

codebrainers

Data science: statystyka i narzędzia do obliczeń rozproszonych - kurs zaawansowany - Małopolski Pociąg do Kariery / Kierunek Rozwój / Graj po Zielone

Numer usługi 2024/11/05/118259/2397403

5 040,00 PLN brutto

5 040,00 PLN netto

90,00 PLN brutto/h

90,00 PLN netto/h

CODEBRAINERS

SPÓŁKA Z

OGRANICZONĄ

ODPOWIEDZIALNOŚĆ

CIA



📍 zdalna w czasie rzeczywistym

📄 Usługa szkoleniowa

🕒 56 h

📅 01.02.2025 do 11.03.2025

## Informacje podstawowe

**Kategoria**

Informatyka i telekomunikacja / Programowanie

**Sposób dofinansowania**wsparcie dla osób indywidualnych  
wsparcie dla pracodawców i ich pracowników**Grupa docelowa usługi**

**Kurs skierowany jest do osób interesujących się analizą danych na poziomie zaawansowanym oraz wykorzystaniem w praktyce narzędzi do obliczeń rozproszonych i analizy dużych zbiorów danych, takich jak Apache Spark i PySpark.**

Zaleca się, aby Uczestnicy posiadali wiedzę z zakresu programowania w języku Python i SQL.

Usługa adresowana również również dla Uczestników Projektu Kierunek – Rozwój, Małopolski Pociąg do Kariery, mBony+, Graj po Zielone (oraz innych programów w ramach 6.6 i 10.17 z terenu woj. Śląskiego), a także dla uczestników innych programów dofinansowań.

Kurs prowadzony jest z naciskiem na budowanie świadomości ekologicznej i wykorzystanie tych umiejętności w celu wspierania zrównoważonego rozwoju. Przyczynia się również do budowania zielonych miejsc pracy w sektorach tradycyjnych.

**Minimalna liczba uczestników**

8

**Maksymalna liczba uczestników**

16

**Data zakończenia rekrutacji**

31-01-2025

**Forma prowadzenia usługi**

zdalna w czasie rzeczywistym

# Cel

## Cel edukacyjny

Kurs uczy praktycznego zastosowania koncepcji statystycznych w celu przewidywania i interpretacji danych, a także wykorzystania narzędzi do obliczeń rozproszonych na dużych zbiorach danych.

Dzięki przekazywaniu również umiejętności ogólnych niezbędnych do pracy w sektorze zielonej gospodarki (zg. z GreenComp), kurs buduje także świadomość ekologiczną i przekazuje kompetencje do rozwoju ekologicznych rozwiązań technologicznych.

## Efekty uczenia się oraz kryteria weryfikacji ich osiągnięcia i Metody walidacji

Efekty uczenia się	Kryteria weryfikacji	Metoda walidacji
Wykorzystuje koncepcje statystyczne w celu przewidywania i interpretacji danych	omawia modele statystyczne oraz stawia własne hipotezy dotyczące analizowanych danych, w tym danych środowiskowych	Test teoretyczny
	analizuje i interpretuje dane z wykorzystaniem narzędzi Scipy, Statsmodel, Numpy i Pandas	Test teoretyczny
Posiada umiejętności ogólne niezbędne do pracy w sektorze zielonej gospodarki	analizuje dane z sektora zielonej gospodarki	Test teoretyczny
	charakteryzuje główne poglądy na temat zrównoważonego rozwoju, zasady środowiskowe 6R w kontekście TIK (rethink, refuse, reduce, reuse, recycle, recover) oraz ocenia wpływ na przyrodę rozwiązań z zakresu cyfrowej gospodarki	Test teoretyczny
Współpracuje w grupie z innymi członkami zespołu	posiada świadomość, iż każde działanie człowieka ma wpływ na środowisko	Test teoretyczny
	komunikuje się z innymi członkami zespołu za pośrednictwem narzędzi kontroli wersji (git)	Test teoretyczny
Analizuje duże zbiory danych z wykorzystaniem narzędzi do obliczeń rozproszonych	korzysta z narzędzi Apache Spark i PySpark w celu analizy dużych zbiorów danych, w tym danych środowiskowych	Test teoretyczny

# Kwalifikacje

## Kompetencje

Usługa prowadzi do nabycia kompetencji.

