

# Raport z realizacji projektu:

## Prognozowanie cen akcji na podstawie danych historycznych.

### Zarys projektu:

Jest to automatyczne narzędzie, gdzie użytkownik podaje nazwę instytucji a dostaje wykres prognozy na 14 dni. Głównym modelem prognozowania jest ARIMA.

### Opis zaimplementowanych funkcji:

#### 1. Działanie programu:

- 1) Pytamy użytkownika o podanie nazwy firmy, której akcje chce przewidywać (funkcja `ask_for_ticker`).
- 2) Pobieramy z alphavantage sugestie co do tego, co to za firma i wyświetlamy je użytkownikowi (dalej `ask_for_ticker`).
- 3) Użytkownik podaje którą pozycję z listy wybiera (dalej `ask_for_ticker`).
- 4) Pobieramy dane z alphavantage wspomnianej firmy (funkcja `download_data`).
- 5) Zaczynamy naukę modeli na danych z ostatnich 300 dni, lecz poza najnowszymi 14 dniami (funkcja `make_model_predictions` z argumentem `train`) i robimy prognozę na te 14 dni.
- 6) Sprawdzamy jak te modele poradziły sobie z prognozą na ostatnie 14 dni, porównując ich przewidywania z rzeczywistymi wartościami (funkcja `save_prediction_summary`).
- 7) Zaczynamy naukę 3 modeli na danych z ostatnich 300 dni (funkcja `make_model_predictions` z argumentem `data['close']`) i robimy prognozę na przyszłe 14 dni.
- 8) Zapisujemy i wyświetlamy wyniki modeli (funkcja `save_forecast`).

9) Rysujemy 2 wykresy, następnie je zapisujemy  
(funkcja `make_plot`), pierwszy przedstawia przewidywania na ostatnie 14 dni, a drugi na przyszłe 14 dni.

## 2. Pliki wynikowe programu:

- lista sugestii alphavantage (`search_list.csv`).
- dane pobrane z alphavantage (`'ticker firmy'.csv`).
- RMSE i MAPE przewidywań 3 modeli dla ostatnich 14 dni (`prediction_summary.csv`).
- prognoza 3 modeli dla przyszłych 14 dni (`forecast.csv`).
- wykresy przewidywań ostatnich i przyszłych 14 dni (`prediction_forecast_plot.png`).

## 3. Używane modele:

1) Model krótkoterminowego trendu:

- Liczymy średnią zwrotów z ostatnich 20 dni

$$\text{zwrot}[t] = \frac{\text{wartość}[t]}{\text{wartość}[t - 1]} - 1$$

- Zakładamy, że kurs będzie rósł codziennie o średni zwrot, czyli:

$$\text{prognoza}[t + n] = \text{wartość}[t] * (1 + \text{średni zwrot})^n$$

- Liczymy średnią zwrotów z ostatnich 100 dni.
- Zakładamy, że kurs będzie rósł codziennie o średni zwrot. 3)  
Model ARIMA (Autoregressive Integrated Moving Average):
- Jest to zaawansowany model do prognozowania szeregów czasowych.
- „AR” odnosi się do tego od jak wielu poprzednich wartości zależy wartość aktualna (przyjmujemy w modelu 2).
- Odnosi się do tego, że zamiast bezpośrednio prognozować wartość, prognozuje różnice wartości, stopień odnosi się do tego jak odległych wartości różnicę rozważamy (przyjmujemy 1).

- Powinno odnosić się do tego, jak bardzo losowość aktualnej wartości zależy od losowości wartości poprzednich (przyjmujemy w modelu 1, czyli że aktualna losowość zależy tylko od poprzedniej).

