



Guía para el profesor

Inteligencia Artificial y Machine Learning
MTFZ2210



Índice

Información general del curso	1
Metodología	2
Evaluación	4
Bibliografía	4
Tips importantes	5
Temario	5
Notas de enseñanza	5
Evidencia	8

Información general del curso

Modalidades

- Clave banner: MTFZ2210
- Modalidad: Connect

Competencia del curso

Determina los fundamentos del *machine learning* y la inteligencia artificial como herramientas fundamentales en las finanzas digitales para la toma de decisiones.





Metodología

Inserte la información aquí.

Este curso ha sido diseñado con la finalidad de ser impartido por un docente líder con experiencia en el ámbito laboral, quien compartirá contigo su conocimiento, experiencia y las mejores prácticas que realiza en su labor profesional.

La experiencia de curso promueve la interacción entre aprendedores de la Universidad Tecmilenio como una forma de enriquecer tu formación contrastando la realidad con la de otros compañeros.

Durante cada sesión, el docente transmite su experiencia y actúa como guía en el proceso de aprendizaje durante la realización de las actividades.

El curso es tetramestral y tiene una distribución semanal; en cada semana se lleva a cabo una sesión. La asistencia a estas sesiones es muy importante para aprendizaje.

Este curso se conforma por 8 temas, y su estructura es la siguiente:

Semana	Módulo	Tema	Evaluable
1	Módulo 1	Tema 1	
		Tema 2	Actividad 1
Tema 3			
Tema 4		Avance de proyecto final	
2			
3	Módulo 1	Tema 5	
		Tema 6	Actividad 2
Tema 7			
Tema 8		Proyecto final	
4			

Las sesiones se dividen en dos bloques, estas son las actividades que se recomiendan realizar:

Bloque 1	Bloque 2
Bienvenida y presentación de agenda.	Recapitulación de lo realizado en el bloque previo.
Actividad de bienestar.	Desarrollo de temas de la semana. <ul style="list-style-type: none"> • Explicación de los temas de la semana con ejercicios prácticos. • Cierre de temas.

Desarrollo de temas de la semana. <ul style="list-style-type: none">• Aplicación en contextos reales (introducción).• Explicación de los temas de la semana con ejercicios prácticos.	Explicación de las actividades que deberán realizarse en la semana (fuera de la sesión).
Receso.	



Evaluación

A continuación, puedes revisar el detalle de la evaluación de este certificado:

Evaluable	Ponderación
Actividad 1	15
Avance del proyecto final	15
Actividad 2	30
Proyecto final	40
Total	100



Temario

Módulo	Tema
Módulo 1	1. Introducción al machine learning
	2. Análisis de los datos
	3. Modelos de predicción con regresión
	4. Modelos de predicción con clasificación
Módulo 2	5. Modelos de predicción con base en redes neuronales
	6. Modelos de predicción con base en árboles de decisión
	7. Interpretación de modelos de aprendizaje automático
	8. Diseño de proyectos de machine learning y estrategias de selección de modelos



Notas de enseñanza

Tema 1 Introducción al machine learning

Notas para la enseñanza del tema:

1. **Contextualizar el tema:** Inicie la sesión con ejemplos cotidianos de inteligencia artificial (IA), como asistentes virtuales, recomendaciones de contenido en plataformas de streaming o aplicaciones de navegación.
2. **Fomentar la participación:** Plantee preguntas abiertas para que los estudiantes compartan experiencias personales con la IA. Esto ayudará a conectar la teoría con la práctica.
3. **Explicar los conceptos clave con claridad:** Asegúrese de definir términos fundamentales como machine learning, aprendizaje supervisado y no supervisado, inteligencia artificial estrecha (ANI) e inteligencia artificial general (AGI).
4. **Utilizar materiales visuales y multimedia:** Apoye la explicación con infografías, videos y demostraciones interactivas que ilustren cómo funcionan los algoritmos de IA.

5. **Promover el pensamiento crítico:** Invite a los estudiantes a reflexionar sobre los beneficios y desafíos éticos de la IA, como la privacidad de datos y el sesgo algorítmico.
6. **Realizar actividades prácticas:** Proponga ejercicios en los que los estudiantes exploren herramientas básicas de IA o analicen casos de uso en diferentes industrias.
7. **Evaluar el aprendizaje de manera dinámica:** Aplique metodologías como debates, estudios de caso y proyectos aplicados para medir la comprensión del tema.

Tema 2 Análisis de los datos

Notas para la enseñanza del tema:

Este tema es fundamental para que los estudiantes comprendan cómo se genera, organiza, almacena y transforma la información en entornos digitales y de análisis de datos. Aquí algunas sugerencias clave para facilitar su enseñanza:

1. **Conecta teoría con práctica desde el inicio.** Usa ejemplos visuales como archivos .csv o capturas de pantallas de plataformas reales (SPSS, Python/Pandas) para que los estudiantes vean cómo se representa la estructura de los datos en distintos contextos.
2. **Explica la lógica detrás de las estructuras.** Aclara qué son las entidades, atributos, tipos de datos y relaciones entre tablas en bases relacionales. Usa diagramas entidad-relación para reforzar la comprensión.
3. **Relaciona OLTP vs OLAP.** Ayuda a distinguir entre el ambiente operativo y el analítico, y por qué ambos requieren estructuras de datos distintas. Utiliza ejemplos como solicitudes de crédito para hacerlo más tangible.
4. **Integra código con propósito.** Al presentar fragmentos en SQL o Python, explícalos paso a paso y muestra el resultado esperado. Resalta buenas prácticas en manipulación de datos: limpieza, transformación, detección de valores nulos y duplicados.
5. **Aborda los tipos de Join con ejemplos visuales.** Utiliza tablas pequeñas que permitan a los estudiantes ver claramente los efectos del inner, left, outer y cross join.
6. **Fomenta el pensamiento crítico.** Invita a los estudiantes a reflexionar sobre cómo una mala estructura de datos puede afectar decisiones, modelos predictivos o análisis de negocio.
7. **Utiliza datasets reales o simulados** para tareas prácticas. Idealmente, trabaja con ellos en clase para aplicar filtros, limpieza, transformación y detección de valores atípicos usando Pandas o SQL.

