

21CSE426T – Financial Machine Learning
Unit 5 Assessment
Group Assignment

Set 1

HOTS:

1. Explain how NLP can enhance the process of financial fraud detection. What are the limitations of using NLP in this context?
2. Describe how word embeddings like Word2Vec differ from traditional bag-of-words models in representing textual data. How might this affect sentiment analysis results?

MCQs:

1. What is the main purpose of Natural Language Processing?
 - A. Data encryption
 - B. Analyzing numerical data
 - C. Enabling computers to understand human language
 - D. Developing operating systems
2. Named Entity Recognition (NER) primarily identifies:
 - A. Synonyms in a text
 - B. Named entities such as names, places, and organizations
 - C. Part-of-speech categories
 - D. Document structure
3. In sentiment analysis, a word embedding technique helps to:
 - A. Count word frequencies
 - B. Convert text into vector format
 - C. Recognize entities in text
 - D. Analyze document structure
4. Tokenization is used to:
 - A. Encrypt data
 - B. Split text into smaller parts
 - C. Organize documents
 - D. Filter irrelevant words
5. Financial sentiment analysis is challenging because:
 - A. Financial texts are often too short
 - B. Financial terms are context-sensitive
 - C. News articles are all unbiased
 - D. Financial data lacks numerical values

Set 2

HOTS:

1. Compare and contrast the use of Naive Bayes and LDA for sentiment analysis in financial news data. How does each model impact the interpretation of results?
2. Explain the significance of event return in sentiment analysis-based trading strategies. How does it differ from simple return calculations?

MCQs:

1. Which of the following is a common step in text preprocessing for NLP?
 - A. Text encryption
 - B. Tokenization
 - C. Data sorting
 - D. Error correction
2. Word2Vec is a type of:
 - A. Text classification algorithm

- B. Word embedding model
- C. Language translation tool
- D. Syntax parser
- 3. NLP can be used in finance to analyze:
 - A. Loan repayments
 - B. Market sentiment from news
 - C. Database management
 - D. File compression
- 4. Which Python package is widely used for building NLP applications and has modules for tokenization, stemming, and more?
 - A. NumPy
 - B. pandas
 - C. NLTK
 - D. Matplotlib
- 5. What is the primary advantage of using LSTM for financial sentiment analysis?
 - A. Lower memory requirement
 - B. Handles long-term dependencies in text
 - C. Provides real-time analysis
 - D. Uses fewer data points

Set 3

HOTS:

1. Discuss how an NLP-powered chatbot could transform customer service in financial institutions. What ethical considerations should be addressed?
2. Propose a pipeline using spaCy for preprocessing financial news data, including tokenization, stop-word removal, and entity recognition. How would you evaluate the effectiveness of each step?

MCQs:

1. In a trading strategy, what is one role of NLP-based sentiment analysis?
 - A. Predicting transaction fees
 - B. Determining asset liquidity
 - C. Informing buy/sell decisions
 - D. Managing account balances
2. Lemmatization differs from stemming by:
 - A. Generating unique non-existent words
 - B. Returning root forms that are actual words
 - C. Ignoring suffixes and prefixes
 - D. Applying only to English text
3. In sentiment analysis, what does TextBlob primarily use to determine sentiment polarity?
 - A. Deep learning models
 - B. Naive Bayes classifiers
 - C. Regular expressions
 - D. Rule-based systems
4. Which model is commonly used for topic modeling in unsupervised sentiment analysis?
 - A. Naive Bayes
 - B. Latent Dirichlet Allocation (LDA)
 - C. LSTM

D. Decision Tree

5. In a chatbot application, NLP is used primarily to:
- A. Display web content
 - B. Generate predefined responses
 - C. Interpret and respond to user queries
 - D. Edit database records

Set 4

HOTS:

1. Analyze the challenges in using traditional sentiment lexicons for financial sentiment analysis. How could these lexicons be adapted for financial contexts?
2. Develop an argument for or against the use of LSTM-based sentiment analysis in real-time trading systems. Consider computational complexity and potential impacts on trading decisions.

