

UNIVERSIDAD AUTÓNOMA DE
CHIHUAHUA



UNIDAD ACADÉMICA:
FACULTAD DE INGENIERÍA

PROGRAMA ANALÍTICO DE LA
UNIDAD DE APRENDIZAJE:

NATURAL LANGUAGE
PROCESSING

DES: INGENIERÍA	
Programa académico	Ingeniería en Ciencia de Datos y Matemáticas Aplicadas
Tipo de materia (Obli/Opta):	Optativa
Clave de la materia:	OPCM801
Semestre:	8vo.
Área en plan de estudios:	Inteligencia artificial
Total de horas por semana:	4
<i>Teoría: Presencial o Virtual</i>	2
<i>Laboratorio o Taller:</i>	0
<i>Prácticas:</i>	2
<i>Trabajo extra-clase:</i>	
Créditos Totales:	4
Total de horas semestre (x sem):	64
Fecha de actualización:	Febrero 2024
Prerrequisito (s):	BI204 Programación para ingenieros BI102 Cálculo diferencial e integral BI201 Álgebra lineal BI301 Probabilidad BI401 Estadística MC602 Machine learning

DESCRIPCIÓN:

La materia le enseña al estudiante la importancia del *procesamiento de lenguaje natural* (NLP por sus siglas en inglés.) en la vida moderna. Esto debido a que la comunicación actualmente en áreas como *servicio al cliente, traducción automática, generación de reportes médicos automáticos, política*, entre otros; se realiza principalmente por medio de lenguaje natural automático. El estudiante adquirirá conocimientos que le permitirán entender y aplicar técnicas tradicionales y modernas como *Deep Learning*, para la implementación de aplicaciones tales como auto correctores y auto completadores de texto, traductores y generadores automáticos de texto por nombrar algunas.

COMPETENCIAS PARA DESARROLLAR:

B1. EXCELENCIA Y DESARROLLO HUMANO

Promueve el desarrollo humano integral con resultados tangibles obtenidos en la formación de profesionales con conciencia ética y solidaria, pensamiento crítico y creativo, así como una capacidad innovadora, productiva y emprendedora en el marco de la innovación y pertinencia social, con matices éticos y de valores, que desde su particularidad cultural le permitan respetar la diversidad, promover la inclusión, valorar la interculturalidad.

E3. APLICACIÓN DE LA COMPUTACIÓN

Aplica conocimientos de computación en proyectos de ciencia de datos y matemáticas aplicadas, enfocándose en el diseño, análisis y solución de problemas multidisciplinarios, implementando algoritmos y modelos *con el fin de obtener información significativa para la toma de decisiones.*

DOMINIOS (Se toman de las competencias)	OBJETOS DE ESTUDIO (Contenidos necesarios para desarrollar cada uno de los dominios)	RESULTADOS DE APRENDIZAJE (Se plantean de los dominios y contenidos)	METODOLOGÍA (Estrategias, secuencias, recursos didácticos)	EVIDENCIAS (Productos tangibles que permiten valorar los resultados de aprendizaje)
<p>E3. APLICACIÓN DE LA COMPUTACIÓN</p> <p>4.- Implementa algoritmos de inteligencia artificial, aplicando técnicas avanzadas para asegurar su eficiencia y precisión en distintas aplicaciones, considerando el costo computacional y la precisión de los resultados.</p>	<p>1. Introducción al procesamiento de lenguaje natural y repaso de Python.</p> <p>1.1. Importancia del NLP 1.2. Aplicaciones del NLP 1.3. Métodos tradicionales vs modernos. 1.4. ¿Por qué usar Python en machine learning y NLP? 1.4.1. Estructuras básicas: listas, tuplas, diccionarios, etc. 1.4.2. Iteradores vs ciclos. 1.4.3. Algunas librerías básicas para ML y NLP en Python. 1.4.4. POO en Python.</p>	<p>Entiende la importancia del NLP en la actualidad.</p> <p>Comprende y aplica el lenguaje de programación Python en el área de NLP.</p>	<p>Clase interactiva Maestro - Alumno.</p> <p>Recursos tecnológicos institucionales.</p>	<p>Tareas y/o mini proyectos.</p>
<p>B1,2 Propone la solución de problemas con una base interdisciplinar (científica, humanística y tecnológica)</p>	<p>2. Representaciones textuales y espacios vectoriales.</p> <p>2.1. Representaciones dispersas. 2.1.1. Extracción de características. 2.1.2. Preprocesamiento 2.1.3. Sentiment Analysis: implementación de un ejemplo usando Logistic regression. 2.2. Probabilidades en Natural Language Processing (NLP) 2.2.1. Probabilidades condicionales (Regla de Bayes)</p>	<p>Explica la importancia de la representación de texto por medio de entidades numéricas.</p> <p>Implementa aplicaciones básicas de <i>sentiment análisis.</i></p>	<p>Clase interactiva Maestro - Alumno.</p> <p>Recursos tecnológicos institucionales.</p>	<p>Tareas y/o mini proyectos.</p>

