

Отчет о выполнении проекта: “Влияние новостей на финансовые рынки”

куратор: Андрей Новиков

участники: Емельянов Андрей, Попов Матвей, Зелинский Илья,
Летягин Артём, Налепова Елизавета, Маняхин Дмитрий

Датасет новостей

LATEST

Markets
**Americans Turn
Against China
Stocks as
Crackdown Angst
Deepens**

43m ago

Politics
**Afghan Airlift
Deadline Must Be
Extended,
Lawmakers Tell
Biden**

48m ago

Politics
**Trump Calls
Biden's
Afghanistan Exit
the 'Greatest
Embarrassment'**

1h ago



Politics
**Boris Johnson
Faces Tory
Backlash on
Afghanistan
After Recalling
MPs**

Influential
Conservative Party
members criticize the
decision to withdraw

Markets
**Stocks Steady,
U.S. Futures
Waver Amid
Virus Die-
Back**

Read a
move
market



22.07.21

India's Environment Ministry has almost halved the time it takes to process environmental clearance for projects.

22.07.21
Singapore's home price growth slowed in the second quarter after residential property prices are expected to continue to expand.

22.07.21
What keeps breaking the buck, is engaged in financial transformation by JPMorgan Chase & Co. strategist Josh Younger describes stable top four coins have 'broken the buck' (i.e. dipped below the dollar).

22.07.21
Malaysia has confirmed one case of an individual being injected with the virus.

22.07.21
Xponential Fitness Inc., the franchise owner of boutique fitness studio, reported a record number of new franchisees.

22.07.21
Japan's Olympics organizers reported a record number of new franchisees.

22.07.21
India's Environment Ministry has almost halved the time it takes to process environmental clearance for projects.



источники: <https://www.bloomberg.com/europe>

репозиторий: <https://github.com/IlyaZelinskiy/ImpactOfNewsOnFinancialMarkets.git>

Датасет по фин. активам

Дата | цена открытия | цена макс. | цена мин. | цена закрытия | объем торгов

	time	open	high	low	close	volume
2019-08-05	04:15:00	46.708	46.708	45.722	45.722	17555
2019-08-05	04:30:00	45.700	45.708	45.462	45.462	11755
2019-08-05	04:45:00	45.524	45.598	45.524	45.598	18960
2019-08-05	05:00:00	45.742	45.746	45.742	45.746	6540
2019-08-05	05:15:00	45.700	45.840	45.700	45.840	3045

репозиторий: <https://github.com/aeradcs/get-stock-data>

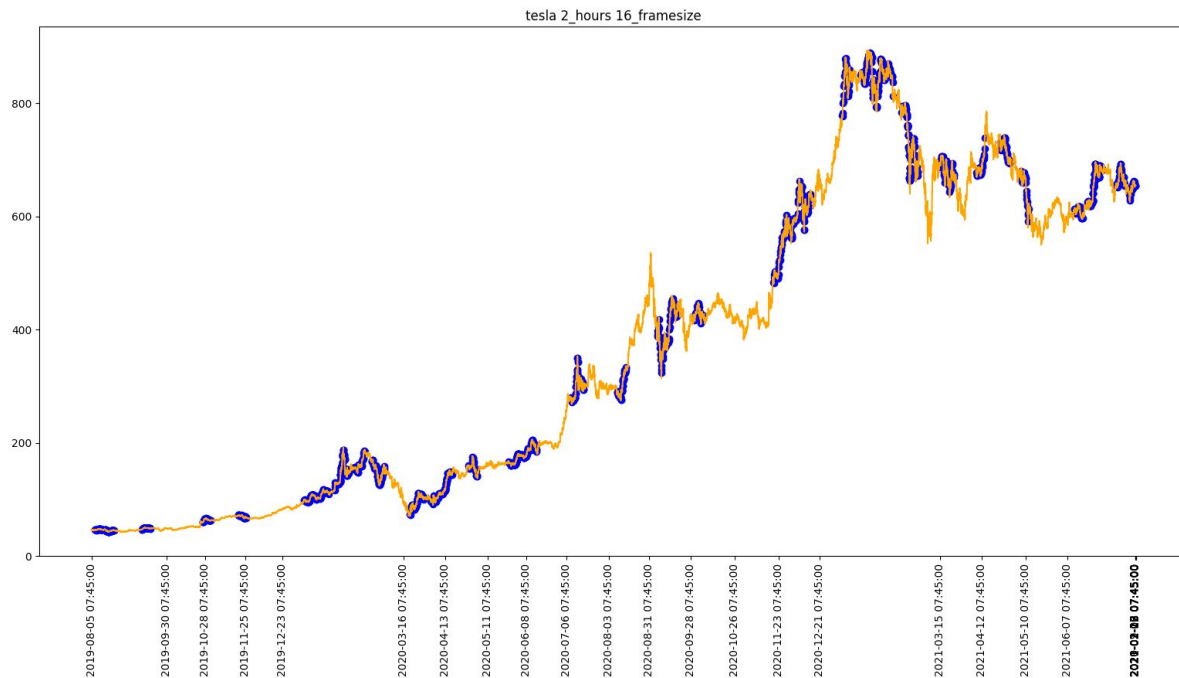
источник: <https://www.alphavantage.co/>

