

PLAN STUDIÓW PODYPLOMOWYCH

Nazwa studiów podyplomowych: „Analiza i inżynieria danych – data science”

Wymiar kształcenia (sem.): dwa semestry

Liczba punktów ECTS konieczna do uzyskania kwalifikacji podyplomowych: 30

Lp.	Nazwa przedmiotu	Rodzaj i wymiar zajęć dydaktycznych			Forma zaliczenia przedmiotu/sposób weryfikacji efektów uczenia się	Punkty ECTS
		Rodzaj zajęć	Zajęcia teoretyczne (godz.)	Zajęcia praktyczne (godz.)		
Semestr I						
1	Wprowadzenie do narzędzi analitycznych	Ćwiczenia		35	zal. / aktywność na zajęciach mini-projekt	5
2	Eksploracja i wizualizacja danych	ćwiczenia		20	zal. oc. / aktywność na zajęciach mini-projekt	3
3	Bazy danych	ćwiczenia		30	zal. oc. / aktywność na zajęciach mini-projekt	5
4	Elementy data mining	ćwiczenia		20	zal. oc. / aktywność na zajęciach mini-projekt	3
Semestr II						
5	Zaawansowane programowanie w języku Python	ćwiczenia		30	zal. oc. / aktywność na zajęciach mini-projekt	4
6	Machine learning	ćwiczenia		40	zal. / aktywność na zajęciach mini-projekt	6
7	Inżynieria Big Data	ćwiczenia		35	zal. oc. / aktywność na zajęciach mini-projekt	4
Łączna liczba godzin		x		210	Łączna liczba punktów ECTS:	30
				210		

Okres zaliczeniowy na studiach podyplomowych: 1 rok

TREŚCI KSZTAŁCENIA

Nazwa studiów podyplomowych: „**Analiza i inżynieria danych – data science**”

Wymiar kształcenia (sem.): dwa semestry

CHARAKTERYSTYKA TREŚCI KSZTAŁCENIA

1. Nazwa przedmiotu: **Wprowadzenie do narzędzi analitycznych**

Cel kształcenia: zapoznanie słuchaczy z wybranym środowiskiem analitycznym, nabycie umiejętności uruchomienia środowiska Python, jak również zdobycie wiedzy na temat architektury oprogramowania pisanego z wykorzystaniem ww. języka.

Treści merytoryczne:

Instalacja i konfiguracja środowiska pracy.

Podstawowe elementy języka Python: organizacja kodu, podstawowe typy danych, instrukcje warunkowe, pętle.

Organizacja kodu: funkcje, moduły, pakiety oraz dokumentacja kodu.

Wprowadzenie do narzędzia Jupyter Notebook.

Podstawowe wykorzystanie pakietów pandas, matplotlib oraz seaborn w środowisku Jupyter Notebook.

Język znaczników Markdown.

Wykorzystanie systemu kontroli wersji Git.

Efekty uczenia się:

Wiedza (słuchacz zna i rozumie): podstawowe funkcje wybranego środowiska programistycznego; metody pracy na danych; zastosowania i funkcjonalność wybranego oprogramowania; budowę i funkcjonalność pakietu używanego do zdalnego nauczania.

Umiejętności (słuchacz potrafi): zainstalować i przygotować środowisko do pracy; napisać prosty program z wykorzystaniem języka Python; dobrać konstrukcje i struktury danych języka Python do realizacji zadanych operacji; korzystać z nowoczesnych rozwiązań technologicznych dostarczonych przez wybraną platformę na potrzeby zdalnego uczenia się.

Kompetencje społeczne (słuchacz jest gotów do): ciągłego doksztalcania się; klarownego omówienia zastosowanych rozwiązań i technologii.

2. Nazwa przedmiotu: **Eksploracja i wizualizacja danych**

Cel kształcenia: zapoznanie słuchaczy z efektywnymi metodami przetwarzania i eksploracyjnej analizy danych z wykorzystaniem zaawansowanych pakietów oraz zaznajomienie z metodami graficznej prezentacji danych.

Treści merytoryczne:

Wczytywanie danych z różnych źródeł.

Przetwarzanie zbiorów - zmiany formatu, brakujące wartości, przekształcanie itp.

Eksploracja danych - filtrowanie, sortowanie, agregacja (biblioteki numpy, pandas).

Wizualizacja danych - przegląd najpopularniejszych bibliotek (matplotlib, seaborn, plotly, bokeh, altair).

Efekty uczenia się:

Wiedza (słuchacz zna i rozumie): sposób wykorzystania środowiska programistycznego języka Python wraz z wybranymi bibliotekami w przygotowaniu, obróbce i przeprowadzeniu analiz danych; możliwości aplikacyjne przedstawionych metod analitycznych i wizualizacji danych; rozmaite techniki prezentacji danych; budowę i funkcjonalność pakietu używanego do zdalnego nauczania.

