ITESM Campus Monterrey

**Proof of Concept: Domain Adaptation for Financial Text Analysis Using Transfer Learning**

Luis Alberto Portilla López

Research Stay - Going beyond Artificial Intelligence: Artificial Emotions

TC3073 | Group 573

## Introduction

Transfer learning and domain adaptation have transformed natural language processing (NLP) by enabling models to generalize knowledge from one domain and apply it to another. This is particularly useful in domains with limited labeled data, where training models from scratch is infeasible. This Proof of Concept (PoC) demonstrates the application of transfer learning for adapting pre-trained language models to financial text analysis tasks.

## Business Problem

The financial sector generates extensive textual data, such as earnings reports, SEC filings, and analyst notes. Effectively analyzing this text is critical for decision-making but faces challenges such as:

* Domain-Specific Jargon: Financial texts often contain specialized terminology and patterns.
* Data Scarcity: Limited availability of labeled datasets for supervised learning in finance.
* Generic Models: Difficulty in achieving high performance with models not adapted to the financial domain.

## Proposed Solution

This PoC proposes leveraging transfer learning to fine-tune pre-trained language models for financial text analysis. The approach includes:

* Model Selection: Use a pre-trained model like BERT or RoBERTa as the base model.
* Dataset: Utilize financial datasets such as SEC filings, earnings call transcripts, or FinBERT datasets for fine-tuning.
* Domain Adaptation: Fine-tune the pre-trained model to capture the specific language and context of financial texts.
* Evaluation: Perform tasks like sentiment analysis of financial reports or classification of SEC filings, using metrics such as accuracy, F1-score, and precision.
* Comparison: Benchmark performance against baseline models trained without transfer learning.

## Expected Outcomes

The implementation is expected to achieve:

* Improved Text Analysis Accuracy: Enhanced performance in domain-specific tasks due to transfer learning.
* Domain Adaptability: Demonstration of the effectiveness of fine-tuning for handling specialized jargon and patterns.
* Scalable Approach: Insights into adapting pre-trained models for other domains with minimal labeled data.

## Conclusion

This PoC highlights the potential of transfer learning and domain adaptation in advancing financial text analysis. By leveraging pre-trained models and fine-tuning them on domain-specific datasets, this approach addresses the limitations of generic models and provides robust solutions for specialized NLP tasks. The findings will inform future applications and research in domain-specific NLP.