Описание точек разладки

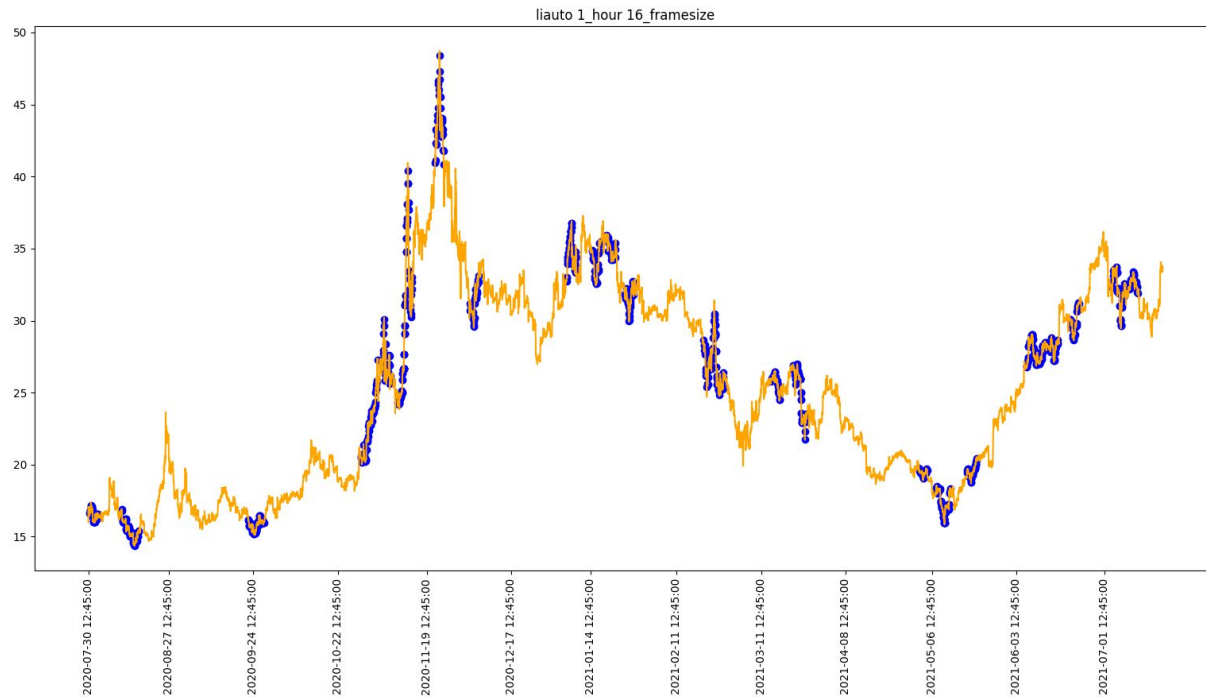
$$y' = ry$$

$$r = \frac{y'}{y} = (\ln y)' \sim \ln(y_n) - \ln(y_{n-1})$$

