

Применение глубокого обучения для оценки финансовых инструментов

Студент:

Леонова Алина Дмитриевна
НФИМд-01-21

Научный руководитель:

к.ф.-м.н, доцент кафедры ИТ
Шорохов Сергей Геннадьевич

Кафедра информационных технологий

Российский университет дружбы народов имени Патриса Лумумбы

Финансовые инструменты



Финансовый инструмент — это юридический термин, используемый в финансах для описания всех требований, других активов и обязательств, которые имеют в качестве своего прямого или косвенного объекта обмен платежными средствами.

Например, финансовыми инструментами являются: ценные бумаги, облигации, деривативы, фьючерсы, опционы, свопы и другие.



Машинное обучение — это общий термин для "искусственного" получения знаний из опыта: искусственная система учится на примерах и может обобщать их после завершения этапа обучения.

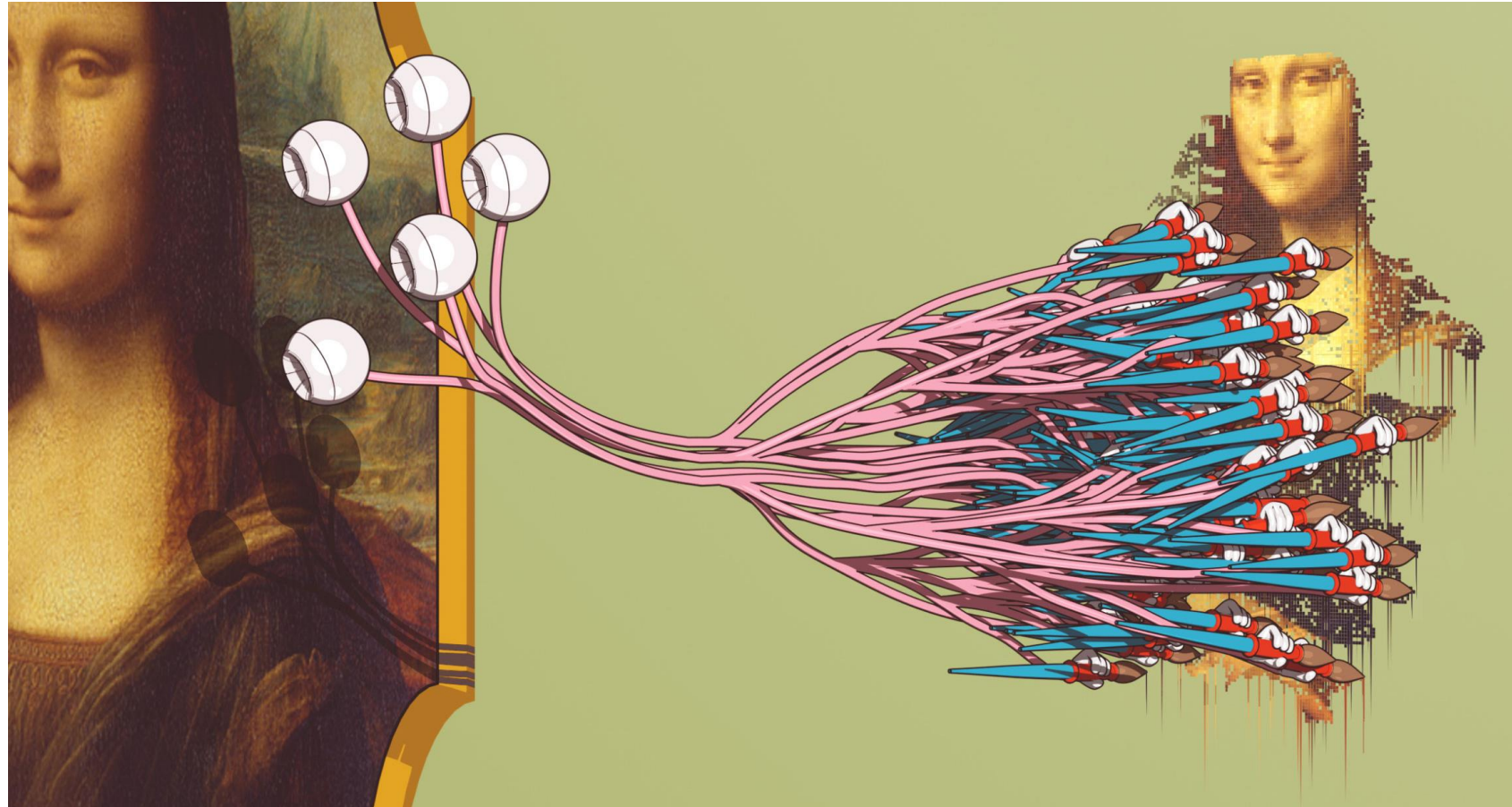


Глубокое обучение и нейронные сети

Глубокое обучение — это метод машинного обучения с помощью искусственной нейронной сети, имитирующей человеческий мозг, состоящий из нейронов, организованных в сеть.

