



B. D. Science

**Задача 15. Модель раннего обнаружения неисправностей
промышленного оборудования**



Команда «B. D. Science»



**Пётр
Ларин**

- Капитан команды
- Data Scientist
- <https://t.me/petrlarin>



**Светлана
Хорольская**

- Data Scientist
- https://t.me/IO_v_OI



**Игорь
Шахматов**

- Data Scientist
- https://t.me/Igor_Shakhmatov_DS



**Марина
Запорожец**

- Data Scientist
- <https://t.me/mazavlia>



**Любовь
Ильина**

- Data Scientist
- https://t.me/Lyubov_Ilyina_Sterkhova



Задачи проекта

- 1 **Определить наличие или отсутствие неисправности M1 на заданных интервалах тестовой выборки**
- 2 **Определять периоды, когда были любые неисправности M3 на протяжении всего тестового интервала.**
- 3 **Определить время до простоя M1 с максимально возможным горизонтом**

Наша гипотеза состоит в том, что, незадолго до наступления аварии М1, на графиках показателей соответствующей машины будут появляться характерные паттерны (сигнатуры), которые сможет распознать одномерная сверточная нейронная сеть.

Для обучения CNN мы подготовили датасет, элементами которого являются фрагменты X_train. Каждый такой фрагмент имеет 16 столбцов, соответствующих показателям одной из машин, на интервале в 10 минут (61 строка).

Один элемент датасета

DT	ЭКСГАУСТЕР 9. ТОК РОТОРА 1	ЭКСГАУСТЕР 9. ТОК РОТОРА 2	ЭКСГАУСТЕР 9. ТОК СТАТОРА	ЭКСГАУСТЕР 9. ДАВЛЕНИЕ МАСЛА В СИСТЕМЕ	ЭКСГАУСТЕР 9. ТЕМПЕРАТУРА ПОДШИПНИКА НА ОПОРЕ 1	ЭКСГАУСТЕР 9. ТЕМПЕРАТУРА ПОДШИПНИКА НА ОПОРЕ 2	ЭКСГАУСТЕР 9. ТЕМПЕРАТУРА ПОДШИПНИКА НА ОПОРЕ 3	ЭКСГАУСТЕР 9. ТЕМПЕРАТУРА ПОДШИПНИКА НА ОПОРЕ 4	ЭКСГАУСТЕР 9. ТЕМПЕРАТУРА МАСЛА В СИСТЕМЕ	ЭКСГАУС ТЕМПЕРА1 МАСЛ МАСЛОБЛ
2019-03-19 14:00:10	0.902000	0.902000	0.475000	337.130000	46.240000	51.920000	40.560000	54.111499	21.174060	39.34
2019-03-19 14:00:20	0.738000	0.738000	0.475000	332.360000	45.430000	50.659107	39.932358	54.350000	19.470000	39.75
2019-03-19 14:00:30	0.902000	0.902000	0.520000	328.288000	46.240000	50.785304	39.752081	53.624738	22.720000	40.56
2019-03-19 14:00:40	0.683333	0.683333	0.460000	324.654000	45.998824	50.704178	40.076582	54.273755	21.900000	38.94
2019-03-19 14:00:50	0.820000	0.820000	0.520000	320.996000	45.430000	50.785304	39.995457	53.624739	21.336311	39.75
...
2019-03-19 14:09:30	0.820000	0.820000	0.390000	180.237500	45.430940	44.051788	37.805041	49.811786	19.470000	37.32
2019-03-19 14:09:40	1.025000	1.025000	0.350000	158.704000	45.430940	43.537986	38.039412	48.680000	21.090000	38.94
2019-03-19 14:09:50	0.820000	0.820000	0.346667	155.323333	45.512066	43.646154	38.130000	49.490000	20.768424	38.13
2019-03-19 14:10:00	19.515000	19.515000	0.376667	155.380000	45.268687	43.646154	37.967298	49.649531	21.092932	38.53
2019-03-19 14:10:10	331.414000	331.414000	0.416000	155.596667	45.349814	43.810000	37.805041	49.325025	20.687299	38.94

61 rows × 16 columns

[illegible]

В датасет попадают все фрагменты, содержащие аварии М1 с **меткой класса “1”** и все фрагменты, предшествующие авариям М1 на протяжении 3 часов – с **меткой класса “2”**. К ним добавляются избранные фрагменты, изолированные от аварий, с **меткой “0”** в количестве, необходимом для сбалансированного датасета.

