Politechnika Warszawska

WYDZIAŁ MATEMATYKI
I NAUK INFORMACYJNYCH

Praca dyplomowa inżynierska

na kierunku Inżynieria i Analiza Danych

System Alertowania Spadków Cen Akcji na Giełdzie przy Użyciu Modelu Językowego oraz Analizy Historycznych Cen

Wojciech Kosiuk

Numer albumu 123456

Szymon Matuszewski

Numer albumu 313435

Michał Mazuryk

Numer albumu 313440

promotor

dr Robert Małysz

WARSZAWA 2023

Spis treści

1 Abstrakt

Ten dokument zawiera opis projektu będącego częścią pracy inżynierskiej studentów wydziału Matematyki i Nauk Informacyjnych Politechniki Warszawskiej: Wojciecha Kosiuka, Szymona Matuszewskiego oraz Michała Mazuryka. Celem projektu jest opracowanie systemu służącego do przewidywania znaczących spadków cen na giełdzie z odpowiednim wyprzedzeniem czasowym przy użyciu modelu językowego analizującego dostępne w Internecie wiadomości z rynku finansowego oraz analizy przeszłych cen akcji bazując na szeregach czasowych. Wynikiem działań ma być model zwracający prawdopodobieństwo potencjalnego spadku ceny akcji w konkretnej perspektywie czasowej.
Potencjalnymi użytkownikami systemu są inwestorzy giełdowi oraz osoby interesujące się tematyką inwestycyjną, dla których ww. system może stanowić wyraźną pomoc przy lepszym prognozowaniu zachowań na giełdzie.
Poniższa dokumentacja zawiera historię zmian, słownik pojęć, specyfikację systemu (w tym wymagania funkcjonalne oraz niefunkcjonalne), harmonogram prac, analizę ryzyka, oraz bibliografię.

Uwagi:

- **Okres przewidywania spadku cen akcji ma zostać ustalony podczas wstępnej analizy danych przez członków zespołu. Może być to kilka okresów (np. 1 dzień, 1 tydzień, 1 miesiąc).**

1.1 Historia zmian

Tabela 1: Historia zmian dokumentu.

Data	Autor	Opis	Wersja
17.10.2023	Wojciech Kosiuk Szymon Matuszewski Michał Mazuryk	Pierwsza wersja opisu architektury projektu	1.0
05.11.2023	Szymon Matuszewski	Konwersja dokumentu do Latex'u	1.1
07.11.2023	Wojciech Kosiuk Szymon Matuszewski Michał Mazuryk	Druga wersja zawierająca propozycje rozwiązania problemu badawczego	2.0

2 Słownik

- **Wymagania funkcjonalne (FR)** - wymagania definiujące funkcję systemu, która opisuje relację między wejściem a wyjściem wynikającym z jego działania. Określone są one z uwzględnieniem projektu systemu.

- **Wymagania niefunkcjonalne (NFR)** - wymagania, dzięki którym możliwa będzie ocena jakości systemu. Określone są one z uwzględnieniem architektury systemu.

- **Akcja** – papier wartościowy potwierdzający prawa o charakterze majątkowym i niemajątkowym, które posiada akcjonariusz względem spółki względem spółki akcyjnej lub komandytowo-akcyjnej.

- **Giełda** - instytucja publiczna, w której kupcy i pośrednicy (maklerzy) kupują lub sprzedają papiery wartościowe i niektóre towary masowe.

- **Uczenie Maszynowe** - (ang. Machine Learning) - podzbiór sztucznej inteligencji (ang. AI) poświęcony algorytmom wykorzystującym proces uczenia się na polepszanie własnych parametrów w celu jak najlepszej predykcji.
- **Model numeryczny** – model uczenia maszynowego, który jako dane wejściowe przyjmuje dane numeryczne.
- **Model językowy (NLP)** - model uczenia maszynowego, który jest zdolny do przetwarzania ciągu danych tekstowych.
- **LLM** - "Large Language Model"w tłumaczeniu duży model językowy. Jest to rodzaj zaawansowanego modelu uczenia maszynowego, który został przeszkolony na dużych zbiorach danych tekstowych i jest zdolny do generowania tekstu, tłumaczenia języków i innych zadań związanych z przetwarzaniem języka naturalnego.
- **Szereg czasowy** – realizacja procesu stochastycznego, której dziedziną jest czas.
- **Tag giełdowy** - zwany także symbolem giełdowym, to unikatowy skrót symbol używany do jednoznacznego zidentyfikowania konkretnej spółki giełdowej
- **FC (w sieciach neuronowych)** - "Fully Connected layer,"oznacza warstwę w pełni połączoną. W warstwie tej każdy neuron jest połączony z każdym neuronem z poprzedniej i następnej warstwy, co oznacza pełne połączenie między nimi.
- **LSTM (w sieciach neuronowych)** - "Long Short-Term Memory,"jest rodzajem warstwy w sieciach neuronowych, szczególnie w rekurencyjnych sieciach neuronowych, która jest zaprojektowana do obsługi sekwencji danych, takich jak szeregi czasowe czy tekst.
- **Baza danych** – miejsce składowania danych na komputerze.
- **Aplikacja przeglądarkowa** - program komputerowy pracujący na serwerze I komunikujący się poprzez sieć komputerową z hostem użytkownika komputera, wykorzystując przy tym przeglądarkę internetową, która jest interaktywnym klientem aplikacji internetowej.
- **Framework** – struktura komponentu programu komputerowego.
- **Flask** – framework służący do projektowania aplikacji przeglądarkowych napisany w języku Python.
- **API** – z angielskiego Application Programming Interface – zestaw reguł i protokołów określający jak poszczególne komponenty programu powinny komunikować się ze sobą.