MCQs:

1. In a trading strategy, what is one role of NLP-based sentiment analysis?
 - A. Predicting transaction fees
 - B. Determining asset liquidity
 - C. Informing buy/sell decisions
 - D. Managing account balances
2. Lemmatization differs from stemming by:
 - A. Generating unique non-existent words
 - B. Returning root forms that are actual words
 - C. Ignoring suffixes and prefixes
 - D. Applying only to English text
3. In sentiment analysis, what does TextBlob primarily use to determine sentiment polarity?
 - A. Deep learning models
 - B. Naive Bayes classifiers
 - C. Regular expressions
 - D. Rule-based systems
4. Which model is commonly used for topic modeling in unsupervised sentiment analysis?
 - A. Naive Bayes
 - B. Latent Dirichlet Allocation (LDA)
 - C. LSTM
 - D. Decision Tree
5. In a chatbot application, NLP is used primarily to:
 - A. Display web content
 - B. Generate predefined responses
 - C. Interpret and respond to user queries
 - D. Edit database records

Set 5

HOTS:

1. Construct a use case for NLP in assessing market volatility through news sentiment. What features would be essential for accurate predictions?
2. Evaluate the benefits and limitations of using TextBlob for initial sentiment analysis in financial contexts. What alternative models could provide more accurate sentiment scores?

MCQs:

1. The main limitation of bag-of-words models in NLP is:
 - A. High computational cost
 - B. Loss of word context and order
 - C. Incompatibility with Python
 - D. Complexity of implementation
2. Why is a deep learning model like LSTM preferred for certain NLP tasks?
 - A. Less training data is required
 - B. It captures sequential dependencies in text
 - C. It has lower computational demands
 - D. It is suitable for rule-based tasks
3. For which purpose would Latent Dirichlet Allocation (LDA) be most suitable?
 - A. Sentiment analysis
 - B. Text classification
 - C. Topic modeling
 - D. Syntax parsing
4. Which sentiment analysis method is most suitable for analyzing stock market conversations?
 - A. TextBlob
 - B. Financial Lexicon-based method
 - C. Basic Naive Bayes
 - D. Movie-based sentiment analysis
5. Which is a characteristic of unsupervised learning in NLP?
 - A. Requires labeled data
 - B. Uses known categories for training
 - C. Discovers patterns without labeled output
 - D. Utilizes reinforcement techniques

Set 6

HOTS:

1. Explain the impact of NLP in automating low-value tasks in financial institutions. How does it improve efficiency and objectivity?
2. Discuss the potential improvements that a financial sentiment analysis model could gain from using BERT or other advanced pretrained models.

MCQs:

1. Which algorithm is frequently used for initial sentiment classification in NLP?
 - A. Decision trees
 - B. K-Nearest Neighbors
 - C. Naive Bayes
 - D. Reinforcement learning

2. What is an advantage of using pre-trained models for NLP tasks?
 - A. Reduced training time and computational cost
 - B. Better grammar correction
 - C. Enhanced visualization of text data
 - D. Improved encryption of data
3. The CountVectorizer is commonly used for which type of NLP model?
 - A. Bag-of-words
 - B. Word embedding
 - C. PoS tagging
 - D. Named Entity Recognition
4. A rule-based chatbot:
 - A. Learns from previous interactions
 - B. Uses a fixed set of responses
 - C. Generates new phrases
 - D. Trains itself over time
5. Dependency parsing is primarily used for:
 - A. Tokenizing text
 - B. Removing stop words
 - C. Calculating sentiment scores
 - D. Understanding relationships between words in a sentence

Set 7

HOTS:

1. Analyze how topic modeling can help uncover hidden themes in financial documents. What challenges might arise in applying topic modeling in this context?
2. Discuss the significance of supervised and unsupervised models in NLP for analyzing large volumes of news data. How does each type of model contribute to insights?

MCQs:

1. Which of the following can help in representing the sentiment score of words within a specific context?
 - A. CountVectorizer
 - B. Word embedding
 - C. Rule-based system
 - D. Syntax parsing
2. A self-learning chatbot can improve through:
 - A. Rule-based responses
 - B. Using predefined answers
 - C. Supervised training on labeled data
 - D. Error correction
3. Which step typically follows tokenization in NLP preprocessing?
 - A. Stop word removal
 - B. Data encryption
 - C. Syntax parsing
 - D. Entity extraction
4. Inference in NLP can best be described as:

- A. Tokenizing data
 - B. Splitting text into paragraphs
 - C. Calculating word frequencies
 - D. Generating decisions from processed text
5. What does the TF-IDF model achieve in text processing?
- A. Counts total words in a text
 - B. Orders sentences based on structure
 - C. Highlights words that are important within a document but not common across documents
 - D. Encrypts textual data