

SYLLABUS ACADEMICO  
**MAGISTER EN DATA SCIENCE**  
**MDS 2021**

**ASIGNATURA** : Taller de Ciencia de Datos para Soluciones Aplicadas de negocios / Casos de Negocios en Ciencia de Datos / Estudio de Casos en Ciencia de Datos / Taller de casos aplicados en Ciencia de datos

**NOMBRE PROFESOR** : Luis Aburto

**EMAIL** : [Luis.aburto@uai.cl](mailto:Luis.aburto@uai.cl)

## 1. INTRODUCCIÓN

Para desarrollar un proyecto en Ciencia de Datos, no solo es necesario conocer bien las herramientas y modelos a usar sobre la data. Resulta muy relevante para el éxito de estos proyectos desarrollar una metodología ordenada y siempre ligado a las mejoras en procesos de negocios, tanto en la industria privada como en organismos gubernamentales. Durante este curso los alumnos harán un recorrido sobre distintas metodologías usadas en la disciplina. Esto se aplicará a distintos casos de negocios de problemas típicos presentes en diversas industrias (segmentación de clientes, recomendación de productos, retención de clientes, optimización de procesos, entre otros). Uno de los grandes objetivos de este curso es que los alumnos comprendan la importancia del contexto de negocio para resolver problemas de Ciencia de Datos. De la misma forma, se evaluarán distintas herramientas evaluando el compromiso entre capacidad predictiva y explicativa. Por último, se desarrollarán problemas de estadística prescriptiva, combinando el poder predictivo de modelos de machine learning junto con herramientas de optimización.

## 2. OBJETIVOS DE LA ASIGNATURA

### Objetivo General:

Desarrollar casos de uso de proyectos en Data Science ligando aspectos esenciales en la aplicación y rediseño de procesos de negocio.

### Objetivos Específicos:

- Entender las distintas metodologías existentes en Ciencia de Datos
- Recomendar acciones de marketing basadas en herramientas de clustering para segmentación de clientes.
- Evaluar como las técnicas de procesamiento mejoran la capacidad de respuesta de modelos de machine learning
- Evaluar técnicas de ingeniería de atributos para la mejora de modelos predictivos
- Evaluar la mejor decisión en campañas de retención de clientes incorporando costos y optimización.

- Usar técnicas de optimización y machine learning para recomendar decisiones de operación minera.
- Conocer diversas industrias y problemas de negocio donde existen oportunidad de aplicar proyectos de Ciencia de Datos

### 3. METODOLOGIA

El desarrollo del curso se llevará a cabo mediante las siguientes actividades:

- Clases expositivas: realizadas por el profesor donde se revisarán los aspectos conceptuales, técnicas y modelos, y aspectos prácticos.
- Ejercicios prácticos: ejercicios prácticos en un lenguaje de programación R que permita entender los conceptos en cada sesión.
- Controles de lectura y Trabajos Grupales: Se realizarán controles de lectura de papers complementarios a la materia. De la misma forma, se realizarán trabajos grupales sobre data y casos de negocios para discusión y análisis.

### 4. EVALUACION

El curso tendrá las siguientes evaluaciones:

- Controles de Lectura (20%)
- Talleres Grupales (30%)
  - Acciones comerciales en Home Improvement
  - Data Mining Cup Distribución Eléctrica
- Examen (50%)

En caso de reprobado alguno de los requisitos, el alumno **tendrá derecho a una instancia de examen recuperativo** a fecha a estipular por la Coordinación Académica y profesor, optando a la **calificación mínima de aprobación del curso** (a no ser que el Syllabus diga otra cosa).

Si el alumno, después de realizar el examen recuperativo, no cumple con alguno de estos dos requisitos, tendrá que cursar la actividad curricular en una segunda oportunidad cuando se dicte nuevamente el curso.

#### INASISTENCIA A EVALUACIONES

En caso que el alumno no pueda asistir a una evaluación, deberá justificar con **certificado médico o laboral para acceder a una segunda y última instancia de evaluación** optando a la nota máxima de aprobación.

Si la ausencia se produce el día del examen, y **ésta no es justificada o es rechazada**, el alumno tendrá derecho a rendir una prueba de recuperación, que se realizará el mismo día del examen de repetición, **pero optando a la nota máxima de examen de un 4.0.** (cuatro coma cero décimas).

Si el alumno nuevamente está ausente en el examen de repetición, éste quedará pendiente hasta la próxima ocasión en que la asignatura o módulo se dicte.

## 5. BIBLIOGRAFIA

Wickham, H., & Grolemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. " O'Reilly Media, Inc."

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

Davenport, T. H. (2006). Competing on analytics. Harvard business review, 84(1), 98.

Do Your Data Scientists Know the 'Why' Behind Their Work? HBR

Berinato, S. (2019). Data science & the art of persuasion. Harvard Business Review, 97(1), 126-137.

Shan, C. (2015). The Data Science Handbook: Advice and Insights from 25 Amazing Data Scientists. Data Science Bookshelf.

Davenport, T. H., Harris, J. G., & Morison, R. (2010). Analytics at work: Smarter decisions, better results. Harvard Business Press.

Siegel, E. (2013). Predictive analytics: The power to predict who will click, buy, lie, or die. John Wiley & Sons.

Duhigg, C. (2012). How companies learn your secrets. *The New York Times*, 16(2), 1-16.

Scherer, M. (2012). Inside the secret world of the data crunchers who helped Obama win. Time. com, 7.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27-34.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). The CRISP-DM user guide. In *4th CRISP-DM SIG Workshop in Brussels in March* (Vol. 1999).

