

DATA SCIENCE AND MACHINE LEARNING APPLIED TO FINANCIAL MARKETS

Modulo III

CIENCIA DE DATOS

Profesor: Fernando Ortega Camargo

Objetivo

En este módulo, los participantes explorarán en profundidad las herramientas fundamentales y los conceptos clave de la ciencia de datos, con un enfoque práctico orientado al análisis y la comprensión de datos financieros. El objetivo principal es brindar a los asistentes una base sólida que les permita aplicar metodologías modernas de ciencia de datos en contextos reales, especialmente en entornos financieros donde la calidad, volumen y estructura de los datos representan desafíos únicos.

A lo largo del curso, se abordarán técnicas de análisis exploratorio, visualización de datos multivariados, limpieza y transformación de información, así como herramientas estadísticas básicas que permiten extraer patrones significativos y detectar anomalías o comportamientos no evidentes a simple vista. Se hará especial énfasis en el uso de bibliotecas especializadas de Python como Pandas, NumPy, Matplotlib, Seaborn y SciPy, que son ampliamente utilizadas en la industria financiera para el procesamiento de grandes volúmenes de información.

Temario

1 BASES DE DATOS Y SQL

1.1. Fundamentos de las bases de datos relacionales

- Que es SQL

- Que son las bases de datos Relacionales

- Componentes de las bases de datos

- Sistemas de Gestion de bases de datos

- MySQL y Workbench

1.2 Creación de bases de datos

- Diagrama entidad Relación

- Llaves primarias y foráneas

- Cardinalidad

- Normalización

1.3 Consultas para el análisis de datos

- Consultas básicas en SQL

- JOINS

- Consultas intermedias

- Consultas avanzadas

2 PYTHON CIENTÍFICO

2.1. Herramientas científicas con Scipy

- Integración numérica

- Interpolación numérica

- Ajuste de curvas

- Raíces de polinomios

- Solución de sistemas lineales

- Método de Newton Raphson

2.2. Optimización

- Máximos y mínimos

- Descenso de gradiente

3 ANÁLISIS DE DATOS

3.1. End to end en ciencia de datos

3.2 Análisis del problema

- Planteamiento y objetivo
- Selección de métricas de desempeño
- Validación de asunciones

3.3 Exploración de datos

- Características estructurales de la información
- Complejidad del dataset
- Importancia del test en el ML
- Correlaciones de atributos
- Combinaciones de atributos

3.4 Preparación de datos

- Tratamiento de NA's
- Datos categóricos
- Escalamiento de variables
- Custom Transformers
- Pipelines
- Desbalance de datos (SMOTE, ROS, ADASYN)

4 ANÁLISIS DEL LENGUAJE NATURAL

4.1 Expresiones Regulares

4.2 Introducción al procesamiento del lenguaje

- Que es NLP y como se aplica en ciencia de datos
- Tipos de tareas
- Corpus y tokens

4.3 Representación del lenguaje : Texto a vectores

- Tokenización
- Limpieza de texto: lowercasting, stopwords, stemming, lemmatization
- Busca de patrones con regex
- N-Grams
- TF-IDF: importancia de términos en corpus

4.4 Análisis de sentimiento

- Fundamentos
- Visualización de polaridad: TextBlob, Vader

4.5 Modelos probabilísticos clásicos

- Naive Bayes aplicado a clasificación de texto

4.6 Fundamentos de secuencias

- Por que el texto es secuencial
- Limitaciones de bolsas de palabras y ngrams
- Introducción a Word embeddings
- Introducción a modelos secuenciales
- Introducción a la atención y transformers

5 WebScraping

5.1 Fundamentos de WebScraping

- Que es? Ventajas y desventajas.
- Estructura de HTML, CSS y el DOM
- Identificación de etiquetas y estructura jerárquica
- Lenguaje Xpath

Ética en el webscraping

- Uso de time sleep, user agents.
- Uso ético y limites legales

5.2 Web Scraping con Python

- Uso de time sleep, user agents.
- Uso ético y limites legales
- Método request y parseo de información lxml
- Búsqueda de información por clases y atributos
- Extracción y limpieza de información
- Acopio de información
- Extracción de texto, números y enlaces
- Limpieza de información con regex
-

5.3Automatización de sitios dinámicos

- Selenium y diferencia con request
- localización de elementos
- Interacción con la web

6 ANÁLISIS DE DATOS FINANCIEROS

6.1 ¿Qué son los datos financieros?

- Carga de datos
- Limpieza de datos

6.2 Exploración y visualización

- Visualización de precios, velas y retornos
- Medias móviles
- Volatilidad histórica
- Visualización de correlaciones entre activos

6.3 Retornos logarítmicos VS simples

- Retornos logarítmicos VS simples
- Distribuciones de precios y retornos
- Covarianza, correlación y portafolios
- Teoría de portafolios

6.4 Series de tiempo financieras

- Estacionariedad, diferenciación y autocorrelación
- Modelos ARIMA
- Tests estadísticos (Dickey Fuller, ACF/PACF)
- Forecasting

Curso

Horarios: Clase 7: 10 pm – 10 pm
Descanso : 8:30 pm – 8:45 pm
Inicio : 08/09/25 Fin: 13/10/25

Evaluación: Asistencia 20% / Exámen Final 80% (última clase)

Adicional: Tareas 20% Extra (5 Papers – Código Github)

Envío de tareas: ortcamfer@gmail.com

Asunto:

Modulo3_ITAM_NombreAA_NombreTarea

Modulo3_ITAM_FernandoOC_TareaBases

Github: <https://github.com/forteg>