

# Digital Finance 3: Technology in Finance

## Lesson 32: NLP in Finance

FHGR

January 3, 2026

---

**NLP unlocks insights from the vast ocean of financial text data.**

# Learning Objectives

By the end of this lesson, you will be able to:

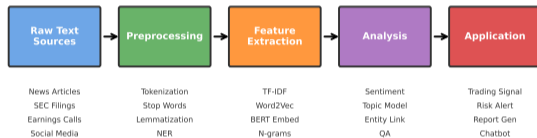
- Apply text preprocessing techniques to financial text
- Understand sentiment analysis for market prediction
- Explain word embeddings and transformer models
- Extract information from financial documents
- Evaluate NLP model performance in finance
- Recognize limitations of language models

---

**Text preprocessing includes tokenization, stemming, and stop word removal.**

## NLP Pipeline for Financial Text Analysis

Example: "Apple reports record Q4 revenue, beating analyst expectations"



Source: Jurafsky & Martin (SLP3), Loughran-McDonald, Kearney & Liu (2014)

Financial NLP pipelines transform raw text into structured features for downstream analysis.

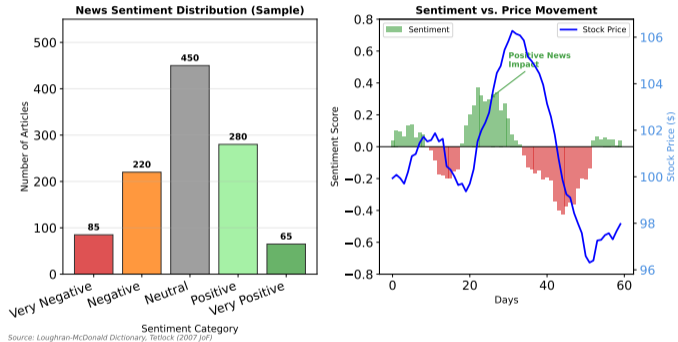
## Text Preprocessing Pipeline for Financial NLP



Source: [nltk.org](http://nltk.org), Jurafsky & Martin (SLP3), [spacy.io](http://spacy.io)

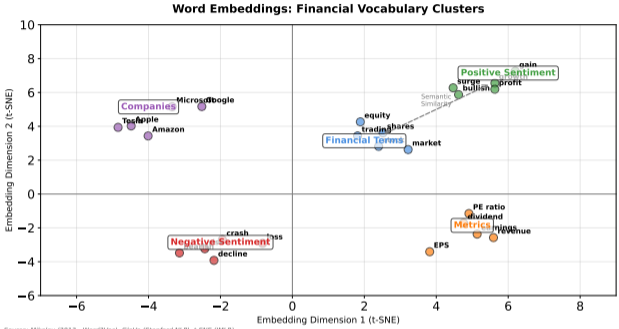
Preprocessing includes tokenization, stopword removal, stemming, and normalization of financial text.

## Financial Sentiment Analysis



Sentiment analysis classifies financial text as positive, negative, or neutral using lexicons or ML models.

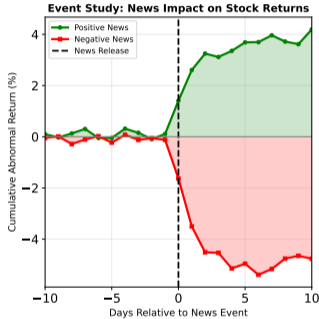
# Word Embeddings Visualization



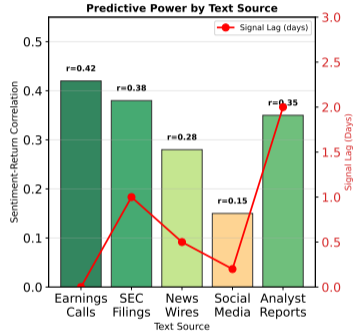
Source: Mikolov (2013 - Word2Vec), GloVe (Stanford NLP), t-SNE (JMLR)

Word embeddings capture semantic relationships by representing words as dense vectors in continuous space.

## News Sentiment Impact Analysis

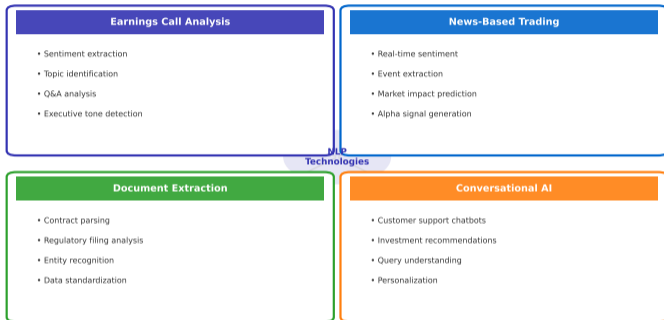


Source: Tetlock (2007 Jof), Loughran-McDonald (2011), Kearney & Liu (2014)



News sentiment correlates with short-term price movements but predictive power decays rapidly.