8. **Refuerza el concepto de *tidy data*.** Asegúrate de que comprendan cómo estructurar datasets para facilitar su análisis, visualización y modelado.
9. **Incluye contexto de aplicación.** Muestra cómo estas estructuras se usan en la industria financiera, en APIs o en sistemas web para conectar con áreas de interés profesional de los estudiantes.
10. **Cierra con una reflexión práctica.** Pide a los estudiantes identificar problemas comunes en la estructura de un dataset y proponer soluciones con base en lo aprendido.

Tema 3 Modelos de predicción con regresión

Notas para la enseñanza del tema:

El tema de modelos de regresión es clave para introducir a los estudiantes en el análisis predictivo y la toma de decisiones basadas en datos, especialmente dentro del campo financiero. Aquí algunas sugerencias para su enseñanza:

1. **Conecta con un caso real desde el inicio.** Utiliza el ejemplo de Rubén y su interés por relacionar edad con riesgo crediticio como punto de partida. Esto ayuda a contextualizar la regresión simple de manera clara y relevante.
2. **Explica los conceptos básicos con ejemplos visuales.** Usa diagramas de dispersión y gráficos de regresión para que los estudiantes identifiquen patrones y entiendan cómo se modelan relaciones entre variables.
3. **Aclara el rol de las variables.** Destaca la distinción entre variable dependiente e independiente, y refuerza este concepto con ejemplos prácticos como edad vs riesgo o rezagos en series de tiempo vs tasa de interés.
4. **Enseña con código paso a paso.** Integra fragmentos de código Python y explica en clase cómo se construye el modelo desde cero: definición de la función, error, gradiente descendente y entrenamiento. Esto favorece tanto la comprensión como la práctica.
5. **Diferencia entre modelos lineales y no lineales.** Expón sus ventajas, limitaciones y niveles de interpretabilidad, y da ejemplos de cuándo conviene usar cada uno.
6. **Relaciona con el ámbito financiero.** Resalta casos de uso comunes como análisis de riesgo crediticio, predicción de tasas, valuación de activos y gestión de portafolios. Esto motiva a los estudiantes al ver la aplicabilidad del tema en el mundo real.
7. **Fomenta la experimentación.** Motiva a los alumnos a modificar el modelo de Ana: cambiar el número de rezagos, probar otros datos económicos o incluir más variables independientes para mejorar predicciones.

8. **Apoya la interpretación de resultados.** Enséñales a leer gráficas de error, curvas de predicción vs datos reales y cómo evaluar si un modelo está bien entrenado.
9. **Promueve el pensamiento crítico.** Pregunta: ¿Qué otras variables podrían influir en el riesgo crediticio? ¿Cuándo un modelo simple no es suficiente? ¿Cómo elegir un buen conjunto de variables?
10. **Cierra con un mini proyecto o simulación.** Pide a los estudiantes construir un modelo de regresión básico con datos reales del sector financiero o educativo, interpretarlo y presentar sus hallazgos.

Tema 4 Modelos de predicción con clasificación

Notas para la enseñanza del tema:

Este tema introduce herramientas fundamentales para predecir comportamientos categóricos en el ámbito financiero, como el riesgo crediticio o la retención de clientes. Aquí algunas sugerencias clave para una enseñanza clara y significativa:

1. **Conecta con un escenario realista desde el inicio.** Introduce el caso de Susana, la asesora de crédito, para motivar el uso de modelos de clasificación como apoyo en decisiones complejas. Esto da contexto práctico y ayuda a los estudiantes a visualizar la aplicación del tema.
2. **Explica claramente la diferencia entre regresión y clasificación.** Refuerza que, mientras la regresión predice valores numéricos, la clasificación se enfoca en etiquetas o categorías (por ejemplo: "cumple" o "no cumple").
3. **Introduce los modelos desde lo simple a lo complejo.** Comienza con la regresión logística como modelo lineal generalizado (GLM) y avanza hacia modelos no lineales como árboles de decisión, random forest o redes neuronales.
4. **Utiliza ejemplos visuales para explicar fronteras de decisión.** Muestra gráficamente cómo se separan las clases y cómo cambia la forma de la frontera con modelos más complejos.
5. **Enseña el uso e interpretación de la matriz de confusión.** Relaciona los cuatro posibles escenarios (TP, TN, FP, FN) con ejemplos del mundo financiero para que los estudiantes comprendan sus implicaciones en la toma de decisiones.
6. **Expón métricas de evaluación con claridad.** Ayuda al grupo a entender qué significan *accuracy*, *precision*, *recall* y *F1 score*, y cuándo es mejor priorizar una métrica sobre otra, especialmente en conjuntos de datos desbalanceados.
7. **Introduce el concepto de Weight of Evidence (WoE).** Usa el ejemplo de segmentación por edad para explicar cómo WoE transforma variables categóricas en valores numéricos útiles para modelos lineales.

8. **Demuestra el proceso paso a paso.** Desde la exploración de la variable objetivo hasta la transformación de datos, entrenamiento y evaluación del modelo. Apoya el proceso con fragmentos de código en Python que puedan ejecutar en clase.
9. **Destaca los desafíos comunes.** Aborda problemas como el desbalanceo de clases, la falta de datos históricos y la calidad de las variables independientes. Proporciona estrategias como oversampling o undersampling para enfrentarlos.
10. **Propón una actividad final aplicada.** Pide a los estudiantes construir un modelo de clasificación con un pequeño dataset real o simulado, aplicando todo el ciclo: preparación, modelado, evaluación y comunicación de resultados.

