

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

МОСКОВСКИЙ ИНСТИТУТ ЭЛЕКТРОНИКИ И МАТЕМАТИКИ
им. А.Н. ТИХОНОВА

Абдурасулзода Далер

**НЕЙРОСЕТЕВОЙ ВЕРОЯТНОСТНЫЙ КАЛЬКУЛЯТОР ОТКАЗОВ
УСТАНОВОК ЭЛЕКТРОЦЕНТРОБЕЖНЫХ НАСОСОВ**

Выпускная квалификационная работа по направлению подготовки
09.03.01. Информатика и вычислительная техника
студента образовательной программы
«Информатика и вычислительная техника»

Студент

Абдурасулзода Д.

подпись

И.О. Фамилия

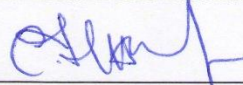
Руководитель

Профессор, д. т. н. Стукач О.В.

должность, звание, И.О. Фамилия

«УТВЕРЖДАЮ»

Академический руководитель образовательной
программы
«Информатика и вычислительная техника»



С. Н. Полесский

« 14 » декабря 2020 г.

**ЗАДАНИЕ
на выпускную квалификационную работу бакалавра**

студенту группы БИВ174 Абдурасулзода Далеру

1.1 Тема работы

Тема (рус.):

Нейросетевой вероятностный калькулятор отказов установок
электроцентробежных насосов.

Тема (англ.):

The Neural Probability Calculator of the Electric Submersible Pump Installation
Fails.

2. Требования к работе

2.1. Общие требования к объекту разработки

Необходимо разработать программу на языке высокого уровня, которая рассчитывает вероятность отказа установок электроцентробежных насосов, используя метод нейронных сетей. Программа должна позволять вводить входные значения параметров, обеспечивать корректную обработку исключительных ситуаций, которые вызваны неверным форматом или недопустимыми значениями входных данных. При вводе параметров любого из отказов обучающей выборки вероятность отказа должна получиться приближенной к единице. Программа предоставляет возможность чтения данных по отказам, предоставленных пользователем, из файла формата CSV и записи полученных результатов в отдельный файл. Также должна быть предусмотрена возможность изменения обучающей выборки пользователем.

2.2. Требования к документации

В отчёте о выполнении ВКР должны быть отражены исследовательская и практическая части работы. Текст ВКР должен быть оформлен в соответствии с ГОСТ.

Техническая документация к разработанному программному обеспечению должна быть оформлена в соответствии с ГОСТ. Практическая часть работы

должна быть воспроизводима в соответствии с документацией. Разработанные программные коды и прочие материалы должны быть размещены в репозитории.

2.3. Требования к программному обеспечению и техническим средствам
Персональный компьютер с ОС Windows версии не ниже 7.

3. Содержание работы

3.1. Изучение различных моделей нейронных сетей.

3.2. Проектирование нейронной сети, получение вероятностных оценок отказов.

3.3. Разработка программного интерфейса, позволяющего использовать возможности калькулятора вероятности отказов.

3.4. Тестирование разработанной программы.

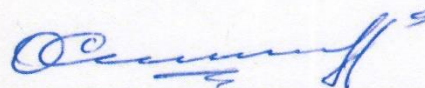
3.5. Составление технической документации.

3.6. Оформление ВКР.

4. Сроки выполнения этапов работы

Проект ВКР представляется студентом в срок до:	«10» февраля 2021 г.
Первый вариант ВКР представляется студентом в срок до:	«26» апреля 2021 г.
Итоговый вариант ВКР представляется студентом руководителю до загрузки работы в систему «Антиплагиат» в срок до:	«10» мая 2021 г.

Задание выдано «9» декабря 2020 г.



О. В. Стукач

Задание принято к исполнению «9» декабря 2020 г.



Д. Абдурасулзода

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»




МОСКОВСКИЙ ИНСТИТУТ ЭЛЕКТРОНИКИ И МАТЕМАТИКИ
им. А.Н. ТИХОНОВА

**График сдачи этапов
выпускной квалификационной работы бакалавра**

студента группы БИВ 174 Абдурасулзода Далера

Тема работы

Нейросетевой вероятностный калькулятор отказов установок
электроцентробежных насосов

Дата представления проекта ВКР	«10» февраля 2021 г.		О.В.Стукач
Дата представления первого варианта ВКР	«26» апреля 2021 г.		О.В.Стукач
Дата представления итогового варианта ВКР	«13» мая 2021 г.		О.В.Стукач

АННОТАЦИЯ

Целью работы является разработка вероятностного калькулятора отказов установок электроцентробежных насосов. В ходе выполнения работы был разработан алгоритм получения вероятностного значения отказа, включающий в себя модель нейронной сети. Реализация функций добавления, считывания и хранения данных позволяет использовать нейронные сети для получения вероятности отказа. Для оценки точности алгоритма были реализованы различные по структуре и способам обучения нейросети, и результаты работы этих моделей были сопоставлены между собой. При внесении параметров насоса в текущее время программа выдает значение вероятности отказа. Разработанная программа может быть использована сотрудниками нефтедобывающей компании.

ABSTRACT

The aim of the work is to develop a probabilistic calculator of failures of electric submersible pump installations. In the course of the work, an algorithm for obtaining a probable failure was developed, which includes a neural network model. The implementation of adding, read and storing data allows using neural networks to get the probability of failure. For the efficiency of the algorithm, neural networks of various structures and methods of training were implemented, and the results of the operation of these models were compared with each other. When entering the pump parameters at the current time, the program gives the value of the probability of failure. The developed program can be used by employees of an oil company.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	8
1 Обзор литературы.....	10
1.1 Методы машинного обучения.....	10
1.2 RCM-методы	12
1.3 Анализ существующих решений	13
1.4 Выводы	15
2 Выбор оптимальной топологии нейросети.....	16
2.1 Обработка исходных данных	16
2.2 Проектирование однослойной нейронной сети	18
2.3 Проектирование нейронных сетей со скрытыми слоями.....	20
3 Динамика вероятностей отказа	23
3.1 Реализация функций для подсчета вероятностей	23
3.2 Получение динамики вероятностей отказа.....	24
4 Оптимизация модели нейросети.....	27
4.1 Синтез данных для модели.....	27
4.2 Оптимизация параметров модели нейронной сети.....	27
4.3 Результаты оптимизации модели.....	28
5 Нейросеть со значениями ежедневных показателей на входе.....	30
5.1 Описание принципа модели	30
5.2 Результаты работы модели.....	30
6 Графический интерфейс	32
ЗАКЛЮЧЕНИЕ	35
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	37

ВВЕДЕНИЕ

Повышение эффективности производства при использовании установок электроцентробежных насосов (далее - УЭЦН) является одной из важнейших задач для нефтяной промышленности. Иногда использование таких установок может привести к различным сбоям и отказам. Поэтому проблема предотвращения таких отказов всегда была и остается актуальной. Для ее решения необходимо наличие алгоритмов, позволяющих своевременно рассчитывать вероятность любого сбоя. Заблаговременное отключение насоса для планового технического обслуживания скважины до наступления отказа может предотвратить критическую ситуацию. Поэтому операторам производственного контроля важно иметь информацию о вероятности отказа при определенных параметрах. В рамках выпускной квалификационной работы предлагается реализовать алгоритм для получения вероятности отказа УЭЦН. Неисправности установок могут быть вызваны многими факторами. Однако в работе рассматривается только один из них – отказ, причиной которого являются солевые отложения. Эти соли возникают, когда насос неизбежно нагревается и выпадает в виде накипи. Данное явление приводит к уменьшению диаметра перекачивающего патрубка, уменьшению объемов производства, отклонению режима работы насоса от нормативного. В результате это становится причиной поломки насоса.

В настоящее время для мониторинга состояния скважин используются специальные программные средства цифровой платформы поддержки добычи, которые протоколируют данные по техническим режимам работы оборудования. Недостатком данного подхода является тот факт, что выводы по результатам мониторинга практически невозможно формализовать, никаких расчетов вероятностных оценок не происходит. Решение о прекращении использования установки принимается оператором на основе интуиции и опыта, что может привести к неблагоприятным последствиям.

В данной работе предлагается использование нейронных сетей для прогнозирования наступления отказа и получения его вероятностного значения.

Управление, прогнозирование и классификация являются одними из основных задач, которые можно решить с помощью нейронной сети. Основным преимуществом нейросетевых моделей является возможность воспроизводить сложные нелинейные зависимости с большим количеством переменных. Ключевой особенностью нейросети является способность к обучению. В процессе обучения определяются связи и зависимости между исходными параметрами. К главным преимуществам нейронных сетей можно отнести скорость обработки данных и качество модели.

