

Научная статья
УДК 004.896:662.27
DOI 10.35266/1999-7604-2023-4-6

НЕЙРОСЕТЕВОЕ ПРОГНОЗИРОВАНИЕ ВХОДНЫХ ПАРАМЕТРОВ ПРИ ДОБЫЧЕ НЕФТИ

Евгений Дмитриевич Семёнов¹, Михаил Яковлевич Брагинский²,
Дмитрий Викторович Тараканов^{3✉}, Инесса Леонидовна Назарова⁴

¹ Сургутнефтегаз, Сургут, Россия

^{2,3,4} Сургутский государственный университет, Сургут, Россия

¹ semunoff90@gmail.com, <https://orcid.org/0009-0007-2994-4119>

² braginskiy_mya@surgu.ru, <https://orcid.org/0000-0003-1332-463X>

³ sprtdv@mail.ru ✉, <https://orcid.org/0000-0003-1851-1039>

⁴ nazarova_il@surgu.ru, <https://orcid.org/0009-0001-0340-2609>

Аннотация. В статье рассматривается применение искусственных нейронных сетей для прогнозирования технологических параметров добычи нефти. Для решения задачи были использованы аппараты искусственных нейронных сетей на основе архитектуры долгой краткосрочной памяти (LSTM) и управляемых рекуррентных блоков (GRU). Полученные результаты нейросетевого прогнозирования подтвердили эффективность применения рекуррентных нейронных сетей, особенно модели LSTM для прогнозирования временных рядов.

Ключевые слова: нейронные сети, добыча нефти, машинное обучение, LSTM, GRU, рекуррентная нейронная сеть, прогнозирование

Для цитирования: Семёнов Е. Д., Брагинский М. Я., Тараканов Д. В., Назарова И. Л. Нейросетевое прогнозирование входных параметров при добыче нефти // Вестник кибернетики. 2023. Т. 22, № 4. С. 42–51. DOI 10.35266/1999-7604-2023-4-6.

Original article

NEURAL NETWORK FORECASTING OF INPUT PARAMETERS IN OIL DEVELOPMENT

Evgeny D. Semyonov¹, Mikhail Ya. Braginsky², Dmitry V. Tarakanov^{3✉},
Inessa L. Nazarova⁴

¹ Surgutneftegas, Surgut, Russia

^{2,3,4} Surgut State University, Surgut, Russia

¹ semunoff90@gmail.com, <https://orcid.org/0009-0007-2994-4119>

² braginskiy_mya@surgu.ru, <https://orcid.org/0000-0003-1332-463X>

³ sprtdv@mail.ru ✉, <https://orcid.org/0000-0003-1851-1039>

⁴ nazarova_il@surgu.ru, <https://orcid.org/0009-0001-0340-2609>

Abstract. The article examines use of artificial neural networks for forecasting of technological parameters of oil development. Artificial neural networks based on the long short-term memory architecture and gated recurrent units are used to solve the problem. The findings of neural network forecasting prove the efficacy of recurrent neural networks, especially the long short-term memory one, for forecasting of time series.

Keywords: neural networks, oil development, machine learning, LSTM, GRU, recurrent neural network, forecast

For citation: Semyonov E. D., Braginsky M. Ya., Tarakanov D. V., Nazarova I. L. Neural network forecasting of input parameters in oil development. *Proceedings in Cybernetics*. 2023;22(4):42–51. DOI 10.35266/1999-7604-2023-4-6.

ВВЕДЕНИЕ

Текущие методы разработки месторождений нефти и газа в естественных условиях позволяют добывать эти ресурсы лишь в течение 3–5 лет. Это очень короткий срок, поэтому для увеличения нефтеотдачи используются методы искусственного воздействия на пласт. Одним из таких методов является закачка воды. Со временем пластовое давление снижается, что приводит к снижению дебитов скважин. Нагнетаемая вода восполняет эти потери, восстанавливая и поддерживая давление. Это, в свою очередь, увеличивает нефтеотдачу, так как вода вытесняет нефть из пласта. При длительной закачке вода может вымыть оставшиеся частицы нефти [1]. Расчеты для регулирования подачи нагнетающей жидкости зависят от множества факторов, включая количество жидкости и давление закачки, коэффициенты охвата пласта заводнением и коэффициенты вытеснения. Эти коэффициенты, в свою очередь, зависят от данных, полученных в процессе добычи [2, 3], поэтому сложно точно оценить предстоящие затраты на нагнетание жидкости в следующем периоде времени.

Параметры, характеризующие процесс добычи нефти и газа, а также результаты добычи за прошедшее время можно использовать для обучения нейронной сети. Нейронная сеть может использовать эти данные при построении прогноза для принятия решений о регулировании добычи нефти.

В настоящее время управление добычей нефти и газа основывается на результатах, полученных в прошлом. Этот метод эффективен для каждого конкретного месторождения, но он не учитывает изменения пластовых условий.

