UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE CIENCIAS MATEMÁTICAS

SÍLABO

1. INFORMACIÓN GENERAL

1.1 Curso	:	Tópicos en Inteligencia Artificial
1.2 Código	:	AI367
1.3 Programa Académico	:	Computación Científica
1.4 Créditos	:	5
1.5 Prerrequisitos	:	AI264. Aprendizaje Profundo. (9^{no} Sem)
1.6 Tipo de curso	:	Obligatorio
1.7 Área de Estudios	:	Estudios Generales
1.8 Número de Semanas	:	16
1.9 Horas por semana	:	3 HT; 4 HP;
1.10 Ciclo	:	I
1.11 Modalidad de aprendizaje	:	Presencial
1.12 Periodo Académico	:	2026-I
1.13 Inicio del semestre	:	
1.14 Fin del semestre	:	

2. PROFESORES

Atención previa coordinación con el profesor

3. INTRODUCCIÓN AL CURSO

La Inteligencia Artificial es un campo en rápida evolución. Este curso explora temas avanzados y contemporáneos que están dando forma actualmente a la industria y la investigación. Se pone énfasis en la IA Generativa, Modelos de Lenguaje Grandes (LLMs) y Aprendizaje por Refuerzo Avanzado. Los estudiantes analizarán las arquitecturas detrás de los sistemas modernos de IA, como los Transformers, y discutirán los desafíos técnicos y sociales de desplegar agentes inteligentes autónomos a gran escala.

4. OBJETIVOS

- Comprender la arquitectura y el proceso de entrenamiento de Modelos de Lenguaje Grandes.
- Dominar los principios de la IA Generativa para síntesis de texto e imágenes.
- Implementar algoritmos avanzados de Aprendizaje por Refuerzo para la toma de decisiones.
- Aplicar técnicas de ingeniería de prompts y ajuste fino a modelos existentes.
- Evaluar críticamente los últimos artículos de investigación y tendencias en IA.

5. RESULTADOS DEL ESTUDIANTE

ABET-1) Habilidad para identificar, formular y resolver problemas técnicos o científicos ampliamente definidos mediante la aplicación de conocimientos de matemáticas y ciencias y/o temas técnicos en áreas relevantes para la disciplina. (Usage)

6. TEMAS

<table>
<tr><th colspan="2">Unidad 1: Fundamentos de Aprendizaje Automático (6 horas)</th></tr>
<tr><th colspan="2">Resultados esperados: ABET-1</th></tr>
<tr><th>Temas</th><th>Aprendizaje esperado (Learning Outcomes)</th></tr>
<tr>
<td>

Definición y ejemplos de una amplia variedad de tareas de aprendizaje automático:

Aprendizaje supervisado:

Clasificación
Regresión

Aprendizaje por refuerzo
Aprendizaje no supervisado:

Agrupamiento

Ideas fundamentales:

Teorema de no hay almuerzo gratis: ningún aprendiz puede resolver todos los problemas; las decisiones de diseño de representación tienen consecuencias.
Fuentes de error e indecidibilidad en el aprendizaje automático

Un aprendizaje supervisado simple basado en estadísticas como regresión lineal o árboles de decisión:

Enfocarse en cómo funcionan sin entrar en detalles matemáticos o de optimización; suficiente para entender y usar implementaciones existentes correctamente

El problema del sobreajuste/control de la complejidad de la solución (regularización, poda - solo intuición):

El compromiso entre sesgo (subajuste) - varianza (sobreajuste)

</td>
<td>

Describir las diferencias entre los tres estilos principales de aprendizaje (supervisado, por refuerzo y no supervisado) y determinar cuál es apropiado para un dominio de problema particular [Explicar]
Diferenciar los términos: IA, aprendizaje automático y aprendizaje profundo [Evaluar]
Formular una aplicación como un problema de clasificación, incluyendo las características de entrada disponibles y la salida a predecir (por ejemplo, identificar caracteres alfabéticos a partir de entrada de cuadrícula de píxeles) [Aplicar]
Identificar el sobreajuste en el contexto de un problema y curvas de aprendizaje y describir soluciones al sobreajuste [Evaluar]
Explicar cómo funciona el aprendizaje automático como un proceso de optimización/búsqueda [Explicar]

