

Prognozowanie inflacji w Polsce

Przygotowali:

Szymon Budziak

Dominik Ciołczyk

Karol Wyrębkiewicz

Jakie były cele projektu?

- Cel główny

Zrozumienie i prognozowanie inflacji w Polsce, z uwzględnieniem wpływu różnych danych makroekonomicznych.

- Cel biznesowy

Stworzenie wiarygodnego modelu prognozującego inflację, który może być używany przez instytucje finansowe, inwestorów i decydentów politycznych do podejmowania świadomych decyzji.

- Cele poboczne

- Analiza historycznych danych dotyczących inflacji w Polsce
- Opracowanie metodologii do przewidywania inflacji
- Testowanie i walidacja modeli predykcyjnych oraz wybranie najlepszego
- Prezentacja wyników w przystępnej formie dla różnych grup odbiorców

Jakie dane zostały zgromadzone?

Dane zostały zgromadzone z kilku źródeł i zawierały:

- Miesięczne wskaźniki HICP dla Polski, Niemiec oraz strefy Euro
- Miesięczną inflację bazową w Polsce
- Miesięczne wskaźniki makroekonomiczne dla Polski
- Miesięczne wskaźniki CPI dla strefy Euro
- Miesięczne wskaźniki CPI dla Niemiec

Dlaczego w prognozowaniu inflacji w Polsce zostały uwzględnione dane z Niemiec oraz strefy Euro?

- Silne powiązania gospodarcze - Polska ma silne powiązania handlowe i gospodarcze z krajami ze strefy Euro, a ich głównym partnerem handlowym są Niemcy
- Wpływ na eksport i import - strefa Euro jest głównym rynkiem zbytu dla polskich towarów i usług
- Polityka monetarna - polityka monetarna prowadzona przez EBC ma pośredni wpływ na Polskę
- Przepływ kapitału - ruchy kapitałowe mogą wpływać na kursy walutowe oraz poziom inflacji

Z jakiej literatury skorzystaliśmy?

- Phillips Curve Inflation Forecasts - James H. Stock and Mark W. Watson
- Comments on “Phillips Curve Inflation Forecasts” by James H. Stock and Mark W. Watson - Adrian Rodney Pagan
- Prognoza inflacji w Polsce oparta na autoregresji - Antoni Wiliński, WSB w Gdańsku
- Prognozowanie inflacji w Polsce na podstawie modeli autoregresji wektorowej - Szymon Wójcik, Główny Urząd Statystyczny

Jak wyglądały dane ostateczne?

- Ostateczne dane zostały ograniczone od początku 2019 do końca 2023 - 4 lata.
- Wymiary danych po takim ograniczeniu, miały wymiary:
 - 77 kolumn
 - 59 wierszy
- Pomimo tego, część danych była wybrakowana - dokonanie wypełnienia brakujących danych.
- Standaryzacja danych - Standard Scaler
- Zastosowanie feature selection do wybrania najistotniejszych cech - 12:
 - PCA (Principal Component Analysis)
 - RFE (Recursive Feature Elimination) z regresją liniową
 - Feature importance przy pomocy Random Forest Regressor
- Początkowa analiza była dokonywana na pojedynczej kolumnie HICP, natomiast dalsze analizy dokonywane były na cechach wybranych przy pomocy feature selection.

Wykorzystane języki programowania i biblioteki

Języki programowania wykorzystane w projekcie, oraz biblioteki:

- Python
 - Numpy, Pandas
 - Matplotlib, Seaborn, Plotly
 - Darts, statsmodels
 - scikit-learn, Tensorflow
 - własna implementacja niektórych modeli
- R
- Statistica (nie do końca język, tylko program)



Jakie modele udało się sprawdzić?

Modele które udało się sprawdzić, to:

- **Naive Seasonal**

Model, który powtarza sezonowe dane. Prognozuje on przyszłe wartości serii czasowej zakładając, że będą one równe wartościom z poprzedniego sezonu.

- **ARIMA**

Statystyczny model szeregów czasowych używany do prognozowania przyszłych wartości na podstawie przeszłych obserwacji, uwzględniając autoregresję, różnicowanie w celu uczynienia danych stacjonarnymi oraz średnią ruchomą.

- **XGBoost**

Algorytm uczenia maszynowego oparty na gradient boosting, który jest szczególnie skuteczny w zadaniach klasyfikacji i regresji dzięki swojej zdolności do obsługi dużych zbiorów danych i skomplikowanych modeli. Posiada architekturę opartą na drzewach decyzyjnych.