Tema 5 Modelos de predicción con base en redes neuronales

Notas para la enseñanza del tema:

Este tema introduce una de las herramientas más potentes del aprendizaje automático: el deep learning, mediante redes neuronales. Su aplicación práctica en variables económicas como tipo de cambio o tasas de interés es altamente relevante para la industria financiera. A continuación, algunas sugerencias clave para guiar tu clase:

1. **Comienza con la historia de José.** Usa su caso como introducción para generar empatía y contexto real: un profesional con experiencia operativa que quiere crecer hacia decisiones más estratégicas usando tecnología.
2. **Explica desde lo biológico hacia lo computacional.** Muestra la analogía entre neuronas reales y artificiales. Esto ayuda a los estudiantes a entender por qué se llaman "redes neuronales" y cómo funcionan de forma simplificada.
3. **Define los componentes esenciales del perceptrón.** Aclara qué son los nodos de entrada, funciones de activación, pesos y salidas. Usa ejemplos gráficos o animaciones si es posible.
4. **Construye el concepto de red neuronal multicapa (MLP).** Explica las capas ocultas y cómo agregarlas mejora la capacidad de modelado, aunque sacrifica interpretabilidad.
5. **Introduce el concepto de deep learning con claridad.** Muestra por qué las redes profundas (con muchas capas) son útiles para identificar patrones complejos, como el comportamiento del tipo de cambio.
6. **Explica los algoritmos clave: *feed-forward* y *back-propagation*.** No necesitas entrar en matemáticas avanzadas, pero sí en el flujo de información y ajuste de parámetros. Resalta que estos procesos ya están automatizados en librerías como TensorFlow o Keras.
7. **Muestra la arquitectura del modelo paso a paso.** Apóyate en el código para que los estudiantes comprendan cómo se prepara el conjunto de datos, se entrena la red y se generan las predicciones.

8. **Integra visualizaciones útiles.** Los gráficos de series de tiempo y de predicción vs valor real permiten validar visualmente el funcionamiento del modelo. Enséñales a interpretarlos correctamente.
9. **Enfatiza la utilidad del modelo en escenarios reales.** Por ejemplo, cómo una empresa puede anticipar el comportamiento del dólar para planear pagos a proveedores. Esto fortalece la conexión entre el modelo y su aplicación práctica.
10. **Cierra con una reflexión estratégica.** Motiva a los estudiantes a pensar cómo este tipo de modelos puede cambiar la manera en que se toman decisiones financieras, y qué competencias adicionales (programación, interpretación de datos, pensamiento crítico) necesitan desarrollar para aplicarlos profesionalmente.

Tema 6 Modelos de predicción con base en árboles de decisión

Notas para la enseñanza del tema:

Este tema es ideal para introducir a los estudiantes en la toma de decisiones basada en datos, especialmente en contextos de incertidumbre como el análisis de riesgo crediticio. Los árboles de decisión y sus variantes permiten modelar relaciones complejas entre múltiples variables, facilitando decisiones más informadas. Aquí algunas recomendaciones clave:

1. **Introduce el tema con el caso de Julián.** El rol de un analista de riesgos en una institución financiera es muy cercano a la vida profesional de los estudiantes. Utilízalo para mostrar cómo una herramienta de machine learning puede facilitar decisiones en entornos inciertos.
2. **Explica el concepto de árbol de decisión con ejemplos visuales simples.** Ayuda a los estudiantes a entender cómo se construyen los árboles a partir de divisiones sucesivas de datos, y la función de nodos raíz, nodos de decisión y hojas.
3. **Diferencia entre modelos de clasificación y regresión.** Usa ejemplos claros (como predecir si un crédito será exitoso o estimar el nivel de deuda) para ayudarles a identificar qué tipo de modelo se necesita según la variable objetivo.
4. **Demuestra cómo se construye un árbol paso a paso.** Utiliza un conjunto de datos pequeño, identifica las variables relevantes y muestra cómo se dividen los datos en función de los mejores puntos de corte (basados en Gini, entropía o error cuadrado medio).
5. **Incorpora práctica con código.** Los estudiantes deben ver cómo se entrena un árbol de decisión real en Python, desde la carga de datos, preparación, entrenamiento y evaluación con métricas como *accuracy*, *precision* y *recall*.

6. **Enseña los conceptos de *bagging*, *random forest* y *boosting*.** Explica sus ventajas: mayor generalización, menor overfitting y mejor rendimiento predictivo. Usa analogías (como el comité de expertos vs. un solo juez) para facilitar la comprensión.
7. **Muestra los resultados de forma comparativa.** Contrasta el rendimiento de un árbol individual con modelos ensemble para que los estudiantes valoren el impacto de estas técnicas en la precisión de las predicciones.
8. **Destaca la utilidad de los árboles en el sector financiero.** Subraya que, por su capacidad de manejar datos tabulares y su interpretabilidad, son ampliamente utilizados para análisis de riesgo, calificación de créditos y detección de fraudes.
9. **Promueve el pensamiento analítico y ético.** Haz preguntas como: ¿qué pasaría si una variable está sesgada? ¿cómo asegurar que el modelo no discrimine? ¿cómo validar que las predicciones se mantengan actualizadas?
10. **Cierra con una actividad práctica o un reto.** Pide que apliquen un modelo de árbol a un conjunto de datos real o simulado y que interpreten sus decisiones a partir del árbol generado.