</td>
</tr>
<tr><td colspan="2">Lecturas : [Goodfellow2016], [Foster2022]</td></tr>
</table>

Unidad 2: Aprendizaje Automático: Datos, Representaciones y Evaluación (5 horas)	
Resultados esperados: ABET-1	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Trabajar con Datos: • Preprocesamiento de datos: ◦ Importancia y peligros de las elecciones de preprocesamiento • Manejo de valores faltantes (imputación, marcar como faltante): ◦ Implicaciones de imputar versus marcar como faltante • Codificación de variables categóricas, codificación de datos de valor real • Normalización/estandarización • énfasis en datos reales, no ejemplos de libros de texto ▪ Representaciones: • Espacios de hipótesis y complejidad • Expansión de características de base simple, como elevar al cuadrado características univariadas • Representaciones de características aprendidas ▪ Evaluación del aprendizaje automático: • Separación de conjuntos de entrenamiento, validación y prueba • Métricas de rendimiento para clasificadores • Estimación del rendimiento en prueba en datos retenidos • Ajuste de los parámetros de un modelo de aprendizaje automático con un conjunto de validación • Importancia de entender qué está haciendo un modelo, dónde están sus peligros/limitaciones y las implicaciones de sus decisiones	▪ Aplicar dos o más algoritmos de aprendizaje estadístico simple a una tarea de clasificación y medir la precisión de los clasificadores [Aplicar] ▪ Implementar un algoritmo de aprendizaje estadístico y el proceso de optimización correspondiente para entrenar el clasificador y obtener una predicción en nuevos datos [Aplicar] ▪ Explicar los procedimientos adecuados de evaluación de AA, incluyendo las diferencias entre rendimiento en entrenamiento y prueba, y qué puede salir mal con el proceso de evaluación llevando a informes inexactos del rendimiento de AA [Explicar] ▪ Comparar dos algoritmos de aprendizaje automático en un conjunto de datos, implementando la metodología de preprocesamiento de datos y evaluación (por ejemplo, métricas y manejo de divisiones entrenamiento/prueba) desde cero [Aplicar]
Lecturas : [**Goodfellow2016**], [**Foster2022**]	

Unidad 3: Redes Neuronales y Ética (5 horas)	
Resultados esperados: ABET-1	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Redes neuronales básicas: • Fundamentos de comprensión de cómo funcionan las redes neuronales y su proceso de entrenamiento, sin detalles de los cálculos • Introducción básica a redes neuronales generativas (por ejemplo, modelos de lenguaje grandes) ▪ ética para el Aprendizaje Automático: • Enfocarse en datos reales, escenarios reales y estudios de caso • Sesgo de conjunto de datos/algorítmico/de evaluación y consecuencias no deseadas	▪ Describir el proceso de entrenamiento de redes neuronales y las representaciones aprendidas resultantes [Explicar]
Lecturas : [**Goodfellow2016**], [**Foster2022**]	

Unidad 4: Formulación del Aprendizaje Automático y Evaluación (3 horas)	
Resultados esperados:	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Formulación del aprendizaje automático simple como un problema de optimización, como regresión lineal por mínimos cuadrados o regresión logística: • Función objetivo • Descenso de gradiente • Regularización para evitar sobreajuste (formulación matemática) ▪ Conjuntos de modelos: • Combinación de mayoría ponderada simple ▪ Evaluación de rendimiento: • Otras métricas para clasificación (por ejemplo, error, precisión, exhaustividad) • Métricas de rendimiento para regresores • Matriz de confusión • Validación cruzada: ◦ Ajuste de parámetros (búsqueda en cuadrícula/aleatoria, mediante validación cruzada)	▪ Comparar y contrastar varias técnicas de aprendizaje (por ejemplo, árboles de decisión, regresión logística, Bayes ingenuo, redes neuronales y redes de creencia), proporcionando ejemplos de cuándo cada estrategia es superior [Comparar] ▪ Evaluar el rendimiento de un sistema de aprendizaje simple en un conjunto de datos del mundo real [Evaluar] ▪ Caracterizar el estado del arte en teoría del aprendizaje, incluyendo sus logros y limitaciones [Explicar] ▪ Explicar el problema del sobreajuste, junto con técnicas para detectar y manejar el problema [Explicar] ▪ Explicar el triple compromiso entre el tamaño de un espacio de hipótesis, el tamaño del conjunto de entrenamiento y la precisión del rendimiento [Explicar]
Lecturas : [**Goodfellow2016**]	