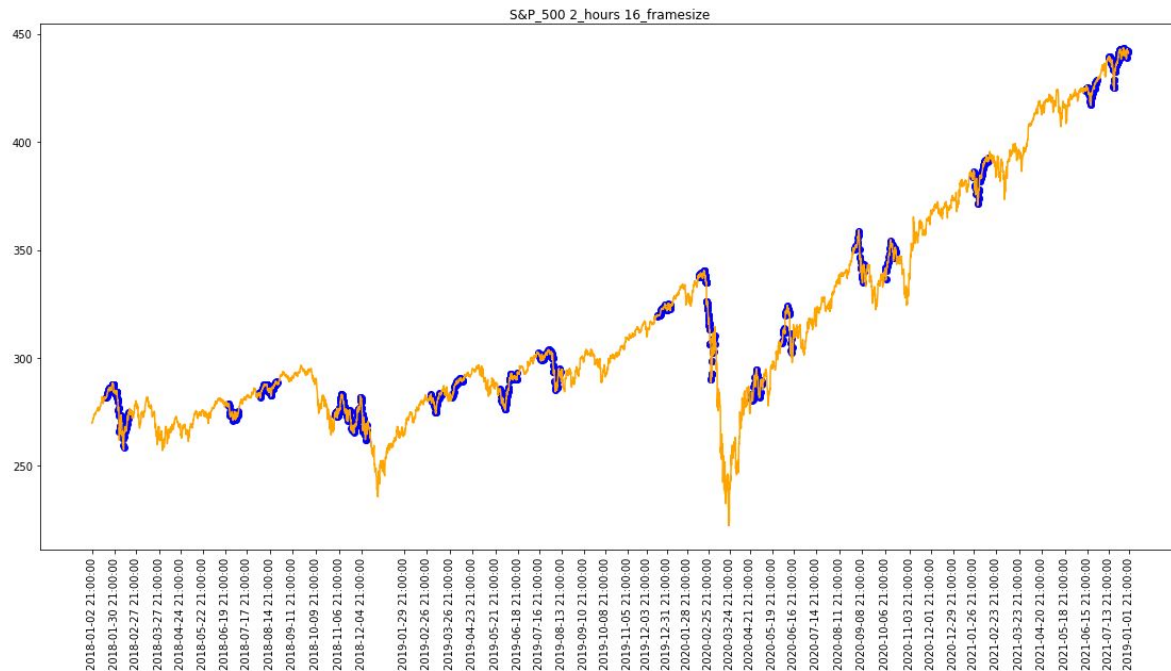
Пуск точек разладки (Tesla)



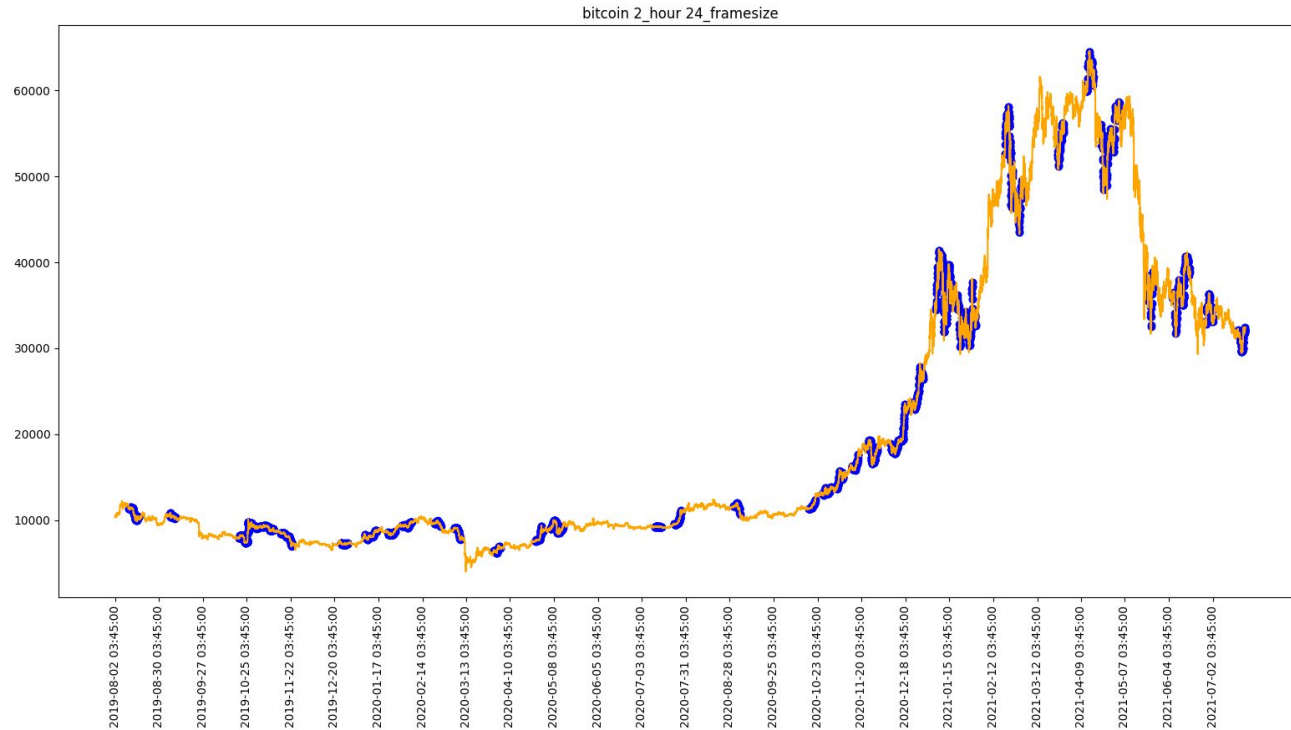
Пуск точек разладки (Li Auto)