Цели и задачи представленного исследования заключаются в создании искусственной

нейронной сети (ИНС) для прогнозирования параметров добычи нефти, исследовании точности прогноза в зависимости от периода добычи и в изучении эффективности использования рекуррентной нейронной сети в решении задачи экстраполяции [4–7].

МАТЕРИАЛЫ И МЕТОДЫ

Один из подходов к прогнозированию временных рядов, включая многофакторные, является использование рекуррентных нейронных сетей (РНС), таких как LSTM (Long Short-Term Memory). РНС шаг за шагом обрабатывает последовательность данных, сохраняя внутреннее состояние, полученное при обработке предыдущих элементов. В данном случае сеть LSTM может быть обучена на многофакторных временных рядах добычи нефти, где пользователь сам выбирает показатели, после чего модель обучается и выдает результаты прогнозирования.

При реализации рекуррентных нейронных сетей были использованы промышленные данные с трех скважин на Федоровском месторождении: одна из них является нефтяной скважиной, а две другие – нагнетательными. Из доступных параметров были выбраны количество закаченной нагнетательной жидкости в час и давление закачки. Прогнозируемым параметром является дебит нефти в час, который имеет высокую практическую значимость для оценки производительности и эффективности месторождения.

Для обучения модели было использовано 10 тысяч записей, что соответствует более чем 416 дням данных. Это позволяет модели учитывать разнообразные сезонные и долгосрочные паттерны, которые могут повлиять на производство и добычу нефти.

Структура данных представлена в табл. 1.

Таблица 1

Структура промышленных данных

Временной ряд	Единица измерения
Количество закаченной воды в нагнетательную скважину (для двух скважин)	Баррелей в час
Давление в скважине (для двух скважин)	Бар
Дебит нефти	Баррелей в час

Примечание: составлено авторами.

Тестовые данные представлены в виде временных рядов и записаны в файл формата CSV с разделителем – запятой.

Для обучения сети вначале были подготовлены данные с помощью библиотеки Python «pandas». После загрузки данные были подвёрнуты масштабированию так, чтобы они представляли рациональные числа от 0 до 1, так как этого требует нейронная сеть.

Для обучения предоставленные данные были преобразованы в принятую LSTM форму [8]. Необходимо было преобразовать

данные в трехмерный формат, где первое измерение – это количество записей или строк в наборе данных, которое в нашем случае равно 10 000, второе – это количество временных шагов, равное 24, и последнее – количество индикаторов, равное 5.

Реализованная модель LSTM (рис. 1, 2) является последовательной моделью с несколькими слоями. Модель состоит из четырех слоев LSTM, за которыми следует плотный слой, предсказывающий дебет нефти.

```
#Загрузка данных из файла
data = pd.read_csv(r'C:\...\dataOil.csv')
#Масштабирование
scaler = MinMaxScaler(feature_range=(0, 1))
data = scaler.fit_transform(data)
#Деление данных на обучающую выборку и тестовые данные
train_size = int(len(data) * 0.6)
test_size = len(data) - train_size
train, test = data[0:train_size, :],
data[train_size:len(data), :]
look_back = 5
trainX, trainY = createDataset(train, look_back)
testX, testY = createDataset(test, look_back)

#Массив моделей
models = [CreateNetworkAI(5, look_back, 1, 100, 0.2),
...
CreateNetworkAI(5, look_back, 1, 1000, 0.2)]
#Названия моделей
nameModel = ["LSTM 1, 100, 0.2", ..... "LSTM 1, 1k, 300"]
#Кол-во эпох для каждой модели
countsEpoch = [100, ....., 300]
testPredicts = []

for index, value in enumerate(models):
    #Обучение модели
    model.fit(trainX, trainY, epochs=countsEpoch[index], verbose=2)
    #Прогноз на тестовых данных
    testPredict = model.predict(testX)
    testPredicts.append(testPredict)

    #Обратное масштабирование полученных данных
    testYunScaled = scaler.inverse_transform(testY)
    testPredictunScaled = scaler.inverse_transform(
        dataForUnScale(testPredict))

    #Вывод графика
    CreatePlotResult(testYunScaled, testPredictunScaled,
        ['P1', 'W1', 'P2', 'W2', 'Oil'],
        nameModel[index], 0, 508)
    #Использование метрик для полученных результатов
    metrics(testYunScaled[:, 4], testPredictunScaled[:, 4])
```

Рис. 1. Листинг 1. Модель ИНС LSTM

Примечание: составлено авторами.

```
def CreateNetworkAI(CountInput, LookBach,
                    CountLongMemLayer, CountUnits, dropOut):
    CountLayFlag = CountLongMemLayer - 1
    model = Sequential()
    model.add(LSTM(units=CountUnits, return_sequences=True,
                   input_shape=(LookBach, CountInput)))
    model.add(Dropout(dropOut))

    while CountLayFlag:
        model.add(LSTM(units=CountUnits,
                       return_sequences=(CountLayFlag - 1 != 0)))
        model.add(Dropout(dropOut))
        CountLayFlag -= 1

    model.add(Dense(units=1))
    model.compile(optimizer='adam', loss='mean_squared_error')

    return model
```

Рис. 2. Листинг 2. Модель ИНС LSTM

Примечание: составлено авторами.

