**Question 1 (4 points)**

Sentiment analysis will always give us either a positive or a negative indication of customer opinion, but nothing in between.

Question 1 options:

|  |  |
| --- | --- |
|  | True |
|  | False |

**Question 2 (4 points)**

What analytical step helps us understand WHY a customer needs to have the information they are requesting?

Question 2 options:

|  |  |
| --- | --- |
|  | creating framed analytical questions |
|  | discovering the information need of a customer |
|  | discovering the context of the problem |
|  | discovering the necessary data to solve the problem |

**Question 3 (4 points)**

\_\_\_\_\_\_\_\_\_\_\_\_ is a technique to compare the frequency of words among documents in a corpus. (Select the best answer)

Question 3 options:

|  |  |
| --- | --- |
|  | Sentiment analysis |
|  | Keyword analysis |
|  | Similarity scoring |
|  | N-gram analysis |

**Question 4 (4 points)**

Which of the following are NOT techniques typically employed in text data mining:

Select 2 correct answer(s)

Question 4 options:

|  |  |
| --- | --- |
|  | hierarchical clustering |
|  | logistic regression |
|  | coding qualitative data |
|  | name entity recognition |
|  | topic recognition |
|  | sentiment analysis |

**Question 5 (4 points)**

An excellent indicator of why the customer is asking for information is: (Select the best answer)

Question 5 options:

|  |  |
| --- | --- |
|  | important leading indicators |
|  | a gap in a KPI |
|  | a business need |
|  | a gap in the business context |

**Question 6 (4 points)**

The definitions of data variables in a database can be found in \_\_\_\_\_\_\_\_\_\_\_\_

Question 6 options:

|  |  |
| --- | --- |
|  | the data report |
|  | the data dictionary |
|  | metadata |
|  | the SQL tables |

**Question 7 (4 points)**

A date is considered a qualitative type of data variable.

Question 7 options:

|  |  |
| --- | --- |
|  | True |
|  | False |

**Question 8 (4 points)**

When using coding to analyze text a good rule of thumb for the number of codes that need to be generated is:

Question 8 options:

|  |  |
| --- | --- |
|  | 8-10 |
|  | 10-15 |
|  | 15-20 |
|  | 4-5 |
|  | 6-8 |

**Question 9 (4 points)**

Which of the following is NOT an example of a visual programming environment.

Select 1 correct answer(s)

Question 9 options:

|  |  |
| --- | --- |
|  | IBM SPSS |
|  | SAS enterprise guide |
|  | Rapidminer |
|  | Orange3 |
|  | Excel |

**Question 10 (4 points)**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ is a technique used to locate and classify name found in a text into pre-defined categories.

Question 10 options:

|  |  |
| --- | --- |
|  | topic modeling |
|  | keyword extraction |
|  | sentiment analysis. |
|  | Named entity extraction |

**Question 11 (4 points)**

What character is used as a delimiter between data fields in a CSV encoded file?

Question 11 options:

|  |  |
| --- | --- |
|  | a space |
|  | a comma |
|  | a tab |
|  | a colon |

**Question 12 (4 points)**

Which text data mining technique maybe used to compare a company's webpage to a competitors to try to improve its ranking in the search engines. (Select all that applies):

Select 2 correct answer(s)

Question 12 options:

|  |  |
| --- | --- |
|  | keyword analysis |
|  | term frequency analysis |
|  | name entity recognition |
|  | sentiment analysis |

**Question 13 (4 points)**

Which binary character encoding technique is preferred for text data mining:

Question 13 options:

|  |  |
| --- | --- |
|  | Unicode UTF-16 |
|  | Unicode UTF-8 |
|  | Baudot code |
|  | ASCII-7 code |

**Question 14 (4 points)**

In keyword analysis the number of words we consider important is determined by:

Question 14 options:

|  |  |
| --- | --- |
|  | reaction |
|  | deduction |
|  | introspection |
|  | inspection |

**Question 15 (4 points)**

The most critical step in a data mining project is:

Question 15 options:

|  |  |
| --- | --- |
|  | discovering the information in need of the customer |
|  | cleaning the data |
|  | selecting the proper analytical algorithm |
|  | framing the right analytical questions |

**Question 16 (4 points)**

Which of the following are required characteristics of a flat file? (Select all that applies as absolutely necessary)

Select 4 correct answer(s)

Question 16 options:

|  |  |
| --- | --- |
|  | there should be no empty rows in the middle of the table |
|  | the first row always has the variable names |
|  | there should be nothing to the right of the last column |
|  | there should be no cell formatting or coloring |
|  | there should be no empty columns in the middle of the table |
|  | there are no empty variable names |
|  | there should be nothing below the last row |

**Question 17 (4 points)**

Most text documents are characterized by what kinds of information (select all that apply):

Question 17 options:

|  |  |
| --- | --- |
|  | metadata |
|  | indexing |
|  | topics |
|  | formatting information |
|  | contextual content |

**Question 18 (4 points)**

In qualitative data analysis a code is most often a \_\_\_\_\_\_\_\_\_\_ that symbolically assigns summative salient essence capturing evocative attributes to a portion of language-based data. (Select all that applies):

Select 3 correct answer(s)

Question 18 options:

|  |  |
| --- | --- |
|  | number |
|  | word |
|  | keyword |
|  | short phrase |

**Question 19 (4 points)**

Which of the following techniques maybe used to visualize the context of the text data (select all that applies):

Select 3 correct answer(s)

Question 19 options:

|  |  |
| --- | --- |
|  | Terms Berry plots |
|  | word trees |
|  | pie charts |
|  | bar plots |
|  | Word clouds |

**Question 20 (4 points)**

Which of the following are techniques NOT typically employed in text data mining:

Select 2 correct answer(s)

Question 20 options:

|  |  |
| --- | --- |
|  | term frequency analysis |
|  | sentiment analysis |
|  | coding qualitative data |
|  | name entity recognition |
|  | topic recognition |
|  | hierarchical clustering |
|  | logistic regression |

**Question 21 (4 points)**

The following are popular applications of sentiment analysis: (select all that applies):

Select 3 correct answer(s)

Question 21 options:

|  |  |
| --- | --- |
|  | social media monitoring |
|  | inventory control |
|  | voice of the customer |
|  | hypothesis testing |
|  | customer feedback |

**Question 22 (4 points)**

\_\_\_\_\_\_\_\_\_\_\_\_ is a technique used in text similarity scoring.

Question 22 options:

|  |  |
| --- | --- |
|  | TF-IDF |
|  | keyword extraction. |
|  | topic modeling |
|  | NER |

**Question 23 (4 points)**

In text data mining a set of static texts that are the input data to our model is called a: (Select the best answer)

Question 23 options:

|  |  |
| --- | --- |
|  | bag of words |
|  | text data file |
|  | flat file |
|  | corpus |

**Question 24 (4 points)**

A \_\_\_\_\_\_\_\_\_ is a graphical technique used to gain insight on what customers are saying from a text field of a customer survey.

Question 24 options:

|  |  |
| --- | --- |
|  | Word cloud |
|  | box plot |
|  | Bar plot |
|  | pie chart |

**Question 25 (4 points)**

Which algorithm is most often used in topic modeling tools:

Question 25 options:

|  |  |
| --- | --- |
|  | latent semantic analysis |
|  | sentiment analysis |
|  | text similarity scoring |
|  | logistic regression |