



МАТЕМАТИЧЕСКА ГИМНАЗИЯ „АКАД. КИРИЛ ПОПОВ” - ПЛОВДИВ

4001 Пловдив, ул. „Чемшир” № 11, тел.: 032/643157; www.omg-bg.com; e-mail: omg_plovdiv@abv.bg

Lexonic

Мултимоделен подход към прогнозирането на
цените на акциите

АВТОРИ:

Петър Христов Желев

ЕГН: 0551284483

Адрес: “Прогрес” №4, Пловдив

Телефон: 0876159949

E-mail: pz66296122@edu.mon.bg

Училище: МГ „Акад. Кирил Попов”

Клас: X

Мартин Димитров Щилиянов

ЕГН:

Адрес: Ул. Прослава №56, Пловдив

Телефон: 0876531312

E-mail: ms97494228@edu.mon.bg

Училище: МГ „Акад. Кирил Попов”

Клас: X

РЪКОВОДИТЕЛ:

Дарина Брънчева

Телефон: 0888/992-447

e-mail: darina.brancheva@edu.mon.bg

Длъжност: Старши учител по информатика и информационни технологии

Съдържание

1. Резюме	3
2. Цели	4
3. Основни етапи в реализирането на проекта	5
4. Ниво на сложност на проекта	6
5. Логическо и функционално описание на решението	6
6. Реализация	9
7. Описание на приложението	9
8. Бъдещи разработки и идеи за реализация на проекта.....	10
9. Използвана литература и ресурси.....	10

РЕЗЮМЕ

Днес повече от всякога, имайки предвид глобалната пандемия и военния конфликт между Украйна и Русия, хората не са се нуждаели от инструмент, който да им помогне да преценят сегашния пазар.

Нашият проект „Lexonic“ представлява уеб платформа, създадена с цел да предостави анализ на всички, които се занимават с търговията на акции, както и на тези, които просто се интересуват от стоковата борса. Проектът предоставя както актуална информация за състоянието на всяка акция от 500-те акции в S&P 500, така и предположения за цената от аналитици и съвкупност от Machine Learning модели. Като имаме предвид функциите на нашата услуга, можем уверено да кажем, че нашият проект представя цялостна картина на пазара, от една страна, взима предвид исторически промени в цената на акциите, а от друга - мнението на обществото не само за определената акция, която търсят, а и за състоянието на пазара като цяло.

Отличителното на нашия проект от другите платформи, които предлагат същите услуги, е че те го правят чрез аналитици и стратегии, а по този начин е невъзможно както разглеждането на пазара от обективна гледна точка, така и преглеждането на всичката, достъпна до тях, информация. Чрез употребата на няколко модела, всеки за определената цел, както и употребата на подхода, описан по-долу, ние успяваме да получим обективен и цялостен поглед върху състоянието на пазара. Имайки предвид сложността на проекта, има още много място за развитие, даже и след представянето на проекта.

1. Цели

Проектът ни „Lexonic“ представлява програма, която предоставя на крайните потребители обективен анализ за състоянието на пазара, чрез употребата на иновативни технологии за финансово моделиране, обединяващи няколко модела, базирани Изкуствен интелект and Machine Learning. Неговата цел е да служи като инструмент в ръцете на търговците на пазара за акции, която да представи състоянието на пазара в цялата му същност с всичките му особености. В много малко моменти на човешката история, в икономическия свят е имало толкова разногласия за посоката, която ще поемат икономиките на отделни държави и световните пазари, което се дължи предимно на неизвестните последствия от глобалната пандемия и настъпващите международни конфликти, както и предишни конфликти, тепърва оставящи диря в икономическото развитие на някои страни. Именно затова смятаме, че нашият проект ще е от значителна полза за тези, които искат да се възползват от сегашната ситуация или да инвестициите си в капиталовия пазара.

