

Raúl Eduardo Paz Zanini¹, Germán Ezequiel Lescano¹ y Lilia Eugenia Palomo¹
{rauleduardo.pazzanini, germanezequiel.lescano, lilia.palomo}@ucse.edu.ar
Facultad de Ciencias para la Innovación y el Desarrollo (FCID)¹
Universidad Católica de Santiago del Estero (UCSE)

CONTEXTO

El proyecto se enmarca en el creciente desarrollo del Procesamiento del Lenguaje Natural, impulsado por los modelos basados en arquitectura Transformer, que han revolucionado la forma de analizar y comprender textos. Surge ante la necesidad de evaluar el desempeño de estos modelos en idioma español, dado que la mayoría de los avances se han centrado en el inglés.

En este contexto, la investigación, financiada por la Facultad de Ciencias para la Innovación y el Desarrollo de la Universidad Católica de Santiago del Estero, promueve la articulación entre docencia e investigación.

Asimismo, busca generar conocimiento aplicable tanto en el ámbito académico como en aplicaciones reales, favoreciendo el desarrollo de herramientas más inclusivas, eficientes y adaptadas a las particularidades lingüísticas del español.

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO/OBJETIVOS

El proyecto se estructura en torno a la evaluación y comparación de modelos Transformer preentrenados aplicados a la clasificación de textos en español, un campo donde aún se evidencia limitada experimentación.

El objetivo general consiste en analizar el rendimiento de distintos modelos para determinar sus fortalezas, limitaciones y posibles aplicaciones. Para ello, se establecen diversos objetivos específicos que guían el desarrollo de la investigación.

En primer lugar, se busca comprender en profundidad los fundamentos teóricos de la arquitectura Transformer y su funcionamiento en tareas de procesamiento del lenguaje natural. Asimismo, se propone revisar el estado del arte de estos modelos en el contexto del idioma español, identificando avances, brechas y oportunidades de mejora.

Otro eje central es la selección de corpus adecuados, que cuenten con etiquetas clasificatorias y representen distintas problemáticas reales. A partir de ello, se implementará un pipeline completo que incluirá etapas de preprocesamiento, entrenamiento y validación de modelos. Posteriormente, se evaluarán los resultados mediante métricas estandarizadas, permitiendo realizar comparaciones objetivas.

Finalmente, se analizarán los resultados obtenidos con el fin de elaborar recomendaciones que orienten la aplicación de estos modelos en distintos contextos, contribuyendo tanto al desarrollo científico como a su transferencia práctica en ámbitos académicos y profesionales.

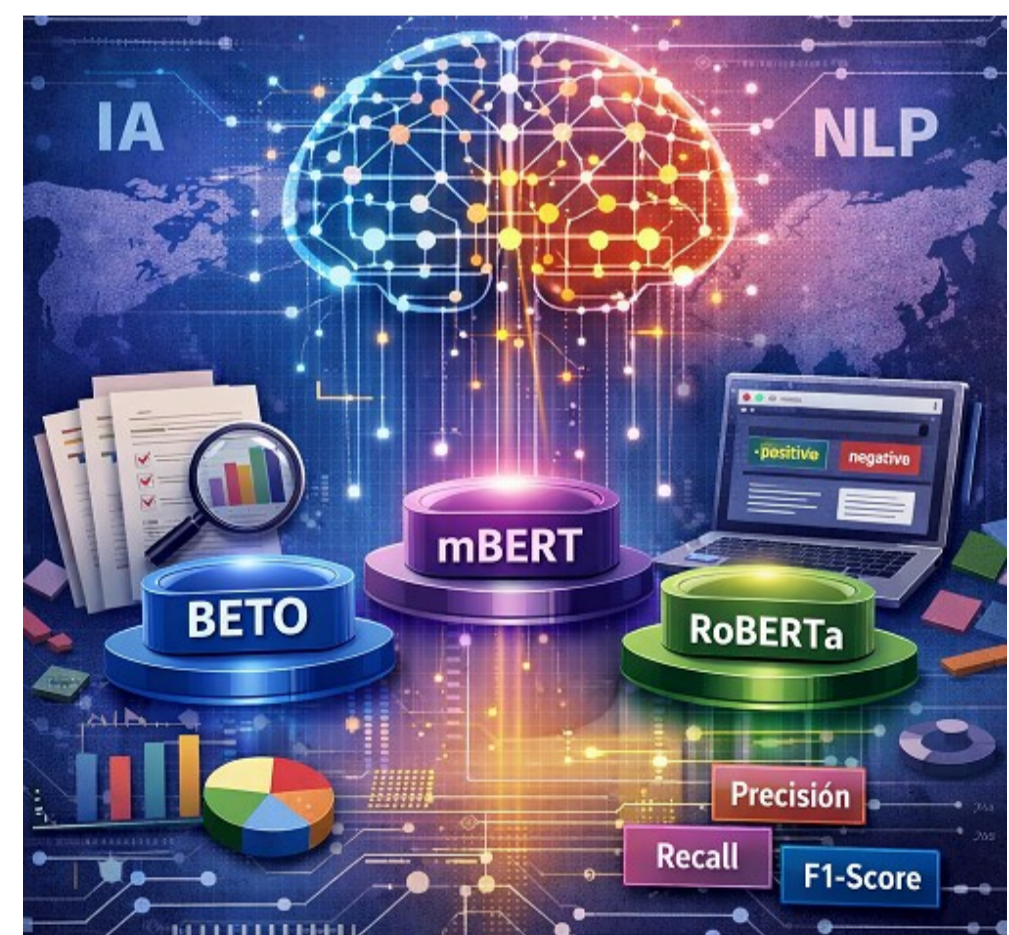
RESULTADOS ESPERADOS

Los resultados esperados del proyecto se orientan a generar aportes significativos en los planos técnico, pedagógico y social.

En el aspecto técnico, se prevé la elaboración de un ranking comparativo del desempeño de distintos modelos Transformer aplicados a la clasificación de textos en español, utilizando métricas estandarizadas como precisión, recall y F1-score. Asimismo, se espera formular recomendaciones prácticas para la selección de modelos en función del tipo de tarea, características del corpus y disponibilidad de recursos computacionales, lo que permitirá optimizar su implementación en diversos contextos.

En el plano pedagógico, los resultados contribuirán a la actualización de contenidos en asignaturas vinculadas a bases de datos y análisis numérico, incorporando avances recientes en inteligencia artificial.

Finalmente, en el ámbito social, el proyecto busca democratizar el acceso a herramientas de procesamiento del lenguaje en español, promoviendo el uso responsable y eficiente de la inteligencia artificial, así como la difusión del conocimiento generado en espacios académicos y científicos.



FORMACIÓN DE RECURSOS HUMANOS

El proyecto integrado por un equipo de tres docentes investigadores y prevé la incorporación de estudiantes de grado en actividades de investigación, busca fortalecer la formación académica y práctica, promoviendo el desarrollo de competencias en inteligencia artificial, análisis de datos y trabajo colaborativo en contextos científicos.