

Masterclass - Analityk Danych 2.0: Excel, Power BI, SQL, Python i AI

Opis szkolenia

Masterclass to **intensywny, zaawansowany kurs analityczny**. Szkolenie kompleksowo przygotowuje do zawodu analityka danych, bazując na zróżnicowanym pakiecie narzędzi analitycznych. Program uwzględnia również zastosowanie sztucznej inteligencji w procesie analizy danych.

Narzędzia stosowane podczas kursu:

- Microsoft Excel (wraz z Power Query, PowerPivot oraz Copilot)
- Microsoft Power BI / Fabric (wraz z Power Query oraz Copilot)
- Microsoft SQL Server, PostgreSQL
- Python
- Microsoft Visual Studio Code, Google Colaboratory
- OpenAI ChatGPT oraz API
- Google NotebookLM
- Anthropic Claude

Cele szkolenia

Celem szkolenia jest przygotowanie uczestników do samodzielnej pracy analitycznej poprzez rozwój wiedzy i praktycznych umiejętności w zakresie pracy z danymi. W szczególności szkolenie ma na celu:

- zrozumienie roli analizy danych w podejmowaniu decyzji biznesowych,
- poznanie pełnego procesu analizy danych – od zdefiniowania problemu biznesowego po prezentację wyników,
- opanowanie narzędzi wykorzystywanych w pracy analityka danych,
- rozwój umiejętności pracy z danymi pochodzącymi z różnych źródeł,
- poznanie metod modelowania danych i przygotowywania ich do analizy,
- rozwój umiejętności wizualizacji danych i tworzenia raportów analitycznych,
- poznanie możliwości wykorzystania sztucznej inteligencji w analizie danych,
- przygotowanie uczestników do wdrażania rozwiązań analitycznych w organizacji.

Umiejętności

Dzięki szkoleniu uczestnik będzie:

- pozyskiwać dane z różnych źródeł, takich jak pliki, bazy danych, systemy informatyczne czy interfejsy API,

- przygotowywać i przekształcać dane do dalszej analizy, w tym czyścić dane oraz integrować je z wielu źródeł,
- pracować z relacyjnymi bazami danych i tworzyć zapytania w języku SQL,
- budować modele danych w narzędziach Excel i Power BI,
- tworzyć analizy, raporty i interaktywne dashboardy wspierające podejmowanie decyzji biznesowych,
- wykorzystywać język Python do analizy danych, automatyzacji pracy oraz przetwarzania większych zbiorów danych,
- stosować podstawowe metody statystyczne i elementy uczenia maszynowego w analizie danych,
- wykorzystywać narzędzia sztucznej inteligencji do wspomaganie analizy danych i automatyzacji pracy analitycznej,
- interpretować wyniki analiz i formułować wnioski biznesowe,
- prezentować wyniki analiz w sposób czytelny i zrozumiały dla interesariuszy.

Dodatkowo uczestnicy poznają zasady wdrażania projektów analitycznych w przedsiębiorstwie, zarządzania jakością danych oraz komunikowania wyników analiz w organizacji.

Program obejmuje także elementy rozwoju kariery, które pomagają przygotować się do pracy w zawodzie analityka danych i budować dalszą ścieżkę zawodową w obszarze data analytics.

Przygotowanie uczestników

Od uczestników naszego intensywnego szkolenia z analizy danych **wymagana jest znajomość:**

- Microsoft Excel w stopniu przynajmniej dobrym
- podstaw teoretycznych statystyki
- podstaw relacyjnych baz danych (nie jest wymagana znajomość konkretnego języka dostępu do bazy danych ani konkretnej platformy bazodanowej).

Nie jest wymagane wykształcenie kierunkowe.

- **Uczestnik musi posiadać konto Google (wystarczające jest konto darmowe).**

W ramach kursu prezentowane będą narzędzia wymagające płatnych subskrypcji lub licencji. Celem jest przekazanie Uczestnikom jak najpełniejszego obrazu dostępnych na rynku rozwiązań. **Centrum Szkoleniowe Comarch zapewni Uczestnikom płatną subskrypcję narzędzia AI Claude Pro.**

Szczegółowy program szkolenia

Moduł 1. Organizacja i przebieg procesu analizy danych

1.1. Wprowadzenie do analizy danych

Ten moduł pełni funkcję fundamentu całego szkolenia – porządkuje pojęcia, ustawia wspólne rozumienie analizy danych oraz pokazuje jej realną wartość w kontekście biznesowym. Dzięki temu uczestnicy nie tylko uczą się narzędzi, ale rozumieją, dlaczego i kiedy ich używać. Wprowadzenie do roli analityka oraz aspektów etycznych pozwala od początku kształtować właściwe podejście do pracy z danymi, co jest kluczowe w środowisku profesjonalnym.