- **Prophet**

Model do prognozowania szeregów czasowych (opracowany przez Facebooka), który jest łatwy w użyciu i skuteczny w uwzględnianiu sezonowości, trendów oraz świąt. Modeluje on dane czasowe jako kombinację trendu i sezonowości, stosując dopasowanie krzywych za pomocą regresji.

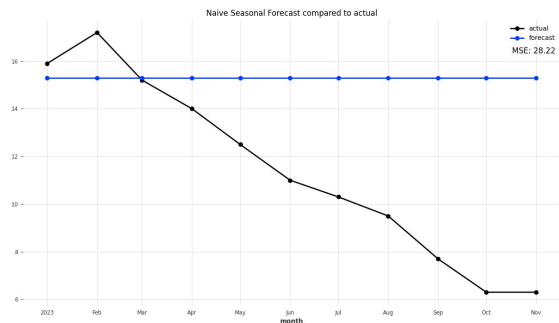
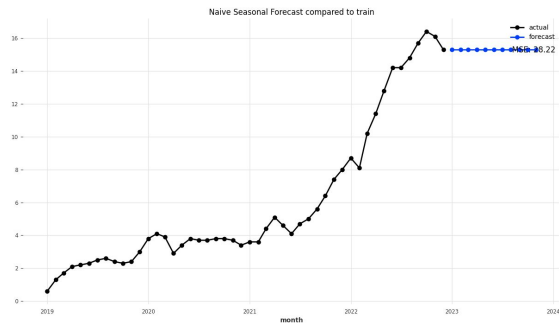
- **LSTM**

Rodzaj sieci neuronowej zaprojektowanej do przetwarzania i prognozowania danych sekwencyjnych, działającej poprzez utrzymywanie i manipulowanie długoterminową i krótkoterminową pamięcią za pomocą specjalnych komórek pamięci i mechanizmów bramek.

Jak radziły sobie poszczególne modele?

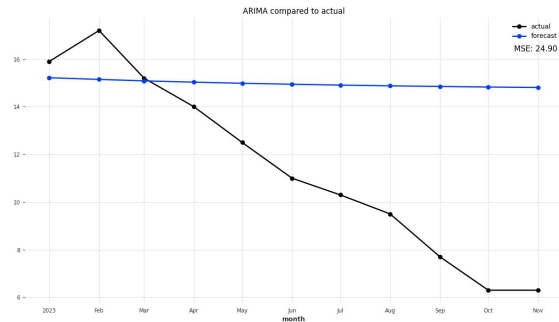
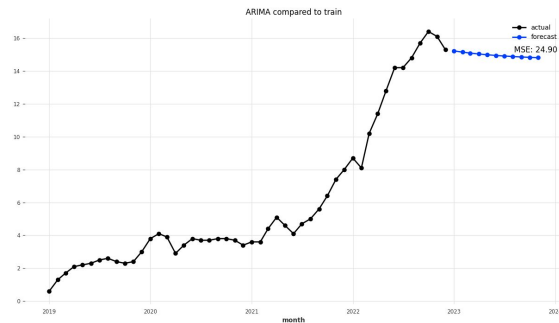
Naive Seasonal