1 Обзор литературы

Для создания алгоритма расчета вероятности выхода из строя насосов УЭЦН ранее были проанализированы научные статьи в этой области. Существуют различные подходы для решения поставленной задачи. К ним можно отнести: методы машинного обучения, статистического анализа, предиктивной аналитики, вероятностные методы, RCM-методы. Рассмотрим подробнее преимущества и недостатки каждого из методов.

1.1 Методы машинного обучения

Существует множество разнообразных моделей машинного обучения, которые можно использовать для решения большого числа задач. Данные модели могут отличаться способом оперирования различными типами данных и определением неочевидных закономерностей.

Выделим основные методы машинного обучения:

- графические и генеративные модели;
- бустинг;
- Байесовское обучение;
- обучение на базе ядер;
- дерево решений;
- нейронные сети.

В графических [1] и генеративных [2] методах архитектура модели задается со свободными параметрами. Для настройки этих параметров применяется градиентный спуск. Данные методы требуют небольших вычислительных мощностей, однако достаточно быстро переобучаются.

Алгоритмы бустинга [3] могут обеспечить необходимый уровень точности прогнозирования за счет многократного использования с применением различных способов, однако данные алгоритмы могут быть несовершенными. Высокую скорость работы и хорошую автоматизацию можно считать преимуществами данного метода.

Байесовское обучение [4] хорошо применимо для модели с небольшим количеством входных данных. Данный метод требует участия человека для задания

априорного значения обучения. Также Байесовские оценки не учитывают старение скважинного оборудования, что можно отнести к недостатку этого метода.

Принцип обучения на базе ядер [5] заключается в том, что используется ядро, удовлетворяющее необходимым условиям и служащее эталоном при обучении. Архитектурная сложность требует высокой производительности, поэтому прогнозирование при большом объеме данных зачастую не реализуется.

Дерево отказов является одним из самых распространенных методов, которые применяются для прогнозирования различных сбоев. Дерево отказов представляет собой многоуровневую графологическую структуру причинно-следственных взаимосвязей. Данные взаимосвязи могут быть получены в ходе отслеживания аварийных ситуаций в обратном порядке для поиска всех возможных причин их возникновения [6].

Преимущества метода:

- возможность проведения количественного и качественного анализа надежности системы;
- поиск ненадежных мест в явном виде;
- иллюстрация поведения объекта и процесса его работы.

Недостатки метода:

- реализация метода потребляет много ресурсов;
- отсутствие возможности учитывания частичного отказа элементов оборудования;
- метод предполагает рассмотрение системы лишь в определенный момент времени, динамика событий отражается не столь успешно.

С помощью применения метода дерева отказов можно получить подробный анализ причин отказов оборудования системы. Данный анализ проводится для каждого периода функционирования как системы в целом, так и отдельных ее частей [6]. Главным преимуществом прогнозной аналитики является возможность предсказания абсолютно разных явлений. К недостаткам можно отнести то, что для реализации данного метода требуется много ресурсов. По этой причине применение

данного метода является достаточно трудной задачей, однако он применим практически везде.

Нейронные сети являются воплощением математической модели. К основным преимуществам можно отнести возможность работы с пропусками в исходных данных, возможность проектирования нетривиальных связей и легкую адаптацию к меняющимся условиям. К недостаткам данного метода можно считать тот факт, что веса, поступающие на вход нейронов, практически не интерпретируемы. Из-за этого сложно объяснить полученное предсказание. Также к минусам применения нейронных сетей можно отнести то, что модель сильно зависит от исходной выборки. От нее зависит точность модели и время обучения.

Если рассматривать методы машинного обучения обобщенно, то можно выделить общие преимущества и недостатки.

Преимущества:

- способность работать, если исходные данные неполные;
- установление нетривиальных закономерностей между входными и выходными данными;
- адаптация к новым условиям и входным данным;
- простота и контролируемость процесса обучения;
- способность работать с большим объемом данных.

Недостатки:

- оперирования с весовыми значениями, которые являются абстрактными и не интерпретируемыми;
- многие алгоритмы могут переобучаться;
- большинство методов машинного обучения требует больших вычислительных ресурсов.

1.2 RCM-методы

Для решения задач прогнозирования также может быть применен метод RCM [7]. Согласно этой технологии, предлагается проведение технического обслуживания, ориентированного на надежность. Данная методология основывается на гипотезе о том, что поддержание оборудования в идеальном исправном состоянии не есть сама

цель. Согласно этому принципу, важнее обеспечить надежность протекания критических производственных и технологических процессов. То есть ремонту подлежит только то оборудование, которое действительно в этом нуждается [8]. Определим преимущества и недостатки данного подхода.

Преимущества:

- анализ по методу RCM способен предсказать отказ, исходя из совокупности параметров, и позволяет значительно сократить плановые работы;
- использование данного метода позволяет разработать план мероприятий по обслуживанию установки.

Недостатки:

- для достижения результата требуется достаточно много времени;
- реализация методов достаточно затратна.

1.3 Анализ существующих решений

Рассмотрим, как ранее описанные методы использовались в научных исследованиях в области предсказания вероятности отказов УЭЦН.

Например, в [9] и [10] использовалось дерево решений. Это один из распространенных инструментов интеллектуального анализа данных и прогнозной аналитики. Такие методы помогают в решении задач классификации и регрессии. В этом исследовании не только предсказано наличие неисправности, которая может привести к вероятному отказу насоса, но также определен тип этой неисправности. Алгоритмы, использованные в данной работе, позволили получить точность результатов, близкую к 90%. Это говорит о том, что эти алгоритмы подходят для практического применения при диагностике неисправностей. Однако следует отметить, что в данной работе основной задачей была именно классификация насосов на исправные и неисправные, а не прогнозирование вероятности их выхода из строя.

Подобные методы машинного обучения использовались в работе [11]. Однако особенность этого исследования заключалась в том, что оно включало также практический этап, в котором проводились испытания на экспериментальном стенде. Изменения условий, приводящие к отказу насоса, моделировались физически. На следующем этапе полученные данные обрабатывались и пропускались через

алгоритм, определяющий вероятностные оценки. Точность обнаружения неисправностей составляет от 76% до 82%. В целом полученные аналитически результаты согласуются с поведением, наблюдаемым в эксперименте.

Совершенно иной подход был использован в исследованиях [12] и [13]. В этих работах для решения проблемы диагностики неисправностей во время работы УЭЦН был использован метод опорных векторов. Этот метод широко используется в задачах классификации. Его суть заключается в том, что каждый объект считается вектором в p -мерном пространстве. Задачей является разделение объектов двух классов гиперплоскостью размерности $(p-1)$. Результаты исследования показали, что применение этого метода достаточно эффективно для поиска неисправностей.

Очень похожая проблема исследовалась в работе [14]. Целью данной работы было получение вероятностных оценок «полета» погружного насоса. Авария типа «полет» — это отказ, при котором происходит самопроизвольное распадение скважинного оборудования. Данный тип сбоя является наиболее опасным. Для решения этой проблемы автор применил байесовский метод, используемый в теории надежности. Этот метод позволяет определить вероятность возникновения события с различными параметрами, которые присутствуют, когда это событие происходит. Авторы использовали реальные данные о таких отказах, доля которых по отношению ко всем типам отказов крайне мала. Используя этот метод, можно было получить вероятностное значение такого отказа для каждого случая. По мнению авторов данной работы, использование байесовского метода позволяет с достаточной точностью оценить состояние оборудования.

Большинство из вышеперечисленных методов использовались вместе в исследовании [15]. В данной работе были рассмотрены три типа неисправностей: перекося вала, дисбаланс лопастей насоса и механическое трение. Основное внимание в этом исследовании уделялось получению вероятностных оценок неисправностей на основе анализа характера вибрации. Основная идея работы заключалась в сравнении методов классификации между собой и выборе лучшего классификатора. По результатам можно заметить, что классификатор «случайного леса» с использованием шаблонов решений достиг лучшей точности по сравнению со всеми

классификаторами. Это объясняется тем, что у него самое высокое медианное значение и высокие показатели других статистических показателей.

1.4 Выводы

Следует отметить, что для прогнозирования вероятности отказа нефтедобывающего оборудования были описаны различные подходы: методы машинного обучения, статистического анализа, предиктивной аналитики, вероятностные методы, RCM-методы. Высокая эффективность при применении методов статистического анализа и вероятностных методов может быть достигнута только при наличии модели формирования отказов, описывающей процессы отклонения от нормального. Разработка данных моделей для установок электроцентробежных насосов является крайне сложной задачей, так как необходимо большое количество исходных данных по реальным отказам. Для выполнения данной работы имеются данные по отказам лишь 35 установок, поэтому можно считать, что вероятностные методы и методы статистического анализа в данной работе сложно применимы. Методы предиктивной аналитики, как правило, основываются на статистических методах, которые требуют глубокого понимания каждого отказа. По этой причине можно считать, что данные методы также непригодны для решения текущей задачи.

