UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE CIENCIAS MATEMÁTICAS

SÍLABO

1. INFORMACIÓN GENERAL

1.1 Curso	:	Procesamiento de Lenguaje Natural
1.2 Código	:	AI262
1.3 Programa Académico	:	Computación Científica
1.4 Créditos	:	4
1.5 Prerrequisitos	:	AI263. Introducción al Aprendizaje de Máquina. (8^{vo} Sem)
1.6 Tipo de curso	:	Electivo
1.7 Área de Estudios	:	Estudios Generales
1.8 Número de Semanas	:	16
1.9 Horas por semana	:	2 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

Este curso introduce los fundamentos teóricos y prácticos del Procesamiento de Lenguaje Natural (NLP), cubriendo desde técnicas clásicas hasta modelos modernos basados en deep learning. Los estudiantes aprenderán a implementar pipelines completos de NLP para aplicaciones como análisis de sentimientos, traducción automática y chatbots.

4. OBJETIVOS

- Comprender la representación computacional del lenguaje humano.
- Implementar algoritmos para procesamiento de texto en Python.
- Aplicar modelos de NLP a problemas reales usando frameworks modernos.

5. RESULTADOS DEL ESTUDIANTE

ABET-2) Habilidad para formular o diseñar un sistema, proceso, procedimiento o programa para el propósito previsto. (Usage)

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

Unidad 1: Fundamentos de NLP (15 horas)	
Resultados esperados: ABET-1,ABET-2	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Tokenización y representación de texto ▪ Modelos de n-gramas y suavizado ▪ Extracción de características (TF-IDF, word embeddings)	▪ Preprocesar texto para modelos de NLP [Usar] ▪ Explicar limitaciones de modelos estadísticos clásicos [Familiarizarse]
Lecturas : [**Jurafsky2021**], [**Manning1999**]	

Unidad 2: Modelado de Lenguaje (15 horas)	
Resultados esperados: ABET-1,ABET-2	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Word2Vec, GloVe y FastText ▪ Redes neuronales recurrentes (RNNs) para NLP ▪ Modelos de secuencia a secuencia (seq2seq)	▪ Entrenar embeddings de palabras [Usar] ▪ Implementar RNNs básicas para generación de texto [Evaluar]
Lecturas : [**Goldberg2017**], [**Mikolov2013**]	

Unidad 3: Transformers y NLP Moderno (15 horas)	
Resultados esperados: ABET-1,ABET-2	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Arquitectura Transformer (atención auto-supervisada) ▪ BERT y modelos pre-entrenados ▪ Fine-tuning para tareas específicas	▪ Aplicar modelos pre-entrenados con HuggingFace [Usar] ▪ Evaluar métricas de desempeño en NLP [Evaluar]
Lecturas : [**Vaswani2017**], [**Devlin2019**]	

Unidad 4: Aplicaciones Prácticas (15 horas)	
Resultados esperados: ABET-1,ABET-2	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Análisis de sentimientos con transformers ▪ Sistemas de pregunta-respuesta ▪ Despliegue de modelos con Flask/FastAPI	▪ Construir un pipeline completo de NLP [Usar] ▪ Optimizar modelos para producción [Evaluar]
Lecturas : [**Jurafsky2021**], [**Chollet2021**]	

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