1. Pernyataan yang benar mengenai data mining berikut ini adalah :   
a. Data mining merupakan proses iterative dan interaktif untuk menemukan pola atau model yang sahih, baru, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive databases)   
b. Data mining is the nontrivial extraction of implicit, previously unknown and potentially useful information from data   
c. Data mining is the search for relationships and global patterns that exist in small databases but are ‘hidden’ among the vast amount of data, such as a relationship between patient data and their medical diagnosis   
**d. Jawaban a dan b benar**    
  
2. Manakah yang benar mengenai teknik searching :   
a. Searching adalah salah satu teknik database yang dilakukan untuk memeriksa serangkaian item yang memiliki sifat-sifat yang inginkan   
b. Tindakan untuk menemukan satu item tertentu baik yang diketahui keberadaannya maupun tidak   
c. Memasukkan kata dalam suatu program komputer untuk memperbaiki informasi yang ada dalam database   
**d. Semua jawaban benar**   
  
3. Kapankah waktu yang tepat untuk melakukan data reduction :   
a. pada saat grouping dan summing data-data yang masih dalam jumlah besar   
**b. pada saat data preprocessing pada rangkaian proses KDD (Knowledge Discovery Databases)**   
c. Pada saat memasukkan data dalam suatu program komputer   
d. Pada saat indexing   
  
4. OLAP dapat digunakan untuk membuat rangkuman dari …. yang berbeda, rangkuman baru dan mendapatkan respon secara online.   
a. Triple dimensi data   
b. Satu dimensi data   
**c. Multidimensi data**   
d. Query data   
  
5. Berikut ini adalah karakteristik-karakteristik penting dari tool data mining, kecuali :   
a. data preparation facilities   
b. selection of data mining operation (algorithms)   
c. product scalability and performance   
**d. facilities for implementation of result**   
  
6. Salah satu karakteristik data warehouse adalah time-variant, yaitu :   
a. Data tidak dapat diperbaharui atau diubah tetapi hanya dapat di tambah dan dilihat   
b. Data yang disusun menurut subjek berisi hanya informasi yang penting bagi pemrosesan decision support   
**c. Penyimpanan data selama 5-10 tahun atau lebih, data dapat digunakan untuk perbandingan atau perkiraan**   
d. Data terletak pada berbagai aplikasi yang terpisah   
  
7. Pada tahap data preprocessing terdapat beberapa alat dan metode yang berbeda yang digunakan untuk preprocessing antara lain sampling yaitu :   
**a. Menyeleksi subset representatif dari populasi data yang besar**   
b. Menghilangkan noise dari data   
c. Memanipulasi data mentah untuk menghasilkan input tunggal   
d. Semua jawaban benar   
  
8. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah di mengerti oleh pihak yang berkepentingan. Hal tersebut dilakukan pada tahap

**a. Interpretation**    
b. evaluation   
c. jawaban a dan b benar   
d. jawaban a dan b salah   
  
9. Proses enrichment yaitu proses ‘memperkaya’ data yang sudah ada dengan :   
a. data / informasi lain yang agak relevan   
**b. data / informasi lain yang diperlukan untuk KDD seperti data atau informasi eksternal**   
c. jawaban a dan b benar   
d. jawaban a dan b salah

 10. Kebutuhan akan data mining dikarenakan :   
a. ketersediaan data yang melimpah, kebutuhan akan informasi sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis   
b. informasi sebagai asset perusahaan yang penting   
c. ketersediaan teknologi informasi dalam skala yang terjangkau   
**d. semua jawaban benar**   
  
11. Pattern evaluation adalah suatu proses :   
**a. Untuk menemukan informasi yang bernilai melalui knowledge base**   
b. Untuk membuang data yang tidak konsisten dan noise   
c. Untuk mentransformasikan data menjadi bentuk yang sesuai untuk di mining   
d. Tidak ada jawaban   
  
12. Berikut ini yang termasuk penugasan pengklasifikasian adalah :   
a. memprediksi tingkat keganasan suatu sel tumor (ganas atau tidak ganas)   
b. pengkategorian jenis berita (berita keuangan, cuaca, hiburan, dll)   
**c. jawaban a dan b benar**   
d. jawaban a dan b salah   
  
