

Introducción a la gestión de datos para generar valor en la industria aseguradora

Mediante el desarrollo de análisis de papers y casos prácticos apoyado con herramientas de inteligencia artificial, este curso busca exponer al alumno a problemáticas reales en la industria de seguros con el objetivo de desarrollar habilidades básicas en análisis de datos, creación de productos y evaluación de proyectos que usen los datos como herramienta para agregar valor a la empresa.

Objetivo General

Al término del curso los alumnos serán capaces de aplicar diferentes metodologías de inteligencia de negocios en el análisis de datos para la creación de productos y evaluación de proyectos.

Objetivos del curso:

- Identificar los fundamentos de la inteligencia de negocios
- Distinguir los diferentes fundamentos en el tratamiento de los datos
- Interpretar datos con el uso de estadística y análisis exploratorio
- Comparar diferentes modelos de clasificación en la medición de riesgos
- Experimentar con regresiones en la estimación de precios
- Aplicar técnicas de clustering y análisis RFM en la segmentación de clientes
- Construir sistemas con resiliencia en la gestión operacional
- Aplicar diferentes fundamentos de infraestructura en la construcción de sistemas con resiliencia

Público Objetivo: Gerentes, C-levels, Subgerentes y mandos medios que vayan a liderar o comenzar iniciativas que busquen hacer uso de los datos para generar valor en sus industrias.

Prerrequisitos: El curso requiere nivel escolar de matemáticas, estadística y excel a nivel básico. NO requiere habilidades de programación previa, pues si bien se entregarán ejemplos prácticos con código ejecutado, el foco no estará en el código sino en las estrategias de solución y en las aplicaciones posibles en la industria.

Preparación recomendada:

Repasar resolución de problemas básicos de matemáticas y estadística

Crear cuenta en Chat-GPT, Llama Meta, Claude y Codeium

Cantidad de Horas: 40 horas cronológicas

Estructura del curso:

Módulo 1: ¿Cómo usar las LLM para apoyarte en el trabajo con el uso de datos?

- ¿Cómo hacer un buen prompt?
- ¿Cómo validar resultados?
- ¿Cómo crear agentes especializados con GPT 4.0? (cuenta pagada)
- ¿Cómo mejorar los resultados usando varias LLM al mismo tiempo? La importancia de mantener la ética en el uso de LLM
 - Lo correcto e incorrecto
 - Alcances y límites de su uso

Módulo 2: Fundamentos de inteligencia de negocios y el uso de modelos para agregar valor usando los datos

- ¿Qué es inteligencia de negocios?
- La diferencia entre medir procesos y resultados
- ¿Qué es y cómo crear un buen KPI? Análisis del paper Competing on analytics
- ¿Qué significa machine learning, Deep learning e IA? ¿En qué se diferencian y cuándo se usan?
- ¿Qué es un Data product? ¿Cuáles son sus componentes y cuándo debo construir uno?
- ¿Cómo evaluar una iniciativa asociada al uso de los datos?
- Análisis del podcast:

Módulo 3: Fundamentos para el tratamiento de datos

- Análisis del paper CRISPDM Towards a Standard Process Model for Data Mining
- y cómo adaptarlo a las aplicaciones actuales
- Tipos de datos a los que me voy a enfrentar y sus usos más comunes
- ¿Qué significa Extracción, limpieza y transformación de datos?
- ¿Por qué necesito hacer exploración los datos? ¿cuáles son los métodos más comunes?
- ¿Por qué necesito hacer limpieza? ¿cuáles son los métodos más comunes?
- ¿Cuándo necesito hacer transformación de los datos? ¿cuáles son los métodos más comunes?
- Caso práctico 1: Ver cómo se aplican las técnicas de limpieza y transformación más comunes sobre un dataset asociado a la venta de seguros de vida (<https://www.kaggle.com/datasets/anmolkumar/health-insurance-cross-sell-prediction/data>)
- Caso práctico 2: Visualización y presentación de resultados utilizando Tableau Public o Metabase Open Source

Módulo 4: Estadística básica, experimentación y validación de hipótesis

- Conceptos fundamentales y técnicas más utilizadas de la estadística descriptiva: medidas de tendencia central, análisis univariado, bivariado, multivariado y test de hipótesis
- Caso práctico 1: Hacemos exploración de datos usando lo aprendido sobre el dataset de seguros de vida
- ¿Cómo saber si los productos que estoy vendiendo son los adecuados? ¿Puedo medir mis esfuerzos de comercialización? ¿cómo?
- ¿Qué es un A/B testing? análisis del paper: A/B testing: A systematic literature review
- Caso práctico 2: Realizamos un análisis de experimentación usando datasets simples de marketing (<https://www.kaggle.com/datasets/faviovaz/marketing-ab-testing/data>) o (<https://www.kaggle.com/datasets/amirmotefaker/ab-testing-dataset/data>)

Módulo 5: Medición del riesgo usando modelos de clasificación

- ¿Qué significa riesgo y por qué se mide?
- ¿Qué es un modelo de clasificación y para qué se puede usar?
- Análisis de caso práctico usando el paper de riesgo de pago de clientes en una empresa chilena
- Análisis de caso práctico usando el paper sobre detección de falsos reclamos al seguro de automóviles de una compañía de USA
- Caso práctico 1: Cómo simplificar la medición del riesgo puede incrementar la adquisición del producto usando datos de una competencia en Kaggle (<https://www.kaggle.com/competitions/prudential-life-insurance-assessment/overview>)
- Caso práctico 2: Visualización y presentación de resultados utilizando Tableau Public o Metabase Open Source