Смятаме, че нашият проект подпомага на потребителите чрез предоставяне на:

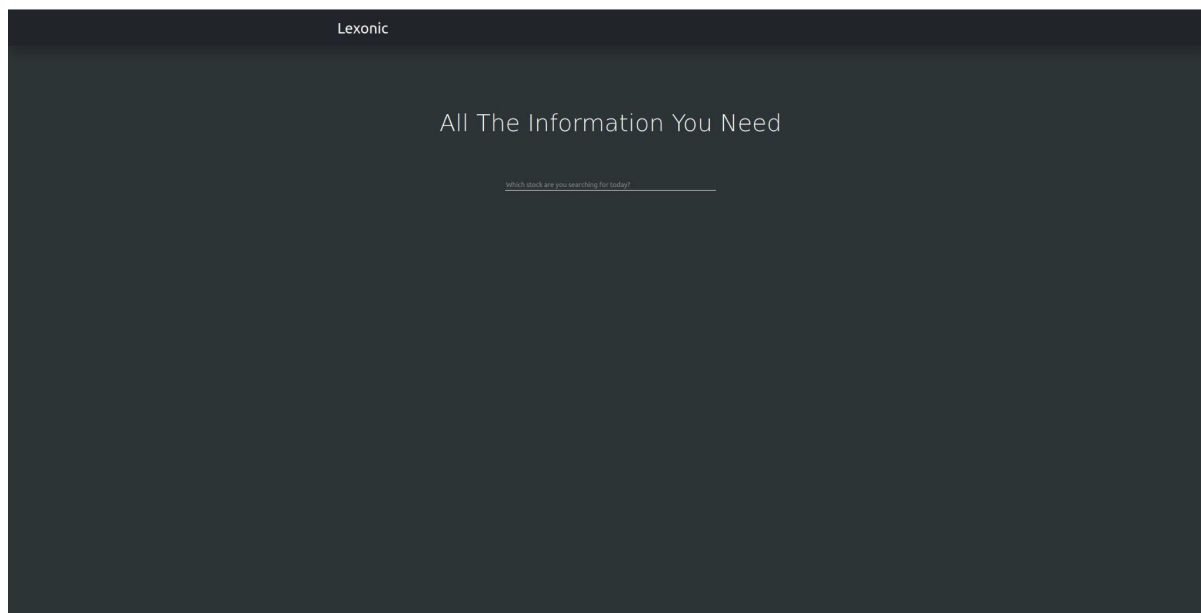
- Прогноза за цената на акциите;
- Прогноза на аналитици за цената на акциите;
- Графика на цената за последната година;
- Мнението на обществото в социалните мрежи за пазара;
- Всички важни подробности за състоянието на акцията;
- Най-скорошните новини за определена акция.

Ние предоставяме цялостен единен портал, чрез който потребителите да си създават представа както за пазара, така и за мнението на обществото, и да получават достъп до най-скорошните новини за дадена компания.

В сфера като търговията на акции, при която даже и малка преднина може да преобърне тактиките на всички други институции, които се занимават с търговия, е крайно важно да се продължи изследването и тестването на модели като нашия. Именно затова в заглавието сме отбелязали този проект като подход, а не архитектура, тъй като има много място за развитие. За да се осъществи това развитие обаче, са нужни изключителни количества данни, които да се обработят, и хардуер за тренирането на нови модели. Поради тази причина този проект е доказване на концепцията.

2. Основни етапи при реализирането на проекта

При реализирането на проекта преминахме през следните етапи:



2.1 Избор на темата

Преди три години, Мартин спореше с наш съученик за това дали акцията на Старбъкс ще се качи, или ще падне. Проблемът, който още тогава забелязахме, беше че и от двете страни имаше както липса на информация за определени теми, свързани с акцията, така и лично отношение към самата компания и кафето им. Макар и този спор да не завърши с разбирателство, от него произлезе нещо по-добро - идеята за създаване на система, която да предсказва цената на акции, индекси и т.н . И от нашия ранен интерес към стоковата борса и компютърните науки се стигна до момент, в който спряхме само да говорим за идеята и започнахме реализацията на подобен проект, който да слива двата ни интереса в едно, като ситуацията, в която се намира света в момента, само ни мотивира.