Также в модель были добавлены слои отсева. Слой отсева добавляется, чтобы избежать переобучения, которое является феноменом, когда модель машинного обучения работает лучше на обучающих данных по сравнению с тестовыми данными [8].

Для увеличения надежности в модель был добавлен плотный слой в конце. Число нейронов в плотном слое установлено равным 1, так как на выходе предсказывается 1 значение.

Для обучения используется среднеквадратичная ошибка в качестве функции потерь, а для уменьшения потерь или оптимизации алгоритма используется оптимизатор adam [9]. Среднеквадратичная ошибка измеряет среднее квадратов ошибок, т. е. среднеквадратич-

ную разность между расчетными и фактическими значениями.

Adam – это алгоритм оптимизации, который можно использовать вместо классической процедуры стохастического градиентного спуска для итеративного обновления весов сети на основе обучающих данных. Скорость обучения поддерживается для каждого веса сети (параметра) и адаптируется отдельно по мере развития обучения [9, 10].

Обучение проходило 50 эпох и заняло примерно 30 минут на персональной ЭВМ.

Структура обучающей выборки (рис. 3, выделено красным) и фактических данных, прогноз которых необходимо получить (рис. 3, выделено синим), представлены на рис. 3.

Time	P1	W1	P2	W2	Oil
0	8	153	10	172	92
60	9	184	10	180	104
120	8	169	8	176	93
180	8	149	10	149	89
240	9	162	8	164	127
300	9	161	9	151	105
360	8	161	9	176	124
420	10	187	8	155	128
480	9	183	10	186	82
540	9	168	9	159	116
600	9	172	10	185	117
660	8	174	8	164	83
720	9	158	9	159	111
780	9	179	8	165	128
840	10	151	9	152	128
900	8	167	10	152	116
960	10	176	9	166	117
1020	8	165	10	151	117
1080	10	175	10	153	121
1140	9	155	8	186	123
1200	9	160	8	157	86
1260	8	152	8	150	130
1320	10	150	9	171	83
1380	8	164	8	185	118
0	10	174	8	180	86
60	10	184	8	154	127
120	8	178	9	183	98
180	8	148	8	156	107

Рис. 3. Структура данных

Примечание: составлено авторами.

Аналогично реализации ИНС с моделью LSTM была реализована ИНС с моделью GRU для сравнительного анализа. Результаты

обучения и работы реализованных РНС представлены на рис. 4, 5.

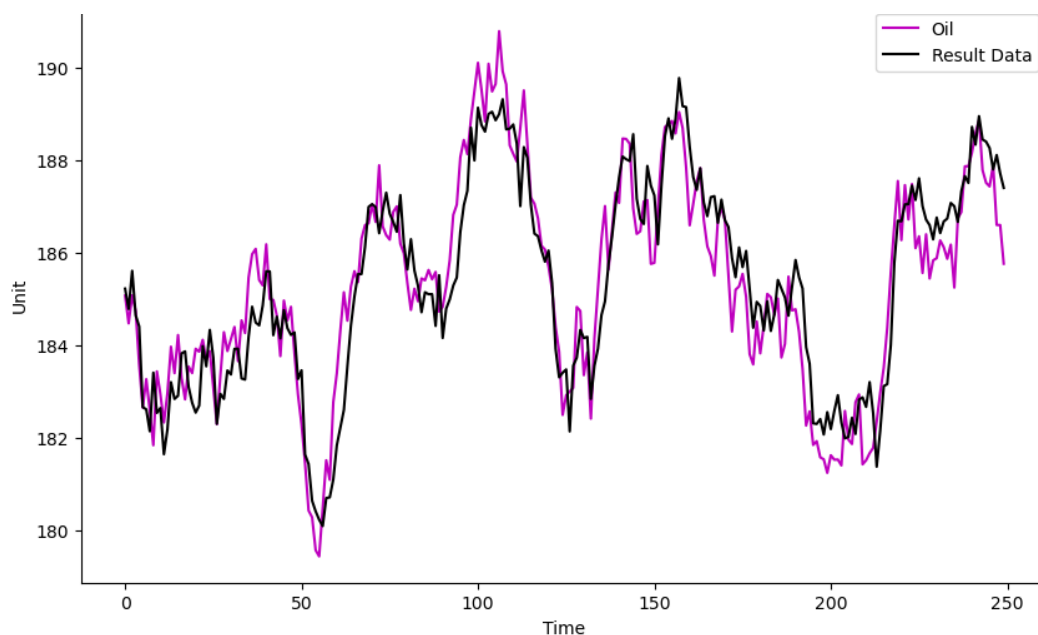


Рис. 4. Единственный слой LSTM, 500 нейронов

Примечание: составлено авторами.