### Warunki uznania kompetencji

**Pytanie 1. Czy dokument potwierdzający uzyskanie kompetencji zawiera opis efektów uczenia się?**

Tak, dokument potwierdzający uzyskanie kompetencji zawiera szczegółowy opis efektów uczenia się, co pozwala jasnozrozumieć osiągnięte kompetencje oraz postępy w nauce.

**Pytanie 2. Czy dokument potwierdza, że walidacja została przeprowadzona w oparciu o zdefiniowane w efektach uczenia się kryteria ich weryfikacji?**

Tak, dokument potwierdza, że walidacja została przeprowadzona zgodnie z zdefiniowanymi w efektach uczenia się kryteriami weryfikacji, co gwarantuje rzetelność i wiarygodność procesu oceny osiągniętych kompetencji.

**Pytanie 3. Czy dokument potwierdza zastosowanie rozwiązań zapewniających rozdzielanie procesów kształcenia i szkolenia od walidacji?**

Tak, dokument potwierdza zastosowanie rozwiązań zapewniających rozdzielanie procesów kształcenia i szkolenia od walidacji. Dzięki temu procesowi zapewniona jest obiektywność i uczciwość oceny osiągniętych kompetencji.

## Program

**Szkolenie skierowane jest do osób interesujących się analizą danych na poziomie zaawansowanym oraz wykorzystaniem w praktyce narzędzi do obliczeń rozproszonych i analizy dużych zbiorów danych** - w tym z sektora zielonej gospodarki - rozwój kariery w tym kierunku daje obecnie ogromne możliwości i mocną pozycję na rynku pracy.

Kurs uczy praktycznego zastosowania koncepcji statystycznych w celu przewidywania i interpretacji danych, a także wykorzystania w praktyce narzędzi takich jak Apache Spark i PySpark.

W szkoleniu mogą wziąć udział zarówno osoby, które myślą o przyszłej pracy w roli data scientist, jak również osoby chcące zdobyć nowe umiejętności a zajmowanych już stanowiskach pracy, kadra kierownicza, właściciele firm, czy też specjaliści sektora zielonej gospodarki - dla których dostęp do analiz jest kluczem do podejmowania trafnych decyzji w bieżącej działalności.

Z racji dynamicznego rozwoju branż zielonych technologii, kurs uczy tworzenia modeli uczenia maszynowego w sposób pozytywnie wpływający na środowisko i wspierający zrównoważony rozwój oraz pozwala na zrozumienie roli technologii cyfrowych we wspieraniu zielonej gospodarki.

--

### **STRUKTURA I PROGRAM KURSU:**

Szkolenie pozwala na poznanie praktycznego zastosowania koncepcji statystycznych w analizie danych oraz zdobycie umiejętności przewidywania i interpretacji danych, z wykorzystaniem narzędzi takich jak Scipy, Statsmodels, Numpy i Pandas.

W drugiej części kursu uczestnicy uczą się w praktyce wykorzystania narzędzi takich jak Apache Spark i PySpark - w celu analizy dużych zbiorów danych, co jest kluczowe w obszarach takich jak prognozowanie trendów biznesowych, analiza danych środowiskowych i predykcja zmian klimatu, przetwarzanie transakcji finansowych, czy też rozwój systemów wspierających zieloną gospodarkę (np. systemów energetyki odnawialnej, czy systemów inteligentnej i zrównoważonej mobilności lokalnej).

Szkolenie obejmuje łącznie ponad 80h nauki, na które składa się:

- 56h lekcyjnych (45 min) = w przeliczeniu 42h zegarowe (60 min)) prowadzonych na żywo (on-line), na platformie webinarowej, w formie wirtualnej klasy, w formule live-coding - przez cały czas z trenerem;
- dodatkowa samodzielna praca własna kursantów w domu (ćwiczenia, projekty), z możliwością konsultacji na platformie Slack - praca ta pozwala utrwalić zdobytą podczas zajęć wiedzę i nie jest wliczana do czasu trwania usługi.