13. Pada metode decision tree, ada berapa cara untuk menspesifikasikan kondisi uji menurut tipe atribut?   
a. ada 3, yaitu nominal, ordinal dan kontine   
b. ada 2, yaitu pembagian 2 cara (2-way split) dan banyak cara (multi-way split)   
**c. jawaban a dan b benar**   
d. jawaban a dan b salah   
  
14. Menggunakan banyak partisi dengan nilai yang berbeda adalah salah satu teknik pembagian (splitting) yang disebut dengan :   
a. binary split   
**b. multi-way split**   
c. nominal split   
d. ordinal split   
  
15. Discretization dilakukan untuk :   
a. mempertimbangkan semua kemungkinan pembagian dan mencari pemotongan terbaik   
b. melakukan perhitungan yang lebih intensif   
**c. membentuk atribut kategorikal ordinal**    
d. menemukan interval dan frekuensi yang sama   
  
16. Dalam Konsep Data Mining terdapat 3 kebutuhan bisnis, Kecuali :   
a. Penambahan maupun peningkatan kapasitas produk   
**b. Penambahan biaya operasi perusahaan**   
c. Pengurangan biaya operasi perusahaan   
d. Peningkatan efektifitas pemasaran dan keuntungan   
  
17. Perbedaan antara basis data relasional dengan OLAP :   
a. identifikasi atribut  
**b. pada basis data relasional data dimasukkan ke dalam tabel, sementara pada OLAP digunakan representasi array multidimensional**

c. pada basis data relasional digunakan representasi array multidimensional sedangkan pada OLAP data dimasukkan ke dalam tabel  
d. jawaban a dan b benar   
  
18. Di dalam Data Mining terdapat permasalahan bisnis yang umum dihadapi, Kecuali :   
a. Bagaimana menyajikan advertensi kepada target yang tepat sasaran   
b. Mengklasifikasikan artikel-artikel secara otomatis   
c. Mengelompokkan pengunjung web yang memiliki kesamaan karateristik tertentu   
**d. Memprediksi kelakuan di masa yang lalu**   
  
19. Berikut ini yang termasuk 2 langkah untuk mengkonversi data tabular ke dalam   
array multidimensi :   
a. identifikasi atribut-atribut mana yang merupakan atribut dimensi dan atribut mana yang merupakan atribut target/tujuan  
b. Mencari nilai dari setiap entri dalam array multidimensi dengan menjumlahkan nilai-nilai (dari atribut tujuan)  
**c. Jawaban a dan b benar**  
d. Jawaban a dan b salah   
  
20. Konsep dasar yang perlu dipikirkan pelaku bisnis sebagai solusi permasalahan, yaitu :   
a. Perumusan Target, Personalisasi, Asosiasi ( Analisis Keranjang Pasar ), Manajemen Pengetahuan.   
b. Pengelompokkan, Estimasi & Prediksi, Pohon Keputusan   
c. Jawaban a dan b salah   
**d. Jawaban a dan b benar**   
  
21. Pernyataan yang benar mengenai data mining berikut ini adalah :   
a. Data mining merupakan proses iterative dan interaktif untuk menemukan pola atau model yang sahih, baru, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive databases)   
b. Data mining is the nontrivial extraction of implicit, previously unknown and potentially useful information from data   
c. Data mining is the search for relationships and global patterns that exist in small databases but are ‘hidden’ among the vast amount of data, such as a relationship between patient data and their medical diagnosis   
**d. Jawaban a dan b benar**

22. Manakah yang benar mengenai teknik searching :   
a. Searching adalah salah satu teknik database yang dilakukan untuk memeriksa serangkaian item yang memiliki sifat-sifat yang inginkan   
b. Tindakan untuk menemukan satu item tertentu baik yang diketahui keberadaannya maupun tidak   
c. Memasukkan kata dalam suatu program komputer untuk memperbaiki informasi yang ada dalam database   
**d. Semua jawaban benar**   
  
23. Kapankah waktu yang tepat untuk melakukan data reduction :   
a. pada saat grouping dan summing data-data yang masih dalam jumlah besar   
**b. pada saat data preprocessing pada rangkaian proses KDD (Knowledge Discovery catabases)**   
c. Pada saat memasukkan data dalam suatu program komputer   
d. Pada saat indexing   
  