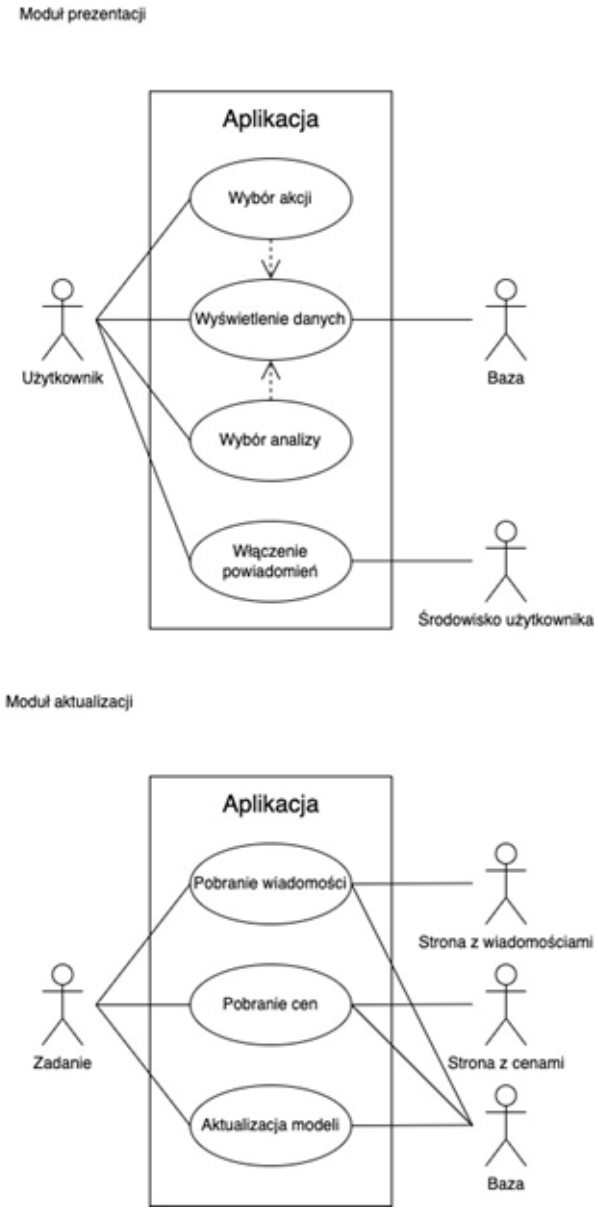
3 Specyfikacja

3.1 Streszczenie

System jest przeznaczony do ostrzegania użytkownika o potencjalnych spadkach cen akcji na giełdzie, które mogą zostać wykorzystane do polepszenia jego inwestycji. Udostępnia interfejs graficzny z podstawowymi informacjami dotyczącymi najnowszych wiadomości

nt. kryptowaluty lub spółki akcyjnej oraz wspomagającymi wykresami analitycznymi. Dodatkowo, użytkownik będzie miał dostęp do kluczowych statystyk warunkujących prawdopodobieństwo alertu. Końcowym beneficjentem systemu ma być potencjalny inwestor zainteresowany daną spółką akcyjną/kryptowalutą.

3.2 Wymagania funkcjonalne



Rysunek 1: Opis przypadków użycia aplikacji.

Tabela 2: Opis wymagań funkcjonalnych aplikacji.

