Zagadnienia do egzaminu dyplomowego na studiach II stopnia

ANALIZA DANYCH – BIG DATA

1. Techniki uczenia nadzorowanego i bez nadzoru.
2. Selekcja zmiennych i obserwacji do modelu.
3. Metody klasyfikacji danych.
4. Modele drzew decyzyjnych.
5. Modele bazujące na drzewach decyzyjnych.
6. Sztuczne sieci neuronowe.
7. Metody grupowania danych.
8. Modele nieparametryczne analizy czasu trwania.
9. Modele parametryczne analizy czasu trwania.
10. Modele semiparametryczne analizy czasu trwania.
11. Podejście bayesowskie w analizie czasu trwania.
12. Procedura budowy modeli predykcyjnych.
13. Przykładowe modele uczenia maszynowego wykorzystywane w zagadnieniach regresji oraz zagadnieniach klasyfikacji.
14. Jakość danych.
15. Imputacja danych.
16. Imputacja wielokrotna.
17. Analizy wzdłużne.
18. Regresja kwantylowa.
19. Regresja adaptacyjna.
20. Analiza skupień.
21. Modele służące do oceny wartości klienta w czasie.
22. Proces analizy danych w kontekście wizualizacji.
23. Dobór metod wizualizacji danych.
24. Silnik decyzyjny w procesach automatyzacji oceny kredytowej.
25. Wybór i przygotowanie danych do budowy modeli scoringowych.
26. Karta scoringowa.
27. Analiza opłacalności procesu akceptacji wniosków kredytowych.
28. Problem estymacji ryzyka dla klientów odrzuconych.
29. Rozwiązania big data w obszarze repozytoriów danych.
30. Specyfika środowisk analitycznych stosowanych w big data.

Zagadnienia do egzaminu dyplomowego na studiach II stopnia

ANALIZA DANYCH – BIG DATA

31. Wybrane algorytmy stosowane w analityce big data.
32. Algorytm MapReduce.
33. Typowe wyzwania związane z danymi big data.
34. Przetwarzanie rozproszone.
35. Kwestie etyczne związane z big data.
36. 3V i 5V w kontekście big data.
37. Dane ustrukturyzowane i dane nieustrukturyzowane w kontekście big data.
38. Porównanie architektury Lambda i architektury Kappa.
39. Pojęcie i zastosowania biznesowe hurtowni danych.
40. Różnice pomiędzy wsadowym i strumieniowym sposobem przetwarzania danych.
41. Biznesowe zastosowania analizy danych w czasie rzeczywistym.
42. Techniki iteracji.
43. Koncepcja funkcji w programowaniu: typowe elementy i cele.
44. Wykonanie programu poprzez interpretację kodu źródłowego.
45. Zarządzanie bibliotekami (pakietami) w środowisku języka Python.
46. Triada CIA (poufność, integralność, dostępność) w bezpieczeństwie informacji.
47. Atak w cyberprzestrzeni typu spear phishing.
48. Podstawowe rekomendacje dla MŚP w zakresie cyberbezpieczeństwa.
49. Sposoby wykonywania złączeń tabel w relacyjnej bazie danych.
50. Klasyfikacja funkcji języka SQL działających na pojedynczych wierszach.
51. Powody tworzenia perspektyw w relacyjnej bazie danych. Klauzule specyficzne dla polecenia tworzącego perspektywy.
52. Operacje na zbiorach - składnia poleceń i znaczenie uzyskanych wyników.
53. Typy problemów, które rozwiązują podzapytania. Klauzule występowania podzapytań, operatory.