24. OLAP dapat digunakan untuk membuat rangkuman dari …. yang berbeda, rangkuman baru dan mendapatkan respon secara online.   
a. Triple dimensi data   
b. Satu dimensi data   
**c. Multidimensi data**   
d. Query data   
  
25. Berikut ini adalah karakteristik-karakteristik penting dari tool data mining, kecuali :   
a. data preparation facilities   
b. selection of data mining operation (algorithms)   
c. product scalability and performance   
**d. facilities for implementation of result**   
  
26. Salah satu karakteristik data warehouse adalah time-variant, yaitu :   
a. Data tidak dapat diperbaharui atau diubah tetapi hanya dapat di tambah dan dilihat   
b. Data yang disusun menurut subjek berisi hanya informasi yang penting bagi pemrosesan decision support   
**c. Penyimpanan data selama 5-10 tahun atau lebih, data dapat digunakan untuk perbandingan atau perkiraan**   
d. Data terletak pada berbagai aplikasi yang terpisah   
  
27. Pada tahap data preprocessing terdapat beberapa alat dan metode yang berbeda yang digunakan untuk preprocessing antara lain sampling yaitu :   
**a. Menyeleksi subset representatif dari populasi data yang besa**r   
b. Menghilangkan noise dari data   
c. Memanipulasi data mentah untuk menghasilkan input tunggal   
d. Semua jawaban benar   
  
28. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah di mengerti oleh pihak yang berkepentingan. Hal tersebut dilakukan pada tahap a. Interpretation   
b. evaluation   
**c. jawaban a dan b benar**   
d. jawaban a dan b salah   
  
29. Proses enrichment yaitu proses ‘memperkaya’ data yang sudah ada dengan :   
a. data / informasi lain yang agak relevan   
**b. data / informasi lain yang diperlukan untuk KDD seperti data atau informasi eksternal**   
c. jawaban a dan b benar   
d. jawaban a dan b salah

30. Kebutuhan akan data mining dikarenakan :   
a. ketersediaan data yang melimpah, kebutuhan akan informasi sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis   
b. informasi sebagai asset perusahaan yang penting   
c. ketersediaan teknologi informasi dalam skala yang terjangkau   
**d. semua jawaban benar**   
  
31. Pattern evaluation adalah suatu proses :   
**a. Untuk menemukan informasi yang bernilai melalui knowledge base**   
b. Untuk membuang data yang tidak konsisten dan noise   
c. Untuk mentransformasikan data menjadi bentuk yang sesuai untuk di mining   
d. Tidak ada jawaban   
  
32. Berikut ini yang termasuk penugasan pengklasifikasian adalah :   
a. memprediksi tingkat keganasan suatu sel tumor (ganas atau tidak ganas)   
b. pengkategorian jenis berita (berita keuangan, cuaca, hiburan, dll)   
**c. jawaban a dan b benar**   
d. jawaban a dan b salah   
  
33. Pada metode decision tree, ada berapa cara untuk menspesifikasikan kondisi uji menurut tipe atribut ?   
**a. ada 3, yaitu nominal, ordinal dan kontine**   
b. ada 2, yaitu pembagian 2 cara (2-way split) dan banyak cara (multi-way split)   
c. jawaban a dan b benar   
d. jawaban a dan b salah   
  
34. Menggunakan banyak partisi dengan nilai yang berbeda adalah salah satu teknik pembagian (splitting) yang disebut dengan :   
a. binary split   
**b. multi-way split**   
c. nominal split   
d. ordinal split

35. Discretization dilakukan untuk :   
a. mempertimbangkan semua kemungkinan pembagian dan mencari pemotongan terbaik   
b. melakukan perhitungan yang lebih intensif   
**c. membentuk atribut kategorikal ordinal**   
d. menemukan interval dan frekuensi yang sama   
  
36. Dalam Konsep Data Mining terdapat 3 kebutuhan bisnis, Kecuali :   
a. Penambahan maupun peningkatan kapasitas produk   
**b. Penambahan biaya operasi perusahaan**   
c. Pengurangan biaya operasi perusahaan   
d. Peningkatan efektifitas pemasaran dan keuntungan   
  
