Q20: Which ones are metrics for IR evaluation?

Chọn một hoặc nhiều hơn:

☑ A. F1

☑ B. Precision

☑ C. Recall

☐ D. F2

Q26: Which ones are applications of Entertainment Graph?

Chọn một hoặc nhiều hơn:

☑ A. Discover key factors in determining whether a movie will be nominated for awards

☑ B. Distinguish popular movies from poorly ranked movies

☐ C. Discover Miss Universe

☑ D. Predict of upcoming movie popularity

Q23: Which kind of data is graph?

Chọn một hoặc nhiều hơn:

☑ A. Economic networks

☑ B. Internet

☐ C. Image

☑ D. Social networks

Q19: What does tf.idf stand for?

Chọn một:

◉ A. Term Frequency – Inverse Document Frequency

○ B. Term Frequency – Inverse Document Format

○ C. Temporary Frequency – Inverse Dual Frequency

○ D. Template Frequency – Interface Dual Format

Q22: Which ones are text classification metrics:

Chọn một hoặc nhiều hơn:

☑ A. F1 measure

☐ B. Visualization

☑ C. Accuracy

☑ D. Confusion matrix

Q5: What is the mode of the following dataset: {6,5,4,5,7,6,6,5,6}?

Chọn một:

◉ A. 6

○ B. 7

○ C. 4

○ D. 5

Q32: The temperature on Tuesday was 37, on Wednesday was 36, on Thursday as 32 degrees. What is the temperature on Friday if we use moving average forecasting with k=2?

Chọn một:

○ A. 34.3333

◉ B. 34

○ C. 35.3333

○ D. 35

Q13: Which ones are Examples of outliers?

Chọn một hoặc nhiều hơn:

☑ A. Drops

☐ B. Trip plan

☑ C. Spikes

☑ D. Trend changes

Q15: Which ones are techniques of Trajectory Exploration?

Chọn một hoặc nhiều hơn:

☑ A. Cleaning

☑ B. Map-Matching

☑ C. Segmentation

☐ D. Level shifts

Q4: What is the median of the following dataset: {6,5,4,5,7,6,6,5,6}?

Chọn một:

◉ A. 6

○ B. 5

○ C. 4

○ D. 7

Q25: Which ones are applications of Cyber Security Graph?

Chọn một hoặc nhiều hơn:

☑ A. Predict computers without proper authorization

☐ B. Face detection

☑ C. Identify intruder machines

☑ D. Provide knowledge of computer viruses propagation

Q29: How many steps are there in the data science pipeline? Provide the details of each step.

Chọn một:

○ A. 3 steps - Get the data, Analyze the data, Visualize the result

◉ B. 4 steps - Ask an interesting question, Get the data, Analyze the data, Visualize the result

○ C. 2 steps - Analyze the data, Visualize the result

○ D. 3 steps - Ask an interesting question, Get the data, Analyze the data

Q2: Which ones are application of Data Analytics?

Chọn một hoặc nhiều hơn:

☑ A. Flu Monitoring

☑ B. Election Predictions

☐ C. Front-end Design

☑ D. Product Recommendation

Q28: Which ones are applications of Community Analysis?

Chọn một hoặc nhiều hơn:

☑ A. Communities for Digital Marketing

☑ B. Communities for Recommender Systems

☑ C. Communities for User Behavior Analytics

☐ D. Centrality

Q16: What does POI stand for?

Chọn một:

○ A. People of integrity

○ B. People of Interest

◉ C. Points of Interest

○ D. Points of integrity

Q10: Which ones are Time Series Applications?

Chọn một hoặc nhiều hơn:

☑ A. Economic and Sales Forecasting

☑ B. Moving Object's Analysis

☑ C. Budget and Stock Market Analysis

☐ D. Exchange Bitcoin

Q7: What is the standard deviation of the following dataset: {6,5,4,5,7,6,6,5,6}?

Chọn một:

○ A. 0.31

○ B. 0.19

◉ C. 0.83

○ D. 0.69

Q18: Which ones are actions of text normalization?

Chọn một hoặc nhiều hơn:

☑ A. Remove stops words

☐ B. Find duplicate words

☑ C. Remove special character and symbols

☑ D. Stemming

Q35: Which ones are types of Recommender Systems?

Chọn một hoặc nhiều hơn:

☐ A. Centrality

☑ B. Rating-bases recommendation

☑ C. Content-bases recommendation

☐ D. Link Prediction

Q30: What are 4Vs in big data?

Chọn một:

○ A. Volumn, Velocity, Vulnerability, Vehicle

◉ B. Volumn, Velocity, Variety, Veracity

○ C. Volumn, Velocity, Variety, Vulnerability

○ D. Vast, Vehicle, Volumn, Variety

Q1: Which pipeline is correct?

Chọn một:

○ A. Get the data, Analyze the Data, Communicate and visualize the results, Ask an interesting question.

◉ B. Ask an interesting question, Get the data, Analyze the Data, Communicate and visualize the results.