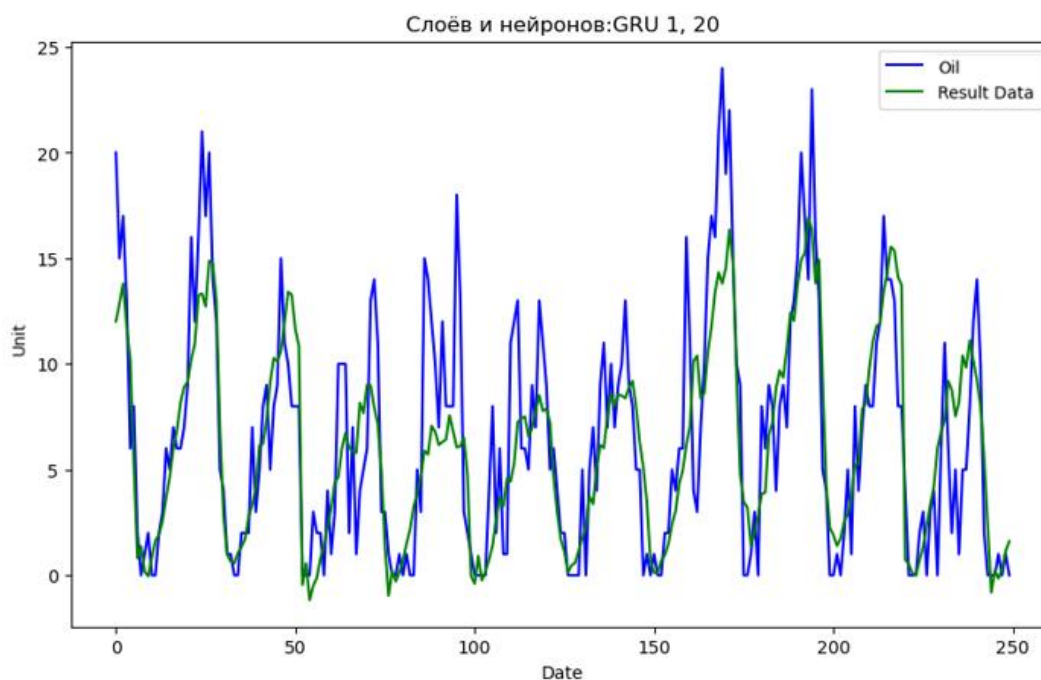


Рис. 5. Единственный слой GRU, 20 нейронов

Примечание: составлено авторами.

По вышеприведенным графикам видно, что в обоих случаях сохраняется тенденция при прогнозировании, но в случае GRU прогнозируемые данные сильнее отличаются

от фактических, чем у модели LSTM. Данные результаты подтверждают эффективность метода и модели LSTM в решении поставленной задачи.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

В начальной стадии проведения исследования было создано 12 нейронных сетей с различными параметрами: количество слоев долгосрочной памяти варьировалось от 1 до 4, а размер выходного пространства – от 20 до 50.

Анализируя графики, можно сделать вывод, что повышение количества слоев долгосрочной памяти приводит к снижению точности предсказаний. В данном контексте для решения поставленной задачи достаточно использовать лишь один слой. Одновременно с этим увеличение размерности выходного пространства приводит к повышению точности прогнозирования.

В ходе второго этапа экспериментов было проведено обучение нейронных сетей с последовательным увеличением размерности выходного пространства. Наблюдаемое повышение точности на графиках подтверждает значимость выбранного параметра. Однако для обеспечения чистоты эксперимента была

проведена удачная попытка привести обучающую модель к снижению точности вследствие избыточного количества нейронов.

По итогам анализа данных были выбраны эффективные параметры для количества слоев долгосрочной памяти и количества нейронов в этих слоях: 1 слой, содержащий 300 нейронов.

Качество моделей LSTM и GRU одинаковые, но GRU проигрывает по времени обучения, и LSTM имеет явно определенную ячейку памяти, которая позволяет сохранять информацию на длительные промежутки времени. Это особенно полезно для моделирования зависимостей в долгосрочных временных рядах, где важно учитывать более удаленные зависимости. Поэтому в дальнейшем проводилось исследование зависимости количества эпох на точность прогноза у модели LSTM.

График предсказанных значений и метрики выбранной модели представлены на рис. 6.

