



Programa de curso

# Procesamiento de lenguaje natural utilizando aprendizaje automático

Escuela de Ingeniería en Computación del Tecnológico de Costa Rica.

**Contacto:**

Para mayor información del programa ingrese al siguiente link:

<https://www.tec.ac.cr/fundatec/programa-ciencias-datos-escuela-computacion>

**Coordinadora del Programa:**

Máster Alicia Salazar Hernández, [asalazar@itcr.ac.cr](mailto:asalazar@itcr.ac.cr)

**Asistente Administrativa del Programa:**

Licda. Sarela Gómez Brenes, [sagomez@itcr.ac.cr](mailto:sagomez@itcr.ac.cr)

## Procesamiento de lenguaje natural con aprendizaje automático

<p><b>Descripción</b></p>	<p>La cantidad de texto disponible en formato electrónico es impresionante, aumenta cada día y es pieza fundamental para automatizar procesos en múltiples áreas como: ciencia, tecnología, industria, comercio, innovación, educación, entre otras. Donde se explore con un poco de detalle es factible encontrar una necesidad que puede ser resuelta, en parte, con esos grandes volúmenes de texto. Sin embargo, la complejidad del lenguaje natural hace difícil el acceso e integración de esos datos con la información disponible.</p> <p>El procesamiento de lenguaje natural (NLP por sus siglas en inglés) es una de las grandes áreas de investigación de las ciencias de la computación, la inteligencia artificial y la lingüística. El principal objetivo del NLP es desarrollar técnicas y algoritmos para procesar, analizar e interpretar automáticamente el lenguaje humano.</p> <p>Recientemente, con el fuerte desarrollo del aprendizaje profundo, el campo de investigación ha evolucionado de forma muy positiva. Aplicaciones como búsquedas en Internet, agentes virtuales de servicio al cliente, traducciones, generación de textos, resúmenes de textos, análisis de sentimientos, entre otras aplicaciones, se han visto beneficiadas y potenciadas con las mejoras realizadas a algoritmos de aprendizaje profundo o redes neuronales artificiales.</p> <p>Como ejemplo de los temas a abordar en el curso, un agente virtual de servicio al cliente basado en búsquedas semánticas con la tecnología de transformers, puede ser accedido y ejecutado en el siguiente cuaderno de Jupyter: <a href="#">INGRESE AQUÍ</a></p>
<p><b>Requisitos de Ingreso</b></p>	<ul style="list-style-type: none"> <li>▪ Personas con algún tipo de grado en Estadística, Ingeniería en Computación, Ingeniería en Sistemas de Información, Matemáticas, Lingüística, entre otras áreas.</li> <li>▪ Es importante que la persona interesada cuente con conocimientos en Python o en su defecto haber programado en algún otro lenguaje y que, por cuenta propia, pueda aprender Python.</li> </ul>
<p><b>Tipo de Curso</b></p>	<p>Aprovechamiento Tipo teórico/práctico Para obtener el certificado correspondiente, es necesario tener una asistencia efectiva de más del 80% a las lecciones, y sus evaluaciones con un promedio mayor o igual a 70.</p>

<b>Cantidad de Horas Lectivas</b>	32 Horas lectivas virtuales
<b>Objetivos</b>	<p><b>Objetivo General:</b></p> <ul style="list-style-type: none"> <li>▪ Al finalizar el curso, el estudiante contará con todas las bases para realizar proyectos que involucren actividades de NLP combinadas con aprendizaje automático.</li> </ul> <p><b>Objetivos Específicos:</b></p> <ul style="list-style-type: none"> <li>▪ Desarrollar un claro entendimiento del flujo de trabajo requerido por proyectos de NLP con aprendizaje automático.</li> <li>▪ Estar en capacidad de aplicar el paradigma de aprendizaje supervisado a problemas de NLP.</li> <li>▪ Conocer la problemática y las principales aplicaciones del aprendizaje automático al NLP.</li> </ul>
<b>Metodología de la enseñanza</b>	El curso utilizará una metodología de aprender haciendo, basado en el desarrollo de proyectos de forma tal que, mediante el desarrollo de casos de estudio y proyectos prácticos, se pueda afianzar los conocimientos adquiridos durante el transcurso de las diferentes lecciones.
<b>Contenidos del programa</b>	<p><b>Introducción</b></p> <ul style="list-style-type: none"> <li>• Flujo de trabajo en procesamiento de lenguaje natural (NLP)</li> <li>• Áreas de aplicación en NLP (análisis de sentimientos, extracción de información, clasificación, etc)</li> <li>• Introducción al NLP con Python</li> </ul> <p><b>Procesamiento del lenguaje</b></p> <ul style="list-style-type: none"> <li>• Tokenización</li> <li>• Etiquetado gramatical (POS por sus siglas en inglés)</li> <li>• Corpus</li> <li>• Análisis de dependencias</li> <li>• Modelos del lenguaje y representación de las palabras</li> </ul> <p><b>Clasificación de textos binaria, multiclase y multinivel</b></p> <ul style="list-style-type: none"> <li>• Introducción</li> <li>• Algoritmos: <ul style="list-style-type: none"> <li>• Naive Bayes</li> <li>• Linear regression</li> <li>• Logistic Regression</li> <li>• Support Vector Machines (SVM)</li> </ul> </li> </ul> <p><b>Extracción de información</b></p> <ul style="list-style-type: none"> <li>• Introducción</li> </ul>