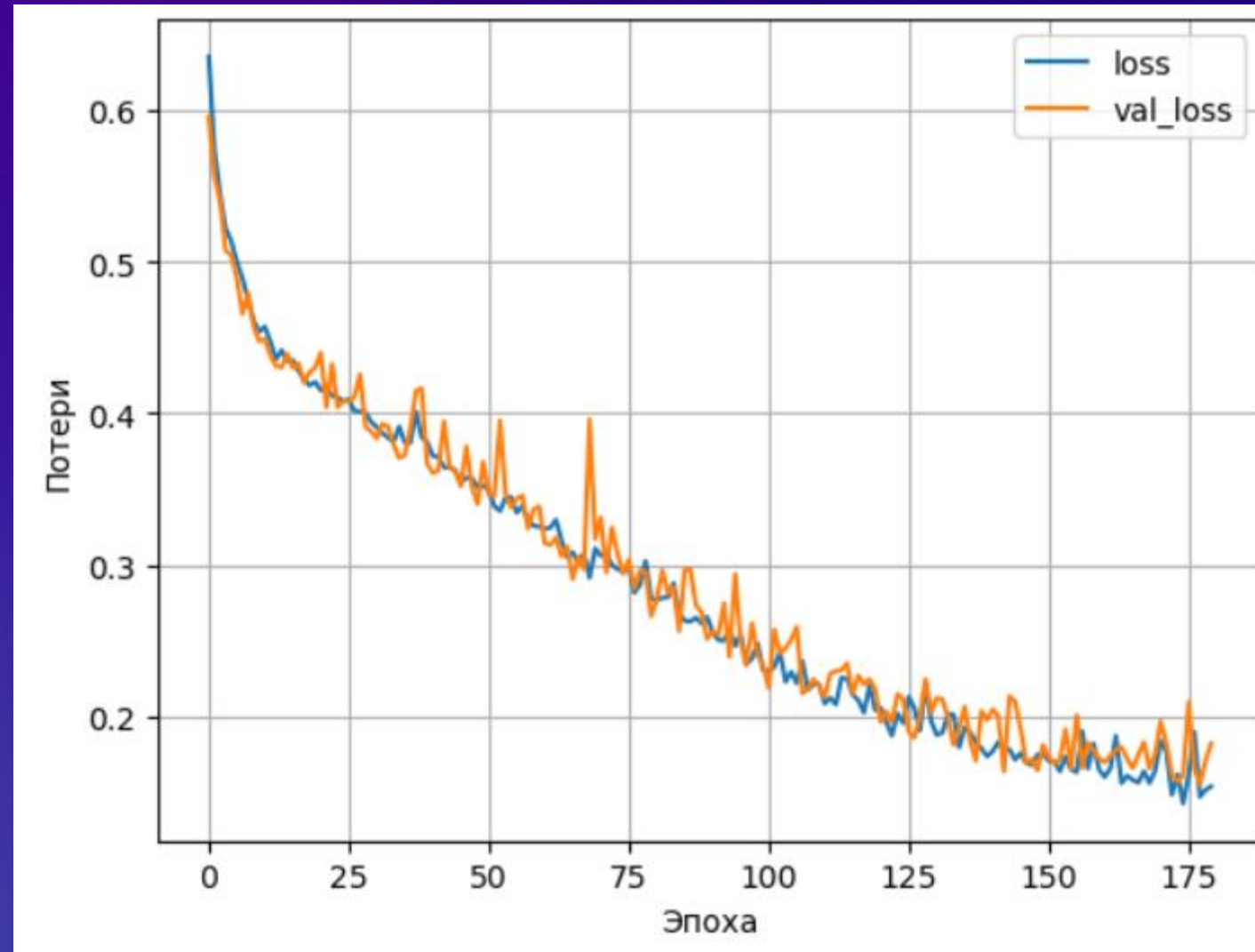


Архитектура CNN

```
model_m = Sequential()  
model_m.add(Conv1D(80, 6, activation='relu',  
input_shape=(61, 16)))  
model_m.add(Conv1D(80, 6, activation='relu'))  
model_m.add(GlobalAveragePooling1D())  
model_m.add(Dropout(0.23))  
model_m.add(Dense(160, activation='relu'))  
model_m.add(Dense(3, activation='softmax'))
```

```
model_m.compile(loss=  
'categorical_crossentropy', optimizer='adam',  
metrics=['accuracy'])
```

F1-score у данной CNN ~ 0.83-0.91



Epoch 180/180 84/84 - 1s - loss: 0.1536 - accuracy: 0.9423 - val_loss: 0.1820 - val_accuracy: 0.9412 - 1s/epoch - 14ms/step



На вход обученной CNN подаются фрагменты из X_{test} , предшествующие пропускам на интервале в 3 часа (18 интервалов). Таким образом, на каждый пропуск мы получаем 18 меток класса, из которых методом голосования определяется наличие аварии в пропуске.