2.2 Определяне на структура на проекта

Работихме, изграждайки системата отдолу-нагоре, като започнахме да четем научни доклади, обяснения по темата в интернет. След което започнахме да намираме източници на данни и съответно за новини, както и избора на база данни. В последствие се наложи и имплементацията на самия модел, както и промяната му измежду различните версии. След като завършихме работата с данните и модела се насочихме към конфигурацията на уеб сървър и създаването интерактивен и интуитивен дизайн, който може

да бъде използван от всеки.

2.3 Изработване на проекта:

- а) Прочитане на научни доклади и добиване на идея за архитектурата;
- б) Подбиране на източници за информация и създаване на базата данни;
- в) Тестване на различни ML модели;
- г) Тестване и проверяване на изправността на компонентите на проекта;
- д) Разработка на графичен интерфейс;
- е) Коригиране на грешки, които пречат на проекта да работи по възможно най-добрия начин.

3. Ниво на сложност на проекта

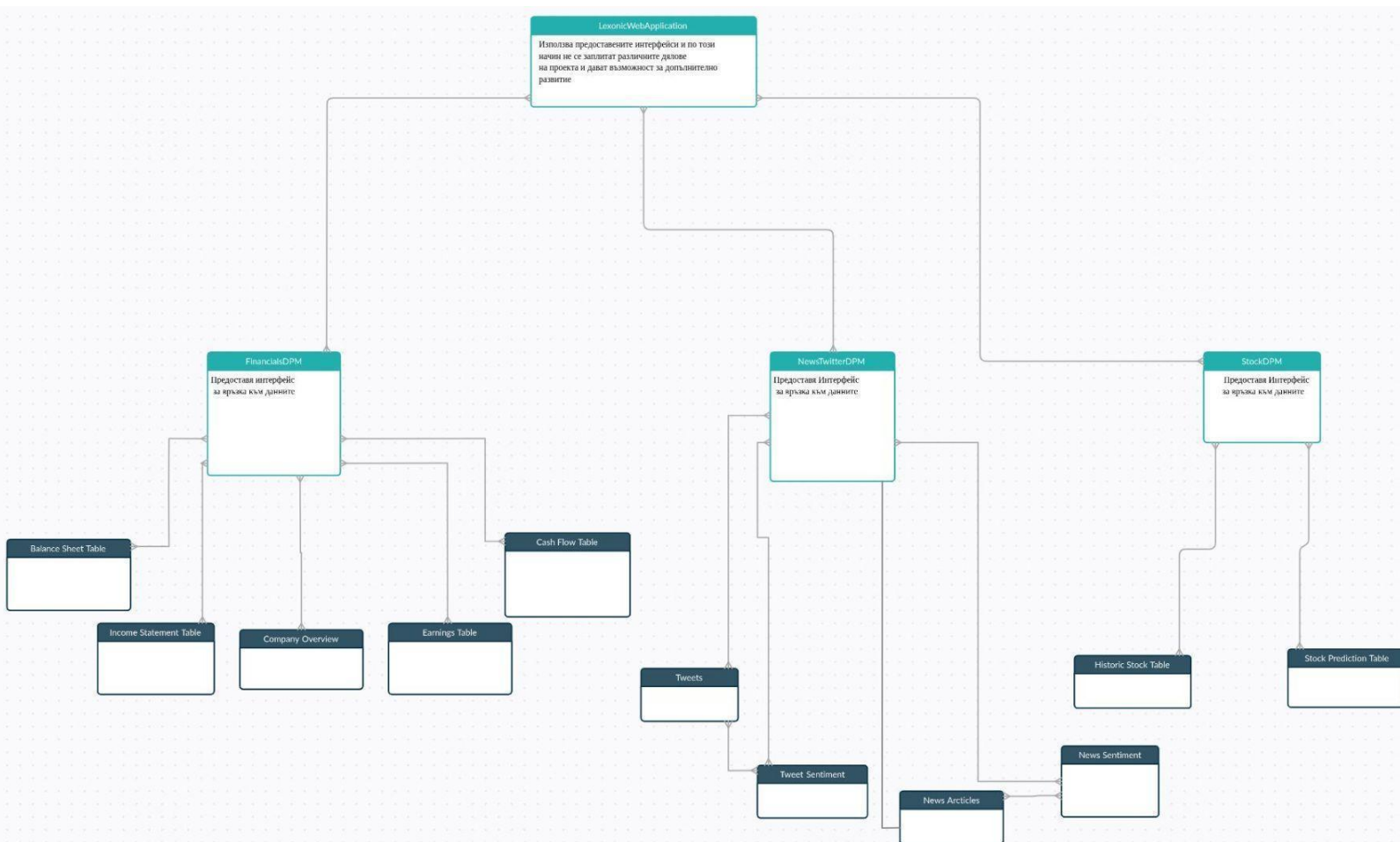
По време на разработване на проекта научихме много неща и се сблъскахме с много проблеми. Най-големият проблем беше интегрирането на ML моделите в C# код, като решихме да викаме моделите директно от Python скрипт. Други проблеми включват дизайна на ML моделите, връзката между софтуера и базата данни и графичен интерфейс.

