ITESM Campus Monterrey

**Proof of Concept: Sentiment Analysis Using Supervised Learning with Word Embeddings**

Luis Alberto Portilla López

Research Stay - Going beyond Artificial Intelligence: Artificial Emotions

TC3073 | Group 573

## Introduction

Sentiment analysis is a critical application of natural language processing (NLP), enabling organizations to analyze and respond to customer feedback, social media trends, and product reviews. Supervised learning approaches, combined with advanced text representation techniques like word embeddings, have significantly improved sentiment classification accuracy. This Proof of Concept (PoC) demonstrates how supervised models can be trained using word embeddings to achieve superior performance in sentiment analysis tasks.

## Business Problem

The ability to understand and classify sentiment in text data is essential for making data-driven decisions in areas like marketing, customer service, and product development. However, sentiment analysis faces several challenges:

* Semantic Understanding: Traditional text representations like bag-of-words fail to capture semantic relationships.
* Scalability: Handling large volumes of diverse text data efficiently.
* Accuracy: Ensuring consistent and reliable sentiment classification across various domains.

## Proposed Solution

This PoC proposes a supervised learning pipeline for sentiment analysis, utilizing word embeddings as text representations. The approach includes:

* Text Representation: Use pre-trained word embeddings such as Word2Vec or GloVe to capture semantic relationships.
* Supervised Models: Train classifiers like Logistic Regression, Support Vector Machines (SVMs), and Neural Networks on labeled sentiment datasets such as IMDb reviews or Twitter Sentiment Analysis.
* Feature Engineering: Experiment with combining embeddings with additional features (e.g., sentence length, POS tags).
* Evaluation: Measure performance using metrics like accuracy, precision, recall, and F1-score.

## Expected Outcomes

The implementation is expected to achieve:

* Improved Sentiment Classification: Higher accuracy and consistency using advanced text representations.
* Scalable Solutions: Demonstrate the applicability of the approach to diverse domains and datasets.
* Insightful Analysis: Provide a deeper understanding of how word embeddings enhance supervised learning models.

## Conclusion

This PoC outlines a comprehensive approach to supervised sentiment analysis, emphasizing the role of word embeddings in improving model performance. By addressing the challenges of semantic understanding and scalability, this solution offers a framework for developing robust sentiment analysis systems. The findings will inform future research and practical applications in text classification tasks.