Tema 7 Interpretación de modelos de aprendizaje automático

Notas para la enseñanza del tema:

Este tema es clave para que los estudiantes aprendan a **entender los resultados de los modelos de machine learning**, tanto en su forma más simple (regresión lineal) como en su forma más compleja (modelos no lineales), y cómo **comunicar estos resultados de forma clara y justificada** ante directivos, inversionistas o clientes.

A continuación, algunas recomendaciones para abordarlo en clase:

1. Empieza con el caso de Jaime

El contexto del director financiero enfrentando preguntas del consejo ayuda a mostrar por qué **interpretar modelos predictivos no es solo técnico, sino estratégico**. Usa esta historia para conectar el aprendizaje con la toma de decisiones en escenarios reales.

2. Reafirma la diferencia entre modelos lineales y no lineales

Utiliza gráficos simples para que los estudiantes comprendan:

- Cómo los modelos lineales generan predicciones con base en **combinaciones ponderadas** de variables independientes.
- Cómo los modelos no lineales (como redes neuronales o XGBoost) **capturan mejor la variabilidad** de los datos, pero a costa de **perder interpretabilidad**.

3. Explica por qué los modelos deben ser interpretables

Haz hincapié en que, aunque un modelo sea preciso, si no se puede explicar **cómo y por qué toma una decisión**, es menos útil en contextos empresariales, regulatorios o éticos.

4. Utiliza ejemplos numéricos y visuales

- Muestra cómo leer los coeficientes en una regresión lineal (por ejemplo: un aumento de 1 unidad en los ingresos \rightarrow +0.8 en ventas).
- Aplica ejemplos con variables numéricas, binarias y categóricas usando codificación *one-hot*.
- Enfatiza cómo **el signo y la magnitud de un coeficiente** nos dan información clara sobre el impacto de una variable.

5. Introduce el concepto de modelos explicativos globales vs. locales

- **Modelos lineales** son globales: los coeficientes aplican igual para todos.
- **SHAP** es local: la explicación se genera **caso por caso**, permitiendo entender cada predicción de forma específica.

6. Demuestra el uso del algoritmo SHAP con visualizaciones

- Utiliza el ejemplo del dataset de precios de casas para mostrar una **gráfica tipo waterfall** (aportación marginal de cada variable).
- Luego muestra la **gráfica tipo beeswarm** para ver el comportamiento general de las variables sobre todo el dataset.

Esto permitirá a los estudiantes **ver y comparar** cómo los diferentes modelos explican sus decisiones.

7. Haz énfasis en cómo se comunica una predicción explicada

Jaime necesita defender su modelo frente a un consejo. Enseña a los estudiantes a **convertir la salida del modelo en lenguaje de negocio**: por ejemplo, "El 65% del impacto en el cambio de utilidad proviene del ingreso medio de los clientes".

8. Promueve el análisis ético y la responsabilidad en modelos

¿Qué pasa si una variable como edad o género tiene un peso alto? ¿Cómo se justifica su uso? ¿Cómo se puede ajustar el modelo? Este es un buen momento para introducir conceptos como **equidad, sesgo algorítmico y responsabilidad social**.

9. Conecta la interpretación con la toma de decisiones bajo escenarios

Usa la parte final del caso para mostrar cómo **los escenarios hipotéticos cambian la entrada del modelo**, y cómo SHAP ayuda a anticipar qué variables tienen más peso y por qué.

10. Cierra con una práctica aplicada

Pide a los estudiantes interpretar los resultados de un modelo lineal y otro no lineal para un mismo problema (ventas, riesgo, crédito). Que expliquen cómo cada variable influye y comparen la claridad de los resultados en ambos enfoques.

Tema 8

Diseño de proyectos de machine learning y estrategias de selección de modelos

Notas para la enseñanza del tema:

Este tema ofrece una excelente oportunidad para conectar la teoría de machine learning con aplicaciones estratégicas reales en el sector financiero. Aquí algunas sugerencias para su desarrollo en clase:

- 1. Contextualiza el reto con el caso de Genaro**
Inicia la sesión presentando el caso de Genaro, destacando su papel como director financiero y su responsabilidad en validar modelos predictivos para el ERP del banco. Esto permitirá que los estudiantes comprendan la relevancia práctica del tema desde el inicio.
- 2. Fomenta el análisis crítico de modelos existentes**
Anima al grupo a reflexionar sobre qué hace que un modelo sea “veraz” en la toma de decisiones crediticias. Utiliza preguntas guía como:
 - ¿Cómo sabemos si un modelo aporta valor real al proceso?
 - ¿Qué implicaciones tiene elegir un modelo con sesgo alto o baja capacidad de generalización?
- 3. Subraya la importancia de la partición de datos**
Dedica un bloque a explicar visualmente el uso de conjuntos de entrenamiento, validación y prueba. Si es posible, realiza una actividad práctica o simulación sencilla con *scikit-learn* para ilustrar cómo afecta cada conjunto al rendimiento final del modelo.
- 4. Promueve la comparación entre modelos simples y complejos**
Haz énfasis en la importancia de partir con un modelo de referencia (baseline), como la regresión logística o lineal. Esto ayudará a justificar la elección (o no) de modelos más sofisticados como random forest o redes neuronales.
- 5. Introduce técnicas de optimización de hiperparámetros**
Muestra cómo usar Grid Search y Random Search, resaltando sus ventajas y limitaciones. Considera integrar una demostración en vivo o compartir un notebook con ejemplos.
- 6. Conecta con el diseño end-to-end de proyectos de ML**
Presenta el flujo completo de un proyecto de machine learning (desde la carga del dataset hasta la evaluación del modelo final) para que los estudiantes comprendan la dimensión técnica y operativa del proceso.
- 7. Reflexiona sobre la escalabilidad y gobernanza**
Cierra la sesión resaltando que la selección del modelo ideal para un ERP no solo depende de su rendimiento, sino también de su interpretabilidad, escalabilidad y mantenimiento a largo plazo.