4. Логическо и функционално описание на решението

Проектът "Lexonic" представлява приложна програма, чиято цел е да подпомогне работата на търговци на акции и хора, притеснени от текущата ситуация на пазарите. Реализацията включва иновативна технология за финансово моделиране обединяваща няколко модела базирани на изкуствен интелект и Machine Learning. Проектът е изработен с помощта на редица библиотеки като Tensorflow, Keras, Pytorch, HuggingFace Transformers, ASP.NET, Microsoft SQL server. Относно графичния интерфейс сме използвали HTML, CSS, JavaScript и React.js.

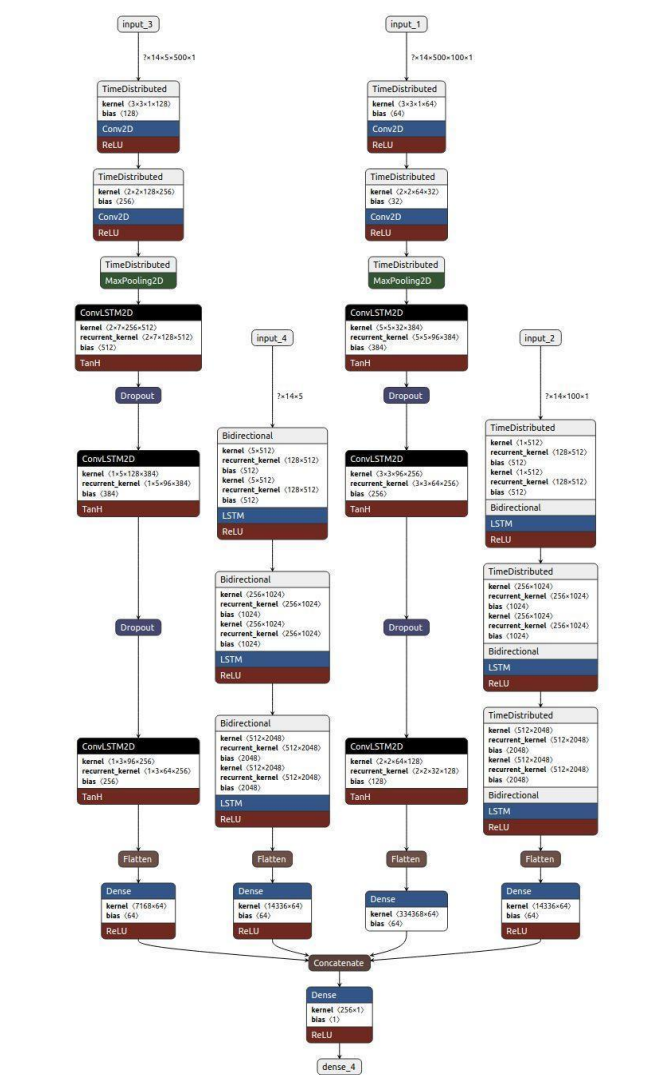
Проектът използва сървър, предоставен от Microsoft, а именно Microsoft SQL Server. Решихме да използваме тази база по няколко причини, като първата е, че и двамата имаме опит с тази база, но освен това ни гарантира Всеобхватност, Последователност, Независимост и Стабилност. Понеже даваме на потребителите данни от таблици е изключително важно свойството всеобхватност в една база, която ни гарантира, че даже и да правим промени, те ще бъдат извършени по начин, по който няма да бъдат афектирани тези, които четат данните. Освен това свойство на базата, тя гарантира, че ако сървърът спре, софтуерът даде грешка, данните в базата няма да бъдат заличени, както със In-Memory Databases.