- Czym jest analiza danych – definicja, cele, znaczenie biznesowe
- Rola analityka danych w organizacji
- Rodzaje analiz: opisowa, diagnostyczna, predykcyjna, preskrypcyjna
- Przegląd narzędzi analitycznych stosowanych w kursie
- Etyka i odpowiedzialność w pracy z danymi

1.2. **Metodyka CRISP-DM**

Metodyka CRISP-DM porządkuje cały proces analityczny – od zdefiniowania problemu biznesowego po wdrożenie rozwiązania. Jej znajomość pozwala uniknąć chaosu projektowego, zwiększa skuteczność pracy z danymi i ułatwia komunikację z interesariuszami.

- Omówienie faz metodyki CRISP-DM:
 - Zrozumienie biznesowe (Business Understanding)
 - Zrozumienie danych (Data Understanding)
 - Przygotowanie danych (Data Preparation)
 - Modelowanie (Modeling)
 - Ewaluacja (Evaluation)
 - Wdrożenie (Deployment)
- Zastosowanie CRISP-DM w praktycznych projektach analitycznych
- Iteracyjność procesu – powroty między fazami

1.3. **Sztuczna inteligencja w analizie danych**

Sztuczna inteligencja, w szczególności generatywne modele językowe i agenci AI, istotnie zmieniają sposób pracy analityków danych, automatyzując wiele zadań i wspierając proces analizy. Zrozumienie ich działania oraz ograniczeń pozwala efektywnie i bezpiecznie wykorzystywać je w praktyce projektowej.

- Podstawowe pojęcia z zakresu AI – generatywna AI i jej zastosowania. Duży model językowy oraz agent AI.
- Zarządzanie kontekstem i konwersacją. Rola pamięci agenta AI w analizie danych.
- Narzędzia wykorzystywane przez agentów AI w analizie danych – generowanie kodu, przeszukiwanie internetu.
- Dostosowywanie agentów AI do własnych potrzeb – personalizacja, umiejętności (skills).
- Bezpieczeństwo i prywatność danych.
- Praktyczna analiza danych z użyciem agenta AI – wprowadzanie danych do kontekstu, konwersacja, praca z treścią generowaną przez agenta AI.

1.4. **Sztuczna inteligencja w organizacji procesu analizy**

Wykorzystanie AI w organizacji procesu analizy pozwala usprawnić planowanie, dokumentowanie oraz definiowanie problemów biznesowych. Dzięki temu analityk może skupić się na kluczowych aspektach merytorycznych, redukując czas poświęcany na zadania organizacyjne i formalne.

- Wykorzystanie narzędzi AI (np. ChatGPT, Copilot) do planowania projektów analitycznych
- Automatyzacja dokumentowania procesów analizy za pomocą AI
- AI jako wsparcie w definiowaniu pytań biznesowych i hipotez

Moduł 2. Pozyskiwanie i przekształcanie danych

2.1. Formy magazynowania danych

Analityk danych pracuje z danymi pochodzącymi z różnych źródeł i w różnych formatach. Zrozumienie, gdzie i jak dane są przechowywane, pozwala je poprawnie odczytać, przetwarzać i analizować.

- Pliki płaskie: CSV, TXT, JSON, XML – charakterystyka, wady i zalety
- Arkusze kalkulacyjne jako źródło danych
- Relacyjne bazy danych – zasady działania
- Hurtownie danych i koncepcja Data Warehouse
- Jeziora danych (Data Lake) – wprowadzenie
- Chmurowe źródła danych – przegląd rozwiązań
- Interfejsy API jako źródło danych

2.2. Power Query – pobieranie i przekształcanie danych

Większość pracy analityka polega na przygotowaniu danych do analizy. Power Query to narzędzie, które pozwala automatycznie pobierać, czyścić i przekształcać dane, dzięki czemu oszczędza czas i zmniejsza liczbę błędów.