В данной работе предлагается использовать нейронные сети, так как в данной узкой предметной области этот метод еще не применялся. Данный подход позволит получить неочевидные закономерности влияния параметров УЭЦН на вероятность отказа. Предполагается спроектировать различные по структуре и способам обучения модели нейронных сетей и выбрать более подходящую для решения конкретной задачи. Работоспособность модели будет проверена на архивных данных, собранных в одной из нефтедобывающих компаний. Недостатком использования нейронных сетей является тот факт, что подобные отказы случались не так много. Это значит, что объем исходной выборки будет невелик и достаточно сложно будет добиться высокой точности модели. Однако данный недостаток можно компенсировать, если грамотно синтезировать входные данные на основе исходной выборки.

2 Выбор оптимальной топологии нейросети

Для определения достоверности полученных результатов необходимо иметь достаточно большую статистику по эксплуатации УЭЦН. Однако объем исходной выборки не позволяет это сделать. По этой причине в данной работе предлагается реализовать различные по топологии и способам обучения нейронные сети и сравнивать результаты их работы между собой. Это позволит выбрать модель нейросети с оптимальной для решения поставленной задачи архитектурой.

2.1 Обработка исходных данных

Для исследования предлагается использовать данные по скважинам, собранные ежедневно сотрудниками нефтедобывающей компании. Эти данные представляют собой значения физических параметров какого-либо компонента УЭЦН. Также имеется таблица мер центральной тенденции этих параметров, посчитанных для каждой скважины. Чтобы алгоритм, подсчитывающий вероятность отказа, учитывал динамику изменения параметров было принято решение использовать в качестве входного сигнала нейронной сети данные из таблицы мер центральной тенденции каждого параметра. Всего имеется 35 установок, в 28 из которых случился отказ.

Исходные данные для проектирования нейронной сети хранятся в файле формата .xls. Для чтения данных использовалась библиотека Pandas на высокоуровневом языке программирования Python. Данная библиотека хорошо подходит для первичной обработки и анализа данных. На рисунке 1 представлен код для чтения исходных данных из файла.


```
[ ] from tensorflow import keras
import pandas as pd
import numpy as np
pd.set_option('display.max_columns', None)

[ ] cols = ['id', 'date_start', 'date_end', 'day_diff', 'gaz_amount', 'gaz_mean',
major_cols = ['liquid_tm', 'liquid', 'energy', 'water', 'nd', 'pressure_zad',
minor_cols = ['_amount', '_mean', '_median', '_min', '_max', '_var', '_std']
add_cols = []
for major_col in major_cols:
    for minor_col in minor_cols:
        add_cols.append(major_col+minor_col)
cols += add_cols
cols += ['prim', 'target', 'ungerm', 'comment']

[ ] data = pd.read_excel('data_otkaz-.xls', skiprows=1, names=cols)
```

Рис. 1. Код для чтения данных из файла.

На рисунке 2 можно наблюдать фрагмент исходных данных из файла формата .xls, а рисунок 3 представляет собой фрагмент считанных данных.

	A	B	C	D	E	F	G	H	I	J	K
1					Газ ТМ						
2	Скважина	Дата пуска	Дата отключения	МРП	Кол-во значений	Среднее арифметическое	Медиана	Min	Max	VAR	СКО
3	108	24.03.2016	23.05.2016	60	61	0	0	0	0	-	0
4	110	27.04.2015	02.06.2017	767	768	2812,88	1004	0	45582	253%	7115,4
5	239	18.02.2016	21.03.2018	32	33	2385,15	3495	0	5065	89%	2134,4
6	315	12.08.2015	12.03.2016	213	214	-	-	0	0	-	-
7	487	01.01.2015	02.04.2015	91	92	781,32	775	400	1068	9%	73,17

Рис. 2. Фрагмент исходных данных.

Сводная таблица, содержащая меры центральной тенденции включает для каждого физического параметра следующие значения:

- срок эксплуатации;
- среднее арифметическое;
- медианное значение;
- минимальное значение;
- максимальное значение;
- дисперсию и среднеквадратичное отклонение.

```
[ ] filtered_data.head()
```

	day_diff	gaz_mean	gaz_median	gaz_min	gaz_max	gaz_var	gaz_std
0	60	0	0	0.0	0	-	0
1	767	2812.88	1004	0.0	45582	2.53	7115.4
2	32	2385.15	3495	0.0	5065	0.89	2134.4
3	213	-	-	0.0	0	-	-
4	91	781.32	775	400.0	1068	0.09	73.17

Рис. 3. Фрагмент считанных данных.

После считывания данные были обработаны и нормализованы, что позволило начать проектирование нейронных сетей на готовых данных. Как видно из рисунка 3, некоторые данные в таблице пропущены, поэтому обработка данных включала в себя также заполнение этих пропусков.

2.2 Проектирование однослойной нейронной сети

Для проектирования нейронных сетей использовалась библиотека Keras, написанная на языке Python. Для начала была спроектирована однослойная нейронная сеть без скрытых слоев. Обучение модели производилось методом кросс-валидации. Особенностью данного метода является то, что исходная выборка разбивается на несколько групп, одна из которых остается для тестирования модели, а другие используются для ее обучения. Стоит отметить, что несмотря на то, что в нейросеть подаются 84 входных сигнала, что говорит о масштабе ее структуры, исходных данных было достаточно мало. Это данные о 35 установках, из которых в 28 случился отказ. По этой причине разбиение исходных данных методом кросс-валидации можно считать оправданным.

Точность данной модели составила 0.57, что является достаточно низким значением. Данный факт вызвал интерес о том, как изменится результат, если обучить нейронную сеть на всей выборке. Чтобы исследовать данный вопрос данная нейронная сеть была обучена на всей исходной выборке.

Результаты работы моделей, обученных двумя методами представлены на рисунках 4 и 5.



Рис. 4. Результат работы однослойной нейронной сети, обученной на всей выборке.

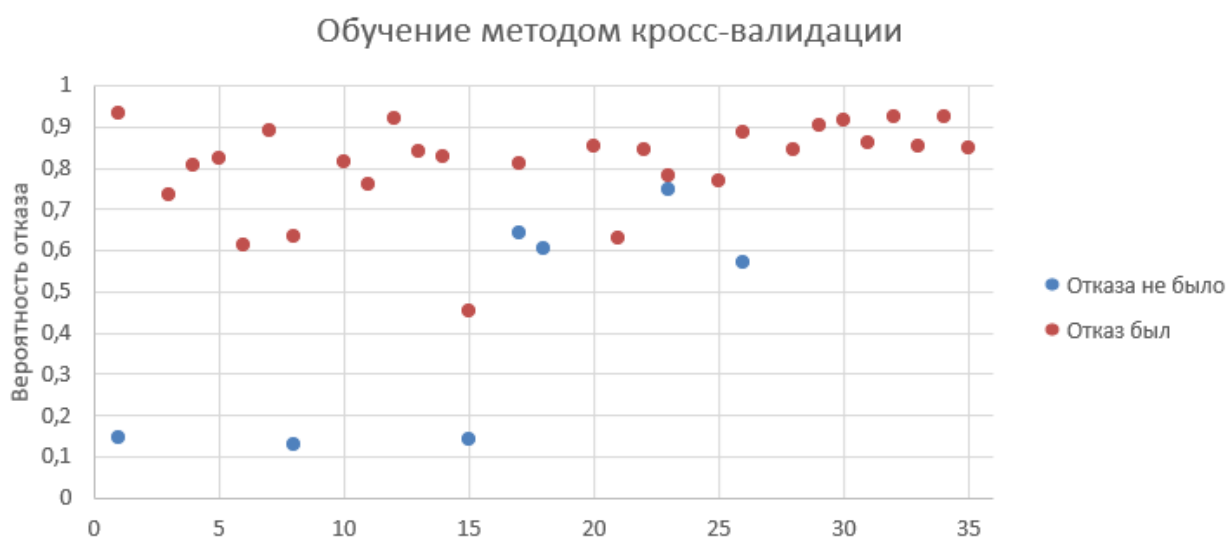


Рис. 5. Результат работы нейронной сети, обученной по методу кросс-валидации.