Архитектурата на таблиците и връзката им между проектите е описана с графиката по-долу, самите колони на таблицата не са включени, понеже не се събират:



В основата на проекта са самите ML модели за изграждането на иновативния финансов алгоритъм. В текущата версия на проекта ние използваме 3 модела. В основата на 2 от 3те модела стои популярната архитектура за Sentiment Analysis - БЕРТ. Моделът, който използваме, за да преценяваме ефекта на новините върху борсата, се нарича ФинБЕРТ. Той е трениран на новини, но освен това е трениран на доклади от анализатори, както и доклади от дадената компания към регулативните институции на дадената държава, в която са базирани. В нашия случай, тъй като работим само с 500-те акции на S&P 500, те ще бъдат на един език и моделът няма да бъде афектиран от употребата на повече от един език. За разлика от много NLP модели, БЕРТ прави разлика в контекста и това е една от основните причини да се използва толкова много. Освен тези качества БЕРТ се предоставя на потребителите трениран и с възможността да бъде трениран за определена цел както ФинБЕРТ е трениран специфично за финансови новини и доклади. По същия начин моделът, използван за извличане на емоциите на хората от туйтове, Туит БЕРТ е трениран за тази определена цел и ни дава висока точност за специфичния домейн.

Архитектурата на третия модел може да видите на графиката по-долу:



Чрез графиката си проличава защо го наричаме мулти-моделен подход.

От една страна, чрез Input_3 разглеждаме промените в цените на всичките акции и ги гледаме като едно цяло. По този начин можем едновременно да намерим закономерности в целия пазар, а от друга страна, моделът може да види какво се случва в целия пазар и да даде по-точен резултат за определената акция.

Input_4 е другата крайност на Input_3, като се разглежда какво се случва в определената акция, като използва Bidirectional слоеве, защото, за да предположим цената, трябва да търсим закономерности, гледайки назад във времето. Благодарение на разделението между модела, който разглежда всичките акции, и този, който разглежда една акция, ние успяваме изцяло да премахнем човешкия фактор в преценяването в това какво влияе на цената на акцията.

Аналогично на гореописаната част на модела Input_1 и Input_2 са отговорни за търсене на закономерности в сентимента на финансовите новини. По същия начин използваме Bidirectional слоеве. Идеята на целия този модел с различните модули, събрани в последствие на едно, е да създадем цялостен снимък на пазара, за да се създаде представа, за точно това, което търсим, а именно посоката на цената.