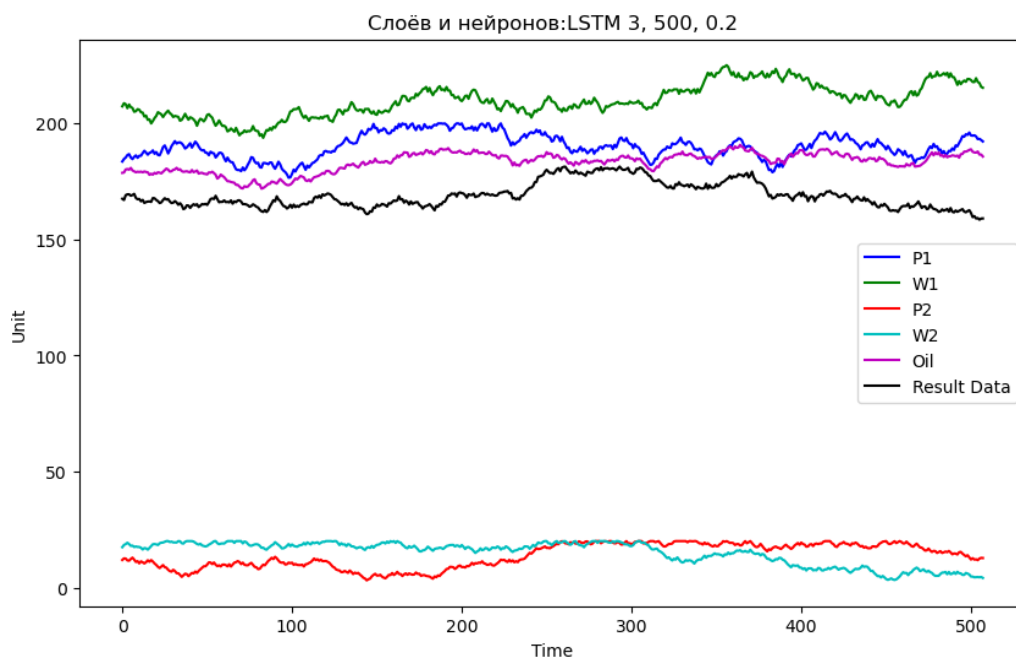


Рис. 6. Фрагмент автоматического отчета исследования

Примечание: составлено авторами.

Тест на стационарность:

Т-статистика = -8,423,

Р-значение = 0,000.

Критические значения:

1 %: -3,4319957562514243 – данные стационарны с вероятностью 99 %;

5 %: -2,862267205086051 – данные стационарны с вероятностью 95 %;

10 %: -2,5671570992175385 – данные стационарны с вероятностью 90 %;

MAD: 1,9625;

MSE: 6,4877;

Стандартная ошибка: 2,5471.

В целях обеспечения надлежащего контроля над экспериментом, а также в связи с нестационарным характером данных было проведено повторное исследование с исполь-

зованием различных параметров ИНС. Характеристики различных моделей, а также их результаты в терминах точности приведены в табл. 2.

Таблица 2

Параметры и результаты точности ИНС

№	Слоев	Нейронов	Отсев	Эпох	MAD	MSE	Стд. ошибка
1	1	100	0,2	100	1,187	2,4807	1,575
2	1	300	0,2	100	1,5975	3,5804	1,8922
3	1	500	0,2	100	1,268	2,2494	1,4998
4	3	500	0,2	100	14,928	285,837	16,906
5	1	500	0,1	100	1,545	3,4076	1,846
6	1	500	0,5	100	1,0529	1,7216	1,3121
7	1	500	0,8	100	1,5423	3,7508	1,9367
8	1	500	0,2	200	0,8363	1,1116	1,0543
9	1	500	0,2	300	1,1261	2,3053	1,5183
10	1	1 000	0,5	300	1,5606	3,5932	1,8956
11	1	500	0,5	200	0,8759	1,2283	1,1083
12	1	1 000	0,2	300	0,8394	1,2134	1,1016

Примечание: составлено авторами.

Результаты исследования на промышленных данных позволили сделать несколько важных выводов. Установлено, что увеличение количества нейронов или эпох обучения приводит к повышению точности прогнозирования. Однако при дальнейшем увеличении этих параметров возникает проблема переобучения, что приводит к ухудшению обобщающей способности модели.

Чтобы решить эту проблему, был изучен механизм вероятности отсева нейронов, который позволяет случайным образом исключать определенные нейроны во время обучения. Однако при увеличении вероятности отсева нейронов также возникает проблема недообучения, когда модель теряет способность извлекать важные признаки из данных.

В результате экспериментов были определены несколько наиболее подходящих параметров для обучения модели. В частности,

использование 500 нейронов и проведение 200 эпох обучения позволило достичь высокой точности прогнозирования. Кроме того, установка вероятности отсева нейронов на уровне 20 % также дала хорошие результаты.

Таким образом, были получены эффективные значения параметров, которые могут использоваться для эффективного обучения модели на основе промышленных данных. Эти результаты могут иметь важное практическое значение для прогнозирования и оптимизации процессов в промышленной сфере.

Графические представления результатов прогнозирования худшей модели приведены на рис. 7, 8. Анализ рис. 6 позволяет наглядно оценить отклонение прогнозируемых значений от фактических данных. Лучшие результаты представлены на графиках рис. 9, 10.

