**1. Analiza Sentymentu na Danych z Twittera**

**Wprowadzenie**

W ramach analizy sentymentu danych pobieranych z platformy Twitter, zdecydowaliśmy się skorzystać z biblioteki NLTK (Natural Language Toolkit). W celu oceny sentymentu poszczególnych tweetów użyjemy modułu SentimentIntensityAnalyzer dostępnego w ramach tej biblioteki.

**Metoda Analizy**

Każdy tweet zostanie przetworzony pod kątem sumarycznego sentymentu związanej z nim opinii na temat określonej kryptowaluty. W trakcie analizy, każdy tweet zostanie oceniony, a następnie przypisana zostanie ocena sentymentu, którą uzyskamy dzięki wykorzystaniu SentimentIntensityAnalyzer.

**Wygenerowanie Pliku Wynikowego**

W wyniku działania programu, zostanie wygenerowany plik tekstowy zawierający oceny sentymentu dla każdego tweetu. Ponadto, program obliczy średni sentyment dla całej bazy tweetów dotyczących jednej kryptowaluty.

**Proces Weryfikacji Programu**

Plik tekstowy uzyskany w wyniku analizy posłuży jako narzędzie weryfikacyjne poprawności funkcjonowania programu. Analiza wynikowego sentymentu pozwoli nam również na zbadanie ewentualnych zależności między zmianami cen rynkowych danej kryptowaluty a opinią publiczną wyrażoną w tweetach.

**Korzystanie z NLTK**

Do implementacji analizy sentymentu wykorzystaliśmy materiały dostępne pod adresem: <https://www.nltk.org/howto/sentiment.html>. Ta dokumentacja była pomocna w zrozumieniu i skutecznym wykorzystaniu funkcji biblioteki NLTK.

**2. Przykłady schematów Pump and Dump**

**Czym jest Pump and Dump**

Jest to specyficzny schemat manipulacji rynku, który obejmuje szybki wzrost cen, a następnie – po uzyskaniu zamierzonego zysku – szybką sprzedaż. Schematy te zawierają dwie grupy ludzi: pierwsza z nich to osoby wewnętrzne, które kupują wystarczającą ilość tokenów po bardzo niskiej cenie. Następnie promują projekt za pomocą kanałów komunikacyjnych (Twitter, Discord, Reddit) i przekonują drugą grupę do kupna danej kryptowaluty. W ten sposób rośnie hype, a wraz z nim – cena. Wtajemniczone osoby w tym momencie sprzedają zakupione wcześniej tokeny, a nieświadomi niczego użytkownicy je skupują po zawyżonej cenie. Następnie kurs tej kryptowaluty spada często jeszcze szybciej niż wzrósł

**E-Coin**

6 lutego 2018 mało znana kryptowaluta wzrosła o 4742%. Startupowi kryptowalutowemu wystarczył jeden dzień, aby przebić się przez TOP-100 z ostatnich miejsc w TOP-500 pod względem kapitalizacji, a następnie wejść do TOP-20.

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Opis wygenerowany automatycznie

**Viking Swap**

VikingsChain, Viking Swap i Space Vikings. Na początku listopada 2021 r. monety te wzrosły odpowiednio o 350%, 3700% i 600% po tym, jak Musk opublikował kilka tweetów na temat Księżyca i Wikingów. Elon zasugerował, że Wikingowie jako pierwsi wylądowali na Księżycu. A ktoś próbował nawet narysować paralelę do tego żartu z klasycznym zdaniem optymistów „Na Księżyc”.

Obraz zawierający tekst, zrzut ekranu

Opis wygenerowany automatycznie

**FuturoCoin**

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Opis wygenerowany automatycznie

Brak aktywności przed skokiem 13.02.2020 r. Wzrost ceny o ponad 8000%, a następnie gwałtowny spadek.

**Podsumowanie**

Oznaki Pump and Dump:

* gwałtowny wzrost i pojawienie się dużych zleceń kupna bez istotnego fundamentalnego powodu,
* pojawienie się rekomendacji zakupu kryptowaluty na czatach, forach, portalach społecznościowych,
* notowania rosną tylko na jednej stronie, podczas gdy na innych giełdach kryptowalut cena się nie zmienia.

**3. Istniejące już projekty wiążące analizę sentymentu z danymi rynkowymi (Forex, giełda, krypto, itp)**

**Projekt1: Predicting-Forex-with-Sentiment-Analysis**

<https://github.com/MagdaleneHo/Predicting-Forex-with-Sentiment-Analysis?fbclid=IwAR18wO3O4jYKnKPM3eFxNCKQ3vpGbxYuscLG6F--hTDlM_j1qx69PFyFnRQ>

Projekt "Predicting Forex with Sentiment Analysis" bada wpływ analizy sentymentu na prognozowanie ruchów na rynku Forex, koncentrując się na parach walutowych EUR/USD i GBP/USD.