Как видно из рисунков 4 и 5, в целом модели достаточно неплохо разделяют насосы, в которых случился отказ от установок, которые не выходили из строя. Также можно заметить, что результаты обучения модели на всей выборке выглядят более предпочтительными. То есть условная граница между установками с высокой вероятностью отказа и низкой более заметна в случае обучения модели на всей выборке. Однако стоит понимать, что обучение на всей выборке может быть не совсем корректным решением, так как сложно предсказать, какие результаты даст эта модель при функционировании на новых данных. Исходя из этих соображений, для проектирования нейронных сетей с более сложной структурой было принято решение использовать для обучения модели метод кросс-валидации.

2.3 Проектирование нейронных сетей со скрытыми слоями

Следующей задачей являлось проектирование нейросетей со скрытыми слоями. С помощью добавления различных скрытых слоев можно наблюдать, как меняется точность модели, то есть определить как наличие и структура скрытых слоев влияет на точность модели. Изначально была спроектирована модель с одним скрытым слоем на 64 нейрона. Точность этой модели составила 0.8, что значительно больше оценки качества однослойной модели без скрытых слоев. Добавление еще одного скрытого слоя на 64 нейрона не изменило точность модели. Нейронная сеть с одним скрытым слоем на 32 нейрона также показала идентичную точность. А вот дальнейшее уменьшение количества нейронов в модели с одним скрытым слоем до 16 нейронов привело к снижению точности модели до значения 0.74. Из данных фактов можно заметить следующие тенденции, что добавление скрытого слоя положительно влияет на качество работы модели. С целью оптимизации размера модели и, соответственно, экономии времени на обучение оправдано выбрать модель с одним скрытым слоем. Оптимальным количеством нейронов в скрытом слое является 32 нейрона. То есть получаем, что наилучшей моделью является нейронная сеть с одним скрытым слоем на 32 нейрона. Данные для сравнения моделей приведены в таблице 1.

Таблица 1

Сравнение точности моделей нейронной сети

Модель	Точность
Однослойная без скрытых слоев	0.57
С одним скрытым слоем на 64 нейрона	0.80
С двумя скрытыми слоями на 64 нейрона	0.80
С одним скрытым слоем на 32 нейрона	0.80
С одним скрытым слоем на 16 нейронов	0.74

Обучим наилучшую модель методом кросс-валидации и на всей выборке.

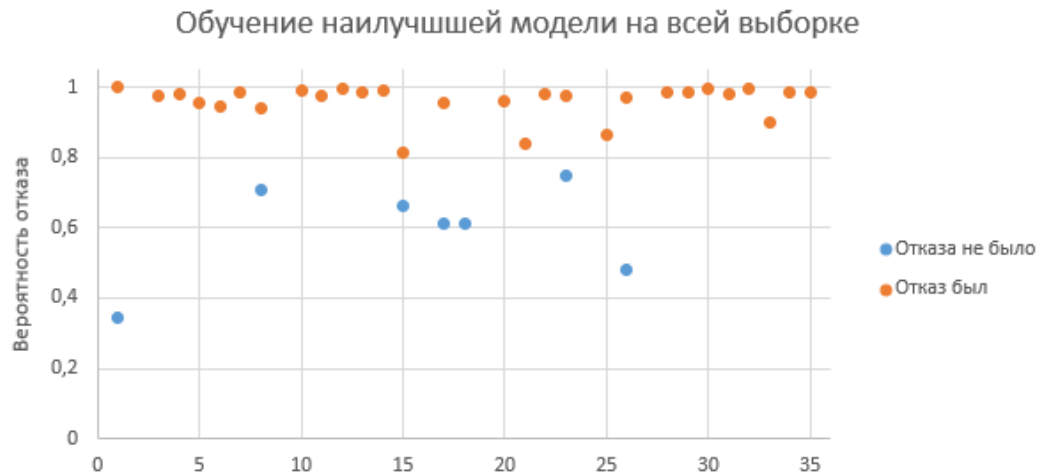


Рис. 6. Результат работы модели с оптимальной топологией, обученной на всей выборке.

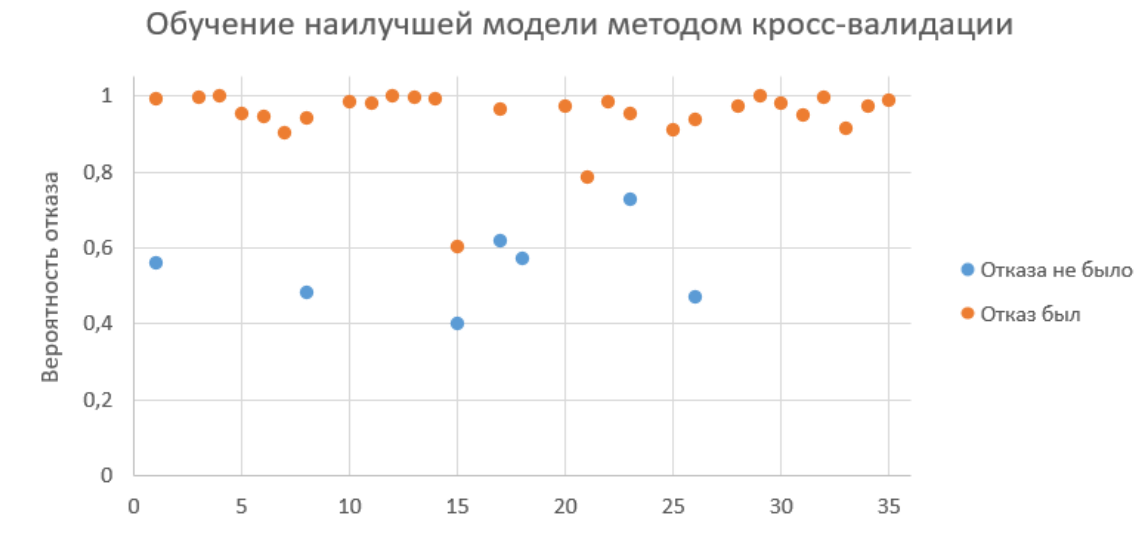


Рис. 7. Результат работы модели с оптимальной топологией, обученной методом кросс-валидации.

Как видно из рисунков 6 и 7, результаты работы моделей с оптимальной архитектурой сохраняют тенденции, замеченные при проектировании однослойной нейронной сети. Обе модели достаточно точно разделяют насосы на две группы. Что касается модели, обученной на всех данных, она позволяет более точно провести условную границу между этими группами. Если сравнивать данные результаты с теми, что были получены при проектировании однослойной нейронной сети, то можно заметить, что диапазон между отказавшими установками и насосами, работавшими без сбоя, стал больше. Данный факт объясняет разницу в точности моделей однослойной нейросети и сети с одним скрытым слоем на 32 нейрона.

Исходя из полученных результатов вероятностей отказа УЭЦН, можно утверждать, что применение нейронных сетей является эффективным способом для решения данной задачи.

3 Динамика вероятностей отказа

3.1 Реализация функций для подсчета вероятностей

Необходимо отметить, что модель нейронной сети на вход принимает не текущие данные в конкретный момент времени. На нейросеть подаются средние, минимальные, максимальные значения, параметры разброса каждой физической характеристики насоса и срок эксплуатации. Это означает, что алгоритм расчета вероятности учитывает динамику показателей насоса, что является очень важным фактором. При введении оператором текущих показателей по определенному насосу, данные каждый раз должны обновляться. Для этого нужно реализовать функцию, которая в качестве входных данных принимает текущие характеристики определенной скважины, записывает эти данные в файл и на выходе выдает параметры с учетом истории данных по этому насосу. Эти параметры уже в свою очередь будут проходить через нейронную сеть.

```
def add_row(full_data, arr):
    """
    Params:
    full_data: pd.DataFrame - raw records data from each pump
    arr: new record to add
    Return:
    full_data: pd.DataFrame - data with new record
    data_for_model: np.array - aggregations for each sensor
    """
    full_data.loc[full_data.shape[0]] = arr
    full_data.to_csv('FULLDATASETADDED.csv', index=None)
    id = arr[-1]
    filtered_data = full_data[full_data['id'] == id]
    filtered_data_gr = filtered_data.groupby(['id'])
    amount = filtered_data['gaz'].count()
    aggregation_cols = ['mean', 'median', 'min', 'max', 'var', 'std']
    agg_dict = {major_col:aggregation_cols for major_col in major_cols}
    agg_values = list(filtered_data_gr.agg(agg_dict).values[0])
    for i in range(4, 84, 6):
        agg_values[i] /= 100
    data_for_model = np.array([amount] + agg_values)
    return full_data, data_for_model
```

Рис. 8. Функция добавления новых данных по текущей установке.