2.2.1. Power Query w Excel

- Łączenie z różnymi źródłami danych (pliki, bazy danych, usługi online)
- Edytor Power Query – interfejs i podstawowe operacje
- Czyszczenie danych: usuwanie duplikatów, obsługa braków, zmiana typów
- Przekształcenia kolumn: podział, scalanie, wyodrębnianie wartości
- Filtrowanie i sortowanie danych
- Operacje na tabelach: dopisywanie (Append), scalanie (Merge)
- Grupowanie i agregacja danych
- Tworzenie kolumn niestandardowych i warunkowych
- Parametryzacja zapytań
- Język M – podstawy składni i zastosowania

2.2.2. Power Query w Power BI

- Różnice i rozszerzenia w stosunku do Power Query w Excel
- Pobieranie danych z usług chmurowych i baz danych
- Zaawansowane transformacje danych
- Profilowanie danych – analiza jakości źródła
- Zarządzanie zapytaniami i optymalizacja wydajności
- Tworzenie funkcji w języku M

2.3. Język SQL i relacyjne bazy danych

SQL to podstawowy język pracy z danymi w bazach danych. Większość danych w firmach jest przechowywana w relacyjnych bazach danych, dlatego umiejętność ich obsługi i pisania zapytań jest kluczowa dla analityka.

2.3.1. Podstawy SQL Server

- Architektura SQL Server – instancje, bazy danych, schematy
- SQL Server Management Studio (SSMS) – nawigacja i konfiguracja
- Tworzenie baz danych i tabel
- Typy danych w SQL Server
- Klucze główne, obce, indeksy

2.3.2. Zapytania SQL

- Instrukcja SELECT – pobieranie danych, filtrowanie (WHERE), sortowanie (ORDER BY)
- Funkcje agregujące: COUNT, SUM, AVG, MIN, MAX
- Grupowanie danych: GROUP BY, HAVING
- Złączenia tabel: INNER JOIN, LEFT JOIN, RIGHT JOIN, FULL JOIN, CROSS JOIN
- Podzapytania (subqueries) i wyrażenia CTE (Common Table Expressions)
- Funkcje okna (Window Functions): ROW_NUMBER, RANK, DENSE_RANK, LAG, LEAD, SUM OVER
- Operacje na typach danych: daty, łańcuchy znaków, konwersje
- Instrukcje DML: INSERT, UPDATE, DELETE
- Tworzenie widoków (Views)
- Procedury składowane – podstawy tworzenia i wywoływania
- Wprowadzenie do problematyki optymalizacji zapytań – plany wykonania, indeksy

2.3.3. SQL a Power Query / Power BI

- Importowanie danych z SQL Server do Power Query i Power BI
- Tryb DirectQuery vs Import – kiedy stosować

- Pisanie natywnych zapytań SQL w Power BI

2.4. Sztuczna inteligencja w pozyskiwaniu i przekształcaniu danych

Sztuczna inteligencja coraz częściej wspiera pracę analityka, szczególnie w przygotowaniu danych. Pozwala szybciej wykonywać powtarzalne zadania i ułatwia pracę osobom mniej doświadczonym.

- Wykorzystanie AI do automatycznego rozpoznawania struktury danych
- Generowanie zapytań SQL i formuł M za pomocą narzędzi AI
- AI w procesie czyszczenia i uzupełniania brakujących danych
- Copilot w Excel / Power BI – automatyzacja przygotowania danych

Moduł 3. Tworzenie modeli danych

3.1. Modelowanie danych w Excel (Power Pivot)

Power Pivot pozwala tworzyć zaawansowane modele danych bez potrzeby używania skomplikowanych narzędzi bazodanowych. Dzięki temu analityk może pracować na dużych zbiorach danych i tworzyć bardziej zaawansowane analizy bezpośrednio w Excelu.

- Wprowadzenie do Power Pivot – włączenie dodatku, interfejs
- Importowanie danych z różnych źródeł do modelu danych
- Tworzenie relacji między tabelami
- Model gwiazdy (Star Schema) i model płatka śniegu (Snowflake Schema)
- Wprowadzenie do języka DAX:
 - Miary (Measures) vs kolumny obliczeniowe (Calculated Columns)
 - Funkcje podstawowe: SUM, AVERAGE, COUNT, DISTINCTCOUNT
 - Funkcje kontekstowe: CALCULATE, FILTER, ALL, ALLEXCEPT
 - Funkcje inteligencji czasowej: TOTALYTD, SAMEPERIODLASTYEAR, DATEADD
 - Funkcje iteracyjne: SUMX, AVERAGEX, RANKX
- Hierarchie i tabele dat (Date Tables)
- KPI – kluczowe wskaźniki efektywności w modelu danych

3.2. Modelowanie danych w Power BI

Power BI jest jednym z najważniejszych narzędzi do analizy i wizualizacji danych. Dobrze zaprojektowany model danych w Power BI decyduje o poprawności wyników, wydajności raportów i łatwości ich rozbudowy.

