Internship report

Name : Chris Anil

Internship title : Real time translation model

## Introduction

During the course of this internship with NullClass, I was tasked with implementing a machine translation model that translates English to French. The goal was to improve translation quality using beam search decoding and evaluate it with BLEU scores. This project allowed me to dive deep into NLP workflows and model evaluation techniques.

## Background

I had already worked on basic English to French and English to Spanish translation using LSTM models. Task 1 pushed me to integrate advanced decoding techniques like beam search, making it more performance-driven. The main objective was to apply beam search on a pre-trained English-to-French model and evaluate the output accuracy — without relying on a GUI.

## Learning Objectives

- Understand the internal mechanics of beam search decoding

- Apply evaluation metrics like BLEU score for NMT quality comparison

- Work with Keras models, tokenizers, and sequence padding

- Organize and prepare machine learning deliverables for GitHub submission

## Activities and Tasks

- Loaded and tested a pre-trained English-to-French LSTM model

- Integrated beam search decoding alongside greedy decoding

- Created a structured `.ipynb` notebook showcasing both approaches

- Evaluated translations using BLEU scores and smoothing techniques

- Prepared supporting files: model weights, tokenizers, and sequence length

## Skills and Competencies

- Machine Learning (NLP): Worked with sequence-to-sequence models using TensorFlow/Keras

- Python: Handled file operations, decoding logic, and data manipulation

- Evaluation Metrics: Implemented BLEU scoring with NLTK for comparative analysis

- Debugging & File Handling: Resolved multiple file-not-found and model-loading issues independently

- Documentation: Maintained a clean Jupyter notebook for reproducibility and submission

## Feedback and Evidence

The final notebook `TASK1\_FIXED\_EngToFrench.ipynb` includes both decoding methods and BLEU comparisons for five test English sentences. All the required support files (`.h5`, tokenizers, sequence length JSON) were generated and organized systematically. The notebook ran without any errors after handling import paths and loading issues, showing complete functionality.

## Challenges and Solutions

- FileNotFoundError: The model filename had been renamed during download. I manually renamed it and verified visibility using Python's os module.

- Model Loading Issues: Initially, dummy `.h5` files lacked config metadata. I replaced them with valid trained models.

- Import Errors: Forgot to import `load\_model` in one cell; I resolved it by placing all imports in the topmost cell.

Each challenge helped me better understand debugging in Python and TensorFlow environments.

## Outcomes and Impact

By completing this task, I now have hands-on experience with:

- Implementing beam search in real NLP workflows

- Evaluating machine translation quality using standard metrics

- Preparing submission-ready ML projects that meet industry standards

This task gave me the confidence to work independently on model evaluation and deployment-ready code.

## Conclusion

This internship task taught me more than just model implementation. It pushed me to think about reproducibility, evaluation, and project organization. I feel better equipped to contribute to real-world NLP projects.

I would like to sincerely thank NullClass for offering this learning opportunity. The task was well-structured, and I genuinely appreciate the focus on self-research and independent execution. This experience has added a valuable skill set to my development journey.