Grupy liczą maksymalnie 8-16 osób i są jednymi z najmniejszych grup na rynku.

#### **Statystyka, analiza statystyczna, modelowanie statystyczne, hipotezy i wnioskowanie:**

- statystyka opisowa, rachunek prawdopodobieństwa
- zmienne losowe i wnioskowanie statystyczne
- rozkłady statystyczne I testowanie hipotez
- czyszczenie/preprocessing danych - skalowanie, kodowanie kategorii, imputacja brakujących danych
- eksploracyjna analiza danych
- podstawy algebry liniowej do Uczenia Maszynowego
- regresja Liniowa - Maksymalne prawdopodobieństwo, metoda najmniejszych kwadratów, regularyzacja
- analiza szeregów czasowych, sarimax
- analiza danych środowiskowych (np. dane dotyczące zmian klimatycznych)

#### **Narzędzia do obliczeń rozproszonych i analizy dużych zbiorów danych:**

- wprowadzenie do Big Data
- Spark SQL
- PySpark
- MLlib
- analiza dużych zbiorów danych (dane biznesowe, środowiskowe)
- optymalizacja zbiorów danych

#### **Wykorzystanie umiejętności cyfrowych w celu wsparcia zrównoważonego rozwoju:**

- zielona gospodarka - charakterystyka poglądów dotyczących zrównoważonego rozwoju, Europejskie ramy kompetencji w zakresie zrównoważonego rozwoju (GreenComp)
- zasady środowiskowe 6R (refuse, reduce, reuse, recover, recycle, rethink)
- zastosowanie narzędzi do obliczeń rozproszonych w celu analizy danych środowiskowych (np. predykcja zmian klimatycznych)

**Całość zajęć prowadzona jest na żywo online.**

Poza zajęciami na żywo (online, wirtualna klasa), uczestnicy otrzymują dodatkowe zadania do pracy w domu, z możliwością kontaktu z prowadzącym również poza zajęciami (na platformie Slack).

**Aby osiągnąć zakładany cel realizacji usługi, uczestnik powinien być obecny w trakcie zajęć zdalnych w czasie rzeczywistym.**

--

Kurs uczy praktycznego zastosowania koncepcji statystycznych w celu przewidywania i interpretacji danych oraz analizy dużych zbiorów danych - co jest niezbędne w celu bp. predykcji zmian klimatycznych, czy też optymalizacji procesów i redukcji zużycia zasobów . Big data wspiera optymalizację zużycia energii w budynkach i infrastrukturze oraz śledzenie emisji i odpadów, umożliwia automatyzację i personalizację, wspiera rozwój energii odnawialnej i monitorowanie zużycia wody, energii i pozostałych zasobów.

Dzięki przekazywaniu umiejętności ogólnych niezbędnych do pracy w sektorze zielonej gospodarki (zgodnie z GreenComp), szkolenie przyczynia się również do tworzenia tzw. "zielonych miejsc pracy" zarówno w sektorach zielonej gospodarki, jak również w sektorach tradycyjnych.

Wiedza zdobyta podczas szkolenia wykorzystywana może być m.in. w celu realizacji inwestycji opisanych m.in. w Rozp. nr 2021/1056 PEiR(UE) ustanawiającym Fundusz Sprawiedliwej Transformacji, tj. m.in. w przypadku wdrażania technologii oraz systemów i infrastruktury zapewniającej czystą energię, redukcję emisji gazów cieplarnianych, inwestycji w energię odnawialną i w efektywność energetyczną, inteligentną i zrównoważoną mobilność lokalną, poprawę efektywności energetycznej, cyfryzację i łączność cyfrową.

