

STUDIA PODYPLOMOWE PRZETWARZANIE DANYCH – BIG DATA

Efekty uczenia się dla zajęć:

Nazwa zajęć: **Narzędzia pracy analityka**

Efekty uczenia się dla przedmiotu (opis): Po ukończeniu przedmiotu student zna zasady funkcjonowania systemów i narzędzi będących podstawą warsztatu programistycznego analityka danych, takich jak Linux, Bash, ssh, rozproszony system kontroli wersji Git, powłoka PowerShell.

Nazwa zajęć: **Zaawansowany SQL**

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna i rozumie podstawowe oraz zaawansowane koncepcje związane z relacyjnymi bazami danych i językiem SQL, włącznie z możliwościami analitycznymi tego języka i metodami optymalizacji zapytań; potrafi formułować złożone zapytania analityczne z wykorzystaniem języka SQL oraz właściwie wykorzystywać metody optymalizacji zapytań takie jak indeksy; jest gotów do twórczego rozwiązywania problemów stawianych analitykowi i pogłębiania wiedzy w oparciu o różne źródła.

Nazwa zajęć: **Analiza i wizualizacja danych w Pythonie**

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna podstawy języka Python: zmienne, podstawowe typy danych i struktury danych oraz operacje sterujące i pracę na plikach; potrafi analizować dane przy wykorzystaniu biblioteki pandas i numpy; obsługuje narzędzia Jupyter do pracy z tzw. Notebookami; wizualizuje dane z wykorzystaniem pakietu matplotlib.

Nazwa zajęć: **Hurtownie danych**

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna i rozumie zadania i architekturę współczesnej hurtowni danych; potrafi zamodelować hurtownię w modelu wielowymiarowym, wyróżnić fakty i wymiary; potrafi skonstruować proces ETL z wykorzystaniem wybranego narzędzia, przygotować dane i zasilić nimi hurtownię, w tym hurtownię w chmurze; potrafi użyć wybranego programu do szybkiego przygotowania danych do dalszej analizy; jest gotowy do twórczego wykorzystania poznanych narzędzi oraz do krytycznej oceny napotkanych rozwiązań.

Nazwa zajęć: **Raportowanie danych**

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna dobre praktyki raportowania danych, rozumie w jaki sposób narzędzia do raportowania danych korzystają ze zbioru danych (w szczególności hurtowni danych) i jak przygotować dane do wizualizacji; potrafi dobrać typ wykresu do danych; potrafi przygotować raport w **wybranych** narzędziach; zna najpopularniejsze narzędzia raportujące dostępne na rynku i wie jakie są kryteria doboru narzędzia do wizualizacji.

Nazwa zajęć: **Uczenie maszynowe**

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna i rozumie podstawowe pojęcia statystyczne związane z teorią estymacji i z teorią testowania hipotez; potrafi przeprowadzić analizę regresji, klasyfikację danych, potrafi wykonać redukcję wymiaru i przeprowadzić analizę skupień; ma świadomość etycznych i społecznych szans i zagrożeń związanych z uczeniem maszynowym.

Nazwa zajęć: Przetwarzanie rozproszone

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna i rozumie podstawowe zagadnienia przetwarzania rozproszonego, takie jak skalowanie mocy obliczeniowej (wertykalne, horyzontalne), algorytm MapReduce, podstawy Apache Hadoop i Apache Spark; zna architektury systemów rozproszonych (klastry, sieci P2P, SSI), oraz podstawowe typy usług chmurowych (IaaS, PaaS, SaaS, FaaS), potrafi porównać systemy on-premise i chmurowe.

Nazwa zajęć: Uczenie głębokie

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna i rozumie podstawowe pojęcia związane z sieciami neuronowymi oraz głębokimi sieciami neuronowymi; potrafi skonstruować i dokonać ewaluacji płytkiej sieci neuronowej; potrafi skonstruować neuronową sieć konwolucyjną oraz neuronową sieć LSTM; wykorzystuje biblioteki Keras i TensorFlow do konstrukcji sieci neuronowych oraz bibliotekę H2O do automatycznej konstrukcji modeli głębokich; jest gotów do ciągłego pogłębiania swojej wiedzy w tym dynamicznie się rozwijającym obszarze.

Nazwa zajęć: Przetwarzanie danych w chmurze

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna i rozumie podstawy systemów przetwarzania danych w chmurze, założenia i paradygmaty architektury; zna koncepcję jeziora danych (Data Lake), budowę, warstwy, różnice implementacji on-premise vs. in cloud; zna różne formaty przechowywania danych AVRO, Parquet, CSV; potrafi dobrać odpowiednie rozwiązanie do określonej sytuacji (Raw / Processed zone).

Nazwa zajęć: Inżynieria danych w Pythonie

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna i rozumie zaawansowane możliwości języka Python niezbędne w pracy inżyniera danych; potrafi programować obiektowo, zna podstawy programowania funkcyjnego; w pogłębionym stopniu poznaje pakiety Numpy i Pandas; zna pySpark; potrafi tworzyć aplikacje interaktywne;

Nazwa zajęć: Przetwarzanie strumieniowe i NoSQL

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna i rozumie zasady przetwarzania strumieniowego danych i wykorzystania różnych związanych z tym narzędzi jak Kafka, Flume, Storm, Spark Streaming; potrafi dostosować narzędzie do napotkanego problemu; zna podstawowe koncepcje baz NoSQL, w szczególności baz danych typu key-value i baz dokumentowych; potrafi tworzyć bazy i analizować je; potrafi pracować z różnymi bazami NoSQL – przede wszystkim HBase, a także Cassandra, Hive/Impala oraz Neo4j; rozumie możliwości i ograniczenia tego typu baz danych.

Nazwa zajęć: Generatywna sztuczna inteligencja

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna i rozumie podstawowe zagadnienia związane z GenAI; potrafi zastosować różne strategie *prompting-u*, tworzy i wystawia modele w chmurze, zwracając uwagę na kwestie kosztów; rozumie proces RAG (Retrieval Augmented Generation); zna metody NLP, duże modele językowe i modele text-to-image; zapoznaje się z rolą baz wektorowych.

Nazwa zajęć: Data Governance

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna i rozumie fundamenty zarządzania danymi (Data Governance), rozumie znaczenie katalogowania danych i zarządzania metadanymi (w tym *data lineage*, *data traceability*), potrafi zastosować odpowiednie metody zapewnienia i podnoszenia jakości danych, rozpoznaje podstawowe kategorie i metryki jakości danych; poznaje podstawy zarządzania danymi referencyjnymi oraz postawy systemów MDM (Master Data Management).

Nazwa zajęć: Laboratorium inżynierii danych w chmurze

Efekty uczenia się dla przedmiotu (opis): po ukończeniu modułu student/ka zna, rozumie i potrafi porównać różnorodne technologie pozwalających na przetwarzanie dużych zbiorów danych; zna podstawy Apache Nifi oraz Elasticsearch i Kibany; potrafi przetwarzać dane strumieniowe z Twittera na zadany temat za pomocą technik NLP (Natural Language Processing – Biblioteki NLTK); potrafi pracować w Notebook Jupyter; potrafi pracować samodzielnie i w grupie nad zadaniem projektem; rozumie potrzebę ciągłego pogłębiania swojej wiedzy.