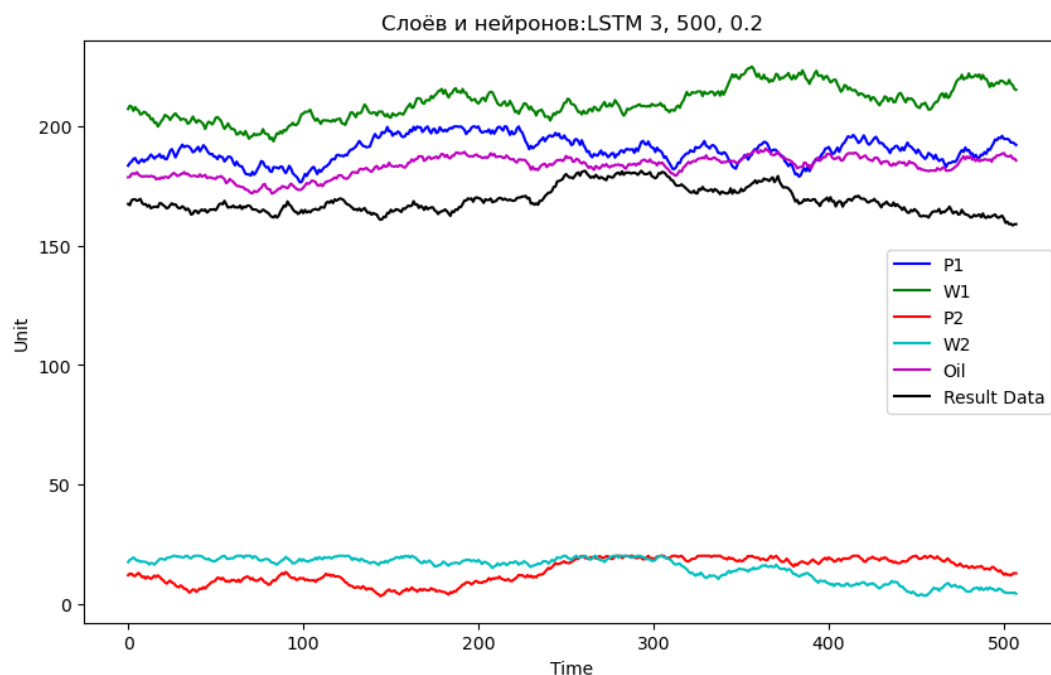


Рис. 7. График тестовых данных и результат прогноза
Примечание: составлено авторами.

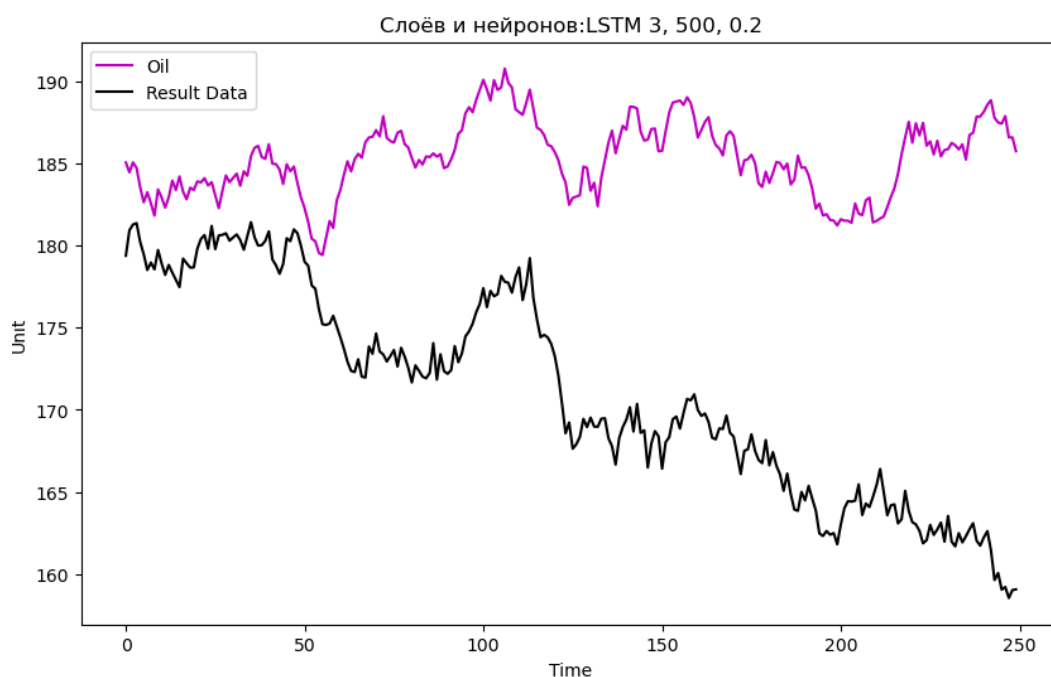


Рис. 8. График тестовых данных и результат прогноза. Худшая модель
Примечание: составлено авторами.