Pedersen, C. L., & Ritter, T. (2020). Use This Framework to Predict the Success of Your Big Data Project. Harvard Business Review.

Redman, T. (2019). Do Your Data Scientists Know the 'Why' Behind Their Work? Harvard Business Review.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77-84.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87.

## 6. CURRICULUM RESUMIDO DEL PROFESOR

**Luis Aburto** es PhD. en Sistemas de Ingeniería de la Universidad de Chile. Además, es Ingeniero Industrial y Master en Gestión de Operaciones de la misma casa de estudios. Luis es profesor asistente de la Facultad de Ingeniería y Ciencias de la Universidad Adolfo Ibáñez. Su investigación se centra en temas de machine learning, forecasting, pricing, operaciones y marketing cuantitativo. Ha participado también en la Universidad Católica de Lovaina (KULEUVEN) como parte del proyecto de intercambio de NeEDS (Network of European Data Scientists) Project.

Previamente, fue profesor en el Departamento de Ingeniería Industrial de la Universidad de Chile, participando también como miembro del Centro de Estudios del Retail. Además, fue socio fundador de Penta Analytics, empresa de Inteligencia de Negocios y Data Science aplicado a distintas industrias como Retail, Banca, Seguros, Distribución Eléctrica, Minería, Manufactura y Gobierno. Ha desarrollado e implementado soluciones analíticas en varios países de Latinoamérica y Estados Unidos, en temas como Segmentación de Clientes y Estimación de valor potencial, Diseño y Evaluación de Promociones, Modelos de Retención de Clientes, Modelos de Detección de Fraude, Pricing y Revenue Management.

## 7. PROGRAMA

### SESION 1

**Tema: Introducción a Metodologías de Ciencia de Datos y Modelos No supervisados**

**Objetivos:**

- *Revisar metodologías para desarrollo de proyectos analíticos y sus diferencias*
- *Aplicar herramientas de clustering para segmentación de clientes y revisar problemas más frecuentes*
- *Extraer conclusiones y recomendaciones de negocio en base a segmentación de clientes*

**Contenidos:**

- Metodologías analíticas: KDD, CRISP-DM, SEMMA, Wickham
- Elementos a considerar para realizar clustering y segmentación de clientes
- Análisis y tipos de segmentación de clientes

**Lectura:**

- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27-34.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). The CRISP-DM user guide. In *4th CRISP-DM SIG Workshop in Brussels in March* (Vol. 1999).
- Davenport, T. H. (2006). Competing on analytics. *Harvard business review*, 84(1), 98.

**Caso:**

- **Segmentación de clientes para banca personas**

- **Análisis de segmentación de clientes para home improvement (taller)**

## SESIÓN 2

### **Tema: Modelos No Supervisados Recomendación de Productos para Retail**

**Objetivos:** Revisar elementos de negocio en aplicación de modelos de recomendación de productos

**Contenidos:**

- Elementos de diseño en reglas de asociación
- Correlación espuria y definición de soporte mínimo
- Elementos de negocio para la aplicación de modelo de recomendaciones

**Lectura:**

- Duhigg, C. (2012). How companies learn your secrets. *The New York Times*, 16(2), 1-16.
- Scherer, M. (2012). Inside the secret world of the data crunchers who helped Obama win. *Time. Com*, 7.

**Caso:**

- Recomendación de packs para promociones de supermercado
- Recomendación de productos para venta en canal tradicional

## SESIÓN 3

### **Tema: Modelos supervisados de clasificación**

**Objetivos:** Revisar casos de aplicación de modelos de clasificación y elementos para una correcta evaluación e implementación en los negocios

**Contenidos:**

- Evaluación de modelos de clasificación
- Optimización de costos para diseño de cut-off en modelos de clasificación
- Elementos de negocios a incorporar en la implementación de modelos supervisados

**Lectura:**

- Pedersen, C. L., & Ritter, T. (2020). Use This Framework to Predict the Success of Your Big Data Project. *Harvard Business Review*.
- Redman, T. (2019). Do Your Data Scientists Know the 'Why' Behind Their Work? *Harvard Business Review*.
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87.

**Caso:**

- Predicción de fuga para clientes de Seguros
- Diseño de campañas de retención para clientes de distribución eléctrica en Alemania (taller)

## SESIÓN 4

## **Tema: Prescriptive Analytics: Uso de herramientas de machine learning y optimización**

### **Objetivos:**

- *Analizar combinación de modelos predictivos y herramientas de optimización para apoyo en toma de decisiones*
- *Revisar herramientas de AB testing y diseño experimental para evitar sesgos en la construcción de modelos*

### **Contenidos:**

- Casos aplicados de modelos predictivos y optimización para recomendación de decisiones
- Elementos de diseño experimental y AB testing para control y análisis de recomendaciones.
- Uplift modeling

### **Lectura:**

- Berinato, S. (2019). Data science & the art of persuasion. Harvard Business Review, 97(1), 126-137.
- Gallo, A. (2017). A refresher on A/B Testing. Harvard Business Review.  
<https://hbr.org/2017/06/a-refresher-on-ab-testing>
- Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77-84.

### **Casos:**

- Estimación de demanda y sensibilidad de precios para optimización de rendimiento de categorías de productos
- Modelo de cobranza y diseño experimental para evaluación de acciones
- Optimización de procesos mineros y estimación de recuperación en lixiviación
- Modelo de predicción de fallas de máquinas y optimización de turnos de mantenimiento