

Proyecto Práctico

Redes Neuronales

Universidad del Valle
Escuela de Ingeniería de Sistemas y Computación
2025-2

Este proyecto tiene como objetivo permitir a los estudiantes analizar el desempeño de las diferentes arquitecturas de red neuronal vistas en el curso y seleccionar las arquitecturas más adecuadas para dar solución a un **problema de clasificación de texto con múltiples etiquetas**, reforzando así los conocimientos obtenidos.

Inicialmente, **grupos de trabajo**, de **máximo 3 estudiantes**, deberán **seleccionar un conjunto de datos** entre las opciones indicadas (**ver sección 1**) ó proponer un conjunto de datos con característica similares.

Se realizará una sustentación parcial del proyecto para retroalimentación temprana y una entrega final con todo corregido. El proyecto además de las habilidades técnicas también evalúa la forma en que los estudiantes documentan el desarrollo de este y cómo presentan los resultados de manera verbal y/o escrita, ayudándoles a desarrollar habilidades de comunicación. **En la sección 2 se describen con detalle los entregables del proyecto y en la sección 3, la rubrica de evaluación a utilizar.**

Finalmente, para el desarrollo del proyecto se empleará el lenguaje de programación Python y librerías de aprendizaje automático como Keras, Tensorflow or Pytorch.

1. Conjuntos de Datos

Nombre	Amazon Reviews Dataset
Descripción	Reseñas reales de productos calificadas desde 1 a 5 estrellas, recolectada entre 2015 y 2019 en Inglés, Japonés, Alemán, Francés, Español y Chino.
Tamaño	200000 muestras por lenguaje.
Ejemplo	Etiqueta: 3 starts Título: <i>Not what I expected</i> Texto: <i>"If you're into thick tshirts, then this is the perfect shirt for you. It's also cut small. The length is good, but not good enough for me to keep it."</i>
Fuente	https://www.kaggle.com/datasets/mexwell/amazon-reviews-multi https://aclanthology.org/2020.emnlp-main.369/

Nombre	Twitter US Airline Sentiment Dataset
Descripción	Mensajes de Twitter en inglés sobre las principales aerolíneas estadounidenses clasificados por sentimientos como: positive, negative, neutral. Se recopilieron datos de Twitter en febrero de 2015
Tamaño	14640 muestras
Ejemplo	Etiqueta: <i>positive</i> Texto: “@VirginAmerica plus you’ve added commercials to the experience... tacky.”
Fuente	https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment

Nombre	20 newsgroup
Descripción	Colección de publicaciones de grupos de noticias agrupados en <i>20 categorías</i>
Tamaño	20000 muestras
Ejemplo	Path: cantaloupe.srv.cs.cmu.edu!das-news.harvard.edu!husc-news.harvard.edu!husc11.harvard.edu!jiu1 Newsgroups: comp.graphics (Etiqueta) Subject: Technical Help Sought Message-ID: <1993Apr6.112102.22553@husc3.harvard.edu> From: jiu1@husc11.harvard.edu (Haibin Jiu) Date: 6 Apr 93 11:21:00 EDT Organization: Harvard University Science Center Nntp-Posting-Host: husc11.harvard.edu Lines: 9 <i>Hi! I am in immediate need for details of various graphics compression techniques. So if you know where I could obtain descriptions of algorithms or public-domain source codes for such formats as JPEG, GIF, and fractals, I would be immensely grateful if you could share the info with me. This is for a project I am contemplating of doing.</i> <i>Thanks in advance. Please reply via e-mail if possible.</i> --hBJ
Fuente	https://archive.ics.uci.edu/dataset/113/twenty+newsgroups https://www.kaggle.com/datasets/crawford/20-newsgroups/data

2. Entregables

El proyecto consta de dos entregas (código + presentación oral) y un informe escrito. El máximo número de personas por grupo es 3.

Entrega 1 (10%)	<p>Tiene como objetivo el entrenamiento y evaluación de un modelo con el dataset seleccionado usando una arquitectura de Perceptrón Multicapa.</p> <p>Se debe entregar:</p> <ul style="list-style-type: none">• Código utilizado para obtener la solución empleando la librería keras, tensorflow o pytorch• Presentación de máximo 15 minutos, resumiendo la entrega, las diapositivas contienen introducción, condiciones de entrenamiento, evaluación, conclusiones, bibliografía y demo ejecutable. Para la elaboración de la presentación tome en cuenta los tips indicados en: https://youtu.be/3-2pwIEZR9U?si=4s_DuLDVTcvEWZum
Entrega 2 (10%)	<p>Tiene como objetivo el entrenamiento y evaluación de dos modelos con el dataset seleccionado usando una arquitectura de RNN sin memoria, y una red RNN con memoria (LSTM o GRU) o Transformer.</p> <p>Se debe entregar:</p> <ul style="list-style-type: none">• Código utilizado para obtener la solución empleando la librería keras, tensorflow o pytorch• Presentación de máximo 15 minutos, resumiendo la entrega, las diapositivas contienen introducción, condiciones de entrenamiento, evaluación, conclusiones, bibliografía y demo ejecutable. Para la elaboración de la presentación tome en cuenta los tips indicados en: https://youtu.be/3-2pwIEZR9U?si=4s_DuLDVTcvEWZum
Informe Escrito (15%)	<p>Reporte técnico describiendo el desarrollo de las dos entregas realizadas. Las partes que componen la estructura básica del reporte técnico y que se esperan en el reporte final del proyecto son:</p> <ul style="list-style-type: none">• Resumen: No más de 200 palabras describiendo el contenido del trabajo.• Tablas de contenido y figuras• Introducción: Se debe procurar que no se extienda más de dos páginas detallando ámbito, objetivos y resultados del trabajo.• Contenido: Debe contener tantas secciones como hagan falta para explicar el desarrollo del trabajo. Es necesario que exista una excelente capacidad de síntesis, precisión en el uso del lenguaje, facilidad de seguimiento de la exposición y que maneje información completa. Todas las secciones deben estar perfectamente ligadas y deben tener un objetivo claro para describir de manera breve pero completa el proceso para la elaboración del trabajo.• Conclusiones: Debe resumir los puntos fundamentales del trabajo y servir de síntesis a aquellas personas interesadas en el trabajo.• Bibliografía: Deben citarse las fuentes consultadas para el desarrollo del trabajo.

3. Rúbricas

Cada una de las entregas será evaluada con la siguiente rúbrica:

<i>Calidad en la presentación</i>	<i>Calidad en las respuestas</i>	<i>Código funcional</i>	<i>Calidad en las pruebas en el código</i>	<i>Elementos adicionales agregados</i>
1.2	1.2	1.2	1.2	0.2

El informe final será evaluado con la siguiente rúbrica:

<i>Calidad en la escritura. Contiene todas las secciones indicadas</i>	<i>Calidad en la descripción de los modelos utilizados</i>	<i>Calidad en el análisis de los resultados</i>	<i>Calidad en Gráficas, Figuras y Bibliografía</i>	<i>Elementos adicionales agregados</i>
1.5	1.5	1.5	0.3	0.2