# Sentiment Analysis CyED

* Juan David Garzón Diaz
* Geovanny Alexander Quintero Velez
* Johan Felipe Jojoa Bucheli

## Introduction

## Sentiment analysis, a subfield of natural language processing, involves discerning the sentiment expressed in textual data, typically categorizing it as positive or negative. In this project, the focus is on developing a sentiment analysis model using supervised learning with two types of neural networks: vanilla Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. The significance of sentiment analysis lies in its applications across various domains, providing valuable insights into the emotions and opinions expressed in textual content.

## Objectives

The primary objectives of this project are centered around the development and evaluation of sentiment analysis models. The dataset utilized is sourced from three distinct websites: Amazon, IMDb, and Yelp. Each sentence in the dataset is annotated with a binary label, indicating whether it conveys a positive sentiment (labeled as 1) or a negative sentiment (labeled as 0). The overarching goals include constructing robust sentiment analysis models using RNNs and LSTMs and conducting a comprehensive evaluation of their performance. Additionally, the project aims to identify and optimize the hyperparameters that contribute to the models' effectiveness.

## Dataset details

The sentiment dataset used is the "Sentiment Labelled Sentences Dataset" obtained from the UC Irvine Machine Learning Repository. The dataset consists of labeled sentences from three websites: Amazon, IMDb, and Yelp. Each website contributes 500 positive and 500 negative sentences, resulting in a balanced dataset. The dataset is prepared for supervised learning, with sentences labeled as 1 for positive and 0 for negative sentiments.

## Preprocessing steps

Preprocessing steps involve loading the data from multiple files, concatenating them into a single DataFrame, and shuffling the data for randomness. The NLTK library is used for tokenization, lowercasing, and removing stopwords. The preprocessing function, preprocess\_text, tokenizes the sentences, converts them to lowercase, and removes stopwords, ensuring that only meaningful words are retained for analysis. The final DataFrame includes the original sentences as well as a new column, 'processed\_sentence,' containing the preprocessed text.

## RNN and LSTM sentiment analysis models

## Performance evaluation

## Turing neural networks implementation

## Comparative analysis

## Conclusion