ID	**Profil**	**Nazwa**	**Opis**	**Odpowiedź systemu**
User	Użytkownik	Wyświetlenie danych	Utworzenie najnowszych wykresów o wybranej przez użytkownika akcji/walucie danych akcji	Wyświetlenie informacji, wykresów pozwalających analizować przyszłe wartości akcji
		Wybór akcji	Możliwość wybrania z listy akcji, waluty interesującej użytkownika	Wyświetlenie danych w oparciu o wybraną akcję
		Włączenie powiadomień	Dodanie alertu do systemu użytkownika, informującego go o wykryciu przyszłego spadku dla wybranej przez niego akcji	Uruchomienie procesu na systemie użytkownika
		Wybór analizy	Wybór prezentowanych informacji o akcjach, w ujęciu dziennym, tygodniowym lub miesięcznym	Załadowanie odpowiedniego modelu do wybranego podejścia i wyświetlenie danych
Job	Zadanie	Pobranie wiadomości	Pobranie wiadomości dotyczących konkretnych akcji z wybranych stron informacyjnych	Pobranie poprzez api strony nowych wiadomości i zapis ich do bazy
		Pobranie cen	Pobranie najnowszych cen akcji/walut	Pobranie poprzez api strony najnowych cen akcji/walut I zapis ich do bazy
		Aktualizacja modeli	Rekalibracja modeli na podstawie nowo pobranych informacji	Wyliczenie cech i ich zapis do bazy oraz uruchomienie skryptów aktualizujących modele

3.3 Wymagania niefunkcjonalne

Tabela 3: Opis wymagań niefunkcjonalnych aplikacji.

Wymagania	**Numer wymagania**	**Opis**
Utility	1	Interfejs aplikacji powinien być stworzony w sposób intuicyjny, tak aby osoba nietechniczna była w stanie w łatwy sposób z niej korzystać
Reliability	2	Aplikacja powinna być dostępna dla użytkowników przeglądarki internetowej
	3	System powinien być dostępny przez 24 godziny 7 dni w tygodniu, żeby inwestor mógł na bieżąco otrzymywać alerty o potencjalnych spadkach cen
Performance	4	System powinien być odpowiednio przetestowany, aby zminimalizować ryzyko jego niepoprawnego działania
	5	System powinien dostarczać godzinne oraz dzienne alerty odnośnie wiadomości finansowych i cen akcji/kryptowaluty
Maintenance	6	System powinien być skalowalny w celu łatwej obsługi większej ilości akcji oraz dłuższej historii cen
	7	Dokumentacja techniczna powinna być szczegółowo opisana w celu łatwego rozwiązywania prostych problemów i obsługi aplikacji
	8	System powinien być zgodny z obowiązującymi przepisami dotyczącymi przechowywania i przetwarzania danych oraz regulacjami rynku finansowego

4 Harmonogram prac

Tabela 4: Harmonogram prac przy tworzeniu systemu.

Zadanie	**Data rozpoczęcia**	**Data zakończenia**
Kolekcja danych cen oraz wiadomości	02.10.2023	09.10.2023
Utworzenie szkicu aplikacji przeglądarkowej	02.10.2023	30.10.2023
Wstępna analiza danych	09.10.2023	05.11.2023
Kamień milowy I	**02.10.2023**	**05.11.2023**

Transformacja wiadmości przy użyciu modeli NLP na dane numeryczne	17.10.2023	19.11.2023
Kompozycja danych	19.11.2023	26.11.2023
Rozpozanie kluczowych predyktorów	26.11.2023	03.12.2023
Zaprojektowanie bazy danych	26.11.2023	03.12.2023
Utworzenie funkcji do modelowania (template)	03.12.2023	10.12.2023
Kamień milowy II	**17.10.2023**	**10.12.2023**
Modelowanie dla różnych okresów czasu oraz różnych spółem/kryptowalut	10.12.2023	24.12.2023
Wprowadzenie funkcjonalności wysyłania powiadomień w aplikacji	17.12.2023	24.12.2023
Kamień milowy III	**10.12.2023**	**24.12.2023**
Utworzenie łączników pomiędzy danymi, modelami i aplikacją	24.12.2023	03.01.2024
Utworzenie modułu wizualnego prezentacji wyników	24.12.2023	03.01.2024
Integracja komponentów w aplikacji	03.01.2024	10.01.2024
Opracowanie wizualnego wyglądu aplikacji	03.01.2024	10.01.2024
Kamień milowy IV	**24.12.2023**	**10.01.2024**
Testy funkcjonalności	10.01.2024	17.01.2024
Definiowanie wniosków	10.01.2024	17.01.2024
Kamień milowy V	**10.01.2024**	**17.01.2024**

5 Analiza ryzyka

Tabela 5: Opis zagrożeń wewnętrznych przy tworzeniu systemu.

