de marzo de 2022. Vigente hasta el 30 de marzo de 2025, o hasta que se apruebe una nueva versión del mismo.-

en Economía Empresarial (Plan de Estudios del Cincuentenario (50º)).

Se desarrolla mediante el sistema de cursos en el primer cuatrimestre. Se dicta 1 curso y la cantidad prevista de alumnos es 20.

Relación de la asignatura con otras materias

Esta materia tiene como principal objetivo introducir al alumno en la extracción de conocimiento a partir de los datos. Se pretende mostrar la importancia en el entorno empresarial y brindar las herramientas básicas para, a partir de un requerimiento concreto, extraer información a partir de los datos.

Para su efectivo desarrollo, el alumno no sólo debe conocer los contenidos de la materia Econometría y Modelización (de la cual es correlativa), sino también los conceptos impartidos en las áreas de análisis de datos y gestión empresarial.

Las técnicas y metodologías aprendidas, serán necesarias para la comprensión y aplicación en problemáticas de diferentes áreas de la empresa tales como logística, operaciones, recursos humanos, marketing y ventas, entre otras.

Contenidos mínimos previstos en el Plan de Estudios

De acuerdo con el plan de estudios vigente, los contenidos mínimos previstos para este curso son:

Introducción

Etapas del Descubrimiento de Conocimiento en Bases de Datos (KDD: *Knowledge Discovery Databases*). Pre-procesamiento de datos. El proceso de *Data Mining*. Postprocesamiento de datos. Utilización del conocimiento.

Preprocesamiento de datos

Limpeza de datos. Datos faltantes: métodos de imputación. Integración y reducción de datos.

Data Mining

Conceptos generales. Tipos de análisis para “extracción” de información: reglas de asociación, clasificación y predicción, agrupamiento (jerárquico y no jerárquico).

Postprocesamiento de datos Big-Data

Introducción a la problemática de trabajar con datos de alta dimensionalidad.

Aplicaciones en el área empresarial

Conocimientos y habilidades, mínimos y básicos, que debiera poseer el alumno

Dadas las características planteadas para el curso, es recomendable que el alumno tenga conocimientos de Estadística, tanto descriptiva como inferencial, incluyendo nociones de modelos estadísticos.

Los alumnos deberán tener práctica en la utilización de herramientas informáticas tales como Word, Excel e InfoStat (versión libre) como mínimo. El objetivo de su manejo y utilización se asocia con el desarrollo de las capacidades para poder presentar informes y realizar síntesis. Se requiere que los alumnos puedan integrar estas diferentes herramientas para que los asistan en el proceso de análisis de datos, síntesis de ideas y de resultados y en la generación de conclusiones. Resultan sumamente importantes las habilidades de comunicación escrita y oral.

Se requiere, además, la capacidad de leer y comprender el idioma inglés ya que algunos materiales de referencia, así como de otros recursos, están principalmente en ese idioma.

Aportes de la asignatura a la formación del futuro profesional

Una gran variedad de los sectores profesionales basan la toma de decisiones en datos, información y conocimiento. La ciencia de datos, análisis de datos o inteligencia de negocios descripta con distintos términos (*data science, data analysis, business intelligence*) es un área que actualmente tiene una creciente demanda de profesionales calificados en todo el mundo.

Esta materia le brinda al alumno los conceptos básicos del proceso completo de extracción de información a partir de los datos, haciendo hincapié en una formación en análisis y solución de problemas de la empresa. Le proporciona un perfil empresarial y tecnológico con capacidades prácticas de análisis de datos y de manejo de herramientas, dentro de un marco científico aplicado a cuestiones empresariales.

Lo introduce, además, en los diferentes conceptos más utilizados con los cuales se identifica la extracción de conocimiento a partir de los datos: ciencia de datos (*data science*), minería de datos (*data mining*), inteligencia de negocios (*business intelligence*), analítica de negocios (*business analytics*).

Los alumnos tendrán los conocimientos necesarios para comprender todo el proceso que se inicia desde el planteo de un requerimiento hasta la publicación de los resultados obtenidos y dispondrán de una base sólida sobre modelos y algoritmos para extraer conocimiento de los datos.