Данные, отображающие текущие показатели насоса и вводимые оператором, обрабатываются с помощью выше реализованной функции. Результатом работы этой функции является получение массива данных, который подается на вход нейронной сети. То есть осталось лишь реализовать функцию, которая в качестве входного параметра будет получать этот массив данных и, используя модель нейронной сети,

выдаст значение вероятности отказа по данной установке. Фрагмент данной функции представлен на рисунке 9.

```
def raw_predict(arr, scaler, model):  
    """  
    Get probability from raw data and model  
    """  
    arr = arr.reshape(1, -1)  
    arr_scaled = scaler.transform(arr)  
    predict = model.predict(arr_scaled)  
    return predict[0][0]
```

Рис. 9. Функция для расчета вероятности отказа.

Теперь можно считать, что алгоритм для расчета вероятности отказа УЭЦН получен. Так как для исследования имеется история данных по некоторым установкам, это позволяет выявить закономерность, как будет изменяться значение вероятности по мере приближения к моменту отказа данного насоса.

3.2 Получение динамики вероятностей отказа

Данные по ежедневным показателям насосов имеются лишь по 8 насосам. Эти данные снимались сотрудниками нефтедобывающей компании ежедневно. Только лишь одна установка из этих восьми в ходе эксплуатации работала безотказно, в остальных насосах случился отказ. Для чтения данных по каждой скважине, снятых ежедневно, также использовалась библиотека Pandas. Данные были считаны и обработаны, что позволило начать проведение анализа динамики вероятностей. В идеальном случае значение вероятности у установки, которая работала без отказа, в ходе эксплуатации должна быть относительно мала. А у установок, в которых произошел отказ, значение вероятности должно стремиться к единице при приближении к моменту отказа.

Проверим сначала единственную установку, которая работала без отказа. Фрагмент кода, с помощью которого получается массив значений вероятности по данной установке представлен на рисунке 10.


```

# pump 524
likelihood = []
_test = pump_524
full_data_ = full_data.copy()
for i in _test:
    full_data_, pm = add_row(full_data_, i)
    likelihood.append(raw_predict(pm, std_scaler_final, model_final))
plt.plot(likelihood)
plt.ylabel('Probability')
plt.xlabel('Day since start')
plt.show()

```

Рис. 10. Фрагмент кода для получения динамики вероятности отказа в ходе эксплуатации.

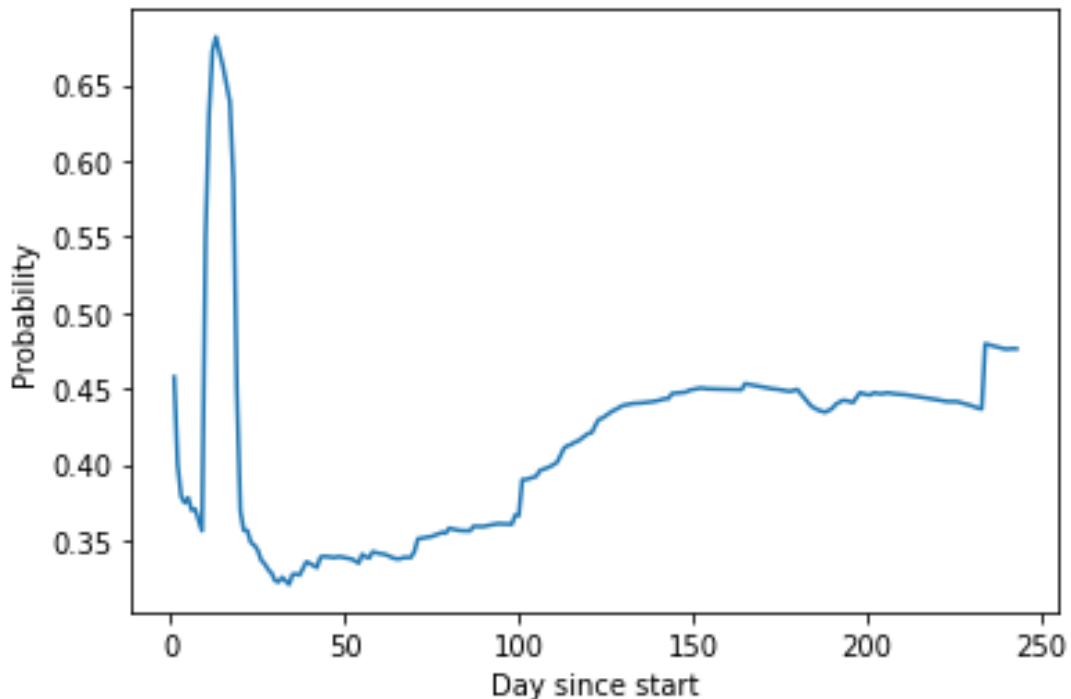


Рис. 11. Динамика вероятностей отказа в ходе эксплуатации установки, работавшей без сбоя.

На рисунке 11 видно, что вероятность отказа данного насоса не превышала значения 0.50 на всём сроке эксплуатации за исключением первых нескольких дней. Резкое возрастание вероятности отказа в первые дни свидетельствует о необходимости совершенствование модели. Однако в целом график достаточно адекватно описывает динамику вероятностей отказа установки, работавшей без сбоя.

Рассмотрим теперь установки, в которых всё же произошёл отказ.

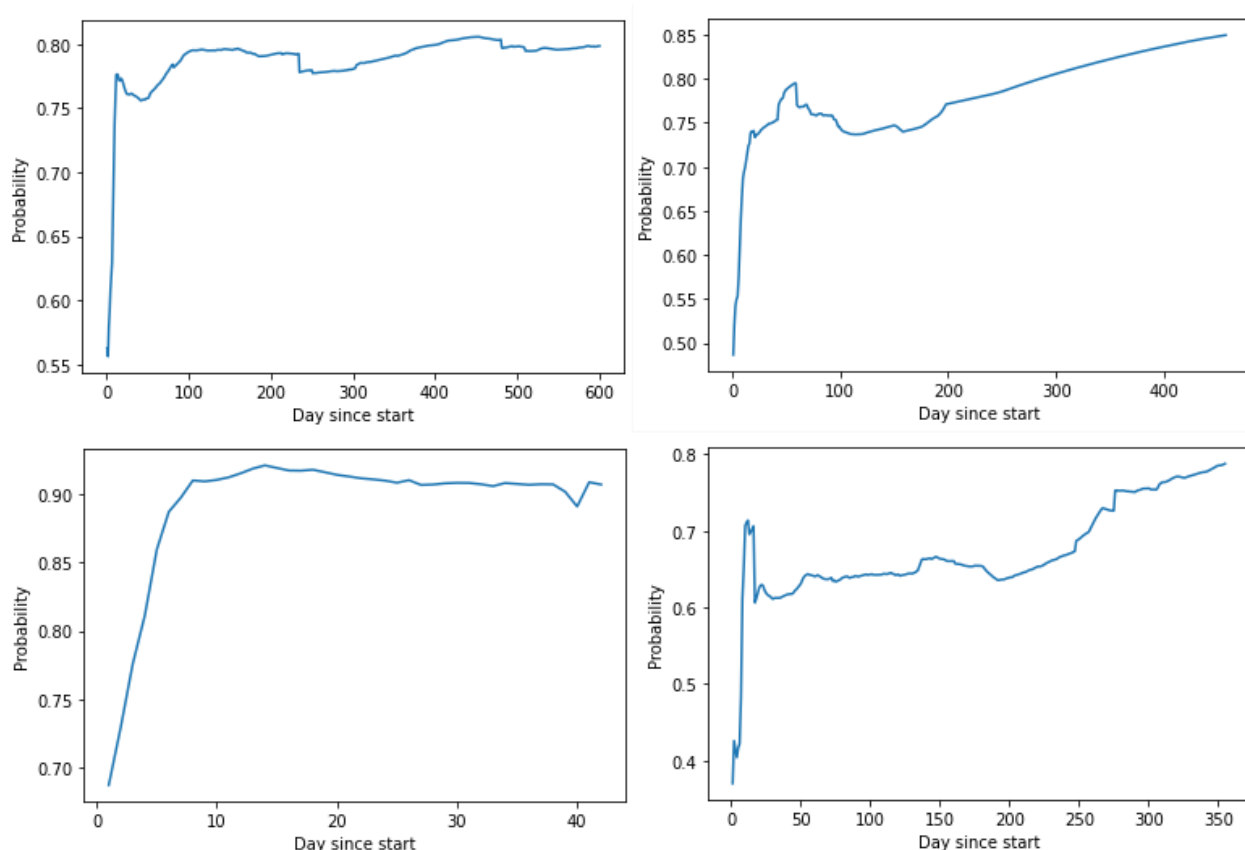


Рис. 12. Динамика вероятностей отказа в ходе эксплуатации установок, отказавших в конечном счете.