Unidad 5: Aprendizaje Profundo, RL y Ética (3 horas)	
Resultados esperados:	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Aprendizaje profundo: • Redes feed-forward profundas (solo intuición, sin matemáticas) • Redes neuronales convolucionales (solo intuición, sin matemáticas) • Visualización de representaciones de características aprendidas de redes profundas • Otras arquitecturas (NN generativas, NN recurrentes, transformers, etc.) ▪ Visión general de los métodos de aprendizaje por refuerzo ▪ Dos o más aplicaciones de algoritmos de aprendizaje automático: • Por ejemplo, medicina y salud, economía, visión, lenguaje natural, robótica, juego ▪ ética para el Aprendizaje Automático: • Continuar enfocándose en datos reales, escenarios reales y estudios de caso • Privacidad • Equidad • Propiedad intelectual • Explicabilidad	▪ Visualizar el progreso del entrenamiento de una red neuronal a través de curvas de aprendizaje en un kit de herramientas establecido (por ejemplo, TensorBoard) y visualizar las características aprendidas de la red [Aplicar] ▪ Dada una aplicación real de aprendizaje automático, describir problemas éticos respecto a las elecciones de datos, pasos de preprocesamiento, selección de algoritmo y visualización/presentación de resultados [Aplicar]
Lecturas : [**Goodfellow2016**]	

Unidad 6: Procesamiento del Lenguaje Natural (20 horas)

Resultados esperados: ABET-1

Temas	**Aprendizaje esperado (*Learning Outcomes*)**

▪ Gramáticas deterministas y estocásticas
▪ Algoritmos de análisis sintáctico:
  • CFGs y analizadores de tabla (por ejemplo, CYK)
  • CFGs probabilísticos y CYK ponderado
▪ Representación del significado/Semántica:
  • Representaciones del conocimiento basadas en lógica
  • Roles semánticos
  • Representaciones temporales
  • Creencias, deseos e intenciones
▪ Métodos basados en corpus
▪ N-gramas y HMMs
▪ Suavizado y retroceso
▪ Ejemplos de uso: Etiquetado POS y morfología
▪ Recuperación de información:
  • Modelo de espacio vectorial:
  ◦ TF & IDF
  • Precisión y exhaustividad
▪ Extracción de información
▪ Traducción de lenguaje
▪ Clasificación de texto, categorización:
  • Modelo de bolsa de palabras
▪ Aprendizaje profundo para PLN:
  • RNNs
  • Transformers
  • Incrustaciones multimodales (por ejemplo, imágenes + texto)
  • Modelos de lenguaje generativos
	▪ Definir y contrastar gramáticas deterministas y estocásticas, proporcionando ejemplos para mostrar la adecuación de cada una [Definir]
	▪ Simular, aplicar o implementar algoritmos clásicos y estocásticos para analizar lenguaje natural [Aplicar]
	▪ Identificar los desafíos de representar el significado [Analizar]
	▪ Listar las ventajas de usar corpus estándar. Identificar ejemplos de corpus actuales para una variedad de tareas de PLN [Listar/Enumerar]
	▪ Identificar técnicas para recuperación de información, traducción de lenguaje y clasificación de texto [Analizar]
	▪ Implementar una transformación TF/IDF, usarla para extraer características de un corpus, y entrenar un algoritmo de aprendizaje automático estándar usando esas características para hacer clasificación de texto [Aplicar]

Lecturas : [**Tunstall2022**], [**Russell2020**]	