- Tworzenie i zarządzanie relacjami w Power BI Desktop
- Konfiguracja kierunkowości filtrowania (Single vs Bi-directional)
- Role Level Security (RLS) – bezpieczeństwo na poziomie wierszy

- Zaawansowane miary DAX w Power BI:
 - Zmienne (VAR / RETURN)
 - Tabele wirtualne (SUMMARIZE, ADDCOLUMNS, CROSSJOIN)
 - Obsługa relacji wiele-do-wielu
 - Grupy obliczeń (Calculation Groups) – wprowadzenie
 - Obliczenia wizualne
- Optymalizacja modelu danych pod kątem wydajności oraz rozmiaru
- Najlepsze praktyki w projektowaniu modeli semantycznych
- Modelowanie typowych scenariuszy biznesowych

3.3. Zaawansowane funkcje pracy z danymi w Excel

Excel jest jednym z podstawowych narzędzi pracy analityka. Znajomość zaawansowanych funkcji pozwala szybciej przetwarzać dane, automatyzować obliczenia i ograniczać pracę ręczną.

- Funkcje wyszukiwania: XLOOKUP, INDEX/MATCH, VLOOKUP
- Funkcje tablicowe i dynamiczne tablice: FILTER, SORT, UNIQUE, SEQUENCE
- Funkcje tekstowe i daty w zaawansowanych scenariuszach
- Funkcja LET – zmienne w formułach
- Funkcja LAMBDA – tworzenie własnych funkcji wielokrotnego użytku
- Zaawansowana walidacja danych
- Tabele strukturalne i odwołania strukturalne
- Połączenie Excela z zewnętrznymi źródłami danych

3.4. Sztuczna inteligencja w modelowaniu danych

Sztuczna inteligencja wspiera analityka w projektowaniu modeli danych, podpowiada rozwiązania i pomaga wykrywać błędy. Dzięki temu można szybciej tworzyć poprawne i wydajne modele.

- Wykorzystanie AI do sugerowania optymalnej struktury modelu danych
- Generowanie formuł DAX za pomocą Copilot i narzędzi AI
- Diagnostyka i wykrywanie problemów w strukturze modelu z użyciem agenta AI
- Wykorzystanie agentów AI oraz dedykowanych AI funkcji arkuszowych w Excel (COPILOT, LABS.GENERATIVEAI)

Moduł 4. Analiza i wizualizacja danych w Excel

Excel pozwala nie tylko przygotować dane, ale także je analizować i prezentować w czytelnej formie. Umiejętność wizualizacji pomaga lepiej zrozumieć dane i skuteczniej komunikować wyniki.

4.1. Tabele przestawne (Pivot Tables)

Tabele przestawne to jedno z najważniejszych narzędzi do szybkiej analizy danych w Excelu. Pozwalają w prosty sposób podsumować duże zbiory danych i odkrywać zależności.

- Tworzenie i konfiguracja tabel przestawnych
- Grupowanie danych w tabelach przestawnych
- Pola obliczeniowe i elementy obliczeniowe
- Fragmentatory (Slicers) i osie czasu (Timelines)
- Połączenie tabel przestawnych z modelem danych Power Pivot
- Tabele przestawne oparte na miarach DAX

4.2. Wizualizacja danych w Excel

Dobra wizualizacja pozwala szybko zrozumieć dane i wyciągać wnioski. Analityk musi umieć przedstawić wyniki w sposób czytelny i zrozumiały dla innych.