MSE: 28.22

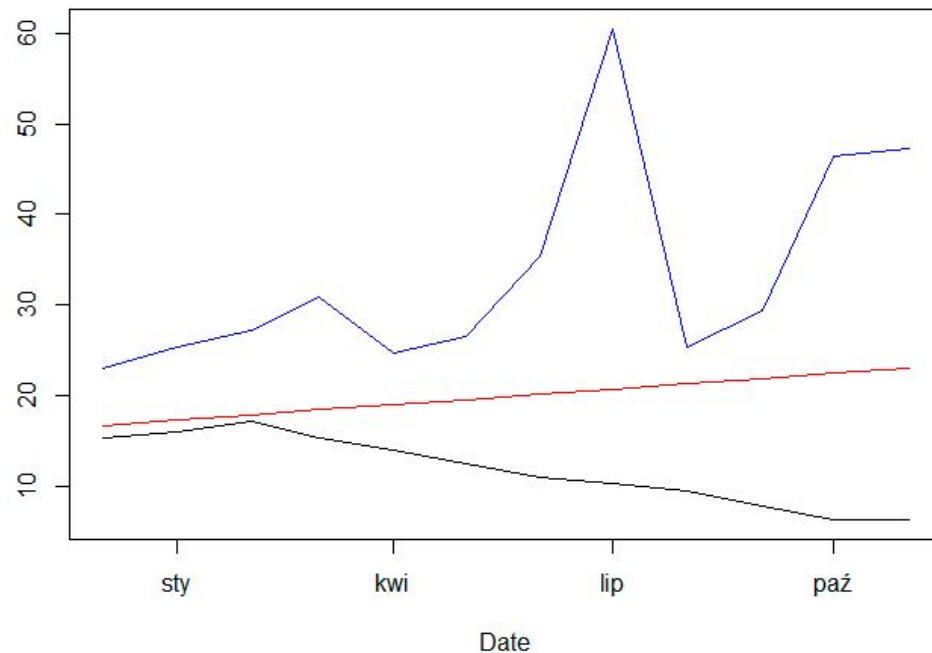
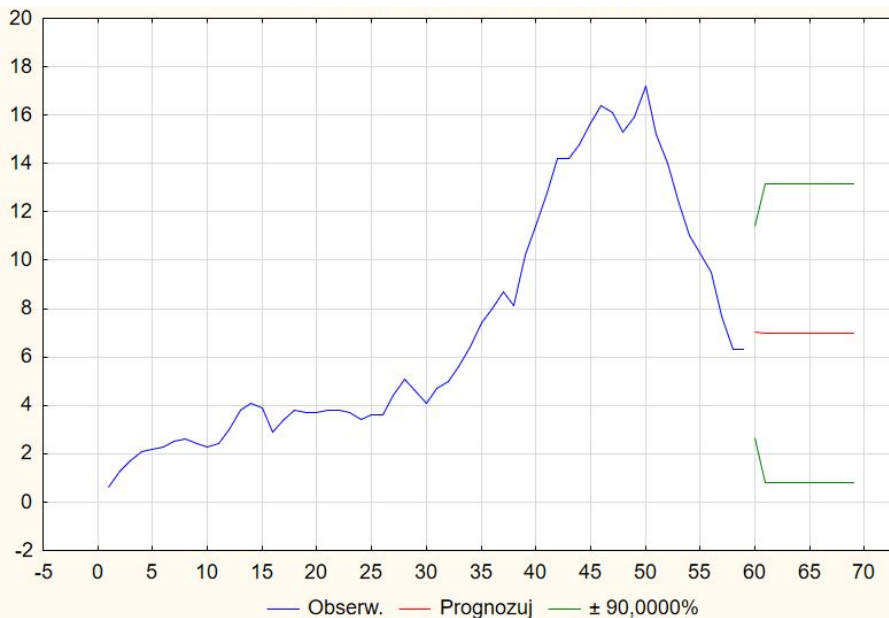


