Zał. nr 5 do ZW 8/2020

Załącznik nr … do programu studiów

|  |
| --- |
| WYDZIAŁ ……… / STUDIUM………………  **KARTA PRZEDMIOTU**   * 1. **Nazwa przedmiotu w języku polskim** Przetwarzanie danych masowych   2. **Nazwa przedmiotu w języku angielskim** Large scale data processing   3. **Kierunek studiów (jeśli dotyczy):** Sztuczna inteligencja   4. **Specjalność (jeśli dotyczy): ……………………..**   **Poziom i forma studiów: ~~I~~ / II stopień / ~~jednolite studia magisterskie\*~~, stacjonarna /**  **~~niestacjonarna~~\***  **Rodzaj przedmiotu: ~~obowiązkowy~~ / wybieralny / ~~ogólnouczelniany~~ \***  **Kod przedmiotu …………….**  **Grupa kursów ~~TAK~~ / NIE\*** |
|  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Wykład | Ćwiczenia | Laboratorium | Projekt | Seminarium |
| Liczba godzin zajęć zorganizowanych w Uczelni (ZZU) | 30 |  | 30 |  |  |
| Liczba godzin całkowitego nakładu pracy studenta (CNPS) | 60 |  | 60 |  |  |
| Forma zaliczenia | Egzamin / ~~zaliczenie na ocenę~~\* | Egzamin / zaliczenie na ocenę\* | ~~Egzamin~~ / zaliczenie na ocenę\* | Egzamin / zaliczenie na ocenę\* | Egzamin / zaliczenie na ocenę\* |
| Dla grupy kursów zaznaczyć kurs końcowy (X) |  |  |  |  |  |
| Liczba punktów ECTS | 2 |  | 2 |  |  |
| w tym liczba punktów odpowiadająca zajęciom  o charakterze praktycznym (P) |  |  | 2 |  |  |
| w tym liczba punktów ECTS odpowiadająca zajęciom wymagającym bezpośredniego udziału nauczycieli lub innych osób prowadzących zajęcia (BU) | 1,2 |  | 1,2 |  |  |

\*niepotrzebne skreślić

|  |
| --- |
| **WYMAGANIA WSTĘPNE W ZAKRESIE WIEDZY, UMIEJĘTNOŚCI I KOMPETENCJI SPOŁECZNYCH**   1. Umiejętność programowania w języku Python 2. Znajomość modeli uczenia maszynowego 3. Posiadanie podstawowych umiejętności z zakresu wykorzystania systemów z rodziny Unix 4. Znajomość oraz umiejętność wykorzystania paradygmatów programowania (obiektowy, funkcyjny) |

\

|  |
| --- |
| **CELE PRZEDMIOTU**  C1 Zapoznanie studentów z podstawowymi mechanizmami przetwarzania danych masowych  C2 Zapoznanie studentów z metodami wdrażania i utrzymywania aplikacji uczenia maszynowego w środowiskach produkcyjnych  C3 Przedstawienie najważniejszych narzędzi do przetwarzania danych masowych oraz zarządzania zasobami obliczeniowymi  C4 Nabycie umiejętności implementacji aplikacji do przetwarzania danych masowych oraz jej wdrożenia w ramach systemu produkcyjnego |

|  |
| --- |
| * + - 1. **PRZEDMIOTOWE EFEKTY UCZENIA SIĘ**   Z zakresu wiedzy:  KSI\_W08 - Zna różnorodne zastosowania technik analizy multimediów w zagadnieniach związanych z pozyskiwaniem informacji z danych masowych i złożonych  Z zakresu umiejętności:  KSI\_U03 - Potrafi sformułować problem do rozwiązania, zebrać i oczyścić dane, dobrać metodę oraz przeprowadzić eksperymenty a otrzymane wyniki poddać krytycznej analizie, dokonać ich interpretacji i prezentacji  KSI\_U04 - Potrafi dokonać oceny rozwiązania w zakresie pozyskania danych, ich przetwarzania oraz analizy a także ekstrakcji wiedzy a także zaproponować jego ulepszenie.  Z zakresu kompetencji społecznych:  KSI\_K01 - Jest gotów do krytycznej oceny odbieranych treści, ma świadomość znaczenia wiedzy w rozwiązywaniu problemów. |

|  |  |  |
| --- | --- | --- |
| **TREŚCI PROGRAMOWE** | | |
| * + 1. **Forma zajęć - wykład** | | * + - * 1. **Liczba godzin** |
| Wy1 | Wprowadzenie do problemu przetwarzania danych masowych. Zasady zaliczenia. | 2 |
| Wy2 | Taksonomia metod przetwarzania danych masowych | 2 |
| Wy3 | Podstawowe metody zrównoleglania algorytmów uczenia maszynowego. Przetwarzanie synchroniczne i asynchroniczne | 2 |
| Wy4 | Spark - przetwarzanie danych z wykorzystaniem paradygmatu Map-reduce - przetwarzanie wsadowe | 2 |
| Wy5 | Spark - przetwarzanie danych z wykorzystaniem paradygmatu Map-reduce - przetwarzanie strumieniowe | 2 |
| Wy6 | Flink - przetwarzanie danych w sposób strumieniowy | 2 |
| Wy7 | Flink - przetwarzanie danych w sposób wsadowy | 2 |
| Wy8 | Produkcyjne aspekty utrzymywania i wdrażania aplikacjami | 2 |
| Wy9 | Platformy zarządzania zasobami obliczeniowymi - wprowadzenie, OpenStack | 2 |
| Wy10 | Platformy zarządzania zasobami obliczeniowymi - Kubernetes | 2 |
| Wy11 | Metody automatyzacji zarządzania produkcyjnymi aplikacjami | 2 |
| Wy12 | Języki do przetwarzania danych masowych | 2 |
| Wy13 | Przykładowe metody z rodziny Gradient Boosting Machine | 2 |
| Wy14 | Zaawansowane metody zrównoleglania algorytmów uczenia maszynowego | 2 |
| Wy15 | Recap - podsumowanie wykładu | 2 |
|  | Suma godzin | 30 |