37. Perbedaan antara basis data relasional dengan OLAP :   
a. identifikasi atribut  
**b. pada basis data relasional data dimasukkan ke dalam tabel, sementara pada OLAP digunakan representasi array multidimensional**  
c. pada basis data relasional digunakan representasi array multidimensional sedangkan pada OLAP data dimasukkan ke dalam tabel  
d. jawaban a dan b benar

38. Di dalam Data Mining terdapat permasalahan bisnis yang umum dihadapi, Kecuali :   
a. Bagaimana menyajikan advertensi kepada target yang tepat sasaran   
b. Mengklasifikasikan artikel-artikel secara otomatis   
c. Mengelompokkan pengunjung web yang memiliki kesamaan karateristik tertentu   
**d. Memprediksi kelakuan di masa yang lalu**   
  
39. Berikut ini yang termasuk 2 langkah untuk mengkonversi data tabular ke dalam   
array multidimensi :   
a. identifikasi atribut-atribut mana yang merupakan atribut dimensi dan atribut mana yang merupakan atribut target/tujuan  
b. Mencari nilai dari setiap entri dalam array multidimensi dengan menjumlahkan nilai-nilai (dari atribut tujuan)  
**c. Jawaban a dan b benar**  
d. Jawaban a dan b salah   
  
40. Konsep dasar yang perlu dipikirkan pelaku bisnis sebagai solusi permasalahan, yaitu :   
a. Perumusan Target, Personalisasi, Asosiasi ( Analisis Keranjang Pasar ), Manajemen   
Pengetahuan.   
b. Pengelompokkan, Estimasi & Prediksi, Pohon Keputusan   
c. Jawaban a dan b salah   
**d. Jawaban a dan b benar**   
  
41. Operasi kunci dari OLAP adalah :   
**a. Formasi dari kubus data**  
b. Korespondensi dari kubus data  
c. Tabel data  
d. Elemen data   
  
42. Mengapa melakukan data mining dari sudut pandang Komersial :   
a. Kecepatan data yang dihimpun dan disimpan (Gbyte/hour)   
b. Data mining untuk reduksi data   
**c. Meledaknya volume data yang dihimpun dan disimpan dalam data warehouse**   
d. Teknik-teknik tradisional tidak fisibel utnuk mengolah data mentah   
  
43. Mengapa melakukan data mining dari sudut pandang keilmuan :   
a. Kecepatan data yang dihimpun dan disimpan (Gbyte/hour)   
b. Data mining untuk reduksi data   
c. Meledaknya volume data yang dihimpun dan disimpan dalam data warehouse   
**d.. Jawaban a dan b benar**   
  
44. Dari sudut pandang komersial dilakuan data mining, kecuali :   
**a. Teknik-teknik tradisional tidak fisibel utnuk mengolah data mentah**  
b. Proses komputasi yang dapat diupayakan   
c. Kuatnya tekanan kompetitif   
d. Meledaknya volume data yang dihimpun dan disimpan dalam data warehouse   
  
45. Sebutkan bidang ilmu yang berkaitan dengan data mining :   
a. Fisika   
b. Kimia

**c. Pemodelan Matematika**

d. Biologi   
  
46. Beberapa solusi dapat diselesaikan dengan data mining dalam penerapannya dalam Analisa Perusahaan dan Manajemen Resiko, Kecuali :   
a. Merencanakan Sumber Daya (Resource Planning)   
**b. Menebak target pasar**   
c. Memonitor Persaingan (Competition)   
d. Merencanakan Keuangan dan Evaluasi Aset   
  
47. Data Mining Engine adalah Suatu proses :   
a. Untuk menemukan informasi yang bernilai melalui knowledge base   
b. Untuk membuang data yang tidak konsisten dan noise   
**c. Untuk mentransformasikan data menjadi bentuk yang sesuai untuk di mining**   
d. Penggabungan data dari beberapa sumber   
  
48. Data Cleaning adalah Suatu Proses :   
a. Untuk menemukan informasi yang bernilai melalui knowledge base   
**b. Untuk membuang data yang tidak konsisten dan noise**   
c. Untuk mentransformasikan data menjadi bentuk yang sesuai untuk di mining   
d. Penggabungan data dari beberapa sumber   
  