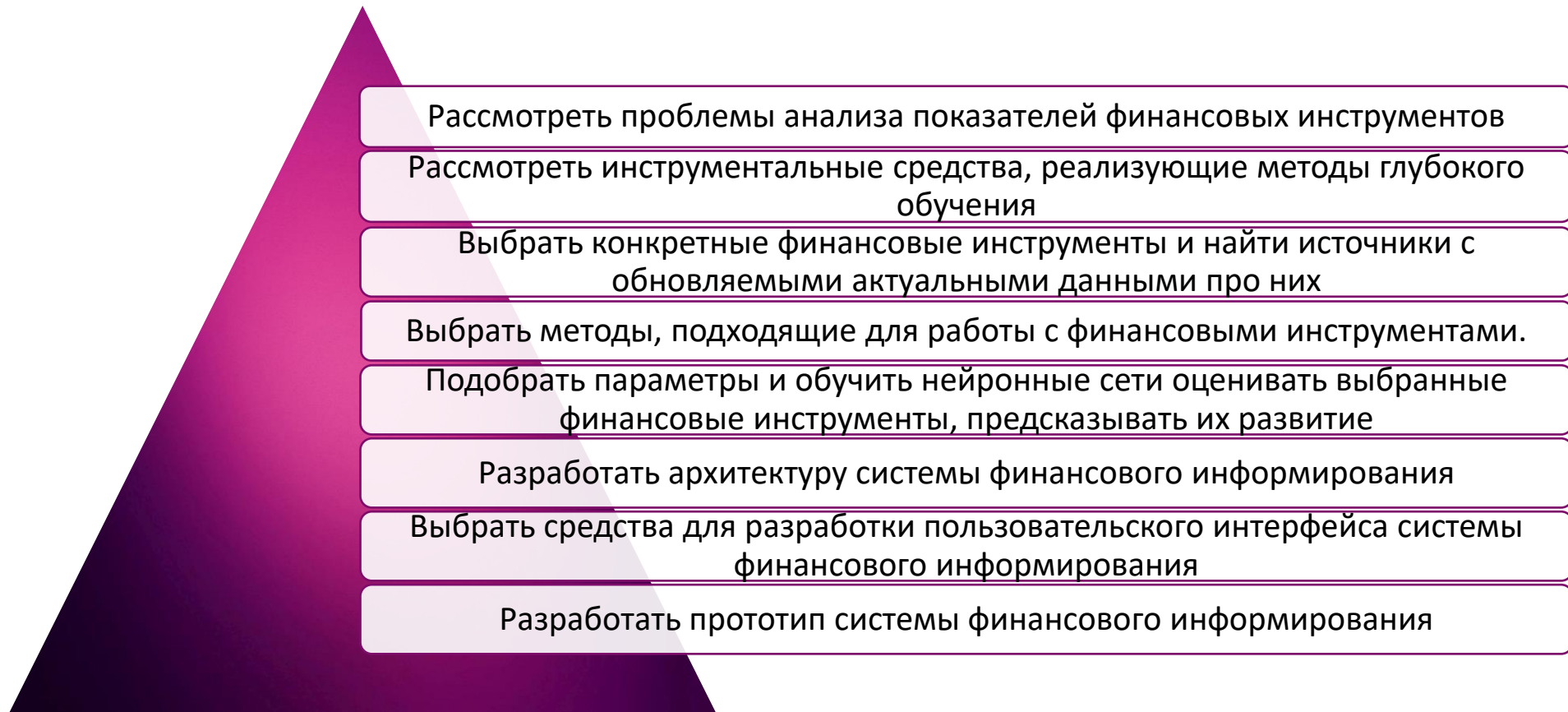


Искусственный интеллект
Машинное обучение
Нейросети
Глубокое обучение



Постановка задачи

Целью работы является исследование применимости средств глубокого обучения к решению задачи оценки финансовых инструментов, а также разработка прототипа системы финансового информирования, обеспечивающего использование реализованных средств.



Структура системы финансового информирования

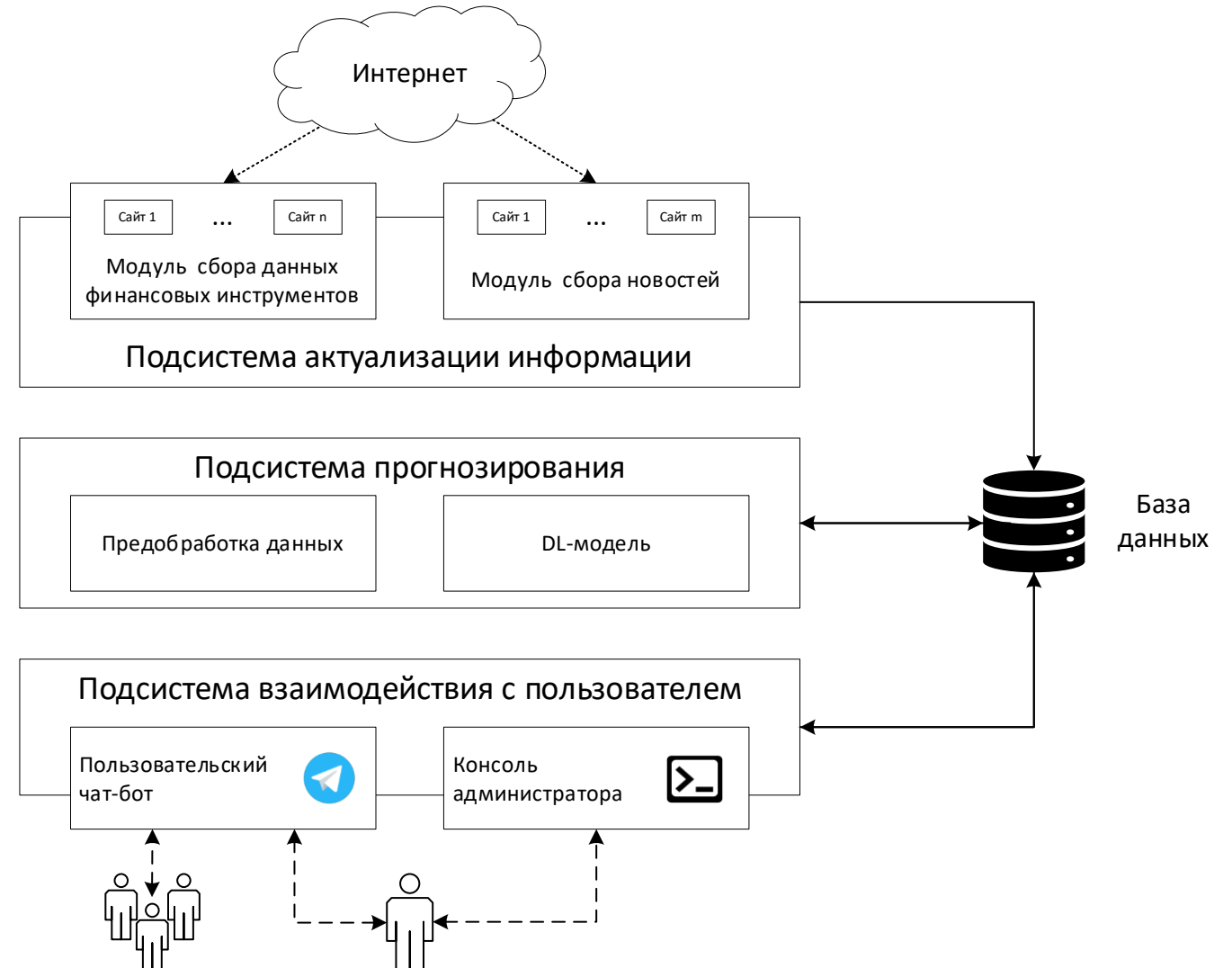
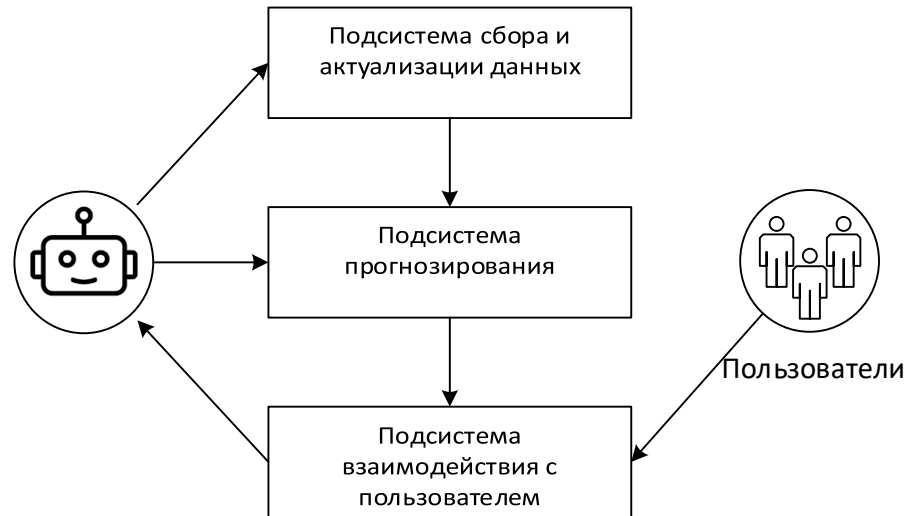
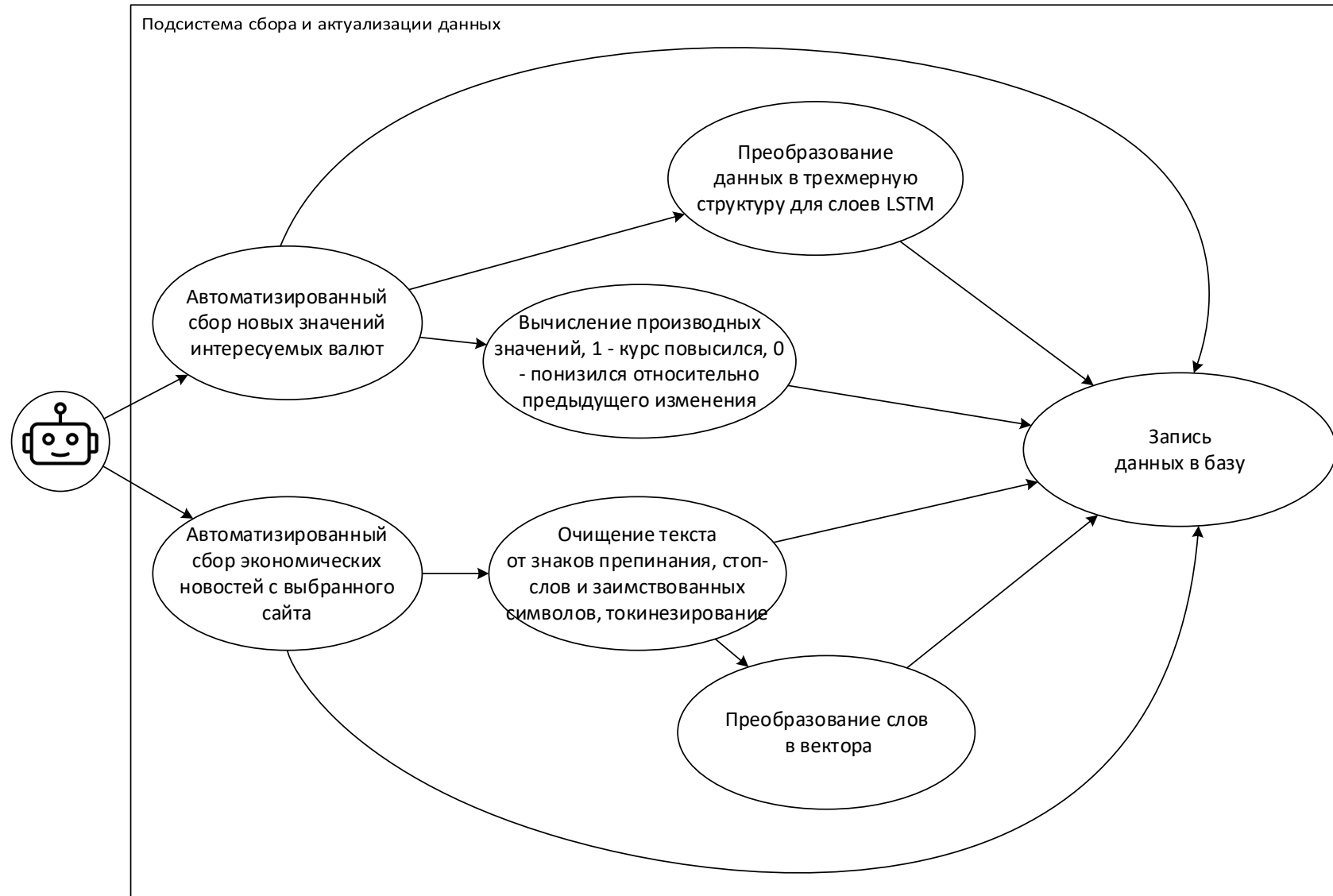