Los conocimientos adquiridos aportan a un perfil profesional que pueda desarrollar alguna de las siguientes funciones: analistas de datos en diferentes áreas

de la empresa (por ejemplo, gestión, *marketing* y ventas, producción, operaciones); responsables, jefes de proyecto o analistas de sistemas de información de inteligencia de negocios; consultores e implantadores de sistemas de inteligencia de negocio; emprendedores, que desean crear negocios basados en la creación o la implantación de sistemas especializados de inteligencia de negocio o de alguno de sus componentes, así como productos y/o servicios basados en datos.

5. OBJETIVOS

General:

El principal objetivo de esta materia consiste en presentarle al alumno el “descubrimiento de conocimiento a partir de los datos” como una herramienta importante para la toma de decisiones. Se explicitan, entonces, las diferentes etapas que permiten descubrir conocimiento a partir de datos y se complementa con la aplicación a problemas concretos de la empresa.

Objetivos específicos:

Se busca que al finalizar el curso, los alumnos sean capaces de:

- comprender y aplicar los conocimientos básicos relacionados con el análisis de datos y *data mining* (en general, ciencia de datos),
- reconocer situaciones empresariales en las cuales la extracción de conocimiento a partir de los datos se convierta en una ventaja competitiva,
- entender la extracción de conocimiento como parte de un proceso mayor que comienza con un requerimiento concreto,
- adquirir competencias profesionales para detectar problemáticas que puedan ser resueltas mediante el análisis de datos,
- integrar conocimientos de diferentes disciplinas para obtener información a partir de datos en el mundo empresarial actual y
- ofrecer una formación aplicada y acercar al estudiante a la realidad profesional.

6. PROPUESTA DE CONTENIDOS

UNIDAD 1. Introducción y Conceptos básicos

Dato. Información. Conocimiento.

Ciencia de datos (Data Science). Analítica de Datos (*Data Analytics*). Minería de Datos (*Data*

Mining). Inteligencia de Negocios (*Business Intelligence*). Big Data. Descubrimiento de Conocimiento en Bases de datos (KDD: *Knowledge Discovery Databases*). Gestión del Conocimiento (*Knowledge Management*).

Etapas de KDD. Identificación del problema. Pre-procesamiento de datos. El proceso de Data Mining. Postprocesamiento de datos. Utilización del conocimiento. Ciclo de vida del dato.

El rol de la ciencia de datos en el entorno empresarial.

Objetivos específicos:

Al finalizar estos temas, se espera que los alumnos sean capaces de:

- ☐ reconocer la necesidad de la extracción de conocimiento a partir de datos en contextos empresariales,
- ☐ entender los conceptos habitualmente usados y relacionados en el contexto de ciencia de datos y
- ☐ construir una definición de los principales conceptos asociados a la extracción de conocimiento a partir de datos, de la relación que existe entre ellos y delimitar los alcances de estos conceptos.

Actividades:

- Lectura y discusión de:
 - Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996).
 - Han, Jiawei; Kamber, Micheline; Pei, Jian. (2012). Capítulo 1.
- Construcción de Glosario de los conceptos básicos.

UNIDAD 2: Preprocesamiento de datos

Datos. Tipos de datos. Calidad. Necesidad del preprocesamiento de datos. Técnicas estadísticas básicas de descripción. Visualización de datos. Técnicas de preprocesamiento de datos. Limpieza. Datos faltantes: métodos de imputación. Transformación. Integración y reducción de datos.

Aplicaciones. Introducción al software QlikView.

Objetivos específicos:

Al finalizar estos temas, se espera que los alumnos sean capaces de: ☐

- identificar la necesidad de comprobar la calidad de los datos,

- ☒ sintetizar las principales características de las técnicas de preprocesamiento de datos y
- ☒ reconocer las situaciones que requieren de su aplicación.

Actividades:

- Lectura y discusión de: Han, Jiawei; Kamber, Micheline; Pei, Jian. (2012). Capítulos 2 y 3.
- Construcción de Glosario de los conceptos básicos.
- Aplicaciones.