Umiejętności (słuchacz potrafi): pozyskiwać i przetwarzać dane za pomocą wybranego środowiska programistycznego; wczytać dane do programu, określić jakość danych, dokonać podstawowych manipulacji na danych; stosować wybrane metody przeprowadzania eksploracji danych; przygotować zestawienie danych w postaci tabelarycznej i graficznej, a następnie dokonać opisu uzyskanych wyników; korzystać z nowoczesnych rozwiązań technologicznych dostarczonych przez wybraną platformę na potrzeby zdalnego uczenia się.

Kompetencje społeczne (słuchacz jest gotów do): uznania ograniczenia własnej wiedzy dotyczącej analizy danych i rozumie potrzebę dalszego kształcenia; brania odpowiedzialności za przedstawioną interpretację wyników; komunikowania ważnych wyników i osiągnięć społeczeństwu; przestrzegania przepisów dotyczących ochrony powierzonych do analizy danych.

3. Nazwa przedmiotu: **Bazy danych**

Cel kształcenia: przedstawienie słuchaczom podstaw baz danych i języków zapytań, architektury systemów baz danych oraz metod projektowania baz danych.

Treści merytoryczne:

Relacyjne bazy danych - język SQL,
Programowanie baz danych PL/SQL

Efekty uczenia się:

Wiedza (słuchacz zna i rozumie): zasady projektowania baz danych; podstawowe własności języka zapytań SQL; zasady komunikacji języków programowania z serwerami baz danych; budowę i funkcjonalność pakietu używanego do zdalnego nauczania.

Umiejętności (słuchacz potrafi): projektować i implementować relacyjną bazę danych; budować oraz modyfikować konstrukcję zapytań do baz danych; importować dane zewnętrzne do bazy; korzystać z nowoczesnych rozwiązań technologicznych dostarczonych przez wybraną platformę na potrzeby zdalnego uczenia się.

Kompetencje społeczne (słuchacz jest gotów do): uznania ograniczenia własnej wiedzy dotyczącej baz danych i odczuwa potrzebę dalszego kształcenia; samodzielnego poszerzania umiejętności tworzenia i modyfikacji baz danych.

4. Nazwa przedmiotu: **Elementy data mining**

Cel kształcenia: zapoznanie słuchacza z metodami przeprowadzania eksploracji danych za pomocą poznanych metod data mining.

Treści merytoryczne:

Wprowadzenie do standardu CRISP-DM.

Podstawowe metody statystyczne:

- badanie rozkładu klas decyzyjnych,
- częstości wartości,
- miary rozproszenia, tendencji centralnej,
- obliczanie korelacji pomiędzy zmiennymi oraz wpływu atrybutów warunkowych na klasę decyzyjną (positive ratio).

Wybrane techniki pracy z danymi:

- standaryzacja,
- normalizacja,
- uzupełnianie uszkodzonych danych,
- konwersja wartości symbolicznych do numerycznych.

Analiza sygnałów i szeregów czasowych

Podstawowe metody regresji liniowej i nieliniowej oraz prognozowania szeregów czasowych

Przetwarzanie danych tekstowych: normalizacja i wektoryzacja

Zastosowanie języka Python do eksploracji, analizy i przetwarzania danych.

Efekty uczenia się:

Wiedza (słuchacz zna i rozumie): specyfikę poszczególnych metod i modeli data mining; istotę prezentowanych algorytmów; poszczególne etapy w procesie odkrywania wiedzy z danych; budowę i funkcjonalność pakietu używanego do zdalnego nauczania.

Umiejętności (słuchacz potrafi): analizować zadane bazy wiedzy przy pomocy wybranego oprogramowania; przygotować dane do wybranej metody data mining; konstruować model klasyfikujący dla zadanej bazy wiedzy; ocenić skuteczność budowanego modelu; korzystać z nowoczesnych rozwiązań technologicznych dostarczonych przez wybraną platformę na potrzeby zdalnego uczenia się.

Kompetencje społeczne (słuchacz jest gotów do): zachowywania ostrożności w wyciąganiu wniosków z eksperymentów, do momentu potwierdzenia tez na wielu danych i przy zastosowaniu metod walidacyjnych; oceny przydatności metod data mining w procesie badania zjawisk masowych;

5. Nazwa przedmiotu: **Zaawansowane programowanie w języku Python**

Cel kształcenia: pogłębienie dotychczas zdobytej wiedzy na temat programowania w języku Python oraz ukształtowanie praktycznych umiejętności programowania.

Treści merytoryczne:

Programowanie zorientowane obiektowo.