SWOT	Zagrożenia	Szanse
Wewnętrzne	1. Stworzenie modelu nieosiągającego satysfakcjonujących wyników. (wysokie ryzyko) Rozwiązanie: - Doszkalanie się każdego z członków zespołu w zakresie pracy z danymi w formie szeregów czasowych 2. Możliwość przekształcenia wypracowanego modelu w nadmiernie złożony system, co utrudni zarządzanie i skalowanie. – spowodowane złożeniem 2 modułów, z których każdy może korzystać z kilku modeli wewnętrznych, a dodatkowo końcowy model musi być stworzony dla każdego z okresów predykcji oraz dla każdej z akcji/waluty/kryptowaluty (średnie ryzyko) Rozwiązanie: - Regularna kontrola skomplikowania modelu oraz wybór prostszych rozwiązań, gdy okażą się bardziej wydajne 3. Trudności w ocenie jakości pojedynczych komponentów złożonego modelu (model językowy oraz model na podstawie przeszłych cen) – spowodowane użyciem jednego końcowego modelu, który łączy moduł językowy i numeryczny (niskie ryzyko) Rozwiązanie: - Badanie wpływu pojedynczych komponentów przy braku zmiany pozostałych - Dodanie modułu wyjaśnialności (XAI) do rozwiązania	1. Brak wysokich wymagań sprzętowych, co ułatwia sprawne rozwijanie projektu. 2. Znajomość technik pracy z dużymi modelami językowymi, co powoduje łatwiejsze debugowanie i bardziej świadome wprowadzanie poprawek do modelu 3. Projektowanie architektury i środowiska z myślą o jego łatwej skalowalności 4. Możliwość dostosowania okresu prognozy – projekt będzie rozwijany tak, aby wytrenowanie nowego modelu dla innego okresu prognozy było proste, co umożliwia zmaksymalizowanie potencjału modelu (korzystanie z okresu prognozy, dla którego model działa najlepiej)

Tabela 6: Opis zagrożeń zewnętrznych przy tworzeniu systemu.

SWOT	Zagrożenia	Szanse
Zewnętrzne	1. Nieprzewidywalne wydarzenia rynkowe i globalne kryzysy finansowe mogą wpłynąć na dokładność modelu predykcyjnego. (wysokie ryzyko) Rozwiązanie: - Ocena modelu w kilku punktach czasowych nieużytych do trenowania w celu jego uogólnienia 2. Małe możliwości poznania dobrych praktyk oraz czego unikać przy pracy z takimi modelami spowodowane brakiem zadowalającej ilości jakościowej literatury w tematyce modeli językowych skierowanych na tematykę finansową ze względu na znaczący rozwój technologiczny w dziedzinie LLM w ostatnim roku. (średnie ryzyko) Rozwiązanie: - Położenie większego nacisku na własny sposób rozwiązania problemu przy wsparciu istniejącej literatury 3. Zależność od dostępności strony dostarczającej dane – portale dostarczające newsy oraz historyczne ceny akcji/kryptowaluty mają określone limity requestów (niskie ryzyko) Rozwiązanie: - Zaplanowanie procesu pobierania i przechowywania danych oraz ich ujednolicenia - Skorzystanie z większej ilości stron udostępniających dane	1. Wysoka dostępność danych finansowych i informacji z rynku. 2. Rosnąca popularność modeli językowych i ich zastosowań, co daje większe możliwości inspiracji oraz łatwiejszego szukania rozwiązań potencjalnych problemów. 3. Potencjał do przyciągnięcia inwestorów i klientów zainteresowanych narzędziem do prognozowania spadków cen akcji. 4. Niska ilość obecnych rozwiązań łączących 2 moduły, stąd łatwiejsza ścieżka przebicia do potencjalnych klientów

6 Propozycja rozwiązania

6.1 Narzędzia

Narzędzia, z których zdecydowaliśmy się skorzystać:

- **Python** (+biblioteki[1]) - uniwersalny wysokopoziomowy język programowania, który zostanie wykorzystany do modelowania, aktualizacji danych oraz przy tworzeniu interfejsu użytkownika.

[1]Kompletny spis użytych bibliotek będzie dostępny dopiero w końcowej fazie projektu ze względu na jego złożoność.

