

# WYDZIAŁ MATEMATYKI I NAUK INFORMACYJNYCH

## KONSPEKT PROJEKTU

### ANALIZA WPŁYWU CZYNNIKÓW EKONOMICZNYCH NA EKSPORT POSZCZEGÓLNYCH TOWARÓW

Autorzy:

Agata Krawczyk  
Urszula Szczęsna  
Filip Szlingiert

Warszawa, 20 Listopada 2025

### Spis treści

1. Cel projektu i potencjalne korzyści z wdrożenia	2
2. Dane	2
2.1. Dane dotyczące eksportu z Polski	2
2.2. Dane makroekonomiczne importerów	3
2.3. Dane dotyczące kursów walut	3
3.1. Składowanie	3
3.2. Przepływ danych	3
3.3. Warstwa analityczna	4
3.4. Pozostałe	4
4. Podział pracy	4

# 1. Cel projektu i potencjalne korzyści z wdrożenia

Celem projektu jest stworzenie systemu do analizowania wpływu różnych czynników ekonomicznych, makroekonomicznych i kursów walut na eksport poszczególnych grup towarów z Polski. Projekt będzie wykorzystywał dane dotyczące eksportu według kodów HS z bazy Comtrade, dane makroekonomiczne oraz historyczne kursy walut tak, aby przeliczyć wartości eksportu na realne ceny.

System będzie umożliwiał agregowanie danych eksportowych według kraju, rodzaju towaru oraz okresu, co umożliwi zidentyfikować czynniki mające największy wpływ na eksport. Analiza pozwoli na ocenę trendów w czasie, zależności między eksportem a sytuacją gospodarczą partnerów handlowych oraz ewentualne przygotowanie rekomendacji dla firm dotyczących rynków docelowych.

## 2. Dane

W projekcie zostaną wykorzystane trzy główne źródła danych, które są publicznie dostępne i legalne do przetwarzania w celach analitycznych. Dane będą pobierane przez API lub dostępne w formatach CSV/JSON.

### 2.1. Dane dotyczące eksportu z Polski

- Pochodzą z [bazy UN Comtrade](#) i obejmują eksport z Polski na przestrzeni ostatnich 10 lat według kodów HS do różnych krajów świata.
- Dane zawierają:
  - kod HS oraz opis towaru
  - kod jednostki
  - ilość
  - waga
  - wartość eksportu w USD
  - kraj importera
  - data importu
- Planowane jest odświeżania danych na poziomie miesięcznym lub rocznym.
- Grupy produktów:
  - 870323 - samochody osobowe
  - 871120 - motocykle
  - 847130 - sprzęt komputerowy
  - 850780 - akumulatory litowo-jonowe/baterie
  - 300490 - leki gotowe
  - 330499 - kosmetyki
  - 100119 - pszenica
  - 080810 - jabłka
  - 020321 - wieprzowina
  - 040690 - sery
  - 220830 - wódka
  - 240220 - papierosy i wyroby tytoniowe
  - 220300 - piwo
  - 220290 - napoje bezalkoholowe

## 2.2. Dane makroekonomiczne importerów

Baza [World Development Indicators](#) zawiera gamę wskaźników makroekonomicznych wielu państw, grup państw i regionów, tj. PKB, populacja, inflacja, wartość importowanych dóbr, wskaźniki rozwoju gospodarczego. Dane zaczęły być gromadzone w 1960 roku, natomiast nie wszystkie wskaźniki oraz nie dla każdego państwa, są one dostępne w pełnych przedziałach czasowych. Dane zostaną pobrane w formacie JSON przy pomocy API.

## 2.3. Dane dotyczące kursów walut

- Dane pochodzą z [bazy Narodowego Banku Polskiego](#) i zawierają kursy walut od 2 stycznia 2002 r. Dostępne są zarówno kursy średnie jak i kursy kupna oraz sprzedaży.
- Dane uśrednione zawierają:
  - datę
  - średni kurs
  - nazwę waluty
  - kod waluty
  - państwo
- Dane kupna i sprzedaży zawierają:
  - datę
  - kurs kupna i sprzedaży
  - nazwę waluty
  - kod waluty
  - państwo
- Dodatkowo dostępne są dane dotyczące ceny złota, które zawierają:
  - Datę publikacji
  - Wyliczona w NBP cena 1g złota w próbie 1000
- Dane będą pobierane przy użyciu API w formacie JSON.

Przykład danych uśrednionych:

```
{  
  "table": "B",  
  "country": "Ekwador",  
  "symbol": "618",  
  "currency": "dolar",  
  "code": "USD",  
  "rates": [  
    {  
      "no": "10/B/NBP/2004",  
      "effectiveDate": "2004-04-28",  
      "mid": 4.011  
    }  
  ]}
```

Przykład danych kupna i sprzedaży:

```
{  
  "table": "C",  
  "currency": "dolar amerykański",  
  "code": "USD",  
  "rates": [  
    {  
      "no": "224/C/NBP/2025",  
      "effectiveDate": "2025-11-19",  
      "bid": 3.6219,  
      "ask": 3.6951  
    }  
  ]}
```

## 3. Stos architektoniczny

### 3.1. Składowanie

- **Apache Hive** - wszystkie dane surowe będą składowane w tabelach Hive.
- **HDFS** - przechowywanie surowych plików parquet, wykorzystywane przez Hive i Spark

### 3.2. Przepływ danych

- **Apache NiFi** - wszystkie dane będą pobierane i wstępnie przetwarzane za pomocą NiFi z ewentualnym wsparciem skryptowym przy trudniejszych operacjach.
- Flow będzie obejmować:
  - pobieranie danych z API
  - mapowanie kolumn
  - standaryzacja nazw krajów
  - standaryzacja jednostek
  - zapis danych do HDFS i Hive

### 3.3. Warstwa analityczna

- **Apache Spark** - przetwarzanie danych wielkoskalowych, agregacje, feature engineering (cena jednostkowa, wartość realna, udział w rynku, )
- **Jupyter Notebook** - interaktywne analizy, wizualizacje

### 3.4. Pozostałe

- Kod źródłowy będzie utrzymywany w repozytorium na **GitHubie**. Rozwiązanie będzie zaimplementowane na lokalnym środowisku.

## 4. Podział pracy

	Agata	Urszula	Filip
Idea projektu	X		
Pozyskiwanie danych	X	X	X
Przygotowanie przepływów danych			X
Składowanie danych		X	
Przetwarzanie danych	X		X
Wsadowa analiza danych		X	
Przygotowanie wizualizacji	X		X
Sporządzenie dokumentacji	X	X	X
Utrzymanie projektu na GitHubie	X	X	X

Tabela 1: Podział obowiązków w projekcie