

## Synthesis of lab 4 (NLP)

### 1. Introduction

In this lab, we explored various Natural Language Processing (NLP) techniques, including classification and regression using RNN models, text generation using GPT-2, and text classification using BERT. Our objective was to gain hands-on experience with different NLP models and understand their applications in real-world scenarios. Here's a synthesis of my key learnings:

### 2. Classification and Regression with RNN Models

We began by implementing RNN-based models to classify and predict text data. This part involved:

- **Data Collection and Preprocessing:** We used web scraping libraries to collect text data from Arabic websites, followed by preprocessing steps such as tokenization, stemming, lemmatization, and stop words removal.
- **Model Training:** We trained RNN, Bidirectional RNN, GRU, and LSTM models to classify and predict the relevance of texts.
- **Evaluation:** The models were evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared metrics. The Bidirectional RNN model performed the best, highlighting its ability to capture context from both directions.

### 3. Text Generation with GPT-2

Next, we focused on text generation using the GPT-2 model:

- **Fine-Tuning GPT-2:** We fine-tuned the pre-trained GPT-2 model on a custom dataset of text prompts. This involved tokenizing the data, creating a DataLoader, and training the model.
- **Text Generation:** After fine-tuning, we generated text based on given prompts. The model demonstrated efficient and coherent text generation capabilities, showcasing the power of transfer learning and pre-trained language models.

### 4. Text Classification with BERT

Finally, we used the BERT model for text classification tasks:

- **Data Preparation:** We loaded and preprocessed a JSON dataset containing Amazon Fashion reviews. This included tokenizing the review text and preparing input tensors for BERT.
- **Model Training:** We fine-tuned the BERT model for classification, training it to predict the overall rating of reviews.

- **Evaluation:** The BERT model achieved high accuracy (98.42%) and F1 score (98.40%), demonstrating its effectiveness in understanding and classifying text data.

## 5. Conclusion

This lab provided a comprehensive overview of various NLP techniques and models, from RNN-based classification and regression to advanced text generation with GPT-2 and classification with BERT. We learned how to preprocess text data, fine-tune pre-trained models, and evaluate their performance. These skills are crucial for numerous NLP applications, including sentiment analysis, document classification, machine translation, and information retrieval. The hands-on experience with different models and datasets has deepened our understanding of NLP and its practical applications.