

# System służący do przewidywania wartości walut - dokumentacja

Mikołaj Marton

August 21, 2025

## Contents

<b>1</b>	<b>Wprowadzenie</b>	<b>1</b>
<b>2</b>	<b>Pobieranie wartości walut</b>	<b>2</b>
<b>3</b>	<b>Baza danych</b>	<b>3</b>
3.1	Tabela <code>Currencies</code>	3
3.2	Tabela <code>ExchangeRates</code>	3
3.3	Tabela <code>Predictions</code>	3
<b>4</b>	<b>Modele predykcyjne</b>	<b>3</b>
4.1	Cechy użyte w modelu	4
4.2	Ocena modeli	4
4.3	Walidacja modeli	4
<b>5</b>	<b>Dashboard</b>	<b>5</b>
5.1	Tytuł dashboardu z data predykcji	5
5.2	Tabela prognozowanych zysków	5
5.3	Wykres zmian wartości waluty	6
5.4	Wybór waluty	6
5.5	Filtr współczynnika ryzyka	6
5.6	Informacje o tabelach i współczynniku ryzyka	6
<b>6</b>	<b>Automatyzacja</b>	<b>7</b>
6.1	Baza danych	7
6.2	Integracja z PowerBI	7
6.3	Automatyzacja uruchamiania skryptów w Windows	7
6.4	Automatyzacja uruchamiania skryptów w Linux	8

## 1 Wprowadzenie

W dynamicznie zmieniającym się świecie finansów, precyzyjne przewidywanie kursów walut stanowi kluczowy element strategii inwestycyjnych. Prezentowany system został zaprojektowany z myślą o zapewnieniu użytkownikom narzędzia umożliwiającego podejmowanie świadomych decyzji inwestycyjnych na podstawie zaawansowanych analiz danych i prognoz.

System codziennie pobiera najnowsze dane z API Narodowego Banku Polskiego (NBP), który dostarcza wiarygodnych i aktualnych informacji o kursach walut. Dzięki zastosowaniu zaawansowanych algorytmów uczenia maszynowego, system jest w stanie przewidzieć wartości walut na następny dzień. Modele przewidywania zostały stworzone indywidualnie dla każdej waluty, co pozwala na dokładne prognozowanie i uwzględnienie specyficznych cech każdej z nich.

Codziennie aktualizowany dashboard prezentuje użytkownikowi wyniki tych prognoz, wskazując waluty, które mogą przynieść zysk przy zakupie. Dashboard jest intuicyjny i interaktywny, co umożliwia łatwe śledzenie zmian i podejmowanie szybkich decyzji inwestycyjnych.

System jest w pełni zautomatyzowany – każdego dnia uruchamiany jest skrypt, który pobiera nowe dane, aktualizuje modele uczenia maszynowego na podstawie najnowszych informacji oraz generuje prognozy na kolejny dzień. Dzięki temu użytkownicy mają pewność, że korzystają z najbardziej aktualnych i precyzyjnych danych.

Niniejsza dokumentacja ma na celu szczegółowe omówienie działania systemu, w tym procesu pobierania danych, mechanizmów uczenia maszynowego oraz sposobu prezentacji wyników na dashboardzie. Każdy z tych elementów zostanie opisany w kolejnych sekcjach, aby zapewnić pełne zrozumienie i umożliwić skuteczne korzystanie z systemu.

## 2 Pobieranie wartości walut

Wartości na temat walut pobierane są codziennie z API Narodowego Banku Polskiego (NBP). NBP dostarcza wiarygodnych i aktualnych informacji o kursach walut, co stanowi fundament dla przeprowadzanych analiz i prognoz. Proces pobierania danych z NBP jest kluczowym elementem systemu, który zapewnia, że użytkownicy mają dostęp do najbardziej aktualnych informacji, co jest niezbędne do podejmowania trafnych decyzji inwestycyjnych.