Данный подход не предусматривает определения конкретного технического места.

Предсказания модели содержатся в файле **submission_1.xlsx**.



Задача

Определять периоды, когда были любые неисправности МЗ (аномальный режим работы техместа) на протяжении всего тестового интервала.

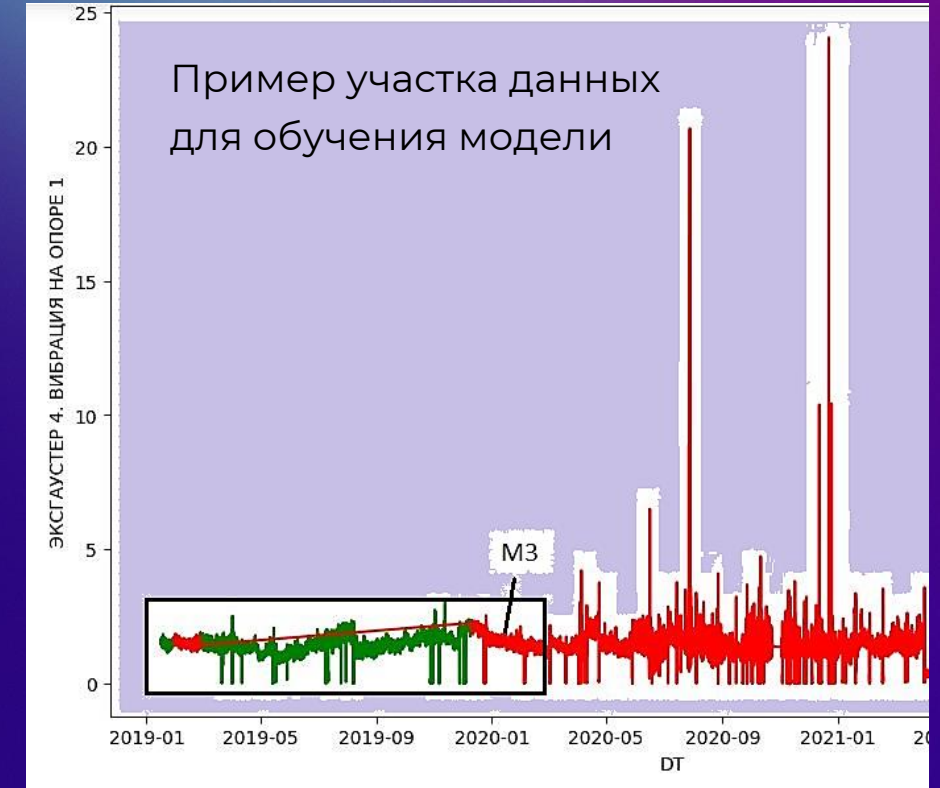
Идея решения

- Определить параметры, влияющие на целевые переменные
- Разработать модель на основе сверточных нейронных сетей и сетей LSTM для прогнозирования МЗ

Формирование выборок для подзадачи №2

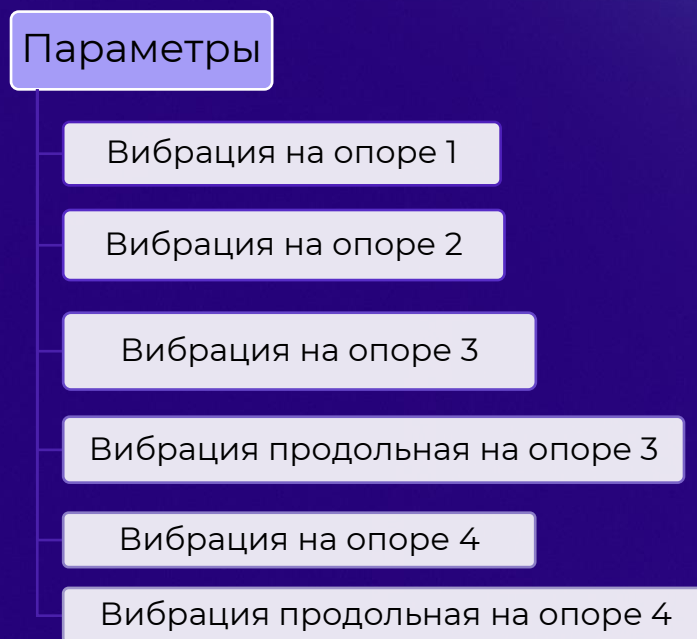
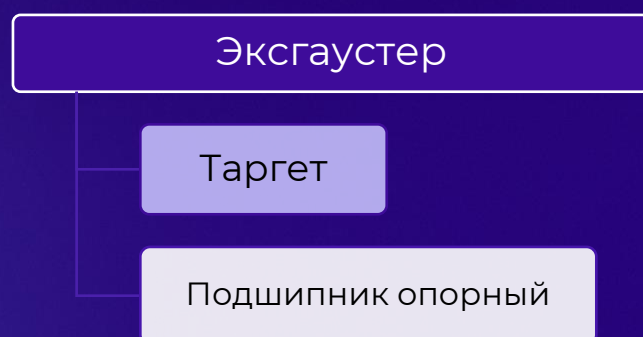
После установления периодов с авариями типа M1, они исключаются или подбираются обучающие данные, не содержащие M1 для таргетов всего эксгаустера