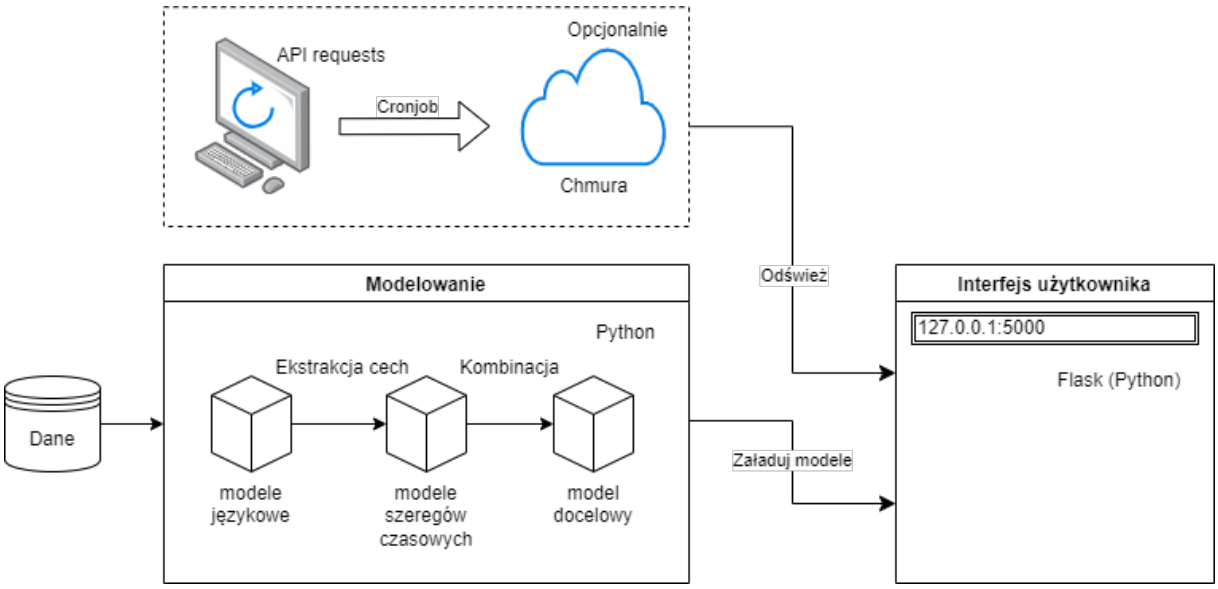
- **Flask** - framework służący do projektowania aplikacji przeglądarkowych napisany w języku Python. [**?**]

- **Baza danych SQLite** - baza danych charakteryzująca się szybką dostpnością danych dla niedużych potrzeb składowych. Wykorzystana zostanie do przechowywania danych potrzebnych do wizualizacji w interfejsie użytkownika z ostatnich 3 miesięcy. [**?**]

- **Wybrane rozwiązanie chmurowe**[2]

6.2 Komunikacja między komponentami

Rozwiązanie będzie składać się z dwóch głównych komponentów: *Modelowanie* oraz *Interfejs graficzny użytkownika*, którym ma być aplikacja przeglądarkowa. Szczególny nacisk kładziemy na pierwszy z komponentów. Dane, które posłużą do modelowania będą zebrane i przechowywane lokalnie na prywatnych maszynach.
Z danych odnośnie wiadomości wyektrachujemy cechy, którymi wzmocnimy modele bazujące na szeregach czasowych, u nas szeregach czasowych ceny akcji. Końcowo powstanie model klasyfikujący spadek ceny w danym przedziale czasowym.
Tak przygowany model docelowy zostanie zintegrowany z interfejsem graficznym użytkownika. Najnowsze dane będą aktualizowane na bieżąco z lokalnego komputera na daną platformę chmurową (opcjonalne) lub zaciągane bezpośrednio przez aplikację. Zamierzamy przechowywać w bazie danych zintegrowanej z aplikacją jedynie dane z ostatnich 3 miesięcy.
Tak przygotowane rozwiązanie pozwoli użytkownikowi na szybką analizę spadków bez konieczności składowania dużej ilości danych.



Rysunek 2: Graficzna wizualizacja schematu proponowanego rozwiązania.