UNIDAD 3: Sistemas Data Warehouse y cubos de datos

DataWarehousing y Online Analytical Processing (OLAP). Definición y conceptos básicos. Necesidad y utilidad. Extracción, Transformación y Carga. Diseño, uso e implementación. Tecnologías de cubos de datos. Conceptos. Importancia. Métodos. Aplicaciones.

Objetivos específicos:

Al finalizar estos temas, se espera que los alumnos sean capaces de:

- conocer los principales conceptos utilizados en lo referente a almacenamientos de datos,
- entender la necesidad e importancia de un almacén de datos,
- familiarizarse con los procesos de extracción, transformación y carga de datos (ETL).

Actividades:

- Lectura y discusión de Han, Jiawei; Kamber, Micheline; Pei, Jian. (2012). Capítulos 4 y 5.
- Aplicaciones.

UNIDAD 4: Data Mining

Conceptos generales.

Tipos de análisis: reglas de asociación, clasificación y predicción, agrupamiento (jerárquico y no jerárquico).

Introducción a las principales técnicas estadísticas y computacionales para la clasificación (supervisada y no supervisada), agrupamiento, reducción de dimensionalidad y predicción. Análisis de componentes principales. Escalamiento multidimensional métrico/coordenadas principales. Análisis de correspondencias múltiples. Análisis de conglomerados (*cluster*).

Análisis discriminante. Árboles de regresión y clasificación. *Machine Learning*. Aprendizaje automático. *Support Vector Machines*. Algoritmos genéticos. Redes neuronales. Aplicaciones.

Objetivos específicos:

Al finalizar estos temas, se espera que los alumnos sean capaces de:

- ☐ introducir a los alumnos en las técnicas estadísticas y computacionales más usadas en la extracción de información a partir de los datos,
- ☐ reconocer las características generales de diferentes técnicas,
- ☐ identificar las técnicas pertinentes de ser aplicada dependiendo de los objetivos y
- ☐ brindar un espacio para la discusión de las similitudes y diferencias entre técnicas estadísticas y computacionales.

Actividades:

- Lectura y discusión de Han, Jiawei; Kamber, Micheline; Pei, Jian. (2012). Capítulos 6, 7, 8, 9, 10, 11 y 12.
- Gironés Roig, Jordi.
- Aplicaciones.

UNIDAD 5: Postprocesamiento de datos

Interpretación, evaluación y visualización.
Generación y transmisión de información.
Generación de conocimiento.

Objetivos específicos:

Al finalizar estos temas, se espera que los alumnos sean capaces de:

- ☐ distinguir los datos de la información y del conocimiento,
- ☐ elaborar informes para dar respuesta en términos de los requerimientos planteados e
- ☐ identificar la importancia de la gestión de la información y el conocimiento en una empresa.

UNIDAD 6: Big-Data

Introducción a la problemática de trabajar con datos de alta dimensionalidad (y datos

masivos). La necesidad de plantear métodos nuevos de preprocesamiento y análisis. Introducción al Big Data. Definición. Variedad, velocidad y volumen. Fundamentos y Usos del Big Data. Introducción a tecnologías que permiten el almacenamiento y procesamiento de datos masivos o big data.

Nuevas tendencias en Big Data.

Aplicaciones (datos de redes sociales, datos geográficos).

Objetivos específicos:

Al finalizar estos temas, se espera que los alumnos sean capaces de:

- ☐ entender los conceptos básicos y las definiciones formales asociadas al *big data*,
- ☐ conocer las tendencias y limitaciones tecnológicas en el tratamiento, procesamiento y almacenamiento de datos masivas y
- ☐ describir, desde una perspectiva general, las oportunidades del uso de *big data* en una organización.

Actividades:

- Lectura y discusión de Hastie, T.; Tibshirani, R.; Friedman, J. (2009). Capítulo 18.

UNIDAD 7: Aplicaciones en el área empresarial

Inteligencia de Negocios. Introducción a los sistemas de Inteligencia de Negocios. Definiciones y marco de referencia. *Business Analytics*. *Data Analytics*. Definición. Reconocimiento de diferentes áreas en las cuales se puede aplicar. Operaciones y Logística (*Operations analytics*). Marketing y Ventas (*Customer analytics*). Gestión económico-financiera (*Finance analytics*). Recursos Humanos (*People analytics*).