49. Digunakan untuk memetakan data dengan prediksi atribut bernilai real, disebut dengan :   
**a. Regresi**   
b. Klasifikasi

c. Prediksi

d. Analisis Time Series   
  
50. Proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk dapat memprediksi kelas dari suatu objek yang labelnya tidak diketahui, disebut dengan :   
a. Regresi  
b. Prediksi

**c. Klasifikasi**

d. Analisis Time Series   
  
51. Model data relasional, Implementasi DBMS relasional adalah ciri-ciri evolusi database pada tahun :   
a. 1960   
**b. 1970**

c. 1980

d. 1990   
  
52. Data mining, data warehouse, database multimedia, dan Web database adalah ciri-ciri evolusi database pada tahun :   
a. 1960   
b. 1970

c. 1980

**d. 1990**   
  
53. Ciri-ciri evolusi database pada tahun 2000 adalah sebagai berikut, Kecuali :   
a. Teknologi web dan sistem informasi global   
b. Data mining dengan berbagai variasi aplikasi   
c. Stream data managemen dan mining   
**d. Pengumpulan data, pembuatan data, IMS dan network DBMS**   
  
54. suatu alat untuk analisa data, yang memecahkan permasalahan penggolongan adalah   
a. Sequence Discovery   
**b. Clustering**   
c. Assosiation Rule   
d. Summarization   
  
55. Salah Satu Teknik Data Mining adalah   
a. Data Warehouse   
**b. Visualisasi Data**   
c. Graphical User Interface   
d. Pattern evaluation   
  
56. Yang bukan termasuk Tipe-Tipe Data Pada Mining adalah   
a. Variabel - Variabel berskala interval   
b. Variabel biner   
**c. Variabel Oktal**   
d. Variabel nominal, ordinal dan rasio   
  
57. Manakah yang termasuk Kategori Pendekatan Clustering adalah   
a. Data Cleaning   
b. Data Integration

c. Data Mining Engine

**d. Algoritma Partisi**   
  
58. Salah satu karakteristik data warehouse adalah Non volatile, yaitu :   
**a. Data tidak dapat diperbaharui atau diubah tetapi hanya dapat di tambah dan dilihat**   
b. Data yang disusun menurut subjek berisi hanya informasi yang penting bagi pemrosesan decision support   
c. Penyimpanan data selama 5-10 tahun atau lebih, data dapat digunakan untuk perbandingan atau perkiraan   
d. Data terletak pada berbagai aplikasi yang terpisah   
  
59. Terdapat beberapa alat dan metode yang berbeda yang digunakan untuk preprocessing seperti Transformation, Yaitu :   
a. Menyeleksi subset representatif dari populasi data yang besar   
b. Menghilangkan noise dari data   
**c. Memanipulasi data mentah untuk menghasilkan input tunggal**   
d. Mengorganisasi data untuk pengaksesan yang lebih spesifik   
  
60. Terdapat beberapa alat dan metode yang berbeda yang digunakan untuk preprocessing seperti Normalization, Yaitu :   
a. Menyeleksi subset representatif dari populasi data yang besar   
b. Menghilangkan noise dari data   
c. Memanipulasi data mentah untuk menghasilkan input tunggal   
**d. Mengorganisasi data untuk pengaksesan yang lebih spesifik**   
  
61. Perbedaan antara basis data relasional dengan OLAP :   
a. identifikasi atribut  
**b. pada basis data relasional data dimasukkan ke dalam tabel, sementara pada OLAP digunakan representasi array multidimensional**  
c. pada basis data relasional digunakan representasi array multidimensional sedangkan pada OLAP data dimasukkan ke dalam tabel  
d. jawaban a dan b benar   
  
62. Berikut ini yang termasuk 2 langkah untuk mengkonversi data tabular ke dalam array multidimensi :   
a. identifikasi atribut-atribut mana yang merupakan atribut dimensi dan atribut mana yang merupakan atribut target/tujuan  
b. Mencari nilai dari setiap entri dalam array multidimensi dengan menjumlahkan nilai-nilai (dari atribut tujuan)  
**c. Jawaban a dan b benar**  
d. Jawaban a dan b salah   
  
63. Operasi kunci dari OLAP adalah :   
**a. Formasi dari kubus data**  
b. Korespondensi dari kubus data  
c. Tabel data  
d. Elemen data   
  
