

## **PoC – Predykcja wyników meczów piłkarskich na podstawie danych uzyskanych z publicznych API**

### **Cel:**

Zbudowanie i wdrożenie modelu uczenia maszynowego, który na podstawie danych o składach, wynikach, kontuzjach i statystykach meczowych będzie przewidywać wyniki przyszłych spotkań piłkarskich (zwycięstwo/remis/porażka).

### **Użyte narzędzia w Azure:**

<b>Obszar</b>	<b>Narzędzie</b>
Gromadzenie danych	Azure Data Factory (web scraping/API), Azure Logic Apps
Przetwarzanie danych	Azure Databricks / Azure Synapse
Składowanie danych	Azure Data Lake Storage Gen2
Trenowanie modelu ML	Azure Machine Learning
Automatyzacja przepływów	Azure Data Factory
Wizualizacja wyników	Power BI
Bezpieczeństwo	Azure RBAC, Azure Key Vault

### **Przebieg:**

#### **1. Zbieranie danych**

Dane historyczne i bieżące pochodzące z publicznych API (np. transfermarkt, football-data.org, Opta) są pobierane automatycznie przez **Azure Data Factory** lub **Logic Apps**.

#### **2. Przechowywanie danych**

Dane surowe i przetworzone są zapisywane w **Azure Data Lake Storage Gen2** w dwóch warstwach:

**Raw** – surowe API/scrapy bez przekształceń,

**Cleaned** – dane oczyszczone, scalone i gotowe do analizy.

#### **3. Przetwarzanie i przygotowanie danych**

Dane są przetwarzane w **Azure Databricks**:

- Czyszczenie danych,
- Inżynieria cech (np. forma drużyny w ostatnich 5 meczach, średni xG przeciwników, liczba kontuzji),

- Łączenie danych z różnych źródeł

#### **4. Budowa i trenowanie modelu**

W **Azure Machine Learning** trenowany jest model ML (np. XGBoost, LightGBM, sieci neuronowe), którego celem jest przewidywanie W/R/P (Win/Draw/Loss),

#### **5. Automatyzacja i integracja**

**Azure Data Factory** cyklicznie:

- Pobiera i przetwarza nowe dane z API,
- Wywołuje endpoint modelu ML dla przyszłych spotkań,
- Zapisuje wyniki predykcji w bazie danych lub w Data Lake.

#### **6. Wizualizacja i raportowanie**

- **Power BI** prezentuje:
  - Predykcje meczów nadchodzącej kolejki,
  - Najbardziej wpływowe czynniki wg modelu,
  - Skuteczność modelu w ostatnich typowaniach.
- Możliwe alerty dla analityków, gdy:
  - Model daje nietypową prognozę,
  - Wykryje wpływową kontuzję w składzie.