Dane dotyczące kursów walut znajdują się w dwóch głównych tabelach udostępnianych przez NBP:

- **Tabela "A"** - Zawiera codzienne kursy walut. Dzięki codziennej aktualizacji, modele przewidywania oparte na danych z tej tabeli charakteryzują się wysoką precyzją. Dane z tabeli "A" obejmują najważniejsze waluty świata, które są kluczowe dla globalnych transakcji finansowych.
- **Tabela "B"** - Zawiera kursy walut udostępniane raz w tygodniu. Z tego powodu, prognozy dotyczące walut z tej tabeli są na ogół mniej dokładne w porównaniu do tych z tabeli "A". Tabela "B" często obejmuje waluty mniej popularne lub te, które nie są przedmiotem codziennych transakcji.

Proces pobierania danych rozpoczyna się każdego dnia od uruchomienia skryptu, który automatycznie łączy się z API NBP. Skrypt ten jest odpowiedzialny za pobieranie najnowszych wartości kursów walut oraz ich przetwarzanie. Dane są następnie weryfikowane pod kątem ich kompletności i dokładności, co zapewnia, że do modeli uczenia maszynowego trafiają wyłącznie sprawdzone informacje.

Aby zapewnić wysoką jakość modeli uczenia maszynowego, pobrano historyczne dane dla każdej waluty od 1 stycznia 2023 roku aż do 1 czerwca 2024 roku. Dane te posłużyły jako baza treningowa dla algorytmów, umożliwiając precyzyjne prognozowanie przyszłych wartości kursów walut. Analiza historycznych danych pozwala modelom na identyfikację wzorców i trendów, które mogą mieć wpływ na przyszłe zmiany kursów.

Każda z tabel jest regularnie aktualizowana:

- Dla tabeli "A" dane są pobierane codziennie, co zapewnia, że modele przewidywania są zawsze oparte na najbardziej aktualnych informacjach.
- Dla tabeli "B" dane są pobierane raz w tygodniu, co oznacza, że prognozy mogą nie być tak precyzyjne, jak w przypadku tabeli "A", jednak nadal dostarczają cennych informacji dla inwestorów zainteresowanych mniej popularnymi walutami.

Automatyzacja procesu pobierania danych eliminuje ryzyko błędów ludzkich i gwarantuje, że modele uczenia maszynowego są zawsze aktualne. Skrypty są zaprojektowane tak, aby wykrywać wszelkie anomalie w danych, co pozwala na szybkie reagowanie na wszelkie nieprawidłowości.

Dzięki takiemu podejściu, użytkownicy systemu mogą podejmować świadome decyzje inwestycyjne, mając pewność, że dane, na których się opierają, są aktualne i precyzyjne. Proces ten jest kluczowym elementem systemu, który nieustannie dąży do zapewnienia najwyższej jakości analiz i prognoz kursów walut.

## 3 Baza danych

Baza danych systemu zawiera trzy główne tabele: **Currencies**, **ExchangeRates** oraz **Predictions**. Każda z tych tabel pełni specyficzną rolę w przechowywaniu i zarządzaniu danymi dotyczącymi kursów walut oraz prognoz.

### 3.1 Tabela Currencies

Tabela **Currencies** zawiera informacje o poszczególnych walutach. Atrybuty tej tabeli są następujące:

- **currency\_id** - Unikalny identyfikator waluty.
- **name** - Nazwa waluty.
- **code** - Trzyliterowy kod waluty.
- **rmape** - Miara dokładności modelu dla danej waluty.
- **table\_val** - Wartość przyjmująca "a" lub "b", wskazująca, z której tabeli NBP pochodzi dana waluta.

Wartości w tabeli **Currencies** zostały uzupełnione w momencie tworzenia systemu i nie są aktualizowane.

### 3.2 Tabela ExchangeRates

Tabela **ExchangeRates** przechowuje kursy walut pobierane z NBP. Atrybuty tej tabeli są następujące:

- **rate\_id** - Unikatowy identyfikator wartości waluty dla danego dnia.
- **currency\_id** - Identyfikator waluty, odniesienie do **currency\_id** w tabeli **Currencies**.
- **date** - Data pobrania wartości dla danej waluty.
- **value** - Wartość kursu dla danej waluty.

Wartości w tabeli **ExchangeRates** są aktualizowane za każdym razem, gdy pobierane są nowe dane z API NBP.