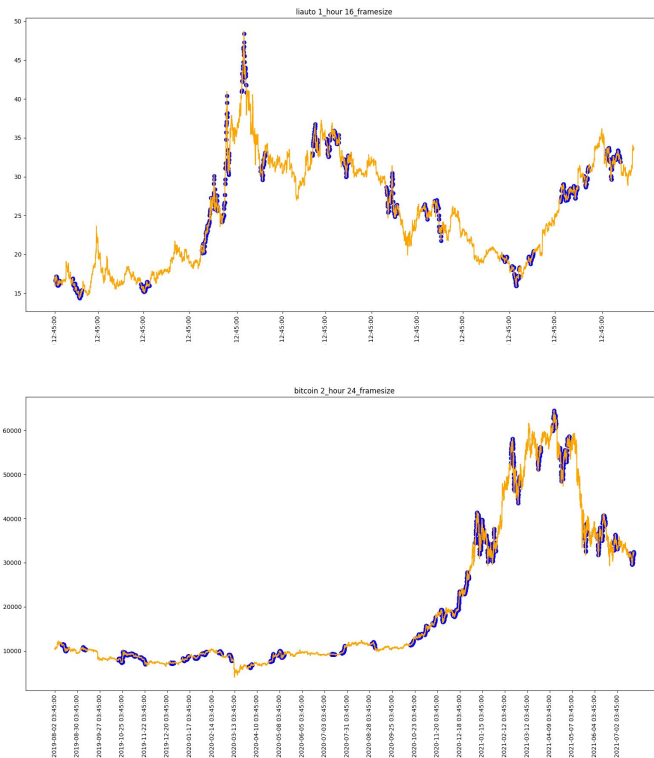
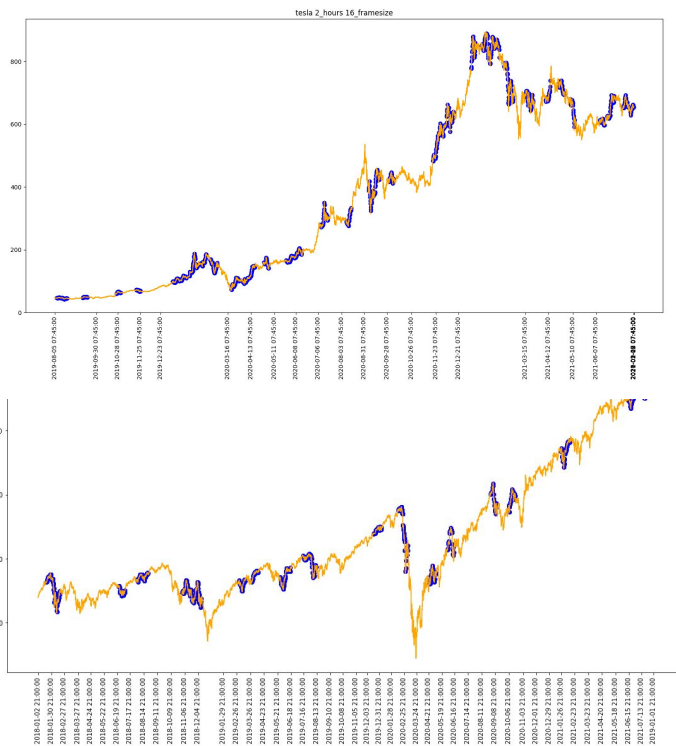
Поиск точек разладки (S&P500)