Диаграмма прецедентов подсистемы сбора и актуализации данных



Подсистема сбора и актуализации данных

Курсы валют на момент открытия биржи с сайта mfd.ru

	Date	USD	USD_ch	USD_ch_bi	EUR	EUR_ch	EUR_ch_bi	CNY	CNY_ch	CNY_ch_bi
0	11.01.2022	75.1315	NaN	NaN	85.1315	NaN	NaN	11.7907	NaN	NaN
1	12.01.2022	74.8355	-0.2960	0.0	84.8784	-0.2531	0.0	11.7446	-0.0461	0.0
2	13.01.2022	74.5277	-0.3078	0.0	84.6709	-0.2075	0.0	11.7088	-0.0358	0.0
3	14.01.2022	74.5686	0.0409	1.0	85.4556	0.7847	1.0	11.7246	0.0158	1.0
4	15.01.2022	75.7668	1.1982	1.0	86.8894	1.4338	1.0	11.9457	0.2211	1.0
...
318	22.04.2023	81.4863	-0.1325	0.0	89.3495	-0.1143	0.0	11.8111	-0.0359	0.0
319	25.04.2023	81.2745	-0.2118	0.0	89.4589	0.1094				
320	26.04.2023	81.5499	0.2754	1.0	90.0332	0.5743				
321	27.04.2023	81.6274	0.0775	1.0	90.1436	0.1104				
322	28.04.2023	81.5601	-0.0673	0.0	90.2023	0.0587				