Actividades y evidencias

A continuación, se presentan las instrucciones para las actividades y evidencias que los estudiantes deberán desarrollar a lo largo del curso. Estas indicaciones tienen como objetivo facilitar su orientación durante cada etapa del proceso formativo, asegurando que las evidencias recojan de manera clara los aprendizajes esperados. Se recomienda revisar cada actividad previamente y brindar acompañamiento oportuno a los estudiantes para resolver dudas y fomentar la participación activa.

Actividad 1

Objetivo de la actividad

Aplicar técnicas de aprendizaje supervisado para predecir el nivel de riesgo crediticio de clientes y evaluar su impacto en la toma de decisiones financieras.

Descripción

Los aprendedores asumirán el rol de analistas de riesgo en una institución financiera. Usarán un conjunto de datos proporcionado para entrenar un modelo de aprendizaje supervisado que clasifique a los solicitantes de crédito en categorías de riesgo (alto, medio o bajo). Posteriormente, interpretarán los resultados y propondrán estrategias para mejorar el modelo.

Instrucciones

1. **Exploración de datos:**

Haz clic [aquí](#) para descargar el *dataset*. Analiza las características principales de los datos, identificando valores faltantes, duplicados o atípicos. Preséntalo detallado con visualizaciones claras y precisas.

2. **Preparación del dataset:**

- Limpia los datos utilizando las herramientas aprendidas (por ejemplo, eliminación de duplicados, tratamiento de valores faltantes).
- Estandariza las variables relevantes y crea un "*tidy dataset*" con justificación técnica clara.

3. **Implementación del modelo:**

- Usa Python y bibliotecas como *scikit-learn* para entrenar un modelo de clasificación supervisada.
- Realiza un análisis comparativo entre diferentes algoritmos (por ejemplo, árbol de decisión, regresión logística).
- Documenta claramente el código y las decisiones técnicas.

4. **Evaluación de resultados:**

- Genera métricas como precisión, sensibilidad y especificidad para evaluar el desempeño del modelo.
- Interpreta los resultados, identificando fortalezas y limitaciones del modelo basado en los datos utilizados. Interpretándolas de manera crítica y relacionándolas con los datos.

5. Propuesta de mejora

Redacta un informe claro y profesional donde propongas mejoras al modelo o cambios en los datos que podrían optimizar el desempeño y las decisiones basadas en él.

Criterios de evaluación

Realiza la entrega de tu avance con base en los criterios de evaluación que se muestran en la rúbrica.

Entregable (s)

1. Notebook de Python con código comentado.
2. Informe de una página con el análisis de resultados y propuestas de mejora.

Actividad 2

Objetivo de la actividad

Comprender y aplicar los fundamentos de *machine learning* para diseñar un modelo predictivo básico que analice datos financieros y soporte la toma de decisiones.

Descripción

El aprendiz implementará un modelo de regresión lineal utilizando un conjunto de datos financieros simulado. A través del modelo, evaluará la relación entre factores como tasas de interés y niveles de riesgo para predecir escenarios financieros futuros.

Instrucciones

1. Utiliza el archivo simulado de datos financieros que contiene las variables: tasa de interés, nivel de riesgo, y rendimiento proyectado. Haz clic [aquí](#) para descargar.
2. Analiza estadísticamente el conjunto de datos utilizando una herramienta como Python o Excel para identificar patrones y relaciones.
3. Utiliza una librería como Scikit-learn (Python) para implementar un modelo de regresión lineal.
 - Entrena el modelo con el 70% de los datos.
 - Usa el 30% restante para validar el modelo.
4. Calcula métricas como el error cuadrático medio (MSE) y la precisión del modelo.
5. Presenta un análisis profundo de cómo las variables analizadas influyen en el rendimiento proyectado y plantea cómo estos resultados pueden influir en decisiones financieras.
6. Incluye gráficos claros y bien diseñados que visualizan correctamente los patrones en los datos y las predicciones del modelo. Selecciona adecuadamente las visualizaciones y apoya la interpretación y análisis de los resultados.

Criterios de evaluación

Realiza la entrega de tu avance con base en los criterios de evaluación que se muestran en la rúbrica.

Entregable (s)

- Código fuente utilizado.
- Reporte en formato PDF que incluya:
 - Gráficos relevantes de los datos y predicciones.
 - Métricas del modelo y su interpretación.
 - Reflexiones sobre cómo los resultados pueden apoyar decisiones financieras.

Evidencia 1

Objetivo de la actividad

Aplicar técnicas de *machine learning* para construir y evaluar modelos predictivos que permitan analizar el riesgo crediticio de clientes en una institución financiera.

Descripción

En esta evidencia, el aprendedor deberá explorar un conjunto de datos financieros y crear tres tipos de modelos predictivos: uno basado en regresión lineal simple, otro en regresión lineal multivariable y un modelo avanzado no lineal. El propósito es identificar patrones en las variables y evaluar cuál de los modelos es más efectivo para predecir el nivel de riesgo crediticio.

Instrucciones

Antes de comenzar, crea un conjunto de datos ficticio siguiendo estas instrucciones, haz clic [aquí](#) para descargar.

- Este código generará un archivo llamado `credit_risk_dataset.csv` con 1000 observaciones y 5 variables relacionadas con la evaluación de riesgo crediticio.
 1. **Análisis exploratorio inicial:**
 - Realiza un análisis exhaustivo exploratorio de los datos para identificar variables relevantes, detectar valores atípicos o faltantes y entender las relaciones entre las variables.
 - Genera visualizaciones (gráficas de dispersión, histogramas, correlaciones) para apoyar tu análisis.
 2. **División del conjunto de datos:**
 - Divide el conjunto de datos en dos subconjuntos: entrenamiento (70%) y prueba (30%).
 - Justifica claramente por qué esta división es importante para la validación de los modelos.
 - Registra detalladamente a los tamaños de los subconjuntos y presenta el código utilizado.
 3. **Modelo de regresión lineal simple:**
 - Escoge y justifica una variable independiente significativa para predecir el nivel de riesgo.
 - Entrena un modelo de regresión lineal simple utilizando el conjunto de entrenamiento.

