

Machine learning para las finanzas

Tecnología e Informática | Tecnología

Descripción del Curso

Este curso de Tecnología está diseñado para estudiantes a partir de 17 años que desean explorar el diseño, la ejecución y la comunicación de proyectos de aprendizaje automático (ML) aplicados a escenarios reales o simulados. A lo largo de 4 semanas, los alumnos desarrollan un proyecto completo que va desde la definición del problema y la selección de métricas, hasta la construcción de un pipeline de ML, su entrenamiento y evaluación, y la presentación de resultados con recomendaciones estratégicas para un negocio. El enfoque es práctico y orientado a la aplicación, con énfasis en la toma de decisiones fundamentadas, la documentación clara y la capacidad de comunicar hallazgos a audiencias técnicas y no técnicas. Contenido de Actividades:

- **Actividad 1: Propuesta de proyecto** Elaboración de una propuesta de proyecto con problema, datos y métricas.
- **Actividad 2: Construcción del pipeline** Implementación de un pipeline sencillo, entrenamiento y evaluación en un conjunto de datos real o simulado.
- **Actividad 3: Presentación final** Presentación oral y entrega de informe con recomendaciones y consideraciones estratégicas.

Objetivo:

- Capacidad para diseñar un proyecto completo de ML, justificar decisiones y comunicar resultados de forma profesional.
- Calidad de la documentación, claridad de la presentación y relevancia de las recomendaciones para el negocio.

Especificaciones:

4 semanas

Competencias

- Analizar problemas reales y definir abordajes de ML adecuados, considerando datos, métricas y viabilidad.
- Diseñar, implementar y evaluar un pipeline de ML simple, con interpretabilidad y control de calidad de resultados.
- Justificar decisiones técnicas y operativas con argumentos claros y fundamentados en datos.
- Comunicar resultados de manera efectiva: informes técnicos, presentaciones orales y visualización de resultados para audiencias diversas.
- Trabajar de forma colaborativa, gestionar tiempos y recursos, y aplicar buenas prácticas de documentación y trazabilidad.
- Desarrollar pensamiento crítico, ética y responsabilidad en el manejo de datos y modelos en contextos empresariales.

Requerimientos

- Conocimientos básicos de programación (preferentemente Python) y conceptos de estadística aplicados a ML.
- Ordenador personal con acceso a internet y capacidad para ejecutar entornos de desarrollo de ML (por ejemplo Python, Jupyter, bibliotecas como scikit-learn, pandas, numpy).
- Entorno de control de versiones (Git) y disponibilidad para trabajar en las actividades programadas.
- Acceso a datasets reales o simulados para la construcción y evaluación del pipeline.
- Compromiso para realizar entregas en las fechas estipuladas y participar en las presentaciones finales.

Unidades del Curso

Unidad 1: Unidad 1: Introducción al Machine Learning y Finanzas

Objetivos de Aprendizaje

- Definir qué es machine learning, aprendizaje supervisado y no supervisado, y sus diferencias.
- Describir ejemplos de aplicaciones financieras donde ML puede aportar valor (predicción, detección de fraude, gestión de riesgo).
- Identificar los componentes de un flujo de trabajo de ML: datos, modelos, entrenamiento, evaluación e interpretabilidad.

Contenidos Temáticos

1. Conceptos básicos de ML y finanzas
 1. Descripción corta: Introducción a conceptos como modelo, entrenamiento, validación, overfitting y sesgo, y cómo se relacionan con decisiones financieras.
2. Aplicaciones financieras del ML
 1. Descripción corta: Casos de uso como predicción de precios, detección de fraude, gestión de riesgo y trading algorítmico a alto nivel.
3. Herramientas y entorno de desarrollo
 1. Descripción corta: Introducción a Python, notebooks, pandas y scikit-learn para comenzar a trabajar con ML en finanzas.

Unidad 2: Unidad 2: Preparación de datos financieros

Objetivos de Aprendizaje

- Identificar fuentes de datos financieros y sus características (formatos, frecuencias, calidad).
- Aplicar técnicas de manejo de valores faltantes (imputación) y detección de outliers.
- Crear y dividir conjuntos de datos en entrenamiento, validación y prueba, manteniendo la reproducibilidad.

Contenidos Temáticos

1. Calidad y fuentes de datos financieros
 1. Descripción corta: Tipos de datos (series temporales, datos de operaciones, microdatos) y consideraciones de calidad.
2. Limpieza de datos y manejo de valores faltantes
 1. Descripción corta: Técnicas simples de imputación (media, mediana, moda) y enfoques más avanzados para series temporales.
3. Detección y manejo de outliers
 1. Descripción corta: Métodos basados en z-score, IQR y consideraciones de finanzas (outliers plausibles vs. atípicos).
4. Preparación de conjuntos de entrenamiento y prueba
 1. Descripción corta: División aleatoria y estratificada, importancia de la reproducibilidad y criterios de partición para series temporales.

Unidad 3: Unidad 3: Regresión para finanzas (predicción de precios o valores)

Objetivos de Aprendizaje

- Implementar regresión lineal y exploratoria para una variable financiera objetivo.
- Analizar la relación entre características y la variable objetivo, e interpretar coeficientes.
- Evaluar el rendimiento de modelos de regresión con métricas adecuadas (RMSE, MAE, R^2).