По графикам вероятностей отказа данных установок (рисунок 12) можно заметить, что резкое возрастание значения вероятности происходит слишком рано. Данный факт также свидетельствует о том, что модель нейронной сети требует совершенствование. В реальности резкое возрастание вероятностного значения должно происходить непосредственно перед отказом установки.

Несмотря на то, что при сравнении значений вероятности разных установок модель предоставила достаточно точные значения, ее недостаток стал проявляться при исследовании динамики вероятностей отказа по отдельным скважинам. Можно предположить, что данный недостаток объясняется малым количеством данных в исходной выборке. Чтобы проверить данную гипотезу было решено произвести синтез данных на основе значений показателей, собранных ежедневно для некоторых скважин. Суммарно имеется около трех тысяч строк с ежедневными значениями показателей насоса. Это позволяет синтезировать новые данные, которые в дальнейшем также будут использоваться при обучении моделей нейронной сети.

4 Оптимизация модели нейросети

4.1 Синтез данных для модели

Учтем, что в ходе эксплуатации УЭЦН первые 20%, 40%, 60% срока эксплуатации насосы работали без сбоя, то есть факта отказа не было. А в последние 20% срока эксплуатации отказ все же случился. Следовательно, по данным промежуткам эксплуатации можно агрегировать данные, которые будут использоваться как входной сигнал нейронной сети. Ведь имея ежедневные значения параметров, можно без труда высчитать меры центральной тенденции переменных за определенный промежуток эксплуатации.

Фрагмент кода для синтеза данных для нейронной сети представлен на рисунке 13.

```
def between_n(df, n):  
    """  
    Count statistics to input into model based on top n % of the data  
    Params:  
        df : pd.DataFrame with all days for one pump, cols = ['gaz', 'liquid_tm',  
        n : top n % of data to count statistics for model  
    Return:  
        data_for_model : np.array, shape = (1, 85) (cols amount)  
    """  
    start = n - 0.15 if n == 0.95 else n - 0.2  
    shape = df.shape[0]  
    top = round(shape * n)  
    bottom = round(shape * start)  
    top_df = df.iloc[bottom:top, :]  
    top_next = df.iloc[top, :]  
    _, data_for_model = add_row(top_df, top_next)  
    return data_for_model.reshape(1, -1)
```

Рис. 13. Синтез данных для нейронной сети.

4.2 Оптимизация параметров модели нейронной сети

Увеличение качества модели нейронной сети можно добиться также подбором оптимальных параметров при обучении. При подборе параметров обучения нейронной сети нужно понимать, что изменения большинства параметров может приводить как к ухудшению, так и к улучшению модели. От правильного выбора параметров зависит не только то, насколько быстро модель будет сходиться к

оптимальному состоянию. При подборе шага обучения, времени обучения и других параметром основным критерием служила минимизация ошибки.

4.3 Результаты оптимизации модели

Для проверки эффективности оптимизации модели нейронной сети за счет синтеза данных и подбора оптимальных параметров обучения были также получены графики изменения вероятности отказа в ходе эксплуатации.

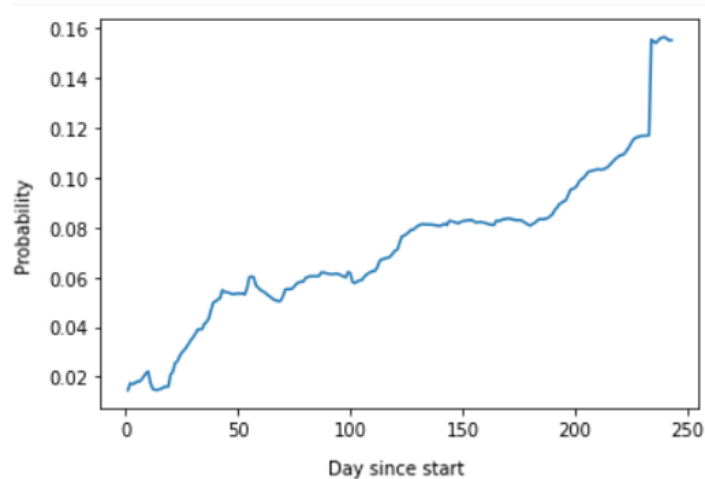


Рис. 14. Полученная после оптимизации модели нейросети динамика вероятностей отказа установки, работавшей без сбоя.

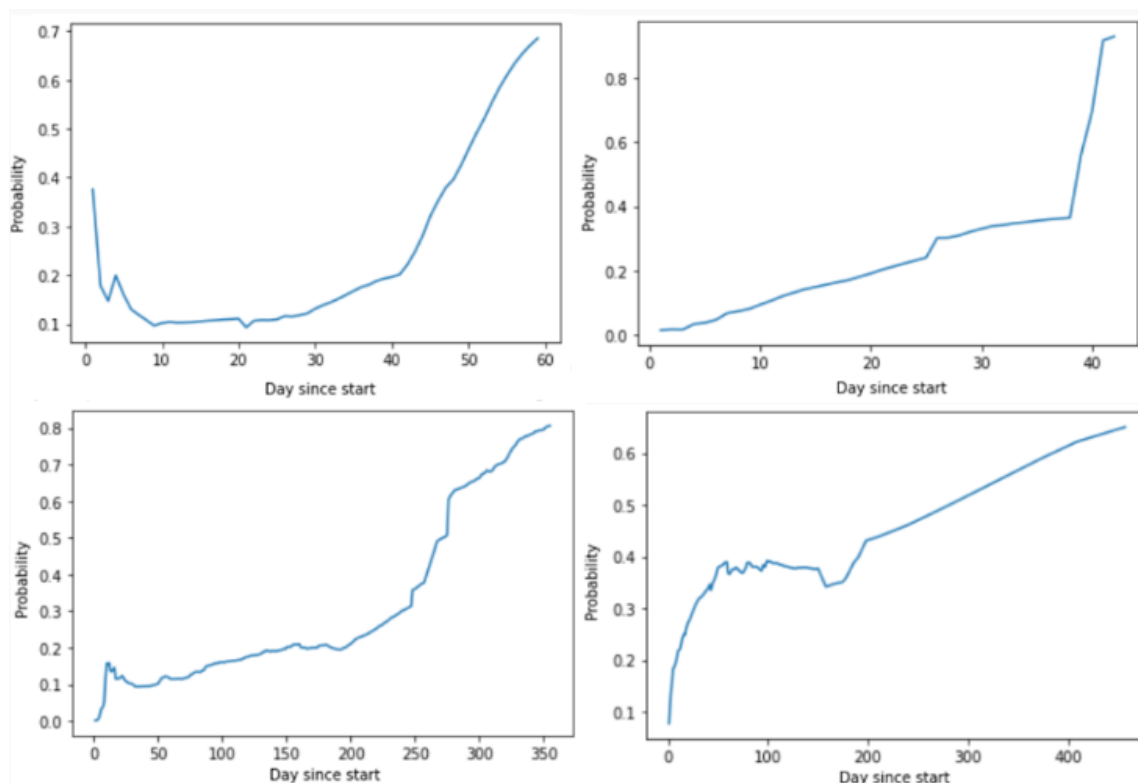


Рис. 15. Полученная после оптимизации модели динамика вероятностей отказа установок, отказавших в конечном счете.

Проанализировав графики, представленные на рисунках 14 и 15, можно сделать вывод, что оптимизация модели путем синтеза данных по мерам центральной тенденции и подбора параметров обучения позволила повысить качество модели. До оптимизации модели резкое возрастание значения вероятности отказа происходила в первые дни эксплуатации насоса, а после оптимизации видно, что стремительное возрастание вероятности происходит ближе к моменту сбоя. Исходя из полученных графиков, можно считать, что данная модель нейронной сети дает достаточно точное представление о вероятности отказа УЭЦН.

5 Нейросеть со значениями ежедневных показателей на входе

5.1 Описание принципа модели

Ранее было проведено исследование того, как изменится качество модели при добавлении в нейронную сеть различных скрытых слоев. Однако во всех этих моделях количество входных нейронов оставалось одинаковым. На вход модели подавались значения из мер центральной тенденции каждого физического параметра установки. Поэтому возникла идея проанализировать то, как изменится качество модели, если на вход нейросети подавать не агрегированные данные по средним, минимальным, максимальным показателям и мерам разброса, а сами значения переменных, снятые сотрудниками нефтедобывающей отрасли ежедневно.

