

**Curso y clave:**

**Ciencia de Datos Aplicada a las Finanzas**

**FZ4026.50**

**Nombre del Profesor:**

**Alfredo Ramírez García**



|  |  |
| --- | --- |
| **Nombre del curso:** | Ciencia de Datos Aplicada a las Finanzas |
| **Ubicación y hora del curso:** | EGADE Business School, Monterrey Lunes 6:00 pm a 9:00 pm |
| **Profesor:** | Alfredo Ramírez García |
| **Correo electrónico:** | alfredo.ramirez@tec.mx |
| **Teléfono:** | 55 1938 7663 |
| **Apoyo fuera de clase:** | En cualquier horario, con previa cita. |

Descripción del curso:

En este curso estudiaremos la metodología para la implementación de soluciones de ciencia de datos para la toma de decisiones basada datos (data-driven). Durante el curso desarrollará a detalle cada una de las fases de la metodología de ciencia de datos, que comprende desde la identificación del problema de negocio hasta la implementación y mejorar de los algoritmos seleccionados, se revisarán temas relacionados con la manipulación y visualización de grandes bases de datos y se explicarán y pondrán en práctica los modelos de mayor utilización en la industria para predicción y clasificación de datos.

Este curso tiene como propósito que el estudiante aprenda los fundamentos de la aplicación de técnicas y algoritmos de ciencia de datos desarrolladas para la identificando patrones o relaciones aparentes y ocultas en los datos que permitan generar valor para el negocio, que el alumno desarrolle la habilidad de identificar problemas reales de la empresa que puedan ser resueltos a través un proceso de ciencia de datos, que comprenda que para lograr implementaciones de ciencia de datos exitosas se requiere trabajo colaborativo interdisciplinario y a través de la cadena de valor del negocio.

Objetivos de curso:

Al finalizar el curso el alumno será capaz de:

* Explicar la importancia de la Ciencia de Datos en los modelos de negocios y los mercados financieros.
* Desarrollar su propia metodología de ciencia de datos.
* Manejar y visualizar bases de datos de grandes dimensiones.
* Conocer y poner en práctica las principales técnicas, algoritmos y modelos para clasificar, predecir o agrupar datos.
* Formular problemas, implementar modelos, y realizar recomendaciones

Metodología de aprendizaje:

* Clases semanales
* Actividades en línea individuales o grupales.
* Lectura semanal del material de IBM Skills Academy y otras fuentes.
* Laboratorios individuales o en grupo.
* Elaboración de reportes individuales o en grupo.

Responsabilidades del alumno y participación en clase:

* Leer el material de IBM Skills Academy previo a la clase.
* La participación en clase es muy importante para el desarrollo de los laboratorios, casos y ejercicios que se desarrollarán durante las sesiones.
* Es importante que los alumnos con conocimientos previos de programación tengan la iniciativa de apoyar a sus compañeros de otras disciplinas en estos temas.
* Profundizar los temas vistos en clases de manera individual.
* La asistencia y puntualidad son importantes, se requiere 80% de asistencia a clase.

Reglas de comportamiento en el salón de clase:

El alumno deberá observar y cumplir el reglamento académico del EGADE y el Tecnológico de Monterrey, sobre todo el apartado que se refiere a alumnos de posgrado.

Honestidad académica:

Se observará que se cumpla el acuerdo de ética que los alumnos y el personal adquieren al entrar a la institución.

Asesoria complementaria:

El alumno cuenta con toda la asesoría complementaría que requiera para entender a profundidad los temas revisados en clase, de manera que logre el máximo aprovechamiento de la materia y desarrolle una base sólida de conocimiento de ciencia de datos que facilite su proceso de aprendizaje.

Criterios de evaluación:

| *Criterio* | *Porcentaje* |
| --- | --- |
| Actividades previas | 10% |
| Actividades en equipo | 35% |
| Skills Academy | 30% |
| Proyecto Final (Equipo) | 25% |

Entrega de reportes, laboratorios, etc.:

Todas las entregas:

* Deben incluir nombre, número de cuenta, título y nombre de la actividad.
* Se realizarán de manera online.
* Se deben entregar en tiempo para ser calificados sobre la escala de 100 pts. Se recibirán trabajos posteriormente a la fecha de la entrega con una penalización de 20 pts.

Bibliografía sugerida:

- Provost, Foster, Data science for business, Sebastopol, Calif. : O'Reilly, 2013, 9781449361327

- Witten I. H, Data mining : practical machine learning tools and techniques, Fourth edition., Cambridge, MA : Morgan Kaufmann, ©2017, spa, 9780128042915

- Durante el curso se revisan temas de diferentes ramas de la Estadística, Matemáticas, Finanzas, Machine Learning; por lo que se recomendará material específico para cada tema visto durante las sesiones.

Material de apoyo:

-IBM Skills Academy: Data Science Practitioner Course.