Contenidos Temáticos

1. Regresión lineal y relaciones financieras
 1. Descripción corta: Construcción de modelos lineales simples y evaluación de supuestos básicos (linealidad, homocedasticidad, normalidad de errores).
2. Regresión con características no lineales
 1. Descripción corta: Introducción a transformaciones y modelos simples no lineales para capturar relaciones complejas (polinomial, árboles simples).
3. Evaluación de modelos de regresión
 1. Descripción corta: Cálculo e interpretación de RMSE, MAE y R^2 ; comparación entre modelos para seleccionar el mejor.

Unidad 4: Unidad 4: Clasificación para finanzas (detección de fraude y señales)

Objetivos de Aprendizaje

- Aplicar técnicas de clasificación binaria (por ejemplo, regresión logística) a casos financieros simples.
- Comprender conceptos de desequilibrio de clases y métricas de rendimiento específicas de clasificación.
- Interpretar resultados de clasificación y considerar impactos operativos y de negocio.

Contenidos Temáticos

1. Clasificación binaria y regresión logística

1. Descripción corta: Construcción de un clasificador simple y evaluación de probabilidades y umbrales.

2. Métodos de clasificación adicionales

1. Descripción corta: Breve visión de árboles de decisión, Random Forest y métodos simples para necesidades financieras.

3. Evaluación de clasificación

1. Descripción corta: Matriz de confusión, precisión, recall y AUC como métricas clave en finanzas.

Unidad 5: Unidad 5: Evaluación de modelos y comparación de enfoques

Objetivos de Aprendizaje

- Identificar métricas de evaluación adecuadas para regresión y clasificación.
- Aplicar validación cruzada o particionamiento robusto para comparar modelos.
- Interpretar resultados de evaluación para tomar decisiones sobre el modelo final.

Contenidos Temáticos

1. Métricas de rendimiento para regresión

1. Descripción corta: RMSE, MAE y R^2 y cuándo utilizarlas en contextos financieros.

2. Métricas de rendimiento para clasificación

1. Descripción corta: Precisión, recall, F1 y AUC; interpretación en términos de costos y riesgos.

3. Comparación de modelos y selección

1. Descripción corta: Métodos para comparar modelos (validación cruzada, pruebas estadísticas simples) y criterios de selección.

Unidad 6: Unidad 6: Interpretación y comunicación de resultados

Objetivos de Aprendizaje

- Explicar la interpretabilidad de modelos simples y la importancia de la transparencia.
- Desarrollar presentaciones y reportes claros para diferentes públicos (técnico, directivo, cliente).
- Identificar limitaciones, supuestos y riesgos que acompañan a los resultados de ML en finanzas.

Contenidos Temáticos

1. Interpretabilidad y explicabilidad

1. Descripción corta: Métodos simples para entender qué influye en las predicciones y cómo comunicarlo (importancias de características, visualizaciones básicas).

2. Presentación de resultados

1. Descripción corta: Cómo estructurar informes y presentaciones para distintos públicos y objetivos de negocio.

3. Limitaciones y supuestos

1. Descripción corta: Discusión de supuestos del modelo, sesgos potenciales y escenarios en que el modelo podría fallar.

4. Visualización de resultados

1. Descripción corta: Uso de gráficos para comunicar tendencias, errores y comparaciones de modelos.

Unidad 7: Unidad 7: Ética, sesgos y mitigación en ML para finanzas

Objetivos de Aprendizaje

- Identificar fuentes de sesgo en datos y modelos financieros.
- Reconocer riesgos legales, regulatorios y de cumplimiento asociados a ML en finanzas.
- Proponer estrategias de mitigación, gobernanza de datos y evaluación de impacto ético.

Contenidos Temáticos

1. Sesgos y riesgos en ML financiero

1. Descripción corta: Tipos de sesgo, efectos en decisiones de negocio y clientes, y cómo pueden surgir en datos y modelos financieros.

2. Consideraciones éticas y cumplimiento

1. Descripción corta: Privacidad, transparencia, consentimiento y regulación aplicable a ML en finanzas.

3. Mitigación y gobernanza

1. Descripción corta: Prácticas de mitigación, auditoría, registros de decisiones y planes de respuesta ante fallos.

Unidad 8: Proyecto de ML para una situación financiera real

Objetivos de Aprendizaje

- Formular un problema financiero real como un problema de ML bien definido (objetivo, métricas y criterio de éxito).
- Seleccionar, preparar y justificar un conjunto de datos adecuado para el proyecto.
- Construir un pipeline básico de ML, entrenar y evaluar, y comunicar resultados y recomendaciones a stakeholders.

Contenidos Temáticos

1. Formulación del proyecto y datos

1. Descripción corta: Definir objetivo, hipótesis, métricas, alcance y riesgos; identificar fuentes de datos y restricciones.

2. Diseño de pipeline y modelado

1. Descripción corta: Selección de técnicas adecuadas, construcción de pipeline, entrenamiento y validación.

3. Evaluación y recomendaciones

1. Descripción corta: Análisis de resultados, limitaciones, impacto en decisiones financieras y recomendaciones finales.

4. Presentación del proyecto

1. Descripción corta: Preparación de informe y presentación a un comité, con explicación clara de datos, modelo y resultados.