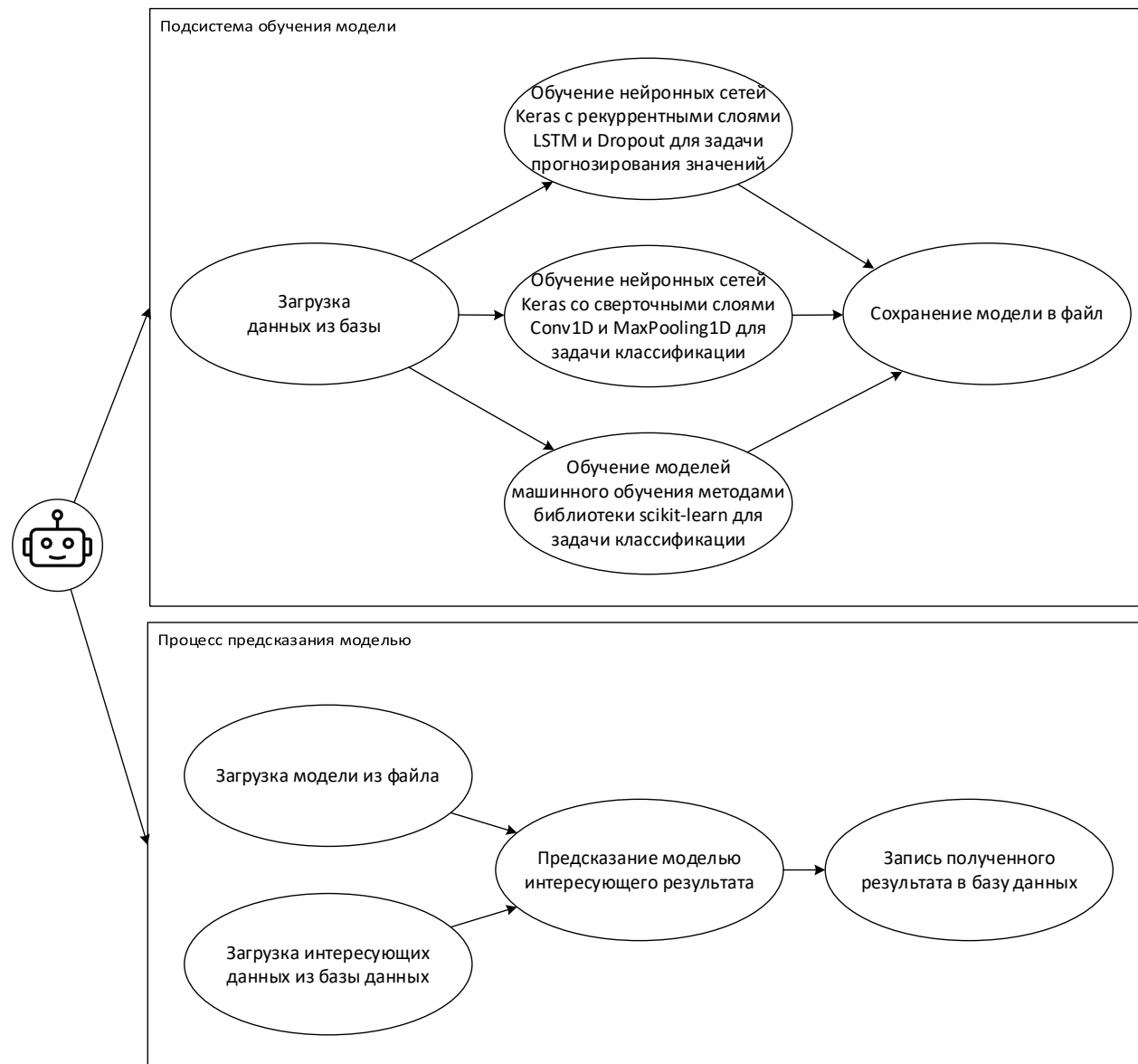
323 rows x 10 columns

Результат парсинга новостей по тегу экономика с сайта РИА Новости с использованием инструментов Selenium и Beautiful Soup

	date	views	name	url
0	01.01.2022	18238	На Украине ввели предельную надбавку к цене на...	https://ria.ru/20220101/gaz-1766394128.html
1	01.01.2022	2085	Запасы газа на Украине за год сократились почт...	https://radiosputnik.ria.ru/20220101/gaz-17663...
2	01.01.2022	15134	СМИ: часть крупного металлургического завода У...	https://ria.ru/20220101/zavod-1766379413.html
3	01.01.2022	32239	В Британии заявили о катастрофичном снижении у...	https://ria.ru/20220101/padenie-1766374961.html
4	01.01.2022	22886	Брюссель намерен присвоить "зеленый" статус ат...	https://ria.ru/20220101/energetika-1766368303....
...
25076	27.04.2023	45090	Россия организовала Западу жесткое приземление	https://ria.ru/20230427/prizemlenie-1867983358...
25077	27.04.2023	52676	Сработал эффект домино. В США поплыл еще один ...	https://ria.ru/20230427/bank-1867967185.html
25078	27.04.2023	1130	Аналитики рассказали, какую карту разыграет кр...	https://ria.ru/20230427/dividendy-1868007554.html
25079	27.04.2023	4177	В России почти полностью прекратились госзакуп...	https://ria.ru/20230427/windows-1868019655.html
25080	27.04.2023	9520	Российский экспорт в Индию впервые превысил 40...	https://ria.ru/20230427/eksport-1868017979.html

25081 rows x 4 columns

Диаграмма подсистемы прогнозирования



Используемые архитектуры нейросетей

Нейроны в нейронных сетях сгруппированы в слои разных типов, вот типовая архитектура свёрточной нейронной сети (CNN)