64. Seleksi sekumpulan sel dari seluruh array multidimensi dengan menetapkan nilai spesifik untuk satu atau lebih dimensi :   
a. Dicing  
b. Roll-up  
c. Drill-down  
**d. Slicing**   
  
65. Salah satu contoh proses roll-up yaitu :   
**a. pada data penjualan, kita dapat mengumpulkan data penjualan dari seluruh tanggal (harian) dalam satu bulan**  
b. membagi total data penjualan bulanan ke dalam total penjualan harian  
c. jawaban a dan benar  
d. jawaban a dan b salah   
  
66. Berikut ini yang termasuk karakteristik Rule-Based Classifier, kecuali :   
a. aturan mutually exclusive : pengklasifikasi (classifier) mengandung aturan mutually exclusive jika aturan-aturan independen satu sama lain dan setiap record paling banyak dilingkupi oleh satu aturan  
b. aturan exhaustive : classifier memiliki lingkupan exhaustive jika mencatat untuk setiap kemungkinan kombinasi nilai atribut  
c. pada aturan exhaustive, setiap record paling sedikit dilingkupi oleh satu aturan  
**d. aturan mutually exclusive : pengklasifikasi (classifier) mengandung aturan mutually exclusive jika aturan-aturan dependen satu sama lain dan setiap record paling banyak dilingkupi oleh satu aturan**   
  
67. Suatu aturan dapat disederhanakan namun hal ini tentunya akan menimbulkan efek yaitu   
a. aturan tidak lagi bersifat mutually exclusive sehingga satu record dapat memicu lebih dari satu aturan  
b. aturan tidak lagi bersifat exhaustive sehingga satu record dapat tidak memicu satupun aturan  
**c. jawaban a dan b benar**  
d. jawaban a dan b salah   
  
68. Ada 2 (dua) skema pengurutan aturan yaitu :   
**a. Rule-based ordering dan Class-based ordering**  
b. Direct ordering dan indirect ordering  
c. Learn-one-rule dan learn-two-rule  
d. Semua jawaban benar   
  
69. Berikut ini adalah aspek-aspek dari sequential covering, kecuali :   
a. pertumbuhan aturan (rule growing), instance elimination  
b. Evaluasi aturan (rule evaluation), criteria berhenti (Stopping criterion)  
**c. Rule deleting dan rule pruning**  
d. Extraction   
  
70. Mengapa kita menghilangkan positive instances?   
a. untuk memaksimumkan pengukuran entropy  
**b. untuk memastikan bahwa aturan selanjutnya akan berbeda**  
c. untuk mencegah underestimating keakurasian aturan  
d. semua jawaban benar   
  
71. Mengapa kita menghilangkan negative instances?   
a. untuk memaksimumkan pengukuran entropy   
b. untuk memastikan bahwa aturan selanjutnya akan berbeda   
**c. untuk mencegah underestimating keakurasian aturan**\*   
d.semua jawaban benar   
  
72. Pernyataan yang benar mengenai stopping criterion :   
**a. menghitung hasil yang didapat (gain)**  
b. jika gain signifikan, maka jangan perhatikan aturan yang bar  
c. serupa dengan pasca pemangkasan (post pruning) pada pohon keputusan  
d. membandingkan tingkat error pada himpunan validasi   
  
73. Pernyataan yang benar mengenai data mining berikut ini adalah :   
a. Data mining merupakan proses iterative dan interaktif untuk menemukan pola atau model yang sahih, baru, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive databases)   
b. Data mining is the nontrivial extraction of implicit, previously unknown and potentially useful information from data   
c. Data mining is the search for relationships and global patterns that exist in small databases but are ‘hidden’ among the vast amount of data, such as a relationship between patient data and their medical diagnosis   
**d. Jawaban a dan b benar**   
  
74. Manakah yang benar mengenai teknik searching :   
**a. Searching adalah salah satu teknik database yang dilakukan untuk memeriksa serangkaian item yang memiliki sifat-sifat yang inginkan**   
b. Tindakan untuk menemukan satu item tertentu baik yang diketahui keberadaannya maupun tidak   
c. Memasukkan kata dalam suatu program komputer untuk memperbaiki informasi yang ada dalam database   
d. Semua jawaban benar   
  