Customer Analytics (inteligencia de negocios aplicada al área de marketing y ventas).

Objetivos. Fases. Beneficios.

Aplicaciones.

Objetivos específicos:

Al finalizar estos temas, se espera que los alumnos sean capaces de:

- ☐ comprender los conceptos asociados a la inteligencia de negocios, ☐ reconocer a la “inteligencia de negocios” como una aplicación de la extracción de conocimiento a partir de bases de datos en un entorno empresarial, ☐ identificar casos de aplicación en cuestiones particulares de la empresa y ☐ desarrollar un caso desde el planteamiento o

detección de una necesidad concreta hasta la obtención de conclusiones útiles para el negocio y para tomar decisiones (incluyendo la descripción de los métodos y técnicas de análisis de datos).

Actividades:

- Lectura y discusión de
Jordi Conesa i Caralt; Núria Braulio Gil; Josep Curto Díaz.
Abelló Gamazo, Alberto; Samos Jiménez, José; Curto Díaz, Josep.
Han, Jiawei; Kamber, Micheline; Pei, Jian. (2012). Capítulo 13.
- Exposición de un caso aplicado al contexto empresarial.

7. PROPUESTA METODOLÓGICA

En este curso se brindan herramientas para la extracción de conocimiento a partir de bases de datos.

Se centrará especialmente en dar una visión general e introductoria de todo el proceso a través de casos prácticos y utilizando herramientas presentes en el mercado.

Esta materia cuenta con 2 clases semanales de 2 horas cada una, las cuales se dictarán en la modalidad teórica-práctica y con 2 horas semanales a cumplimentar de manera asincrónica.

Clases teórico-prácticas

En cada una de las clases se expondrán los temas y conceptos propuestos, teniendo presente la necesidad que conlleva a estudiar cada uno de dichos temas. El

abordaje de los mismos será realizado a partir de un ejemplo o de un caso concreto en el cual se ponga de manifiesto la necesidad de aplicación de determinadas herramientas (y conceptos).

En un paso posterior, se procederá a la conceptualización. Para ello, se trabajará con variedad de materiales: apuntes específicos elaborados con fines *didácticos* (por ejemplo, filminas), lectura de artículos recomendados, casos prácticos, casos de estudio, enlaces a páginas que resulten de interés, actividades propuestas, entre otras.

Adicionalmente a los materiales indicados como obligatorios, en el transcurso de la asignatura se pueden recomendar otros recursos como lecturas adicionales, webs de referencia, y blogs especializados. La dinámica del curso puede implicar puntualmente la búsqueda por parte del estudiante de bibliografía o recursos adicionales.

Se prevé la ejemplificación y el debate de problemas concretos de la empresa con la finalidad de “interpretar” cada uno de los conceptos, produciendo de esta manera, una retroalimentación constante entre teoría y práctica.

Actividades asincrónicas

Durante todo el cuatrimestre se contará con un espacio virtual. El objetivo del uso de este espacio es el de complementar el espacio presencial de manera tal que se logre una integración de los conocimientos.

Se plantean las siguientes actividades:

- Obligatorias con evaluación:

- o Cuestionarios de auto-evaluación a realizarse al finalizar cada uno de los temas.

Cada uno de ellos estará disponible durante el tiempo que se desarrolle el tema y los alumnos podrán completarlo en el momento que lo deseen. Estos cuestionarios tienen un doble objetivo, por un lado, que los alumnos puedan corroborar los conocimientos que tienen sobre un tema determinado, y por otro, que los docentes puedan relevar los principales inconvenientes surgidos en la comprensión de los mismos.

Estos cuestionarios son obligatorios pero no se tendrá en cuenta el puntaje obtenido en la nota final de la materia. Se pedirá que los alumnos hayan respondido un mínimo de 3 (tres) cuestionarios para poder acceder a la promoción de la materia.

- o Entrega de trabajo a formar parte de la evaluación de la segunda parte de la materia.

- Optativas:

- o Glosario de conceptos.

- o Foro de consultas. Para cada uno de los temas, habrá un foro para que los alumnos puedan plasmar las dudas y consultas.