	<p>2.2.2. Naïve Bayes para Sentiment Analysis.</p> <p>2.3. Modelos espacio-vectoriales</p> <p>2.3.1. Representaciones vectoriales de palabras y documentos</p> <p>2.3.2. Distancia euclidiana y similitud de coseno.</p> <p>2.3.3. Visualización de vectores n-dimensionales: Principal Component Analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE)</p>			
<p>B1,2 Propone la solución de problemas con una base interdisciplinar (científica, humanística y tecnológica)</p>	<p>3. Modelos probabilísticos.</p> <p>3.1. Creando un modelo para autocorrección de palabras</p> <p>3.1.1. Probabilidades de palabras</p> <p>3.1.2. Planteando el problema como alineación de secuencias.</p> <p>3.1.3. Solución como programación dinámica.</p> <p>3.2. Part of Speech Tagging (POS).</p> <p>3.2.1. Cadenas de Markov.</p> <p>3.2.2. Modelos ocultos de Markov</p> <p>3.3. Modelos de Lenguaje</p> <p>3.3.1. N-gramas</p> <p>3.3.2. Probabilidad condicional de secuencias de palabras.</p> <p>3.3.3. Un modelo de lenguaje generativo.</p> <p>3.3.4. Métricas para evaluación de modelos de lenguaje.</p> <p>3.4. Word Embeddings.</p> <p>3.4.1. One-Hot vectors.</p> <p>3.4.2. Significado n-vectorial embebido.</p> <p>3.4.3. Métodos de Word Embeddings.</p> <p>3.4.3.1. No-contextuales.</p> <p>3.4.3.1.1. Word2vec, GloVe, fastText.</p> <p>3.4.3.2. Contextuales.</p> <p>3.4.3.2.1. BERT, GPT-X, etc.</p> <p>3.4.3.3. EL método Skipgram.</p>	<p>Entiende modelos de lenguaje básicos por medio de conceptos probabilísticos.</p> <p>Comprende el concepto de vectores de palabras (Word embeddings) como representaciones de texto con significado.</p>	<p>Clase interactiva Maestro - Alumno.</p> <p>Recursos tecnológicos institucionales.</p>	<p>Tareas y/o mini proyectos.</p>

