

FINANCIAL NEWS ENTITY EXTRACTION



Group Members

Supriya

Y.Dedeepya

Y.Sahithi

P.Evangeline

Tasneem



Problem statement

Financial news contains large volumes of unstructured text. Manually identifying key entities such as companies, currencies, stock tickers, and financial events is slow and inconsistent.

Objectives

- To develop an automated NER system for financial news articles.
- To accurately extract key entities such as companies, tickers, currencies, and financial events.
- To train and evaluate the model using financial-domain datasets like FiNER.
- To generate structured outputs that support financial analysis and decision-making.



Proposed solution

Develop an automated Financial News NER system that processes news articles, identifies domain-specific entities, and outputs structured information. The system will use machine learning or deep learning models trained on financial datasets like FiNER to ensure high accuracy and domain relevance.

Workflow

1. Preprocess Data - Collect and clean financial news text.
2. Train Model - Use FiNER/custom data to train the NER model.
3. Extract Entities - Identify companies, tickers, currencies, and events.
4. Output Results - Generate structured results for analysis.



Dataset

- FiNER-139 Dataset - A financial-domain NER dataset containing annotated news articles.
- Includes labels such as companies, stock tickers, financial events, currencies, and economic terms.

Methodology

- Data Preparation - Clean text, tokenize, and format for NER training.
- Model Training - Train a machine learning or deep learning NER model using financial datasets.
- Entity Extraction - Apply the model to new financial news to detect key entities.
- Result Generation - Output extracted entities in structured form (tables/JSON).



Technology used

- Python for data processing and model development
- NLTK / spaCy for text preprocessing
- Scikit-learn / TensorFlow / PyTorch for NER model training
- Pandas & NumPy for data handling
- Jupyter/Colab for experimentation and testing



Results & Findings

The screenshot shows a user interface for analyzing financial news text. At the top, there is a text input field labeled "Enter Financial News Text" containing the following text:

Fed Raises Interest Rates by 0.25% to Combat Inflation
2024-01-23
The Federal Reserve announced a 0.25 percentage point increase in the federal funds rate on Wednesday, bringing the target range to 5.25% to 5.50%. Fed Chair Jerome Powell stated that the decision reflects ongoing efforts to bring inflation down to the 2% target. The move marks the 11th rate increase since March 2022. Stock markets reacted negatively, with the S&P 500 dropping 1.2% and the Dow Jones Industrial Average falling 350 points. The 10-year Treasury yield rose to 4.35%.

Below the text input is a blue button labeled "Analyze Entities".

Under the heading "Analysis Results", there is a section titled "Your Entered Text:" which displays the same text as the input field, with entities highlighted in different colors (e.g., "Fed" in blue, "0.25%" in green, "Wednesday" in red). Below this is another block of text showing the analyzed entities extracted from the input.

- The NER model successfully extracted key financial entities such as companies, tickers, currencies, and financial events.
- Achieved high accuracy and consistent entity detection on financial news articles.
- Model performed better on domain-specific data (FiNER-139) compared to general datasets.
- Structured outputs (tables/JSON) improved readability and supported faster financial analysis.

Conclusion

- The project successfully built an NER system that accurately extracts key financial entities from news articles.
- Using domain-specific data improved model performance and reliability.
- The system helps transform unstructured financial text into structured insights for faster decision-making.



Future Scope

- Integrate real-time financial news for live entity extraction.
- Expand to additional entity types like sentiment and market movements.
- Deploy as a web or API-based tool for broader accessibility.
- Enhance accuracy using advanced transformer models (BERT/FinBERT).

THANK YOU

