



Internship Report

Internship at Manipal Institute of Technology (MIT),
Manipal Academy of Higher Education (MAHE), India

Intern: **Mohamed Ahmed Mansour**

Supervisor: **Prof. Ramakrishna**

Duration: 6 Weeks (Ongoing)

Abstract

This report documents my 6-week internship at Manipal Institute of Technology (MIT), Manipal Academy of Higher Education (MAHE), India, under the supervision of Prof. Ramakrishna. The primary focus was the development of a **Dual-Branch Fake News Detection Framework** that combines semantic reasoning (BERT) with factual verification (TransE + Knowledge Graphs).

The internship also included studying **polysemy in NLP**, implementing traditional machine learning baselines, and experimenting with a **news category classification** model. Weekly progress involved moving from theoretical study to building practical ML/DL pipelines and producing a first draft of a research paper.

The experience enhanced my technical expertise in natural language processing and deep learning, improved my research and documentation skills, and prepared me for future academic and industry challenges.

Table of Contents

1. Introduction
2. Weekly Work Summary
3. Skills and Tools Learned
4. Datasets Used
5. Challenges and Solutions
6. Conclusion and Future Work
7. Acknowledgements
8. References

1. Introduction

This internship provided the opportunity to work on an advanced NLP project under academic guidance at MIT MAHE. The main research problem was to detect fake news by combining **semantic analysis** and **knowledge-based reasoning**. The framework integrates **BERT** embeddings for contextual understanding with **TransE** embeddings from triplets aligned to a knowledge graph. This dual-branch design seeks to balance *how news is written* with *what facts it conveys*.

The report is structured to present weekly progress, technical and soft skills learned, datasets used, challenges encountered, solutions applied, and future directions.

2. Weekly Work Summary

Week 1

- Read two research articles recommended by Prof. Ramakrishna.
- Studied **polysemy in NLP** and its effect on interpretation of news.
- Learned preprocessing: cleaning, tokenization, stopword removal.
- Repository: Polysemy NLP Project

Week 2

- Explored stemming vs. lemmatization, POS tagging, n-grams.
- Learned about embeddings: One-Hot, TF-IDF, Word2Vec, BERT.
- Built **first version** of the framework, trained BERT + TransE (failed fusion).

Week 3

- Improved pipeline to develop **version 0.2** (fusion module, still unstable).
- Built ML baselines for fake news detection: Logistic Regression, SVM, Random Forest.
- Repository: Fake News Detection

Week 4

- Created **version 0.3** with REBEL triplet extraction + fuzzy matching.
- Stable prototype combining BERT + TransE via fusion.
- Repository: Dual-Branch Framework

Week 5

- Developed **News Category Classification** model.
- Added GUI (Gradio), license, documentation for framework.
- Wrote first draft of research paper.
- Repository: News Category Classification

Week 6 (Ongoing)

- Writing this report and improving research paper.

3. Skills and Tools Learned

Technical: Preprocessing, tokenization, stemming/lemmatization, embeddings, ML (LogReg, SVM, RF), DL (BERT, TransE), REBEL, Gradio, Hugging Face, PyTorch, OpenKE.

Soft: Research reading, debugging, project documentation, collaboration, academic writing, time management.

4. Datasets Used

- **FakeNewsNet:** aggregated dataset (PolitiFact + GossipCop).
- **PolitiFact:** fact-checking political news dataset.
- **GossipCop:** celebrity/entertainment fake/real news dataset.
- **LIAR:** short statements labeled for truthfulness.

5. Challenges and Solutions

- **Failed pipeline (early versions):** fixed by restructuring workflow, adding robust fusion.
- **KG alignment issues:** solved with REBEL + fuzzy matching.
- **Compute constraints:** solved by tuning hyperparameters and using ML baselines as guides.

6. Conclusion and Future Work

This internship was an academic and technical growth journey. It combined research, coding, and project building. The highlight was the dual-branch framework, integrating semantics and factual reasoning for fake news detection.

Future Work: improve fusion, extend to multimodal (text+images), expand datasets, and polish research paper for publication.

7. Acknowledgements

I sincerely thank **Prof. Ramakrishna** for his guidance, valuable feedback, and constant encouragement throughout this internship. I am grateful to **MIT MAHE** for providing the opportunity and resources for this research. I also appreciate the support of peers and the research environment, which helped me overcome challenges and make steady progress.

8. References

- Polysemy NLP Project
- Fake News Detection
- Dual-Branch Framework
- News Category Classification
- Datasets: FakeNewsNet, PolitiFact, GossipCop, LIAR