|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | |  | |
| Test końcowy | | | *PS-06-01-Z1/03*  Ważne od: 2021.07.02 | |
| ***Big Data - podstawy analizy danych opartej o duże zbiory danych***  ***Data: 12-13.05.2025***  ***Wykładowca: Marcin Albiniak***  ***TES\_250512\_2551500PYT\_140408KC\_*** | | | | |
| ***Dane uczestnika testu*** | ***Wyniki – punktacja*** | |
| ***Imię:*** Kliknij lub naciśnij tutaj, aby wprowadzić tekst. | ***Maksymalna: 11 pkt.*** | |
| ***Nazwisko:*** Kliknij lub naciśnij tutaj, aby wprowadzić tekst. | ***Uzyskana:*** Kliknij lub naciśnij tutaj, aby wprowadzić tekst.***pkt.*** | |
| ***Data wypełnienia testu:*** Kliknij lub naciśnij, aby wprowadzić datę. |  | |
| ***Miejscowość:*** Kliknij lub naciśnij tutaj, aby wprowadzić tekst. |  | |

**1) Środowiska do programowania w języku Python to :**

Datalore

PyCharm

Intellij

Visual Studio Code

Liczba punktów: ***\_\_\_\_\_***

**2) Która z poniższych odpowiedzi najlepiej opisuje pojęcie „Big Data” i jego znaczenie?:**

Big Data to zbiór niewielkich plików tekstowych przechowywanych lokalnie, wykorzystywanych do backupu danych osobowych.

Big Data to technika szyfrowania dużych baz danych w czasie rzeczywistym.

Big Data odnosi się do bardzo dużych, zróżnicowanych i szybko zmieniających się zbiorów danych, których analiza pozwala na uzyskanie wartościowych informacji wspierających decyzje biznesowe.

Big Data to specjalistyczny typ bazy danych SQL stworzony przez firmę Google do obsługi serwisów społecznościowych.

Liczba punktów: ***\_\_\_\_\_***

**3) Numpy jest pakietem, który zwiera w sobie:**

deklarację struktur n-wymiarowych

funkcje matematyczne: podstawowe i złożone

wykresy złożone

ramki

Liczba punktów: ***\_\_\_\_\_***

**4) Podstawowymi strukturami w Pandas są:**

DataFrame

Tuple

Series

Set

Liczba punktów: ***\_\_\_\_\_***

**5)** **zadania jakie można wykonać za pomocą matplotlib:**

regresja

klasyfikacja

rozwiązywanie układów nieliniowych n-równań

kreślenie wykresów

Liczba punktów: ***\_\_\_\_\_***

**6) Jak należy przygotować dane do analizy:**

znormalizować dane

podzielić na cząstkowe zbiory:

przygotować jeden duży zbiór danych

dane przekształcić na wartości logiczne: true lub false

Liczba punktów: ***\_\_\_\_\_***

**7) Które z poniższych stwierdzeń są prawdziwe w kontekście porównania przetwarzania wsadowego i strumieniowego?**

Przetwarzanie strumieniowe umożliwia analizę danych w czasie rzeczywistym, podczas gdy przetwarzanie wsadowe analizuje dane z pewnym opóźnieniem.

Przetwarzanie wsadowe wymaga mniej zasobów obliczeniowych niż strumieniowe, ponieważ działa ciągle i nie wymaga buforowania danych.

Przetwarzanie wsadowe najlepiej sprawdza się w zadaniach, gdzie nie ma potrzeby natychmiastowej reakcji na dane, np. w raportach dziennych lub miesięcznych.

Przetwarzanie strumieniowe polega na przechowywaniu dużych plików danych i ich okresowej analizie w blokach.

Liczba punktów: \_\_\_\_\_

**8) Które z poniższych narzędzi i technik są powszechnie wykorzystywane w analizie dużych zbiorów danych w języku Python?:**

Pandas — biblioteka do manipulacji i analizy danych, szczególnie danych tabelarycznych.

Matplotlib — biblioteka do analizy danych w czasie rzeczywistym i przetwarzania równoległego.

Dask — narzędzie umożliwiające przetwarzanie dużych zbiorów danych z równoległością i podziałem danych na fragmenty.

SQLite — system bazodanowy wykorzystywany do dynamicznego trenowania modeli uczenia maszynowego w czasie rzeczywistym.

Liczba punktów: \_\_\_\_\_

**9) Które z poniższych stanowią typowe wyzwania związane ze skalowalnością i wydajnością systemów Big Data?:**

Trudność w zapewnieniu niskich opóźnień przy jednoczesnym przetwarzaniu ogromnych wolumenów danych.

Niska dostępność bibliotek open source wspierających przetwarzanie danych na dużą skalę.

Koszty związane z równoważeniem obciążenia i zarządzaniem infrastrukturą rozproszoną.

Ograniczenie liczby możliwych formatów danych do analizy w systemach Big Data.

Liczba punktów: \_\_\_\_\_

**10 )Polecenie import dotyczy:**

strings

modułów

rejestrów

bibliotek

Liczba punktów: \_\_\_\_\_

**11) Które z poniższych zagadnień są kluczowymi wyzwaniami w zakresie bezpieczeństwa i prywatności danych w systemach Big Data?:**

Trudność w zapewnieniu zgodności z regulacjami prawnymi, takimi jak RODO lub HIPAA.

Brak potrzeby stosowania szyfrowania danych ze względu na ich dużą objętość.

Ryzyko nieautoryzowanego dostępu do danych przechowywanych w rozproszonych środowiskach obliczeniowych.

Ograniczona możliwość przechowywania danych historycznych w systemach Big Data z powodów technicznych.

Liczba punktów: \_\_\_\_\_