## Harmonogram

Liczba przedmiotów/zajęć: 18

Przedmiot / temat zajęć	Prowadzący	Data realizacji zajęć	Godzina rozpoczęcia	Godzina zakończenia	Liczba godzin
<b>1 z 18</b> Statystyka opisowa, rachunek prawdopodobieństwa, zmienne losowe i wnioskowanie statystyczne (on-line, na żywo, wykład + ćw.)	Bartek Bilski	01-02-2025	09:00	12:00	03:00
<b>2 z 18</b> Przerwa	Bartek Bilski	01-02-2025	12:00	12:30	00:30
<b>3 z 18</b> Rozkłady statystyczne I testowanie hipotez (on-line, na żywo, wykład + ćw.)	Bartek Bilski	01-02-2025	12:30	15:30	03:00
<b>4 z 18</b> Czyszczenie/pre processing danych - skalowanie, kodowanie kategorii, imputacja brakujących danych (on-line, na żywo, wykład + ćw.)	Bartek Bilski	06-02-2025	18:00	21:00	03:00
<b>5 z 18</b> Eksploracyjna analiza danych (on-line, na żywo, wykład + ćw.)	Bartek Bilski	11-02-2025	18:00	21:00	03:00
<b>6 z 18</b> Podstawy algebry liniowej do Uczenia Maszynowego (on-line, na żywo, wykład + ćw.)	Bartek Bilski	13-02-2025	18:00	21:00	03:00

Przedmiot / temat zajęć	Prowadzący	Data realizacji zajęć	Godzina rozpoczęcia	Godzina zakończenia	Liczba godzin
<b>7 z 18</b> Regresja Liniowa - Maksymalne prawdopodobieństwo, metoda najmniejszych kwadratów, regularyzacja (on-line, na żywo, wykład + ćw.)	Bartek Bilski	15-02-2025	09:00	12:00	03:00
<b>8 z 18</b> Przerwa	Bartek Bilski	15-02-2025	12:00	12:30	00:30
<b>9 z 18</b> Analiza Szeregow czasowych, sarimax (on-line, na żywo, wykład + ćw.)	Bartek Bilski	15-02-2025	12:30	15:30	03:00
<b>10 z 18</b> Praktyczne zastosowanie analizy statystycznej, analiza danych środowiskowych (np. dane dotyczące zmian klimatycznych) (on-line, na żywo, wykład + ćw.)	Bartek Bilski	18-02-2025	18:00	21:00	03:00
<b>11 z 18</b> Wprowadzenie do Big Data - Spark (on-line, na żywo, wykład + ćw.)	Bartek Bilski	20-02-2025	18:00	21:00	03:00
<b>12 z 18</b> Spark/ Databricks- Spark SQL i analiza danych, GreenComp, zasady 6R, wykorzystane TIK w zielone gospodarce (on-line, na żywo, wykład + ćw.)	Bartek Bilski	27-02-2025	18:00	21:00	03:00

Przedmiot / temat zajęć	Prowadzący	Data realizacji zajęć	Godzina rozpoczęcia	Godzina zakończenia	Liczba godzin
<b>13 z 18</b> Spark/Databricks-Spark SQL I analiza danych (on-line, na żywo, wykład + ćw.)	Bartek Bilski	01-03-2025	09:00	12:00	03:00
<b>14 z 18</b> Przerwa	Bartek Bilski	01-03-2025	12:00	12:30	00:30
<b>15 z 18</b> Spark/Databricks- Mllib (on-line, na żywo, wykład + ćw.)	Bartek Bilski	01-03-2025	12:30	15:30	03:00
<b>16 z 18</b> Spark/Databricks- Mllib (on-line, na żywo, wykład + ćw.)	Bartek Bilski	06-03-2025	18:00	21:00	03:00
<b>17 z 18</b> Optymalizacja zbiorów danych, zastosowanie narzędzi do obliczeń rozpr. w celu analizy danych środ. (np. predykcja zmian klimatycznych) (on-line, na żywo, wykład + ćw.)	Bartek Bilski	11-03-2025	18:00	20:00	02:00
<b>18 z 18</b> Walidacja umiejętności (test w formie cyfrowej)	-	11-03-2025	21:00	22:00	01:00

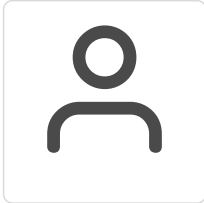
## Cennik

### Cennik

Rodzaj ceny	Cena
Koszt przypadający na 1 uczestnika brutto	5 040,00 PLN
Koszt przypadający na 1 uczestnika netto	5 040,00 PLN
Koszt osobogodziny brutto	90,00 PLN