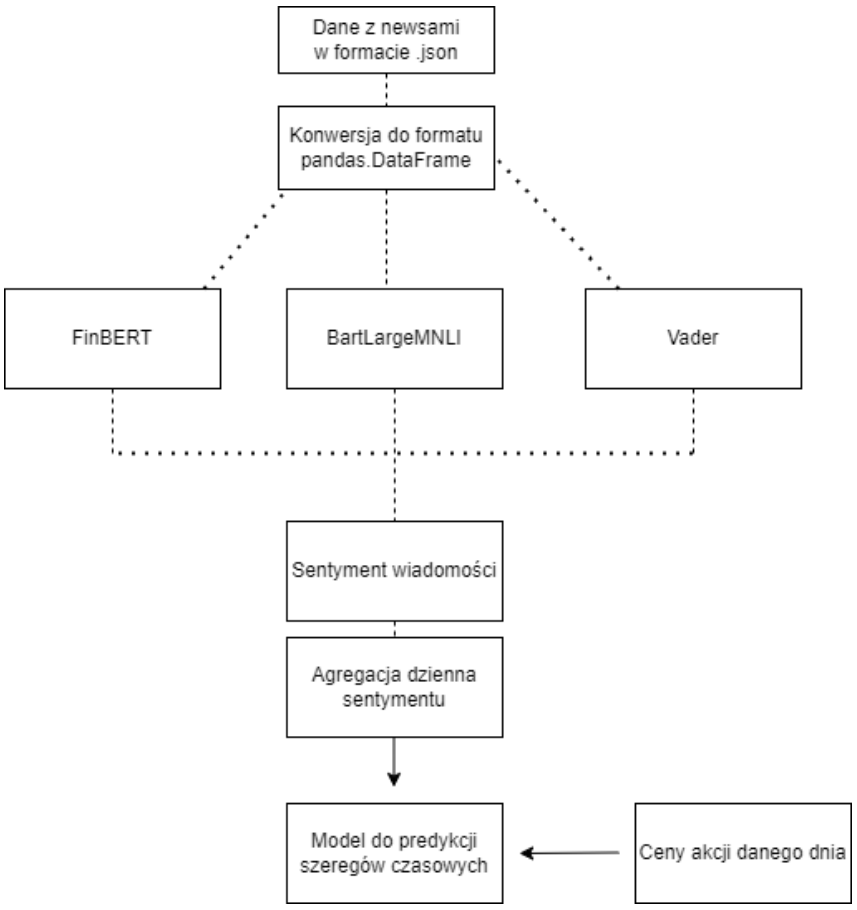
[2]Zastrzegamy sobie możliwość zmiany koncepcji aktualizowania danych w bazie lokalnej.

6.3 Modelowanie

6.3.1 Wiadomości

Podstawowym źródłem wiadomości będzie portal AlphaVantage. Wiadomości dostarczone przez portal są wyfiltrowane tylko do tych, obejmujących wybrane przez nas spółki. Modelowanie newsów polega na predykcji sentymentu wynikającego z treści wiadomości na podstawie dostępnych modeli językowych. Główny model to FinBERT, model BERT dotrenowany na danych finansowych do predykcji sentymentu, rozumianego jako wpływ newsa na wartość akcji danej spółki. Dodatkowo, użyty zostanie model BartLargeMNLI do zadań typu zero-shot classification, gdzie klasy nie są z góry zdefiniowane, a ich dobór jest częścią modelowania. Jako, że jest to model językowy, potrafi zaklasyfikować on tekst do podanych klas z pewnym prawdopodobieństwem. W naszym przypadku, używać będziemy klas powiązanych z wpływem newsa na cenę akcji (bullish, bearish, neutral). Trzecim modelem jest Vader - model do analizy sentymentu opierający się na analizie gramatycznej i leksykalnej. Nie jest to LLM, stąd czas klasyfikacji jest krótki. Ryzkiem jest brak konkurencyjności z dużymi modelami językowymi, natomiast ze względu na jego odmienną strukturę działania, wierzymy, że może dać inne spojrzenie na dane.

Poniżej przedstawiony jest schemat dziennego przetwarzania i modelowania newsów:



Rysunek 3: Dzienne przetwarzanie newsów finansowych

6.3.2 Ceny akcji

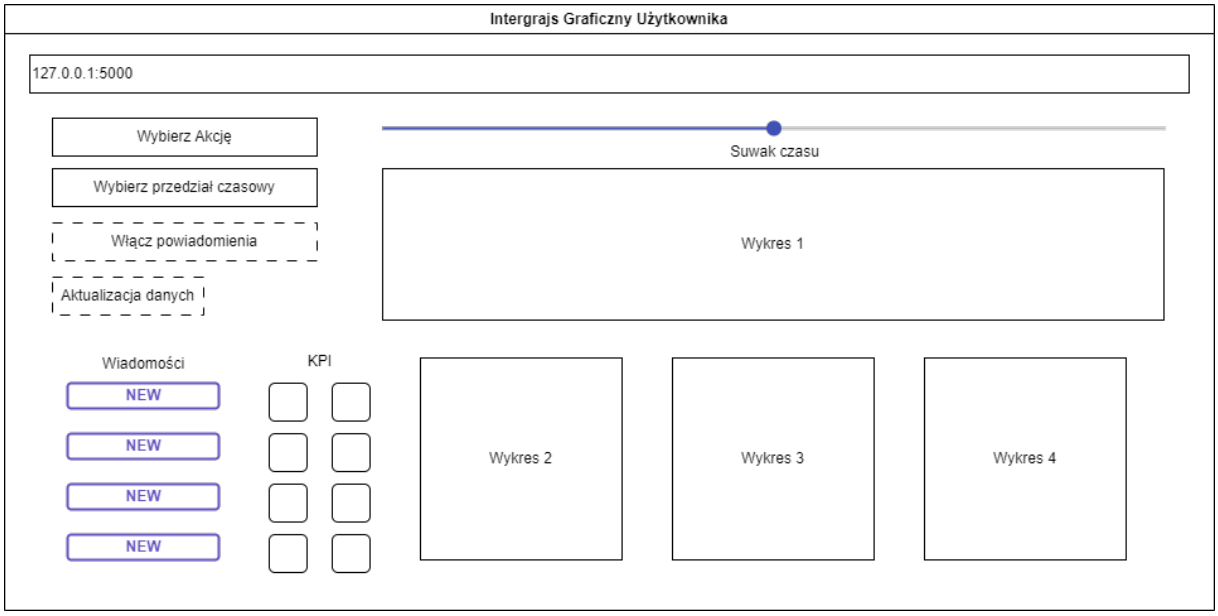
Dane finansowe o statystykach dziennych wybranej przez nas spółki w raz z informacjami o spółkach możliwie skorelowanych, będą w połączeniu z agregowanymi statystykami o sentymencie newsów tworzyć dane do modelu. Zakładamy finalnie użycie po trzech modeli dla danej spółki, odpowiednio zwracających cenę akcji w odstępie dnia, tygodnia i dwóch tygodni. Każdy z nich wytrenujemy od podstaw, sprawdzając modele:

- do klasycznych danych tabularycznych - regresja liniowa, catboost, xgboost;
- typowe do szeregów czasowych - arima, var;
- sieci neuronowe z wykorzystaniem warstw konwolucyjnych, LSTM oraz FC

6.4 Aplikacja webowa

6.4.1 Koncepcja interfejsu użytkownika

Użytkownik będzie miał do wyboru model, z którego będzie chciał skorzystać. Zrobi to przy pomocy zaznaczenia odpowiedniej akcji spółki/waluty/kryptowaluty oraz zaznaczenia przedziału czasowego, którym będzie zainteresowany. Suwak czasowy pozwoli na manipulację wyświetlanymi wykresami, których domyślnie ma być około 4, aby nie przeciążyć użytkownika. Wyświetlane będą jeszcze informacje na temat najświeższych wiadomości oraz kluczowe wskaźniki warunkujące działanie modelu. Opcjonalnie może zostać dodana możliwość włączenia powiadomień ostrzegawczych oraz przycisk zaktualizowania danych.



Rysunek 4: Graficzna wizualizacja projektu aplikacji przeglądarkowej. Miejsce poszczególnych komponentów na wizualizacji jest umowne. Przerywaną linią zaznaczone są elementy opcjonalne.

7 Raport danych

7.1 Ceny akcji

7.1.1 Wstępna analiza danych

Główne informacje finansowe zostaną uzyskane poprzez API yahoo finanse, które dla danego indeksu giełdowego zwraca w formie .json jego statystyki w formie dziennej.

Tabela 7: Schemat ramki danych z cenami akcji

Date	Open	High	Low	Close	Adj Close	Volume
2023-10-01	330.01	334.50	329.45	332.34	331.00	5 149 322
2023-10-02	332.21	337.23	331.55	336.01	333.91	7 421 266
...	...	...	...	...	...	...

Tabela 8: Opis ramki danych z cenami akcji

Index	Nazwa kolumny	Typ	Opis
0	Date	datetime	data przebiegu cen
1	Open	double	cena, po której akcja została otwarta na początku dnia
2	High	double	najwyższa cena, jaką akcja osiągnęła tego dnia
3	Low	double	najniższa cena, jaką akcja osiągnęła tego dnia
4	Close	double	cena, po której akcja została zamknięta na końcu dnia
5	Adj Close	double	cena, po której akcja została zamknięta na końcu dnia uwzględniająca wszelkie dostosowania, takie jak dywidendy i podziały akcji
6	Volume	integer	liczba akcji, które zostały wymienione w tym dniu

7.1.2 Wstępne przygotowanie danych

W naszej ocenie, istotne jest użycie w raz z informacjami cen akcji spółki, cen spółek możliwie skorelowanych, aby dostrzegać trendy i sytuacje na rynku oraz dane o substytutach dla naszej wybranej spółki. Dane dla wszystkich wybranych tagów giełdowych są pobierane i konwertowane do formy tabularycznej, a następnie łączone po dacie. Następnie aby zachować ciągłość daty, należy uzupełnić wartości dla weekendów, ponieważ wtedy giełda nie funkcjonuje i nie pojawiają się nowe wartości. Kolejnym etapem jest połączenie agregatów danych z wiadomości po dniu. Z powodu charakteru danych czasowych należy również zastosować operację okienkowania, czyli podziału danych na określone, nakładające się na siebie fragmenty o ustalonym rozmiarze, zależnym od modelu.