Moduły i pakiety.

Obsługa plików.

Dekoratory.

Wyrażenia lambda.

Usuwanie błędów, testowanie.

Wyrażenia regularne.

Efekty uczenia się:

Wiedza (słuchacz zna i rozumie): zaawansowane mechanizmy w języku Python; strukturę języka programowania; budowę i funkcjonalność pakietu używanego do zdalnego nauczania.

Umiejętności (słuchacz potrafi): używać zaawansowanych konstrukcji składniowych języka Python; modyfikować istniejące duże programy w Pythonie; samodzielnie rozwiązywać problemy na każdym etapie przygotowania i realizacji programów i projektów w języku Python; korzystać z nowoczesnych rozwiązań technologicznych dostarczonych przez wybraną platformę na potrzeby zdalnego uczenia się.

Kompetencje społeczne (słuchacz jest gotów do): tworzenia czytelnych i wydajnych programów; samodzielnego wyszukiwania informacji w literaturze; formułowania pytań, służących pogłębieniu własnej wiedzy dotyczącej wybranego języka programowania

6. Nazwa przedmiotu: **Machine learning**

Cel kształcenia: wprowadzenie wybranych algorytmów stosowanych w robotyce mobilnej oraz IoE (Internecie Wszechrzeczy), w tym: modelowanie mapy, lokalizacja na mapie, sterowanie serwomechanizmami, śledzenie obiektów, planowanie ruchu, wygładzanie ruchu.

Treści merytoryczne:

Uczenie nadzorowane i regresja liniowa.

Statystyki Bayesowskie.

Drzewa decyzyjne.

Uczenie nienadzorowane.

Sieci neuronowe.

Modele generatywne i autokodery.

Algorytmy i metody uczenia modeli głębokich.

Wybrane problemy klasyfikacji, detekcji, regresji.

Efekty uczenia się:

Wiedza (słuchacz zna i rozumie): podstawowe pojęcia z zakresu uczenia maszynowego; podstawowe metody klasyfikacji; budowę i funkcjonalność pakietu używanego do zdalnego nauczania.

Umiejętności (słuchacz potrafi): stosować podejście uczenia maszynowego lub sztucznej sieci neuronowej do praktycznego problemu; konstruować model klasyfikujący dla zadanej bazy wiedzy; ocenić skuteczność budowanego modelu; wyprowadzać wnioski na podstawie eksperymentów; przygotować harmonogram dobierania metod w zależności od ich skuteczności; weryfikować postawione tezy badawcze i demonstrować rozwiązania; korzystać z nowoczesnych rozwiązań technologicznych dostarczonych przez wybraną platformę na potrzeby zdalnego uczenia się.

Kompetencje społeczne (słuchacz jest gotów do): uznania ograniczenia własnej wiedzy, odczuwając potrzebę poszerzania swojej wiedzy i praktycznych umiejętności w zakresie metod uczenia maszynowego;

7. Nazwa przedmiotu: Inżynieria Big Data

Cel kształcenia: zapoznanie słuchaczy z dostępnymi rozwiązaniami technologicznymi i nowoczesnymi metodami przetwarzania danych.

Treści merytoryczne:

Integracja Python z bazami danych.

Wprowadzenie do Big Data.

Architektura i technologie Big Data.

Platforma Apache Hadoop.

Podstawy Apache Spark.

Batchowe i strumieniowe przetwarzanie danych.

Budowa przepływów danych z użyciem Apache Airflow.

Uczenie maszynowe w Big Data.

Efekty uczenia się:

Wiedza (słuchacz zna i rozumie): metody pozyskiwania, porządkowania i przechowywania dużych zbiorów danych; podstawowe metody i narzędzia do przetwarzania dużych zbiorów danych; budowę i funkcjonalność pakietu używanego do zdalnego nauczania.

Umiejętności (słuchacz potrafi): budować modele analizy danych w oparciu o różnorodne narzędzia; dobrać metodę analizy dużych zbiorów danych do wybranego problemu; dokonać analizy i prezentacji zgromadzonych danych i pozyskanych informacji; korzystać z nowoczesnych rozwiązań technologicznych dostarczonych przez wybraną platformę na potrzeby zdalnego uczenia się.

Kompetencje społeczne (słuchacz jest gotów do): uznania ograniczenia własnej wiedzy dotyczącej analizy danych i rozumie potrzebę dalszego kształcenia; wzięcia odpowiedzialności za przedstawioną interpretację wyników, posiadając przy tym świadomość wpływu stosowanych metod na ich precyzję; komunikacji ważnych wyników i osiągnięć społeczeństwu; przestrzegania przepisów dotyczących ochrony powierzonych do analizy danych.