Unidad 7: Fundamentos de Aprendizaje Automático (6 horas)	
Resultados esperados: ABET-1	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Definición y ejemplos de una amplia variedad de tareas de aprendizaje automático: • Aprendizaje supervisado: ◦ Clasificación ◦ Regresión • Aprendizaje por refuerzo • Aprendizaje no supervisado: ◦ Agrupamiento ▪ Ideas fundamentales: • Teorema de no hay almuerzo gratis: ningún aprendiz puede resolver todos los problemas; las decisiones de diseño de representación tienen consecuencias. • Fuentes de error e indecidibilidad en el aprendizaje automático ▪ Un aprendizaje supervisado simple basado en estadísticas como regresión lineal o árboles de decisión: • Enfocarse en cómo funcionan sin entrar en detalles matemáticos o de optimización; suficiente para entender y usar implementaciones existentes correctamente ▪ El problema del sobreajuste/control de la complejidad de la solución (regularización, poda - solo intuición): • El compromiso entre sesgo (subajuste) - varianza (sobreajuste)	▪ Describir las diferencias entre los tres estilos principales de aprendizaje (supervisado, por refuerzo y no supervisado) y determinar cuál es apropiado para un dominio de problema particular [Explicar] ▪ Diferenciar los términos: IA, aprendizaje automático y aprendizaje profundo [Evaluar] ▪ Formular una aplicación como un problema de clasificación, incluyendo las características de entrada disponibles y la salida a predecir (por ejemplo, identificar caracteres alfabéticos a partir de entrada de cuadrícula de píxeles) [Aplicar] ▪ Identificar el sobreajuste en el contexto de un problema y curvas de aprendizaje y describir soluciones al sobreajuste [Evaluar] ▪ Explicar cómo funciona el aprendizaje automático como un proceso de optimización/búsqueda [Explicar]
Lecturas : [**Sutton2018**]	

Unidad 8: Aprendizaje Automático: Datos, Representaciones y Evaluación (5 horas)	
Resultados esperados: ABET-1	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Trabajar con Datos: • Preprocesamiento de datos: ◦ Importancia y peligros de las elecciones de preprocesamiento • Manejo de valores faltantes (imputación, marcar como faltante): ◦ Implicaciones de imputar versus marcar como faltante • Codificación de variables categóricas, codificación de datos de valor real • Normalización/estandarización • énfasis en datos reales, no ejemplos de libros de texto ▪ Representaciones: • Espacios de hipótesis y complejidad • Expansión de características de base simple, como elevar al cuadrado características univariadas • Representaciones de características aprendidas ▪ Evaluación del aprendizaje automático: • Separación de conjuntos de entrenamiento, validación y prueba • Métricas de rendimiento para clasificadores • Estimación del rendimiento en prueba en datos retenidos • Ajuste de los parámetros de un modelo de aprendizaje automático con un conjunto de validación • Importancia de entender qué está haciendo un modelo, dónde están sus peligros/limitaciones y las implicaciones de sus decisiones	▪ Aplicar dos o más algoritmos de aprendizaje estadístico simple a una tarea de clasificación y medir la precisión de los clasificadores [Aplicar] ▪ Implementar un algoritmo de aprendizaje estadístico y el proceso de optimización correspondiente para entrenar el clasificador y obtener una predicción en nuevos datos [Aplicar] ▪ Explicar los procedimientos adecuados de evaluación de AA, incluyendo las diferencias entre rendimiento en entrenamiento y prueba, y qué puede salir mal con el proceso de evaluación llevando a informes inexactos del rendimiento de AA [Explicar] ▪ Comparar dos algoritmos de aprendizaje automático en un conjunto de datos, implementando la metodología de preprocesamiento de datos y evaluación (por ejemplo, métricas y manejo de divisiones entrenamiento/prueba) desde cero [Aplicar]
Lecturas : [**Sutton2018**]	

Unidad 9: Redes Neuronales y Ética (5 horas)	
Resultados esperados: ABET-1	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Redes neuronales básicas: • Fundamentos de comprensión de cómo funcionan las redes neuronales y su proceso de entrenamiento, sin detalles de los cálculos • Introducción básica a redes neuronales generativas (por ejemplo, modelos de lenguaje grandes) ▪ ética para el Aprendizaje Automático: • Enfocarse en datos reales, escenarios reales y estudios de caso • Sesgo de conjunto de datos/algorítmico/de evaluación y consecuencias no deseadas	▪ Describir el proceso de entrenamiento de redes neuronales y las representaciones aprendidas resultantes [Explicar]
Lecturas : [**Sutton2018**]	

Unidad 10: Aplicaciones e Impacto Social de la IA (6 horas)