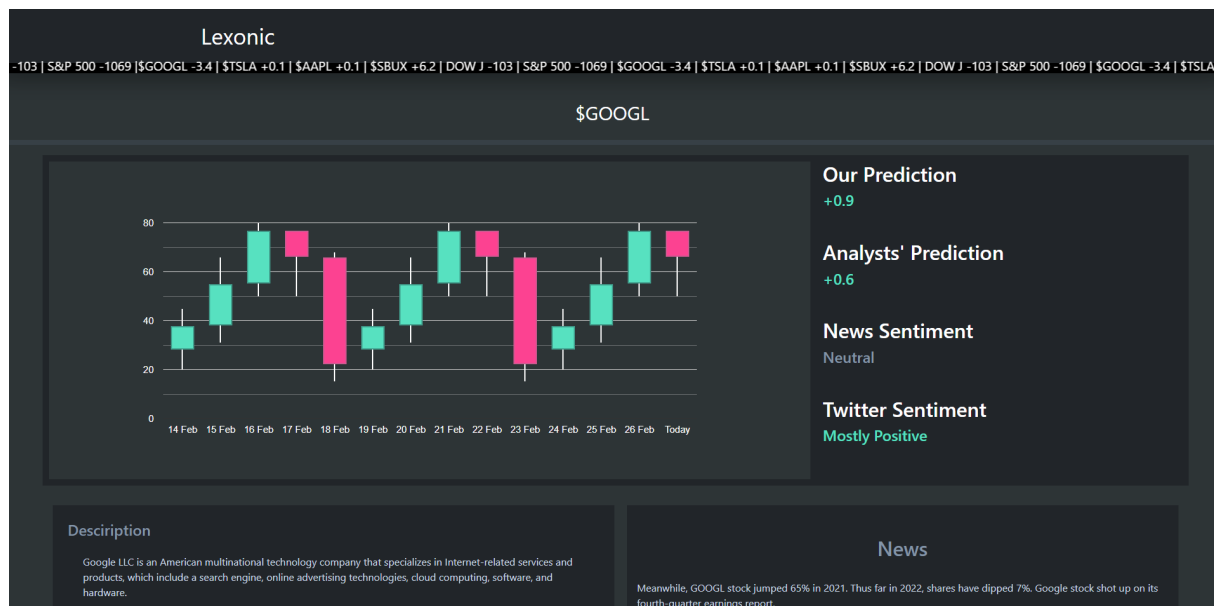
5. Реализация

Проектът е реализиран чрез комбинация от C# и Python. За създаването му използвах JetBrains Rider и JetBrains PyCharm. За създаването на таблиците и графичния интерфейс използвах JetBrains Rider. За контрол на версиите използвах Git и GitHub.

6. Описание на приложението

Проектът ни напълно покрива нуждите на хората, които искат да анализират цената на определена акция и да видят цялостно състояние на капиталовия пазар.. Графичният потребителски интерфейс е разделен на 3 основни части.

Първата, която се намира в началото на страницата, се състои от логото на нашия проект, името на потърсената акция и лента, която показва промените в отделните сектори на пазара.



Втората част представлява графика, която показва промяната в цената на търсената акция в последните 14 дена и вдясно от нея са очакванията за цената на акцията на следващия ден въз основа на прогноза от нашия финансов модел, както и настроението на скорошните новини и на обществото мнение в социалните мрежи.

В третата част има описание на акцията и нейните показатели в сегашния пазар. Вдясно от нея са най-съществените скорошни новини, свързани с акцията или самата компания.

7. Бъдещи разработки

Още със създаването на проекта ни беше ясно, че ще има много бъдещи разработки и именно затова Microservices архитектурата, която сме избрали е идеална. Позволява ни да добавяме модули към проекта, без да трябва да пренаписваме другите. Както е в случая - имаме модул, който тегли данните за акциите и ги обработва, имаме модул, който тегли новините и туитовите и ги обработва, и модул, за финансовите данни на компаниите. Всичките тези модули не общуват помежду си, а само със самия Web Application.

Ясно е, че само с новини, туйитове и историческите промени в цената няма как точно да предположим конкретната цена на акция, особено имайки предвид, че нямаме достатъчно данни. Затова още от сега сме започнали да четем научни доклади за начина, по който влияят различните фактори, като новоприети закони, новоизбрани политици и най-вече по-добър достъп до тези и всички останали, нужни за функционирането на проекта, данни. Освен това сме предвидили цялостно иновативен модел, който след известно тестване ще пуснем в продължън кода.

8. Използвана литература

Araci, Dogu Tan, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models." ArXiv, 27 August 2019,

<https://arxiv.org/pdf/1908.10063.pdf>

2. Lee, Si Woon, et al. "Stock market forecasting with super-high dimensional time-series data using ConvLSTM, trend sampling, and specialized data augmentation.", Available online 8 July 2020,

<https://www.sciencedirect.com/science/article/abs/pii/S0957417420305285>.

3. Ren, Yinghao, at al. "Impact of News on the Trend of Stock Price Change: an Analysis based on the Deep Bidirectional LSTM Model.", Available online 27 July 2020,

<https://www.sciencedirect.com/science/article/pii/S1877050920315829>

4. Devlin, Jacob, et al. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." ArXiv, 24 May 2019,

<https://arxiv.org/pdf/1810.04805.pdf>

5. Bustos, O., Pomares-Quimbaya, A."Stock market movement forecast: A Systematic review.", Available online 20 April 2020,

<https://www.sciencedirect.com/science/article/abs/pii/S0957417420302888>