1)Прогнозирование направления движения цены биржевых инструментов по новостному потоку.

Мотивируемое тем, что флуктуации цен на бирже, сильно зависящие от политической, географической и т.д. обстановок,  интересные не только при скальпинге. Для среднесрочных торгов и инвестиций такие данные также имеют большую роль, позволяя корректировать вложения. Как правило, крупные изменения в политике, природные катаклизмы и все события которые изменяют распределение цен котировок, освещаются в прессе.

2)исследование строится вокруг постоянных изменений цен биржевых котировок, новостей, и алгоритма NMF вектора.

3)Требуется на основе большого количество новой информации (предоставляемой в разрозненном текстовом виде)  касающейся компаний, перечисленных на фондовом рынке, предсказать повышение, понижение либо стабилизацию цен на акции, ценные бумаги и т.д. Необходимо разработать модель, которая также учитывает недавнее движение акций, и так называемую “неожиданную прибыль”(*отчет о прибылях и убытках компании, значительно отличающийся (в положительном или отрицательном направлении) от ожиданий аналитиков (согласованного прогноза)*

4) **Методы исследования**

В работе приведены другие, которые как улучшают уже существующие, так и вводят новые методы обработки естественного языка.

Так в Xie et al. (2013) вводится дерево представлений об информации в новостях, в Bollen et al. (2010) использованы данные из Twitter'a.

Bar-Haim et al. (2011) распознают лучших экспертов-инвесторов, а Leinweber and Sisk (2011) исследуют влияние новостей и времени усвоения новостей в событийной торговле.

В Kogan et al. (2009) приводится предсказание риска по финансовым отчётам и в Engelberg (2008) - закономерность о том, что лингвистическая информация (возможно из-за когнитивной нагрузки при обработке) имеет более долгосрочную предсказуемость цен, нежели количественная информация.

5) **Решаемая в данной работе задача**

Построить и исследовать модель прогнозирования направления движения цены. Задано множество новостей S и множество временных меток T, соответствующих времени публикации новостей из S. 2. Временной ряд P, соответствующий значению цены биржевого инструмента, и временной ряд V, соответствующий объему продаж по данному инструменту, за период времени T'. 3. Множество T является подмножеством периода времени T'. 4. Временные отрезки w=[w0, w1], l=[l0, l1], d=[d0, d1], где w0 < w1=l0 < l1=d0 < d1. Требуется спрогнозировать направление движения цены биржевого инструмента в момент времени t=d0 по новостям, вышедшим в период w.

6) **Предлагаемое решение**

8К - отчеты компаний об их внутренних событиях. Данная отчетность выходит строго в период между закрытием торгов в один день и их открытием на следующий день.

Из отчета 8К убираются все HTML-теги, таблицы и прочее.

Используется метод NMF вектора.

Вычитается из цен сегодняшнего открытия торгов вчерашние цены закрытия торгов с поправкой на индекс.

Берется текст отчета 8К и на выходе нейронной сети функция, принимающая три значения :

\*UP-цена открытия следующего дня больше на 1+% от предыдущего дня - “изменение индекса”

\*DOWN- цена открытия следующего дня меньше на 1+% от предыдущего дня - “изменение индекса”

\*STAY - цена открытия следующего дня в пределах +/-1%  от предыдущего дня - “изменение индекса”

7) **Работа, описывающая наиболее близкое решение** https://drive.google.com/file/d/12KsFJNEADfXYLlV0Ler19A1E9N1pGa3-/view?usp=sharing

8) **Плюсы метода:**

* Большой объем данных
* Он более доступен небольшим инвесторам, чем real-time trading tools, которыми пользуются большие трейдинговые компании
* Он показывает accuracy на 10% больше, baseline, который использует только финансовые фичи(см. Статью в пункте 7)
* смотрят “изменение цены”-”изменение индекса” => чистое влияние
* все дивидендные гэпы убирали

**Минусы:**

* Исследование проведено на рынке США, где отчеты выходят не в торговое время => вся информация отражается мгновенно в цене акции от открытии
* результаты не имеют значения на практике => невозможно извлечь финансовую прибыль
* Метод не улавливает такие эффекты, как: slippage, transaction costs, borrowing costs

9) -

10) Эксперимент будет проведен на финансовых данных: данные о котировках (с интервалом в один тик) нескольких финансовых инструментов (GAZP, SBER, VTBR, LKOH) за 2 квартал 2017 года с сайта Finam.ru; для каждой точки ряда известны дата, время, цена и объем. И на текстовых данных: экономические новости за 2 квартал 2017 года от компании Форексис; каждая новость является отдельным html файлом.