### 3.3 Tabela Predictions

Tabela **Predictions** zawiera prognozowane wartości kursów walut. Atrybuty tej tabeli są następujące:

- **prediction\_id** - Unikalny identyfikator predykcji.
- **currency\_id** - Id waluty, dla której dokonano predykcji, odniesienie do **currency\_id** w tabeli **Currencies**.
- **date** - Data, dla której wykonano predykcję.
- **value** - Wartość predykcji kursu waluty.

Wartości w tabeli **Predictions** są aktualizowane za każdym razem, gdy wykonywana jest nowa predykcja. Baza danych została zaprojektowana w sposób umożliwiający efektywne przechowywanie, aktualizacje i dostęp do danych dotyczących kursów walut oraz prognoz. Struktura bazy danych zapewnia integralność danych i pozwala na szybkie przetwarzanie informacji niezbędnych do generowania dokładnych prognoz kursów walut.

## 4 Modele predykcyjne

Do predykcji kursów walut wykorzystano model XGBoost, który jest zaawansowanym algorytmem uczenia maszynowego. Jego zastosowanie pozwala na nowoczesne podejście do prognozowania wartości szeregów czasowych, wykorzystując nie tylko bezpośrednie wartości waluty, ale także inne cechy wydobyte z tych wartości w procesie inżynierii cech.

## 4.1 Cechy użyte w modelu

W procesie tworzenia modeli uwzględniono następujące cechy:

- **shift\_1** - Wartość kursu z dnia poprzedniego. Pozwala na uwzględnienie bezpośredniej zależności czasowej.
- **shift\_2** - Wartość kursu sprzed dwóch dni. Umożliwia analize krótkoterminowych trendów.
- **rolling\_2\_mean** - Średnia ruchoma z dwóch poprzednich dni. Pomaga w wygładzeniu danych i identyfikacji ogólnych trendów.
- **rolling\_3\_mean** - Średnia ruchoma z trzech poprzednich dni. Umożliwia analize nieco dłuższych trendów.
- **rolling\_std\_2** - Ruchome odchylenie standardowe z dwóch poprzednich dni. Pozwala na ocenę zmienności kursu w krótkim okresie.
- **rolling\_std\_3** - Ruchome odchylenie standardowe z trzech poprzednich dni. Umożliwia ocenę zmienności kursu w nieco dłuższym okresie.
- **value\_lag\_1** - Różnica między wartością z dnia poprzedniego a wartością sprzed dwóch dni. Wskazuje na zmiany w trendzie.
- **value\_lag\_2** - Różnica między wartością sprzed dwóch dni a wartością sprzed trzech dni. Pozwala na identyfikację wcześniejszych zmian w trendzie.
- **trend, sezonowość i rezydua** - Wartości wyliczone przy użyciu metody STL (Seasonal-Trend Decomposition using LOESS). Umożliwiają analize długoterminowych trendów, cyklicznych wzorców oraz nieregularnych fluktuacji.
- **expanding\_mean\_1** - Średnia ruchoma w rozszerzającym się oknie. Pomaga w identyfikacji długoterminowych trendów w danych historycznych.

## 4.2 Ocena modeli

Do oceny modeli wykorzystano metrykę RMAPE (Root Mean Absolute Percentage Error). Metryka ta pokazuje różnice w wartościach w sposób procentowy, co jest bardzo przydatne, gdy osoba korzystająca z dashboardu jest osobą nietechniczną nie mającą wprawy w korzystaniu z innych bardziej popularnych metryk. MAPE jest łatwa do interpretacji i zrozumienia, ponieważ można ją łatwo przedstawić jako współczynnik ryzyka. Użyto RMAPE, ponieważ pozwala sprowadzić wartości błędów dla poszczególnych modeli do jednego układu odniesienia i porównywać je ze sobą.

Wzór na RMAPE jest następujący:

$$\text{RMAPE} = \sqrt{\frac{1}{n} \sum_{t=1}^n \left( \frac{A_t - F_t}{A_t} \right)^2}$$

gdzie  $A_t$  oznacza rzeczywistą wartość, a  $F_t$  prognozowaną wartość.