Из данных исключаются интервалы, не содержащие информации о дате устранения неисправности



Формирование выборок для подзадачи №2

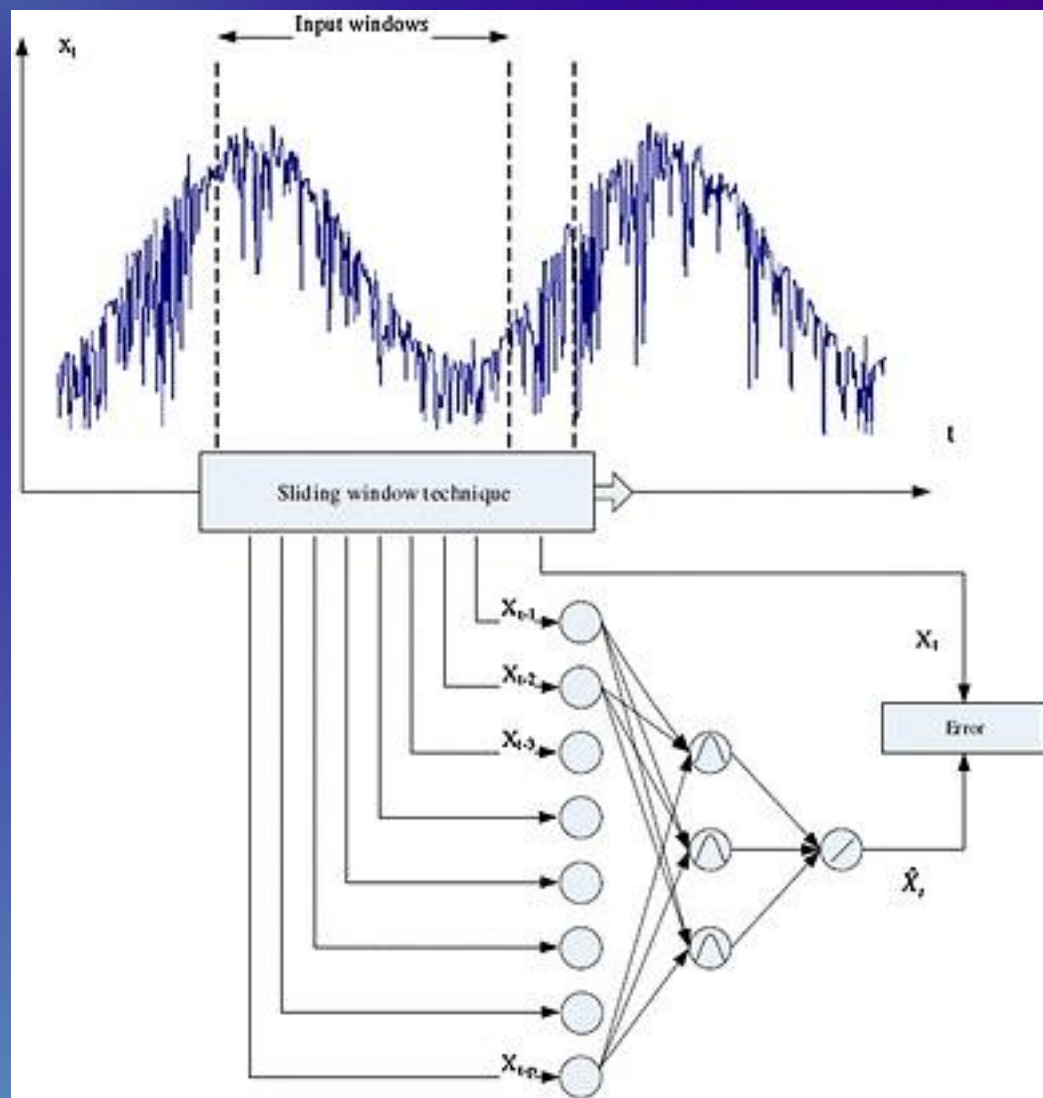
Пример отбора комбинации параметров и соответствующих прогнозируемых переменных