Resultados esperados: ABET-1

Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Al menos una aplicación de IA a un problema y campo específico, como medicina, salud, sostenibilidad, redes sociales, economía, educación, robótica, etc. (elegir al menos uno para el Núcleo CS): • Formular y evaluar una aplicación específica como un problema de IA: ◦ Cómo lidiar con problemas mal especificados o mal planteados • Disponibilidad/escasez y limpieza de datos: ◦ Limpieza y preprocesamiento básico de datos ◦ Sesgo del conjunto de datos • Sesgo algorítmico • Sesgo de evaluación • Evaluación de las implicaciones sociales de la aplicación ▪ Modelos generativos profundos desplegados: • Visión general de alto nivel de modelos generativos de imágenes profundas (por ejemplo, a 2023, DALL-E, Midjourney, Stable Diffusion, etc.), sus usos y sus limitaciones/peligros. • Visión general de alto nivel de modelos de lenguaje grandes (por ejemplo, a 2023, ChatGPT, Bard, etc.), sus usos y sus limitaciones/peligros. ▪ Visión general del impacto social de la IA: • ética • Equidad • Confianza/explicabilidad • Privacidad y uso de datos de entrenamiento • Autonomía humana y supervisión/regulaciones/requisitos legales • Sostenibilidad	▪ Dado un dominio de aplicación y problema del mundo real, formular una solución de IA para ello, identificando datos/entrada apropiados, preprocesamiento, representaciones, técnicas de IA y métricas/metodología de evaluación [Aplicar]
Lecturas : [**Russell2020**], [**Bostrom2014**]	

Unidad 11: Aplicaciones Adicionales e Impacto Social de la IA (6 horas)	
Resultados esperados: ABET-1	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Una o más aplicaciones adicionales de IA a un amplio conjunto de problemas y diversos campos, como medicina, salud, sostenibilidad, redes sociales, economía, educación, robótica, etc. (elegir un área diferente de la elegida para el Núcleo CS): • Formular y evaluar una aplicación específica como un problema de IA: ◦ Cómo lidiar con problemas mal especificados o mal planteados • Disponibilidad/escasez y limpieza de datos: ◦ Limpieza y preprocesamiento básico de datos ◦ Sesgo del conjunto de datos • Sesgo algorítmico • Sesgo de evaluación • Evaluación de las implicaciones sociales de la aplicación ▪ Profundidad adicional sobre modelos generativos profundos desplegados: • Introducción a cómo funcionan los modelos generativos de imágenes profundas (por ejemplo, a 2023, DALL-E, Midjourney, Stable Diffusion) incluyendo discusión de atención • Introducción a cómo funcionan los modelos de lenguaje grandes (por ejemplo, a 2023, ChatGPT, Bard) incluyendo discusión de atención • Idea de modelos fundamentales, cómo usarlos, y los beneficios/problemas con entrenarlos a partir de grandes datos ▪ Análisis y discusión del impacto social de la IA: • ética • Equidad • Confianza/explicabilidad • Privacidad y uso de datos de entrenamiento • Autonomía humana y supervisión/regulaciones/requisitos legales • Sostenibilidad	▪ Analizar el impacto social de una o más aplicaciones específicas de IA del mundo real, identificando problemas respecto a ética, equidad, sesgo, confianza y explicabilidad [Evaluar] ▪ Describir algunos de los modos de falla de los modelos generativos profundos actuales para lenguaje o imágenes, y cómo esto podría afectar su uso en una aplicación [Describir]
Lecturas : [**Russell2020**], [**Bostrom2014**]	

7. PLAN DE TRABAJO

7.1 Metodología

Se fomenta la participación individual y en equipo para exponer sus ideas, motivándolos con puntos adicionales en las diferentes etapas de la evaluación del curso.

7.2 Sesiones Teóricas

Las sesiones de teoría se llevan a cabo en clases magistrales donde se realizarán actividades que propicien un aprendizaje activo, con dinámicas que permitan a los estudiantes interiorizar los conceptos.

7.3 Sesiones Prácticas

Las sesiones prácticas se llevan en clase donde se desarrollan una serie de ejercicios y/o conceptos prácticos mediante planteamiento de problemas, la resolución de problemas, ejercicios puntuales y/o en contextos aplicativos.

8. SISTEMA DE EVALUACIÓN

********* EVALUATION MISSING ********

9. BIBLIOGRAFÍA BÁSICA