MULTI-HOP CLAIM VERIFICATION

Team 7:
Gireeshee Pendela(gp590)
Veera Jeeshitha Kolla(vk536)

Problem Statement

- The goal of this project is to build an automated system that can verify the factual correctness of a given claim by analyzing textual information.
- Each claim must be classified into one of three categories:
    1. Supported: Evidence confirms the claim
    2. Refuted: Evidence contradicts the claim
    3. Not Enough Info(NEI): Evidence is insufficient or missing

NLP Task:

- It is a NLP text classification task
- The model reads the claim written in natural language
- It predicts whether the claim is supported by known facts

DATASET:

- FEVER - Large-scale fact verification dataset(~185k claims)
- Each claim labeled as supported/ refuted/ NEI
- Based on real evidence from Wikipedia
- Standard benchmark for claim verification research

APPROACH & PROGRESS

Training Setup:

- The FEVER dataset is split into 70% training, 15% validation, and 15% testing to build, tune, and evaluate the models reliably.

Feature Representation:

- Each claim is converted to numerical form using TF-IDF, capturing important words and short phrases that help differentiate between Supported, Refuted, and NEI labels.

Baseline Models Implemented:

- We trained multiple classical machine learning models including Logistic Regression, Support Vector Machine (SVM), and Random Forest, using TF-IDF features for claim classification.

Evaluation Strategy:

- Model quality is measured using Accuracy and Macro-F1, to ensure fair evaluation across all three classes.

Current Progress:

- Data preprocessing, training, and evaluation are successfully completed, and the system can now classify new claims with the baseline models.

RESULTS & NEXT STEPS:

Model	Accuracy	Macro-F1
Logistic Regression	0.694	0.688
SVM (Best)	**0.719**	**0.718**
Random Forest	0.711	0.700

KEY OBSERVATIONS:

- SVM performed the best among the classical models on our TF-IDF features.
- Random Forest performed better than expected, possibly because the sampled dataset captured some nonlinear patterns.
- Logistic Regression was stable but slightly behind the other two models.

NEXT STEPS:

- Fine-tune BERT/RoBERTa for improved contextual understanding.
- Compare transformer results with our baselines using the same metrics.

THANK YOU