

25/03 600 FINANCIAL DATA

Group Discussion

Course Overview ▾

M1: Fixed Income Data ▾

M2: Equities and Cryptocurrencies ▾

M3: Working with Portfolios and Tick Data ▾

M4: Alternative Data ▴

FD Forum M4

LESSON 1: SIMILARITY MEASURES

☐ Required Readings☐ Lesson Notes

LESSON 2: SEARCH ENGINE DATA

☐ Required Readings☐ Lesson Notes

LESSON 3: APPLICATIONS OF ALTERNATIVE DATA

☐ Required Readings☐ Lesson Notes

LESSON 4: SOCIAL MEDIA

☐ Lesson Notes

ASSESSMENTS

☐ FD Collaborative Review Task 2☐ FD Practice Quiz M4☐ FD Graded Quiz M4

M5: News Data and Sentiment Analysis ▾

[Home](#) > [My Courses](#) > [Financial Data](#) > [M4: Alternative Data](#) > FD Graded Quiz M4

FD Graded Quiz M4

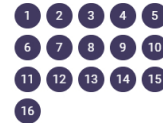
Question 1

According to Taborda et al (SA-MAIS), what is the primary limitation of using only domain-specific dictionaries for sentiment analysis?

- ☐ They are too expensive to implement
- ☒ They are created based on specific and relatively small datasets
- ☐ They only work with positive sentiments
- ☐ They cannot process financial terms

Time left:
0:18

QUESTIONS



Question 2

According to "Chapter 6: Vector Semantics and Embeddings", how would you apply the skip-gram model to capture the semantic relationship between 'doctor' and 'hospital'?

- ☐ Calculate their direct co-occurrence count only
- ☒ Train the model on contexts where these words co-occur within the window size
- ☐ Measure their document-level co-occurrence only
- ☐ Count the total frequency of both words in the corpus

Question 3

What would be a robust approach to cryptocurrency sentiment analysis using Reddit data?

- ☐ Counting total comments
- ☐ Analyzing only price correlations
- ☐ Using only standard sentiment dictionaries
- ☒ Combining sentiment scores with market-specific terminology analysis

Question 4

What is the primary difference between CountVectorizer and TfidfVectorizer in text analysis?

- ☐ CountVectorizer is faster but less accurate
- ☒ TfidfVectorizer considers term importance across corpus while CountVectorizer only counts term frequency
- ☐ CountVectorizer works with numbers while TfidfVectorizer works with text
- ☐ TfidfVectorizer can only work with English text

Question 5

Given a document with 100 terms where 'trading' appears 5 times and appears in 3 out of 10 total documents, what is its TF-IDF score? (Using natural log)

- ☒ 0.0602
- ☐ 0.0924
- ☐ 0.0729
- ☐ 0.0453

Question 6

Given vectors $X = [1 \ 1 \ 1]$ and $Y = [2 \ 2 \ 2]$, calculate their Minkowski distance with $p=3$.

- ☒ 1.44
- ☐ 1.91
- ☐ 1.73
- ☐ 2.08

Question 7

What are the key limitations of StockTwits sentiment analysis?

- ☐ Sentiment analysis is not available for all stocks
- ☐ The platform only allows positive sentiment expressions
- ☐ The system automatically determines sentiment
- ☒ Sentiment tags reflect subjective opinions that may include personal biases

Question 8

Which similarity measure would be most appropriate for comparing company descriptions in regulatory filings?

- ☐ Minkowski Distance
- ☐ Manhattan Distance
- ☐ Euclidean Distance
- ☒ Cosine Similarity with TF-IDF

Question 9

What distinguishes Cambridge Dictionary's definition of social media from other definitions presented?

- ☐ Its focus on user interactions
- ☐ Its emphasis on content sharing
- ☒ Its emphasis on technological platforms over social aspects
- ☐ Its description of community building

Question 10

How would you best design a system to monitor emerging market risks using Google Trends data?

- ☐ Track only the most searched financial terms
- ☒ Integrate rising queries with risk-related terms and validate with market indicators
- ☐ Focus solely on regional search patterns
- ☐ Monitor only negative sentiment keywords

Question 11

If two terms have LSA coordinates of (0.5, 0.3) and (0.4, 0.2) respectively, what is their cosine similarity? (use natural log)

- ☐ 0.677
- ☐ 0.582
- ☐ 0.897
- ☒ 0.997

Question 12

For two stocks with daily returns [0.02 -0.01 0.03] and [0.03 0.01 0.02], calculate their cosine similarity.

- ☒ 0.79
- ☐ 0.97

☐ 0.86

☐ 0.65

Question 13

According to Taborda et al (SA-MAIS), why is domain-specific sentiment analysis considered more challenging than general sentiment analysis for financial tweets?

☐ Financial tweets are longer than regular tweets

☐ Financial tweets are written in multiple languages

☒ Financial tweets contain technical jargon and polysemic terms with context-dependent meanings

☐ Financial tweets contain more emojis

Question 14

How does Named Entity Recognition enhance TF-IDF analysis in financial contexts?

☐ It improves calculation speed

☐ It reduces preprocessing requirements

☒ It identifies specific companies and relationships

☐ It simplifies document normalization

Question 15

Based on your understanding of social media analytics and market behavior, what characteristic of social media data provides the most comprehensive insight into market sentiment?

☐ The number of active users

☐ The volume of daily posts

☒ The multi-dimensional nature of user interactions that combine explicit opinions with implicit behavioral patterns

☐ The frequency of updates

Question 16

Which approach would be most effective for developing a new trading strategy using Google Trends data?

☐ Use only historical search trends without market data

☒ Integrate rising queries analysis with traditional technical indicators

☐ Focus exclusively on regional search patterns

☐ Design a strategy based solely on highest search volume terms

SUBMIT