- Evalúa el modelo calculando el error cuadrado medio (MSE) en los conjuntos de entrenamiento y prueba.
 - Justifica la selección de la variable independiente con investigación propia.
4. **Modelo de regresión lineal multivariable:**
- Entrena un modelo multivariable utilizando todas las variables relevantes del conjunto de datos.
 - Calcula el MSE en ambos conjuntos y compara su rendimiento con el modelo simple.
 - Indica si se necesitaron transformaciones adicionales de los datos y describe el proceso.
5. **Modelo avanzado no lineal:**
- Selecciona un modelo no lineal del catálogo de scikit-learn (por ejemplo, Random Forest o Support Vector Machines).
 - Justifica la selección del modelo con una breve investigación sobre su funcionamiento.
 - Entrena el modelo y evalúa su rendimiento en ambos conjuntos de datos.
6. **Comparación y conclusiones:**
- Genera una tabla comparativa con los resultados de los tres modelos (MSE de entrenamiento y prueba).
 - Indica cuál modelo sería más útil para una institución financiera y por qué.

Criterios de evaluación

Realiza la entrega de tu avance con base en los criterios de evaluación que se muestran en la rúbrica.

Evidencia final

Objetivo de la actividad

Profundizar en la interpretación de modelos predictivos aplicados en el ámbito financiero, integrando herramientas como SHAP para mejorar la comprensión de los resultados generados por algoritmos avanzados y su impacto en decisiones estratégicas.

Descripción

Esta evidencia el aprendiz utilizará modelos de aprendizaje automático previamente desarrollados en la **evidencia 1** y aprenderá a interpretar sus resultados de manera efectiva,

considerando las relaciones entre variables y resultados. Esto incluye aplicar herramientas como SHAP y realizar análisis de sensibilidad para explorar cómo las variaciones en los insumos afectan las predicciones del modelo.

Instrucciones

1. Revisión de fundamentos

- Elabora un informe que explique con claridad cómo interpretar modelos lineales y no lineales en *machine learning*, destacando los métodos y herramientas disponibles como SHAP.
- Proporciona ejemplos de interpretación en contextos financieros, como la evaluación de riesgos o la predicción de utilidades.
- Bien fundamentados y con lenguaje técnico adecuado.

2. Definición de un escenario financiero

- Selecciona uno de los escenarios prácticos relacionados con la interpretación de resultados, como:
 - Análisis de impacto de variables macroeconómicas en el rendimiento financiero.
 - Identificación de factores clave en la clasificación de clientes.
 - Evaluación de la contribución de variables en un modelo de predicción de riesgos.
- Describe el escenario y por qué es crítico para la toma de decisiones.

3. Implementación de herramientas de interpretación

- Aplica SHAP para analizar un modelo predictivo previamente desarrollado (o uno simple creado para este fin).
- Genera visualizaciones como gráficos de cascada y enjambre para mostrar la contribución de cada variable independiente al resultado del modelo.

4. Análisis de sensibilidad

- Realiza un análisis exhaustivo que explore cómo los cambios en las variables independientes afectan las predicciones.
- Usa gráficos comparativos para ilustrar los impactos y discutir las implicaciones financieras.

5. Informe técnico

- Presenta un informe técnico bien estructurado, con gráficos relevantes, interpretaciones claras y conclusiones sólidas que relacionan los hallazgos con el escenario planteado.

6. Presentación y reflexión

- Prepara una presentación visual que explique los resultados del análisis, destacando la importancia de interpretar correctamente los modelos para la toma de decisiones.

- Reflexiona sobre las limitaciones de las herramientas de interpretación y propone mejoras o ajustes futuros.

Criterios de evaluación

Realiza la entrega de tu avance con base en los criterios de evaluación que se muestran en la rúbrica.

Entregable (s)

1. Informe inicial: Documento en PDF (2-3 páginas) explicando los fundamentos y ejemplos de interpretación de modelos en finanzas.
2. Código fuente: Archivo en Python (.py o equivalente) con análisis de SHAP y sensibilidad.
3. Resultados del análisis: PDF (3-5 páginas) con gráficos e interpretaciones clave.
4. Presentación final: Archivo en PowerPoint o PDF (10 diapositivas) con resumen de metodología, resultados y conclusiones.



Anexos

Anexo 1. Rúbrica de Actividad 1

Criterios de evaluación	Nivel de desempeño			%
	Altamente competente 100%-86%	Competente 85%-70%	Aún sin desarrollar la competencia 69%-0%	
Exploración de datos	20-17	16-14	13-0	20
	Identifica correctamente valores faltantes, duplicados y atípicos, presentando un análisis detallado con visualizaciones claras y precisas.	Identifica valores faltantes, duplicados y atípicos, pero el análisis o las visualizaciones carecen de profundidad.	No identifica correctamente los problemas de los datos ni proporciona análisis significativo o visualizaciones.	
Preparación del dataset	20-17	16-14	13-0	20
	Limpia y estandariza los datos de manera precisa, creando un <i>dataset</i> ordenado (" <i>tidy dataset</i> ") con justificación técnica clara y detallada	Limpia y estandariza los datos, aunque con errores menores o justificación técnica poco detallada	No realiza la limpieza o estandarización adecuada de los datos o no entrega un " <i>tidy dataset</i> ".	
Implementación del modelo	25-22	21-18	17-0	25
	Entrena y compara al menos dos algoritmos con configuraciones óptimas, documentando claramente el código y las decisiones técnicas.	Entrena y compara algoritmos, pero no optimiza configuraciones o documenta parcialmente el código.	No implementa modelos funcionales o carece de documentación técnica adecuada.	
	20-17	16-14	13-0	