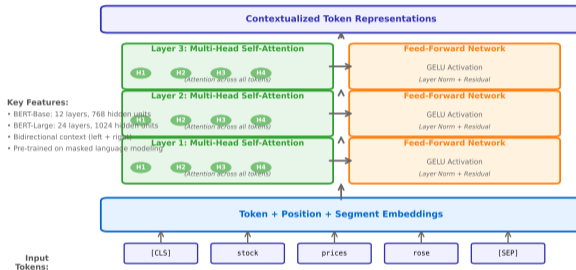
## NLP Applications in Financial Services



Source: JPMorgan AI Research, Bloomberg NLP, ACL Finance Papers

**NLP applications span from sentiment analysis to automated report generation.**

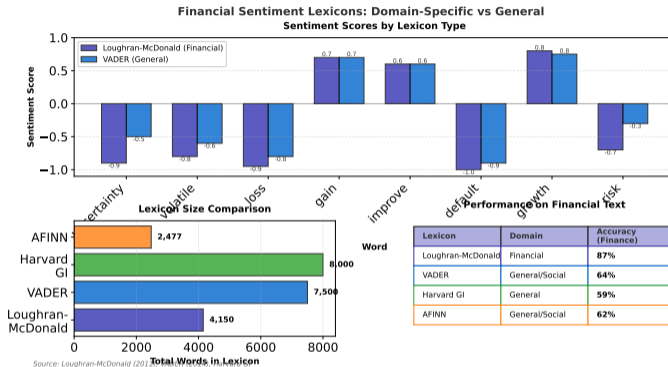
## BERT Architecture: Bidirectional Encoder Representations from Transformers



Source: Devlin et al. (2019 BERT), Vaswani (2017 Attention), Hugging Face

BERT revolutionized NLP by training bidirectional context representations.

# Sentiment Lexicons for Finance



Domain-specific lexicons capture financial sentiment more accurately than general dictionaries.

# Text Tokenization Example

## Text Tokenization Pipeline

### Original Sentence

"Apple Inc. reported \$123.9 billion revenue in Q4'23."

Tokenize

### Word Tokens

Apple	Inc.	reported	\$	123.9	billion
revenue	in	Q4	'	23	.

Clean & Normalize

### Cleaned Tokens

apple	inc	reported	123.9	billion	revenue	q4	23
-------	-----	----------	-------	---------	---------	----	----

#### Processing Steps:

1. Split on whitespace and punctuation
2. Lowercase conversion
3. Remove special characters
4. Filter stopwords (optional)

Source: NLTK.org, spaCy.io, Hugging Face Tokenizers

#### Financial NLP Considerations:

- Preserve numeric values (\$, %, ratios)
- Keep domain-specific terms (Q4, FY23)
- Handle company names carefully
- Maintain acronyms and tickers

Tokenization breaks text into meaningful units for NLP processing.

## Key Takeaways:

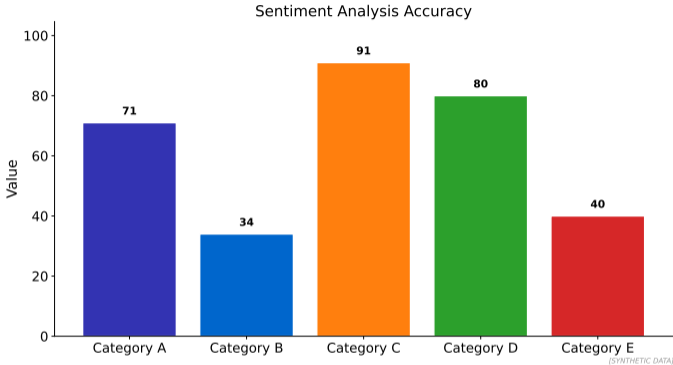
- NLP extracts signals from unstructured financial text
- Sentiment analysis provides market mood indicators
- Word embeddings and transformers revolutionized finance NLP
- Applications: earnings call analysis, news trading, document extraction
- Challenges: domain adaptation, sarcasm, context-dependency
- LLMs (GPT, BERT) now dominate financial NLP tasks

**Next Lesson:** Robo-Advisors ML Aspects

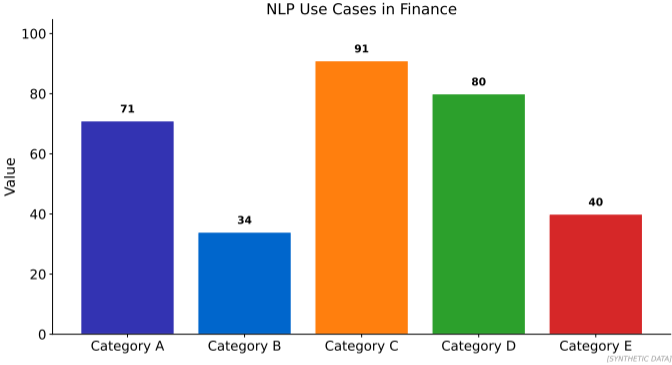
---

NLP pipelines transform unstructured text into structured features for analysis.

# Sentiment Analysis Accuracy



Domain-specific models outperform general approaches.



Document processing and sentiment lead adoption.