	<ul style="list-style-type: none"> <li>• Reconocimiento de nombres de entidades (NER)</li> <li>• Algoritmos             <ul style="list-style-type: none"> <li>• Hidden Markov Models (HMM)</li> <li>• Conditional Random Fields (CRF)</li> </ul> </li> </ul> <p><b>Preguntas y respuestas</b></p> <ul style="list-style-type: none"> <li>• Introducción</li> <li>• Fuentes de datos</li> <li>• Recuperación de información</li> <li>• Preprocesamiento y comprensión de las preguntas</li> </ul> <p><b>Fundamentos de redes neuronales artificiales - RN</b></p> <ul style="list-style-type: none"> <li>• Introducción</li> <li>• Perceptrón</li> <li>• Optimización</li> </ul> <p><b>Redes neuronales recurrentes - RNN</b></p> <ul style="list-style-type: none"> <li>• Introducción</li> <li>• Redes neuronales recurrentes y LSTM (Long Short-Term Memory)</li> <li>• LSTM bidireccionales</li> </ul> <p><b>Transformers</b></p> <ul style="list-style-type: none"> <li>• Introducción</li> <li>• Algunas aplicaciones</li> </ul>
<p><b>Materiales</b></p>	<p>Todo el material a utilizar en el desarrollo del curso ha sido preparado por profesores de La Escuela de Computación, y se encontrará de forma digital y podrá ser accedido en todo momento por el estudiante.</p>
<p>Bibliografía recomendada</p>	<p>Eisenstein, J. (2018). Natural Language Processing. <a href="https://github.com/jacobeisenstein/qt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf">https://github.com/jacobeisenstein/qt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf</a></p> <p>Jurafsky, D. and Martin, J. (2020). <a href="#">Speech and Language Processing (3rd ed. Draft)</a></p> <p>Goldberg, Y. (2015). <a href="#">A Primer on Neural Network Models for Natural Language Processing</a></p>

Goodfellow, I., Bengio, J., and Courville, A. (2016). Deep Learning. The MIT Press.

Rao, D., McMahan, B. (2019). Natural Language Processing with PyTorch. O'Reilly Media, Inc.

Zhang, A., Zachary, C. Lipton, Li, M. and Smola, A. (s.f.). Dive into Deep Learning. Open Source Book. <https://d2l.ai/index.html>

## Políticas de nuestro Programa:

### CONGELAMIENTOS:

En caso de que algún estudiante decida congelar el monto de cualquiera de los módulos de los programas de Educación Continua que ofrece la Escuela de Ingeniería en Computación deberá, solicitar el formulario de congelamiento al correo: [sagomez@itcr.ac.cr](mailto:sagomez@itcr.ac.cr), [asalazar@itcr.ac.cr](mailto:asalazar@itcr.ac.cr), posteriormente remitirlo a estos dos correos y debe esperar al visto bueno del Coordinador General del Programa.

El congelamiento se hace solamente por **motivos de fuerza mayor**. Este congelamiento tiene una vigencia de un bimestre, es decir, debe de ingresar al siguiente grupo que inicie. En caso de no presentar su reingreso antes de este lapso perderá el monto cancelado, correspondiente al módulo congelado. En caso de reingreso y si hubiese un cambio de precio, se debe cancelar la diferencia en relación a la inversión del módulo cancelado y la inversión del módulo en que indique se incorpora.

### DEVOLUCIONES:

Las devoluciones se realizarán debidamente justificadas (**por motivos de fuerza mayor**) y deberán ser aprobadas por la coordinación del programa. Dicha solicitud de devolución puede ser presentada **ÚNICAMENTE** los días anteriores al inicio de las lecciones del módulo correspondiente, si se presenta el primer día de clases o posterior, **NO PROCEDE** dicha devolución, sin excepción. Favor no comprometer al personal administrativo.

Para dicha aprobación el estudiante debe presentar una carta dirigida a la Coordinadora del Programa (Máster Alicia Salazar Hernández) justificando el motivo de la devolución de dinero, también deberá indicar; nombre, número de cédula, curso al que pertenece, monto que canceló y el número de cuenta para la transferencia en caso de que sea una empresa quien realice el depósito, deberá incluir los mismos datos y la firma del encargado de la inscripción o de quien realice el depósito con número de cédula jurídica, y razón social, en ambos casos **está sujeta a la aprobación de la Coordinación del Programa**, si la misma se aprueba tendrá 8 días para ser efectiva.

Independientemente al motivo de la solicitud de devolución, se procede con el cobro administrativo del 35% aplicado al monto cancelado, quedando a disposición de la de la coordinación del programa.

Deberá remitir la carta debidamente firmada a los correos: [correo:sagomez@itcr.ac.cr](mailto:correo:sagomez@itcr.ac.cr), [asalazar@itcr.ac.cr](mailto:asalazar@itcr.ac.cr).