7.2 Wiadomości

7.2.1 Wstępna analiza danych

Dane dotyczące wiadomości ze świata finansowego na temat konkretnej spółki są dostarczane w formacie .json. Po konwersji do ramki danych, końcowa forma pliku z newsami jest taka, jak w tabeli poniżej:

Tabela 9: Schemat ramki danych z wiadomościami

Index	Title	Summary	Time Published	Other Columns
0	News Title 1	News Summary 1	01-11-2023 13:16:52	...
1	News Title 2	News Summary 2	02-11-2023 22:56:33	...
2	...	...	...	...

Wśród pozostałych istotnych kolumn możemy wyróżnić kolumnę Ticker Relevance Score mówiącą w jakim stopniu news dotyczy danej spółki, a także kolumny opisujące sentyment danej wiadomości wyliczany wewnętrznie przez firmę AlphaVantage, z których wstępnie nie planujemy korzystać.
Poniżej znajduje się spis wszystkich kolumn wraz z ich typem wewnętrznym (wynik wykonania metody pandas.DataFrame.info() na ramce danych z wiadomościami finansowymi) wraz z kolumną "Opis"stanowiącą krótki opis danej kolumny:

Tabela 10: Opis ramki danych z wiadomościami

Index	Nazwa kolumny	Typ	Opis
0	Title	object	tytuł newsa
1	URL	object	link
2	Summary	object	streszczenie newsa
3	Overall Sentiment Score	float64	ogólny sentyment wypowiedzi [-1,1]
4	Overall Sentiment Label	object	label sentymentu wypowiedzi
5	Ticker Relevance Score	float64	trafność newsa w kontekście spółki [-1,1]
6	Ticker Sentiment Score	float64	sentyment wypowiedzi w kierunku spółki [-1,1]
7	Ticker Sentiment Label	object	label sentymentu wypowiedzi w kierunku spółki
8	Time Published	object	czas publikacji newsa

7.2.2 Wstępne przygotowanie danych

Głównym problemem w przygotowaniu danych dotyczących newsów jest zaplanowanie ich pobierania ze względu na limity requestów do API strony dostarczającej wiadomości finansowe. Wybraliśmy stronę AlphaVantage, która zapewnia limity wystarczające do prawidłowego wytrenowania modelu i codziennego funkcjonowania aplikacji. Surowe dane są dostarczane w formacie json. Dane te są konwertowane do formy tabularycznej przy zachowaniu używanych feature'ów. Przechodzą one następnie przez modele językowe, gdzie z każdego z nich odpowiednie kolumny dodawane są do końcowej ramki danych.

8 Bibliografia

Literatura

[1] Dogu Araci, *FinBERT: Financial Sentiment Analysis with Pre-trained Language Models*, University of Amsterdam, 2019, arXiv:1908.10063.

[2] Liapis, Charalampos M., Aikaterini Karanikola, and Sotiris Kotsiantis. 2021. *A Multi-Method Survey on the Use of Sentiment Analysis in Multivariate Financial Time Series Forecasting* Entropy 23, no. 12: 1603. https://doi.org/10.3390/e23121603

[3] Smith, S., O'Hare, A. *Comparing traditional news and social media with stock price movements; which comes first, the news or the price change?.* J Big Data 9, 47 (2022). https://doi.org/10.1186/s40537-022-00591-6

[4] Kedar, S. V. . (2021). *Stock Market Increase and Decrease using Twitter Sentiment Analysis and ARIMA Model.* Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(1S), 146–161. https://doi.org/10.17762/turcomat.v12i1S.1596

[5] Junaid Maqbool, Preeti Aggarwal, Ravreet Kaur, Ajay Mittal, Ishfaq Ali Ganaie, *Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach*, Procedia Computer Science, Volume 218, 2023, Pages 1067-1078, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2023.01.086.

[6] Pratyush Muthukumar, Jie Zhong *A Stochastic Time Series Model for Predicting Financial Trends using NLP* Department of Computer Science, University of California (2021). https://arxiv.org/pdf/2102.01290.pdf

[7] Georgios Makridis, Philip Mavrepis, Dimosthenis Kyriazis *A deep learning approach using natural language processing and time-series forecasting towards enhanced food safety* Springer Science, Business Media LLC (2022). https://link.springer.com/content/pdf/10.1007/s10994-022-06151-6.pdf?pdf=button