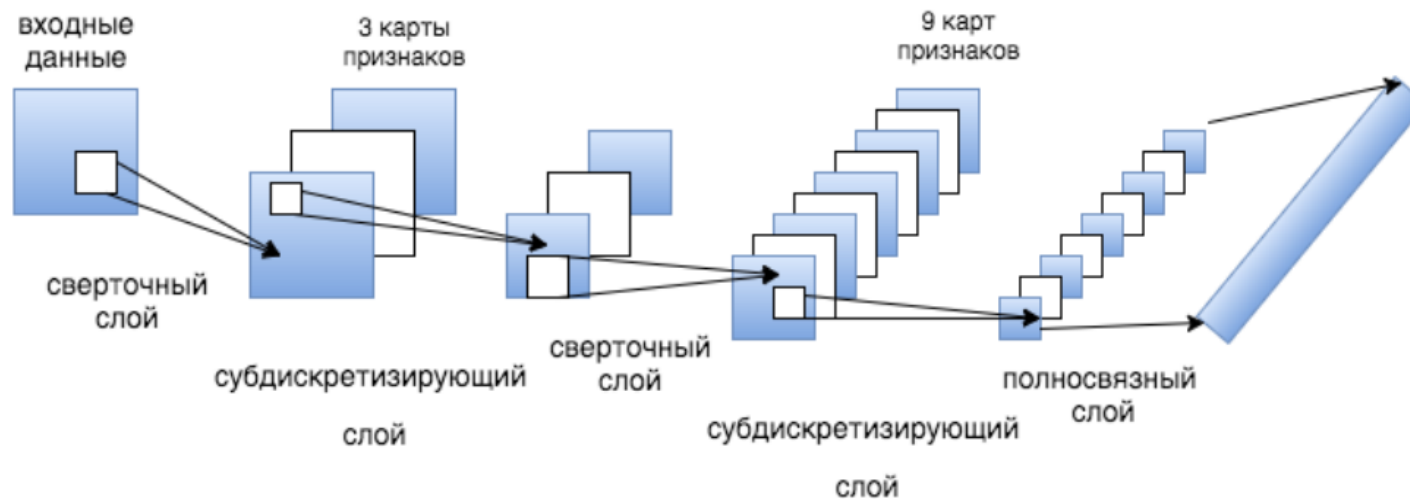
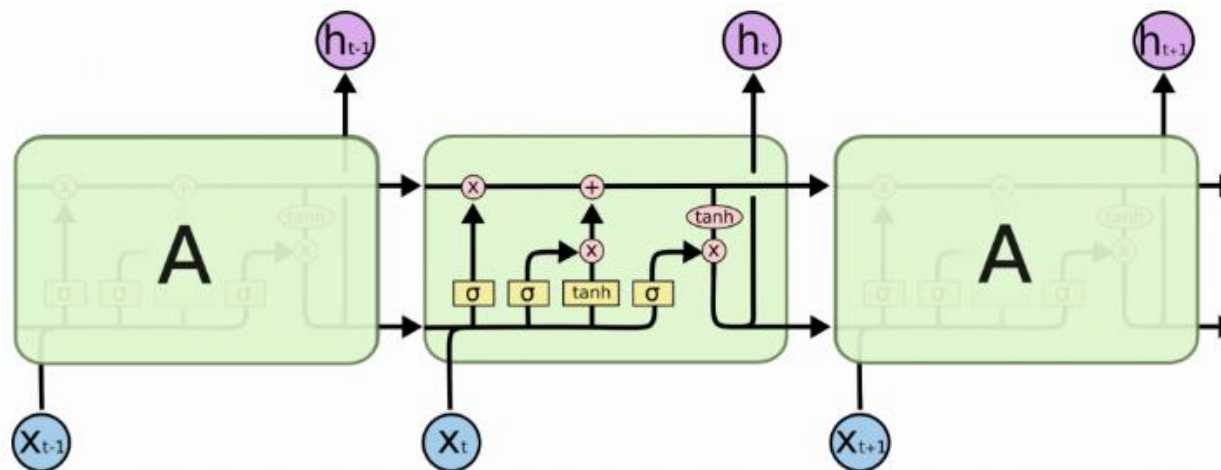
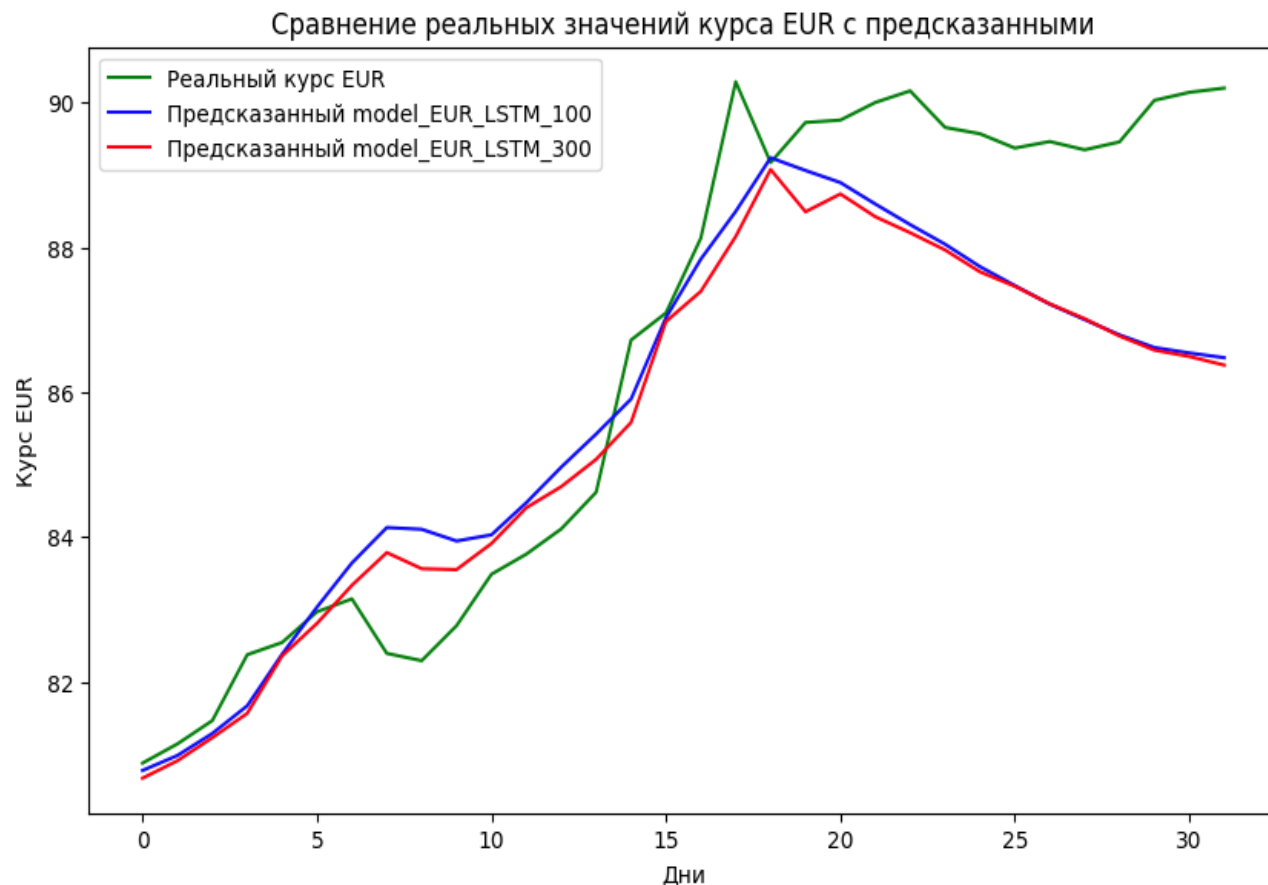


Схема нейронной сети долгой краткосрочной памяти (LSTM)



Прогнозирование курса валют как временного ряда (LSTM)



	USD_LSTM_100	USD_LSTM_300
explained_variance_score	0.818028	0.765815
mean_squared_log_error	0.000185	0.000222
mean_squared_error	1.212222	1.445675
root_mean_squared_error	1.101010	1.202362
mean_absolute_error	0.945215	1.013626
median_absolute_error	0.968950	0.909652
max_error	1.964436	2.377704

	EUR_LSTM_100	EUR_LSTM_300
explained_variance_score	0.819169	0.838483
mean_squared_log_error	0.000345	0.000352
mean_squared_error	2.718329	2.788356
root_mean_squared_error	1.648735	1.669837
mean_absolute_error	1.270622	1.289954
median_absolute_error	0.858379	1.079251
max_error	3.724608	3.827331