Итак, имеются данные по ежедневному (i) изменению параметров (j) от отдельных скважин. Обозначим их $x_i^{(j)}$.

Основное уравнение:

$$\sum w_j x_i^{(j)} \rightarrow P$$

для каждого i, где P - вероятность отказа.

Если посмотреть на значения $x_i^{(j)}$, то можно видеть, что оператор считает насос отказавшим, если почти все параметры не меняются день ото дня. Это практически всегда последние 10 дней до остановки.

Значит, $\sum w_j x_i^{(j)} \rightarrow 0$ в первые дни, $\sum w_j x_i^{(j)} \rightarrow 1$ в последние 10 дней.

Следовательно, для построения сети функцией результата делаем вероятность отказа, а задача обучения сети - найти весовые коэффициенты w.

5.2 Результаты работы модели

Для проверки работоспособности модели также получим динамику вероятности отказа в ходе эксплуатации насоса.

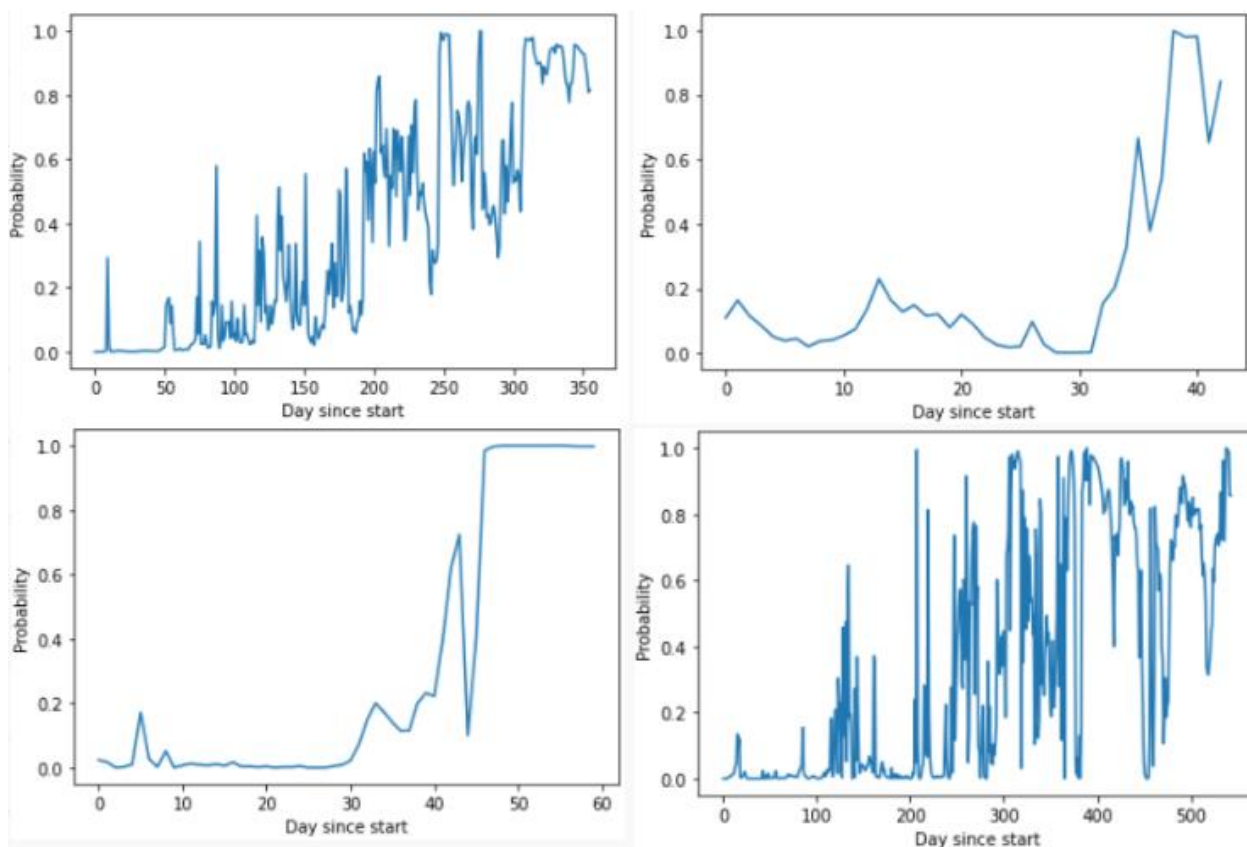


Рис. 16. Полученная на основе новой модели динамика вероятности отказа установок, отказавших в конечном счете.

Результаты данного эксперимента, продемонстрированные на рисунке 16, можно интерпретировать следующим образом: наблюдается общий тренд вверх, то есть вероятность постепенно увеличивается, а колебание объясняется среднеквадратичным отклонением исходных параметров.

6 Графический интерфейс

Для того чтобы операторы нефтедобывающей отрасли имели возможность использовать калькулятор отказов, был разработан графический интерфейс программы, представленный на рисунке 17.

Вероятность отказа УЭЦН

Выбрать данные для обучения модели

Сохранить результаты

№ насоса	488	Обв-ХАЛ	38,8499	Рлин	4,6666
Qгаз тм	3,1684	Нд	669,1428	КВЧ-ХАЛ	5,14285
Qж тм	0,2526	Рзатр	3	I	8,3333
Qж	14,5	Рбуф	2,5714	F	33,3333
Wакт	1,5601	Дшт	12	Qн	2,7614

Получить вероятность

Вероятность отказа составляет 13,34%

Рис. 17. Главное окно программы.

Интерфейс программы был разработан на языке программирования Python в среде разработки PyCharm с помощью библиотеки PyQt5. Данная библиотека позволяет реализовать интерфейс, удовлетворяющий функционалу данного приложения.

Для реализации интерфейса использовались следующие виджеты:

- кнопки (QPushButton);
- формы для ввода (QLineEdit);
- надписи (QLabel);
- файловые окна (QFileDialog);
- диалоговые окна (QMessageBox).

При нажатии на кнопку «Выбрать данные» появляется файловое окно, в котором предлагается выбрать файл формата .csv или .xls с исходными данными для обучения нейросети (рисунок 18).

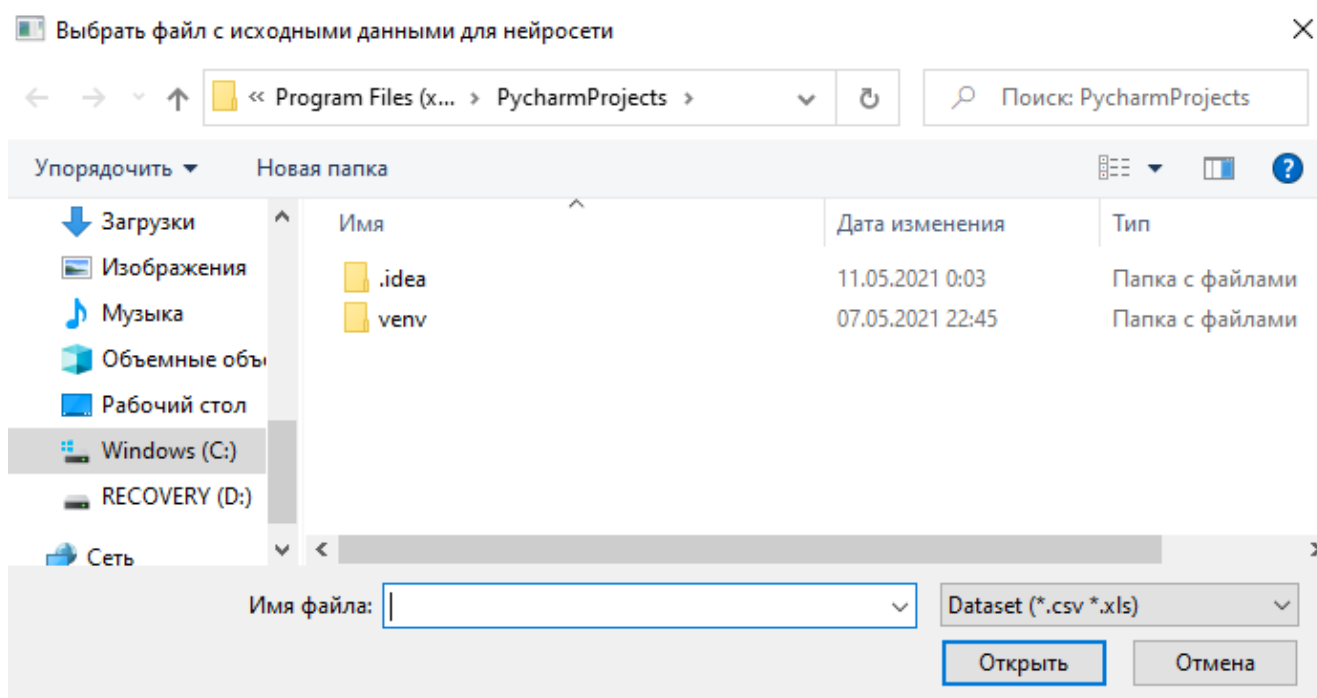


Рис. 18. Файловое окно для выбора файла с исходными данными.