Evaluación de resultados	Genera métricas completas (precisión, sensibilidad y especificidad), interpretándolas de manera crítica y relacionándolas con los datos.	Genera métricas básicas y ofrece una interpretación adecuada pero no crítica o con escasa relación con los datos.	No genera métricas relevantes o no las interpreta adecuadamente.	20
Informe	15-13	12-11	10-0	15
	Presenta un informe claro y profesional, con propuestas innovadoras de mejora basadas en el análisis de los resultados.	El informe cumple con los requisitos mínimos, pero las propuestas de mejora son generales o poco fundamentadas.	No entrega un informe o las propuestas de mejora carecen de sustento técnico o relevancia.	
TOTAL				100%

Anexo 2. Rúbrica de Actividad 2

Criterios de evaluación	Nivel de desempeño			%
	Altamente competente 100%-86%	Competente 85%-70%	Aún sin desarrollar la competencia 69%-0%	
Análisis estadístico de datos	15-13	12-11	10-0	15
	Presenta un análisis estadístico exhaustivo, identificando patrones clave en las variables y justificando con gráficos claros y relevantes. Utiliza técnicas estadísticas avanzadas de manera precisa y contextualiza los resultados en el ámbito financiero.	Realiza un análisis estadístico adecuado, identificando algunos patrones relevantes y apoyándose en gráficos funcionales. Aunque correcto, el análisis podría incluir más profundidad o mayor conexión con el contexto financiero.	El análisis estadístico es insuficiente, con errores significativos o falta de gráficos. No identifica patrones relevantes o no contextualiza los resultados en el ámbito financiero.	
Implementación del modelo	25-22	21-18	17-0	25
	Implementa correctamente el modelo de regresión lineal usando Scikit-learn. El código es funcional, eficiente y bien documentado. Utiliza correctamente el conjunto de entrenamiento y prueba (70%/30%). Incluye técnicas complementarias (por ejemplo, normalización de datos) que mejoran el desempeño del modelo.	Implementa el modelo de regresión lineal utilizando Scikit-learn, aunque podría tener detalles técnicos por mejorar (por ejemplo, optimización del código o documentación). Cumple con la división de datos 70%/30%, pero sin añadir elementos adicionales que mejoren el modelo.	La implementación del modelo es incompleta o presenta errores que afectan significativamente su funcionalidad. No respeta la división de datos 70%/30% o carece de documentación adecuada en el código.	
	20-17	16-14	13-0	

Evaluación del modelo (métricas)	Calcula correctamente métricas relevantes como MSE y precisión,	Calcula métricas básicas como MSE y precisión, pero la interpretación es limitada o poco crítica.	No calcula métricas relevantes o las interpretaciones son incorrectas. Carece de análisis crítico sobre	20
---	---	---	---	-----------

	interpretándolas de manera crítica y relacionándolas con el contexto financiero. Propone mejoras basadas en los resultados obtenidos.	Aunque funcional, no profundiza en cómo los resultados podrían ser mejorados o aplicados en un contexto financiero.	los resultados o su impacto en el contexto financiero.	
	20-17	16-14	13-0	
Gráficos y visualización	Incluye gráficos claros y bien diseñados que visualizan correctamente los patrones en los datos y las predicciones del modelo. La selección de visualizaciones es adecuada y apoyan la interpretación y análisis de los resultados.	Presenta gráficos funcionales que representan los datos y predicciones, aunque podrían ser más claros o mejor diseñados. Algunas visualizaciones no aportan información significativa o podrían ser reemplazadas por alternativas más efectivas.	Los gráficos son confusos, mal diseñados o inexistentes. No aportan valor al análisis o interpretación de los resultados.	20
	20-17	16-14	13-0	
Reflexión y aplicación financiera	Presenta una reflexión profunda sobre cómo los resultados del modelo pueden influir en decisiones financieras. Expone conexiones claras y fundamentadas entre las variables analizadas y sus implicaciones prácticas, demostrando un sólido entendimiento del contexto financiero.	La reflexión incluye aspectos clave, pero carece de profundidad o fundamentación en algunos puntos. Identifica algunas implicaciones prácticas, aunque podría desarrollar más conexiones entre los resultados y su aplicación financiera.	La reflexión es superficial, con escasa o nula conexión entre los resultados del modelo y las decisiones financieras. No demuestra un entendimiento claro del impacto financiero.	20

TOTAL	100%
--------------	-------------

Anexo 3. Rúbrica de Evidencia 1

Criterios de evaluación	Nivel de desempeño			%
	Altamente competente 100%-86%	Competente 85%-70%	Aún sin desarrollar la competencia 69%-0%	
Análisis exploratorio inicial	10-9	8-7	6-0	10
	Realiza un análisis exhaustivo e identifica variables clave, valores atípicos y relaciones significativas. Incluye visualizaciones claras, relevantes y bien interpretadas.	Realiza un análisis adecuado, pero no exhaustivo. Identifica algunas variables clave y relaciones generales, aunque las visualizaciones son limitadas o carecen de interpretación completa.	El análisis es superficial o incompleto. No identifica variables clave ni relaciones significativas, y las visualizaciones son irrelevantes o ausentes.	
División del conjunto de datos	10-9	8-7	6-0	10
	Justifica claramente la división del conjunto de datos en entrenamiento y prueba. Detalla tamaños correctos y presenta el código funcional.	Justifica parcialmente la división de los datos. Aunque presenta tamaños y código, estos pueden contener errores menores o falta de claridad.	No justifica la división, los tamaños son incorrectos o no presenta el código necesario.	
Modelo de regresión lineal simple	15-13	12-11	10-0	15
	Selecciona y justifica la variable independiente de manera fundamentada. Entrena el modelo correctamente y evalúa con MSE bien calculado para ambos conjuntos.	Selecciona la variable y justifica parcialmente. Entrena el modelo y evalúa con MSE, aunque presenta errores o falta claridad en las justificaciones.	No justifica la selección de la variable, el modelo no es funcional, o los cálculos del MSE son incorrectos o ausentes.	
Modelo de	20-17	16-14	13-0	