Programas Estadísticos

Se trabajará con el programa estadístico InfoStat (versión libre). Se explorarán, además, herramientas de reporte y análisis tales como QlikView y Tableau, entre otras.

8. NÚCLEO CENTRAL DE ACTIVIDADES Y/O TRABAJOS PRÁCTICOS

Se presenta a continuación el Plan de Actividades que detalla las tareas a

desarrollar en cada una de las clases del cuatrimestre.

PLAN DE ACTIVIDADES ASIGNATURA: INTRODUCCIÓN AL ANÁLISIS DE DATOS Y DATA MINING

A continuación se presentan el cronograma de la materia (considerando un total de 15 semanas ya que su dictado es cuatrimestral).

SEMANA	UNIDAD	TEMA A DESARROLLAR	ACTIVIDADES PRESENCIALES	ACTIVIDADES ASINCRÓNICAS	BIBLIOGRAFÍA RECOMENDADA
1	1	<p>Introducción y conceptos básicos Dato. Información. Conocimiento. Ciencia de datos (Data Science). Analítica de Datos (<i>Data Analytics</i>). Minería de Datos (<i>Data Mining</i>). Inteligencia de Negocios (<i>Business Intelligence</i>). Big Data. Descubrimiento de Conocimiento en Bases de datos (KDD: <i>Knowledge Discovery Databases</i>). Gestión del Conocimiento (<i>Knowledge Management</i>).</p>	<p>Exposición del profesor.</p> <p><u>Actividades:</u> Lectura y discusión de: - Fayyad, Piatetsky Shiapiro, Smyth y Uthurusamy. (1996). - Han, Kamber y Pei, Jian. (2012). Capítulo 1.</p>	<p>☑ Foro de consultas (optativo). ☑ Glosario (optativo). ☑ Cuestionario auto evaluación (obligatorio).</p>	<p>☑ Apuntes de cátedra. ☑ Fayyad, Piatetsky Shiapiro, Smyth y Uthurusamy (1996). ☑ Han, Kamber y Pei, Jian. (2012). Capítulo 1. ☑ Marr, B. (2016).</p>
2	1	<p>Introducción y conceptos básicos Etapas de KDD. Identificación del problema. Pre-procesamiento de datos. El proceso de Data Mining. Postprocesamiento de datos. Utilización del conocimiento. Ciclo de vida del dato.</p>			

3	2	<p>Preprocesamiento de datos Datos. Tipos de datos. Calidad. Necesidad del preprocesamiento de datos. Técnicas estadísticas básicas de descripción. Visualización de datos. Técnicas de preprocesamiento de datos. Limpieza. Datos faltantes: métodos de imputación.</p>	<p>Exposición del profesor. <i>Actividades:</i> - Lectura y discusión de: Han, Kamber y Pei. (2012). Capítulos 2 y 3. - Aplicaciones.</p>	<p>☑ Glosario (optativo). ☑ Cuestionario auto evaluación (obligatorio). ☑ Foro de consultas (optativo).</p>	<p>☑ Apuntes de cátedra. ☑ Han, Kamber, y Pei. (2012). Capítulos 2 y 3.</p>
		<p>Transformación. Integración y reducción de datos. Aplicaciones. Introducción al software QlikView.</p>			
4	3	<p>UNIDAD 3: Sistemas Data Warehouse y cubos de datos DataWarehousing y Online Analytical Processing (OLAP). Definición y conceptos básicos. Necesidad y utilidad. Extracción, Transformación y Carga. Diseño, uso e implementación. Tecnologías de cubos de datos. Conceptos. Importancia. Métodos. Aplicaciones.</p>	<p>Exposición del profesor. <i>Actividades:</i> - Lectura y discusión de Han, Kamber y Pei. (2012). Capítulos 4 y 5. - Aplicaciones.</p>	<p>☑ Cuestionario auto evaluación (obligatorio). ☑ Foro de consultas (optativo).</p>	<p>☑ Apuntes de cátedra. ☑ Han, Kamber y Pei. (2012). Capítulos 4 y 5.</p>
5		<p>Consulta general</p>			