○ C. Ask an interesting question, Analyze the Data, Get the data, Communicate and visualize the results.

○ D. Get the data, Analyze the Data, Ask an interesting question, Communicate and visualize the results

Question: What happens if a term t appears in all documents?
Ans: idf = 0 -> that term is just redundant like stop-word

Q33: The last period's forecast was 70 and demand was 60. What is the simple exponential smoothing forecast with alpha of 0.4 for the next period?

Chọn một:

◉ A. 66

○ B. 64

○ C. 65

○ D. 26

Q8: Which ones are analysis techniques on Time Series data?

Chọn một hoặc nhiều hơn:

☑ A. Anomaly Detection

☑ B. Segmentation

☑ C. Exploration

☐ D. Point-of-interest Analysis

Question: What is the mean, median, mode, variance, standard deviation, 25-th percentile of the following dataset: 7,9,3,4,2,8,5,6,13,1.
Ans:
Mean = 5.8.
Median = 5.5.
There is no mode because there are no data values that occur most frequently.
Variance = 11.76.
Standard deviation = 3.429.
25-th percentile = 3.

Q6: What is the variance of the following dataset: {6,5,4,5,7,6,6,5,6}?

Chọn một:
- ⦿ A. 0.69
- ◯ B. 0.19
- ◯ C. 0.31
- ◯ D. 2.07

Question: Provide an application that you think the big data should be applied? Please explain why you make that choice.
Ans:
Big data can be applied to the education field. We can develop a system that can track the students' activities on e-learning such as BKeL. The system can know how many times students access a page on BKeL and can have an evaluation of the overall performance of students.
Reason: Because there are still lots of students fail to pass the course every semester

Q27: Which ones are Graph Analysis Techniques?

Chọn một hoặc nhiều hơn:
- ☑ A. Centrality
- ☑ B. Node Classification
- ☑ C. Link Prediction
- ☐ D. Text alignment

Q3: What is the mean of the following dataset: {6,5,4,5,7,6,6,5,6} ?

Chọn một:
- ◯ A. 4.56
- ◯ B. 7
- ⦿ C. 5.56
- ◯ D. 6

Q24: Which ones are applications of Social Network Graphs?

Chọn một hoặc nhiều hơn:
- ☐ A. Identify intruder machines
- ☑ B. Identify the most influential people
- ☑ C. Conduct marketing campaigns
- ☑ D. Recommend friends

Q17: Which ones are Trajectory Anomaly Detection?

Chọn một hoặc nhiều hơn:
- ☑ A. Disasters
- ☐ B. Human interaction
- ☑ C. Traffic accidents
- ☑ D. Taxi drivers' malicious detour

Question: Explain the differences between content-based recommendation and collaborative filtering recommendation. Give the example for each type.

Ans:
Content-based uses user profile and item profile.
Example: If the customer wants to watch a movie, the system will base on the user profile to find what type of films the customer likes and base on the film description to find all films that match the customer.
Collaborative filtering uses user ratings.
Example: If the customer wants to watch a movie, the system will base on the customer's past ratings on watched film to find the movie that has a similar type to the film having the highest rating.

Q11: Choose the right answer:

Chọn một:

○ A. Trends: Persistent, overall upward and downward pattern

○ B. Random variation: Persistent and unsystematic.

○ C. Random variation: Persistent and residual fluctuations

○ D. Seasonality: Regular pattern of up and down fluctuations and erratic

Q12: Which ones are techniques of Forecasting?

Chọn một hoặc nhiều hơn:

☑ A. Moving Average

☑ B. Exponential Smoothing

☐ C. Persistent

☑ D. Seasonality

Q9: Which ones are Operations on Trajectory data?

Chọn một hoặc nhiều hơn:

☑ A. Exploratory Analysis

☐ B. Forecasting

☑ C. Activity Analysis

☑ D. Anomaly Detection

Q21: Which ones are text classification applications?

Chọn một hoặc nhiều hơn:

☑ A. Gmail tabbed categories

☑ B. Spam email detection/filtering

☐ C. Identify the most influential people

☑ D. Tagging content or products using categories

Question: What happens if term t is not in the corpus?
Ans: $|C_t| = 0$ -> we cannot compute idf

Q34: Which ones are recommender Systems – Challenges?

Chọn một hoặc nhiều hơn:

☑ A. Attacks

☑ B. Cold-Start Problem

☐ C. Defenses

☑ D. Data Sparsity

Q31: Explain the differences between content-based recommendation and collaborative filtering recommendation. Give the example for each type.

Chọn một:

○ A. content-based uses user rating, collaborative filtering filters user profile

○ B. content-based users user profile, collaborative filtering uses user ratings

◉ C. content-based uses user profile and item profile, collaborative filtering uses user ratings

○ D. content-based uses user rating, collaborative filtering uses user ratings

Q14: Which ones are Trajectory data applications?

Chọn một hoặc nhiều hơn:

☑ A. Routes recommendation

☐ B. Process and Quality Control

☑ C. Human interaction

☑ D. City traffic planning