## Użyte Narzędzia, Biblioteki:

### Narzędzia i biblioteki:

1. **Pandas** → to biblioteka do analizy danych w Pythonie. Umożliwia manipulację danymi w formie tabelarycznej, podobnie jak arkusze kalkulacyjne. Główne struktury danych to:
  - DataFrame: dwuwymiarowa tabela z wierszami i kolumnami.
  - Series: jednowymiarowa tabela, podobna do listy.
2. **Numpy** → inaczej mówiąc Numerical Python, jest to biblioteka do obliczeń numerycznych. Dostarcza:
  - ndarray: wielowymiarowe tablice o stałym typie danych.
  - Wiele funkcji matematycznych i statystycznych, które są szybkie i efektywne.
3. **Matplotlib** → to biblioteka do tworzenia wykresów w Pythonie. Najpopularniejszy moduł to pyplot, który pozwala na szybkie tworzenie wykresów. Umożliwia tworzenie różnorodnych wykresów, takich jak:
  - Liniowe
  - Słupkowe
  - Histogramy
  - Wykresy kołowe
4. **Statsmodels** → to biblioteka do modelowania statystycznego. Jest przydatna w bardziej zaawansowanej analizie statystycznej i ekonometrycznej. Każda z tych bibliotek jest niezwykle użyteczna w analizie danych, a często są one używane razem, aby uzyskać pełniejszy obraz danych i przeprowadzić kompleksową analizę. Umożliwia:
  - Analizę regresji (liniowej, logistycznej itp.)
  - Testy statystyczne
  - Analizę szeregów czasowych
  - Estymację modeli ekonometrycznych

## Napotkane problemy:

Największym napotkanym problemem w projekcie było szukanie odpowiednich parametrów dla modelu ARIMA, tak aby uzyskane wyniki miały sens. Polegało to na starannym doborze parametrów modelu, takich jak autoregresja, różnicowanie i średnia ruchoma, aby wyniki były wiarygodne i miały znaczenie w kontekście analizowanego problemu. Dobrze dobrane parametry mogły znacznie poprawić jakość prognoz i interpretację wyników.

## Możliwe rozszerzenie funkcjonalności:

### 1. Dodanie wskaźników technicznych:

Wprowadzenie analizy technicznej, na przykład poprzez wskaźniki jak RSI (Relative Strength Index) czy MACD (Moving Average Convergence Divergence), może poprawić jakość prognoz poprzez uwzględnienie dodatkowych czynników rynkowych.

### 2. Zastosowanie metod uczenia maszynowego:

Rozszerzenie programu o modele uczenia maszynowego, takie jak drzewa decyzyjne, lasy losowe czy sieci neuronowe, mogłoby zwiększyć dokładność prognoz poprzez uwzględnienie bardziej złożonych wzorców w danych.

### 3. Interaktywna wizualizacja:

Dodanie interaktywnych wykresów przy użyciu bibliotek takich jak Plotly lub Bokeh. Pozwoli to użytkownikom lepiej analizować wyniki i dostosowywać zakresy czasowe czy różne parametry wizualizacji.

## Źródła informacji:

1. **Źródło danych:** <https://www.alphavantage.co/documentation/>
2. **Podstawy uczenia maszynowego:** <https://www.kaggle.com/learn/intro-to-machine-learning>
3. **Inspiracja:** <https://neptune.ai/blog/predicting-stock-prices-using-machine-learning>
4. **ARIMA model:**  
<https://www.investopedia.com/terms/a/autoregressive-integrated-moving-averagearima.asp#:~:text=An%20autoregressive%20integrated%20moving%20average%2C%20or%20ARIMA%2C%20is,it%20predicts%20future%20values%20based%20on%20past%20values>

## **Wnioski:**

### 1. Prognozowanie z wykorzystaniem ARIMA:

Model ARIMA okazał się skuteczny w prognozowaniu krótkoterminowych zmian cen akcji, wykorzystując dane historyczne.

### 2. Porównanie modeli:

Dzięki porównaniu różnych modeli (krótkoterminowy, długoterminowy oraz ARIMA), możliwe było określenie, który z nich najlepiej przewiduje ceny akcji w różnych kontekstach.

### 3. Automatyzacja procesu:

Program automatyzuje proces prognozowania, co ułatwia użytkownikom uzyskiwanie prognoz bez głębokiej znajomości analizy danych.

### 4. Wykorzystanie danych z Alpha Vantage:

Integracja z API Alpha Vantage pozwala na bieżąco aktualizować dane oraz zwiększa dostępność różnych informacji o spółkach.

---

## **Autorzy:**

Dominik Górski 52679

Julia Flaszyńska 54149

Krystian Galus 52676

Paulina Izienicka 52690