7	4	<p>UNIDAD 4: Data Mining Conceptos generales. Tipos de análisis: reglas de asociación, clasificación y predicción, agrupamiento (jerárquico y no jerárquico).</p>	<p>Exposición del profesor. <i>Actividades:</i> - Lectura y discusión de: Han, Kamber y Pei. (2012). Capítulos 6, 7, 8, 9, 10, 11 y 12. Gironés Roig, Jordi. - Aplicaciones.</p>	<p>☑ Cuestionario auto evaluación (obligatorio). ☑ Foro de consultas (optativo).</p>	<p>☑ Han, Kamber y Pei. (2012). Capítulos 6, 7, 8, 9, 10, 11 y 12. ☑ Gironés Roig, Jordi. Lectura sugerida: ☑ Hastie, Tibshirani y Friedman. (2009). ☑ James, Witten, Hastie y Tibshirani (2013).</p>
8	4	<p>UNIDAD 4: Data Mining Introducción a las principales técnicas estadísticas y computacionales para la clasificación (supervisada y no supervisada), agrupamiento, reducción de dimensionalidad y predicción. Análisis de componentes principales. Escalamiento multidimensional métrico/coordenadas principales. Análisis de correspondencias múltiples.</p>			

9	4	<p>UNIDAD 4: Data Mining Análisis de conglomerados (<i>cluster</i>). Análisis discriminante. Árboles de regresión y clasificación. Aplicaciones.</p>			
10	4	<p>UNIDAD 4: Data Mining <i>Machine Learning</i>. Aprendizaje automático. <i>Support Vector Machines</i>. Algoritmos genéticos. Redes neuronales.</p>			

11	5	UNIDAD 5: Postprocesamiento de datos Interpretación, evaluación y visualización. Generación y transmisión de información. Generación de conocimiento.	Exposición del profesor.		
12	6	UNIDAD 6: Big-Data Introducción a la problemática de trabajar con datos de alta dimensionalidad (y datos masivos). La necesidad de plantear métodos nuevos de preprocesamiento y análisis. Introducción al Big Data. Definición. Variedad, velocidad y volumen. Fundamentos y Usos del Big Data. Introducción a tecnologías que permiten el almacenamiento y procesamiento de datos masivos o big data. Nuevas tendencias en Big Data. Aplicaciones (datos de redes sociales, datos geográficos).	Exposición del profesor. <i>Actividades:</i> Lectura y discusión de Hastie, Tibshirani y Friedman. (2009). Capítulo 18	<input type="checkbox"/> Cuestionario auto evaluación (obligatorio). <input type="checkbox"/> Foro de consultas (optativo).	<input type="checkbox"/> Hastie, Tibshirani y Friedman. (2009). Capítulo 18
13	7	UNIDAD 7: Aplicaciones en el área empresarial Inteligencia de Negocios. Introducción a los	Exposición del profesor. <i>Actividades:</i> Lectura y discusión de	<input type="checkbox"/> Foro de consultas (optativo).	<input type="checkbox"/> Conesa i Caralt, Braulio Gil, Curto Díaz.

		<p>sistemas de Inteligencia de Negocios. Definiciones y marco de referencia. <i>Business Analytics. Data Analytys</i>. Definición. Reconocimiento de diferentes áreas en las cuales se puede aplicar. Operaciones y Logística (<i>Operations analytics</i>). Marketing y Ventas (<i>Customer analytics</i>). Gestión económico-financiera (<i>Finance analytics</i>). Recursos Humanos (<i>People analytics</i>). <i>Customer Analytics</i> (inteligencia de negocios aplicada al área de marketing y ventas). Objetivos. Fases. Beneficios. Aplicaciones.</p>	<p>Conesa i Caralt, Braulio Gil y Curto Díaz. Abelló Gamazo, Samos Jiménez y Curto Díaz. Han, Kamber y Pei. (2012). Capítulo 13. Elaboración de ejemplos con los alumnos. Cierre general de la materia. Integración general de conceptos.</p>		<p>☒ Abelló Gamazo, Samos Jiménez y Curto Díaz. ☒ Han, Kamber y Pei. (2012). Capítulo 13</p>

9. CONDICIONES DE APROBACIÓN

Las condiciones de aprobación se ajustan al reglamento de Enseñanza y Promoción (RCA N° 080/2017) y a su modificación (RCA N° 140/2019).