ARIMA

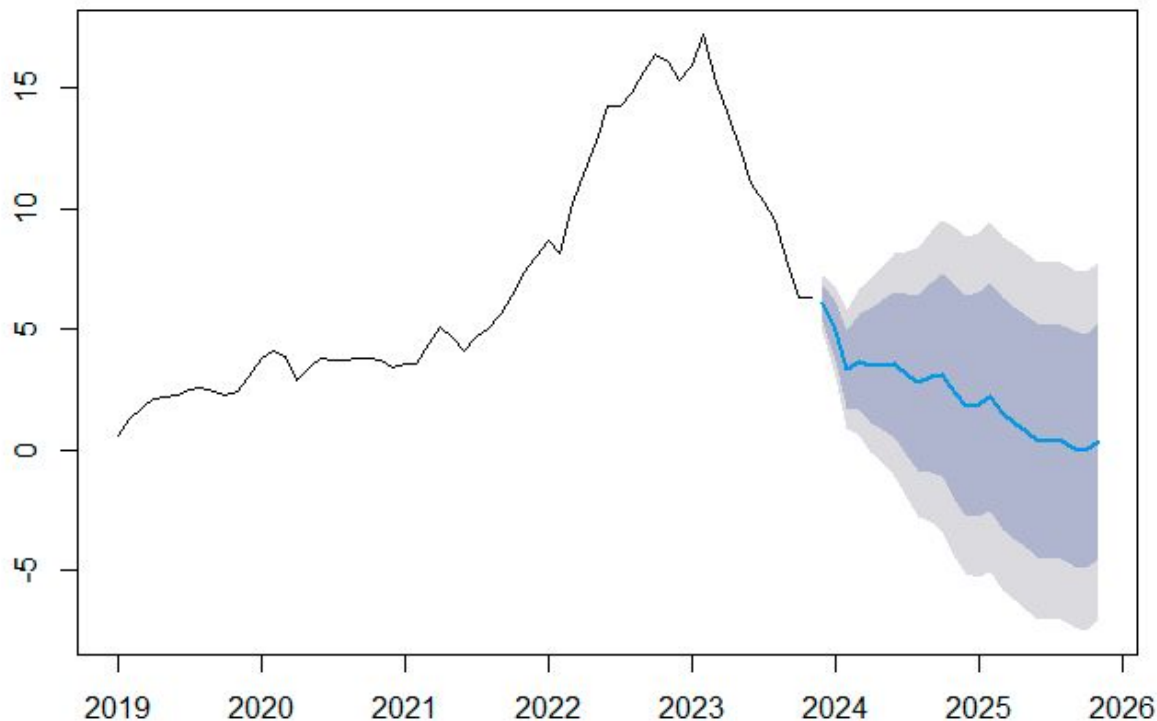
MSE: 24.90



Predykcja za pomocą modelu ARIMA



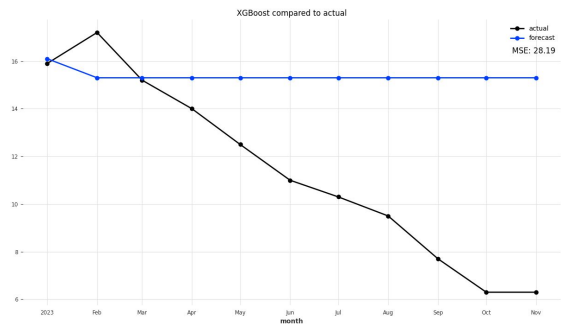
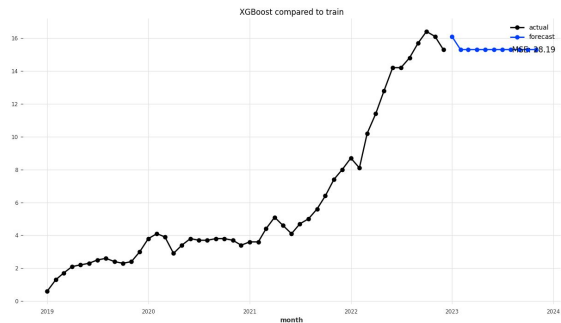
Predykcja po transformacji danych



Jak radziły sobie poszczególne modele?

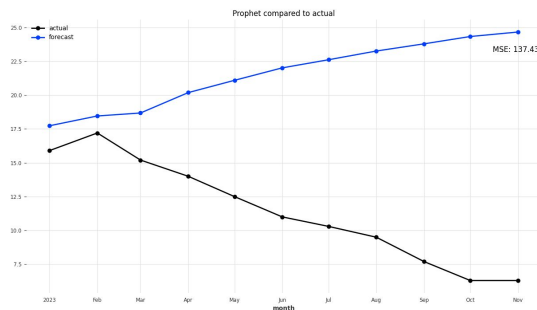
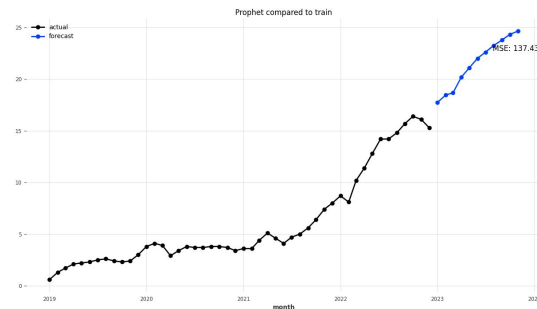
XGBoost

MSE: 28.19



Prophet

MSE: 137.43

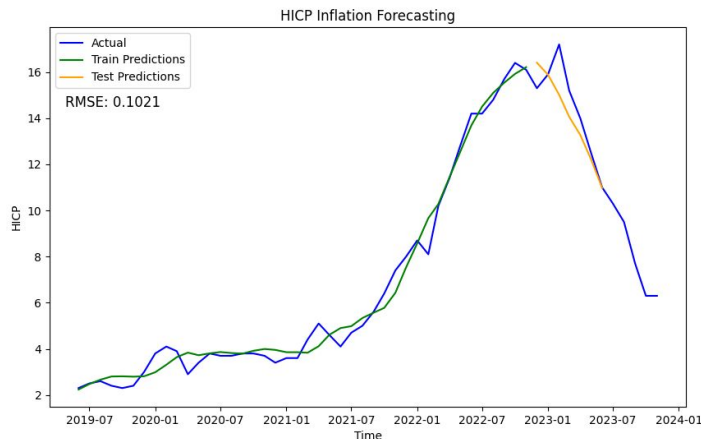


Jak radziły sobie poszczególne modele?