## Prowadzący

Liczba prowadzących: 3



1 z 3

### Bartek Bilski

Programista języka Python oraz Data Scientist. Obecnie Data Scientist w Allegro, wcześniej Data Scientist w Kokoro Global, Data Analyst w cord. Absolwent studiów magisterskich Data Science na Uniwersytecie w Bath, z zapleczem w matematyce i statystyce. Entuzjasta sztucznej inteligencji, nauki i brazylijskiego jiu jitsu. Autor publikacji z zakresu Data Science, jak również kilku książek edukacyjnych.

Doświadczenie, m.in.: obecnie, Data Scientist, Allegro, 2023 - 2024, Data Scientist, Kokoro Global, 2022, Data Analyst, Cord, 2020 - 2021, Mortgage Advisor. Wykształcenie: University of Bath (Data Science)(2022) oraz University of West of England (Matematyka i Statystyka)(2017). Dodatkowe szkolenia, m.in.: Azure Databricks, Databricks Data + AI World Tour London, IRX @ DTX + UCX.

Posiada dośw. w zakresie ziel. komp. W okresie ostatnich 5 lat: m.in.: analizy i projekt segmentacji(GivEnergy Ltd., magazyny energii), skupiający się na odnawialnych źródłach energii i działaniach proekologicznych, z wykorzystaniem ML oraz modeli LLM (GPT).



2 z 3

### Szymon Budziak

Data Scientist w Transmission Dynamics Poland. Aktywny uczestnik wielu Hackatonów i wykładowca w Kole Naukowym BIT AI. Programista i inżynier systemów ACK Cyfronet przy komputerach HPC. Pasjonat programowania, sztucznej inteligencji, uczenia maszynowego oraz inwestowania. Prywatnie również trener tenisa ziemnego.

Wykształcenie: Computer Science, Engineer's degree, Università degli Studi di Napoli Federico II (2023), inż., Informatyka, AGH w Krakowie (2024). Dodatkowe szkolenia, m.in.: Python for Data Science and Machine Learning, Machine Learning Specialization by Andrew Ng.

Posiada dośw. w zakresie ziel. komp. W okresie ostatnich 5 lat: m.in.: twórca aplikacji przewidującej efektywność energetyczną w gospodarstwach domowych USA na podstawie danych energetycznych (z wykorzystaniem AI), w oparciu o technologie Python oraz Tensorflow do analizy danych energetycznych, a także API do zbierania informacji o zużyciu energii. Aplikacja miała na celu wspieranie decyzji o optymalizacji zużycia energii w gospodarstwach domowych i promowanie bardziej zrównoważonych praktyk w gospodarce.



3 z 3

### Mikołaj Kucharski

Obecnie Big Data Engineer w Allegro, wcześniej Data Scientist w Capgemini oraz Data Engineer w Ernst & Young. Zawodowo zajmuje się zagadnieniami z pogranicza data engineeringu, data science i sztucznej inteligencji w branży konsultingowej. Głównymi narzędziami jego pracy są Python, SQL, pyspark i technologie chmurowe, przede wszystkim Microsoft Azure. Jego obszarem zainteresowań są w głównej mierze konwolucyjne sieci neuronowe i ich zastosowanie w analizie obrazów. Poza



rozwojem zawodowym stawia również na rozwój fizyczny trenując brazylijskie jiu jitsu. Wolne chwile lubi spędzać przy dobrym filmie lub książce.

Doświadczenie, m.in.: obecnie, Big Data Engineer, Allegro, 2023 - 2024, Data Science, Capgemini, 2023, Data Engineer, EY, 2021 - 2023, Assistant Geotechnical Designer, Stump Franki, 2020 - 2021, Geotechnical Engineer, Jacobs, 2019 - 2020, Geotechnical Specialist. Wykształcenie: mgr, Informatyka, Polsko-Japońska Akademia Technik Komputerowych w Warszawie (2024), mgr inż, Civil Engineering - Geotechnical Engineering, SGGW w Warszawie (2021), inż, Civil Engineering - Geotechnical Engineering, SGGW w Warszawie (2022).