Celem projektu jest przewidywanie ruchu na rynku Forex na podstawie sentymentu.

Kluczowe elementy projektu:

1. **Trzy główne części projektu:**

* Pobieranie nagłówków wiadomości z różnych źródeł (CNBC, NY Times, Reuters, Guardian)
* Ocenianie sentymentu nagłówków wiadomości na poziomie słów i zdań
* Używanie binarnej oceny codziennego ruchu na rynku Forex (EUR-USD/GBP-USD) jako zmiennej docelowej, gdzie sentyment jest zmienną niezależną

1. **Dane:**
   * Forex Data: Pobrane z otwartego źródła (2019-2020, interwał dzienny).
   * Źródła wiadomości: Reuters, CNBC, The Guardian, NY Times.
2. **Analiza sentymentu:**
   * **Analiza na poziomie słów:** wykorzystuje pre-processing, w tym tokenizację, normalizację, filtrowanie słów i obsługę negacji. Ocenia sentyment, uwzględniając znaczenie konkretnych słów.
   * **Analiza na poziomie zdań**: Używa pakietu Sentimentr z R do obliczenia sentymentu tekstu na poziomie zdania. Ocena sentymentu odbywa się dla całego zdania, bez analizy poszczególnych słów. Wyniki analizy sentymentu są następnie łączone z danymi rynku Forex dla przeprowadzenia analizy prognozowej.
3. **Modele predykcyjne:**
   * Algorytmy uczenia maszynowego: Support Vector Machine, Logistic Regression, Decision Tree, KNN, Gaussian Naïve Bayes.
   * Trenowanie i walidacja z zastosowaniem zasady Pareto (80/20).

**Projekt 2: Stock-Prediction**

<https://github.com/alisonmitchell/Stock-Prediction?fbclid=IwAR35OqHQivK1fVSa8vhTFzx_oUUWn_pT9QvvGo5AgoOQAJcAkWp-Thtlcus>

Projekt koncentruje się na opracowaniu strategii przewidywania rynku akcji, łącząc analizę czasową, techniczną i sentymenu. Celem jest trafne określenie odpowiednich momentów kupna i sprzedaży.

1. **Analizy:**
   * **Eksploracyjna analiza danych:** Dane giełdowe z yfinance obejmują ceny (Open, High, Low, Close), liczba akcji w obrocie (Volume) i cenę zamknięcia (Adjusted Close). Skupiając się głównie na Adjusted Close, uwzględnia się działania korporacyjne dla precyzyjnej analizy zwrotów. Przekształcenie danych umożliwia obliczenia zwrotów, a macierze kowariancji/korelacji pomagają w wyborze.
   * **Analiza techniczna:** Analiza techniczna to wykorzystanie wykresów i wskaźników technicznych do identyfikacji sygnałów handlowych i wzorców cenowych. Badane są różne strategie techniczne, korzystając z najpopularniejszych wskaźników: Moving Averages, Moving Average Convergence Divergence (MACD), Stochastic Oscillator, Relative Strength Index (RSI), Money Flow Index (MFI), Rate of Change (ROC), Bollinger Bands, and On-Balance Volume (OBV).
   * **Time Series**: To zestaw punktów danych ułożonych w czasie, istotny w prognozowaniu trendów na rynku akcji. Aby modelować, czasami stosuje się transformacje, sprawdzając stacjonarność za pomocą testu ADF.
   * **Analiza sentymentu**: artykuły będą pobierane z Investing.com z użyciem Selenium i Beautiful Soup
2. **Modele:**
   * Użycie różnych modeli uczenia maszynowego, w tym Naive Bayes, Gradient Boosting, a także zaawansowane sieci neuronowe (RNN, LSTM, GRU).
3. **Algorytmy i techniki:**
   * Prognozowanie rynku akcji będzie realizowane jako problem regresji do przewidywania cen 'n' dni w przyszłość i problem klasyfikacji do przewidywania kierunku zmian cen. Wykorzystane będą różne modele, takie jak Gaussian Naive Bayes i Gradient Boosting, z użyciem technik TimeSeriesSplit do uniknięcia uprzedzeń w danych treningowych.
4. **Biblioteki i Źródła Danych:**
   * Wykorzystanie popularnych bibliotek Python (Numpy, Pandas, Keras, TensorFlow, itp.) oraz danych z Yahoo! Finance i Investing.com.

**Projekt 3: financial\_sentiment\_analysis**

<https://github.com/cerenkasap/financial_sentiment_analysis>

(kolejny przykład już bez opisu)