LSTM (pierwszy model)

RMSE: 0.1021

```
model = Sequential([  
    Input(shape=(seq_length, 1)),  
    LSTM(units=50, activation='relu'),  
    Dense(units=1)  
])
```



LSTM (rozbudowana architektura)

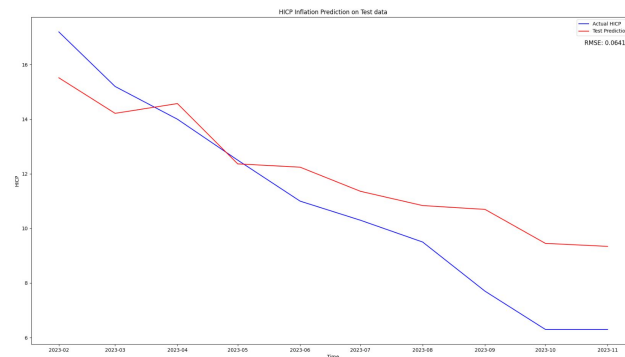
MSE: 0.0494

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 1, 50)	12,800
lstm_1 (LSTM)	(None, 50)	20,200
dropout (Dropout)	(None, 50)	0
dense (Dense)	(None, 1)	51

Total params: 33,051 (129.11 KB)

Trainable params: 33,051 (129.11 KB)

Non-trainable params: 0 (0.00 B)



Jakie mogą być dalsze badania?

- Opracowanie model hybrydowego (np. ARIMA-LSTM)
- Zastosowanie metod zespołowych
- Zastosowanie bardziej zaawansowanych architektur sieci neuronowych względem LSTM (np. GRU dla optymalizacji czasu predykcji albo CNN-LSTM dla lepszej zdolności do efektywnej ekstrakcji cech z danych sekwencyjnych)
- Zastosowanie wiedzy makroekonomicznej-Dynamiczne Stochastyczne Modele Równowagi Ogólnej (DSGE)
- Modele Bayesowskie z gęstością prawdopodobieństwa
- Rozszerzenie zbioru danych o dane pozyskane np. z użyciem Technik Przetwarzania Języka Naturalnego (NLP) lub oparte na prognozach ankietowych

Wnioski

- Predykcja inflacji jest bardzo trudna nawet przy uwzględnieniu wielu czynników makroekonomicznych oraz wydarzeń historycznych
- Z naszych analiz najlepiej radziły sobie modele LSTM oraz ARIMA, jednak trochę brakowało, aby ich wyniki mogły być uznane za wiarygodne, z perspektywy eksperta w tej dziedzinie
- Modele univariate, takie jak ARIMA, są proste do wdrożenia i często dają dobre wyniki w stabilnych warunkach ekonomicznych. Mogą być mniej skuteczne w zmieniających się warunkach, ponieważ opierają się tylko na danych historycznych dotyczących samej inflacji.
- Modele wielowymiarowe, takie jak VAR, mogą uwzględniać więcej zmiennych ekonomicznych, co może poprawić dokładność prognoz. Wymagają więcej danych i są bardziej złożone do wdrożenia, ale mogą lepiej uwzględniać złożone interakcje między różnymi czynnikami ekonomicznymi.

Bibliografia

- Dane inflacyjne
 - <https://data.ecb.europa.eu/search-results?searchTerm=Inflation>
 - <https://nbp.pl/statystyka-i-sprawozdawczosc/inflacja-bazowa/>
 - <https://stat.gov.pl/wskazniki-makroekonomiczne/>
 - <https://tradingeconomics.com/poland/consumer-price-index-cpi>
 - <https://www-genesis.destatis.de/genesis/online?operation=abruftabelleBearbeiten&levelindex=1&levelid=1709382443109&auswahloperation=abruftabelleAuspraegungAuswaehlen&auswahlverzeichnis=ordnungsstruktur&auswahlziel=werteabruf&code=61111-0002&auswahltext=&werteabruf=Value+retrieval#abreadcrumb>
 - <https://fred.stlouisfed.org/series/CPHPTT01EZM659N>
- Modele
 - https://unit8co.github.io/darts/generated_api/darts.models.forecasting.baselines.html
 - https://unit8co.github.io/darts/generated_api/darts.models.forecasting.arima.html
 - https://unit8co.github.io/darts/generated_api/darts.models.forecasting.prophet_model.html
 - https://unit8co.github.io/darts/generated_api/darts.models.forecasting.xgboost.html
 - <https://arxiv.org/abs/2105.06756>