- Zasady skutecznej wizualizacji danych
- Wykresy standardowe: kolumnowe, liniowe, kołowe, punktowe
- Wykresy zaawansowane: waterfall, combo, histogram, box plot
- Wykresy przestawne (Pivot Charts)
- Sparklines – mikrowykresy w komórkach
- Formatowanie warunkowe jako narzędzie wizualizacji
- Tworzenie interaktywnych dashboardów w Excel
- Mapy 3D (Power Map) – wizualizacja danych geograficznych

4.3. Analiza statystyczna w Excel

Analiza statystyczna pozwala lepiej zrozumieć dane, wykrywać zależności i przewidywać przyszłe wyniki. Jest podstawą bardziej zaawansowanej analizy danych.

- Statystyki opisowe: średnia, mediana, odchylenie standardowe, percentyle
- Dodatek Analysis ToolPak – przegląd narzędzi
- Analiza korelacji i regresji
- Analiza scenariuszy: Solver, tabele danych, menedżer scenariuszy
- Prognozowanie z wykorzystaniem wbudowanych funkcji (FORECAST)

4.4. Sztuczna inteligencja w analizie danych w Excel

Sztuczna inteligencja wspiera analizę danych, automatyzując wiele zadań i pomagając szybciej znaleźć istotne wnioski. Dzięki temu analityk może skupić się na interpretacji wyników, a nie tylko na obliczeniach.

- Funkcje AI w Excel: XLOOKUP z inteligentnymi sugestiami
- Copilot w Excel – automatyczne generowanie formuł, wykresów i analiz
- Analiza danych z użyciem funkcji IDEAS (Analyze Data)
- Generowanie podsumowań i insightów przez AI
- Automatyczne wykrywanie trendów i anomalii

Moduł 5. Analiza i wizualizacja danych w Power BI

Power BI to narzędzie stworzone do zaawansowanej analizy i wizualizacji danych. Umożliwia tworzenie interaktywnych raportów i dashboardów, które wspierają podejmowanie decyzji biznesowych.

5.1. Interfejs i struktura Power BI Desktop

Znajomość interfejsu Power BI pozwala sprawnie poruszać się po narzędziu i efektywnie tworzyć raporty. To podstawa dalszej pracy z analizą i wizualizacją danych.

- Widoki: raport, dane, model
- Obszar filtrów, panele wizualizacji i pól
- Zarządzanie stronami raportu

5.2. Tworzenie wizualizacji w Power BI

Wizualizacje w Power BI pozwalają przekształcić dane w czytelne i interaktywne raporty. Dzięki nim można łatwo odkrywać zależności i prezentować wyniki w sposób zrozumiały dla odbiorców.

- Wizualizacje standardowe: wykresy kolumnowe, liniowe, kołowe, mapowe
- Wizualizacje zaawansowane: Treemap, Funnel, Waterfall, Ribbon, Decomposition Tree
- Tabele i macierze – formatowanie warunkowe, paski danych
- Obliczenia wizualne – formuły DAX na poziomie wizualizacji
- Karty (Cards) i wskaźniki KPI
- Wizualizacje niestandardowe z AppSource
- Tooltips – niestandardowe podpowiedzi
- Drillthrough i drilldown – nawigacja w danych
- Zakładki (Bookmarks) i przyciski nawigacyjne

5.3. Interaktywność i zaawansowane funkcje raportów

Interaktywność raportów pozwala użytkownikom samodzielnie analizować dane i dostosowywać widok do swoich potrzeb. Dzięki temu raporty są bardziej użyteczne i wspierają podejmowanie decyzji.

- Fragmentatory i filtry – konfiguracja i synchronizacja
- Interakcje między wizualizacjami (Edit Interactions)

- Parametry What-if i dynamiczne miary
- Grupy obliczeń – dynamiczne przełączanie miar
- Row-Level Security – konfiguracja i testowanie
- Tworzenie kompletnych dashboardów analitycznych
- Dobre praktyki projektowania raportów (UX/UI)

5.4. Publikowanie i udostępnianie raportów

Tworzenie raportu to tylko część pracy analityka. Kluczowe jest jego udostępnienie i zapewnienie, że dane są aktualne oraz dostępne dla odpowiednich osób.

- Power BI Service – publikowanie raportów w chmurze
- Obszary robocze (Workspaces) i aplikacje
- Odświeżanie danych – harmonogramy, bramy danych (Gateway)
- Udostępnianie raportów i dashboardów
- Eksport danych i raportów

5.5. Sztuczna inteligencja w Power BI

Funkcje AI w Power BI pozwalają szybciej analizować dane i odkrywać zależności bez konieczności budowania skomplikowanych modeli. Ułatwiają także interpretację wyników i ich prezentację.