La asignatura puede aprobarse en carácter promocional o bien mediante examen final, considerando el rendimiento de los alumnos en el transcurso de la cursada. Se tendrán dos (2) instancias de evaluación parciales, distinguiéndose las siguientes situaciones:

- En caso de obtener **nota promedio mayor o igual a 6 (seis)** y de haber cumplimentado las actividades asincrónicas obligatorias (mínimo de 3 cuestionarios de auto-evaluación), el alumno quedará habilitado para ser evaluado en una instancia integradora de los conocimientos y prácticas fundamentales de la Asignatura. Si en la instancia integradora, obtiene una nota mayor o igual a 6 (seis), tendrá la ***promoción de la materia***. En caso de obtener una nota menor (o de estar ausente), ***aprobará la materia*** y quedará habilitado para rendir examen final en carácter de alumno regular. ***Importante: En caso de que el alumno no haya cumplido con las actividades asincrónicas obligatorias, no estará en condiciones de promocionar.***
- En caso de haber rendido las dos instancias y obtener **nota promedio mayor o igual que 4 (cuatro) y menor que 6 (seis)**, el alumno ***aprobará la materia*** y quedará habilitado para rendir examen final en carácter de alumno regular.
- En caso de no quedar encuadrado en los casos anteriores o de mediar ausencia sin justificación en alguna de las instancias de evaluación parciales, el alumno tendrá la posibilidad de rendir un recuperatorio global. Si en el mismo obtiene una nota mayor o igual a 4 (cuatro), ***aprobará la materia*** y quedará habilitado para rendir examen final en carácter de alumno regular. Si obtiene una nota menor a 4 (cuatro), desaprobará la materia y deberá recursarla.

La asistencia mínima obligatoria será del 60% computable sobre la totalidad de las clases efectivas.

Las evaluaciones parciales consistirán en exámenes individuales escritos y en la entrega de un trabajo (la cual se realizará mediante la plataforma virtual), en los que se pretende relevar el nivel de conocimiento y habilidad adquiridos por el alumno a lo largo de la cursada. Se trata de dos instancias de evaluación teórico – práctica.

El examen integrador consistirá en una situación teórico-práctica en la cual el alumno deberá aplicar e integrar distintos temas desarrollados durante la cursada.

Se utilizará la siguiente escala de calificaciones:

Puntos Porcentuales	Escala	Descripción
0	0 (cero)	Reprobado
01 a 20	1 (uno)	Aplazado
21 a 30	2 (dos)	Aplazado
31 a 49	3 (tres)	Aplazado
50 a 53	4 (cuatro)	Suficiente
54 a 59	5 (cinco)	Bueno
60 a 68	6 (seis)	Bueno
69 a 77	7 (siete)	Muy bueno
78 a 86	8 (ocho)	Muy bueno
87 a 95	9 (nueve)	Distinguido
96 a 100	10 (diez)	Sobresaliente

Tendrá vigencia el programa del año en curso.

10. CONTENIDOS A REVALIDAR

Al ser una asignatura que aborda cuestiones metodológicas y tecnológicas para el tratamiento y análisis de datos en un contexto de cambio constante, se deberán revalidar todos aquellos contenidos que se modifiquen por cuestiones de actualidad o relevancia profesional. Además, también deberán hacerlo con respecto a aquellos contenidos que se incorporen a los programas vigentes y difieran del programa correspondiente al momento de su cursada.

11. BIBLIOGRAFÍA

a) *Requerida:*

Abelló Gamazo, Alberto; Samos Jiménez, José; Curto Díaz, Josep. *La factoría de información corporativa*. CC-BY-NC-ND • PID_00203541.

Bradley, Efron; Hastie, Trevor. (2016). *Computer Age Statistical Inference. Algorithms, Evidence, and Data Science*. Cambridge University Press.

Conesa i Caralt, Jordi; Braulio Gil, Núria; Curto Díaz, Josep. *Customer Analytics. Mejorando la inteligencia del cliente a través de los datos*. Barcelona: Universitat Oberta de Catalunya.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining*. Menlo Park, CA: AAAI Press.