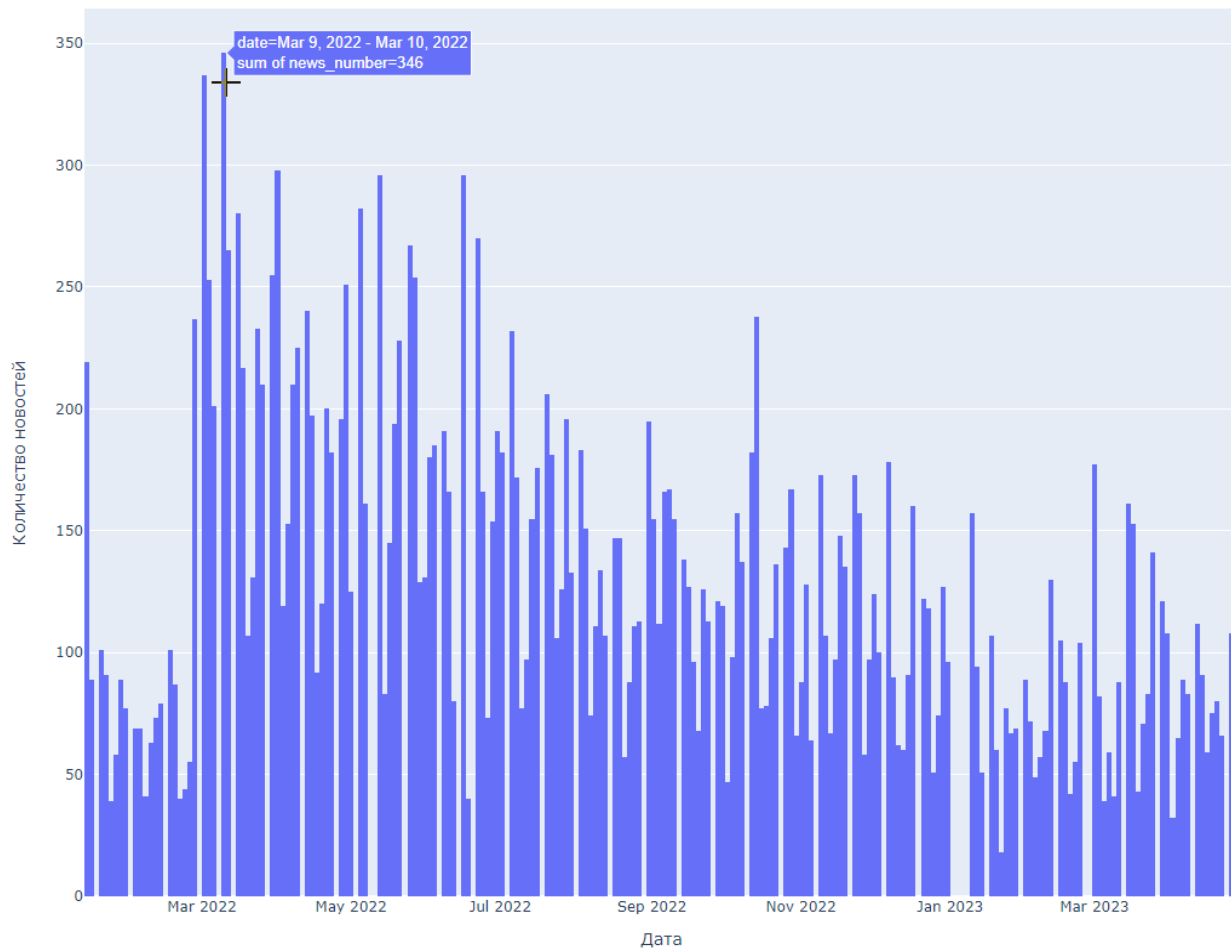
	CNY_LSTM_100	CNY_LSTM_300
explained_variance_score	0.799693	0.812195
mean_squared_log_error	0.000150	0.000139
mean_squared_error	0.023393	0.021708
root_mean_squared_error	0.152947	0.147337
mean_absolute_error	0.131409	0.129307
median_absolute_error	0.136820	0.135501
max_error	0.254086	0.263512

Предсказание направления изменения курса валют по новостям (предобработка)

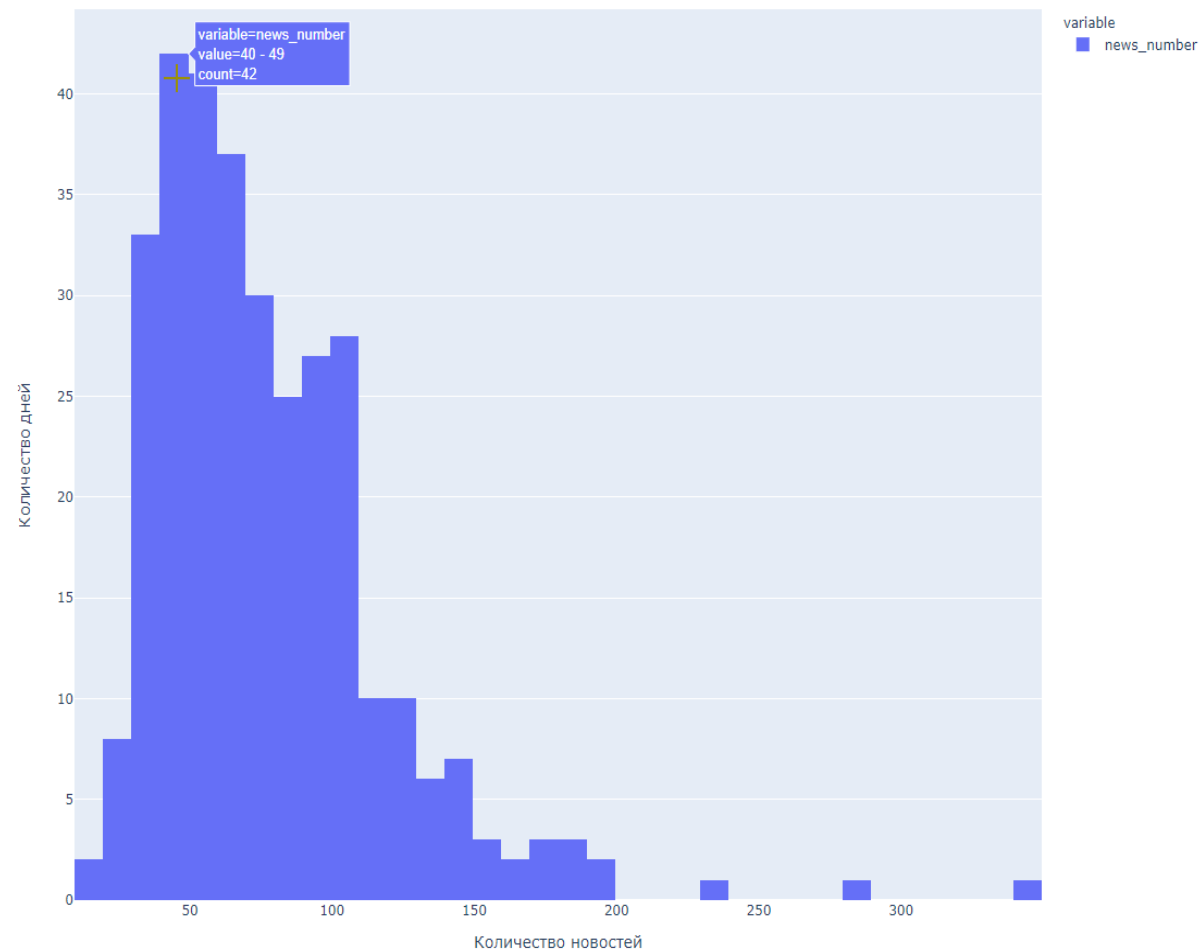
	date	news_number	top1	top2	top3	top4	top5	top6	top7	top8	...	top31
0	12.01.2022	189	ДНР начала поставлять в Россию колбасу	Пушков назвал условие для конца транзита газа ...	Daily Express: Путин переиграл Европу в ситуац...	Не "Северным потоком" единым. Что еще с размах...	Россия не обязана поставлять весь газ в Европу...	На Украине допустили "воровство" российского г...	Запад встревожен из-за сделки России и Китая п...	Байден дал возможность России зарабатывать три...	...	Fitch рассказало, как "Сила Сибири — 2" может ...
1	13.01.2022	30	Путин предложил проиндексировать пенсии выше и...	Покупать или подождать: что будет с ценами на ...	Допрасходы на индексацию пенсий в 2022 году со...	Аналитик назвал срок, на который имеет смысл п...	Премьер Молдавии рассказала о переговорах с "Г...	"Молдовагаз" предупредил о рисках прекращения ...	Молдавия попросит "Газпром" об отсрочке оплаты...	Продажи обновленной Lada Vesta начнутся к сере...	...	NaN
2	14.01.2022	42	В ФРГ назвали истинную причину задержки с запущ...	В Госдуму внесли законопроект о закрытии дел д...	Путин увел Россию с "тупого" пути развития, за...	Греф заявил о готовности к возможным санкциям ...	Курс рубля снизился после заявления Рябкова	В Литву прибыл новый танкер со сжиженным приро...	Шторм на рынке труда: у кого самые высокие зар...	Боррель высказался о сертификации "Северного п...	...	Бюджет по итогам 2021 года исполнен с профицит...
3	15.01.2022	47	В Польше признали потери от "капризной" борьбы...	WSJ: Европа откажется от жестких санкций проти...	"Газпром" подал в суд на польскую компанию PGNiG	Мясо, овощи, хлеб: какие еще продукты вскоре п...	Иллюзии об экономическом сотрудничестве с Запа...	Первый зампред ЦБ Швецов покинет пост	"Газпром" сообщил о ремонте газопровода "Валда...	Цены на недвижимость в Ливане упали на 70%	...	Правительство готовит план поддержки системы з...
4	18.01.2022	58	Китай украл у Америки самое ценное	Поворот на Север: Россия отказалась выбирать м...	"Пора спуститься с небес на землю": поляки под...	Захарова оценила запрос Германии к Нидерландам...	Из списка поручений Путина уберут пункты по за...	Холодильник с телевизором схлестнулись в смерт...	"Газпром" не забронировал мощности для транзит...	По тормозам: когда ждать падения цен на машины	...	Додон предложил меры по выходу Молдавии из газ...
...
317	22.04.2023	35	Forbes опубликовал рейтинг самых обедневших ми...	США решили снова подключить Россию к SWIFT. Эк...	Выходят из-под контроля. Украина снова подстав...	СМИ: запрет на экспорт в Россию авто отразится...	Сбербанк по итогам 2022 года выплатит рекордные...	Посол Украины в Турции призвал страну искать а...	Банк ТВС предупредил об изменениях в связи с а...	В Ленинградской области рассказали о работе за...	...	Иранские бизнесмены предложили создать в Астра...
318	25.04.2023	61	Медведев прокомментировал идею G7 о запрете эк...	СМИ: ЦБ призвал коллекторов и банки скоординир...	Силуанов объяснил, почему Россия перешла на ра...	СМИ: Минфин и Центробанк завершили проект о до...	В Южной Осетии назвали размер долга перед Росс...	Песков считает, что Россия будет развиваться т...	Министры сельского хозяйства G7 призвали к див...	Henkel рассчитывает на обратный выкуп активов	Эксперт оценил запасы оборудования в попавших ...