- Key Influencers – automatyczna analiza czynników wpływających
- Decomposition Tree – interaktywna analiza przyczynowo-skutkowa
- Smart Narratives – automatyczne generowanie opisów tekstowych
- Anomaly Detection – wykrywanie odchyleń w danych szeregów czasowych
- Copilot w Power BI – tworzenie raportów i miar za pomocą języka naturalnego
- Automatyczne generowanie insightów (Quick Insights)

Moduł 6. Język Python w analizie i wizualizacji danych

Python to jedno z najważniejszych narzędzi w analizie danych. Umożliwia automatyzację pracy, analizę dużych zbiorów danych oraz tworzenie zaawansowanych modeli i wizualizacji.

6.1. Podstawy języka Python

Podstawy Pythona są niezbędne, aby móc wykorzystywać ten język w analizie danych. Pozwalają zrozumieć, jak pisać kod i automatyzować pracę z danymi.

- Konfiguracja środowiska: Python, Jupyter Notebook, VS Code
- Wykorzystanie Google Colaboratory
- Typy danych, zmienne, operatory

- Struktury danych: listy, krotki, słowniki, zbiory
- Instrukcje warunkowe i pętle
- Funkcje – definiowanie, argumenty, wartości domyślne
- Praca z plikami (odczyt, zapis)

6.2. Python w analizie danych – biblioteki

Biblioteki takie jak NumPy i Pandas są podstawą analizy danych w Pythonie. Umożliwiają szybkie przetwarzanie, analizę i łączenie danych z różnych źródeł, także z baz danych.

6.2.1. NumPy

- Tablice wielowymiarowe (ndarray)
- Operacje matematyczne i statystyczne na tablicach
- Indeksowanie, wycinanie, filtrowanie

6.2.2. Pandas

- DataFrame i Series – tworzenie, indeksowanie, selekcja danych
- Importowanie danych: CSV, Excel, JSON, SQL
- Czyszczenie danych: brakujące wartości, duplikaty, typy danych
- Transformacje: merge, join, concat, pivot, melt, stack/unstack
- Grupowanie i agregacja (groupby)
- Funkcje okna (rolling, expanding, shift)
- Praca z seriami czasowymi (datetime, resample)
- Optymalizacja pamięci i wydajności

6.2.3. Połączenie Pythona z SQL Server

- Biblioteki: pyodbc, sqlalchemy
- Wykonywanie zapytań SQL z poziomu Pythona
- Zapisywanie wyników do bazy danych
- Automatyzacja raportowania z bazy danych

6.3. Wizualizacja danych w Pythonie

Wizualizacja danych w Pythonie pozwala tworzyć bardziej zaawansowane i interaktywne wykresy niż w podstawowych narzędziach. Dzięki temu analityk może lepiej analizować dane i prezentować wyniki.

6.3.1. Matplotlib

- Tworzenie wykresów: liniowe, słupkowe, kołowe, punktowe
- Formatowanie: tytuły, osie, legendy, adnotacje

- Subplots – wiele wykresów na jednym rysunku
- Eksport wykresów do plików

6.3.2. Seaborn

- Wykresy statystyczne: heatmap, pairplot, boxplot, violinplot
- Wizualizacja rozkładów i korelacji
- Stylizacja i palety kolorów

6.3.3. Plotly

- Interaktywne wizualizacje
- Wykresy 3D i mapy
- Dashboardy z użyciem Plotly Dash – wprowadzenie

6.4. Analiza dużych zbiorów danych z wykorzystaniem Apache Spark

Przy dużych zbiorach danych standardowe narzędzia mogą być niewystarczające. Apache Spark pozwala przetwarzać dane w sposób rozproszony, co umożliwia analizę bardzo dużych wolumenów danych.

- Wprowadzenie do Big Data i Apache Spark
- Architektura Spark: Driver, Executor, Cluster Manager
- PySpark – konfiguracja i podstawy
- Spark DataFrames – tworzenie, transformacje, akcje
- Spark SQL – zapytania na dużych zbiorach danych
- Przetwarzanie danych w trybie wsadowym i strumieniowym
- Integracja PySpark z Pandas

7. Uczenie maszynowe w analizie danych

Uczenie maszynowe pozwala nie tylko analizować dane historyczne, ale także przewidywać przyszłe zdarzenia i automatycznie wykrywać wzorce. To ważny krok w kierunku bardziej zaawansowanej analizy danych.

