

Instituto Tecnológico de Costa Rica
Maestría en Computación
Curso: Procesamiento de Lenguaje Natural (NLP)
Semestre II-2025

Proyecto final de NLP

Contenido

1. Administrativos	1
2. Descripción del proyecto	1
Introducción	1
Objetivos de aprendizaje	1
Definición del anteproyecto (entrega semana 8 - 17 de setiembre)	2
Definición del proyecto	2
3. Guía para ejecutar el proyecto	3
4. Rúbrica	3

1. Administrativos

Entrega: A través del TEC-digital. Un archivo .zip que contenga los datos o referencia a estos, la documentación solicitada y el código en Python bien documentado.

Modo de trabajo: Grupos de 2 personas máximo (se motiva a que trabajen en equipo).

Tecnología base a utilizar: Python, PyTorch y las bibliotecas que los estudiantes seleccionen.

2. Descripción del proyecto

Introducción

El **objetivo del trabajo** es poner en práctica **habilidades de investigación y el conocimiento adquirido durante el curso** por medio de aplicaciones seleccionadas por los participantes que permitan **experimentar con algoritmos** de aprendizaje automático aplicados al NLP en sus áreas de interés.

Objetivos de aprendizaje

1. Poner en práctica habilidades de investigación y documentación de resultados.
2. Aplicar el conocimiento teórico-práctico adquirido en el curso en la implementación de un ejemplo aplicado al área de interés de los estudiantes.
3. Experimentar con el flujo completo de trabajo requerido en proyectos de aprendizaje automático aplicado al NLP.
4. Fortalecer capacidades en los estudiantes en el uso de los recursos tecnológicos disponibles libremente.
5. Compartir el conocimiento adquirido por medio de una presentación de 20 minutos al grupo en la última sesión del curso (semana 16).

Definición del anteproyecto (entrega semana 8 - 17 de setiembre)

Como parte del desarrollo del proyecto se debe definir un anteproyecto que contenga las siguientes secciones:

- Justificación
- Antecedentes (en esta sección deben listar al menos **cuatro artículos** relevantes que se relacionen con su proyecto. Se debe mencionar el objetivo del artículo, explicar la relación con su proyecto e incluir la referencia)
- Marco teórico (que describa una selección de los algoritmos a utilizar)
- Definición del experimento
- Plan de trabajo
- Fuentes de datos que esperan utilizar

Definición del proyecto

Se debe utilizar el formato general para publicaciones de la IEEE (<https://www.overleaf.com/read/cbtdvknjrdsf#0e9468>). Con secciones como:

1. Introducción: que incluya justificación, antecedentes (a partir de los artículos seleccionados bien referenciados y otros), descripción del problema y el aporte de la investigación.
2. Marco teórico con citas a las referencias de las tecnologías utilizadas.
3. Metodología: descripción del flujo de trabajo a aplicar en el proyecto que incluya una descripción y cita de la fuente de los datos a utilizar. En esta sección se debe

describir el procesamiento de los datos realizado, la modelización, el entrenamiento o afinamiento de modelos y la evaluación de los resultados, entre otras secciones

4. Análisis de los resultados y discusión
5. Conclusiones y recomendaciones
6. Referencias
7. Se debe presentar los resultados de su investigación al grupo (la última semana del curso).

3. Guia para ejecutar el proyecto

El grupo de estudiantes deben seleccionar un tema de investigación de interés. Se debe diseñar y ejecutar una investigación que incluya la implementación de un experimento en el tema seleccionado. Una lista de temas sugeridos y otros recursos de interés están disponible en:

https://docs.google.com/document/d/1UTxnMz_2EFygytjazOFzAZgNJOu3jLvA87IP76L5-gY/edit?usp=sharing

Actividades:

- Se recomienda revisar esta lectura para el proyecto (metodología):
 - Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning. Capítulo 11: Practical Methodology.
<https://www.deeplearningbook.org/contents/guidelines.html>
- Investigar
 - Investigar y seleccionar un tema de interés.
 - Buscar una o dos referencias muy cercanas a lo que desean realizar y usarlas de base para diseñar el experimento.
 - Definir un plan de trabajo y cómo van a **innovar en el proyecto**.
 - El proyecto debe involucrar sustancialmente lenguaje humano y aprendizaje automático (ej. redes neuronales).
- Preparar un anteproyecto para presentar en la fecha indicada.
- Implementar el experimento
- Se puede usar ChaGPT y otros modelos como apoyo a la investigación, pero recuerden que la evaluación va a estar en función del aporte del grupo de estudiantes.

4. Rúbrica

Anteproyecto	Puntos
1. Justificación	5
2. Antecedentes	10
3. Marco teórico	20
4. Definición del experimento (metodología)	10
5. Plan de trabajo	5
6. Fuentes de datos que esperan utilizar	5
Proyecto	
1. La introducción incluye todas las secciones definidas anteriormente.	5
2. Marco teórico	5
3. Metodología:	
• Adquisición de datos, incluye depurado y estudio de las características de los datos	10
• Selección y estudio de modelos a utilizar	5
• Evaluación de resultados con las métricas apropiadas (incluyendo el tiempo y los recursos requeridos)	10
4. Se presentaron los resultados y discusión	10
5. Se presentaron las conclusiones y recomendaciones	5
6. Se presentaron las referencias	5
7. Se entrega el código apropiadamente documentado	5
8. Se realizó la presentación del proyecto en clase	10