|  |  |  |
| --- | --- | --- |
| **Forma zajęć - ćwiczenia** | | **Liczba godzin** |
| Ćw1 |  |  |
| Ćw2 |  |  |
| Ćw3 |  |  |
| Ćw4 |  |  |
| .. |  |  |
|  | Suma godzin |  |

|  |  |  |
| --- | --- | --- |
| **Forma zajęć - laboratorium** | | **Liczba godzin** |
| La1 | Zapoznanie z programem laboratorium, sposobami oceny, szkolenie BHP. | 2 |
| La2 | Podstawy wykorzystania infrastruktury chmurowej - Unix, Shell | 2 |
| La3 | Podstawy wykorzystania kontenerów - Docker, containerd | 2 |
| La4 | Ewaluacja metod zrównoleglania zadań - Python, wątki i procesy | 2 |
| La5 | Wykorzystanie przetwarzania opartego o broker wiadomości | 2 |
| La6 | Wykorzystanie podejścia MapReduce w przetwarzaniu danych masowych - Spark | 2 |
| La7 | Wykorzystanie podejścia MapReduce w przetwarzaniu danych masowych - Flink | 2 |
| La8 | Implementacja rozproszonego mechanizmu pozyskiwania danych | 2 |
| La9 | Implementacja mechanizmu czyszczenia i ekstrakcji cech | 2 |
| La10 | Wybór i ewaluacja modeli uczenia maszynowego | 2 |
| La11 | Wykorzystanie rozwiązania Kubernetes w przetwarzaniu danych masowych - K3S | 2 |
| La12 | Uruchomienie opracowanej metody z wykorzystaniem środowiska Kubernetes | 2 |
| La13 | Udostępnienie opracowanej metody - API gRPC | 2 |
| La14 | Wykorzystanie opracowanej metody - API REST/GraphQL + SPA | 2 |
| La15 | Prezentacja całościowego rozwiązania | 2 |
|  | Suma godzin | 30 |

|  |  |  |
| --- | --- | --- |
| * + 1. **Forma zajęć - projekt** | | * + - * 1. **Liczba godzin** |
| Pr1 |  |  |
| Pr2 |  |  |
| Pr3 |  |  |
| Pr4 |  |  |
| … |  |  |
|  | Suma godzin |  |

|  |  |  |
| --- | --- | --- |
| **Forma zajęć - seminarium** | | **Liczba godzin** |
| Se1 |  |  |
| Se2 |  |  |
| Se3 |  |  |
| … |  |  |
|  | Suma godzin |  |

|  |
| --- |
| * + 1. **STOSOWANE NARZĘDZIA DYDAKTYCZNE** |
| N1. Wykład - prezentacja multimedialna  N2. Środowisko deweloperskie Python  N3. Biblioteki programistyczne dla języka Python |

**OCENA OSIĄGNIĘCIA PRZEDMIOTOWYCH EFEKTÓW UCZENIA SIĘ**

|  |  |  |
| --- | --- | --- |
| **Oceny** (F – formująca (w trakcie semestru), P – podsumowująca (na koniec semestru) | Numer efektu uczenia się | Sposób oceny osiągnięcia efektu uczenia się |
| F1 | KSI\_U03  KSI\_U04 | Kontrola zrealizowania zadań studentów. Każde zadanie ma ustaloną liczbę punktów. Suma punktów przeliczana na ocenę w następujący sposób:  <50%, 60%) 🡪 dst  <60%, 70%) 🡪 dst+  <70%, 80%) 🡪 db  <80%, 90%) 🡪 db+  <90%, 🡪 bdb |
| P - wykład | KSI\_W08 | Egzamin z pytaniami otwartymi i zamkniętymi. Do każdego pytania przypisana jest liczba punktów. Suma punktów jest przeliczana na końcową ocenę w następujący sposób:  <50%, 60%) 🡪 dst  <60%, 70%) 🡪 dst+  <70%, 80%) 🡪 db  <80%, 90%) 🡪 db+  <90%, 🡪 bdb |

|  |
| --- |
| **LITERATURA PODSTAWOWA I UZUPEŁNIAJĄCA** |
| **LITERATURA PODSTAWOWA:**   1. Leskovec, J., Rajaraman, A., & Ullman, J. D. Mining of massive datasets. Cambridge university press (2014) 2. VanderPlas, J.: Python Data Science Handbook: Essential Tools for Working with Data, O'Reilly Media (2016) 3. S.E. Ahmed (Ed.): Big and Complex Data Analysis, Methodologies and Applications, Springer (2017) 4. Wilhelm, A. F., Kestler, H. A. (Eds.): Analysis of Large and Complex Data. Springer (2016) 5. Provost, F., Fawcett, T.: Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, O'Reilly Media (2013) 6. Zaharia M., Wendell P., Konwinski A., Karau H., Learning Spark, (2015)   **LITERATURA UZUPEŁNIAJĄCA:**   1. L. Jimmy, D. Chris. Data-Intensive Text Processing with MapReduce (2010), http://lintool.github.com/MapReduceAlgorithms/index.html |
| **OPIEKUN PRZEDMIOTU (IMIĘ, NAZWISKO, ADRES E-MAIL)** |
| Tomasz Kajdanowicz, tomasz.kajdanowicz@pwr.edu.pl |