Módulo 6: Estimaciones de precio básicas usando regresiones

- ¿Qué es una regresión y para qué se puede usar?
- Análisis de caso práctico usando el paper que utiliza regresiones para estimar la demanda de seguros de salud en USA
- Caso práctico 1: Entender regresiones lineales simples y múltiples haciendo una estimación del costo de un seguro de salud en USA (<https://www.kaggle.com/datasets/teertha/ushealthinsurancedataset/data>)
- Caso práctico 2: Entender las técnicas de regresión disponibles y cómo evaluarlas usando la competencia de predicción de precios de casas de Kaggle (<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques>)

Módulo 7: Generación de segmentación de clientes usando análisis RFM y técnicas de clustering

- ¿Qué es una segmentación y cómo puede ser útil en la industria de seguros?
- ¿Qué es el análisis RFM? ¿Cuándo lo puedo aplicar y para qué sirve?

- Caso práctico 1: Realizamos una segmentación con un dataset público de transacciones de tarjeta de crédito para entender cómo segmentar a las personas en base a su consumo
(<https://www.kaggle.com/datasets/arjunbhasin2013/ccdata/data>)
(<https://www.kaggle.com/datasets/shivamb/bank-customer-segmentation/data>)
- Caso práctico 2: Visualización y presentación de resultados utilizando Tableau Public o Metabase Open Source

Módulo 8: Construcción de sistemas con resiliencia

- ¿Por qué debo pensar en sistemas y no solo en modelos?
- ¿Qué es la deuda técnica?
- Análisis de paper Hidden technical debt in machine learning systems
- ¿Qué es el drift y cómo abordarlo? se entregan ejemplos
- ¿Qué es el sesgo y cómo abordarlo? se entregan ejemplos Análisis de paper Temporal quality degradation in AI models Gestión operacional de sistemas de ML
 - ¿Cómo monitorear mi sistema?
 - ITIL: Un framework para abordar incidentes en sistemas de ML

Módulo 9: Fundamentos de infraestructura para la construcción de sistemas con resiliencia

- ¿Qué es una API, FTP? ¿en qué se diferencian? ¿Cuándo se usan?
- ¿Qué es un Data Lake y un Data Warehouse? ¿En qué se diferencian? ¿Para qué se usan?
- ¿Cuáles son los tipos de bases de datos más comunes? ¿Cuáles son las características, ventajas y desventajas de cada uno?
- ¿Cuáles son los motores de bases de datos más comunes? ¿Cuáles son las características, ventajas y desventajas de cada uno?
- ¿Qué es un ambiente de trabajo?
- ¿Qué es docker y para qué se usa?
- ¿Qué es Kubernetes y para qué se usa?
- Lectura de paper sobre la importancia en la creación de herramientas para crear áreas con alta performance usando el paper Nailing Prediction: Experimental Evidence on the Value of Tools in Predictive Model Development

Metodología:

Introducción al uso de datos para generar valor en la industria aseguradora, es un programa teórico/práctico basado en clases expositivas, donde el profesor con experiencia práctica en las temáticas que expone presenta los temas y genera discusión en clases y la aplicación a través de un trabajo grupal, que se debe exponer al finalizar el programa.

El rol del docente, durante el desarrollo del curso será de tutor y facilitador del trabajo individual y grupal, asegurando que los contenidos se pongan en acción y se generen aprendizajes significativos en los alumnos, vinculados a su desarrollo en el campo laboral.

Se desarrollarán:

- Cuestionarios y tareas sobre conceptos teóricos
- Participación en discusiones y simulaciones en clase
- Presentaciones de proyectos individuales y en equipo
- Ensayos de reflexión sobre el aprendizaje personal y las prácticas ágiles

Requisitos de Aprobación:

- El programa se aprueba con calificación igual o superior a nota 4 en escala de 1 a 7
- Asistencia superior al 75%

Calificaciones y su ponderación:

- Participación en Clases: tendrá una ponderación del 20%
- Prueba 1: tendrá una ponderación del 20%
- Prueba 2: tendrá una ponderación del 20%
- Prueba 3: tendrá una ponderación del 20%
- Ensayo final: tendrá una ponderación del 20%

Pruebas:

Las pruebas se realizan en línea y los resultados se entregan 1 semana después

Examen Final:

El examen final consistirá en un ensayo escrito. Recibirás la pregunta al menos una semana antes del examen y deberás entregar la solución escrita en un documento electrónico PDF. Los resultados finales estarán disponibles 2 a 3 semanas después de haber sido entregados.

Entregas con retraso: Al trabajo atrasado se le deducirá el 5% por cada período de veinticuatro horas que transcurran después de la fecha de vencimiento. Si circunstancias previstas o imprevistas le impiden completar una tarea a tiempo, puede solicitar una extensión. Las prórrogas deben solicitarse con antelación a la fecha límite. Si la situación justifica una extensión, determinaremos una nueva fecha de entrega para el ensayo según sus circunstancias individuales.

Plagio:

Presentar las ideas de otra persona como propias, ya sea textualmente o reformuladas con sus propias palabras, es una falta académica grave con graves consecuencias. Esto incluye el uso directo de inteligencia artificial (es decir, sin razonamiento propio sobre las ideas entregadas por la IA).