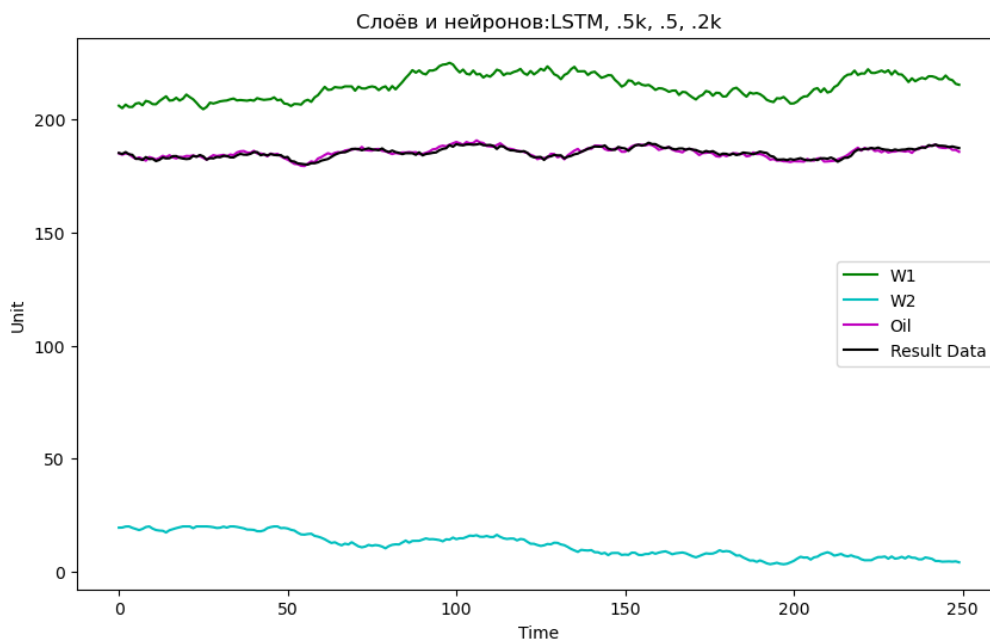


Рис. 9. График результата прогноза. Все параметры. Лучший результат
Примечание: составлено авторами.

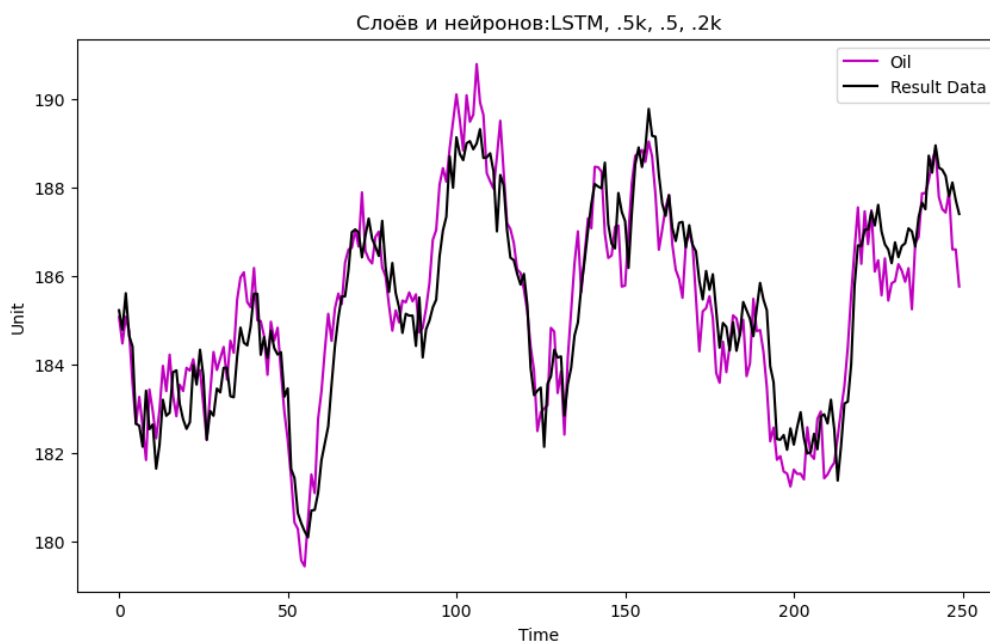


Рис. 10. График результата прогноза. Лучший результат
Примечание: составлено авторами.

ЗАКЛЮЧЕНИЕ

В работе представлены компьютерные эксперименты для определения эффективных параметров обучения нейронной сети. Было выявлено, что использование большего количества слоев приводит к уменьшению точности прогнозирования. Однако увеличение количества нейронов в сети позволяет повысить точность прогноза. При сравнении моделей

LSTM и GRU, модель GRU оказалась незначительно точнее, но, учитывая более длительное время обучения, модель LSTM – предпочтительнее.

Также было проведено исследование зависимости точности модели от количества эпох обучения. Результаты показали, что с увеличением количества эпох точность прогноза повышается.

В целом, данное исследование подтвердило эффективность применения рекуррентных нейронных сетей, особенно модели LSTM для прогнозирования временных рядов. Полученные результаты могут быть полезны для при-

ятия решений и оптимизации процессов в области добычи нефти и других сферах, где требуется точный прогноз временных данных.