Поиск точек разладки (Bitcoin)



Поиск точек разладки



Рубрикация текстов. Эмбединг

news

India's Environment Ministry has almost halved...
Singapore's home price growth slowed in the se...
What keeps breaking the buck, is engaged in fi...
Malaysia has confirmed one case of an individu...
Xponential Fitness Inc., the franchise owner o...

news

[indias, environment, ministry, almost, halved...
[singapores, home, price, growth, slowed, seco...
[keeps, breaking, buck, engaged, financial, tr...
[malaysia, confirmed, one, case, individual, i...
[xponential, fitness, inc, franchise, owner, b...

номер слова
в словаре

количество
повторений
в тексте

0	1	2	3	4	...	18106	18107	18108	18109	18110
2.0	1.0	1.0	1.0	1.0	...	0.0	0.0	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0

Рубрикация текстов. Трехмерные визуализации

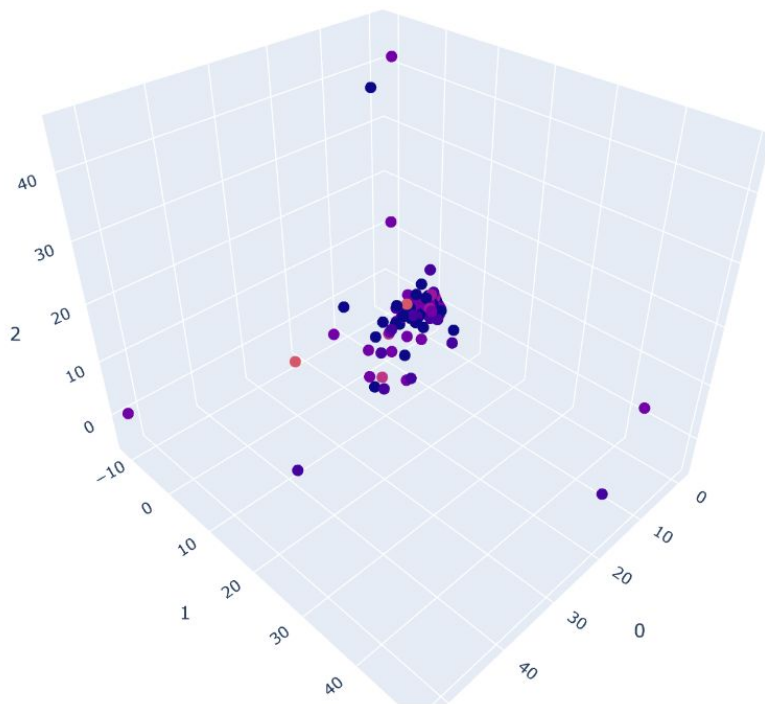
bag of words для эмбединга



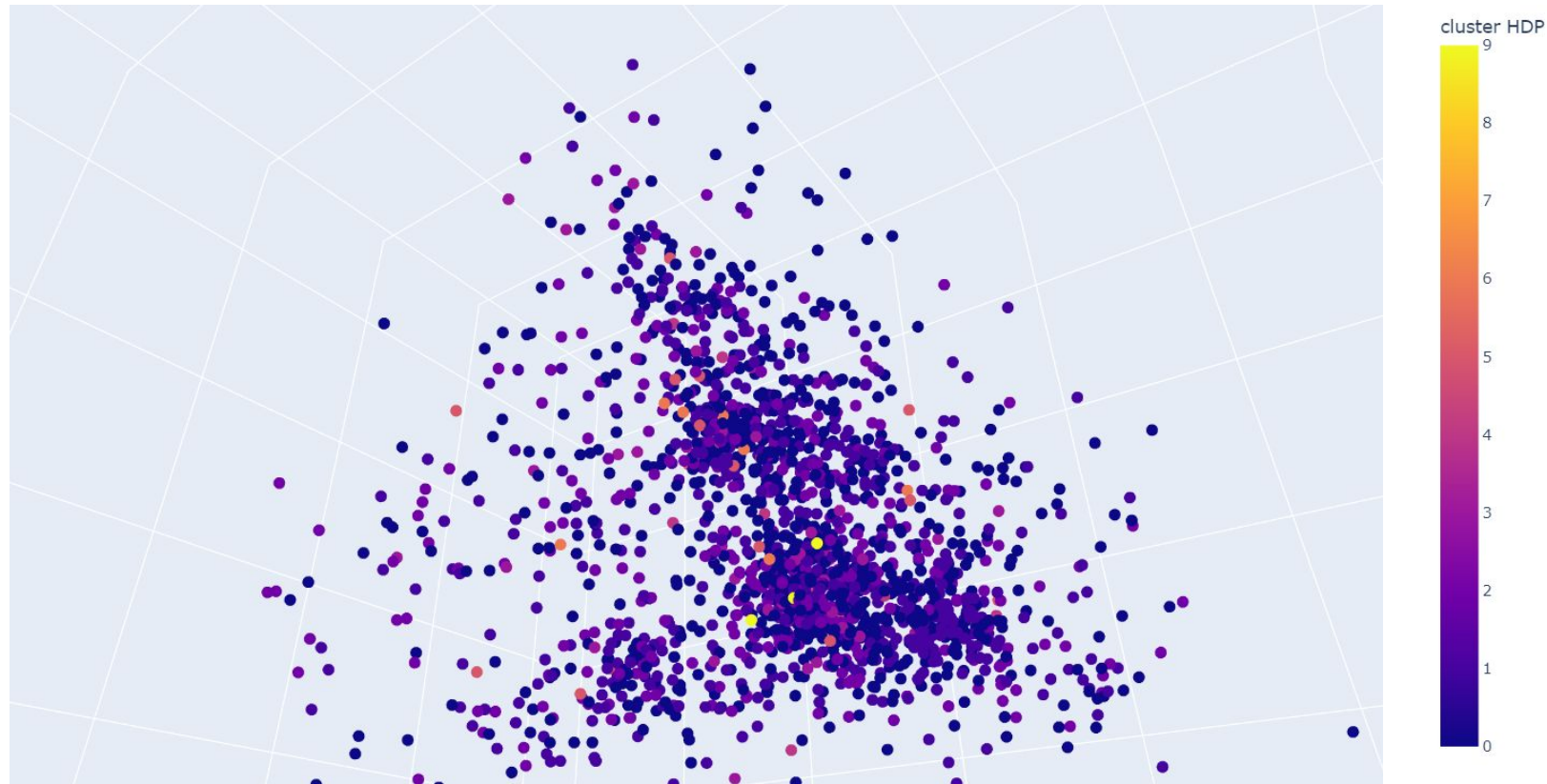
HDP для выделения тем



PCA для понижения размерности
векторов-новостей



Рубрикация текстов. Трехмерные визуализации



RandomForestClassifier бенчмарк

	precision	recall	f1-score
0	0.599	0.982	0.744
1	0.818	0.281	0.419
2	1.000	0.089	0.163
3	0.000	0.000	0.000
4	1.000	0.333	0.500
5	0.000	0.000	0.000
6	0.000	0.000	0.000
8	0.000	0.000	0.000
9	0.000	0.000	0.000

accuracy	0.623
----------	-------

репозиторий: <https://github.com/aeradcs/news-embedding-hdp>

Выделение именованных сущностей из каждой новости

A former trader at **Bank of America Corp.'s Merrill Lynch** unit told a federal jury in **Chicago** that he learned how to manipulate the price of precious metals from two more senior traders in the bank's **New York** office, John Pacilio and Edward Bases. Harnaik Lakhan said he used **Merrill's** internal computer system...

•
•
•



Bank of America Corp.'s Merrill Lynch, Merrill



Chicago, New York



ссылка на код: <https://github.com/ArtemLetyagin/named-entity-recognition.git>

Новости подозрительные на отношение к разладке

$p = (\text{количество новостей в теме, потенциально влияющих на значение котировок}) / (\text{общее число новостей в теме})$

1. по новости предсказываем кластер
2. по кластеру находим p
3. если $p > 0.5$, то новость помечается как потенциально влияющая

Мы готовы ответить
на ваши вопросы