FINANCIAL NEWS ENTITY EXTRACTION

Presented by Group - 5

Our Team

AP23110011513 - Z.Udaya lakshmi

AP23110011328 - M.Vyshnavi

AP23110010953 - CH.Suhass

AP23110010976 - S.Praneeth

AP23110010645 - R.Pujith

Problem Statement & Objectives

Problem Statement

Financial news articles contain critical information like company names, stock tickers, currency values, mergers, acquisitions, and economic events.
However, manually extracting these details is time-consuming and error-prone.

Objectives

- Automatically extract key financial entities from news text
- Build a custom NER model trained from scratch
- Convert financial text into structured insights
- Provide a simple web interface for real-time entity extraction

Proposed Solution & Workflow

Proposed Solution

Develop a complete pipeline that:

- Processes raw financial text
- Tokenizes using a custom-trained tokenizer
- Predicts entities using a scratch-trained NER model
- Displays extracted entities via a Flask web interface

Workflow

1. Dataset Loading (FiNER)
2. Text Preprocessing → Tokenization → Label Alignment
3. Train Token Classification Model (Transformers)
4. Evaluate & Save the Model
5. Deploy using Python Flask
6. User inputs text → System returns detected entities

Dataset & Methodology

Dataset

FiNER Dataset (Financial Named Entity Recognition)
Contains financial news tokens labeled with:

- Company (ORG)
- Currency/Money (MONEY)
- Stock Ticker (TICKER)
- Financial Events (EVENT)
- Date/Time (DATE)
- Non-entity (O)

Methodology

- Load dataset (gtfintechlab/finer-ord)
- Standardize tokens + labels
- Custom train SentencePiece tokenizer (16k vocab)
- Train a Transformer token classification model
- 6 layers
- 256 hidden size
- 7-label output
- Evaluate using SeqEval metrics (Precision, Recall, F1)
- Deploy model via Flask API

Technologies Used

Programming Languages

- Python
- JavaScript

ML Frameworks

- PyTorch
- HuggingFace Transformers
- SentencePiece Tokenizer

Tools & Libraries

- Datasets
- SeqEval
- Evaluate
- Flask (Web Framework)
- HTML, CSS, JS

Environment

- Google Colab (Training)
- Local Deployment (Flask Server)

Results & Findings

Model Results

- Successfully trained NER model from scratch
- Achieved meaningful detection of:
  - Company names
  - Monetary values
  - Stock tickers
  - Acquisition / merger terms
  - Financial events
- Model outputs consistent labels in real testing
- Web interface accurately displays extracted entities

Qualitative Results

Example Input:

“Apple announced a $2 billion acquisition of PayPlus today.”

Output Entities:

- Apple → ORG
- $2 billion → MONEY
- acquisition → EVENT
- PayPlus → ORG

Conclusion & Future Scope

Conclusion

- Developed a complete end-to-end Financial NER System
- Model trained fully from scratch
- Web interface enables easy, interactive analysis
- Helps convert unstructured financial text into actionable insights

Future Scope

- Add offset-based highlighting for perfect UI accuracy
- Integrate stock API to link companies → tickers → market data
- Add graphical dashboards (donut charts, entity frequency trends)
- Deploy as a cloud-based REST API for large-scale analytics

Q&A Session?

THANK YOU!