-Garcia Márquez P. Y B. Lev (2019) Data Science and Digital Business. Springer.

- Kotu, V. Y B. Desphande (2019) Data Science. Morgan Kaufman.

- Vander Plas, J. (2016) Python Data Science Handbook. O´Reilly

- Daddy M. (2019) Business Data Science. McGraw-Hill.

- Kelleher, J. y B. Tierney (2018) Data Science. MIT Press.

Antecedentes académicos y profesionales de los Profesores:

El profesor Alfredo Ramírez García es candidato a Doctor en Ciencias Financieras por EGADE Business School, cursó el programa MEDEX de IPADE Business School y es Ingeniero en Sistemas. Se ha desempeñado como profesor de cátedra en temas de ciencia de datos en EGADE Business School desde el año 2021 y como director de la Maestría en Finanzas de la sede Santa Fé desde junio de 2023.

En el ámbito profesional cuenta con 17 años de experiencia en la industria de TI, entre 2005 y 2014 trabajó para RedIT, empresa proveedora de servicios administrados y centros de datos, como ingeniero de operaciones y gerente de seguridad de la información y continuidad de negocio. Entre 2015 y 2021 laboró en Pink Elephant, en donde se desempeñó como consultor de arquitectura de TI en el desarrollo de proyectos de continuidad de negocio, atendiendo principalmente empresas del sector financiero.

El Dr. del Castillo es profesor de economía y negocios en el Tec de Monterrey, Campus Monterrey desde 2020. Además, ha sido profesor en programas de educación ejecutiva en el Tecnológico de Monterrey. Se incorporó a EGADE Business School como profesor e investigador de tiempo parcial en 2022. Su trayectoria profesional incluye más de diez años de experiencia en investigación e inteligencia de negocios.

Su investigación académica se ha publicado en prestigiosas revistas, como Journal of Product & Brand Management, Retailing & Consumer Services y Economic Research. Además, el Dr. del Castillo ha presentado sus investigaciones en congresos nacionales e internacionales como el FIMEF (Congreso de Investigaciones Financieras), ACACIA (Academia de Ciencias Administrativas), IAM (Academia Iberoamericana de Management), SMLA (Strategic Management in Latin America), CERALE (Centre D'Etudes Et de Recherche Amérique Latine Europe) y WEAI (Western Economic Association International).

El Dr. del Castillo tiene un Ph.D. en Ciencias Administrativas, una maestría en economía de la Escuela de Gobierno y Políticas Públicas del Tecnológico de Monterrey y una licenciatura en economía.

El Ing. Carlos Crespo Elizondo, MSF, es graduado de la Maestría en Ciencias Financieras en Stuart School of Business del Instituto Tecnológico de Illinois. Desde su formación, ha liderado innovaciones en el ámbito de finanzas computacionales, enfocando su talento hacia la construcción de portafolios de inversión cuantitativos y el algotrading, tanto de activos bursátiles como de criptomonedas. Ha brindado asesorías y capacitaciones a casas de bolsa, enriqueciendo sus operaciones algorítmicas y operativas. Esta pasión por las finanzas computacionales también lo llevó a incursionar en el desarrollo de modelos de Business Intelligence, proporcionando soluciones vanguardistas para empresas de tecnología.

Desde 2017, el Ing. Crespo, ha compartido sus conocimientos a través de la docencia, impartiendo cursos de AlgoTrading a nivel postgrado y desde 2022, la clase de Ciencias de Datos Aplicadas a Finanzas. Además de su labor educativa, su aporte a la comunidad financiera ha sido evidente con la creación de la librería de Python "dabu", una herramienta especializada para el mercado de capitales de la Bolsa Mexicana de Valores, que ha beneficiado a múltiples profesionales al proporcionar acceso a Estados Financieros trimestrales y precios de cierre e intradía.

Estructura del curso:

| Sesión | Día | Profesor | Tema general | Temas particulares |
| --- | --- | --- | --- | --- |
| 1.1 | 11/09 | Alfredo Ramírez | Presentación del curso.  Metodología Ciencia de Datos | Presentación.  Explicación del curso.  Creación de cuentas IBM Cloud.  Creación de cuentas de IBM Skills Academy.  Explicación del Material y la Certificación.  Identificación del Problema.  Mención de la importancia de la estadística en las finanzas.  Revisión de las fases de una metodología formal de ciencia de datos. |
| 1.2 | 11/09 | Alfredo Ramírez | Implicaciones de la IA en la estrategia de los negocios. | Adopción de modelos de IA para la construcción de una ventaja competitiva.  Data-driven vs AI-driven  Inteligencia Colaborativa  Metodología de Ciencia de Datos |
| 2.1 | 18/09 | Alfredo Ramírez | Aplicaciones de los modelos supervisados de machine learning | Aplicación de modelos de machine learning supervisados para la creación de credit-scoring.  Regresión lineal.  Regresión logística |
| 2.2 | 18/09 | Alfredo  Ramírez | Caso Lending Club Credit Scoring | Descripción del modelo de Negocio de Lending Club  Exploración del Set de Datos de Crédito de Lending Club |
| 3.1 | 25/09 | Alfredo Ramírez | Caso Lending Club Credit Scoring | Transformación y preparación del set de datos  Exploración del diccionario de datos  Selección de features para ajuste de modelo |
| 3.2 | 25/09 | Alfredo Ramírez | Caso Lending Club Credit Scoring | Creación de los sets de entrenamiento y de pruebas  Aplicación de k-fold cross validation.  Entrenamiento de los modelos (regresión logística)  Análisis de la matriz de confusiones  evaluación de medidas de desempeño de los modelos |
| 4.1 | 02/10 | Alfredo Ramírez | Modelos no lineales | Regresión polinomial  Splines  Smoothing Splines |
| 4.2 | 02/10 | Alfredo Ramírez | Modelos no lineales | Laboratorio de modelos no lineales. |
| 5.1 | 09/10 | Carlos Crespo | Algorithmic Trading | Análisis Técnico  Que es Análisis Técnico  Análisis Técnico: Objetivo similares al Análisis Fundamental?  Las bases del Análisis Técnico: La teoría de DOW.  Las diferentes gráficas del análisis técnico  Que son la Tendencia, Cambio de Tendencia y  Patrones de Continuidad? |
| 5.2 | 09/10 | Carlos Crespo | Algorithmic Trading | Análisis Técnico  Promedios móviles  Bandas de Bollinger  Serie Fibonacci |
| 6.1 | 16/10 | Carlos Crespo | Portfolio Management | Teoría de portafolios  Markowits y la selección del portafolio  Optimización del riesgo y retorno |
| 6.2 | 16/10 | Carlos Crespo | Portfolio Management | Construcción de portafolio cuantitativo |
| 7.1 | 23/10 | Carlos Crespo | Cryptocurrencies y Blockchain | Monedas digitales y el anonimato  Criptografía y las criptomonedas  Descentralización  Mecánica del Bitcoin |
| 7.2 | 23/10 | Carlos Crespo | Cryptocurrencies y Blockchain | Trabajando con un blockchain  Block Hashing  Llave pública y llave privada |
| 8.1 | 30/10 | Carlos Crespo | Cryptocurrencies y Blockchain | Sistemas de pago  Mercados |
| 8.2 | 30/10 | Carlos Crespo | Cryptocurrencies y Blockchain | Algotrading en criptmonedas  Establecer conexión con un exchange  Lectura de información histórica y real através de API’s y/o WebSockets  Aplicación de herramientas de análsis técnico para generar ordenes de compra venta |
| 9.1 | 06/11 | Ernesto del Castillo | Procesos Estocásticos | Definición de proceso estocástico.  Funciones de densidad y funciones de distribución.  Definición de un proceso estacionario y no estacionario.  Raíz unitaria.  Dickey-Fuller Test.  Procesos diferenciados y no diferenciados (integrada). Series de ejemplo para observar estacionariedad (acción, indice, moneda, criptoactivo, commoditie, serie económica,etc.) |
| 9.2 | 06/11 | Ernesto del Castillo | Introducción a las series de tiempo | Series de poca y alta memoria.  Autocorrelación.  Identificar estructura en la serie:  -Estadístico Q\*  -ACF / PACF Definición e interpretación |
| 10.1 | 13/11 | Ernesto del Castillo | Laboratorio de Series de Tiempo | Modelo AR.  Modelo MA.  Modelo ARMA / ARIMA. |
| 10.2 | 13/11 | Ernesto del Castillo | Series de tiempo Estacionales | Modelo SARIMA.  Modelo STL: ARIMA, estacionalidad, ajuste distribución. |
| 11.1 | 20/11 | Ernesto del Castillo | Laboratorio Series de tiempo estacionales | Modelo SARIMA.  Modelo STL. |
| 11.2 | 20/11 | Ernesto del Castillo | Ajuste de distribución a series diferenciadas. | Práctica de diferenciación.  Gráfica series niveles, diferencia, densidad, histogramas.  Ajuste de la distribución a una serie de datos.  Pruebas de bondad de ajuste Kolmogorov-Smirnov.  Prueba de Normalidad Jarque-Bera. |
| 12.1 | 27/11 | Ernesto del Castillo | Simulación Monte Carlo | Modelos paramétricos.  Definición de simulación Monte Carlo.  Estados de la naturaleza.  Números pseudo-aleatorios, diferentes motores.  Ejemplos de problemas de metodología, integración por Monte Carlo.  Laboratorio simulación Monte Carlo (estacionariedad, ajuste, simulación, reescalar al precio). |
| 12.2 | 27/11 | Ernesto del Castillo | Simulación Bootstrap | Modelos No paramétricos.  Definición del modelo bootstrap (técnicas de remuestreo).  Laboratorio simulación bootstrap. |