## 4.3 Walidacja modeli

Do trenowania i testowania modeli zastosowano metodę *walk-forward validation*. Procedura ta polega na iteracyjnym budowaniu modelu na bazie minimalnej próbki danych, a następnie w każdym kolejnym kroku model jest testowany na jednej nowej obserwacji, po czym ta obserwacja jest dodawana do zbioru treningowego i model jest ponownie trenowany. Taka metoda walidacji pozwala na realistyczne odwzorowanie rzeczywistego procesu prognozowania, w którym system przewiduje jedną wartość, a następnie model jest aktualizowany na podstawie nowych danych.

Metoda *walk-forward validation* opiera się na podziale danych na sekwencyjne podzbiory, gdzie każdy krok walidacji polega na przesunięciu okna treningowego o jedną obserwację w przód, przetestowaniu

modelu na nowej obserwacji, a następnie dołączeniu tej obserwacji do zbioru treningowego. Wzory opisujące te metode sa następujące:  $Model_t = Train(Data_{1:t-1})$

$$Prediction_t = Model_t(Data_t)$$

$$Data_{1:t} = Data_{1:t-1} \cup Data_t$$

Wartość RMAPE używana w tabeli **Currencies** jest średnia dla wszystkich stworzonych modeli dla danej waluty. Modele zostały zapisane w dedykowanym folderze, co umożliwia ich łatwe zarządzanie i ponowne wykorzystanie.

## 5 Dashboard

Dashboards stanowią integralną część systemu, umożliwiając użytkownikom wizualizację i interpretację wyników prognoz. Poniżej przedstawiono szczegółowy opis poszczególnych elementów dashboardu, ich funkcje oraz cel umieszczenia ich na interfejsie użytkownika.

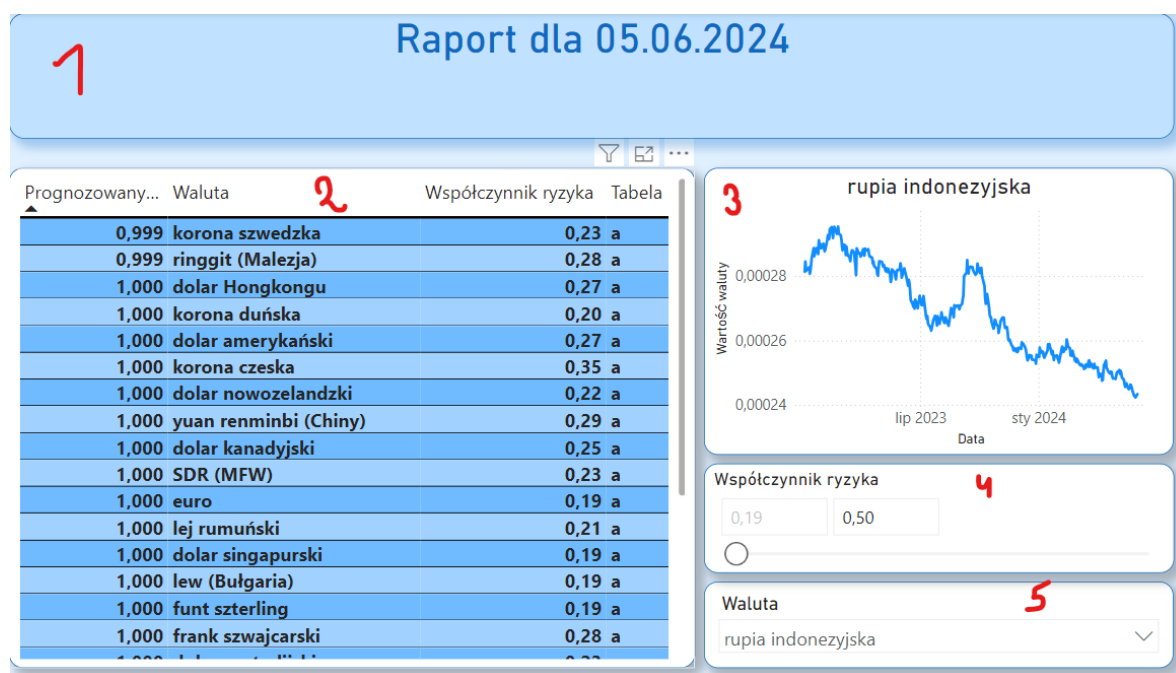


Figure 1: Wyjściowy wygląd dashboardu z zaznaczonymi sekcjami do omówienia

### 5.1 Tytuł dashboardu z data predykcji

Na szczycie dashboardu znajduje się tytuł z data predykcji, która jest automatycznie pobierana z danych i aktualizowana codziennie. Element ten informuje użytkownika o bieżącej dacie, dla której prognozy zostały wygenerowane, co jest kluczowe dla śledzenia i porównywania danych w czasie.