Список источников

1. Куленцан А. Л., Марчук Н. А. Анализ добычи нефти из пластов в Российской Федерации // Российский химический журнал. 2022. Т. 66, № 1. С. 71–75. DOI 10.6060/rcj.2022661.10.
2. Мурашов Б. А., Тепляков Н. Ф. Расчет и оптимизация объектов подготовки и перекачки нефти при оценке проектов и реинжиниринге // ПРОНЕФТЬ. Профессионально о нефти. 2018. № 4. С. 71–74.
3. Шевчук Т. Н., Кашников О. Ю., Мезенцева М. А. и др. Прогноз показателей добычи из пластов баженовской свиты на основе статистических зависимостей и методов машинного обучения // ПРОНЕФТЬ. Профессионально о нефти. 2020. № 4. С. 63–68.
4. Тутыгин В. Искусственный интеллект в нефтегазовой индустрии как фактор развития производственной системы // Science and Innovation. 2023. № 3. С. 168–171.
5. Хасанов И. И., Хасанова З. Р., Шакиров Р. А. и др. Обзор применения нейросетей в области добычи и транспорта нефти и газа // Транспорт и хранение нефтепродуктов и углеводородного сырья. 2022. № 3–4. С. 11–15.
6. Подольский А. К. Применение методов искусственного интеллекта в нефтегазовой промышленности // Современная наука. 2016. № 3. С. 33–36.
7. Семенов Е. Д. Адаптивное прогнозирование входных параметров при добыче нефти с использованием рекуррентной нейронной сети // Проблемы и решения автоматизации XXI века : сб. материалов V Национал. науч.-практич. студ. конф., 25–26 марта 2022 г., г. Сургут. Сургут : СурГУ, 2022. С. 78–83.
8. Будыльский Д. В. GRU и LSTM: современные рекуррентные нейронные сети // Молодой ученый. 2015. № 15. С. 51–54.
9. Kingma D. P., Ba J. L. Adam: A method for stochastic optimization. In: *Proceedings of the 3rd International Conference for Learning Representations*, May 7–9, 2015, San Diego. p. 1–15. DOI 10.48550/arXiv.1412.6980.
10. Melinte D. O., Vladareanu L. Facial expressions recognition for human-robot interaction using deep convolutional neural networks with rectified Adam optimizer. *Sensors*. 2020;20(8):2393.

Информация об авторах

Е. Д. Семёнов – инженер-программист II категории.
М. Я. Брагинский – кандидат технических наук, доцент.
Д. В. Тараканов – кандидат технических наук, доцент.
И. Л. Назарова – аспирант.

References

1. Kulentsan A. L., Marchuk N. A. Analysis of oil production from reservoirs in the Russian Federation. *Rossiiskii khimicheskii zhurnal*. 2022;66(1):71–75. DOI 10.6060/rcj.2022661.10. (In Russian).
2. Murashov B. A., Teplyakov N. F. Calculation and optimization of oil treatment and pumping facilities during project evaluation and reengineering. *PROneft. Professionally about Oil*. 2018;(4):71–74. (In Russian).
3. Shevchuk T. N., Kashnikov O. Yu., Mezentseva M. A. et al. Production forecast for Bazhen formation reservoirs on the basis of statistical analyses and machine learning techniques. *PROneft. Professionally about Oil*. 2020;(4):63–68. (In Russian).
4. Tutugin V. Iskusstvennyi intellekt v neftegazovoi industrii kak faktor razvitiia proizvodstvennoi sistemy. *Science and Innovation*. 2023;(3):168–171. (In Russian).
5. Khasanov I. I., Khasanova Z. R., Shakirov R. A. et al. Overview of the use of neural networks in the field of oil and gas production and transportation. *Transport and Storage of Oil Products and Hydrocarbons*. 2022;(3–4):11–15. (In Russian).
6. Podolsky A. K. The application of artificial intelligence methods in the oil and gas industry. *Sovremennaya nauka*. 2016;(3):33–36. (In Russian).
7. Semenov E. D. Adaptivnoe prognozirovaniye vkhodnykh parametrov pri dobyche nefti s ispolzovaniem rekurrentnoi neuronnoi seti. In: *Proceedings of the V National Research-to-Practice Conference “Problemy i resheniya avtomatizatsii XXI veka”*, March 25–26, 2022, Surgut. Surgut: SurSU; 2022. p. 78–83. (In Russian).
8. Budylsky D. V. GRU i LSTM: sovremennyye rekurrentnye neuronnye seti. *Molodoi uchenyi*. 2015;(15):51–54. (In Russian).
9. Kingma D. P., Ba J. L. Adam: A method for stochastic optimization. In: *Proceedings of the 3rd International Conference for Learning Representations*, May 7–9, 2015, San Diego. p. 1–15. DOI 10.48550/arXiv.1412.6980.
10. Melinte D. O., Vladareanu L. Facial expressions recognition for human-robot interaction using deep convolutional neural networks with rectified Adam optimizer. *Sensors*. 2020;20(8):2393.

Information about the authors

E. D. Semyonov – Second Category Software Engineer.
M. Ya. Braginsky – Candidate of Sciences (Engineering), Docent.
D. V. Tarakanov – Candidate of Sciences (Engineering), Docent.
I. L. Nazarova – Postgraduate.