7.1 Podstawy uczenia maszynowego

Podstawy uczenia maszynowego pozwalają zrozumieć, jak działają modele predykcyjne i jak poprawnie je budować. Bez tej wiedzy trudno tworzyć wiarygodne i użyteczne rozwiązania.

- Uczenie nadzorowane vs nienadzorowane
- Rodzaje algorytmów: klasyfikacja, regresja, grupowanie, reguły asocjacyjne
- Podział danych: zbiór treningowy, walidacyjny, testowy
- Metryki oceny modeli: accuracy, precision, recall, F1-score, RMSE, MAE, R2
- Przeuczenie (overfitting) i niedouczenie (underfitting)

- Walidacja krzyżowa (Cross-Validation)
- Inżynieria cech: kodowanie zmiennych, skalowanie, selekcja cech

7.2. Algorytmy klasyfikacji i regresji

Znajomość podstawowych algorytmów pozwala dobrać odpowiednią metodę do problemu i budować skuteczne modele predykcyjne. To kluczowy element praktycznego wykorzystania uczenia maszynowego.

- Regresja liniowa i logistyczna
- Drzewa decyzyjne i lasy losowe (Random Forest)
- K-najbliższych sąsiadów (KNN)
- Support Vector Machines (SVM) – wprowadzenie
- Gradient Boosting: XGBoost, LightGBM – wprowadzenie

7.3. Algorytmy grupowania i redukcji wymiarów

Algorytmy grupowania i redukcji wymiarów pozwalają odkrywać ukryte struktury w danych oraz upraszczać zbiory danych bez dużej utraty informacji. Są szczególnie ważne, gdy nie mamy gotowych etykiet danych.

- K-Means – grupowanie danych
- DBSCAN
- Analiza skupień hierarchicznych
- PCA (Principal Component Analysis) – redukcja wymiarów

7.4. Biblioteka scikit-learn – funkcje zaawansowane

Zaawansowane funkcje scikit-learn pozwalają budować bardziej uporządkowane, dokładne i zrozumiałe modele. Dzięki nim można automatyzować proces modelowania i lepiej interpretować wyniki.

- Pipeline – budowanie potoków przetwarzania
- Strojenie hiperparametrów (GridSearchCV, RandomizedSearchCV)
- Interpretacja modeli: Feature Importance, SHAP – wprowadzenie

7.5. Sztuczna inteligencja w analizie danych z użyciem języka Python

Sztuczna inteligencja znacząco rozszerza możliwości analizy danych w Pythonie. Umożliwia pracę z tekstem, automatyzację analizy oraz szybsze tworzenie wniosków i raportów.

- Przetwarzanie języka naturalnego (NLP) – podstawy:
 - Tokenizacja i wektoryzacja tekstu
 - Embedding
 - Analiza sentymentu z użyciem gotowych modeli
- Wykorzystanie API modeli językowych (OpenAI, Anthropic) w analizie danych
- Automatyzacja analizy danych z użyciem AI:
 - Generowanie kodu analitycznego przez modele językowe

- Automatyczne podsumowywanie wyników analizy
- Wsparcie eksploracyjnej analizy danych (EDA) przez narzędzia AI
- Wykorzystanie kodu generowanego przez AI w analizie danych wrażliwych
- Generatywna AI w tworzeniu raportów i prezentacji wyników
- Odpowiedzialne stosowanie AI – bias, interpretowalność, etyka

Moduł 8. Wdrażanie analizy danych w przedsiębiorstwie

Analiza danych ma największą wartość wtedy, gdy jest wykorzystywana w praktyce w organizacji. Wdrożenie rozwiązań analitycznych pozwala podejmować lepsze decyzje i usprawniać procesy biznesowe.

8.1. Proces wdrożenia projektów analitycznych

Nawet najlepsza analiza nie ma wartości, jeśli nie zostanie poprawnie wdrożona i zrozumiana przez odbiorców. Proces wdrożenia pozwala przejść od pomysłu do realnego zastosowania w biznesie.

- Od prototypu do produkcji – ścieżka wdrożenia
- Definiowanie wymagań biznesowych i KPI
- Komunikowanie wyników analizy interesariuszom
- Storytelling z danymi – budowanie narracji analitycznej
- Dokumentacja projektu analitycznego

8.2. Ewaluacja wyników analizy

Ocena wyników pozwala upewnić się, że analiza i modele działają poprawnie i przynoszą wartość biznesową. Bez ewaluacji trudno zauważyć błędy i utrzymać jakość rozwiązań w czasie.