### 5.2 Tabela prognozowanych zysków

Centralnym elementem dashboardu jest tabela, w której zestawione są wartości prognozowane dla poszczególnych walut. Tabela ta zawiera cztery kolumny:

- **Prognozowany zysk** - Stosunek wartości prognozy do ostatniej dostępnej wartości dla danej waluty, co pozwala użytkownikowi na szybkie zidentyfikowanie potencjalnych zysków.
- **Nazwa waluty** - Pełna nazwa waluty, umożliwiająca łatwą identyfikację.

- **Współczynnik ryzyka (RMAPE)** - Wartość pokazująca ryzyko związane z niedokładnością modelu. Im wyższy współczynnik ryzyka, tym większe prawdopodobieństwo, że prognoza jest niedokładna lub błędna.
- **Źródło danych** - Informacja o tym, z której tabeli (a lub b) pochodzi dana waluta. Predykcje wartości z tabeli "a" są na ogół dokładniejsze.

Umieszczenie tej tabeli na dashboardzie ma na celu zapewnienie użytkownikowi przejrzystego i szybkiego dostępu do najważniejszych informacji dotyczących prognoz, co ułatwia podejmowanie decyzji inwestycyjnych.

### 5.3 Wykres zmian wartości waluty

Kolejnym istotnym elementem jest wykres przedstawiający zmiany wartości wybranej waluty na przestrzeni dostępnych danych historycznych. Wykres ten pozwala użytkownikowi na wizualną analizę trendów i wzorców w danych historycznych, co może wspomagać interpretację i ocenę prognoz.

### 5.4 Wybór waluty

Elementem numer 5 na dashboardzie jest interaktywny selektor waluty, który umożliwia użytkownikowi wybór waluty, dla której chce zobaczyć szczegółowe dane na wykresie. Funkcja ta zwiększa interaktywność dashboardu i pozwala na łatwe porównywanie różnych walut.

### 5.5 Filtr współczynnika ryzyka

Element numer 4 to filtr założony na tabele prognozowanych zysków, pozwalający na określenie maksymalnej wartości współczynnika ryzyka (RMAPE). Filtr ten umożliwia użytkownikowi ograniczenie wyświetlanych danych do tych walut, które spełniają określone kryteria ryzyka, co pomaga w skoncentrowaniu

### 5.6 Informacje o tabelach i współczynniku ryzyka

W drugiej zakładce dashboardu znajdują się szczegółowe informacje dotyczące tabel "a" i "b" oraz wyjaśnienie współczynnika ryzyka (RMAPE) i jego interpretacji. Ta sekcja ma na celu edukację użytkowników na temat źródeł danych i miar stosowanych w systemie, co jest kluczowe dla zrozumienia i właściwej interpretacji prezentowanych prognoz.

- **Tabela "a"** - Zawiera codziennie aktualizowane wartości walut, co zazwyczaj przekłada się na wyższą dokładność prognoz.
- **Tabela "b"** - Zawiera wartości walut udostępniane raz w tygodniu, co może wpływać na mniejszą dokładność prognoz.
- **Współczynnik ryzyka (RMAPE)** - Szczegółowe wyjaśnienie, czym jest RMAPE, jak jest obliczany oraz jak należy interpretować jego wartości.

**Współczynnik ryzyka:** jest to miara określająca ryzyko związane z nieprzewidywalnością modelu służącego do przewidywania wartości wybranej waluty. Miara ta jest oparta na **RMAPE** (Relatywny absolutny błąd procentowy) i oznacza średnie procentowe odchylenie wartości od średniej dla danej waluty, znormalizowane do wspólnego układu odniesienia w celu umożliwienia porównywania ze sobą wartości dla tej miary między walutami. Poniżej zaprezentowano szacunkową interpretację wartości RMAPE:

**RMAPE < 0.10** - bardzo wysoka dokładność predykcji

**RMAPE < 0.20** - wysoka dokładność predykcji

**0.2 < RMAPE < 0.50** - Rozsądna dokładność predykcji

**RMAPE > 0.5** - Niska dokładność predykcji

**Tabela a i b:** Wartości walut pobierane są cyklicznie z api **Narodowego Banku Polskiego**. Wartości z tabeli **a** są udostępniane codziennie, natomiast wartości z tabeli **b** są udostępniane w tygodniowych interwałach. W związku z czym, prognozy dla walut z tabeli **b** powinny cechować się niższą dokładnością, mimo że nie zawsze będzie to odwzorowane w wyższej wartości współczynnika ryzyka, ponieważ modele zostały wytrenowane na mniejszej ilości danych oraz zmiany w odstępach tygodniowych mogą być trudniejsze do zamodelowania niż zmiany codziennie.

Figure 2: Zakładka z informacjami o tabelach "a" i "b" oraz współczynnika ryzyka

Celem umieszczenia tych informacji jest zapewnienie pełnej transparentności systemu oraz umożliwienie użytkownikom świadomego korzystania z narzędzi analitycznych i prognoz.

## 6 Automatyzacja

W celu zapewnienia płynności działania i aktualizacji danych, system został w pełni zautomatyzowany. Automatyzacja obejmuje tworzenie i zarządzanie bazą danych, integrację z PowerBI oraz codzienne uruchamianie skryptów aktualizacyjnych.

### 6.1 Baza danych

Baza danych została stworzona przy użyciu SQLite3, co umożliwia łatwe zarządzanie i przechowywanie danych. Plik wyjściowy bazy danych został zapisany z rozszerzeniem .db, a następnie przy użyciu sterownika SQLite pobranego ze strony <http://www.ch-werner.de/sqliteodbc/> baza danych została zarejestrowana w administratorze źródeł danych ODBC w systemie Windows.

### 6.2 Integracja z PowerBI

W PowerBI wybrano opcję źródła danych ODBC, co pozwala na automatyczne ładowanie zmian wprowadzanych do bazy danych przez skrypt bezpośrednio do PowerBI. Dzięki temu użytkownicy mogą na bieżąco śledzić aktualne dane i prognozy bez potrzeby ręcznego odświeżania raportów.

### 6.3 Automatyzacja uruchamiania skryptów w Windows

Aby skrypty aktualizacyjne były uruchamiane codziennie, wykorzystano narzędzie *Windows Task Scheduler*. Poniżej przedstawiono kroki, jak skonfigurować zadanie w Task Scheduler:

1. Otwórz *Task Scheduler* w systemie Windows.
2. Kliknij *Create Basic Task* w panelu *Actions*.
3. Wprowadź nazwę i opis zadania, a następnie kliknij *Next*.
4. Wybierz, jak często zadanie ma być uruchamiane (np. *Daily*), i kliknij *Next*.

5. Ustaw czas rozpoczęcia i kliknij *Next*.
6. Wybierz akcje *Start a Program* i kliknij *Next*.
7. Wskaż skrypt do uruchomienia (np. plik .py lub .bat) i kliknij *Next*.
8. Przejrzyj ustawienia i kliknij *Finish* aby zakończyć konfigurację.

## 6.4 Automatyzacja uruchamiania skryptów w Linux

Na platformie Linux automatyzacja skryptów jest realizowana za pomocą narzędzia *Cron*. Poniżej przedstawiono, jak skonfigurować zadanie w Cron:

1. Otwórz terminal.
2. Edytuj plik crontab za pomocą polecenia *crontab -e*.
3. Dodaj nową linię z harmonogramem uruchamiania skryptu. Na przykład, aby uruchamiać skrypt codziennie o godzinie 6:00 rano, dodaj:

```
0 6 * * * /path/to/your/script.sh
```

4. Zapisz i zamknij plik crontab.
5. Sprawdź konfigurację crontab za pomocą polecenia *crontab -l*, aby upewnić się, że zadanie zostało dodane poprawnie.

Automatyzacja uruchamiania skryptów zarówno w Windows, jak i Linux, zapewnia codzienną aktualizację bazy danych i modeli predykcyjnych, co jest kluczowe dla utrzymania aktualności i dokładności prognoz prezentowanych na dashboardzie.