regresión lineal multivariable	Entrena un modelo multivariable óptimo utilizando todas las variables relevantes.	Entrena un modelo multivariable funcional, pero con transformaciones	El modelo presenta errores graves o no incluye variables relevantes. Los	20
	Justifica transformaciones y calcula el MSE con precisión en ambos conjuntos.	limitadas o justificaciones incompletas. Cálculo del MSE adecuado, pero puede carecer de detalle.	cálculos del MSE son incorrectos o no se justifican.	
Modelo avanzado no lineal	20-17	16-14	13-0	20
	Selecciona y justifica el uso de un modelo avanzado adecuado. Entrena correctamente y presenta resultados claros del MSE en ambos conjuntos.	Selecciona un modelo avanzado con justificación limitada. Entrena el modelo con errores menores y presenta resultados del MSE sin detalle suficiente.	No justifica el modelo seleccionado, no lo entrena correctamente o los resultados del MSE son irrelevantes o inexistentes.	
Comparación de modelos y conclusiones	15-13	12-11	10-0	15
	Genera una tabla comparativa clara y bien organizada. Extrae conclusiones relevantes y fundamentadas sobre el mejor modelo para predecir el riesgo crediticio.	Presenta una tabla comparativa básica. Las conclusiones son válidas, pero carecen de profundidad o claridad en la fundamentación.	No presenta una tabla comparativa adecuada. Las conclusiones son irrelevantes, incorrectas o no fundamentadas.	
Calidad del informe técnico	10-9	8-7	6-0	10
	El informe está bien estructurado, utiliza lenguaje técnico adecuado y presenta todos los elementos requeridos con coherencia y precisión.	El informe incluye la mayoría de los elementos requeridos, pero con estructura o lenguaje técnico limitado.	El informe carece de estructura, utiliza lenguaje inadecuado o presenta muchos elementos faltantes.	
TOTAL				100%

Anexo 4. Rúbrica de Evidencia final

Criterios de evaluación	Nivel de desempeño			%
	Altamente competente 100%-86%	Competente 85%-70%	Aún sin desarrollar la competencia 69%-0%	
Informe inicial: Fundamentos y ejemplos	15-13	12-11	10-0	15
	Explica con claridad los métodos de interpretación de modelos lineales y no lineales. Incluye ejemplos financieros relevantes, bien fundamentados y con lenguaje técnico adecuado.	Explica los métodos de interpretación de manera general. Los ejemplos son válidos, pero carecen de profundidad o conexión clara con el contexto financiero.	No explica adecuadamente los métodos de interpretación, o los ejemplos son irrelevantes o ausentes.	
Definición del escenario financiero	10-9	8-7	6-0	10
	Selecciona un escenario relevante y crítico para la toma de decisiones. Describe claramente su importancia y justifica con ejemplos sólidos su relevancia en finanzas.	Selecciona un escenario relacionado, pero con descripción limitada. La justificación es válida pero superficial.	No define un escenario claro, la descripción es inadecuada o falta justificación de su relevancia.	
Implementación de SHAP: Análisis y visualizaciones	25-22	21-18	17-0	25
	Aplica SHAP correctamente para interpretar un modelo predictivo. Genera visualizaciones claras (cascada, enjambre) y las interpreta de forma técnica y comprensible.	Aplica SHAP pero con errores menores en la ejecución o interpretación. Las visualizaciones son funcionales, aunque carecen de claridad en algunos aspectos.	No implementa SHAP de manera efectiva. Las visualizaciones son incorrectas, incompletas o faltantes, con poca interpretación.	
	20-17	16-14	13-0	20

Análisis de sensibilidad	Realiza un análisis exhaustivo sobre cómo los cambios en las variables afectan las predicciones. Usa gráficos comparativos claros y relaciona los hallazgos con implicaciones financieras.	Realiza un análisis adecuado, pero no exhaustivo. Los gráficos son útiles pero limitados en claridad, o las implicaciones financieras son generales.	El análisis de sensibilidad es superficial o inexistente. Los gráficos son inadecuados y las implicaciones financieras no están definidas.	
Resultados del análisis: Informe técnico	20-17	16-14	13-0	20
	Presenta un informe técnico bien estructurado, con gráficos relevantes, interpretaciones claras y conclusiones sólidas que relacionan los hallazgos con el escenario planteado.	Presenta un informe técnico funcional, pero con faltas en la estructura o profundidad de las conclusiones. Los gráficos pueden carecer de claridad o detalle.	El informe es desorganizado, carece de gráficos relevantes o las conclusiones no están alineadas con el escenario planteado.	
Presentación final: Metodología, resultados y conclusiones	10-9	8-7	6-0	10
	Diseña una presentación visualmente atractiva y organizada, destacando la metodología, los resultados y conclusiones. Usa lenguaje técnico claro y accesible.	Presenta un trabajo funcional, pero con estructura o diseño limitado. Los resultados y conclusiones son claros, pero no destacan visualmente.	La presentación es desorganizada o incompleta. Los resultados y conclusiones no están claros o no se relacionan con el análisis previo.	
TOTAL				100%