Гистограммы распределения количества новостей

Количество новостей каждый день



Количество дней для количества новостей



Предсказание направления изменения курса валют по новостям

Результаты применения стандартных моделей машинного обучения библиотеки scikit-learn к задаче классификации

pred_res_USD_bi_top18

	Models - USD_ch_bi - top18	Accuracy	Precision	Recall	F1 Score
0	Logistic Regression	61.538462	62.500000	60.606061	61.538462
1	Support Vector Classifier	63.076923	59.574468	84.848485	70.000000
2	Random Forest	49.230769	50.000000	63.636364	56.000000
3	Gradient Boosting	49.230769	50.000000	60.606061	54.794521
4	K-Nearest Neighbors	63.076923	62.162162	69.696970	65.714286
5	Decision Tree	61.538462	62.500000	60.606061	61.538462

pred_res_USD_bi_top40

	Models - USD_ch_bi - top40	Accuracy	Precision	Recall	F1 Score
0	Logistic Regression	64.615385	64.705882	66.666667	65.671642
1	Support Vector Classifier	70.769231	65.909091	87.878788	75.324675
2	Random Forest	66.153846	67.741935	63.636364	65.625000
3	Gradient Boosting	66.153846	66.666667	66.666667	66.666667
4	K-Nearest Neighbors	49.230769	50.000000	33.333333	40.000000
5	Decision Tree	63.076923	66.666667	54.545455	60.000000

pred_res_CNY_bi_top18

	Models - CNY_ch_bi - top18	Accuracy	Precision	Recall	F1 Score
0	Logistic Regression	60.000000	60.869565	45.161290	51.851852
1	Support Vector Classifier	50.769231	42.857143	9.677419	15.789474
2	Random Forest	47.692308	28.571429	6.451613	10.526316
3	Gradient Boosting	61.538462	66.666667	38.709677	48.979592
4	K-Nearest Neighbors	52.307692	50.000000	45.161290	47.457627
5	Decision Tree	46.153846	43.750000	45.161290	44.444444

pred_res_CNY_bi_top40

	Models - CNY_ch_bi - top40	Accuracy	Precision	Recall	F1 Score
0	Logistic Regression	58.461538	56.666667	54.838710	55.737705
1	Support Vector Classifier	63.076923	76.923077	32.258065	45.454545
2	Random Forest	61.538462	80.000000	25.806452	39.024390
3	Gradient Boosting	69.230769	72.000000	58.064516	64.285714
4	K-Nearest Neighbors	53.846154	51.428571	58.064516	54.545455
5	Decision Tree	60.000000	60.869565	45.161290	51.851852

Предсказание направления изменения курса валют по новостям (USD top-18)

Сводное представление модели со сверточными (Conv1D) слоями и результаты её работы

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 139, 2076)	20760000
dropout (Dropout)	(None, 139, 2076)	0
conv1d (Conv1D)	(None, 139, 64)	664384
max_pooling1d (MaxPooling1D)	(None, 69, 64)	0
dropout_1 (Dropout)	(None, 69, 64)	0
conv1d_1 (Conv1D)	(None, 69, 64)	28736
max_pooling1d_1 (MaxPooling1D)	(None, 34, 64)	0
dropout_2 (Dropout)	(None, 34, 64)	0
global_average_pooling1d (GlobalAveragePooling1D)	(None, 64)	0
dropout_3 (Dropout)	(None, 64)	0
dense (Dense)	(None, 2)	130