Gironés Roig, Jordi. *Algoritmos*. FUOC. Fundació para la Universitat Oberta de Catalunya. CC-BY-NC-ND • PID_00197284.

http://openaccess.uoc.edu/webapps/o2/bitstream/10609/71345/1/Business%20analytics_M%C3%B3dulo%202_Algoritmos.pdf

Han, Jiawei; Kamber, Micheline; Pei, Jian. (2012). *Data Mining. Concepts and Techniques*. Third Edition. Elsevier.

Hastie, T.; Tibshirani, R.; Friedman, J. (2001). *The elements of statistical learning*. Nueva York: Springer.

Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Second Edition Springer. 745 páginas.

James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert (2013). *An Introduction to Statistical Learning with Applications in R*. ISSN 1431-875X. ISBN 978- 1-4614-7137-0 ISBN 978-1-4614-7138-7 (eBook). DOI 10.1007/978-1-4614-7138-7. Springer Science+Business Media New York.

Johnson, R; Wichern, D. (1998). *Applied Multivariate Statistical Analysis*. 4ª Edición. Prentice Hall.

Marr, B. (2016). *Big Data Uncovered: What Does A Data Scientist Really Do?*
<https://www.forbes.com/sites/bernardmarr/2016/01/07/big-data-uncovered-what-does-a-data-scientist-really-do/?sh=76cfc3a7a5bb>.

Software:

InfoStat versión 2019. <http://www.infostat.com.ar>

QlikView. <https://dataiq.com.ar/>

b) Complementaria:

Anderson, D. R.; Sweeney, Dennis J.; Williams, Thomas A. *Estadística para administración y economía*. (2008). Cengage Learning Editores. 10ª edición. 1091 páginas.

Balzarini M.G., Gonzalez L., Tablada M., Casanoves F., Di Rienzo J.A., Robledo C.W. (2008). *Infostat. Manual del Usuario*. Editorial Brujas, Córdoba, Argentina.

Canavos, G. (1988). *Probabilidad y Estadística. Teoría y aplicaciones*. Mc Graw Hill. Interamericana de México. 667 páginas.

Cuadras, C. M. (1996). *Métodos de análisis multivariante*. EUB, S.L.

Davenport, T.H. y Harris, J.G. (2007). *Competing on Analytics: The New Science of Winning*. Nueva York: Harvard Business Press.

Hernández Orallo J., Ramírez Quintana M y Ferri Ramirez C. (2004). *Introducción a la Minería de Datos*. Editorial Pearson Educación SA, Madrid.

Lebart, L.; Morineau, A.; Piron, M. (1995). *Statistique Exploratoire Multidimensionnelle*. Dunod.

Levine, D. M; Krehbiel, T. C.; Berenson, M. L. (2006). *Estadística para administración*. México: Pearson Educación, 4ª Edición. 619 páginas.

Nisbert, R.; Elder, J.; Miner, G. (2009). *Handbook of Statistical Analysis and Data Mining Applications*. Academic Press.

Peña, D. (2002). *Análisis de Datos Multivariantes*. Ed. Mc. Graw Hill.

Pyle D. (2003). Business Modeling and Data Mining. *The Morgan Kaufmann Series in Data Management Systems*. Morgan Kaufmann Publisher.

Putler, D.S. y Krider, R.E. (2012). *Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R*. New York: Chapman and Hall/CRC.

Zhou Z. (2003). Three perspectives of Data Mining. *Artificial Intelligence* (143) pg. 139-146. Elsevier Science B.V.

Links de interés

- <http://hbr.org>
- <https://www.gartner.com/it-glossary>
- Knowledge Discovery Mine <http://www.kdnuggets.com>
- <http://www.information-management.com>
- <http://searchdatamanagement.techtarget.com>
- <http://dataanalysis.blogs.uoc.edu>

- StatSoft Inc. (2003) Data Mining Textbook. Electronic Textbook.
<http://www.statsoft.com/textbook/stdatmin.html>
- DMG - The Data Mining Group, un consorcio mixto para crear estándares e intercambiar modelos predictivos. <http://www.dmg.org>