<p>B1,2 Propone la solución de problemas con una base interdisciplinar (científica, humanística y tecnológica)</p>	<p>4. Modelos secuenciales.</p> <ul style="list-style-type: none"> 4.1. Redes Neuronales en NLP. 4.2. Sentiment Analysis con redes neuronales. 4.3. Redes Neuronales Recurrentes (RNNs). <ul style="list-style-type: none"> 4.3.1. Arquitecturas vector-to-sequence, sequence-to-vector-sequence-to-sequence. 4.3.2. RNNs y el vanishing gradient. 4.3.3. Gated Recurrent Unit (GRU) RNNs. 4.3.4. Long Short-Term Memory (LSTM) RNNs. 4.3.5. Named Entity Recognition (NER) con LSTMs. 	<p>Repasa el entendimiento de las redes neuronales tradicionales.</p> <p>Comprende la arquitectura de las redes neuronales recurrentes y sus diversas variantes.</p>	<p>Clase interactiva Maestro - Alumno.</p> <p>Recursos tecnológicos institucionales.</p>	<p>Tareas y/o mini proyectos.</p>
<p>B1,2 Propone la solución de problemas con una base interdisciplinar (científica, humanística y tecnológica)</p>	<p>5. Modelos de Atención.</p> <ul style="list-style-type: none"> 5.1. Neural Machine Translation (NMT). 5.2. Modelos sequence-to-sequence. <ul style="list-style-type: none"> 5.2.1. Encoders-Decoders y el cuello de botella de información. 5.2.2. Alineación de palabras y atención. 5.2.3. Keys, Queries y Values para recuperación de información en la capa de atención. 5.2.4. Beam search para mejores decoders. 5.3. Bilingual Evaluation Understudy (BLEU score) como medida de evaluación de calidad en traducción de textos. 5.4. Recall-Oriented Understudy for Gisting Evaluation (ROUGE score) como medida de evaluación de calidad en textos generados automáticamente. 5.5. Los Transformers. <ul style="list-style-type: none"> 5.5.1. Aplicaciones: Sumarización de textos, auto-complete, NER, question answering, traducción, chatbots, sentiment analysis, text classification, spell-checking. 5.5.2. Arquitectura básica. 	<p>Comprende el concepto de “atención” en el contexto de las redes neuronales, como única necesidad para un rendimiento efectivo.</p> <p>Entiende y aplica las diferentes métricas para evaluación de modelos de lenguaje generadores de texto automático.</p> <p>Estudia y analiza las ventajas y limitantes de las redes neuronales “transformer”.</p>	<p>Clase interactiva Maestro - Alumno.</p> <p>Recursos tecnológicos institucionales.</p>	<p>Tareas y/o mini proyectos.</p>

	<p>5.5.3. Modelos de lenguaje pre-entrenados basados en Transformers: BERT, GPT-X, T5.</p> <p>5.5.4. Self attention y las matemáticas detrás.</p> <p>5.5.5. Multi-head attention.</p> <p>5.5.6. Transfer learning.</p> <p>5.5.7. Fine-tuning.</p> <p>5.5.8. Algunas aplicaciones.</p> <p>5.5.8.1. Question answering.</p> <p>5.5.8.2. Chat-bots</p> <p>5.5.9. Limitantes de los Transformers</p> <p>5.5.9.1. Secuencias largas.</p> <p>5.5.9.2. Complejidad computacional cuadrática de la capa de atención.</p> <p>5.5.9.3. LSH attention.</p> <p>5.5.9.4. Reversible Layers.</p> <p>5.5.9.5. El Reformer, un Transformer más eficiente.</p>			
--	---	--	--	--

FUENTES DE INFORMACIÓN (Bibliografía, direcciones electrónicas)	EVALUACIÓN DE LOS APRENDIZAJES (Criterios, ponderación e instrumentos)
<ol style="list-style-type: none"> 1. Deep Learning. Ian Goodfellow and Yoshua Bengio and Aaron Courville. 2. The Handbook of Computational Linguistics and Natural Language Processing. Alexander Clark, Chris Fox, Shalom Lappin. 3. Natural Language Processing with Python. Steven Bird, Ewan Klein, Edward Loper. 4. Natural Language Processing with Transformers. Lewis Tunstall, Leandro Von Werra, Thomas Wolf. 	<p>Parcial 1 (30%)</p> <ul style="list-style-type: none"> • Tareas y/o Mini proyecto. (100%) <p>Parcial 2 (30%)</p> <ul style="list-style-type: none"> • Tareas y/o Mini proyecto. (100%) <p>Parcial 3 (40%)</p> <ul style="list-style-type: none"> • Tareas y/o Mini proyecto. (100%)

CRONOGRAMA

Objetos de estudio	Semanas
--------------------	---------