Total params: 21,453,250
Trainable params: 21,453,250
Non-trainable params: 0

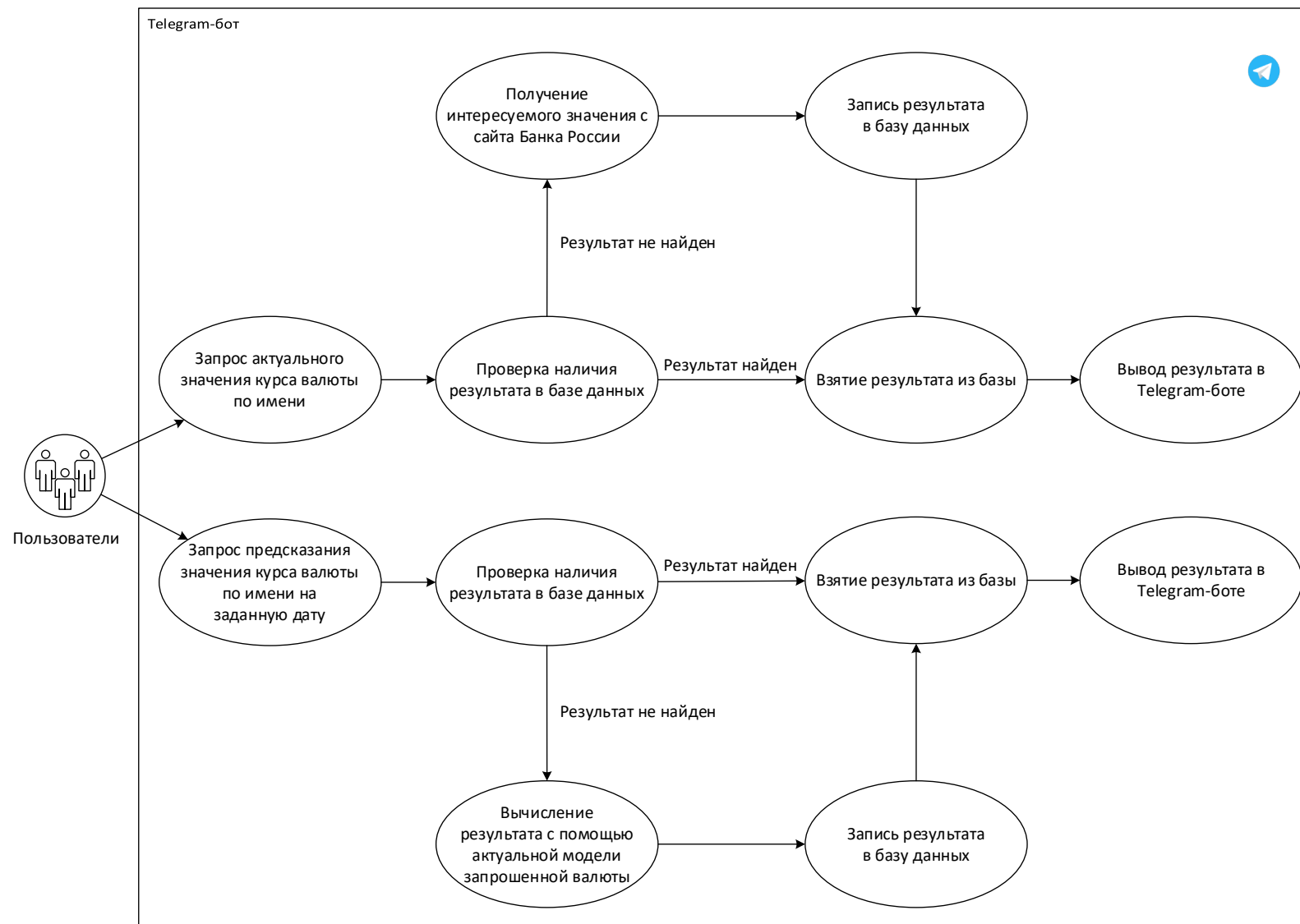


Epoch 29/30
26/26 [=====] - 10s 382ms/step - loss: 0.0022
Epoch 30/30
26/26 [=====] - 10s 381ms/step - loss: 0.0023
3/3 [=====] - 0s 65ms/step

Results
Accuracy 0: 0.5846

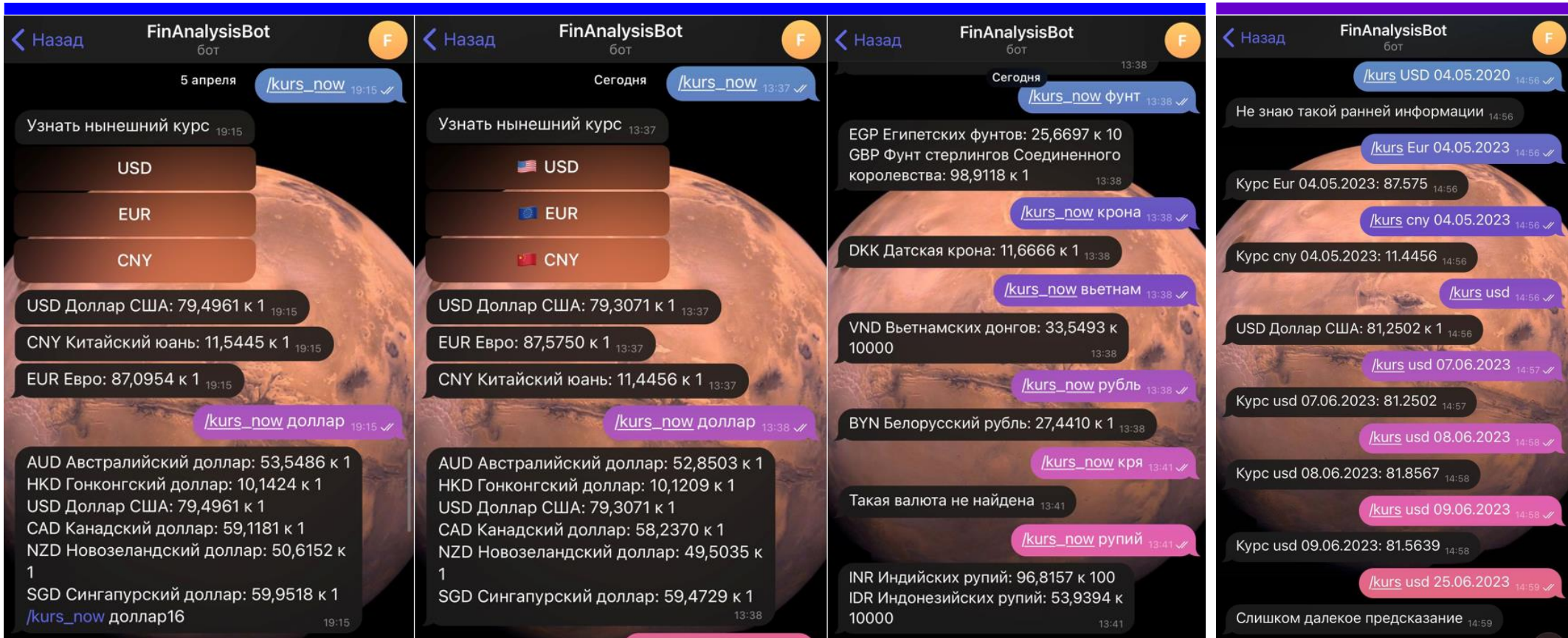
	precision	recall	f1-score	support
0	0.57	0.62	0.60	32
1	0.60	0.55	0.57	33
accuracy			0.58	65
macro avg	0.59	0.59	0.58	65
weighted avg	0.59	0.58	0.58	65

Диаграмма подсистемы взаимодействия с пользователем через Telegram-бот



Подсистема взаимодействия с пользователем через Telegram-бот

Тестирование возможностей работы команд `kurs_now` и `kurs` в боте



5 апреля

4 мая

7 июня

Были получены следующие результаты:

- были созданы модели машинного обучения и нейросетевые модели для предсказания курса наиболее актуальных валют к российскому рублю и для предсказания курса акций нескольких крупных компаний;
- была разработана архитектура системы финансового информирования;
- был создан прототип системы финансового информирования, включающий:
 - автоматизированный сбор интересующих данных;
 - прогнозирование курса валют нейросетями с LSTM слоями;
 - Telegram-бот в качестве пользовательского интерфейса системы.

Апробация работы

- ◆ Часть результатов работы была представлена на конференции ITTMM 2023 (19 апреля 2023 г.).
- ◆ Было получено свидетельство о государственной регистрации программы для ЭВМ.

Программа «Система финансового информирования FinAnalysisBot», № 2023662173.

Спасибо за внимание