Данные преобразуются в последовательность в виде скользящего окна прогнозирования



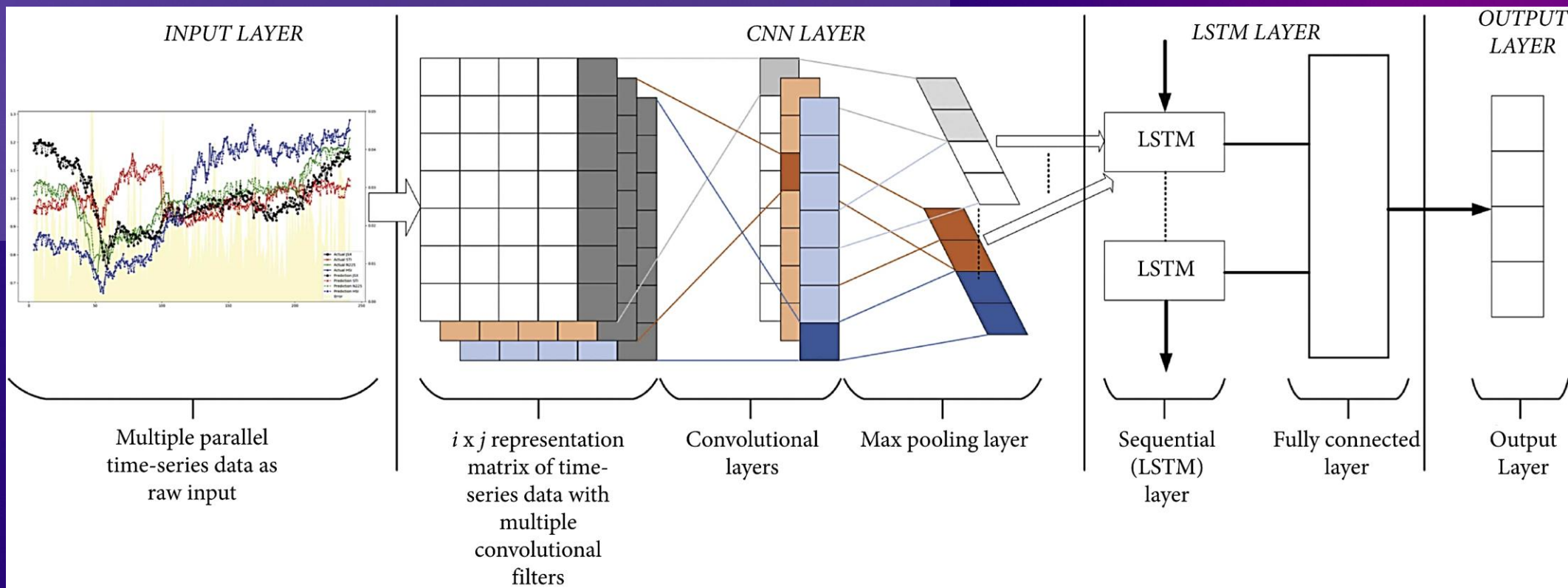
Реализация подхода скользящими окнами



Параметры окна

- Размер
- Сдвиг
- Метка окна (0, 1, 2)

Нейронная сеть для подхода №2





Архитектура нейронной сети для подхода №2



Тип слоя	Название слоя	Количество нейронов	Параметры слоя	Функция активации
Сверточный слой	Conv1D	32	размер фильтра - 5	relu
Сверточный слой	Conv1D	32	размер фильтра - 5	relu
Пулинг-слой	MaxPooling1D	–	размер окна - 3	–
LSTM-слой	LSTM	64	–	–
Слой dropout регуляризации	Dropout	–	коэфф. отключения – 0.2	–
Полносвязный слой	Dense	64	–	–
Полносвязный слой для классификации	Dense	3	–	softmax



1

Метрики CNN по задаче №1

Accuracy ~0.95

F1-score ~ 0.83-0.91

2

По задаче № 2 необходимо дальнейшее развитие модели, а также подбор влияющих параметров и релевантных данных



СПАСИБО ЗА ВНИМАНИЕ!