74. Kapankah waktu yang tepat untuk melakukan data reduction :   
a. pada saat grouping dan summing data-data yang masih dalam jumlah besar   
**b. pada saat data preprocessing pada rangkaian proses KDD (Knowledge Discovery Databases)**   
c. Pada saat memasukkan data dalam suatu program komputer   
d. Pada saat indexing   
  
76. OLAP dapat digunakan untuk membuat rangkuman dari …. yang berbeda, rangkuman baru dan mendapatkan respon secara online.   
a. Triple dimensi data   
b. Satu dimensi data   
**c. Multidimensi data**   
d. Query data   
  
77. Berikut ini adalah karakteristik-karakteristik penting dari tool data mining, kecuali :   
a. data preparation facilities   
b. selection of data mining operation (algorithms)   
c. product scalability and performance   
**d. facilities for implementation of result**   
  
78. Salah satu karakteristik data warehouse adalah time-variant, yaitu :   
a. Data tidak dapat diperbaharui atau diubah tetapi hanya dapat di tambah dan dilihat   
b. Data yang disusun menurut subjek berisi hanya informasi yang penting bagi pemrosesan decision support   
**c. Penyimpanan data selama 5-10 tahun atau lebih, data dapat digunakan untuk perbandingan atau perkiraan**   
d. Data terletak pada berbagai aplikasi yang terpisah   
  
79. Pada tahap data preprocessing terdapat beberapa alat dan metode yang berbeda yang digunakan untuk preprocessing antara lain sampling yaitu :   
**a. Menyeleksi subset representatif dari populasi data yang besar**   
b. Menghilangkan noise dari data   
c. Memanipulasi data mentah untuk menghasilkan input tunggal   
d. Semua jawaban benar   
  
80. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah di mengerti oleh pihak yang berkepentingan. Hal tersebut dilakukan pada tahap

**a. Interpretation**   
b. evaluation   
c. jawaban a dan b benar   
d. jawaban a dan b salah   
  
81. Proses enrichment yaitu proses ‘memperkaya’ data yang sudah ada dengan :   
a. data / informasi lain yang agak relevan   
**b. data / informasi lain yang diperlukan untuk KDD seperti data atau informasi eksternal**   
c. jawaban a dan b benar   
d. jawaban a dan b salah   
  
82. Kebutuhan akan data mining dikarenakan :   
a. ketersediaan data yang melimpah, kebutuhan akan informasi sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis   
b. informasi sebagai asset perusahaan yang penting   
c. ketersediaan teknologi informasi dalam skala yang terjangkau   
**d. semua jawaban benar**   
  
82. Pattern evaluation adalah suatu proses :   
a. Untuk menemukan informasi yang bernilai melalui knowledge base   
b. Untuk membuang data yang tidak konsisten dan noise   
**c. Untuk mentransformasikan data menjadi bentuk yang sesuai untuk di mining**   
d. Tidak ada jawaban   
  
84. Berikut ini yang termasuk penugasan pengklasifikasian adalah :   
a. memprediksi tingkat keganasan suatu sel tumor (ganas atau tidak ganas)   
b. pengkategorian jenis berita (berita keuangan, cuaca, hiburan, dll)   
**c. jawaban a dan b benar**   
d. jawaban a dan b salah   
  
85. Pada metode decision tree, ada berapa cara untuk menspesifikasikan kondisi uji menurut tipe atribut?   
a. ada 3, yaitu nominal, ordinal dan kontine   
b. ada 2, yaitu pembagian 2 cara (2-way split) dan banyak cara (multi-way split)   
**c. jawaban a dan b benar**   
d. jawaban a dan b salah   
  
86. Menggunakan banyak partisi dengan nilai yang berbeda adalah salah satu teknik pembagian (splitting) yang disebut dengan :   
a. binary split   
**b. multi-way split**   
c. nominal split   
d. ordinal split   
  
87. Discretization dilakukan untuk :

a. mempertimbangkan semua kemungkinan pembagian dan mencari pemotongan terbaik

b. melakukan perhitungan yang lebih intensif

**c. membentuk atribut kategorikal ordinal**   
d. menemukan interval dan frekuensi yang sama