# Internship Report

## Internship at MIT MAHE, India

Intern: **Mohamed Ahmed Mansour**

Supervisor: **Prof. Ramakrishna**

Duration: 6 Weeks (Ongoing)

## 1. Introduction

I completed a 6-week internship at MIT MAHE, India, under the guidance of Prof. Ramakrishna. The main project focused on developing a **Dual-Branch Fake News Detection Framework**, which integrates semantic analysis (BERT) and knowledge graph reasoning (TransE). In addition, I explored polysemy in Natural Language Processing (NLP), implemented multiple versions of the framework, and experimented with supporting projects related to fake news and news classification.

## 2. Weekly Work Summary

**Week 1**

- Read two foundational research articles provided by Prof. Ramakrishna.
- Studied the concept of **polysemy in NLP** and its impact on news interpretation.
- Learned **text preprocessing** basics, including cleaning, tokenization, and handling unstructured data.

**Week 2**

- Deepened understanding of **text processing techniques** such as stopword removal, stemming, lemmatization, and POS tagging.
- Explored word representation methods (One-Hot, TF-IDF, embeddings).
- Began work on the **first version** of the dual-branch framework.
- Attempted to train BERT and TransE models, but the pipeline lacked a fusion model and was unsuccessful.

**Week 3**

- Researched to improve the framework pipeline → developed **version 0.2**.
- This version included a fusion model combining BERT and TransE, but it did not work effectively.
- As a side project, implemented a **Fake News Detection system** using traditional ML models (Logistic Regression, SVM, Random Forest).
- This project worked well and achieved promising results.

**Week 4**

- Developed **version 0.3** of the framework.
- Successfully integrated **triplet extraction (REBEL)**, **fuzzy matching**, BERT text branch, and TransE knowledge branch.
- This marked a significant improvement compared to earlier attempts, though optimization challenges remained.

**Week 5**

- Started an additional project on **News Category Classification**.
- Enhanced version 0.3 of the dual-branch framework:
• Added a **GUI** using Gradio.
• Included a proper license and improved project documentation.
- Drafted the **first version of the research paper** describing the framework and experiments.

**Week 6 (Ongoing)**

- Preparing the **internship report** (this document).
- Refining the **research paper** for submission and review.

## 3. Skills and Tools Learned

**Technical Skills:**
- Text preprocessing, tokenization, stemming, lemmatization.
- Machine Learning (Logistic Regression, SVM, Random Forest).
- Deep Learning with **BERT**, **TransE**, and fusion models.
- Knowledge Graph alignment and triplet extraction using **REBEL**.
- Libraries: PyTorch, Hugging Face Transformers, OpenKE.

**Soft Skills:**
- Research reading and summarization.
- Debugging and improving ML pipelines.
- Writing academic papers.
- Documentation and reproducibility in GitHub projects.

## 4. Challenges and Solutions

- **Challenge:** Early versions of the framework failed due to pipeline complexity.
**Solution:** Iteratively improved versions (0.1 → 0.3), incorporated fusion models, and conducted extensive research.

- **Challenge:** Balancing multiple side projects.
**Solution:** Used time management to work on smaller projects (Fake News Detection, News Classification) alongside the main framework.

- **Challenge:** Integrating NLP models with Knowledge Graph embeddings.
**Solution:** Implemented triplet extraction and fuzzy matching modules to ensure semantic-factual alignment.

## 5. Conclusion and Reflection

Over the course of five weeks, I progressed from **theoretical learning (polysemy, preprocessing, embeddings)** to **hands-on implementation (dual-branch framework, ML baselines, classification projects)**. The internship strengthened both my **research foundations** and **practical ML skills**.

The dual-branch framework project remains the highlight, representing a novel approach that combines semantics and factual reasoning for **fake news detection**. This internship has provided me with valuable exposure to research, coding, and academic writing, and

will strongly support my future career in AI and NLP.