Значения параметров насоса необходимо ввести в соответствующие формы для ввода. Входной параметр «Номер насоса» имеет тип целого числа - `int`. Остальные входные параметры имеют тип чисел с плавающей запятой - `float`. При вводе некорректных данных появляется окно, сообщающее об ошибке в определенном поле (рисунок 19).

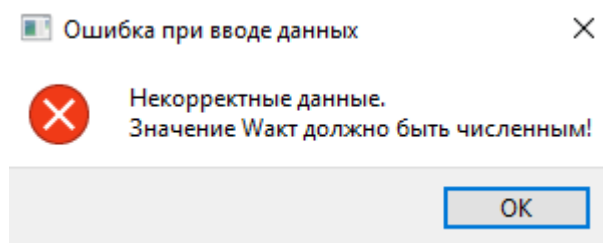


Рис. 19. Окно с сообщением о недопустимом входном параметре.

Кнопка «Получить вероятность» становится активна только после ввода всех параметров. При ее нажатии появляется окно с результатом, то есть со значением вероятности отказа (рисунок 20).

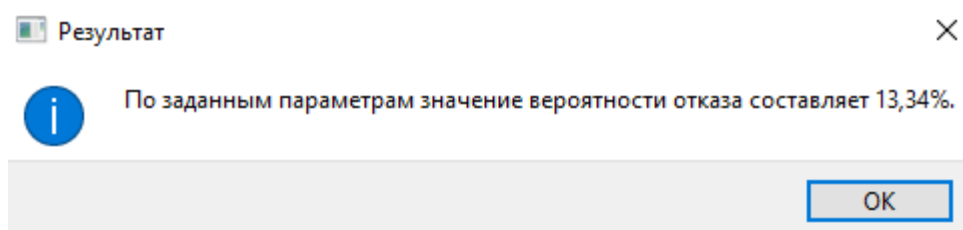


Рис. 20. Окно с результатом работы программы.

Также значение вероятности отображается на главном окне, и кнопка «Сохранить результаты» становится активной. При ее нажатии появляется файловое окно, в котором предлагается выбрать место сохранения и название файла с результатами. В этот файл записываются входные параметры, по которым была посчитана вероятность и само полученное значение вероятности.

Репозиторий проекта с исходным кодом и необходимыми файлами находится по ссылке: https://github.com/daler98/ESP_failure_probability. В нём содержатся файл с исходными данными, по которым обучались нейронные сети, архив файлов с ежедневными показателями для некоторых установок и файлы с кодом на языке Python. Файл `failure_probability.py` содержит код, с помощью которого проводился анализ топологии нейросетей и оптимизация моделей. Для запуска приложения с графическим интерфейсом необходимо распаковать скачанный архив с помощью ZIP-архиватора, и открыть файл `main.exe`, расположенный в директории `dist`.

ЗАКЛЮЧЕНИЕ

По результатам выполнения выпускной квалификационной работы был разработан нейросетевой вероятностный калькулятор отказа установок электроцентробежных насосов. Для выбора оптимальной, с точки зрения точности, топологии были спроектированы различные модели нейронной сети, начиная с однослойной модели и заканчивая нейросетями с несколькими скрытыми слоями. В ходе исследования зависимости точности модели от ее структуры и числа скрытых слоев, оптимальной оказалась модель с одним скрытым слоем на 32 нейрона.

Так как входными данными нейронной сети являются не значения, которые заносятся оператором ежедневно, а сводные данные, учитывающие динамику показателей каждой установки, были разработаны функции, обеспечивающие использование нейронной сети для подсчета вероятности. К ним относятся функция добавления новых данных и записи их в файл, а также функция для подсчета вероятности, на вход которой подается модель нейронной сети.

После реализации калькулятора были проведены различные типы тестирования. Самым важным исследованием, отражающим корректность работы нейронной сети, является наблюдение за динамикой вероятности отказа установки в ходе ее эксплуатации. Данная динамика была получена как для исходной модели, выбранной ранее, так и для оптимизированной на ее основе нейросети. Оптимизация модели происходила за счет добавления в модель при обучении данных, агрегированных на основе ежедневных показателей, и подбора оптимальных параметров обучения. Если сравнить результаты работоспособности модели до и после оптимизации, то необходимо отметить, что точность модели после ее оптимизации стала значительно выше. Это подтверждают графики, демонстрирующие динамику вероятности отказа. Если до оптимизации модели резкий скачок вверх наблюдался в первые дни эксплуатации, то после внесения ранее описанных изменений резкое возрастание вероятности к единице наблюдалось уже ближе к моменту отказа установки.

Также в ходе выполнения работы была спроектирована модель нейронной сети, на вход которой поступают значения физических показателей, снятые ежедневно. Принцип этой модели заключается в том, что первые дни эксплуатации оценивались нулевой вероятностью отказа, а последние десять дней – единичной вероятностью. Данное решение было принято после изучения динамики этих показателей. Полученные результаты работы данной модели значительно отличаются от результатов, полученных ранее. При получении динамики вероятности отказа наблюдаются хаотичные колебания, однако графики имеют общий тренд вверх.

Сравнивая обе полученные модели, необходимо отметить, что каждая из них имеет свои преимущества. Фактически это две разные модели, реализованные на основе различных предположений. Для того чтобы понять, какая из них ближе к реальной практике, необходимо иметь значительно большую статистику по скважинам.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Suvorova A.V. et al. Probabilistic graphical models of individual socially significant behavior on the base of incomplete data // SPIIRAS Proc. SPIIRAS, 2014. Vol. 3, № 22. P. 101.
2. Artamonov Y.S. □ // International Journal of Open Information Technologies. 2017. Vol. 5, № 5. 9–15 p.
3. Маценов А.А. Комитетный бустинг: минимизация числа базовых алгоритмов при простом голосовании // Всероссийская конференция ММРО-13. 2007. Р. 180–183.
4. Tulupyev A., Kharitonov N. Algebraic Bayesian Networks: a Frequentist Approach to Knowledge Pattern Parameters Machine Learning
5. Singh S. et al. SVM based system for classification of microcalcifications in digital mammograms // Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings. 2006. P. 4747–4750.
6. Боровиков С.М., IT-система прогнозирования надёжности сложных электронных систем методом анализа дерева отказов / С.М. Боровиков, А.Е. Епехин // Информационные системы и технологии: управление и безопасность, №2, Т., 2013, с. 140-144.
7. Nowlan F. S., Heap H. F. Reliability-centered Maintenance. — San Francisco: Dolby Access Press, 1978. — 466 p.
8. Баскакова Н.Т., Дорман В.Н. К вопросу об эффективности стратегий управления ремонтами металлургического оборудования // Экономический анализ теория и практика. Общество с ограниченной ответственностью Издательский дом Финансы и кредит, 2016. № 5 (452).
9. Barrios Castellanos M. et al. Fault identification using a chain of decision trees in an electrical submersible pump operating in a liquid-gas flow // J. Pet. Sci. Eng. Elsevier B.V., 2020. Vol. 184. P. 106490.
10. Sakthivel N.R., Sugumaran V., Babudevasenapati S. Vibration based fault diagnosis of monoblock centrifugal pump using decision tree // Expert Syst. Appl. Pergamon,

2010. Vol. 37, № 6. P. 4040–4049.

11. Castellanos M.B., Serpa A.L. Decision tree classifier for parametric fault detection in electrical submersible pumps.
12. Zheng S.-B. et al. Application of support vector machines to sensor fault diagnosis in esp system // Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 04EX826). 2004. Vol. 6. P. 3334–3338.
13. Zhang Y. et al. PCA fault feature extraction in complex electric power systems // Adv. Electr. Comput. Eng. 2010. Vol. 10, № 3. P. 102–107.
14. Атнагулов А.Р. et al. Применение метода байеса при оценке вероятности “полета” погружного насосного агрегата на забой скважины // Сетевое издание «Нефтегазовое дело». 2008. № 1.
15. Oliveira-Santos T. et al. Combining classifiers with decision templates for automatic fault diagnosis of electrical submersible pumps // Integr. Comput. Aided. Eng. IOS Press, 2018. Vol. 25, № 4. P. 381–396.