- Metody oceny jakości analizy
- Walidacja wyników z ekspertami domenowymi
- Testowanie A/B – zasady i zastosowania
- Monitorowanie modeli w produkcji
- Cykl życia modelu analitycznego

8.3. Zarządzanie danymi w organizacji (przegląd)

Dane w organizacji muszą być dobrze zarządzane, bezpieczne i zgodne z przepisami. Bez tego analiza może prowadzić do błędnych wniosków lub naruszeń prawa.

- Data Governance – zasady zarządzania danymi
- Jakość danych – wymiary i metryki
- RODO/GDPR – ochrona danych osobowych w kontekście analityki
- Bezpieczeństwo danych i kontrola dostępu

- AI Act – klasyfikacja rozwiązań AI oraz wymagania względem ich wdrożenia

8.4. Sztuczna inteligencja we wdrażaniu analityki

Sztuczna inteligencja wspiera wdrażanie analityki, automatyzuje procesy i ułatwia dostęp do danych w organizacji. Pomaga także budować kulturę opartą na danych.

- AI w automatyzacji raportowania i dystrybucji wyników
- Chatboty analityczne – konwersacyjny dostęp do danych
- Integracja narzędzi AI z istniejącą infrastrukturą danych
- Trendy i kierunki rozwoju AI w analizie danych
- Budowanie kultury data-driven z wykorzystaniem AI

Moduł 9. Konsultacje z doradcą HR

Umiejętności techniczne to tylko część sukcesu. Ważne jest także umiejętne zaplanowanie kariery, przygotowanie się do rekrutacji i budowanie swojej pozycji na rynku pracy.

9.1. Rozwój kariery analityka danych

- Profil kompetencyjny analityka danych
- Ścieżki kariery: Data Analyst, BI Analyst, Data Scientist, Data Engineer
- Przygotowanie CV i portfolio projektów analitycznych
- Rozmowa kwalifikacyjna – typowe pytania i zadania rekrutacyjne
- Certyfikacje branżowe: Microsoft, Google, Tableau
- Budowanie marki osobistej i networking w branży danych

Metoda realizacji

- **Kurs składa się z 26 zdalnych spotkań na żywo.** Spotkania odbywają się dwa razy w tygodniu. **Każde trwa 4 godziny szkoleniowe.** Spotkania mają formę wykładu oraz demonstracji wykorzystania oprogramowania.

Spotkania są nagrywane. Uczestnicy kursu mają dostęp do wszystkich nagrań na platformie

- **60 godzin szkolniowych materiału e-learningowego (E-Super pakiet dla analityka – <https://www.comarch.pl/szkolenia/e-learning/e-learning-analiza-danych/e-pakiet-dla-analityka-superpremium/>)**
- Dostępne materiały szkoleniowe zawierają dodatkowe nagrania video, pliki przykładowe i ćwiczeniowe oraz filmy instruktażowe omawiające wymagania względem komputera oraz pokazujące instalację i konfigurację w środowisku użytkownika.
- **W trakcie kursu uczestnicy otrzymają do realizacji 13 zadań.** W celu zaliczenia kursu i uzyskania certyfikatu konieczne jest zrealizowanie co najmniej 10 zadań.

Liczba dni, liczba godzin szkoleniowych

Kurs składa się z 26 zdalnych spotkań na żywo. Spotkania odbywają się dwa razy w tygodniu. Każde trwa 4 godziny szkoleniowe (104 godziny pracy z trenerem)

80 godzin szkoleniowych e-learningu

Harmonogram (sobota – 9.00-12.00, środa 18.00-21.00)

1. 17 czerwca
2. 20 czerwca
3. 1 lipca
4. 4 lipca
5. 8 lipca
6. 11 lipca
7. 15 lipca
8. 18 lipca
9. 22 lipca
10. 25 lipca
11. 29 lipca
12. 1 sierpnia
13. 5 sierpnia
14. 8 sierpnia
15. 19 sierpnia
16. 22 sierpnia
17. 26 sierpnia
18. 16 września
19. 19 września
20. 23 września
21. 26 września

5 spotkań w październiku (do ustalenia)