Posiada dośw. w zakresie ziel. komp. W okresie ostatnich 5 lat: m.in. praca nad infr. do analizy KPI z zakresu ESG, praca nad optymalizacją proc. biznesowych w celu redukcji zuż. zasobów, optymalizacja przetw. dużych zbiorów danych w celu red. kosztów i śladu węglowego.

## Informacje dodatkowe

### Informacje o materiałach dla uczestników usługi

#### W ramach szkolenia uczestnik otrzymuje:

- dostęp do materiałów oraz ćwiczeń podsumowujących zdobytą wiedzę (materiały on-line w formie pdf,html, jupyter notebook)
- zbiory danych wykorzystywane podczas ćwiczeń;
- bezpłatną licencję edukacyjną na wybrane IDE JetBrains;
- dostęp do kanałów Slack dedykowanych szkoleniu;
- dostęp do nagrań z odbytych zajęć.

### Warunki uczestnictwa

**Zaleca się, aby Uczestnicy posiadali wiedzę z zakresu programowania w języku Python i SQL.** W szkoleniu mogą wziąć udział zarówno osoby, które myślą o przyszłej pracy w roli data scientist, jak również kadra kierownicza i pracownicy sektora zielonej gospodarki, czy właściciele firm, dla których dostęp do analiz jest kluczem do podejmowania trafnych decyzji w bieżącej działalności.

W przypadku korzystania z dofinansowania, warunkiem uczestnictwa jest zapisanie się przez BUR wraz z podaniem aktualnego ID wsparcia.

### Informacje dodatkowe

Zakres zg. z RSI Woj. Śl. 2030: Techn. Inf. i kom., (i) techn. szt. int. i uczenia masz., (ii) techn. data mining, (iii) techn. zaaw. baz danych i hurtowni danych oraz z RSI Woj. Mał.. Met. i urz. służące do poz. dan.

Usługi real. są w godz. dyd. (1 godz. dyd. = 45 min.).

Przewidziane są przerwy podczas zajęć 6h, które zostały uwzgl. w harm. usługi, jednak nie wliczają się do ilości godzin samej usługi.

Podst. zw. z VAT: Dz.U.2013.1722, art. 3, ust. 1, pkt. 14 - usł. kszt. zaw. lub przekw. zaw., fin. w co najmniej 70% ze środków publ. - każdorazowo wer. w stosunku do Uczestnika.

Zapisanie się w BUR nie jest jednoznaczne z zarezerwowaniem miejsca. Prosimy o dodatkowy kontakt tel. / mail / msg / www w celu potw. miejsca.

Zawarto umowę z WUP w Krakowie w ramach projektu Małopolski Pociąg do Kariery.

Zawarto umowę z WUP w Toruniu w ramach projektu Kierunek Rozwój.

Usługi dedykowane również uczestnikom innych programów dof.

Zdobyte kompetencje dotyczą cyfrowej transformacji.

# Warunki techniczne

Zajęcia prowadzone są w czasie rzeczywistym na platformie Zoom, wraz z dostępem do kanałów grupowych na platformie Slack.

## Minimalne wymagania sprzętowe:

- komputer / laptop / lub inne urządzenie ze stałym dostępem do internetu, wyposażone w kamerę internetową

## Minimalne wymagania dotyczące parametrów łącza sieciowego:

- szybkość pobierania / przesyłania: minimalna 2 Mb/s / 128 kb/s, zalecana: 4 Mb/s / 512 kb/s

## Niezbędne oprogramowanie umożliwiające dostęp do zajęć oraz materiałów:

- przeglądarka internetowa
- Zoom w wersji bezpłatnej dla użytkownika

Uczestnicy otrzymują linki do spotkań przed każdymi zajęciami. Link umożliwiający uczestnictwo w szkoleniu jest aktywny w godzinach wskazanych na karcie usługi.

## Kontakt



**Katarzyna Hauffa**

**E-mail** [biuro@codebrainers.pl](mailto:biuro@codebrainers.pl)

**Telefon** (+48) 607 999 696